TRANSCRIPTOMIC AND METATRANSCRIPTOMIC APPROACHES TO CHARACTERIZING GENES CODING FOR FIBER DIGESTION WITHIN THE RUMEN ECOSYSTEM

PAN WANG
Bachelor of Science, Shandong University, 1998
Master of Science, Shandong University, 2001

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the
Requirements for the Degree

DOCTOR OF PHILOSOPHY

Department of Biological Sciences
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

To my children, Eric and Alan

Abstract

The rumen microbiome constitutes a unique genetic resource of plant fiber degrading microbial enzymes that could be used for agricultural and industrial purposes. *Anaeromyces mucronatus* is a poorly characterized anaerobic lignocellulolytic fungus in the rumen. This thesis aimed at better understanding *A. mucronatus* YE505 and the particle associated rumen microbiota based on transcriptomic and metatranscriptomic approaches. High quality RNA was isolated from the fiber-associated rumen sample based on an improved RNA extraction method. A transcriptomic study was performed to investigate the expression of the fiber degrading system of *A. mucronatus* YE505, and the functional diversity of the fiber-associated eukaryotes from the rumen of muskoxen (*Ovibos moschatus*) was explored by a metatranscriptomic study. Much carbohydrate degradation related protein modules were detected. This study established effective approaches to characterizing the functional contents of rumen eukaryotic microbiome as well as rumen fungi, and identified several candidate genes that merit further investigation.

Acknowledgements

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

First and foremost, I would like to express my utmost gratitude to Dr. Tim McAllister and Dr. Brent Selinger for all their guidance and understanding, who not only took a chance on me and give me an incredible opportunity, but were most supportive to this project. Both Dr. McAllister and Dr. Selinger are always there to encourage and inspire me as I overcame all the obstacles in the completion this research work.

I would like to express my sincere acknowledgement to the members of my supervisory committee Dr. Robert Forster, Dr. Tony Russell and Dr. Steven Mosimann for their many suggestions and support over the years, and to my external examiner Dr. Mostafa Elshahed for his helpful comments and discussions.

I would like to thank the members of the McAllister and Selinger lab both past and present at the Department of Biological Sciences, University of Lethbridge, and Lethbridge Research Centre, Agriculture and Agri-Food Canada, for their kindness and support. I would like to thank Dr. Adrian Tsang and many dedicated collaborators at Centre for Structural and Functional Genomics, Concordia University and Genome Quebec Innovation Centre. I would like to thank the collaborators at Department of Biology and Wildlife, Institute of Arctic Biology, University of Alaska. Without their collaboration, the completion of this project would not have been possible.

Finally I would like to thank my husband, children and parents who have accompanied me along the long journey always offering support and love.

Table of contents

List of Tables

List of Figures

List of Abbreviations

AGPC	acid guanidinium-phenol-chloroform
BAC	Bacterial Artificial Chromosome
bp	Base pair
CAZy	Carbohydrate active enzyme database
CAZymes	Carbohydrate-active enzymes
CBM	Carbohydrate binding module
CD	Catalytic domain
cDNA	complementary DNA
CE	Carbohydrate Esterase
CelD_N	cellulase N-terminal immuno-globulin domain
CMC	carboxymethylcellulose
COG	Cluster of Orthologous Groups
Da	Dalton
DGGE	Denaturing gradient gel electrophoresis
ECCF	Extra-cellular cultural fluid
EF1	Elongation factor 1
EPS	Extracellular polymeric substances
EST	Expressed sequence tag
fn3	fibronectin-3
FPKM	fragments per kilobase per million fragments mapped
Gbp	Giga base pair
GCS	Glucose-cellobiose-starch
GH	Glycoside Hydrolase
GI tract	Gastro-intestinal tract
GT	Glycosyl Transferase
HGT	Horizontal gene transfer
ITS	Internal transcribed spacer
KOG	euKaryotic Orthologous Groups
LC-MS/MS	liquid chromatography coupled with tandem mass spectrometry
LRCI	Liquid Ruminal Contents Isolation
LSU	Ribosomal large subunit
Mbp	Million base pair
mRNA	Messenger RNA
MW	Molecular weight
NCBI	National Center for Biotechnology Information
ncRNA	Non-coding RNA
NGS	Next-generation sequencing
nr database	Non-redundant amino acid database

NRMO database	Trimmed down non-redundant muskoxen amino acid database
ORF	Open reading frame
OTU	Operational taxonomic unit
PCR	polymerase chain reaction
PEP	phosphor enol pyruvate
PL	Polysaccharide Lyase
pNPC	*p*-nitrophenyl-β-d-cellobioside
pNPG	*p*-nitrophenyl-β-d-glucoside
qPCR	quantitative PCR
RF	rumen fluid
RIN	RNA integrity number
RNA-Seq	RNA sequencing
rRNA	Ribosomal RNA
RS	Rumen solids
RT-qPCR	quantitative reverse transcriptase-PCR
SLH	S-layer homolog domain
SRCI	Solid Ruminal Contents Isolation
SSU	Ribosomal small subunit
T-RFLP	Terminal restriction fragment length polymorphism
TGS	Third-generation sequencing
tRNA	Transfer RNA
U	Unit
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
VFAs	Volatile fatty acids
α-NA	α-Naphthyl acetate

Chapter 1 Literature Review

1.1 Introduction

Ruminants are unique in terms of their ability to use high fiber feeds. These animals (e.g., cattle, sheep, goats, deer) are able to obtain nutrients through the transformation of complex polysaccharides in plant cell walls into microbial protein and volatile fatty acids (VFAs) to support their own growth (Russell and Rychlik, 2001). However, ruminants themselves do not produce the enzymes required for the degradation of complex plant cell wall polysaccharides. Rather, they have developed a symbiotic relationship with a wide range of anaerobic microorganisms including bacteria, fungi and protozoa that colonize the digestive tract within a few weeks after birth. These microorganisms ferment plant carbohydrates ingested by the ruminant host yielding VFAs, vitamins and microbial protein as end products. VFAs are in turn used by the ruminant as an energy source (Wallace, 1994). As the microbial populations are able to produce vitamins and serve as a source of protein for the host, ruminants can adapt to nutrient-sparse environments. This property has led to their near global distribution, occupying environments ranging from the equator to the high arctic.

1.2 The Rumen

Ruminants are called foregut fermenters. The uniqueness of the ruminant digestive tract in anatomy is the stomach, which is composed of four compartments: reticulum, rumen, omasum and abomasum. Compared to some other monogastric herbivores, which are hindgut fermenters (e.g., rabbits, horses), fermentation in ruminants occurs in the forestomach comprised of the reticulum and rumen. The rumen is the major site of feed

digestion in the digestive tract, accounting for approximately 70% of the total digestive tract volume (Hobson, 1997). Its capacity varies greatly in adult ruminants, ranging from about 10 L in sheep to about 200 L in cattle, and the range is even greater if wild ruminants such as moose and mouse deer are considered. Forestomach fermentation offers several nutritional advantages over hindgut fermentation, as it allows the opportunity for fermentation end products to be digested and absorbed within the host's lower digestive tract. Fermentation products that are not absorpted through the rumen wall as well as undigested feeds flow from the rumen to the omasum, where omasal leaves provide a large surface area for efficient water and mineral absorption. The omasum also acts as a muscular pump, moving the digesta from the reticulorumen to the abomasum. The abomasum functions as a 'true' stomach, producing enzymes and hydrochloric acid which hydrolyses proteins in a manner similar to the monogastric stomach. The high quality microbial protein derived in the reticulo-rumen from low quality recalcitrant plant sources can be digested efficiently in the abomasum; thereby, meeting a large proportion of the protein requirements of the host (Wallace, 1994). Rumination enables repeated mastication of feed, which enhances the ability of the microbial population to ferment it.

1.3 The rumen ecosystem

The rumen maintains a fastidious anaerobic environment, with relatively constant pH and temperature and mixing of microbes with substrate through rumen contractions. It represents the most active fibrolytic fermentation system currently known (Selinger et al., 1996). A diverse population of obligate anaerobic microorganisms exists in the rumen. Bacteria dominate this ecosystem in number and can reach levels as high as 10^{11} cells per

mL. In contrast, protozoa (10^5 to 10^7 per mL), anaerobic fungi (zoospores up to 10^5 per mL) and archaea (0.3 - 3% of the biomass) are less abundant (Li and Heath, 1992). These microorganisms actively interact with each other to form a symbiotic community.

The rumen microbiota is stable, but also dynamic in nature. Temperature in the rumen is held almost constant at about 39 ℃ with the pH typically ranging between 5.2 and 6.8 (Flint, 1997), but it can decline below 5.0 when cattle are fed high-starch diets (Nagaraja and Titgemeyer, 2007). Strictly anaerobic conditions are required for efficient microbial fermentation in the rumen. Trace amounts of oxygen may enter the rumen either with ingested feeds or via diffusion across the rumen wall from the blood stream, but it is quickly consumed by facultative anaerobic bacteria residing on the rumen epithelium.

Overall, rather than geographical location or even species of ruminant, diet has been found to be the main factor to influence the types and numbers of the predominant rumen microbes in adult ruminants (Hobson, 1997; Stewart et al., 1997). The microbial population changes considerably with changes in diet composition, as well as with the quantity and frequency of consumption. For example, nutrient composition, texture of diet and the presence of additives such as plant secondary metabolites and essential oils can affect the distribution of microbial species within the rumen, and overall digestive activity of the ruminal microbiota (McAllister and Newbold, 2008; McGinn et al., 2004).

Microbial species that occupy the rumen may also be isolated from other environments including in the caeca and large intestines of non-ruminant herbivores and omnivores, and the digestive tract of some insects such as termites. Certain species also

exist in the soil microflora and contribute to the anaerobic decomposition of plant debris (Hobson, 1997).

In a normally functioning rumen, proteins and polymeric carbohydrates, which usually make up the largest part of incoming feed, are fermented by the microbiota to VFAs, ammonia, carbon dioxide and hydrogen. The hydrogen is utilized by methanogens to reduce carbon dioxide to methane. The VFAs are absorbed across the rumen wall and serve as major carbon and energy sources for the host. A portion of the VFAs, undigested feed components, and microbial cells pass from the rumen and enter the lower digestive tract where they can also be absorbed or if undigested, excreted in the feces.

1.3.1 Rumen microorganisms

Currently over a thousand microbial species or operational taxonomic units (OTUs) have been identified in the rumen (Hess et al., 2011), and it has been estimated that only a fraction of these (less than 10%) have been cultivated in the laboratory (Flint et al., 2008; Kim et al., 2011). Consequently, the majority of microorganisms in the rumen have been identified strictly based on molecular techniques. Our present knowledge of the microbial community is primarily based on information gained from the culture of only a small portion of the microbial species present in this unique environment.

1.3.1.1 Rumen bacteria

The rumen bacteria account for the largest portion of microbial biomass in the rumen, exhibit the richest species diversity and are responsible for the majority of ruminal feed degradation (Stewart et al., 1997). The majority of ruminal bacteria are Gram-negative, obligate anaerobes. The rumen bacteria are roughly divided into groups

based on digestive activity or preference for feed components and include species with cellulolytic, hemicellulolytic, amylolytic, pectinolytic, proteolytic, ammonia-producing, sugar-utilizing, acid-utilizing and lipid-utilizing activities (Kamra, 2005).

Microorganisms within the rumen have evolved the capability to efficiently utilize plant cell wall fiber through the synergistic activities of the microbial enzymes. Rumen bacteria and their abilities to degrade plant fiber has been studied for over a century (Hungate, 1947). Fibrolytic bacteria such as *Fibrobacter succinogenes*, *Ruminococcus flavefaciens*, and *Ruminococcus albus* have received much attention during the past several decades, owing to their strong capability to utilize plant fiber and the comparative easiness whereby they can be cultured in the laboratory. Many fiber degrading enzymes have been isolated from these bacteria and their catalytic activities characterized in detail (Krause et al., 2003).

1.3.1.2 Rumen fungi

In contrast to bacteria, the anaerobic fungi were relatively recently discovered in the rumen (Orpin, 1975) in the mid-1970's, even though the flagellated zoospore stage of their life cycle was identified as early as in 1910 and mistakenly classified as a flagellated protozoa. Based on mycelium/zoospore morphological characteristics, as well as molecular markers such as internal transcribed spacer (ITS) sequences, the rumen anaerobic fungi have been grouped into six genera, including polycentric fungi (*Anaeromyces*, *Cyllamyces* and *Orpinomyces*) and monocentric fungi (*Caecomyces*, *Neocallimastix* and *Piromyces*) (Kittelmann et al., 2012; Li and Heath, 1992; Liggenstoffer et al., 2010). All species are placed currently in the family of Neocallimatigales, which is the sole family in the newly erected phylum

Neocallimastigomycota (Table 1.1). Rumen fungi have a life cycle and morphology that is typical of the chytridiomycetes and possess chitin in their cell wall (Orpin and Joblin, 1997), but unlike typical chytrids, they are strict anaerobes.

Anaerobic rumen fungi can ferment a variety of plant cell wall polysaccharides to a number of fermentation end products including formate, acetate, lactate, carbon dioxide and hydrogen. The genetic details of the metabolic pathways that they employ in fermentation are largely unknown. However, instead of the mitochondria that are present in the cells of aerobic organisms, rumen fungi possess hydrogenosomes, a membrane bound mitochondria-like organelle that generates ATP and hydrogen (Akhmanova et al., 1999; Boxma et al., 2004). The hydrogen generated within the hydrogensomes supports the formation of a symbiotic relationship between anaerobic fungi and methanogens (Boxma et al., 2005).

When ruminants are fed fiber rich forage diets, rumen fungi account for about 8 - 20% of the microbial biomass within the rumen (Orpin and Joblin, 1997; Rezaeian et al., 2004). The fungal mycelia penetrate plant tissue as a result of their filamentous growth. Consequently, anaerobic fungi in the rumen as well as in other regions of the gastro-intestinal tract (GI tract) of herbivores are believed to play an active role in fiber degradation (Krause et al., 2003). Recently this group has attracted more research attention due to their distinct characteristics and potential to serve as a source of active fibrolytic enzymes for commercial purposes (Wang et al., 2011). Unlike rumen bacteria, they have the unique capacity to penetrate the cuticular surface and the lignified tissues of plant cell walls, and digest the fiber found in recalcitrant forages such as cereal straw (Orpin and Joblin, 1997). Fungi have been shown to produce a wide range of highly

active plant cell-wall degrading enzymes and are most abundant in the rumen of ruminants fed reclacitrant high-fiber diets (Orpin and Joblin, 1997). Genomic information on rumen fungi is still very limited, mainly due to difficulties in analyzing the extreme AT rich (80 – 85% mol%) coding and non-coding regions of their genome (Brownlee, 1989; Chen et al., 2006; Nicholson et al., 2005).

Some researchers believe the rumen fungi play a pivotal role as the initial/primary colonizers of plant fiber in the rumen (Joblin et al., 2002; Orpin and Joblin, 1997), while others consider their role in the process to be negligible owing to their low population density (Orpin and Joblin, 1997; Tuckwell et al., 2005). Increases in fungal biomass in the rumen of hosts fed poor quality high-fiber diets suggest that fungi may play a prominent role in feed digestion under these conditions (Orpin and Joblin, 1997).

1.3.1.3 Rumen protozoa

Based on cell counts, the amount of protozoa in the rumen is relatively low (10^5 to 10^7 per ml), but due to their large size these unicellular eukaryotes can account for up to 40% of rumen microbial biomass (Flint, 1997). Protozoa were detected in domestic ruminants as early as the nineteenth century (Kamra, 2005). Over 100 species of rumen protozoa have been identified, representing over 25 genera. Based on their morphology, ciliate protozoa have been classified into two groups, i.e., holotrichs and entodiniomorphids. Among the holotrichs, *Isotricha, Dasytricha, Buetschlia* and *Charonina* are widely distributed in the rumen and GI tract of non-rumen herbivores (Williams and Coleman, 1997).

The ciliate protozoa are generally considered as predators within the rumen ecosystem, preying on bacterial cells and fungal zoospores; thereby contributing to

nitrogen recycling in the rumen. Some are able to digest starch, pectin, soluble sugars and other plant particles as energy sources (Williams and Coleman, 1997). Enzymes responsible for cellulose and hemicellulose degradation have also been reported and it has been estimated that protozoa may account for up to 30% of ruminal fiber digestion (Russell and Rychlik, 2001). Many protozoa engulf and store starch granules, thereby modulating the rate of starch fermentation in the rumen (Russell and Rychlik, 2001).

Although protozoa float freely in the rumen fluid, large numbers may also attach to the surface of feed particles. Because they are predators of rumen bacteria, the number of protozoa in the rumen fluctuates inversely with the number of bacteria. In fact, ruminants can survive periods of complete defaunation where no protozoa are detected in the rumen, therefore, unlike bacteria, protozoa are not absolutely essential for rumen fermentation. Considerable effort has been devoted to the development of technologies to eliminate or alter the species composition of the ruminal protozoal population (Firkins et al., 2007). This defaunation process was reported to increase bacterial density, stimulate starch degradation and propionate production, and decrease methanogenesis, but a reduction in fiber digestion has also been reported (Morgavi et al., 2010; Mosoni et al., 2011). At present no defaunation technologies are routinely employed in commercial livestock production (McAllister and Newbold, 2008).

1.3.1.4 Other rumen microorganisms

In addition to the previously described major microbial groups, other organisms exist in the rumen, including methanogens, mycoplasmas and bacteriophages.

Bacteriophages are observed in the rumen in concentrations of 10^{10} per mL of ruminal fluid. Diurnal fluctuations in bacteriophage numbers in ruminal fluid have been

observed, and likely reflect changes in the number of available host bacterial cells, which is influenced by the feeding cycle of the host animal. Bacteriophages have been found in association with cellulolytic, amylolytic, methanogenic and acetogenic rumen bacteria (Klieve et al., 2004). Infection of rumen bacteria by lytic phage may account in part for the high levels of bacterial cell lysis, frequently reported in studies using ruminal fluid (Wells and Russell, 1996). In the future, bacteriophage therapy administered by inoculating the rumen with phage targeted against undesirable bacterial species (e.g., *Streptococcus bovis* or *Escherichia coli* O157:H7) may serve as a means of preventing some digestive diseases or pathogen transmission (Herrera et al., 2009; Rivas et al., 2010).

Although anaerobic mycoplasmas were identified in the rumen about 40 years ago, they are the least studied of the rumen microorganisms (Stewart et al., 1997). Mycoplasmas are commonly co-isolated with protozoa, fungi and methanogens, likely because they lack a cell wall and are insensitive to the antibiotics frequently used in the isolation of these organisms of interest.

Methanogens belong to the domain *Archaea*. Methanogens provide anaerobic ecosystems with a route of hydrogen disposal, enabling reduced cofactors such as NADH to be reoxidized. Thus, they play a critical metabolic role in the recycling of reducing equivalents, enabling rumen microbes to derive energy from the fermentation of carbohydrates, proteins and lipids (Whitford et al., 2001). The greenhouse gas – methane is formed when hydrogen is used to reduce carbon dioxide. This is the case for most species of methanogens, but other substrates including formate, acetate, methylamine and methanol, can also be used as a substrates by some species of methanogens (Stewart et al., 1997).

To date, over 100 species of methanogens have been identified, but only seven of them have been isolated and cultivated from the rumen (Joblin, 2005). Methanogens are estimated to comprise approximately 0.3 – 3% of the rumen microbial biomass (Janssen and Kirs, 2008). Although they make up only a small portion of the rumen microbial biomass, methogens play a crucial role in rumen function and animal nutrition. Efficient H_2 removal eliminates the inhibitory effect of hydrogen accumulation on microbial fermentation and leads to a more favourable pattern of VFA formation nutritionally and to an increased rate of fermentation (McAllister and Newbold, 2008).

The population densities of methanogens in the rumen appear to be influenced by diet, with emissions per unit of feed digested increasing when ruminants are fed high fiber diets (McAllister and Newbold, 2008). The possibility of negative consequences of climate change has led a major international effort by researchers to explore strategies to lessen ruminal methane emissions (McAllister and Newbold, 2008; Morgavi et al., 2010).

1.3.2 Microbe-microbe Interactions

The rumen microbiota is not a random mixture of hundreds of species of microorganisms; rather it is a structured and dynamic ecosystem. Over 80% of rumen microbial cells are attached to solid feed particles and thus establish and function in the form of biofilms (Cheng and McAllister, 1997; Costerton et al., 1987). The members of the rumen microbial community interact extensively. Examples of both synergistic and antagonistic relationships between rumen microorganism have been observed (Orpin and Joblin, 1997; Stewart et al., 1997; Williams and Coleman, 1997). The examples include both synergistic and antagonistic relationships among bacterial species, predation of

ruminal bacteria and fungi by ciliate protozoa and initial plant cell wall invasion by fungal hyphae providing bacteria with access to the interior of plant cells.

Synergism is more than general protocooperation, and there are several examples of this most common relationship identified in the rumen microbiota. Not only do both synergists benefit from collaboration, the resultant substrate consumption or product formation is substantially higher than the sum of activity of the individuals (Nikolaev and Plakunov, 2007). For example, although facultative anaerobic microorganisms only exist in the rumen ecosystem in low abundance, their consumption of oxygen facilitates the growth of their strict anaerobic 'roommates' (Wolin et al., 1997).

Cellulolytic microorgamisms establish synergistic relationships with non-cellulolytic species, an interaction that accelerates the rate of cellulose degradation. The specific adhesion of bacteria to plant fibers is the essential first step in plant cell wall digestion, but optimal rates of cellulose digestion are not achieved unless the organisms are combined with coworkers such as bacteria or fungi (Costerton et al., 1987). As soluble nutrients arising from cellulose digestion accumulate within the biofilm, they become available to the cellulolytic bacteria themselves as well as to heterotrophic secondary colonizers which are attracted by chemotaxis and stimulated to divide and to form structured consortia. For example, cells of *Treponema bryantii* and *Butyrivibrio fibrisolvens* are commonly found in association with adherent cellulolytic bacteria, especially *F. succinogenes* (Costerton, 2007; Costerton et al., 1987; Kudo et al., 1987). While they have no cellulolytic enzymes, *T. bryantii* exhibits a chemotactic response towards butyrate, a fermentation end product generated by cellulolytic bacteria that also inhibits cellulolytic enzymes if it accumulates. Thus by utilizing butyrate for their own

growth, *T. bryantii* enhances the activity of the cellulolytic bacteria by preventing end product accumulation (Costerton, 2007; Kudo et al., 1987).

As stated previously, methanogenic archaea establish synergistic relationships with cellulolytic bacteria, anaerobic fungi and even protozoa, by utilizing hydrogen and formate produced in the course of cellulose fermentation. This prevents the accumulation of reduced coenzyme NADH and stimulates ATP synthesis within the microbial community (McAllister et al., 1994; Nikolaev and Plakunov, 2007).

Competition between various species of ruminal bacteria is also common. For example, *R. albus* 7 produces a bacteriocin with activity against *R. flavefaciens* FD-1 (Chen et al., 2004; Nikolaev and Plakunov, 2007; Odenyo et al., 1994), most probably because these two species compete for the same nutritional source as they are both cellulose degraders and occupy the same niche.

The specific order in which various species colonize the digesta surface is thought to influence the spatial organisation of the rumen microbiota (McAllister et al., 1994), although the initial rate of attachment of rumen bacteria to forage is thought to be similar among colonizing species (Edwards et al., 2007; Edwards et al., 2008a). Fungal spores, on the contrary, are thought to attach to the forage slower than bacteria (Edwards et al., 2008b). But fungal zoospores are able to colonize the lignified tissues preferentially, and the vegetative thalli are better at penetrating plant tissue than are bacteria and protozoa, and provide new attachment sites for the latter groups (Nagpal et al., 2009). Current results infer that utilization of nutrients by primary and secondary colonizers promotes further proliferation and stimulates subsequent development and maturation of the biofilm into a structured consortium (Edwards et al., 2008a).

Organisation of ruminal microorganisms into biofilms has several advantages. Firstly, the self-produced extracellular polymeric substances (EPS) that coat biofilm communities trap nutrients and concentrate enzymes adjacent to their targeted substrate (McAllister et al., 1994). Secondly, competing microbes are excluded from the digestion site. Formation of EPS protects the cellulolytic enzymes on the cell surface from degradation by ruminal proteases (Miron et al., 2001). Thirdly, mature 'stable' multi-species biofilms are retained in the rumen as much as three times longer than planktonic cells, and are resistant to detachment (Edwards et al., 2008a; McAllister et al., 1994), thus this arrangement increases their opportunity to thoroughly digest plant fibers (Miron et al., 2001). Fourthly, adherent microbes are protected from a range of antimicrobials including antibodies, antibiotics and bacteriophages (Costerton et al., 1987). It has also been proposed that biofilms offer protection from predation (Costerton et al., 1987), although many protozoa attach to feed particles (Williams and Coleman, 1997) and some researchers have argued that they have mechanisms to predate attached bacterial communities (Edwards et al., 2008a). Moreover, horizontal gene transfer (HGT) has been documented among rumen microbial members (Keeling and Palmer, 2008; Ricard et al., 2006), which may also be facilitated by the high density of microorganisms associated with biofilms (Flemming and Wingender, 2010).

1.4 Plant cell wall degradation by rumen microorganisms

In nature, the hydrolysis of plant cell wall fiber is carried out by fiber-degrading microorganisms, which include both aerobic and anaerobic fungi and bacteria present in soil and the guts of animals. These microorganisms synthesize a complex collection of

cellulases, hemicellulases and ligninases. The microbial consortium in the rumen is unique and amongst the most active fiber degrading system known (Selinger et al., 1996).

1.4.1 Structure of plant cell walls

The plant cell wall is composed primarily of a group of polymers known as lignocellulose, which comprises about half of the plant biomass and stores a large portion of the solar energy captured through photosynthesis (Sánchez, 2009). It represents the most abundant renewable organic resource on earth. Lignocellulose consists of three major components: cellulose, hemicellulose and lignin, which are strongly intermeshed and chemically bound by non-covalent forces and covalent linkages (Figure 1.1) (Kumar et al., 2008). Cellulose and hemicellulose are macromolecules constructed from different sugar residues; whereas lignin is composed of various polyphenolics.

Cellulose is a major constituent of plants and is a linear biopolymer of D-glucose subunits linked through β – 1, 4 glycosidic bonds. The elemental fibrils are linked together by hydrogen bond and van der Waals forces (Sánchez, 2009). Depending on the degree of hydrogen bonding within and between cellulose molecules, this polysaccharide is found in crystalline or paracrystalline (amorphous) forms. In the latter conformation, cellulose is more susceptible to enzymatic degradation (Krause et al., 2003). In nature, cellulose is associated with other plant polymers, primarily hemicellulose and lignin, and this association may affect its biodegradation (Lynd et al., 2002). Cellulose-hydrolyzing enzymes (i.e. cellulases) are divided into three major groups: cellobiohydrolase (exoglucanase), endoglucanase, and β-glucosidase (Figure 1.2) (Lynd et al., 2002).

Hemicellulose, the second most abundant component of lignocellulosic biomass, is a group of branched heterogeneous polysaccharide composed of pentoses (D-xylose, D-arabinose), hexoses (D-mannose, D-glucose, D-galactose) and sugar acids (4-*O*-methyl-glucuronic, D-galacturonic and D-glucuronic acid) and typically has a lower molecular weight than cellulose (Sánchez, 2009). The subunits are generally linked together by β – 1, 4 –, and sometimes β – 1, 3 – glycosidic bonds. Hemicelluloses link cellulose fibers into microfibrils and cross-link with lignins, creating a complex network that provides structural strength.

Xylan is a common hemicellulose, and is composed of β – 1, 4 glycosidic bond linked D-xylose backbone. The xylose residues can be substituted with acetic acid at the C2 and/or C3 positions, 4-*O*-methylglucuronic acid at the C2 position, and arabinose at the C3 position. The arabinose may be further esterified by phenolic acids, which crosslink xylan and lignin within the matrix (Christov and Prior, 1993). Other common forms of hemicelluloses include mannan and glucomannan, xyloglucan, and β-glucan, based on the composition of the backbone sugar residues (Scheller and Ulvskov, 2010). The backbone of mannan consists entirely of mannose residues, and that of glucomannan is formed by D-glucose and D-mannose residues. Both xyloglucan and β-glucan have a backbone composed of D-glucose residues. In the former, most of the glucose residues are substituted with α – 1, 6 – linked xylose residues, but in latter, the backbone is linked through either β – 1, 4 or β – 1, 3, and in some cases β – 1, 6 glycosidic bonds. Similar to xylan, the backbones of these polymers are commonly acetylated or substituted by sugar/sugar acid residues. The type and degree of substitution depends on the plant species and tissues to which it is composed.

Due to their heterogeneity, the degradation of hemicelluloses involves many kinds of glycoside hydrolases and carbohydrate esterases (Dashtban et al., 2009; Sánchez, 2009). Major enzymes involved in xylan degradation include xylanase, glucuronidase, arabinofuranosidase, acetylxylan esterase and ferulic acid esterase (Figure 1.2). Enzymes such as endomannase, galactosidase, β – mannosidase and β – glucosidase are also involved in hemicellulose degradation.

Lignin is typically a complex polyphenolic 3-dimensional framework containing thousands of phenolic units (Dashtban et al., 2009). It is an amorphous heteropolymer, insoluble in water and optically inactive. It is formed from phenylpropane units joined together by non-hydrolyzable linkages including C-C and aryl-ether linkages (Sánchez, 2009). Oxidation of lignin is catalyzed by ligninases including lignin peroxidase, manganese peroxidase, versatile peroxidase and laccase, and their activities require oxygen. The basidiomycetes "white-rot" fungi are currently the only known efficient lignin degraders (Martínez et al., 2009a). The anaerobic microorganisms in the rumen are not capable of degrading lignin (Weimer et al., 2009).

1.4.2 Carbohydrate active enzymes

As shown in Figure 1.2, many different kinds of enzymes are involved in lignocellulose degradation. Combined, enzymes involved in plant cell wall carbohydrate digestion are called carbohydrate active enzymes (CAZymes). A specialized database, CAZy – the Carbohydrate-Active enZYmes Database, is dedicated to the display and analysis of genomic, structural and biochemical information on carbohydrate active enzymes including catalytic domains (CDs) and carbohydrate-binding modules (CBMs) that degrade, modify, or create glycosidic bonds (Cantarel et al., 2009). This

classification system was first introduced in late 1980s (Henrissat et al., 1989), and now it has become the gold standard for the classification of these kinds of enzymes (Cantarel et al., 2009). This complete classification system groups enzymes into families based on primary structure comparisons of their catalytic domains (Collins et al., 2005). The classification groups continue to grow as new CAZy sequences are identified. As the structure and molecular mechanisms of an enzyme are related to its primary sequence, the CAZy system reflects both structural and mechanistic features. Enzymes within a particular family have a similar three-dimensional structure and similar catalytic mechanism; however, members within one classification family may be very diverse in their substrate specificity.

Currently CAZymes are classified into the following major classes: glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and carbohydrate-binding modules (CBMs). Glycoside hydrolases hydrolyze and/or rearrange glycosidic bonds. This is the largest CAZyme group, consisting of over 100 different families, depending on their sequence and enzyme characteristics. It contains many important enzymes involved in polysaccharide degradation, such as cellulases, xylanases and many other sugar hydrolases. Glycosyl transferases are the enzymes which form glycosidic bonds between sugar residues. Many enzymes involved in polysaccharide synthesis are assigned to this class. Polysaccharide lyases cleave glycosidic bonds through non-hydrolytic mechanisms. Carbohydrate esterases hydrolyze carbohydrate esters, and are actively involved in modifying and removing sugar residues such as during the debranching of xylan molecules. Carbohydrate-binding modules are not catalytic modules, but contiguous amino acid

sequences that exhibit carbohydrate-binding activity. Usually CBMs are found within CAZymes and usually function as the substrate binding domain. Some exceptions are CBMs in cellulosomal scaffoldin proteins and rare instances where putative independent CBMs have been reported (http://www.cazy.org).

1.4.2.1 Enzymes involved in cellulose degradation

Cellulases are very diverse in their sequences, structures and mechanisms, even though cellulose is a structurally simple homopolymer of glucose. All cellulases use one of two mechanisms to degrade cellulose: hydrolysis with retention or hydrolysis with inversion mechanism (Wilson, 2008). The former maintains the stereochemistry of the anomeric hydroxyl group of the sugar subunit, in this case, the glucose residue; while the latter inverts the stereochemistry of the anomeric hydroxyl group. There are two functionally different types of cellulases: endoglucanases (also called endocellulases) and cellobiohydrolases (also called exocellulases). A third enzyme, β-glucosidase, which cleaves di- and oligosaccharides, the products of cellulases, is required to completely hydrolyze cellulose to glucose. Generally speaking, exocellulases are processive enzymes that remain attached to the cellulose chain until it is completely hydrolyzed, whereas endocellulases can be either processive or non-processive (Kurasin and Valjamae, 2011; Sukharnikov et al., 2011).

Exocellulases sequentially cleave di- and oligosaccharides (usually 2 – 4 residues) from the end of the cellulose chain and accordingly, their active sites are in the shape of a tunnel (Sukharnikov et al., 2011). There are two classes of exocellulases: the first attacks the non-reducing end whereas the second attacks the reducing end of cellulose (Lynd et al., 2002). All known fungal and the majority of bacterial exocellulases that are active on

the non-reducing end of the chain are classified in family GH6, but several from bacteria are classified in GH5 and GH9. All exocellulases from aerobic fungi with activity against the reducing end are classified in family GH7, while the bacterial members are in family GH48. It is interesting that the anaerobic fungal members of this class fit in family GH48 instead of GH7 (Wilson, 2008). Actually, recent studies have shown that there is a strong similarity between the CAZymes of the rumen anaerobic fungi and those from rumen bacteria, rather than those from aerobic fungi, suggesting that anaerobic fungi exchanged genes coding for these enzymes through horizontal gene transfer with bacteria (Garcia-Vallve et al., 2000).

On the contrary, endoglucanases randomly bind to the interior of long cellulose chains and cleave the glycosidic bond between sugar residues. They belong to GH families 5, 6, 7, 9 and to over a dozen other families. All known structures of endocellulase CDs have an open active site, a prequisite for their ability to bind the interior region of the cellulose chain (Wilson, 2008).

Processive endocellulases have interesting properties and were first discovered in late 1990s (Irwin et al., 1998; Reverbel-Leroy et al., 1997; Sakon et al., 1997). These enzymes initially bind to an interior point along a cellulose molecule, but instead of releasing from the cellulose fiber after the first cleavage, the non-reducing end of the cellulose chain is shifted to the enzyme's empty -4 to -1 subsites, enabling processive cleaveage of cellotetraose from the non-reducing end of the cellulose chain (Wilson, 2012). These enzymes are currently assigned to either GH9 or GH48 families. Because most anaerobic cellulolytic bacteria do not produce GH6 exocellulases (those working on the non-reducing end), it is believed that the processive endocellulase appears to replace

their activity and play an important role in enabling anaerobic bacteria to degrade cellulose (Wilson, 2012).

Cellulases often act synergistically on crystalline cellulose, with the specific activity of some cellulase mixtures containing four to six enzymes showing an activity that is over 10 times higher than that of any single cellulase in the mixture (Lynd et al., 2002). It seems that synergism only occurs when two cellulases attack different regions of the cellulose molecule, with each cellulase creating new attack sites for other enzymes within the mixture (Sánchez, 2009).

1.4.2.2 Enzymes involved in xylan degradation

With many kinds of side chain modifications, xylan contains a variety of chemical linkages, and thus its degradation requires a number of different enzymatic activities. The breakage of xylan backbone β – 1,4 – xylan linkage requires only a single enzyme: xylanase (aka endo – 1,4 – β – xylanase). Like endoglucanases, endoxylanases randomly cleave the glycosidic bond at the interior of xylan molecules, by either a retaining or inverting mechanism in terms of the anomeric configuration of the reactant xylose residue (Collins et al., 2005). The major xylanase families are GH10 and GH11, while GH5, 7, 8, 43 and a few other families also possess some members that exhibit this activity. Families GH5, 7, 10 and 11 contain enzymes which carry out hydrolysis with a retaining mechanism (Collins et al., 2005). In contrast, enzymes in families GH8 and 43 typically utilize an inverting mechanism (Collins et al., 2005). In general, GH10 xylanases have broader substrate specificity than those of GH11. Specifically, GH10 enzymes not only degrade linear chains of β – 1,4 – linked xylose residues, but also xylan backbones that are highly substituted as well as smaller xylo-oligosaccharides (van den

Brink and de Vries, 2011). The xylo-oligosaccharides released by endoxylanases are further degraded by β-xylosidases, mostly belonging to the GH3 or GH43 families (van den Brink and de Vries, 2011).

To completely degrade xylan, all substitutions on the backbones have to be released. This requires several different enzymes divided over many GH and CE families. Acetylxylan and ferulic acid esterases are two major kinds of carbohydrate esterases (Dashtban et al., 2009). Acetylxylan esterases release acetyl residues from xylan chains. They are distributed into at least eight CE families, including CE1, 4, 5, and 16 (Biely, 2012). The presence of acetylxylan esterases is essential for efficient degradation of the xylan backbone by endoxylanases (van den Brink and de Vries, 2011). The major difference between the CE families lies in the degree to which they hydrolyze different *O*-linked acetyl groups. Families CE1, 4, and 5 have a strong preference for side chains linked to hydroxyl group at C-2 position (2-*O*- linked) of the xylose residue, which is the most common linkage in hemicellulose, while CE16 prefers 3-*O*- and 4-*O*- linked residues (Biely, 2012; Li et al., 2008).

Ferulic acid esterases remove *p*-coumaric acid and ferulic acid, the two cinnamic acids present in xylan. Some of these esterases belong to CE1, while a considerable number have yet to be assigned to a CE family. Ferulic acid esterases have been divided into five types based on substrate specificity, the nature of the product released and the degree of similarity in amino acid sequences (Qi et al., 2011). Some particular groups of esterases show preference for substrates with methoxy substituents such as ferulic acid, while others prefer substrates containing one or two hydroxyl substitutions, such as *p*-coumaric acid (van den Brink and de Vries, 2011).

1.4.3 Cellulose biodegradation by microorganisms

Fungi are predominantly responsible for lignocellulose degradation in the environment with the most rapid degraders belonging to the basidiomycetes (Sánchez, 2009). The aerobic fungus *Hypocrea jecorina*, originally called *Trichoderma reesei*, is the most studied aerobic cellulolytic microorganism (Wilson, 2008).

Since many microorganisms are unable to transport insoluble materials across the cell membrane, the enzymatic degradation of cellulose and hemicellulose needs to occur extracellularly. The produced soluble sugars are then transported inside the cell for further metabolism. Currently, two strategies for plant cell wall digestion by lignocellulolytic microorganisms have been described: (i) free cellulose mechanism: secretion of "free enzymes" extracellularly and (ii) cellulosomal mechanism: enzymes that are maintained in close association with the outer cell envelope layer (Wilson, 2009). Some researchers believe that there is a third strategy as discussed below (Wilson, 2008).

Most aerobic cellulolytic microbes, including bacteria and fungi, secrete sets of individual enzymes, which act synergistically on native cellulose. Many of these enzymes contain one or more CBMs, joined by a flexible linker peptide to the CD. The CBMs may be found on the C- or N- terminus, depending on the enzyme, but the location of a CBM normally does not affect enzyme activity (Wilson, 2012). Processive exocellulases and endocellulases are believed to be important components of the cellulose degrading enzyme complex and often account for more than half of the total cellulose degrading proteins (Wilson, 2009).

Most anaerobic microorganisms utilize a different strategy for cellulose degradation in the form of large multienzyme complexes termed cellulosomes, with molecular

weights of over 1 million Dalton (Da) (Bayer et al., 2008). Only a few of the enzymes in cellulosomes contain CBMs, but the scaffoldin protein to which the enzymes attach contains a CBM that binds to cellulose (Bayer et al., 2008). Processive cellulases are also important cellulosomal components (Wilson, 2009).

A strategy used by two cellulolytic bacteria does not seem to conform to either of the two methods of plant cell wall degradation described above (Wilson, 2008). The Gram-negative anaerobe *F. succinogenes* is one of the major cellulose degraders within the rumen and has been extensively studied (Jun et al., 2007). It grows very rapidly when utilizing cellulose as the sole carbon and energy source, owing to its efficient cellulose degrading mechanism (Fields et al., 2000). Another Gram-negative bacterium *Cytophaga hutchinsonii* is an aerobic cellulolytic bacterium. Both the genomic sequences of *C. hutchinsonii* (Xie et al., 2007) and *F. succinogenes* (Qi et al., 2005; Ransom-Jones et al., 2012) provide strong evidence that their mechanism of cellulose digestion differs from the two strategies previously described. Most of the cellulase genes do not encode for a CBM, nor a dockerin domain or scaffoldin gene. All of the cellulase genes appear to code for endoglucanases, and there are no genes that code for any known exocellulases or processive endocellulases (Qi et al., 2005; Ransom-Jones et al., 2012). These aspects suggest that these organisms do not utilize the free cellulase mechanism or the cellulosome mechanism for the degradation of cellulose. One possible mechanism proposed was that individual cellulose molecules are transported into the periplasmic space where they are degraded by endoglucanases (Ransom-Jones et al., 2012; Wilson, 2008).

1.5 Current advances in omics studies

In the past, investigations focusing on rumen microbial communities were usually directed at describing diversity and richness. Studies investigating the functionality of the rumen ecosystem were directed at isolating the dominant species. In the last decade, studies began to investigate the functionality of the complete microbial consortium by using methods such as real time quantitative PCR (qPCR), cDNA (complementary DNA) libraries, microarrays and more recently, meta-omic approaches (i.e., metagenomics, metatranscriptomics and metaproteomics). These approaches have become possible as a result of recent advances in high-throughput sequencing and mass spectrometry based peptide sequencing as well as computational analyses. Combining these techniques has allowed researchers to determine the microbial genes involved and gene expression by natural communities without the need for cultivation in the laboratory (Rosen et al., 2009).

1.5.1 Next generation sequencing

Prior to the introduction of next-generation sequencing (NGS), the automated Sanger method dominated the DNA sequencing market for almost 20 years, and led to a number of historically outstanding accomplishments, including the sequencing of the human genome. Sanger sequencing is based on the chain-termination method, in which a series of different-sized fragments of DNA are generated from numerous identical copies of one DNA molecule starting at the same location, but ending at different locations with a chain terminating dideoxynucleotide labelled with one of four fluorescent dyes. All the fragments are then resolved in order of the length via capillary electrophoresis and the

original sequence is determined through sequentially "reading" the last chain terminating fluorescently labelled dideoxynucleotide of each fragment (Schadt et al., 2010).

The limitations of automated Sanger sequencing — chiefly low throughput and high cost, created a demand for new and improved technologies for mass sequencing of genomes. In comparison to automated Sanger sequencing, newer methods are referred to as NGS. The major advantage offered by NGS is the ability to produce an enormous volume of data at a very affordable price. At the same time, even newer techniques are emerging and are referred to as third-generation sequencing (TGS), the most promising of which is the single molecular real-time sequencing technology developed by Pacific Biosciences (Pac-Bio) (Schadt et al., 2010) which has just been newly introduced to the market.

Currently several NGS technologies are commercially available including Roche/454, Illumina/Solexa, Life Technologies/SOLiD, and Ion Torrent (Loman et al., 2012). Although these platforms are quite diverse in sequencing biochemistry, conceptually their work flows are similar (Shendure and Ji, 2008). They all rely on a three-stage workflow of library preparation, template amplification and sequencing (Loman et al., 2012).

Generally speaking, an initial fragmentation step is required to generate random, overlapping DNA fragments ranging from 150 base pair (bp) to 800 bp in length by either mechanical or enzymatic fragmentation. Specific adaptors can be ligated to the ends of the fragmented molecules to serve as primer-binding sites for the subsequent template amplification reaction (Loman et al., 2012). Mate pair sequencing is supported by all platforms. In this method, the ends of DNA fragments (typically several kilobases)

are joined together to form circular molecules and subjected to a second fragmentation. Fragments flanking the joint position are then selected and adaptors added (Metzker, 2010). Paired-end sequencing is similar to mate pair sequencing, but DNA fragments are directly sequenced from each end without the need for additional preparation steps (Metzker, 2010). The Illumina platform has direct support for paired-end sequencing. Mate pair and paired-end sequencing provide valuable information about the location of sequences distributed across the genome, facilitating assembly (Loman et al., 2012).

Current commercial NGS platforms immobilize and spatially separate millions or billions of template molecules on a solid surface where they are amplified. Simultaneous solid-phase amplification of the immobilized singular template fragments enables massive parallel sequencing (Shendure and Ji, 2008). Pac-Bio's TGS technology only requires unamplified single DNA molecule templates, with no amplification step required (Schadt et al., 2010).

The actual sequencing procedure is often described as "sequencing by synthesis" and relies on imaging-based data acquisition (Shendure and Ji, 2008). Several biochemistry mechanisms are applied by different platforms (Table 1.2) (Loman et al., 2012), and the enzyme driving the synthesis can be either a DNA polymerase or a DNA ligase. Data are acquired each cycle by imaging fluorescent signals of the full array, which are generated when fluorescently labelled nucleotides are incorporated. Each sequencing run generates millions of short sequences called reads. The detailed sequencing and imaging mechanisms are clearly explained and illustrated in a number of excellent review papers (Mardis, 2008; Metzker, 2010; Shendure and Ji, 2008).

It is likely that multiple platforms will coexist in the marketplace with some exhibiting advantages for particular applications over others, since there is considerable variation in performance including throughput, read length, error rate, as well as, cost and run time (Table 1.2) (Loman et al., 2012). The efficiency of these technologies is rapidly advancing as NGS companies are constantly improving their platforms to enable more rapid and comprehensive sequencing at lower cost.

The NGS technologies dramatically outperform older Sanger-sequencing technologies by a factor of 100 - 1,000 in daily throughput, and reduce the cost of sequencing one million nucleotides to as low as 0.1% of that associated with Sanger sequencing (Kircher and Kelso, 2010). This dramatic improvement in sequencing efficiency at reduced cost has opened up new approaches on how sequencing based technologies can be applied. As a result there has been an exponential increase in publications in which NGS is applied for a vast variety of research purposes. Important applications include: (I) full-genome sequencing, more targeting discovery of mutations or polymorphisms or large-scale comparative and evolutionary studies by sequencing many related organisms or strains within one species (pangenomics) (Metzker, 2010); (II) metagenomics, which targets the whole microorganism ecosystem directly obtained from environmental samples, instead of depending on cultivation of individual microbial species (Mardis, 2008; Turnbaugh et al., 2009); (III) mapping of structural rearrangements, including copy number variation, balanced translocation breakpoints and chromosomal inversions (Shendure and Ji, 2008); (IV) large-scale analysis of DNA methylation (Meissner et al., 2008); (V) ‘Chip-Seq’—genome wide mapping of DNA-protein interactions, by deep sequencing of DNA fragments which are isolated through

chromatin immunoprecipitation (Mardis, 2008; Park, 2009); and (VI) 'RNA-Seq'—sequencing of RNA molecules, which is covered in detail in section 1.5.2.

1.5.2 Transcriptomics, metatranscriptomics and RNA-Seq

The transcriptome is the complete set of transcripts in a cell, and the quantity that has been synthesized for a specific developmental stage or physiological condition (Wang et al., 2009). Transcriptomics, the study of the transcriptome, is essential for interpreting the expressed functional elements of the genome. Metatranscriptomics is a branch of transcriptomics, which studies and correlates the transcriptomes of a group of interacting organisms or species. Since its inception, transcriptomics has quickly become an important and promising tool for ecological studies, especially those focusing on complex communities (Warnecke and Hess, 2009). Although DNA-based genomics and metagenomics provide abundant information on the metabolic and functional capacity of an organism or a microbial community, they cannot differentiate between expressed and non-expressed genes, and thus are not a true reflection of metabolic activities (Sorek and Cossart, 2010). On the contrary, transcriptomics and metatranscriptomics retrieve and sequence RNAs from a species or environmental samples. Thus, they provide the most unbiased perspective on gene transcription *in situ* (Su et al., 2012).

Before the wide application of NGS technologies, hybridization-based technologies such as microarray and Sanger sequencing were applied to assess the transcriptome. Microarrays were usually preferable, as it was not practical to use Sanger sequencing to sequence such a large volume of genetic material (Conway and Schoolnik, 2003; Ozsolak and Milos, 2011). In these studies, only a portion of the transcript was analysed and isoforms were generally indistinguishable from each other. These disadvantages limit the

capability of annotating the structure of transcriptomes. These studies also faced several challenges including the low recovery of high-quality mRNA from environmental samples, the short half-lives of mRNA species, and the need for separation of mRNA from other RNA species (Simon and Daniel, 2011). These limitations have been overcome to a great extent with the improvement of RNA isolation techniques in the past decade, together with the NGS-based RNA-Seq (RNA sequencing) technique (Wang et al., 2009). In contrast to microarray methods, NGS-based approaches directly determine the cDNA, or even RNA sequences. RNA-Seq provides a powerful method for both mapping and quantifying transcriptomes. The experimental procedure of RNA-Seq is similar to other NGS applications. Depending on the purpose of the study, a fraction of the total RNAs are isolated from a species or environmental sample and serves as the starting material for library construction. Several fractionation methods are used, based on the length/size or the traits of the target molecules. For example, studies targeting expression profiles of eukaryotes enrich messenger RNAs (mRNAs) by taking advantage of the fact that mature eukaryotic mRNAs are modified with the addition of polyadenylic acids to the 3' end of the mRNA molecules (poly-A tailed). Oligo-dT primers hybridize to the poly-A tailed RNA fraction and thus selectively enrich the mRNA molecules that typically only account for 5-10% of total RNAs. With this approach, the highly abundant ribosomal RNAs and transfer RNAs are largely eliminated from the sample. Afterwards the RNAs are converted to a library of cDNA fragments with adaptors attached to the ends (Costa et al., 2010). The constructed library is sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (paired-end sequencing) as described in section 1.5.1. Following sequencing, the

resulting reads are either aligned to a reference genome or reference transcriptome, or assembled *de novo* without a reference genome. This generates a genome-scale transcription map that consists of both the transcriptional structure and level of expression of each gene (Martin and Wang, 2011).

RNA-Seq has provided the most promising approach for mapping and quantifying transcriptomes, especially metatranscriptomes (Wang et al., 2009), and offers several key advantages. First, unlike hybridization-based methods, RNA-Seq does not require prior knowledge of what genes might exist or be expressed, and thus is not limited to detecting transcripts that correspond to an existing genomic sequence. After sequencing, the resulting reads can not only be aligned to an existing reference genomic sequence or reference transcripts, but assembled *de novo* without the genomic sequence (Martin and Wang, 2011). This makes it possible to identify novel gene sequences, and to quantify rare transcripts without prior knowledge of a particular gene. Consequently RNA-Seq is particularly attractive for non-model organisms and complex environmental samples where limited existing sequence information is available. Metatranscriptomics have been used to analyze many microbial communities including ocean surface waters (Frias-Lopez et al., 2008), coastal waters (Gilbert et al., 2008), soil samples (Urich et al., 2008) and the human gut (Gosalbes et al., 2011).

A further strength of RNA-Seq is its ability to detect and quantify individual transcript isoforms. Alternative splicing is known to contribute to functional diversity in eukaryotes, but it has not been well studied at the level of the transcriptome, principally because of the difficulty of measuring expression for each isoform (Malone and Oliver, 2011). RNA-Seq approaches provide direct sequence information that spans exon/exon

boundaries and makes it possible to study the expression, diversity and abundance of different isoforms of a gene (Malone and Oliver, 2011). Precise location of transcription boundaries and other RNA processing events can also be obtained (Metzker, 2010; Nagalakshmi et al., 2010).

Additionally, RNA-Seq is able to detect a very large dynamic range of expression level of transcripts (Gilbert and Hughes, 2011). In theory, the sequencing depth used is the only restriction on quantification limit. Deeper sequencing will detect sequences expressed at lower levels and quantify expression levels more accurately (Malone and Oliver, 2011). In contrast, microarrays are not sensitive enough to quantify genes expressed at very low levels whereas those expressed at very high levels can saturate the array (Costa et al., 2010). RNA-Seq has also shown consistent results when compared with qPCR results, with high levels of reproducibility (Wang et al., 2011).

Finally, RNA-Seq usually requires lower amounts of RNA sample compared to Sanger sequencing, because there are no cloning steps involved in library construction (Ozsolak and Milos, 2011). This is a huge advantage, especially for projects where limited amounts of RNA can be isolated.

Taking all of these advantages into account, RNA-Seq is the first sequencing method that allows the entire transcriptome to be surveyed in a very high-throughput manner. At a reasonable cost, RNA-Seq offers single-base resolution for annotation and quantification of gene expression levels at the genome scale.

1.5.3 Limitations and Challenges related to NGS and RNA-Seq

Next generation sequencing technologies have been extensively improved since the introduction of the first commercial platform in 2005 (Loman et al., 2012). Currently in addition to reducing the per-base cost of sequencing by several orders of magnitude, NGS instruments also have fewer infrastructure requirements. However, when compared to Sanger sequencing, NGS is still limited in terms of read-length and accuracy.

Compared to Sanger sequencing's 10^{-4} – 10^{-5} error rate, the error rates of NGS technologies are extremely high at 10^{-2} – 10^{-3} (Kircher and Kelso, 2010). Consequently, even though deep sequencing provides abundant sampling depth compared to traditional approaches, such as DGGE (Denaturing gradient gel electrophoresis), T-RFLP (Terminal restriction fragment length polymorphism), or 16S ribosomal RNA (rRNA) gene clone libraries, the high error rate may result in the overestimation of rare phylotypes (Su et al., 2012). Despite that, direct sequencing of metagenomic DNA is still proposed to be the most accurate approach currently available for assessment of taxonomic composition as it avoids the bias introduced by polymerase chain reaction (PCR) amplification of DNA in approaches such as DGGE and 16S rRNA clone libraries (Su et al., 2012).

Read length for NGS technologies also remains limited. Sanger sequencing can normally reach over 1,000 bp; however, Roche/454 has the longest average read-length of up to 500 bp, while the other NGS technologies only read 100 – 200 bp (Loman et al., 2012). The relatively short read lengths consequently raise bioinformatic challenges as to how best and most efficiently extract biologically meaningful insights from the very large datasets produced (Kircher and Kelso, 2010; Shendure and Ji, 2008).

Downstream data management and bioinformatic analysis are the principal challenges with NGS (Pop and Salzberg, 2008). The massive data sets produced place substantial demand on information technology in terms of data storage, tracking, quality control as well as statistical analysis (Datta et al., 2010; Pop and Salzberg, 2008). Assembling millions of short reads into contigs before alignment to the reference genomic sequence, or mapping the short reads directly to the reference genome raises considerable bioinformatic challenges (Pop and Salzberg, 2008). Special attention must be paid to exon-exon junctions and polyA tails (Costa et al., 2010). Repetitive DNA or extremely AT or GC rich sequences present technical challenges as these regions are ambiguous for sequence alignment (Treangen and Salzberg, 2012).

Although small-scale projects in the kilobase-to-megabase range will still likely use conventional Sanger sequencing, future large-scale projects will likely rely entirely on NGS. As shown in Table 1.2, there are important differences among the NGS platforms that result in advantages with respect to specific applications. Some applications may be more tolerant of short read-lengths than others or differ in their overall accuracy and source of errors such as the rate of insertion-deletion vs substitution errors. Other considerations include the availability of the platform and sequencing costs.

1.5.4 Current advances of omics studies related to rumen fungi and microbiomes focusing on plant fiber degradation

Previous to the wide application of NGS technologies, an excellent study defining the nature of the termite hindgut microbiome generate over 71 million base pairs (Mbps) by applying Sanger sequencing method (Warnecke et al., 2007). This work set a new

standard for metagenomic research that focused on microbial consortia involved in plant fiber degradation. Since, as mentioned, the rumen is a unique resource for the discovery of novel plant cell wall degrading enzymes, a research group applied NGS to analyze the bovine rumen metagenome (Brulc et al., 2009). This study using 454 sequencing technology generated 103 Mbp of sequences from three fiber-adherent and one pooled liquid sample obtained from the rumens of three Angus Simmental Cross steers. Shortly after, the great power of Illumina NGS technologies was demonstrated in a cow rumen metagenomic project aiming at searching for potential novel lignocellulosic degrading enzymes suitable for the cellulosic biofuel industry (Hess et al., 2011). This project is the deepest and most complete rumen metagenomic study currently available and greatly expanded our understanding of rumen microbiota to a new level. From a total of 268 giga base pairs (Gbps), a little over 2.5 million open reading frames (ORFs) were assembled, with over half predicted to represent full-length genes. Over 1,500 OTUs and 27,000 putative CAZymes were identified, a level that was much higher than those identified from previous studies (Brulc et al., 2009; Pope et al., 2010; Warnecke et al., 2007). The functions of the CAZymes were also investigated. Ninety candidate proteins were expressed and 57% were found to be enzymatically active.

In an attempt to identify more unique CAZymes, studies have also been aimed at microbiomes from animals living in unique environments. Since macropods evolved in geographical isolation of other herbivores, they show a wide range of unique adaptations to diets and were proposed to harbour a different microbiome in their foregut than those existing in the bovine rumen. Pope et al. (2010) studied the Tammar wallaby foregut microbiome using Sanger sequencing to characterize 16S rRNA gene clone libraries and

shotgun libraries, with 454 sequencing performed on selected fosmids. The sequencing obtained a total of over 600 Mbp sequences.

When the OTUs detected from the termite (Warnecke et al., 2007), bovine rumen (Brulc et al., 2009) and wallaby were compared, clear host-specificity was obvious, with only a small number of OTUs shared between the bovine and wallaby microbiome, with those in termite microbial community being unique (Pope et al., 2010). This was expected as each unique host has a highly adapted microbiota. At the same time, the microbiomes from three individual bovine rumens fed on the same diet also exhibited considerable diversity (Brulc et al., 2009).

Dai et al. (2012) recently explored the rumen cellulolytic microbiome of Tibetan yak rumen by screening bacterial artificial chromosome (BAC) library for fibrolytic enzyme activities and 223 positive BAC clones were pyrosequenced on Roche/454 platform. About 150 glycoside hydrolase genes were identified with the majority occurring in gene clusters.

Pope et al. (2012) also investigated the rumen microbiome of arctic reindeer. Multiple polysaccharide utilization loci-like systems were found, as well as about 5,000 putative GH gene fragments from over 20 CAZy families, by analyzing the sequences as well as metabolic reconstruction of the Bacteriodales-related clade. A number of cohensin/dockerin modules, which were rarely reported in previous rumen metagenomic studies, were also identified, suggesting that cellulosomes may play an important role in cellulose digestion within these arctic ruminants.

Most of the previous studies have directly targeted the fiber degrading microbiota, and as expected, resulted in most of the sequences being of bacterial origin, as bacteria

represent the majority of the biomass in the rumen microbiota. A newly published paper aiming at transcriptomics and secretomics of *Neocallismatix patriciarum* W5 improved our knowledge towards the anaerobic fungi, a group of active fiber degraders (Wang et al., 2011). Other than sequencing the genome of this fungus, the researchers focused on functional characterization by sequencing transcriptomes expressed under various growth conditions, as well as secretomic analysis based on mass spectroscopic analysis. Both 454 and Illumina sequencing platforms were applied. This study helped to gain a better global understanding of the GHs produced by *N. patriciarum*. A total of 219 putative GHs were classified into 25 GH families. Some highly expressed or potentially full length contig candidates were expressed. At least five novel cellulases displayed activities, and one β-glucosidase and one exocellulase demonstrated high enzyme activities.

1.6 Research objectives

The rumen microbial ecosystem has now been investigated in detail for over half a century. Over the past two decades, great improvements have been made towards understanding the dynamic nature of this unique microbiota and it is one of the most accessible and understood microbial ecosystems (Flint, 1997). The rumen is widely recognized as one of the most unique fibrolytic microbial ecosystem that is second to none in its ability to convert plant cell wall polysaccharide to fermentable sugars.

The overall goal of this thesis project is to increase our understanding of the transcriptome of a representative rumen anaerobic fungus *Anaeromyces mucronatus* YE505 and the metatranscriptome of particle associated microbiota from muskoxen (*Ovibos moschatus*) rumen, with an emphasis on CAZyme coding sequences. Based on

the current incomplete understanding of the rumen microbiota, it is reasonable to propose the hypothesis that a broad range of fibrolytic degrading enzymes yet unknown will be discovered by this project by achieving the following objectives:

1) Establish a fast and reliable RNA isolation method for extracting total RNA from rumen samples, especially from feed particle associated microbiota.

2) Elucidate the transcriptomes from the rumen fungus *A.mucronatus* YE505 grown on various carbon sources.

3) Elucidate the metatranscriptome of the rumen solid associated eukaryotes from muskoxen and study the gene expression profile of these feed particle associated microorganisms.

Metatranscriptomics is a rapidly emerging field and has shown considerable potential as a means of identifying novel biocatalysts (Sorek and Cossart, 2010; Warnecke and Hess, 2009). By applying (meta)transcriptomic analysis, this research will provide a broader and deeper picture of the rumen fungi and the rumen ecosystem of muskoxen, and enable sequence-based approaches to identify genes coding for novel enzymes. It will enhance our understanding of the molecular mechanism of lignocellulose bioconversion. Furthermore, the methods and procedures established in this study will enable more detailed investigation of the impact of the host, diet and other conditions on rumen function at the gene expression level. Knowledge generated by this study will also aid in the industrial conversion of renewable plant fiber biomass to value added economically significant products.

1.7 Tables and Figures

Table 1.1 Currently classified anaerobic fungal species isolated from the gut of herbivores (Borneman and Akin, 1994; Breton et al., 1991; Ozkose et al., 2001; Nagpal et al., 2009).

Genus	Species	Source of isolation
Neocallimastix	*N. frontalis*	Sheep
	N. patriciarum	Sheep
	N. hurleyensis	Sheep
	N. variabilis	Cattle
Piromyces	*P. communis*	Sheep
	P. mae	Horse
	P. dumbonica	Elephant
	P. rhizinflata	Ass
	P. minutus	Deer
	P. spiralis	Goat
	P. citronii	Horse
	P. polycephalus	Water buffalo
Caecomyces	*C. communis*	Sheep
	C. equi	Horse
Orpinomyces	*O. intercalaris*	Cattle
	O. joyonii	Sheep
Anaeromyces	*A. elegans*	Cattle
	A. mucronatus	Cattle
Cyllamyces	*C. aberensis*	Cattle

Table 1.2 Comparison of next-generation sequencing platforms (Loman et al., 2012) (reprinted with permission).

Machine (Manufacturer)	Chemistry	Model read length (base)[*]	Run time	Gb per run	Current approx. cost (US $)[#]	Advantages	Disadvantages
High-end instruments							
454GS FLX+ (Roche)	Pyrosequencing	700-800	23 h	0.7	500k	• Long read lengths	• Appreciable hands-on time • High reagent costs • High error rate in homopolymers
HiSeq 2000/2500 (Illumina)	Reversible terminator	2 × 100	11 d (regular mode) or 2 d (rapid run mode)[§]	600 (regular mode) or 120 (rapid run mode)[§]	750k	• Cost-effectiveness • Steadily improving read lengths • Massive throughput • Minimal hands-on time	• Long run time • Short read lengths
5500xl SOLiD (Life Technologies)	Ligation	75 + 35	8 d	150	350k	• Low error rate • Massive throughput	• Very short read lengths • Long run times
PacBio RS (Pacific Biosciences)	Real-time sequencing	3,000 (max 15,000)	20 min	3 per day	750k	• Simple sample preparation • Low reagent costs • Very long read lengths	• High error rate • Expensive system • Difficult installation
Bench-top instruments							
454 GS Junior (Roche)	Pyrosequencing	500	8 h	0.035	100k	• Long read lengths	• Appreciable hands-on time • High reagent costs • High error rate in homopolymers
Ion Personal Genome Machine (Life Technologies)	Proton detection	100 or 200	3 h	0.01-0.1 (314 chip), 0.1-0.5 (316 chip) or up to 1 (318 chip)	80k (including OneTouch and server)	• Short run times • Appropriate throughput for microbial application	• Appreciable hands-on time • High error rate in homopolymers

Ion Proton (Life Technologies)	Proton detection	Up to 200	2 h	Up to 10 (Proton I chip) or up to 100 (Proton II chip)	145k + 75k for compulsory server	• Short run times • Flexible chip reagents	• Instrument not available at time of writing
MiSeq (Illumina)	Reversible terminator	2 × 150	27 h	1.5	125k	• Cost-effectiveness • Short run times • Appropriate throughput for microbial applications • Minimal hands-on time	• Cost-Read lengths too short for efficient assembly

[*] Average read length for a fragment-based run.

[#] Approximate cost per machine plus additional instrumentation and service contract.

[§] Available only on the HiSeq 250.

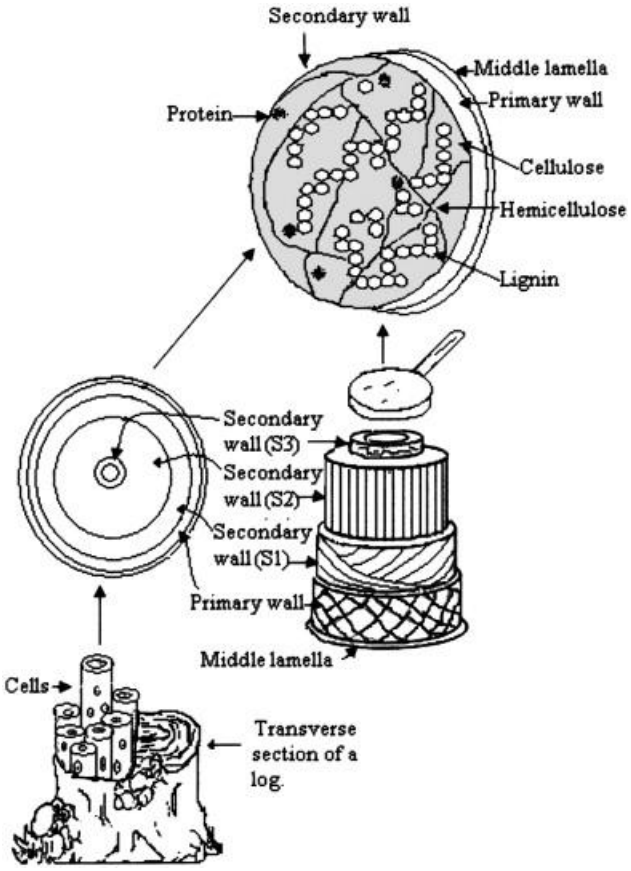

Figure 1.1 Composition of lignocellulosic residues (Sánchez, 2009) (reprinted with permission).

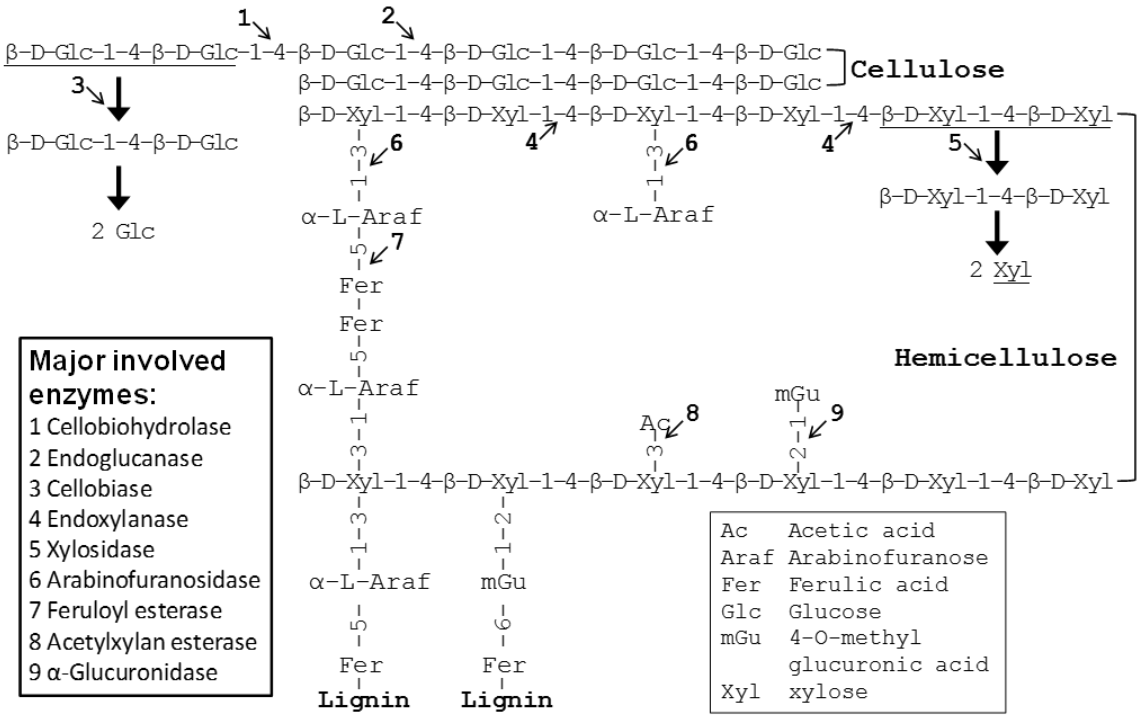

Figure 1.2 Simplified schematic diagram illustrating plant cell wall components and enzymes involved in plant cell wall degradation.

Chapter 2 Isolation of high-quality total RNA from rumen anaerobic bacteria and fungi, and subsequent detection of glycoside hydrolases*

2.1 Introduction

The rumen is a highly specialized fermentation chamber containing an array of unique fibrolytic enzymes produced by bacteria, protozoa and fungi with potential broad application in both research and industry (Selinger et al., 1996). Microbial populations within the rumen have commonly been studied from the perspective of the liquid or solid fraction of rumen contents (Cheng and McAllister, 1997). Solid-associated microorganisms represent the major proportion of total rumen microbes (McAllister et al., 1994; Yu and Forster, 2005), and are estimated to produce up to 90% of the endoglucanase and xylanase activities in the rumen (Miron et al., 2001). Consequently, the study of the solid phase microbial community is likely to yield the most information about rumen micriobial function.

Until recently, research on rumen microbial communities was mainly targeted at describing diversity, while research on function was limited to a relatively small number

* This chapter is an adapted version of the manuscript "Wang, P., Qi, M., Barboza, P., Leigh, M.B., Ungerfeld, E., Selinger, L.B., McAllister, T.A., and Forster, R.J. (2011). Isolation of high-quality total RNA from rumen anaerobic bacteria and fungi, and subsequent detection of glycoside hydrolases. Can. J. Microbiol. **57**(7): 590-598".

of culturable microbial species. Recently, developments in next generation sequencing have made the culture-independent study of complex natural microbial habitats possible (Warnecke and Hess, 2009). However, to describe the genes present and their expression, reliable and repeatable nucleic acid isolation techniques are required.

Although total RNA has been isolated from environmental microbial communities in soil and the human intestinal tract (Peršoh et al., 2008; Sessitsch et al., 2002; Zoetendal et al., 2006), obtaining representative extracts of RNA from rumen contents, especially solid phase, remains challenging. Previous attempts to isolate solid-attached ruminal microorganisms relied on detachment of cells from feed particles. However, attached microorganisms grow in the form of multi-species biofilms that often lyse during the separation step and result in quick RNA degradation. Since rumen fungi penetrate feed particles, they are generally inaccessible to isolation methods that rely on biofilm detachment. It has repeatedly been shown that the recovery of attached microorganisms using a variety of detachment methods is incomplete (Martín-Orúe et al., 1998; Ranilla and Carro, 2003; Trabalza-Marinucci et al., 2006; Whitehouse et al., 1994), and that the detachment method affects the profile of the microbiota recovered from solid particles (Martínez et al., 2009b; Ramos et al., 2009). Furthermore, feed particles are rich in phenolic acids, polysaccharides and proteoglycans that readily form complexes with nucleic acids and inhibit reverse transcription and/or PCR reactions (Monteiro et al., 2001; Sharma et al., 2003). To date, only two RNA isolation procedures from rumen liquids have been reported, and resulted in modest yields of partially intact RNA, with no evidence of RNA contributions from rumen eukaryotes (Béra-Maillet et al., 2009; Kang et al., 2009). The isolation of intact total RNA from rumen solids has yet to be reported.

Here we report a successful isolation of total RNA from both solid and liquid phases of rumen contents. This procedure should facilitate the identification of actively transcribed genes from a variety of feed-associated microbes, including rumen fungi, enabling more representative gene expression profiles of the rumen ecosystem to be compared among individual hosts and with changes in diet and other conditions. Additionally, the technique may facilitate the isolation of intact, high quality total RNA from a variety of environmental sources.

2.2 Methods

2.2.1 Animals and rumen sampling

The overall experimental flowchart used for this study is illustrated in Figure 2.1. For isolation of total RNA from liquid phase rumen samples and for comparison of methods developed in this study and previous studies (Béra-Maillet et al., 2009; Kang et al., 2009), samples of ruminal content were obtained from a ruminally cannulated Holstein cow fed a 40% barley grain - mixed grass hay diet immediately prior to feeding. The cow was housed in a tie-stall barn at the Agriculture and Agri-Food Canada Lethbridge Research Centre in Lethbridge, Alberta, Canada and was cared for in accordance with the guidelines set by the Canadian Council on Animal Care (CCAC, 1993). Immediately after withdrawal of a sample, a liquid phase sample was separated by squeezing through four layers of cheesecloth and the liquid was transported to the laboratory in an insulated vessel. Any remaining large particulate fragments were then separated using a Bodum coffee filter plunger (Bodum Inc., Triengen, Switzerland). Fluid

phase aliquots of 0.2 mL each were placed in 2 mL microfuge tubes and stored at -80 ℃ until processed further (Figure 2.1, pane A).

Solid phase rumen contents were obtained from ruminally cannulated muskoxen at the R. G. White Large Animal Research Station, University of Alaska, Fairbanks, Alaska (Figure 2.1, pane B). All procedures with muskoxen were approved under protocol No. #139821-2 by the Institutional Animal Care and Use Committee at the University of Alaska Fairbanks. Solid phase samples were obtained by placing whole ruminal contents in a heavy walled 250 mL beaker and by separating the particulate and liquid phases using a Bodum coffee filter as described above. Subsamples of solid digesta (~ 2.5 g) were immediately flash-frozen in liquid nitrogen (Figure 2.1, pane B, branch B1). To examine the effectiveness of RNAprotect bacteria reagent (Qiagen, Mississauga, Ontario, Canada), solid phase samples were separated into six portions (~2.5 g each), with 5 mL of RNAprotect immediately stirred into each of the six samples just prior to being flash-frozen in liquid nitrogen. All samples were frozen within 5 min of the sample being withdrawn from the animal (Figure 2.1, pane B, branch B2). Samples were immediately transferred to the laboratory, and stored at -80 ℃ until processed further.

2.2.2 Total RNA isolation

2.2.2.1 Liquid phase rumen sample RNA isolation

The acid guanidinium-phenol-chloroform (AGPC) method (Chomczynski and Sacchi, 1987) served as the standard RNA isolation procedure ("control method"). Two published procedures for rumen sample RNA isolation (Béra-Maillet et al., 2009; Kang et

al., 2009) were tested and they were referred to as method K (Kang et al., 2009) and method B (Béra-Maillet et al., 2009).

An optimized isolation procedure, designated LRCI (Liquid Ruminal Contents Isolation) was developed. All experimental procedures were performed on duplicate subsamples as follows. Microfuge tubes containing 0.2 mL liquid phase rumen subsamples (Figure 2.1, pane A) were withdrawn from the freezer and 1.5 mL of TRIzol reagent was immediately added into each tube. The samples were then allowed to thaw at room temperature. Cells were disrupted by bead beating for 3 min at 300 revolutions per second with 0.2 g of glass beads of size range 0.7–1.1 mm (Sigma-Aldrich, Oakville, Ontario, Canada) at room temperature on TissueLyser (Qiagen). The homogenized sample was allowed to stand at room temperature for 5 min and RNA was isolated following the AGPC method (Chomczynski and Sacchi, 1987). The air-dried RNA pellet was re-dissolved in 100 µL of nuclease-free water (Qiagen). The RNA cleanup was performed by using either RNeasy mini kit or MEGAclear kit according to the manufacturer's instructions.

2.2.2.2 Solid phase rumen sample RNA isolation

A procedure designated SRCI (Solid Ruminal Contents Isolation) was developed. First, rumen solids (RS) were manually ground to a fine powder in liquid nitrogen using a mortar and pestle, and then further ground for 5 min in liquid nitrogen using a Retsch RM100 grinder (Retsch GmbH, Haan, Germany). Ground samples (~200 mg) were placed in 2 mL microfuge tubes and each was mixed with 1.5 mL of TRIzol reagent. The samples were thawed, incubated at room temperature for 5 min and subsequently the

RNA was extracted using the AGPC method (Chomczynski and Sacchi, 1987), as described in the LRCI procedure.

The LCRI procedure with minor modification was also tested on rumen solids, by putting a small piece of frozen sample (~0.2 g) into a 2 mL microfuge tube and 1.5 mL of TRIzol reagent was immediately added. Cell disruption by bead beating used glass beads and the same equipment settings as liquid-phase samples but was carried out twice instead of once, with 5 min of incubation on ice between intervals.

2.2.2.3 Effects of RNAprotect Bacteria Reagent on RNA isolation from solid phase rumen contents

To test the efficacy of RNAprotect to preserve solid-phase rumen RNA, we employed RNAprotect in three different ways during the extraction of RNA from rumen solids and compared the results with the RNAprotect-free SRCI procedure. RNAprotect was added to rumen solids as described above in the sampling section. The six samples were divided into three groups in duplicate for the following three treatments (Figure 2.1, pane B, branch B2). In the first method (treatment I), the ~ 2.5 g sample was ground in liquid nitrogen in the presence of 5 mL of RNAprotect. Subsequently, TRIzol reagent was added and RNA was extracted according to the above SRCI procedure without removing RNAprotect. In the second method (treatment II), the sample was thawed on ice and centrifuged at 4 ℃ for 10 min at 5000 × g. Excess RNAprotect was removed from the sample as recommended in the RNAprotect manual. The pellets were then ground in liquid nitrogen and RNA was subsequently extracted. In the third method (treatment III), the sample was ground in liquid nitrogen in the presence of RNAprotect, but the sample was allowed to completely thaw and was centrifuged at 4 ℃ for 10 min at

5000 × g to remove the supernatant, which contained most of the RNAprotect reagent, and then TRIzol reagent was added to the pellets and RNA was subsequently extracted.

2.2.3 Effects of RNA clean-up kits on RNA quality

Three laboratory kits: RNeasy mini kit (Qiagen), RiboPure kit (Applied Biosystems/Ambion, Streetsville, Ontario, Canada), and MEGAclear kit (Applied Biosystems/Ambion), were tested for their ability to purify SRCI-extracted RNA from solid ruminal contents. The purification procedures were performed according to the manufacturer's instructions.

2.2.4 Evaluation of RNA quantity and quality

RNA purity was estimated by measuring the absorbance ratio at A260/A280 and A260/A230 using a NanoDrop 2000 Spectrophotometer (Thermo Scientific, Waltham, Massachusetts). RNA concentration and integrity were estimated using an Agilent 2100 bioanalyzer (Agilent Technologies, Mississauga, Ontario, Canada) and RNA 6000 Nano kit (Agilent Technologies) according to the manufacturer's recommendations. The prokaryotic total RNA nano assay protocol was used since prokaryotes account for the majority of RNA in rumen contents (Yu and Forster, 2005).

Large subunit/small subunit (LSU/SSU) rRNA peak area ratio and RNA integrity number (RIN) analyses were performed for each RNA sample using the 2100 Expert software version B.02.07 (Agilent Technologies).

2.2.5 Reverse transcription and PCR

Total RNA isolated from solid-phase muskoxen rumen sample was treated with a DNA-free kit (Applied Biosystems/Ambion) following the recommended procedures.

After treatment, total RNA (1 µg) was reverse-transcribed into cDNA using a SuperScript III first-strand synthesis kit (Invitrogen) and random hexamer primers according to the manufacturer's instructions. A reverse transcriptase negative control was also included in all runs and the generated products were used in subsequent PCRs.

Fragments of three glycoside hydrolase genes *celF*, x*ynD*, and c*el3* from *Fibrobacter succinogenes* S85 were selected to detect transcript production. Two sets of primers for each gene were chosen as described by Béra-Maillet et al. (2009) (Table 2.1).

Subsequent PCRs were performed by using Platinum *Taq* DNA polymerase High Fidelity (Invitrogen). To each 25 µL PCR reaction system, 1 µL of first-strand cDNA was added. The PCR conditions for whole length primers (designated W, Table 2.1) consisted of an initial denaturation step for 1 min at 94 ℃, followed by 40 amplification cycles of 15 s at 94 ℃, 30 s at 55 ℃, and 3 min at 68 ℃. The final cycle included elongation at 68 ℃ for 5 min. The PCR conditions for internal primers (designated I, Table 2.1) consisted of an initial denaturation step for 1 min at 94 ℃, followed by 40 amplification cycles of 15 s at 94 ℃ , 30 s at 55 ℃ , and 45 s at 68 ℃. The final cycle included elongation at 68 ℃ for 2 min. A PCR negative control (no addition of first-strand cDNA template) was included with each PCR procedure. The PCR products were visualized on a 1% agarose gel stained with ethidium bromide.

2.3 Results

2.3.1 Optimization of the Liquid Phase RNA isolation method

The extraction efficiency of three AGPC reagents—TRIzol (Invitrogen), TriPure (Roche Diagnostics, Laval, Quebec, Canada) and Tri Reagent (Applied

Biosystems/Ambion), was compared and the RNA obtained showed no apparent differences with respect to RNA yield or quality (data not shown). Therefore these three reagents were considered equivalent in our method, and we elected to use TRIzol in further extractions owing to its widespread use.

The three methods examined here resulted in similar RNA purity, with an A260/A280 ratio of 1.8 – 2.0 and an A260/A230 ratio of approximately 2.0. Among them, method K was the most complicated as it involved cell dissociation, both enzymatic and mechanical lyses and a column purification step (Kang et al., 2009). However, method K yielded <40 μg of RNA per mL of rumen fluid (RF), with a LSU/SSU rRNA ratio of <1 and a RIN of <5, both of which indicated considerable degradation (Table 2.2). In contrast with method K, fewer isolation steps yielded higher amounts of RNA from Method B (152 ± 16 $\mu g \cdot (mL\ RF)^{-1}$) (Béra-Maillet et al., 2009) and the LRCI method (172 ± 14 $\mu g \cdot (mL\ RF)^{-1}$) (Table 2.2). The RNA generated from method B were of higher quality than that from method K, with an rRNA ratio of >1.2 and a RIN of >8. The LRCI procedure isolated RNA with the highest quality and quantity, with rRNA ratios as high as 1.8 and a RIN of >9.4 (Table 2.2). When a column cleanup step was included, the RNA quality was further improved by removing 5S region fragments, without compromising RNA yield. A typical RNA sample isolated from rumen liquids by following this LRCI procedure is shown in Figure 2.2, part A.

2.3.2 Total RNA isolation from rumen solids

The application of our optimized method (Figure 2.1, pane B, branch B1) yielded high-quality RNA from solid rumen samples. A typical electropherogram result is illustrated in Figure 2.2, part B, with an average yield of approximately 110 $\mu g \cdot (g\ RS)^{-1}$,

a LSU/SSU rRNA ratio of 1.9, and a RIN of 9.8. We compared the effect of bead beading and grinding under liquid nitrogen on RNA isolation. Disruption of solid contents by bead beading for 3 min two times at room temperature resulted in a low yield of poor-quality RNA (60 $\mu g \cdot (g\ RS)^{-1}$ with a LSU/SSU rRNA ratio of only 0.6. Therefore, grinding of samples in liquid nitrogen was selected for further RNA extraction from rumen solids.

RNA extracted from solid-phase samples from a muskoxen showed a higher complexity than that from the liquid-phase sample from a cow, as is evidenced by the two major ribosome RNA peaks (Figure 2.2). Each peak showed a combination of slightly different-sized fragments, which was expected, since the sizes of rRNAs differ between bacteria, archaea, and eukaryotes. A band was noticeable in the 5S RNA region and may arise from small RNA fragments or other impurities in rumen contents (Figure 2.3, lane 1). Column purification using either RiboPure or MEGAclear kits reduced the presence of these small fragments more effectively than did the RNeasy mini kit (Figure 2.3, lanes 2 to 4). Brownish material, possibly arising from plant phenolics, remained visible after isopropanol precipitation but was removed after column purification. The MEGAclear kit was chosen and used in further RNA extractions from rumen solids.

Isolated total RNA was stable when stored in nuclease-free water at -80 ℃ for at least 3 months. Up to four freeze-thaw cycles were applied without changing the RNA concentration, LSU/SSU rRNA ratio, or RIN, as indicated by the Bioanalyzer analysis on the stored samples (data not shown).

2.3.3 Effect of RNAprotect bacteria reagent on RNA isolation from rumen solids

RNAprotect was added in preliminary investigations to obtain total RNA from rumen solids, but this approach failed to yield high quality RNA. Subsequently, we investigated three approaches for inclusion as well as exclusion of RNAprotect from our procedure for the extraction of RNA from rumen solids. Samples that were extracted without RNAprotect yielded high quality total RNA, whereas those that included RNAprotect yielded similar quantity, but poorer quality RNA (Figure 2.4). All three methods that included RNAprotect resulted in an rRNA ratio of <1.0. Treatment III (complete thaw of sample after sample disruption, followed by removal of RNAprotect before the addition of TRIzol) seemed to be particularly unfavourable, as one of the samples was almost completely degraded, and had an rRNA ratio of only 0.2 (Figure 2.4, lane 8). This may have occurred because the disrupted cells were thawed without sufficient protection before the addition of TRIzol.

2.3.4 RT-PCR and detection of glycoside hydrolases

Internal fragments of glycoside hydrolase genes (*celF*, *xynD* and *cel3*) were amplified by RT-PCR with the fragment size corresponding to the predicted amplicon length. The 3 kb fragment of *celF* was also detected (Figure 2.5). However, the nearly full length amplicons of *xyn*D and *cel*3 were not detected. All the negative controls that lacked reverse transcriptase did not show bands corresponding to the amplicons of interest, confirming that positive amplicons were reverse transcribed from mRNA. However, negative RT-PCR reactions from the template before DNase I treatment

generated amplicons of all three gene fragments, suggesting that trace amounts of DNA were still present in the RNA sample (data not shown).

2.4 Discussion

The present study describes a rapid and effective method to isolate high quality, highly representative total RNA from both liquid and solid ruminal contents.

The LRCI and SRCI methods were simple as no separation step of microorganisms from feed particles was necessary. This reduced the amount of time the samples were exposed to oxygen (which damages the anaerobic microbes) and decreased the time before the samples were frozen, thus reducing the extent of RNA degradation during the separation procedure. Procedures for the extraction of RNA from rumen solid and liquid contents were virtually identical, with the exception that a more vigorous cell disruption process was employed for solid rumen contents. Microbial cells within the liquid phase of rumen contents were efficiently lysed by simple bead beating at room temperature under the protection of TRIzol reagent. However, the quality and quantity of RNA isolated from solid rumen content by the bead beating procedure was far lower. Thus, it was necessary to disrupt microbial cells within the biofilms and particles of solid contents by grinding in liquid nitrogen, as the solids contained large undigested feed particles. These particles could not be efficiently disrupted by bead beating, and prevented TRIzol reagent from penetrating rapidly for sufficient protection. It is important to efficiently break down the microbial cells that are attached to the surface as well as those that have penetrated into the interior of feed particles, since most fibrolytic microbes colonize in these locations, and it has been shown that the largest diversity of bacteria in the rumen exists in the residual particulate fraction (Kong et al., 2010). Grinding of the whole solid

contents in liquid nitrogen not only allows more RNA of greater quality to be obtained, but would also eliminate to a great extent the bias in the microbial populations recovered from solid particles by detachment procedures (Martínez et al., 2009b; Ramos et al., 2009).

The RIN was initially designed to help estimate the integrity of total eukaryotic RNA. It is determined not only by the ratio of the ribosomal bands alone, but also by the entire electrophoretic trace of the RNA sample (Schroeder et al., 2006). A RIN of 10 represents a perfectly intact RNA sample. Although the Agilent software also gives a calculated RIN value for prokaryotic RNA, this approach has not been extensively validated in terms of its value as an indicator of the integrity of prokaryotic RNA. In our study, we noticed that the 2100 Expert software was occasionally unable to discern the rRNA peak areas accurately, making it necessary to manually adjust the peak recognition. In these cases, the RINs did not always correspond to the rRNA ratio or apparent RNA integrity. For example, two SRCI-isolated RNA samples (Figure 2.4, lane 1 and 2) showed similar concentrations, with the same rRNA ratio of 1.7. But under default software parameters, the software only gave a RIN of 5.9 for the sample in lane 2, as opposed to a RIN of 8.2 for the sample in lane 1. However, the RNAprotect samples (Figure 2.4, lane 3 to 8) had rRNA ratios of < 1.0, but had higher RINs (8.4 to 8.6). As the Agilent Bioanalyzer software does not possess a module for mixed rRNA samples, the generated RINs should be regarded as approximate.

According to the manufacturer, RNAprotect bacteria reagent was designed to prevent both degradation of RNA transcripts and induction of genes, and thus provide immediate stabilization of the gene expression profile of bacteria. However, at least in

this study, this reagent reduced both RNA yield and quality. Immediate freezing and grinding of samples in liquid nitrogen followed by mixing in TRIzol reagent yielded RNA of high quality and stability. The addition of RNAprotect to the procedure increased the complexity of the extraction. It resulted in lower RNA yield and quality, whether or not it was removed prior to the addition of TRIzol. Therefore, in the present study we found no value to including RNAprotect in the RNA extraction procedure for either the liquid or solid fraction of rumen contents.

RT-PCR showed that three typical GH genes from *F. succinogenes* S85 were amplifiable from the isolated total RNA. As was found by Béra-Maillet et al. (2009), the three approximately 200 bp internal fragments of *celF*, *xynD* and *cel3* were all detected. A 3000 bp length of the *celF* gene was amplified by using W series primers (i.e., whole-length primers), but no full length amplicons of *xynD* and *cel3* were generated. A possible explanation for this may be related to the fact that random hexamer primers were used for reverse transcription. These reaction conditions are unavoidably biased towards the generation of fragmented, short length first strand cDNA products as opposed to whole length sequences. It was possible that compared with the other two genes, a high transcription level of *celF* existed in the isolated total RNA and resulted in enough whole length cDNA product to be detectable by PCR amplification. We selected these three genes based upon the previous quantitative reverse transcriptase-PCR (RT-qPCR) results by Béra-Maillet et al. (2009), which showed that these genes are highly expressed in rumen contents. According to these authors, *celF* had the lowest transcript level of the three highly transcribed genes under their experiment conditions. However, our study analyzed solid ruminal contents from muskoxen, in which the distribution of, and relation

to known *Fibrobacter* species is unknown. CelF (previously named EGF) was reported to be one of the major cellulose binding proteins identified in *F. succinogenes* (McGavin and Forsberg, 1988; Mitsumori and Minato, 1995). Highly similar gene fragments have also been found in strains of *Fibrobacter intestinalis* (Béra-Maillet et al., 2004; Qi et al., 2005). Considering that the *Fibrobacter* genus is a major contributor to fibrolytic activity within the rumen (Stewart et al., 1997), the detection of *cel*F gene expression would seem probable in most rumen systems. The detection of the 3 kb *celF* fragment attests to the integrity of the isolated RNA.

Using the SRCI procedure described in this manuscript, sequencing of eukaryotic polyadenylated mRNA isolated from rumen of muskoxen was carried out using the Illumina Genome Analyzer (Illumina, San Diego, California). Detailed sequence analysis is described in Chapter 3.

Metatranscriptomics is a rapidly emerging field and has shown considerable potential as a means of identifying novel biocatalysts (Sorek and Cossart, 2010; Warnecke and Hess, 2009). Our method makes it practical to obtain large quantities of high-quality total RNA, enabling sequence-based approaches to identify genes coding for novel enzymes from environmental samples. The procedure could be easily adapted to other environments, such as compost, leaf cutter ant gardens, or soil, with little difficulty. For the rumen environment, the method enables the investigation of the impact of the host, diet, and other conditions that affect ruminal function and gene expression within this unique ecosystem.

2.5 Tables and Figures

Table 2.1 Primers used for RT-PCR.

Protein	Gene	Primer	Sequence	Product size (bp)
EGF	*celF*	celFWF	GTCCGCATCTGGCTGTGTA	3053
		celFWR	CTTGCCGACCTTGATACCC	
		celFIF	CAAGAACGGTGGCGAATC	186
		celFIR	CGGGTGTTGTCCCAGTAGAG	
XynD	*xynD*	xynDWF	GCCCGCATGACGTACTTT	2505
		xynDWR	GTGCAGCAGCCAATAAACCT	
		xynDIF	GGCAAGAACGATGTGACCTT	200
		xynDIR	TGTCCTTGCGGTAGTCACTG	
Cel3	*cel3*	cel3WF	CATAAAACCGACCCCCAAAT	2156
		cel3WR	ATTGCGCCATTCCTGTTACT	
		cel3IF	AGCGATGGTAAGGTCACTGC	240
		cel3IR	GTGGATGGTGGCGTAGTCC	

Table 2.2 Comparison of RNA yield and quality isolated from the rumen fluid (RF) by using different isolation procedures.

Protocols	Total RNA yield (µg (ml RF)$^{-1}$)	Agilent bioanalyzer analysis		Absorbance ratio	
		LSU/SSU rRNA ratio	RIN value	A_{260}/A_{280}	A_{260}/A_{230}
Method K	38.0	0.7	4.9	1.78	1.90
	39.0	0.8	4.9	1.89	1.82
Method B	140	1.4	8.1	1.99	2.03
	163	1.2	8.2	1.86	1.82
LRCI without column	232	1.6	9.4	2.02	1.95
	201	1.7	9.4	1.88	2.06
LRCI	182	1.8	10	1.91	1.89
	162	1.8	9.5	2.05	2.10

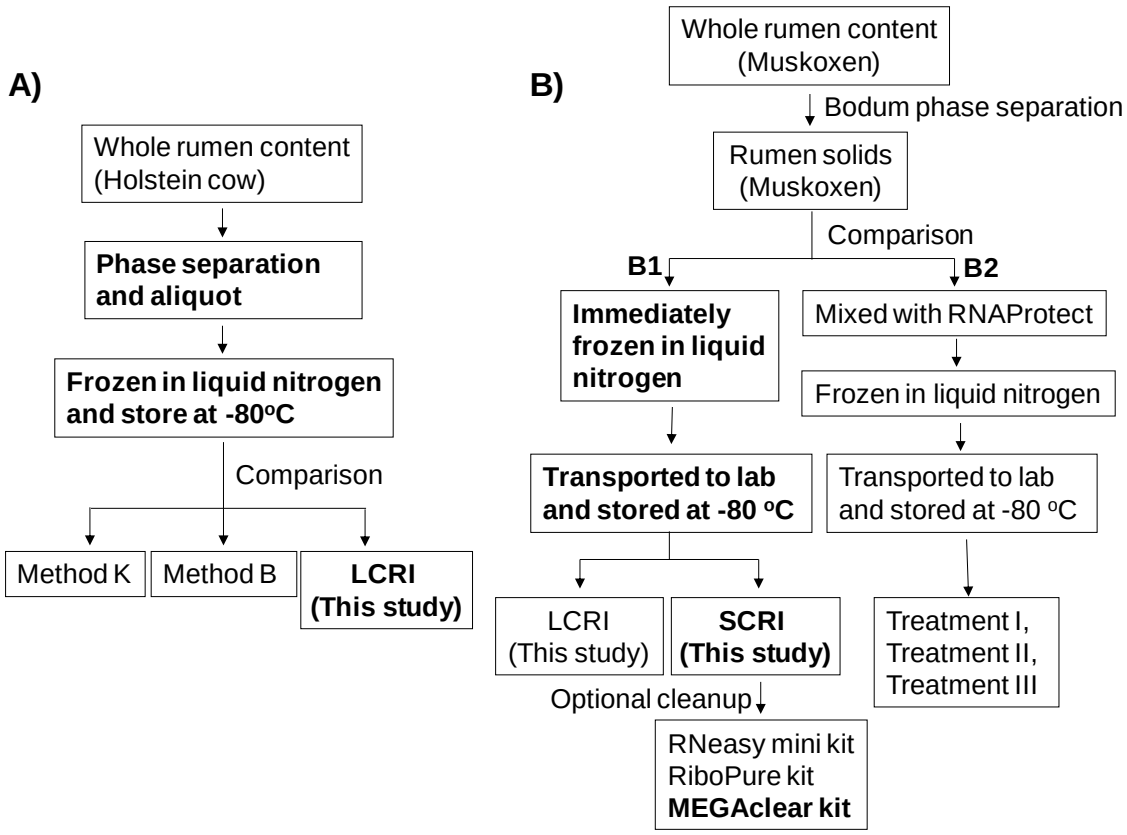

Figure 2.1 RNA isolation experimental flowchart.

Bold indicates the optimal procedural steps established in this study.

LRCI, Liquid Ruminal Contents Isolation;

SRCI, Solid Ruminal Contents Isolation.

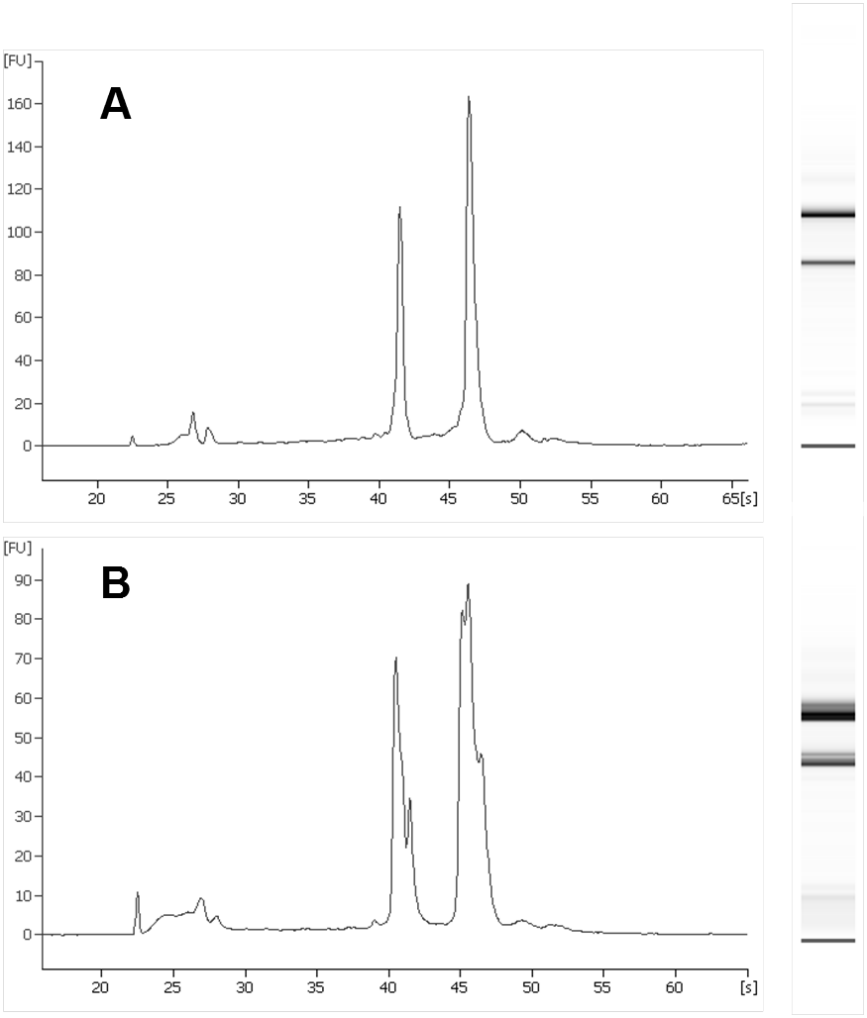

Figure 2.2 Analysis of total RNA integrity extracted from liquid phase and solid phase of ruminal contents.

A: Total RNA extracted from liquid phase rumen contents from a cow using LRCI (Liquid Ruminal Contents Isolation, LSU/SSU ratio: 1.8; RIN (RNA integrity number): 9.5)

B: Total RNA extracted from solid phase rumen contents from a muskoxen fed triticale straw using SRCI (Solid Ruminal Contents Isolation, LSU/SSU ratio: 1.9; RIN: 9.8)

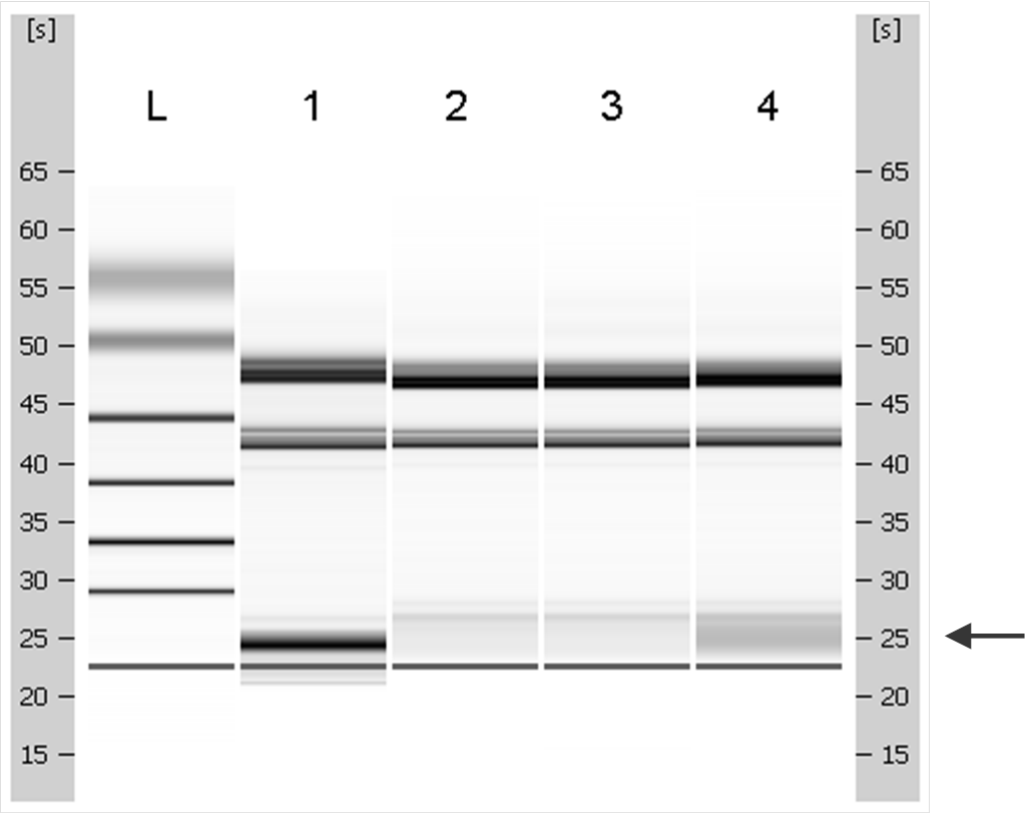

Figure 2.3 Comparison of column purification of RNA.

Total RNA was isolated from solid rumen contents either without column purification or purified using three different commercial kits as described in section 2.2.3. RNAprotect bacteria reagent was not included in the extraction. RNA was analyzed using an Agilent 2100 bioanalyzer and RNA6000 nano chip. The gel pherogram was generated with the 2100 Expert software. The arrow corresponds to the 5S RNA region.

L: RNA ladder (RNA 6000 Nano kit);

Lane 1: Total RNA without column purification;

Lane 2: Total RNA purified with Ribopure kit (Ambion);

Lane 3: Total RNA purified with MEGAclear kit (Ambion), i.e., SRCI (Solid Ruminal Contents Isolation) procedure;

Lane 4: Total RNA purified with RNeasy mini kit (Qiagen).

	SRCI (without RNAprotect)		RNAprotect added					
			Treatment I		Treatment II		Treatment III	
RNA conc. ($\mu g \cdot ml^{-1}$)	**210**	**214**	74	95	53	87	120	87
RNA yield ($\mu g \cdot (g RC)^{-1}$)	**84.0**	**85.6**	59.2	76.0	42.4	69.6	96.0	69.6
LS/SS rRNA ratio	**1.7**	**1.7**	0.8	0.9	0.8	0.9	0.8	0.2
RIN	**8.2**	**5.9**	8.5	8.6	8.5	8.4	8.5	NA

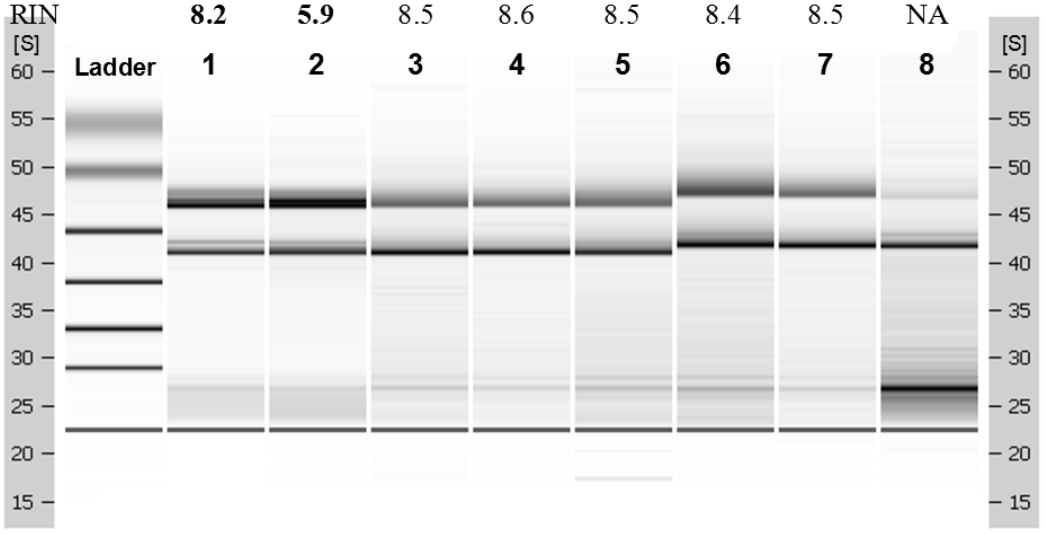

Figure 2.4 Effect of different RNAprotect treatments as described in section 2.2.2.3, on total RNA yield and quality from rumen solids (RS), and comparison with SRCI (Solid Ruminal Contents Isolation) results.

For each treatment there were two samples, hence two lanes per treatment.

Lane Ladder: the RNA ladder (RNA 6000 Nano kit).

NA: the RIN (RNA integrity number) was not provided by the software.

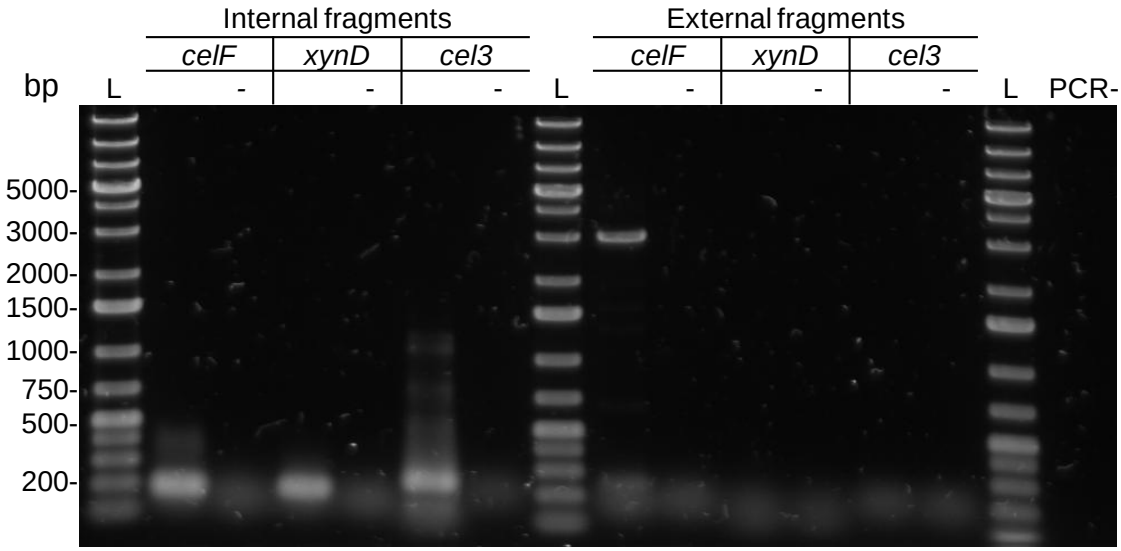

Figure 2.5 RT-PCR amplification of three *Fibrobacter succinogenes* S85 transcripts.

L: DNA ladder;

–: gene-specific negative controls (without addition of reverse transcriptase);

PCR–: PCR negative control.

Chapter 3 Snapshot of the eukaryotic gene expression in muskoxen rumen – a metatranscriptomic approach[†]

3.1 Introduction

Within the gastrointestinal tract of herbivores a diverse group of anaerobic microorganisms, including bacteria, archaea and eukaryotes, produces a vast array of lignocellulolytic enzymes that in turn digest complex plant cell wall polysaccharides and ferment the released simple sugars. The resulting volatile fatty acids and microbial protein are a source of carbon, nitrogen and energy for the host (Flint, 1997; Russell and Rychlik, 2001). Substantial efforts have been made to understand polysaccharide digestion within the rumen through isolation and identification of cellulolytic species, characterization of their enzymes (Krause et al., 2003), and sequencing the genomes of the major culturable rumen bacteria (Berg Miller et al., 2009; Cai et al., 2010; Flint et al., 2008; Kelly et al., 2010; Morrison et al., 2010; Purushe et al., 2010). The recent introduction of massively parallel sequencing technologies has enabled the sequencing of herbivore gut microbiomes, including the foreguts of cattle and wallabies (Brulc et al., 2009; Hess et al., 2011; Pope et al., 2010) and the hindgut of termites (Warnecke et al.,

[†] This chapter is an adapted version of the manuscript "Qi, M.[1], Wang, P.[1], O'Toole, N., Barboza, P.S., Ungerfeld, E., Leigh, M.B., Selinger, L.B., Butler, G., Tsang, A., McAllister, T.A., and Forster, R.J. (2011). Snapshot of the Eukaryotic Gene Expression in Muskoxen Rumen—A Metatranscriptomic Approach. PLoS One **6**(5): e20521".

[1] These authors contributed equally to this manuscript.

2007). These studies have led to the identification of novel cellulolytic enzymes, many of which quite likely arise from microbes that are unculturable in the laboratory (Flint et al., 2008).

Despite the prolific activity directed at understanding the rumen microbiome, there is a distinct lack of information about the eukaryotic component of the rumen metagenome and no rumen fungal or protozoal genomes have been reported. Only a small portion of putative genes described in previous metagenomic studies were attributed to eukaryotes (Brulc et al., 2009; Hess et al., 2011; Pope et al., 2010), although the role of anaerobic fungi (*Neocallimastigomycota*) and rumen protozoa (*Litostomatea*) in rumen cellulose digestion is widely recognized (Orpin and Joblin, 1997; Williams and Coleman, 1997). The paucity of genomic information about anaerobic eukaryotes in the rumen is likely a consequence of 1) the low abundance of eukaryotic DNA in the rumen metagenome; 2) the inadvertent exclusion of eukaryotic species by sample preparation methods; and 3) the failure of bioinformatic analyses to annotate novel eukaryotic gene sequences.

Rumen anaerobic fungi not only produce highly active fibrolytic enzymes targeting the plant cell walls, but they also physically disrupt plant cell walls including the cuticle via penetrating rhizoids. Zoospores, the mobile phase of the fungal life cycle, also preferentially colonize lignin-rich regions of the plant cell wall and upon germination, solubilize these regions to a greater extent than rumen bacteria. Studies have shown that rumen fungi may account for up to 8~20% of the total rumen microbial biomass in ruminants consuming forage (Orpin and Joblin, 1997; Rezaeian et al., 2004). A recent

study demonstrated that anaerobic fungi are widely distributed in both ruminant and non-ruminant herbivores (Liggenstoffer et al., 2010).

The effects of rumen protozoa on fiber digestion are less clear. Previous studies are inconsistent and reports on the effects of protozoa range from a 50% increase to a 15% decrease in fiber digestion (for review, see Williams and Coleman, 1997). Studies on the contribution of rumen protozoa to fiber degradation have also been hampered by the difficulty in cultivating protozoa *in vitro* without the presence of associated bacteria. However, glycoside hydrolase activities and genes have been identified from these organisms (Béra-Maillet et al., 2005; Devillard et al., 1999).

Identification of potent lignocellulolytic and other carbohydrate active enzymes are of great interest for industrial processes, such as cellulosic ethanol production (Percival Zhang et al., 2006; Tilman et al., 2009). In the present study we applied mRNA-Seq technology (Wang et al., 2009) to target the polyadenylated eukaryotic mRNA from the muskoxen (*Ovibos moschatus*) rumen microbial consortium. Muskoxen are arctic ruminants that live primarily in northern Canada, Alaska and Greenland. For the majority of the year, their food sources are limited to forages high in lignocellulose content, due to the very short arctic summer (Barboza et al., 2006). Consequently, we speculated that muskoxen have evolved to harbour a microbial community that efficiently degrades plant cell wall fiber (Crater et al., 2007). The sampled muskoxen were maintained at an isolated wildlife research facility (R.G. White Large Animal Research Station, University of Alaska Fairbanks, Fairbanks, AK) and had not been in contact with domestic ruminants. It was hypothesized that the metatranscriptome approach would lead to the identification of genes coding for novel enzymes and also, for the first time, provide

information on the expression of carbohydrate active enzymes by the eukaryotic community in the rumen.

3.2 Materials and Methods

3.2.1 Ethics statement

The animals were cared for according to procedures that were approved under protocol No. #139821-2 by the Institutional Animal Care and Use Committee at the University of Alaska, Fairbanks, AK.

3.2.2 Muskoxen rumen sampling

Ruminal samples were obtained from two cannulated muskoxen at the University of Alaska (Fairbanks, AK) in September, 2009. The muskoxen were mature male castrates with mean body mass between 245 and 271 kg. During a 1-month period, the muskoxen were fed a triticale (× *Triticosecale*) straw or a brome grass hay (*Bromus* sp.) based high fiber diet, offered twice daily, plus a small amount of protein and mineral supplement once in the morning (335g d^{-1}; Table 3.1). During the last week of the period, rumen samples were obtained in the morning, before the muskoxen were fed. The ruminal contents were transferred to a heavy walled 250 ml beaker and the solid and liquid phases were separated using a Bodum coffee filter plunger (Bodum Inc., Triengen, Switzerland). Subsamples of solid digesta (~ 2.5 g) were immediately flash-frozen in liquid nitrogen. All samples were frozen within 5 min of the sample being withdrawn from the animal. Samples were immediately transferred to the lab, and stored at -80 ℃ until further processing.

3.2.3 RNA extraction and purification

Total RNA was isolated from rumen solids according to the method established in Chapter 2. The quality of total RNA was estimated by running the samples on RNA 6000 nano chip on an Agilent 2100 BioAnalyzer.

3.2.4 RNA sequencing and sequence assembly

mRNA-Seq libraries were constructed from 100 µg of total RNA using the Illumina mRNA-Seq sample preparation kit according to the manufacturer's instructions (Illumina Inc, San Diego, USA). Two samples from two individual muskoxen (one fed triticale straw, one fed brome grass hay) were combined and sequenced using the Illumina Genome Analyzer II system at the McGill University/Genome Quebec Innovation Centre. Obtained sequencing reads were assembled *de novo* using a combination of Velvet (available at http://www.ebi.ac.uk/~zerbino/velvet/) (Zerbino and Birney, 2008) and CAP3 (available at http://pbil.univ-lyon1.fr/cap3.php) (Huang and Madan, 1999) programs. The initial RNA-seq data set in fastq format was split into 14 separate data sets of 2 million reads each. On each of these split data sets, the Velvet suite of programs was run with three different k-mer parameters; k=37, k=45 and k=53. The resultant files were then concatenated into a single contigs file. The program CAP3 was then run on this file, with default parameters. The files with extensions .contigs and .singlets were concatenated into a single file which represented the assembled transcript contigs in the present study. This method of transcript assembly was selected following extensive experimentation and produces more long contigs containing full length transcripts when compared with traditional assembly methods which are more suited to DNA reads. The

contigs were then clustered together at 95% sequence identity over 80% of their lengths using the DNA version of CD-HIT in the Rammcap package (Li, 2009).

To validate contig assemblies, 20 contigs containing different putative carbohydrate active enzyme genes were amplified by reverse transcription PCR using primers designed specifically for each contig. The target contigs and primer sequences are listed in Table 3.2. Briefly, the muskoxen total RNA samples were reverse transcribed using a Superscript III kit and oligo-dT primers (Invitrogen). PCRs were carried out using Platinum *Taq* polymerase Hi Fidelity (Invitrogen) using the conditions recommended by the manufacturer.

3.2.5 Sequence analysis

All sequence analyses, unless otherwise specified were performed using both the reads and the assembled contigs. The databases employed for this analysis were the latest versions available during the analysis period (Jun 2010 to Dec 2010). The genome sequence of the rumen bacterium *Fibrobacter succinogenes* S85 (Accession No. NC_013410), expressed sequence tags (ESTs) of rumen protozoa (Ricard et al., 2006) and rumen fungi *Pyromyces* sp. E2 were retrieved from National Center for Biotechnology Information (NCBI) Genbank databases (http://www.ncbi.nlm.nih.gov/).

3.2.6 rRNA sequence identification

Ribosomal RNA sequences were firstly identified by BLASTN searches against LSU and SSU ribosomal RNA databases (Version 104) from the ARB-Silva database (Pruesse et al., 2007). Subsequently, all the sequences were further analyzed by the rRNA-hmm (Huang et al., 2009) and tRNA (transfer RNA)-scan (Lowe and Eddy, 1997)

programs in the Rammcap package (Li, 2009) using the default settings. The sequences that had E-values equal or less than 10^{-5} (bit score $\geq$ 52) and overlap $\geq$ 50bp to entries in the SSU/LSU database, as well as those identified by rRNA-hmm program (SSU rRNA, LSU rRNA and tRNA) are referred to as non-coding RNA (ncRNA) in this paper.

3.2.7 Binning

The functional based taxonomic assignment was constructed by the Metagenomic Analyzer (MEGAN) software (Huson et al., 2007) based upon the best BLASTX hit to an in-house database named as NRMO database (trimmed down non-redundant muskoxen amino acid database). The NRMO database contained all protein sequences in the Genbank non-redundant amino acid (nr) database that had a match to any of our assembled contigs, with an E-value no greater than 10^{-5}. There were about 230,000 entries in the NRMO database. To validate the NRMO database, 20,000 reads were randomly picked and BLASTed against both the nr database and NRMO database respectively, and compared. The results, especially for the taxonomy distribution at genus level, were very similar to each other (data not shown). Collector's curves were produced from an *ad hoc* Perl script and plotted in Microsoft Excel version 2003.

For the taxonomic community composition analysis based on the amino acid sequences of actin and elongation factor 1 (EF1), related sequences were retrieved from the nr database. Because related protein sequences from rumen fungal and/or rumen protozoa species were not present in the nr database, actin sequences of *Piromyces* E2 ESTs (Genbank Accession numbers: GT912769, GT909886, and GT912949) and EF1 sequences from Rumen protozoa ESTs (Genbanck Accession numbers AM051945, AM053457, AM051946, AM054620, AM051759, and AM053167) were retrieved. The

translated EST sequences were merged into Genbank nr derived sequences. BLASTX searches against each of the two protein family databases were carried out using the non-rRNA reads as the query. The results were analyzed by MEGAN with a bit score cut-off of 52. The taxonomic composition was also estimated by running the software MLTreeMap on the assembled contigs as described (Stark et al., 2010).

Putative full-length ORFs were identified as follows: The assembled contigs were BLASTX-ed against the UniProt database. Contigs were then translated in the proper reading frame based on the BLASTX hits. The resulting amino acid sequences were searched for all full length ORFs of at least 70 amino acids which fully encompass the alignment of the BLAST Hits.

3.2.8 Functional annotation of the coding RNA sequences

The coding RNA sequences were searched using RPS-BLAST against both the KOG and the COG databases and the Genbank nr database. Bacteria-like reads identified by nr BLASTX were further searched against the COG database. The functional roles of the sequences were assigned based on the KOG and COG searches. The matches that had E-values equal or less than 10^{-5} were considered significant.

3.2.9 Carbohydrate active protein annotation

Lignocellulolytic gene containing reads and contigs were identified and classified based on CAZy database (Cantarel et al., 2009) as described by Warnecke and colleagues (Warnecke et al., 2007), with the following modifications. Both HMMER3 (Eddy, 2009) and BLASTX searches were carried out as follows: Step A) Glycoside hydrolase and carbohydrate binding module (CBM) families that have associated Pfam HMMs (v24.0)

(Pope et al., 2010) were used directly for HMMER hmmsearch. Step B) In an attempt to associate Pfam HMMs to CAZy families without such models, all members of these CAZy families were searched against the Pfam-A and Pfam-B databases (v24.0). Results were manually checked and Pfam HMMs were conservatively chosen for a CAZy family only when the following two criteria were met: i) all hits to that Pfam group were from the same CAZy family; ii) At least 80% of group members were identified to conform to the conserved Pfam model. In instances where one Pfam HMM model represented members from two or more closely related CAZy groups, a class of combined CAZy groups was assigned. Step C) For those CAZy families that currently are not represented by a Pfam model, the representative sequences as described by Warnecke et al. (2007) were used in BLASTX searches with a score cut-off of 52. Step D) For CAZy families with neither a Pfam accession nor representative amino acid sequences, an HMM profile was built based on T-coffee (Notredame et al., 2000) alignment of representative members selected from the CAZy web site and used for HMMER3.

3.2.10 Glycoside hydrolase cluster analysis

To compare putative genes coding for carbohydrate active proteins identified in the muskoxen rumen metatranscriptome with other genomes/metagenomes, the percentages of glycoside hydrolase families were calculated. A two-dimensional matrix was constructed, consisting of the GHs that were identified from genomes or metagenomes, wherein each cell in the matrix indicated how often a GH family was seen within a particular sample. Pearson correlation coefficients of each two samples were calculated and transformed into distances and clustered by using the unweighted pair group method

with the arithmetic mean algorithm as previously described (Garcia-Vallve et al., 2000; Qi et al., 2005).

3.2.11 Sequence Data Availability

The sequencing reads are available from the NCBI short read archive under Accession number SRA030623.1.

3.3 Results

3.3.1 General sequence statistics

In the present study, we adopted a metatranscriptomic approach to identify enzymes from the muskoxen rumen microbial community. We used samples from muskoxen fed triticale straw and brome hay for deep sequencing with the goal to obtain transcripts encoding carbohydrate-active enzymes. Total RNA was extracted from rumen solids to ensure isolation of the cellulolytic microbial biofilm as well as RNA from microbes deeply embedded in the plant fiber. After purification, the eukaryotic mRNA was sequenced using the Illumina Genome Analyzer II platform. This approach resulted in a total of 25,900,806 reads, with an average read length of 108 nt, generating a total of 2.8 gigabases of sequence data (Figure 3.1).

Raw sequencing reads were assembled into a total of 59,129 contigs with an average length of 310 bp, for a total of 19 M base pairs. The maximum length of the contigs was 13,498 bp, which contained a single open reading frame of 13,083 bp. It encoded for a 4,354-amino-acid protein that showed 28% identity to a hypothetical protein from a strain of *Orpinomyces* (Nicholson et al., 2005). The size distribution of all the contigs is illustrated (Figure 3.2). Over 12,000 contigs were represented by 100 or

more reads, including 2,551 contigs represented by more than 1000 reads and 545 represented by more than 5,000 reads. To validate the contig assembly, twenty glycoside hydrolase related contigs (≥500 bp) were chosen and primers designed to amplify the target (Table 3.2). Nineteen of the twenty contigs were successfully amplified using at least one set of primers.

Using BLASTN searches against the Silva LSU and SSU ribosomal RNA databases coupled with rRNA-HMM and tRNA-scan, we identified 4.77 million nc RNA reads or 18.4% of the total (Figure 3.1). During the RNA sample preparation step, oligo-dT coated magnetic beads were used to remove a large proportion of rRNA. Assuming ncRNA account for 95% of the original total RNA (Neidhardt and Umbarger, 1996), approximately 99% of this amount was successfully removed as indicated by the percentage of the ncRNA reads identified.

The average GC content of ncRNA reads was 51%. In contrast, the GC content of all potential protein encoding RNA reads was 39.2%, a value that was significantly lower than the ncRNA reads and much lower than those reported in metagenomic studies of other microbial communities associated with ocean, fresh water and various animal environments (average GC% is 49.56 ± 4.9% (Willner et al., 2009)). The average GC content of the assembled contigs was 37.9%, which were also lower than other metagenomic studies (Figure 3.3).

BLASTX searches using the 21 million potential protein encoding reads against the NRMO database indicated 7.8 million reads had at least one significant match (E-value ≤ 10^{-5}) (Figure 3.1). RPS-BLAST searches against the euKaryotic Orthologous Groups (KOG) and the Cluster of Orthologous Groups (COG) identified 6.0 million reads that

may have conserved domains, of which 0.9 million reads were not identified by BLASTX searches. The remaining reads (48% of all reads) did not show any similarity to any of the above mentioned databases.

The first BLASTX match was used to estimate the possible origin of each putative protein coding RNA reads, according to MEGAN analysis (Figure 3.4). About 6.6 million reads (31%) showed highest similarity to proteins from the Eukaryota kingdom. Among these, rumen anaerobic fungi, which belong to the phylum Neocallimastigomycota (Hibbett et al., 2007), were represented by 1.4 million reads. Rumen ciliate protozoa, which belong to the Litostomatea class of the Alveolata group, were represented by 1.1 million reads. At the genus level, the most represented genera were known rumen/anaerobic species, including *Entodinium, Piromyces, Neocallimastix, Trichomonas, Orpinomyces, Entamoeba,* and *Epidinium*, that were represented by over 100,000 reads each (Figure 3.5). A total of 0.7 million reads (3.4%) and 0.02 million reads (0.1%) were binned to Bacteria and Archaea kingdoms, respectively.

3.3.2 Analysis of sequencing coverage

The sequencing coverage was first assessed by looking at the matches to rumen eukaryotic proteins that were present in the nr database. Rumen anaerobic fungal protein sequences (total of 257) were obtained from the Genbank database as of July, 2010. TBLASTN searches were performed using all of these proteins as queries, and matches to 220 of them (with Expect value $\leq$ 1E-20) were found in the muskoxen contigs. Another similar search identified 104 of 107 published rumen protozoal protein sequences in our dataset. The identification of almost all known rumen fungal and protozoal sequences demonstrated the comprehensive coverage of the current sequencing approach.

The coverage was further evaluated by plotting collector's curves based on the number of functional gene categories (matched to the KOG database) and gene accession numbers identified (matched to the NRMO database), as a function of the number of reads (Figure 3.6). Saturating coverage was found for both curves, as roughly 80% of the total richness was found at a point of less than 14% of the sequencing effort.

3.3.3 Functional analysis of the putative protein-coding reads

Based on the RPS-BLAST search results, 6.0 million reads could be assigned to a cluster of the KOG/COG databases. Most of the assignable sequences belonged to the "Translation, ribosomal structure and biogenesis" cluster (45% of all the assigned sequences), while Cytoskeleton (16%) was the second largest cluster (Figure 3.7). Correspondingly, 9 of the top 10 KOG/COGs also belonged to these two categories, with KOG0676 (Actin and related proteins, which were represented by 551,126 reads) and KOG0052 (Translation elongation factor one, 230,087 reads) as the first two abundant KOG groups (Figure 3.8). These results indicate that these groups of genes were actively transcribed. About 18% of the reads that matched the KOG/COG databases were involved in metabolism, including metabolism of carbohydrate (7.9%), energy conversion (4.2%) and metabolism of amino acids (2.6%) (Figure 3.7). KOG/COG groups involved in glycolysis and pyruvate metabolism accounted for the highest read numbers (Table 3.3, Figure 3.9), demonstrating genes belonging to those clusters had a central role in metabolism.

Hydrogenosomes are membrane-bound organelles present in anaerobic eukaryotes including rumen fungi and protozoa (Boxma et al., 2005; Williams, 1986). Hydrogenosomes are distantly related to mitochondria and are the centre of ATP and

hydrogen generation in these microorganisms (Boxma et al., 2005). In the present rumen metatranscriptome dataset, KOG/COGs linking with hydrogenosomes were also identified, including iron-only hydrogenase, malic enzyme, pyruvate:formate lyase, succinyl-CoA synthetase and acetate:succinate CoA-transferase. Combining the highly expressed KOG/COGs, a representative energy metabolism pathway of muskoxen rumen eukaryotes was reconstructed (Figure 3.9).

3.3.4 Lignocellulolytic enzymes

Sequence homology based enzyme annotation was biased toward the identification of known enzymes that were already present in the database. To minimize this bias, we used a more sensitive Pfam-HMM search to identify CAZyme modules (Table 3.4, Table 3.5, Figure 3.10). In all, these analyses identified over 400,000 reads potentially encoding lignocellulolytic enzyme modules, spanning about 110 CAZy families. The read number in each family gives an indication of the expression status of that group. However, it needs to be pointed out that the number of reads that matched to lignocellulolytic enzymes was likely underestimated as the short reads (108 nt, translated into 36 amino acid residues maximum) result in a less robust database search score than that obtained using the full sequence of complete genes.

To circumvent this problem, similar analyses were also performed on the assembled contigs. In total, we identified over 2,500 contigs with a significant match to at least one CAZy module (Table 3.5). Since the use of short contigs may overestimate the total number of enzymes, we further restricted our targets to those contigs longer than 500 base pairs (1,082 in total, Table 3.5, Table S.1, Table S.2). These contigs, especially those assembled from larger numbers of reads, could serve as good candidates of

potentially useful lignocellulolytic enzymes. Only 17% of these contigs were more than 70% identical to sequences in the nr database, while 46% of them had less than 50% identity (Figure 3.11). Seventeen percent of the CAZy module coding sequences identified were most similar to nr database sequences annotated as "(conserved) hypothetical protein", "predicted protein" or "unnamed protein product". These results indicate that rumen eukaryotes produce a large variety of CAZymes with many of them remaining uncharacterized. There were 242 contigs that had two or more distinct CAZy domains. The proposed fungal dockerin CBM10 modules were predominant in these predicted multi domain enzymes, and were found in 190 (78.5%) of these contigs. Glycoside hydrolases from families GH6, GH45, and GH48 were found in 25, 25 and 23 multi-domain contigs respectively, most of which were linked to CBM10 modules.

3.3.4.1 Catalytic modules

Most of the muskoxen rumen microbiome cellulases identified were classified as families GH5, 6, 9, 45 and 48, which were represented by 30 to 51 contigs. Similar to other rumen metagenomic studies (Brulc et al., 2009; Hess et al., 2011; Pope et al., 2010), no contigs showed similarity to family GH7 or GH44. In this study, putative swollenin modules were detected in 16 contigs. Swollenin was reported to dissociate cellulose fiber with no hydrolytic activity (Saloheimo et al., 2002) and has not been previously reported to be associated with anaerobic microorganisms. Hemicellulose degrading enzymes from GH8, GH10, GH11, GH26 and GH53 were also identified. GH10 and GH11 were the predominant families that were represented by 29 and 33 contigs, respectively. Carbohydrate esterases remove the ester bond on the xylan backbone, exposing it to glycoside hydrolases. There were 111 contigs showing similarity to carbohydrate

esterases in the CAZy database. The Family CE4 family was the largest family with 50 contigs. The CE1 family, encoding feruloyl esterases important for lignin solubilization in the rumen, was represented by 13 contigs. Twenty seven contigs were annotated to be polysaccharide lyases (Table 3.5).

GH family profiles recovered from the metatranscriptome of muskoxen rumen were compared to the metagenomes of the termite hindgut (Warnecke et al., 2007), wallaby foregut (Pope et al., 2010) and bovine rumen (Brulc et al., 2009; Hess et al., 2011), as well as the genomes of several anaerobic bacteria and expressed sequence tags (ESTs) of the rumen fungi *Piromyces* sp. E2 (Table 3.5, Figure 3.10). Clustering analysis of the GH family distribution (Figure 3.10) showed that the muskoxen metatranscriptome was most closely related to anaerobic cellulolytic bacteria and the rumen fungus *Piromyces*. There were some major differences between putative genes identified by our study and previous metagenome sequencing studies. Approximately 28% of identified GHs were cellulases in the muskoxen rumen transcriptome, as compared to 8.5% or less in reported rumen metagenome studies. A large number of genes potentially encoding GH6, GH48 and swollenin enzymes identified in the present study were rarely found in other studies of rumen metagenomes. Conversely GH29, 35, 39 and 42 family members described in other rumen metagenome studies (Brulc et al., 2009; Hess et al., 2011; Pope et al., 2010) were not identified in the rumen metatranscriptome of muskoxen (Figure 3.10).

3.3.4.2 Accessory modules

The most abundant accessory module was CBM10, which was identified from 403 contigs longer than 500 bp (Table 3.4). CBM10 has been shown to be associated with many rumen fungal glycoside hydrolases. It was once proposed to be a eukaryotic

counterpart of bacterial cellulosomal dockerin (Steenbakkers et al., 2001), but a recent study (Nagy et al., 2007) suggested that the structure of this domain differs from bacterial dockerins. Dockerin containing cellulases are known to interact with cellulosome scaffolding proteins in bacteria to form the cellulosome structure. However, a cellulosome scaffolding protein has yet to be identified from rumen fungi. The exact function of CBM10 modules remain to be explored.

CBM18, known as a chitin binding module, was the second largest CBM group, and was identified from 108 contigs. Other major CBMs identified included CBM1, CBM6 and CBM13, that presented in 33, 27 and 31 contigs, respectively. Accessory modules that are commonly found in bacterial cellulases, such as bacterial cellulosome dockerin, cohesin, S-layer homolog domain (SLH), cellulase N-terminal immuno-globulin domain (CelD_N) and fibronectin-3 (fn3) modules, were not found in the contigs or reads.

3.3.4.3 Lignocellulolytic gene expression

Both gene diversity and gene expression information can be obtained from metatranscriptomic sequencing analysis. To demonstrate the latter, we summarized the read numbers associated with contigs/genes (Table 3.6, Table S.1, Table S.2). Cellulase and xylanase coding sequences in GH families 6, 11, 45 and 48 were over-represented compared to other families (Table 3.6). Putative GH1 genes, which encode oligosaccharide degrading enzymes, were also over-represented. In addition, a putative family 6 polysaccharide lyase, which has never been reported from eukaryotes, was represented by over 20 k reads.

3.4 Discussion

Cellulase activities are well known to be present in microbial communities found in soil, compost, and herbivore digestive tracts including the rumen. Metagenomic technology coupled with high throughput sequencing has enabled the identification of thousands of sequences encoding for enzymes that degrade plant cell walls (Allgaier et al., 2010; Brulc et al., 2009; Hess et al., 2011). In the present study, we used rumen solid samples from muskoxen fed a highly lignified diet of triticale straw or brome grass hay. Rumen solids are reportedly responsible for 80~90% of the endoglucanase and xylanase activities in the rumen (McAllister et al., 1994; Williams and Strachan, 1984), as attachment and the formation of digestive microbial biofilms is a prerequisite for the ruminal digestion of feed (McAllister et al., 1994). By applying an improved isolation method, high quality RNA was extracted from rumen solids (Chapter 2). Combined with an efficient method of *de novo* assembly of short sequence reads, the present study has provided a comprehensive catalogue of eukaryotic cellulolytic enzymes in the muskoxen rumen, many of which are supported by full-length cDNA information.

To our knowledge, we are the first group to report the whole eukaryotic transcriptome of a rumen microbial community. Metatranscriptomic studies have been previously carried out in various microbiomes focused on marine microbial communities (Frias-Lopez et al., 2008; Gifford et al., 2011; Gilbert et al., 2008; Poretsky et al., 2009), and the gut microbial population of the piglet (Poroyko et al., 2010). Most of these studies used Roche 454 pyrosequencing technology and although the read lengths were longer than that obtained with Illumina sequencing, far fewer total sequences were produced. In fact most of these studies produced less than 500 M bp of sequences and

sample-sequencing depth was low (Gifford et al., 2011). Illumina sequencing in the present study generated 2.8 gigabases of sequencing data, which is at least 6 times that of previous metatranscriptomic studies (Frias-Lopez et al., 2008; Gifford et al., 2011; Gilbert et al., 2008; Poretsky et al., 2009; Poroyko et al., 2010). The deep sequencing coverage obtained in this studyis attested by the observation that the slopes of the collector's curves reached a plateau and that most of the rumen eukaryotic sequences reported in Genbank were identified in the metatranscriptome (331 out of 364 sequences).

Metatranscriptomics has distinct properties when compared to metagenomics, the first being that metatranscriptomic analysis identifies most extensively transcribed genes while metagenomic sequencing identifies the most numerically dominant genes. For example, *Prevotella* is a group of predominant rumen bacteria that at times account for as much as 60% of the total bacteria in the rumen (Kong et al., 2010; Stevenson and Weimer, 2007), even though they play no active role in the digestion of recalcitrant cellulose (Purushe et al., 2010). Indeed, the GH profiles of *Prevotella ruminicola* clustered closely to those of the wallaby and bovine rumen metagenomes, with GHs involved in the degradation of oligosaccharides and hemicelluloses being highly represented, whereas the proportion of GHs related to cellulase were much lower (Figure 3.10). If a gene encoding a GH was present in a microbial species of low rumen abundance, even if highly expressed, it would be unlikely to be recovered by metagenomic sequencing. Such a scenario may be applicable to the rumen anaerobic fungi as they are reported to account for no more than 8% of the total rumen microbial biomass (Orpin and Joblin, 1997).

Secondly, metatranscriptomic analysis may provide insight into the degree of gene expression (Table 3.6), which could help focus gene mining towards those enzymes that

are most active in plant cell wall degradation. Consequently, the yield of potential gene targets for further development may be far higher with metatranscriptomic than metagenomic approaches. A recent rumen metagenomics study using Illumina sequencing technology generated 268 gigabases of metagenomic DNA with about 103 putative carbohydrate active enzymes identified per gigabase (Hess et al., 2011). In our present study, we were able to identify 2500 candidates in 2.8 gigabases of RNA sequence, translating to 893 putative carbohydrate active enzymes per gigabase.

When muskoxen are fed on high fiber diets, cellulolytic microorganisms attach to and penetrate the fiber and expression of many of their cellulolytic enzyme genes is induced. Genes that are highly expressed generate more sequencing reads, increasing sequence coverage, resulting in the assemblage of longer contigs and thereby increases the likelihood of regenerating full length genes or modules. Indeed, of the 59,129 contigs in the present study, there were over 2,500 contigs with lengths over 1.0 kb and 6,022 with lengths between 0.5 to 1 kb (Figure 3.2). Among the contigs longer than 500 bp, over 1,000 of these harboured sequences putatively encoding for carbohydrases. Most putative CAZy genes longer than 500 bp (96%) were represented by 100 or more reads (Table S.1, Table S.2), corresponding to an average sequence coverage of about 139 times. These highly expressed putative CAZy genes are likely to play critical roles in the breakdown of plant fiber by rumen eukaryotes and have potential for use in cellulosic biofuel production as well as other industrial processes.

Cellulases from different families often have different substrate specificity and reaction mechanisms. Efficient degradation of the plant cell wall requires synergistic interactions between enzymes with high activity for different substrates (Lynd et al.,

2002). Not surprisingly, we identified cellulases from a wide range of CAZy families. The range of glycoside hydrolases identified showed remarkable differences as compared to previous rumen metagenomic studies. For example, a large number of putative GH6, GH48, and Swollenin genes were identified. All three GH families are important for the degradation of crystalline cellulose, which is the most recalcitrant part of plant cell walls. However, these GH families were not described by earlier metagenomic approaches in the bovine rumen or wallaby foregut (Brulc et al., 2009; Pope et al., 2010). Even in the most recent deep metagenome sequencing study (Hess et al., 2011), only three putative GH48 genes were identified (Hess et al., 2011), while GH6 and Swollenin were not found. In contrast, our study identified 31 GH6, 33 GH48 and 16 Swollenin sequences from rumen eukaryotes and GH6 and GH48 were among the most actively transcribed families (Table 3.5). These data clearly suggest that rumen eukaryotes play an important role in crystalline cellulose digestion through expression of a large amount of exo-glucanases, which were not detected in other rumen metagenomic studies.

The different CAZy families identified by rumen metagenomic studies and our metatranscriptomic study could be due not only to different nucleic acid sequencing targets, but also to the differences in the samples. Muskoxen could have developed a different plant cell wall degrading rumen microflora for survival in the arctic as compared to domesticated cows and sheep. Indeed, our preliminary analysis has identified differences in the microbial population between muskoxen and domesticated cattle (Forster, RJ, personal communication).

Assessment of the combined rumen/gut microbiome sequencing information across studies would suggest that the rumen seems to lack cellulases from GH7 and GH12

families. So far, all members of GH7, a family of exo-glucanases, have been isolated from aerobic fungi. The GH7 activity of releasing cellobiose from the reducing end of the cellulose chain may be undertaken by GH48 cellulases within the rumen. Enzymes in family GH12 have been shown to have endoglucanase and xyloglucases activities (Gloster et al., 2007). Functional aspects of these enzymes may arise from GH74 enzymes in the rumen (Yaoi et al., 2007).

A total of 3.4% reads showed top BLASTX matches to bacteria (Figure 3.4) and over half of these reads exhibited the highest similarity to proteins from bacteria that are known as predominant rumen/gut residents, such as families of *Fibrobacteraceae, Clostridiaceae, Ruminococcaceae, Prevotellaceae, Bacteroidaceae,* and *Lachnospiraceae* (Figure 3.12). However, these bacteria-like reads are very unlikely to come from bacterial mRNA because only enriched polyA RNAs were sequenced, which were rarely found in bacteria mature mRNA. Phylogenetic binning using protein markers also confirmed the absence of bacteria-derived genes in our dataset (Figure 3.13). In addition, the GC content of the "bacteria-like" genes identified from muskoxen transcriptome were about 40.1%, which is also within the range of rumen eukaryotic coding sequences identified, but much lower than the average GC content of the metagenomic studies which represent the bacterial population (Figure 3.3). These findings imply that bacteria-like coding sequences, most of which are involved in a variety of metabolic functions (Figure 3.7) (including 35% of all putative CAZy genes identified), may have been horizontally transferred into the genome of rumen eukaryotes, most likely from rumen bacteria, a possibility that has been previously raised (Garcia-Vallve et al., 2000; Ricard et al., 2006).

Anaerobic fungi and protozoa are the two major groups of eukaryotes in the rumen (Hungate et al., 1964; Orpin, 1975). In the present study, functional based phylogenetic binning (Figure 3.4), top BLAST matches of the CAZy enzymes (Figure 3.14) and phylogenetic analysis based on SSU ribosome RNA sequence (data not shown), all indicated the presence of both groups. There were significantly (χ^2=348, p<0.0001) more CAZy enzymes matching to rumen fungi than to rumen protozoa (Figure 3.14), indicating that rumen fungi may play a more significant role in fiber digestion in the muskoxen rumen. However, since there are more CAZy genes from rumen fungi than from protozoa in the nr database (101 vs 28, Table 3.5), the differences could be due in part to the number of homologues currently present in the database.

Eukaryotic anaerobic microbes are poorly understood, especially from a molecular perspective. Although this study focused primarily on genes encoding enzymes involved in plant cell wall degradation, the data presented greatly expands our current knowledge of these unique eukaryotes and should provide further insight into their co-evolution, metabolism and function within the rumen microbial community.

3.5 Tables and Figures

Table 3.1 Feed composition.

Feed	DM[1] (g g^{-1})	NDF[2] (g (g DM^{-1}))	ADF[3] (g (g DM^{-1}))	Cellulose (g (g DM^{-1}))	Hemicellulose (g (g DM^{-1}))	Lignin (g (g DM^{-1}))	Nitrogen (g (g DM^{-1}))
Hay	0.875	0.699	0.378	0.308	0.321	0.066	0.0112
Straw	0.878	0.837	0.522	0.437	0.315	0.074	0.00673
SEM	1.82×10^{-3}	0.0143	0.0150	0.0175	2.40×10^{-3}	0.0135	0.0464
P =	0.49	0.02	0.02	0.007	0.19	0.71	< 0.001

[1] DM: dry matter.

[2] NDF: neutral detergent fiber.

[3] ADF: acid detergent fiber.

Table 3.2 Primers used for validating lignocellulolytic enzyme related contigs.

Primer Name	Primer sequence	Target Contig	Contig Length	Position	Theoretical Product Size	CAZY Family
30088_62_outerF	ATGGTGGTGATAACAACTCTGG	Contig30088	1072	62	941	CE1
30088_70_outerR	CCCATTCTACCGTCACCTTC			1003		
30088_265_innerF	AGTCTTAAGAGTAACACCACCC			265	423	
30088_385_innerR	AGATCAAAGGCTGATGGAGCAG			688		
29149_62_outerF	TTACCATTACCTTCACCGTGACCTC	Contig29149	1132	62	967	CE1
29149_104_outerR	TTTCCCAGGCGGCGGTATGG			1029		
29149_200_innerF	GCCCATTCAGCTAAGTTACCC			200	718	
29149_215_innerR	GGGTGGATTCACTCAAGATGA			918		
2424_106_outerF	GAGCACCAACACAAGCACTAG	Contig2424	1033	106	832	CE3
2424_96_outerR	GTGGTATGGGTGGTATGTTCG			938		
2424_280_innerF	CTTGATCAGTACCCATATCAGC			280		
111_106_outerF	GGGTTGTACAGTTGAATACACCG	Contig111	1476	106	1255	GH6
111_116_outerR	CCAGCATCTGGAGCACCTTG			1361		
111_315_innerF	GTGGTATTCCAAGCAAATGTGG			315	856	
111_300_innerR	ACCAGGTACCAGAAGCGTTG			1171		
22047_114_outerF	GGCATGGATAGCACAAAGATTG	Contig22047	1921	114	1694	GH6
22047_114_outerR	TGGAGCTGGTTTCATGGCAG			1808		
22047_450_innerF	TGCTCTTGCCGCTAAGGTCTC			450	1172	
22047_300_innerR	ACCTGGGTGCTTACGGTCAG			1622		
30327-O1	GGAAATGGTTCTTGGGGTGTAG	Contig30327	1485	98	1252	CE1
30327-O2C	CCTGGAGTGTTTGCTTTTGG			1350		
30327-I1	GGGGTCAAAACAACCAAG			420	651	
30327-I2c	GACATACCACCACCCATT			1071		
30515-O1	GACAGCATGGGTAACCTA	Contig30515	1129	93	911	CE6
30515-O2c	CTGCTATTCCACCACTTG			1004		
30515-I1	TTCGGCGTCAGTAGTATCTTCC			391	503	
30515-I2c	GCCGTTCCAGGTTGTGATATTC			894		

29571-O1	CTGTCCCACCACTTGCTAATTG	Contig29571	1198	233	912	CE6
29571-O2c	TCGACATCCCACTTACCATCAG			1145		
29571-I1	TGTTGCTGTTGCTGGTTGTG			339	745	
29571-I2c	AACATGGGTAGCCAAGATCC			1084		
1733-O1	CCAAAACCACCACCTGGAAT	Contig1733	1399	256	1018	GH45
1733-O2c	GGTGAATGGGGTGTTGAAAACG			1274		
1733-I1	AGCAGCAGCGTAACCATAGGA			420	538	
1733-I2c	CGGATTGTGTTGCAGCTTGGA			958		
32-O1	GGGCAACAAGCCAAGGTTAC	Contig32	1284	113	998	GH45
32-O2c	GCTTTAAGGCAGCTGGAAGG			1111		
32-I1	TGATGGTAAGTGGGCCATTG			315	635	
32-I2c	CCACCAGTGTTGGTTGTTTG			950		
29098-O1	GGCTGGTAAGGTCTGTAGAG	Contig29098	1484	71	1228	GH48
29098-O2c	CCATGTCACCACCGAAAC			1299		
29098-I1	GGGACAAGTTGAAGACTACC			539	313	
29098-I2c	TGTTGGATGGAGCCTTAC			852		
Node3576-O1	TCCATCTTCACCAGCTACCTATGC	Contig_Node 3576	2067	123	1534	GH48
NODE3576-O2c	CCCAGAATCTGTGGTAACGGAATC			1657		
NODE3576-I1	GCTCGTGCTATTCAAGGTGCTTAC			508	718	
NODE3576-I2c	CTACCGTCCCAAGAAAGGTTTGTG			1226		
2118-I1	TGGGTCTTGGTTGCATAC	Contig2118	2396	866	1026	GH74
2118-I2c	TGCTGGAGGTGATGTTAC			1892		
2118-O1	TTTCTGGGGCTCCATATC			295	1959	
2118-O2c	GTCTCGGGTATTGTTGTC			2254		
24305-O1	CAAATGGCCTACATGGACTGACC	Contig24305	1399	262	1012	GH74
24305-O2c	CACCTGTACCACCTGCTTCTTTC			1274		
24305-I1	TTGCTGGATTGGCCTTCGGAGGAT			516	436	
24305-I2C	GGCGAAAACACCATTTCCTG			952		
23421-O1	CTCCAGCTATTGCCCAATTCG	Contig23421	1355	65	1127	CE3
23421-O2C	CAGCAGCTCCATTACCACAAC			1192		

23421-I1	GGTATGTTCGGTGGAGGTCAAAG			235	824	
23421-I2C	AGCAGTAGTGGCTGGTGGATTAG			1059		
900-O1	TGGCCCAACAGGAAGAATCAAC	Contig900	1425	108	1218	CBM10/ CE15
900-O2C	TTAGCCGATTGGGAATGCAGAC			1326		
900-I1	ACCGAATCTGTCTTCCACTC			490	662	
900-I1c	CCTAGTGGTAATGGTCCATTCC			1152		
28807-O1	GTGGAGCGACTAAAGCAGTAAG	Contig28807	1617	327	1194	CBM10/ CE15
28807-O2c	CTTGGCTACCCATGTTGTG			1521		
28807-I1	CTAAACGGGAACAACCAG			582	625	
28807-I2c	TGCTCCAACTCCAGATAC			1207		
node7061-O2	CTTGCCTTCCAGCTGTTAATGC	Contig Node 7061	960	44	651	DUF297/ CBM10
node7061-O2c	TCGGTTGGGTAACCGTAAAGAG			695		
node7061-I1	CTTATACGGTGCTACCAAGG			168	373	
Node7061-I2c	AACCAGCAGCCATGTTAC			541		
260-I1	GCCAGAGAAGAAGCTAAAGG	Contig260	1229	301	581	DUF297
260-I2c	AGTAGAGGCAGAACCAGAAC			882		
260-O1	CTGGGCTTTAGGTACTAAGG			141	763	
260-O2c	TAGCAGTGGTAGTGGTCTTC			904		
7694-O1	GTGATGCTCGCAATCTCTAC	Contig7694	1435	88	587	PL09
7694-O2C	CCATCAGCATTAGCACCATAGG			675		
7694-I1	CCCATTTCCATGGCTGAAATGC			114	419	
7694-I2C	TTGCTCCACTAACCCAGATACC			533		

Table 3.3 Metabolic related KOG/COG groups represented by 5000 or more reads in the metatranscriptome from muskoxen rumen eukaryotes.

KOG/COG Description	Reads Number	KOG/COG Category
KOG2670, Enolase	60668	Carbohydrate transport and metabolism
KOG0657, Glyceraldehyde 3-phosphate dehydrogenase	58954	Carbohydrate transport and metabolism
KOG0626, Beta-glucosidase, lactase phlorizinhydrolase, and related proteins	45908	Carbohydrate transport and metabolism
COG0574, PpsA, Phosphoenolpyruvate synthase/pyruvate phosphate dikinase	41292	Carbohydrate transport and metabolism
KOG2099, Glycogen phosphorylase	28858	Carbohydrate transport and metabolism
KOG4153, Fructose 1,6-bisphosphate aldolase	24624	Carbohydrate transport and metabolism
KOG1367, 3-phosphoglycerate kinase	24163	Carbohydrate transport and metabolism
COG2211, MelB, Na+/melibiose symporter and related transporters	17403	Carbohydrate transport and metabolism
KOG0625, Phosphoglucomutase	10338	Carbohydrate transport and metabolism
KOG1643, Triosephosphate isomerase	8573	Carbohydrate transport and metabolism
KOG1369, Hexokinase	7768	Carbohydrate transport and metabolism
COG0057, GapA, Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase	6101	Carbohydrate transport and metabolism
COG2115, XylA, Xylose isomerase	5791	Carbohydrate transport and metabolism
KOG0372, Serine/threonine specific protein phosphatase involved in glycogen accumulation, PP2A-related	5660	Carbohydrate transport and metabolism
KOG3749, Phosphoenolpyruvate carboxykinase	66252	Energy production and conversion
KOG1494, NAD-dependent malate dehydrogenase	29529	Energy production and conversion
KOG1255, Succinyl-CoA synthetase, alpha subunit	22008	Energy production and conversion
COG0281, SfcA, Malic enzyme	17260	Energy production and conversion
KOG2799, Succinyl-CoA synthetase, beta subunit	17208	Energy production and conversion
KOG0749, Mitochondrial ADP/ATP carrier proteins	14650	Energy production and conversion
KOG1257, NADP+-dependent malic enzyme	14255	Energy production and conversion
COG1882, PflD, Pyruvate-formate lyase	12898	Energy production and conversion
KOG0232, Vacuolar H+-ATPase V0 sector, subunits c/c'	12765	Energy production and conversion
KOG1352, Vacuolar H+-ATPase V1 sector, subunit A	10826	Energy production and conversion
KOG0453, Aconitase/homoaconitase (aconitase superfamily)	9878	Energy production and conversion

KOG1351, Vacuolar H+-ATPase V1 sector, subunit B	8407	Energy production and conversion
KOG1447, GTP-specific succinyl-CoA synthetase, beta subunit	6336	Energy production and conversion
KOG2189, Vacuolar H+-ATPase V0 sector, subunit a	5029	Energy production and conversion
KOG2250, Glutamate/leucine/phenylalanine/valine dehydrogenases	15125	Amino acid transport and metabolism
KOG2467, Glycine/serine hydroxymethyltransferase	8025	Amino acid transport and metabolism
KOG2448, Dihydroxy-acid dehydratase	7053	Amino acid transport and metabolism
KOG2263, Methionine synthase II (cobalamin-independent)	6888	Amino acid transport and metabolism
KOG0053, Cystathionine beta-lyases/cystathionine gamma-synthases	6162	Amino acid transport and metabolism
COG2873, MET17, O-acetylhomoserine sulfhydrylase	6133	Amino acid transport and metabolism
KOG2790, Phosphoserine aminotransferase	6060	Amino acid transport and metabolism
KOG1268, Glucosamine 6-phosphate synthetases, contain amidotransferase and phosphosugar isomerase domains	7607	Cell wall/membrane/envelope biogenesis
KOG1506, S-adenosylmethionine synthetase	24216	Coenzyme transport and metabolism
KOG1370, S-adenosylhomocysteine hydrolase	8926	Coenzyme transport and metabolism
KOG0204, Calcium transporting ATPase	19494	Inorganic ion transport and metabolism
KOG0693, Myo-inositol-1-phosphate synthase	25686	Lipid transport and metabolism
KOG0059, Lipid exporter ABCA1 and related proteins, ABC superfamily	9000	Lipid transport and metabolism
KOG0888, Nucleoside diphosphate kinase	8411	Nucleotide transport and metabolism
KOG0055, Multidrug/pheromone exporter, ABC superfamily	5171	Secondary metabolites biosynthesis, transport and catabolism

Table 3.4 List of putative carbohydrate esterases, polysaccharide lyases and carbohydrate binding related modules in the muskoxen rumen eukaryotic metatranscriptome (Muskoxen MT), and comparison of their abundance of selected CAZy modules with those of three other foregut microbiomes: wallaby (Macropod); bovine rumen (Bovine 2009 and Bovine 2011) and the hindgut of termite (Termite).

Domains identified from muskoxen MT contigs that significantly differ from the rest data are indicated in bold.

	MuskOxen Contigs (≥500bp)	Macropod	Bovine 2009	Bovine 2011	Termite
Carbohydrate Esterases					
CE1	13	0	11	NR^	NR
CE2/CE3	**6**	**0**	**1**	**NR**	**NR**
CE4	50	24	4	NR	NR
CE6	**0**	**0**	**0**	**NR**	**NR**
CE7	19	3	2	NR	NR
CE8	0	3	0	NR	NR
CE9	3	14	0	NR	NR
CE10	3	18	0	NR	NR
CE11	4	2	0	NR	NR
CE12	0	6	0	NR	NR
CE13	5	0	NR	NR	NR
CE14	0	1	NR	NR	NR
CE15	**0**	**0**	**NR**	**NR**	**NR**
CE16	**8**	**0**	**NR**	**NR**	**NR**
Polysaccharide lyases					
PL1	**6**	**1**	**0**	**NR**	**NR**
PL2	0	0	0	NR	NR
PL3	**4**	**0**	**0**	**NR**	**NR**
PL5	0	0	0	NR	5
PL6	7	0	0	NR	NR
PL9	**8**	**0**	**0**	**NR**	**NR**
PL10	0	0	0	NR	NR
PL11	**1**	**0**	**0**	**NR**	**5**
PL12+15+17+21	0	0	NR	NR	NR
Carbohydrate Binding Modules					
CBM1	**33**	**0**	**0**	**0**	**0**
CBM2	0	1	0	50	0

CBM3	3	0	0	33	0
CBM4/9/16/22	3	6	0	417	5
CBM6	27	3	0	932	13
CBM10	**403**	**0**	**0**	**1**	**0**
CBM11	0	0	0	NR	3
CBM13	31	2	1	118	0
CBM18	**108**	**0**	**0**	**0**	**0**
CBM20	5	3	0	112	0
CBM21	0	1	0	1	0
CBM26	0	1	0	NR	0
CBM29	**11**	**0**	**0**	**NR**	**0**
CBM30	0	0	0	NR	1~8
CBM32/47	2	6	1	747	4
CBM34	0	2	0	72	0
CBM35	1	0	0	NR	0~1
CBM36	8	0	0	NR	2~13
CBM37	3	0	0	NR	1
CBM48	3	22	0	NR	0~1
CBM50	11	47	NR	1887	NR
CBM51	0	1	NR	173	NR

^ NR: not reported.

Table 3.5 The abundance of contigs coding lignocellulytic enzymes [glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), carbohydrate-binding modules (CBMs), and other related modules] in the muskoxen eukaryotic metatranscriptome (Muskoxen MT) and a comparison of their abundance in the databases of rumen fungal (Ru. fungi) and rumen protozoal genes (Ru. prot.) as well as different anaerobic bacteria, including *Bacteroides fragilis* (Bfra), *Butyrivibrio proteoclasticus* (Bpro), *Clostridium thermocellum* (Cthe), *Fibrobacter succinogenes* (Fsuc), *Prevotella ruminicola* (Prum), *Ruminococcus flavofaciens* (Rfla), and the rumen fungus *Piromyces* sp. E2 ESTs (Pir. ESTs).

CAZY family	Pfam Accession	Muskoxen MT Contigs all sizes	Muskoxen MT Contigs ≥500bp	Muskoxen MT Reads	Pir. ESTs	Ru. fungi	Ru. prot	Bfra	Bpro	Cthe	Fsuc	Prum	Rfla
Glycoside hydrolase catalytic domains													
GH1	PF00232.11	24	17	32628	32	5	0	0	1	2	0	0	0
GH2	PF02836.10	12	10	186	0	0	0	15	8	1	2	8	1
GH3	PF00933.14	50	22	4967	4	1	0	10	10	2	3	12	6
GH4	PF02056.9	0	0	2	0	0	0	0	0	0	0	0	0
GH5	PF00150.11	75	46	2423	9	13	13	0	5	11	12	5	11
GH6	PF01341.10	55	31	26950	20	27	0	0	0	0	0	0	0
GH7	PF00840.13	0	0	0	0	0	0	0	0	0	0	0	0
GH8	PF01270.10	6	6	496	1	0	0	0	1	2	6	1	0
GH9	PF00759.12	52	42	10894	22	3	2	0	3	16	9	1	12
GH10	PF00331.13	71	29	5658	1	2	7	0	6	5	7	3	6
GH11	PF00457.10	45	33	22294	13	36	4	0	0	1	3	0	8
GH12	PF01670.9	0	0	0	0	0	0	0	0	0	0	0	0
GH13	PF00128.17	77	47	4564	1	0	2	6	14	2	3	5	4
GH14	PF01373.10	5	0	144	0	0	0	0	0	0	0	0	0

CAZY family	Pfam Accession	Muskoxen MT Contigs all sizes	Muskoxen MT Contigs ≥500bp	Muskoxen MT Reads	Pir. ESTs	Ru. fungi	Ru. prot	Bfra	Bpro	Cthe	Fsuc	Prum	Rfla
GH15	PF00723.14	0	0	0	0	0	0	0	0	1	0	0	0
GH16	PF00722.14	17	10	1797	0	5	0	6	2	2	4	2	3
GH17	PF00322.10	0	0	0	0	0	0	0	0	0	0	0	0
GH18	PF00704.21	30	20	1905	11	0	0	2	1	4	2	1	1
GH19	PF00182.12	2	0	8	0	0	0	0	0	0	0	0	0
GH20	PF00728.15	1	0	61	0	0	0	12	0	0	0	2	0
GH21	deleted family												
GH22	PF00062.13	0	0	1	0	0	0	0	0	0	0	0	0
GH23	PF01464.13	4	1	105	0	0	0	3	0	2	3	3	0
GH24	PF00959.12	5	2	3	0	0	0	0	0	0	0	1	1
GH25	PF01183.13	31	16	2	0	0	0	1	5	0	0	3	7
GH26	PF02156.8	13	8	1737	0	6	0	2	0	3	5	1	6
GH27	PF02065.11	9	6	85	0	0	0	7	5	0	1	2	1
GH28	PF00295.10	1	0	1	0	0	0	0	2	0	0	5	0
GH29	PF01120.10	0	0	10	0	0	0	9	1	0	0	3	0
GH30	PF02055.9	3	3	0	0	0	0	0	1	2	4	0	1
GH31	PF01055.19	35	4	1411	5	0	0	4	6	0	0	7	2
GH32	PF00251.13	9	3	221	0	0	0	3	3	0	0	4	0
GH33	PF02012.13	4	3	0	0	0	0	5	0	1	1	1	1
GH34	PF00064.11	0	0	0	0	0	0	0	0	0	0	0	0
GH35	PF01301.12	0	0	14	0	0	0	4	2	0	0	2	0
GH36	PF02065	9	6	85	0	0	0	7	5	0	1	2	1
GH37	PF01204.11	0	0	28	0	0	0	0	0	1	0	0	0
GH38	PF01074.15	1	0	80	0	0	0	1	1	0	0	1	0
GH39	PF01229.10	0	0	2	0	0	0	0	1	0	0	0	0

CAZY family	Pfam Accession	Muskoxen MT Contigs all sizes	Muskoxen MT Contigs ≥500bp	Muskoxen MT Reads	Pir. ESTs	Ru. fungi	Ru. prot	Bfra	Bpro	Cthe	Fsuc	Prum	Rfla
GH40	deleted family												
GH41	deleted family												
GH42	PF02449.8	0	0	2	0	0	0	4	3	0	0	2	1
GH43	PF04616.7	81	54	139	75	0	0	10	13	5	13	18	9
GH44	BLAST	0	0	5	0	0	0	0	0	1	1	0	1
GH45	PF02015.9	58	51	14034	10	1	0	0	0	0	4	0	0
GH46	PF01374.11	0	0	0	0	0	0	0	0	0	0	0	0
GH47	PF01532.13	0	0	51	1	0	0	0	0	1	1	0	0
GH48	PF02011.8	54	33	106034	37	2	0	0	0	2	0	0	1
GH49	PF03718.6	0	0	0	0	0	0	0	0	0	0	0	0
GH50	PB001653	0	0	2	0	0	0	0	0	0	0	1	0
GH51	PF06964	1	1	112	0	0	0	1	2	1	1	7	0
GH52	PF03512.6	0	0	0	0	0	0	0	0	0	0	0	0
GH53	PF07745.6	4	0	83	0	0	0	0	2	1	2	2	1
GH54	PF09206.4	0	0	0	0	0	0	0	0	0	1	0	0
GH55	BLAST	0	0	26	0	0	0	0	0	1	0	0	0
GH56	PF01630.11	0	0	0	0	0	0	0	0	0	0	0	0
GH57	PF03065.8	0	0	0	0	0	0	1	1	0	3	1	0
GH58	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
GH59	PF02057.8	0	0	0	0	0	0	0	0	0	0	0	0
GH60	Deleted family												
GH61	PF03443.7	0	0	0	0	0	0	0	0	0	0	0	0
GH62	PF03664.6	0	0	0	0	0	0	0	0	0	0	0	0
GH63	PF03200.9	0	0	31	0	0	0	0	0	0	0	0	0

CAZY family	Pfam Accession	Muskoxen MT Contigs all sizes	Muskoxen MT Contigs ≥500bp	Muskoxen MT Reads	Pir. ESTs	Ru. fungi	Ru. prot	Bfra	Bpro	Cthe	Fsuc	Prum	Rfla
GH64	PB001434	4	3	593	0	0	0	0	0	0	0	0	0
GH65	PF03632.8	0	0	14	0	0	0	2	0	0	0	0	0
GH66	PB003959	0	0	0	0	0	0	0	0	0	0	0	0
GH67	PF07488.5	1	1	639	1	0	0	0	1	0	0	1	0
GH68	PF02435.9	0	0	0	0	0	0	0	0	0	0	0	0
GH69	Deleted: now PL16												
GH70	PF02324.9	0	0	1	0	0	0	1	0	0	0	0	0
GH71	PF03659.7	0	0	0	0	0	0	0	0	0	0	0	0
GH72	PF03198.7	5	0	129	0	0	0	0	0	0	0	0	0
GH73	PF01832.13	2	2	40	0	0	0	1	0	0	0	1	0
GH74	BLAST	22	6	171	1	0	0	0	0	1	1	0	1
GH75	PF07335.4	0	0	0	0	0	0	0	0	0	0	0	0
GH76	PF03663.7	3	1	8	0	0	0	3	0	1	0	1	0
GH77	PF02446.10	22	8	2643	0	0	0	1	1	0	1	2	1
GH78	PF05592.4	5	0	181	0	0	0	3	4	2	0	1	0
GH79	PF03662.7	0	0	0	0	0	0	0	0	0	0	0	0
GH80	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
GH81	PF03639.6	0	0	0	0	0	0	0	0	1	0	0	0
GH82	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
GH83	PF00423.12	0	0	0	0	0	0	0	0	0	0	0	0
GH84	PF07555.6	3	1	182	0	0	0	1	0	0	0	0	0
GH85	PF03644.6	0	0	0	0	0	0	0	0	0	0	0	0
GH86	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
GH87	BLAST	7	3	433	0	0	0	0	0	0	0	0	0
GH88	PF07470.6	9	1	45	0	0	0	3	5	0	2	2	1

CAZY family	Pfam Accession	Muskoxen MT Contigs all sizes	Muskoxen MT Contigs ≥500bp	Muskoxen MT Reads	Pir. ESTs	Ru. fungi	Ru. prot	Bfra	Bpro	Cthe	Fsuc	Prum	Rfla
GH89	PF05089.5	15	6	1191	0	0	0	1	0	0	0	1	0
GH90	PF09251	0	0	0	0	0	0	0	0	0	0	0	0
GH90	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
GH91	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
GH92	PF07971.5	0	0	12	0	0	0	8	0	0	0	8	0
GH93	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
GH94	BLAST	2	1	4308	0	0	0	0	1	3	1	1	2
GH96	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
GH97	PF10566.2	2	2	16	0	0	0	4	0	0	0	6	3
GH98	PF08306.4	0	0	1	0	0	0	0	0	0	0	0	0
GH99	PB000173	0	0	12	0	0	0	0	0	0	0	1	0
GH100	PF04853.5	0	0	1	0	0	0	0	0	0	0	0	0
GH101	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
GH102	PF03562.10	0	0	1	0	0	0	0	0	0	0	0	0
GH103	BLAST	0	0	1	0	0	0	0	0	0	0	0	0
GH104	PF00959.12	5	2	3	0	0	0	0	0	0	0	1	1
GH105	PF07470.6	9	1	45	0	0	0	3	5	0	2	2	1
GH106	PB000239	0	0	1	0	0	0	0	0	0	0	1	0
GH107	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
GH108	PF05838.5	0	0	0	0	0	0	0	0	0	0	0	0
GH109	PF01408	2	0	51	1	0	0	5	6	3	3	4	1
GH110	PB004422	0	0	0	0	0	0	2	0	0	0	0	0
GH111	PWGH111	0	0	0	0	0	0	0	0	0	0	0	0
GH112	PF09508	0	0	1	0	0	0	0	1	0	0	0	0
GH113	PB000085	0	0	0	0	0	0	0	0	0	0	0	0
GH114	PF03537	23	11	343	38	0	0	0	0	0	0	0	0

CAZY family	Pfam Accession	Muskoxen MT Contigs all sizes	Muskoxen MT Contigs ≥500bp	Muskoxen MT Reads	Pir. ESTs	Ru. fungi	Ru. prot	Bfra	Bpro	Cthe	Fsuc	Prum	Rfla
GH115	PB000184	13	10	1541	0	0	0	1	3	0	0	1	0
GH116	PF12215	0	0	0	0	0	0	0	0	0	1	0	0
GH117	PWGH117	16	14	8	0	0	0	5	4	3	4	8	2
GH118	PWGH118	0	0	0	0	0	0	0	0	0	0	0	0
CBMs													
CBM1	PF00734.11	85	33	3697	20	11	0	0	0	0	0	0	0
CBM2	PF00553.12	0	0	2	0	0	0	0	7	0	0	0	0
CBM3	PF00942.11	3	3	1	0	0	2	0	1	20	0	0	3
CBM4	PF02018.10	5	3	0	0	0	3	2	0	8	5	4	19
CBM5	PF02839.7	0	0	1	0	0	0	0	0	0	0	0	0
CBM6	PF03422.8	40	27	84	0	1	0	3	3	15	30	2	6
CBM7	deleted entry												
CBM8	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
CBM9	PF02018.10	5	3	0	0	0	3	2	0	8	5	4	19
CBM10	PF02013.9	908	403	100192	183	84	0	0	0	0	0	0	0
CBM11	PF03425.6	0	0	0	0	1	0	0	0	1	6	0	0
CBM12	PF02839.7	0	0	1	0	0	0	0	0	0	0	0	0
CBM13	PF00652.15	40	31	5	2	0	1	0	3	1	0	0	8
CBM14	PF01607.17	0	0	0	0	0	0	0	0	0	0	0	0
CBM15	PF03426.7	0	0	0	0	0	0	0	0	0	0	0	0
CBM16	PF02018.10	5	3	0	0	0	3	2	0	8	5	4	19
CBM17	PF03424.7	0	0	0	0	0	0	0	0	0	0	0	0
CBM18	PF00187.12	370	108	22914	88	0	0	0	0	0	0	0	0
CBM19	PF03427.6	0	0	0	0	0	0	0	0	0	0	0	0
CBM20	PF00686.12	6	5	84	0	0	0	1	0	0	0	2	0
CBM21	PF03370.6	1	0	52	3	0	0	0	0	0	0	0	0

CAZY family	Pfam Accession	Muskoxen MT Contigs all sizes	Muskoxen MT Contigs ≥500bp	Muskoxen MT Reads	Pir. ESTs	Ru. fungi	Ru. prot	Bfra	Bpro	Cthe	Fsuc	Prum	Rfla
CBM22	PF02018.10	5	3	0	0	0	3	2	0	8	5	4	19
CBM23	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
CBM24	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
CBM25	PF03423.6	0	0	0	2	0	0	0	0	2	0	0	0
CBM26	PB013554	0	0	4	1	0	0	0	1	0	0	0	0
CBM27	PF09212.3	0	0	0	0	0	0	0	0	0	0	0	0
CBM28	PF03424.7	0	0	0	0	0	0	0	0	0	0	0	0
CBM29	BLAST	13	11	2318	9	19	0	0	0	0	0	0	0
CBM30	BLAST	0	0	29	0	0	0	0	0	0	3	0	0
CBM31	PF11606	0	0	0	0	0	0	0	0	0	0	0	0
CBM32	PF00754.18	6	2	27	0	0	0	21	1	2	1	6	2
CBM33	PF03067.8	0	0	0	0	0	0	0	0	0	0	0	0
CBM34	PF02903.7	0	0	0	0	0	0	0	2	1	0	0	0
CBM35	BLAST	0	1	1	0	5	0	0	0	1	1	0	4
CBM36	BLAST	21	8	0	0	0	0	1	1	10	1	2	2
CBM37	BLAST	4	3	5	0	0	2	0	0	2	0	0	15
CBM38	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
CBM39	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
CBM40	PF02973.9	0	0	0	0	0	0	0	0	0	0	0	0
CBM41	PF03714.7	0	0	0	0	0	0	0	1	0	0	0	0
CBM42	PF05270.6	0	0	4	0	0	0	0	0	4	0	0	0
CBM43	PF07983.6	0	0	1	0	0	0	0	0	0	0	0	0
CBM44	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
CBM45	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
CBM46	PF03442	0	0	0	0	0	0	0	0	0	0	0	0
CBM47	PF00754.18	6	2	27	0	0	0	21	1	2	1	6	2

CAZY family	Pfam Accession	Muskoxen MT Contigs all sizes	Muskoxen MT Contigs ≥500bp	Muskoxen MT Reads	Pir. ESTs	Ru. fungi	Ru. prot	Bfra	Bpro	Cthe	Fsuc	Prum	Rfla
CBM48	PF02922.11	6	3	98	7	0	0	2	4	1	2	4	2
CBM49	PF09478.3	0	0	0	0	0	0	0	2	0	0	0	0
CBM50	PF01476.13	18	11	1520	3	0	0	4	3	10	3	5	0
CBM51	PF08305.4	0	0	0	0	0	0	0	0	0	3	0	0
CBM52	PF10645	4	1	101	0	0	0	0	0	0	0	0	0
CBM53	BLAST	0	0	0	0	0	0	0	0	0	0	0	0
CBM54	PWCBM54	0	0	0	0	0	0	0	0	1	0	0	0
CBM55	PWCBM55	0	0	0	0	0	0	0	0	0	0	0	0
CBM56	PWCBM56	0	0	0	0	0	0	0	0	0	0	0	0
CBM57	PF11721	2	0	19	0	0	0	0	1	0	0	0	0
CBM58	PWCBM58	0	0	0	0	0	0	0	0	0	0	0	0
CBM59	PWCBM59	0	0	0	0	0	0	0	0	0	0	0	0
Carbohydrate esterases													
CE1	PF00756	17	13	86	0	2	0	4	6	3	1	7	9
CE2+CE3	PB002673	18	6	85	0	3	0	3	2	1	1	2	3
CE4	PF01522.14	78	50	983	4	0	0	2	5	5	3	2	6
CE5	PF01083.15	0	0	0	0	0	0	0	0	0	0	0	0
CE6	PF03629	19	19	17	0	2	0	2	1	0	5	3	0
CE7	PF05448	1	0	6	0	0	0	2	1	1	1	3	2
CE8	PF01095.12	10	3	96	0	0	0	0	2	1	1	2	1
CE9	PF01979.13	7	3	31	0	0	0	5	3	6	2	1	3
CE10	PF00135	15	4	745	10	0	0	1	5	1	1	3	0
CE11	PF03331.6	0	0	4	0	0	0	1	0	0	1	1	0
CE12	PB008046	10	5	553	0	0	0	1	3	1	5	4	3
CE13	PF03283	0	0	157	0	0	0	0	1	0	0	0	1
CE14	PF02585.10	0	0	5	0	0	0	1	1	1	0	1	0

CAZY family	Pfam Accession	Muskoxen MT Contigs all sizes	Muskoxen MT Contigs ≥500bp	Muskoxen MT Reads	Pir. ESTs	Ru. fungi	Ru. prot	Bfra	Bpro	Cthe	Fsuc	Prum	Rfla
CE15	PWCE015	9	8	4231	0	0	0	1	0	0	1	3	2
CE16	PWCE016	2	0	3	1	0	0	0	0	0	0	0	0
Polysaccharide lyases													
PL1	PF00544.12	10	6	244	0	0	0	0	1	2	6	1	4
PL2	PF06917	0	0	0	0	0	0	0	0	0	0	0	0
PL3	PF03211.6	10	4	3660	2	0	0	0	0	0	0	0	0
PL4	PF06045	1	1	0	0	0	0	0	0	0	0	0	0
PL4	PF09284	0	0	0	0	0	0	0	0	0	0	0	0
PL5	PF05426.5	0	0	0	0	0	0	0	0	0	0	0	0
PL6	PWPL006	8	7	0	0	0	0	0	0	1	0	0	0
PL7+PL18	PF08787	0	0	0	0	0	0	0	0	0	0	0	0
PL8	PF02278.11	0	0	8	0	0	0	1	0	0	0	0	0
PL9	QMPL09	15	8	325	0	0	0	2	4	2	2	0	2
PL10	PF09492	0	0	0	0	0	0	0	0	0	1	1	0
PL11	PWPL011	1	1	113	0	0	0	0	1	1	2	1	7
PL13	PWPL013	0	0	0	0	0	0	0	0	0	0	0	0
PL14	PB002765	0	0	0	0	0	0	0	0	0	1	0	0
PL16	PF07212	0	0	0	0	0	0	0	0	0	0	0	0
PL19	Deleted:No wGH91												
PL20	PWPL020	0	0	0	0	0	0	0	0	0	0	0	0
PL22	PB009195	0	0	0	0	0	0	5	0	0	2	3	0
PL12+15+17 +21	PF07940	0	0	3	0	0	0	2	1	0	0	1	0
Other domains associated with GH catalytic or carbohydrate binding domains													
AXE1	PF05448.5	1	0	6	0	0	0	2	1	1	1	3	2

CAZY family	Pfam Accession	Muskoxen MT Contigs all sizes	Muskoxen MT Contigs ≥500bp	Muskoxen MT Reads	Pir. ESTs	Ru. fungi	Ru. prot	Bfra	Bpro	Cthe	Fsuc	Prum	Rfla
Alpha-amylase_C	PF02806.11	2	2	143	4	0	0	1	3	1	1	1	2
Alpha-L-AF_C	PF06964.5	1	1	112	0	0	0	1	2	1	1	7	0
Alpha-mann_mid	PF09261.4	0	0	15	0	0	0	1	1	0	0	1	0
Bac_rhamnosid_N	PF08531.3	0	0	24	0	0	0	2	3	0	0	1	0
Big_1	PF02369.9	0	0	0	0	0	0	0	0	1	0	0	0
Big_2	PF02368.11	0	0	29	0	0	0	0	2	2	0	4	5
Big_3	PF07523.5	0	0	1	0	0	0	1	1	0	0	0	3
Big_4	PF07532.4	0	0	1	0	0	0	0	4	1	0	0	1
Bgal_small_N	PF02929.10	3	3	24	0	0	0	5	1	0	1	2	1
CBMX	PF06204.4	1	1	793	0	0	0	0	1	2	1	1	2
CelD_N	PF02927.7	0	0	9	0	0	0	0	2	4	6	0	5
CHB_HEX	PF03173.6	0	0	0	0	0	0	0	0	0	0	0	0
CHB_HEX_C	PF03174.6	0	0	0	0	0	0	1	1	0	0	0	0
ChiC	PF06483.4	0	0	0	0	0	0	0	0	0	0	0	0
ChitinaseA_N	PF08329.3	0	0	0	0	0	0	0	0	0	0	0	0
Cohesin	PF00963.11	0	0	1	0	0	0	0	0	8	0	0	2
Dockerin_1	PF00404.11	0	0	1	0	0	0	0	0	72	0	0	10
fn3	PF00041.14	0	0	0	1	0	0	2	1	6	3	2	0
GDE_C	PF06202.7	3	3	357	3	0	0	1	0	2	0	1	0
Glucodextran_B	PF09136.3	0	0	0	0	0	0	0	0	0	0	0	0
Glucodextran_N	PF09137.4	0	0	0	0	0	0	0	0	0	0	0	0

CAZY family	Pfam Accession	Muskoxen MT Contigs all sizes	Muskoxen MT Contigs ≥500bp	Muskoxen MT Reads	Pir. ESTs	Ru. fungi	Ru. prot	Bfra	Bpro	Cthe	Fsuc	Prum	Rfla
GH2N	PF02837.11	9	7	463	0	0	0	17	9	2	2	13	2
GH2	PF00703.14	6	4	21	0	0	0	15	6	1	2	8	1
GH3C	PF01915.15	37	23	2035	2	1	0	8	8	1	1	11	4
GH20b	PF02838.8	0	0	5	0	0	0	8	0	0	0	2	0
GH32C	PF08244.5	1	0	0	0	0	0	1	1	0	0	1	0
GH38C	PF07748.6	0	0	24	0	0	0	1	1	0	0	1	0
GH42M	PF08532.3	0	0	1	0	0	0	0	1	0	0	0	1
GH42C	PF08533.3	0	0	0	0	0	0	0	0	0	0	0	1
GH65N	PF03636.8	0	0	0	0	0	0	1	0	0	0	0	0
GH65C	PF03633.8	0	0	1	0	0	0	1	0	0	0	0	0
GH67N	PF03648.7	0	0	0	0	0	0	1	0	0	0	2	0
GH67C	PF07477.5	1	1	406	0	0	0	0	1	0	0	1	0
GH98C	PF08307.4	0	0	1	0	0	0	0	0	0	0	0	0
Glyco_transf_36	PF06165.4	1	1	333	0	0	0	0	2	3	1	1	2
GT36_AF	PF06205.4	2	1	588	0	0	0	0	1	2	1	1	2
He_PIG	PF05345.5	0	0	0	0	0	0	1	0	0	0	0	0
PKD	PF00801.13	0	0	3	0	0	0	1	0	2	0	0	0
TIG	PF01833.17	7	4	9	0	0	0	0	0	0	0	2	0
SusD	PF07980.4	0	0	11	0	0	0	59	0	0	0	29	0
TonB_dep_rec (SusC)	PF00593.17	0	0	4	0	0	0	82	0	0	1	41	0
Others													
SLH	PF00395	0	0	26	0	0	0	0	0	25	0	0	0
Swollenin	SWOLLEN	22	16	7065	7	0	0	0	0	0	0	0	0

Table 3.6 CAZyme contigs (≥500 bp) identified with 10,000 or more reads in the metatranscriptome from the muskoxen rumen

Contig	Domains	Length	No. of Reads	E-value	Id%	BLASTX Hit Description
Contig21589	GH48; CBM10	2292	52778	0	78.2	cellulase Cel48A precursor [Piromyces sp. E2]
Contig28863	GH11	1148	44380	0	86.8	xylanase [Neocallimastix frontalis]
Contig22047	GH6; CBM10	1921	44342	0	83.1	cellobiohydrolase II-like cellulase CelH [Orpinomyces sp. PC-2]
Contig28684	GH48	898	33408	1.00E-154	81.1	cellulase Cel48A precursor [Piromyces equi]
Contig627	CBM20; PL6	1777	20162	7.00E-14	24.5	Parallel beta-helix repeat protein [Planctomyces maris DSM 8797]
Contig27493	GH16	878	19963	9.00E-74	57.1	Licheninase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig28658	GH48; CBM10	1093	16001	1.00E-145	74.3	cellulase Cel48A precursor [Piromyces sp. E2]
Contig21206	GH45	925	15508	1.00E-68	52.4	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig21311	GH94; GT36_AF	2521	15361	0	72.5	cellobiose phosphorylase [Prevotella ruminicola 23]
Contig21506	GH1	821	14352	1.00E-110	88.6	beta-glucosidase Cel1C [Piromyces sp. E2]
Contig29533	GH1	819	14351	1.00E-138	88.2	beta-glucosidase [Orpinomyces sp. PC-2]
Contig3078	Glucosaminidase	601	13803	1.00E-26	72.5	Muramidase (flagellum-specific) [Eubacterium rectale A1-86 (DSM 17629)]
Contig29741	GH1	586	13272	7.00E-91	84.7	beta-glucosidase Cel1C [Piromyces sp. E2]
Contig29986	CBM1	854	13133	6.00E-33	37.8	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig29325	GH43	1409	12451	1.00E-113	59.0	beta-xylosidase/alpha-L-arabinfuranosidase, putative, gly43F [Cellvibrio japonicus Ueda107]
Contig26982	GH1	1879	11902	0	82.3	beta-glucosidase [Orpinomyces sp. PC-2]
Contig12311	GH1	1369	11613	0	83.7	beta-glucosidase [Piromyces sp. E2]
Contig30005	GH48	1203	11562	0	83.1	cellulase Cel48A precursor [Piromyces sp. E2]
NODE_3576	GH48; CBM10	2067	11488	0	79.8	cellulase Cel48A precursor [Piromyces sp. E2]
Contig29850	GH48; CBM10	1092	11311	1.00E-152	73.8	cellulase Cel48A precursor [Piromyces equi]
Contig30163	GH1	954	11088	1.00E-146	74.1	beta-glucosidase [Orpinomyces sp. PC-2]
Contig30798	GH45	523	10896	3.00E-48	63.6	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig29098	GH48; CBM10	1484	10522	1.00E-145	71.3	cellulase Cel48A precursor [Piromyces equi]
Contig1597	GH6; CBM10	1020	10347	1.00E-117	65.4	exoglucanase Cel6A [Piromyces sp. E2]
Contig27001	GH43	1127	10074	1.00E-115	60.1	beta-xylosidase/alpha-L-arabinfuranosidase, putative, gly43F [Cellvibrio japonicus Ueda107]

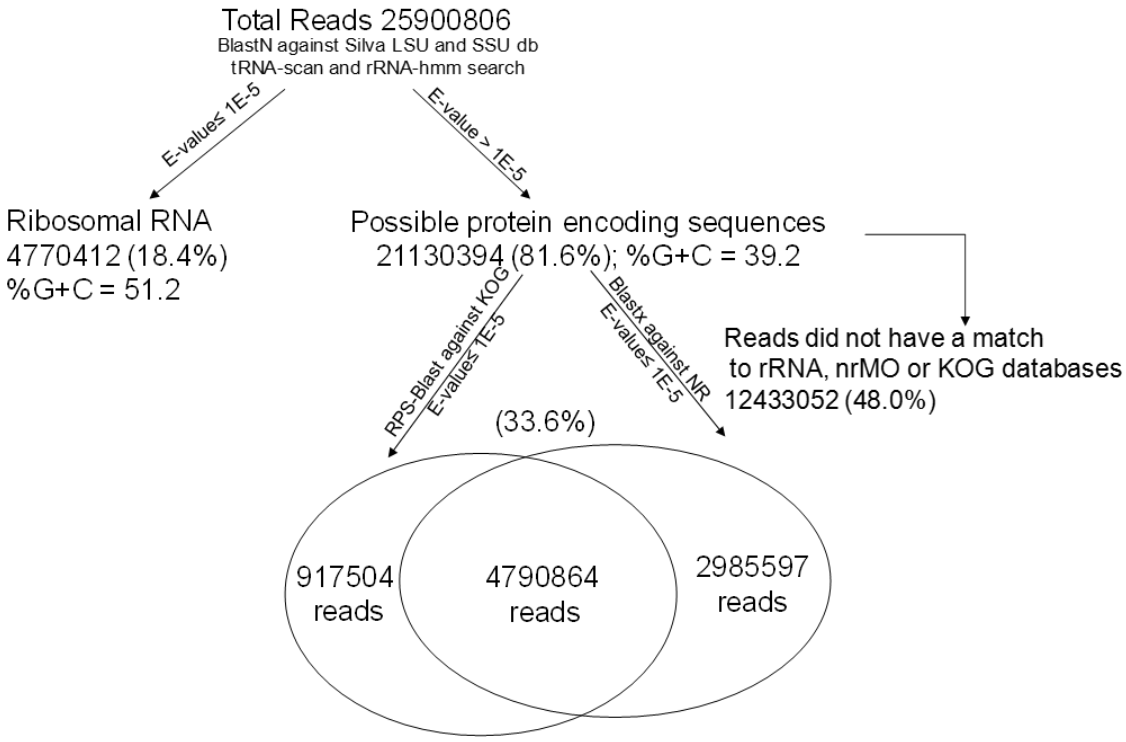

Figure 3.1 Summary statistics for the muskoxen rumen eukaryotic metatranscriptome.

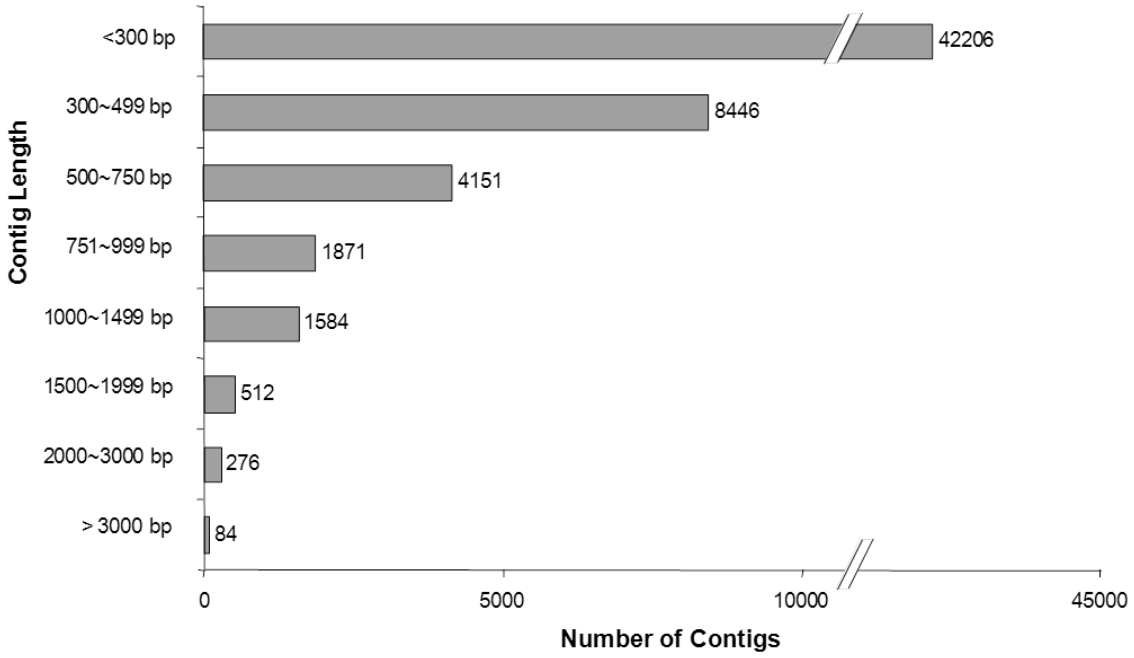

Figure 3.2 Length distribution of muskoxen rumen metatranscriptome contigs.

The number of contigs is indicated on the right side of the bar.

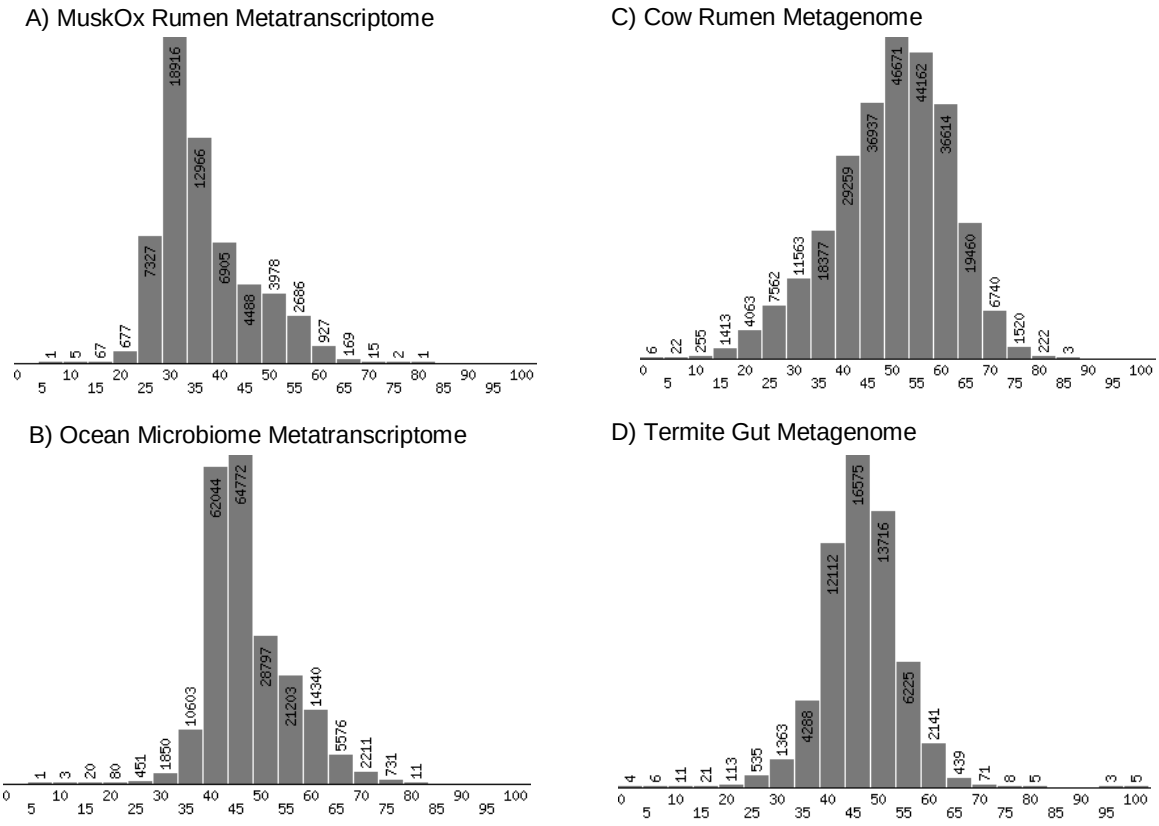

Figure 3.3 GC content analysis of the muskox rumen microbial community metatranscriptome.

The GC molar % of each contig was calculated. Number shown on the column indicates number of contigs with a certain GC range. The data of ocean microbiome metatranscriptome (Poretsky et al., 2009), bovine rumen metagenome (Brulc et al., 2009) and termite gut metagenome (Warnecke et al., 2007) are also shown.

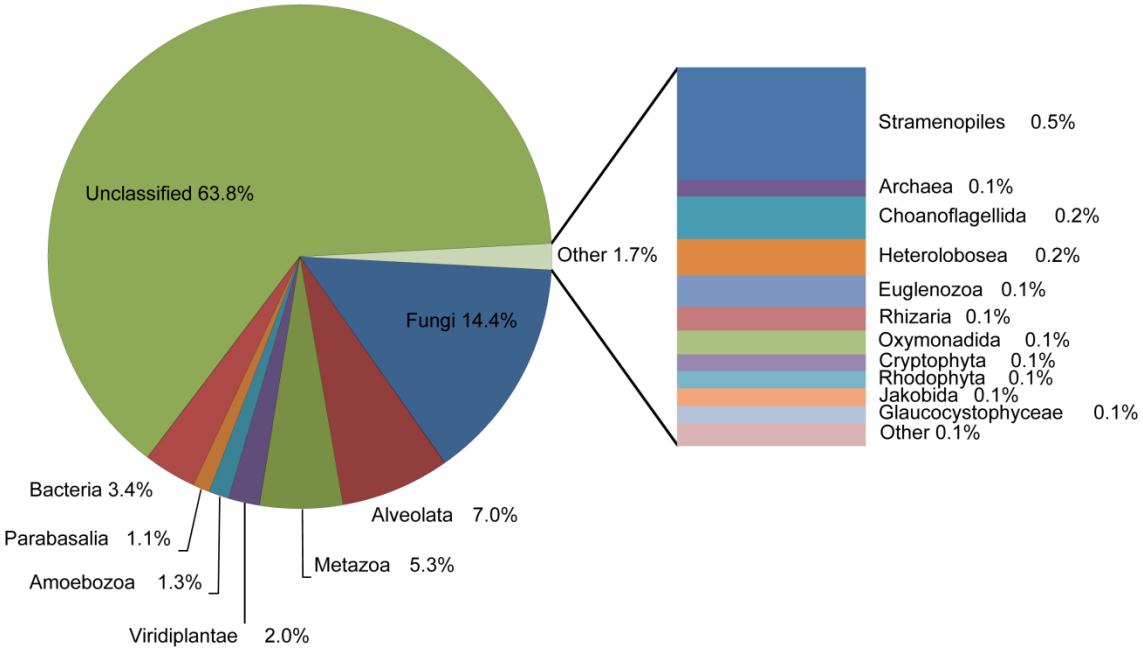

Figure 3.4 Phylogenetic distribution of muskoxen rumen metatranscriptome putative protein encoding reads (a total of 21.1 million) based on MEGAN analysis of top BLASTX hits against the NRMO database. The percentages of the major phylogenetic groups are indicated.

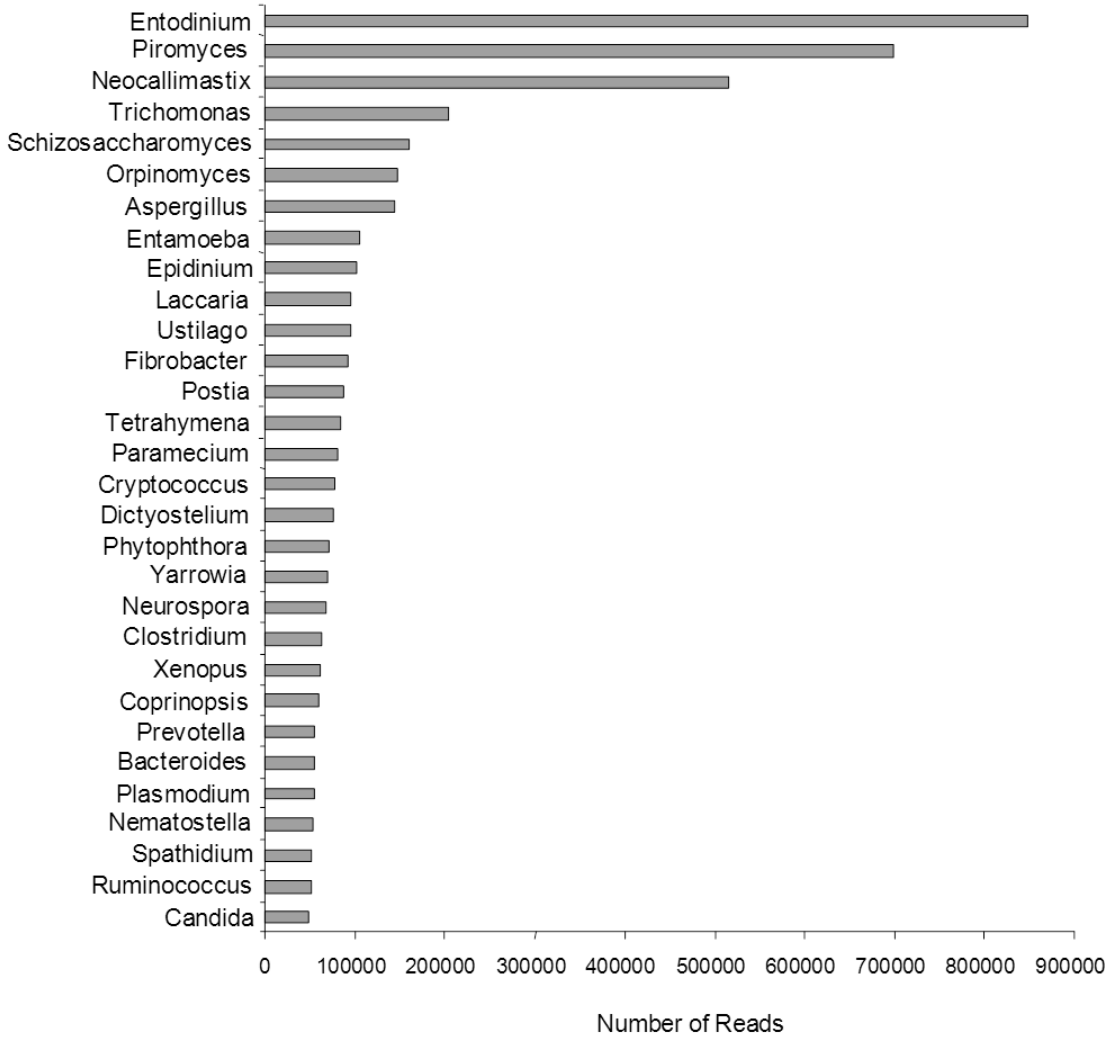

Figure 3.5 Top 30 phylogenetic bins of the muskoxen rumen metatranscriptome as determined by comparison against NCBI's non-redundant amino acid (nr) database. Ranks are determined by the highest total reads number at the genus level.

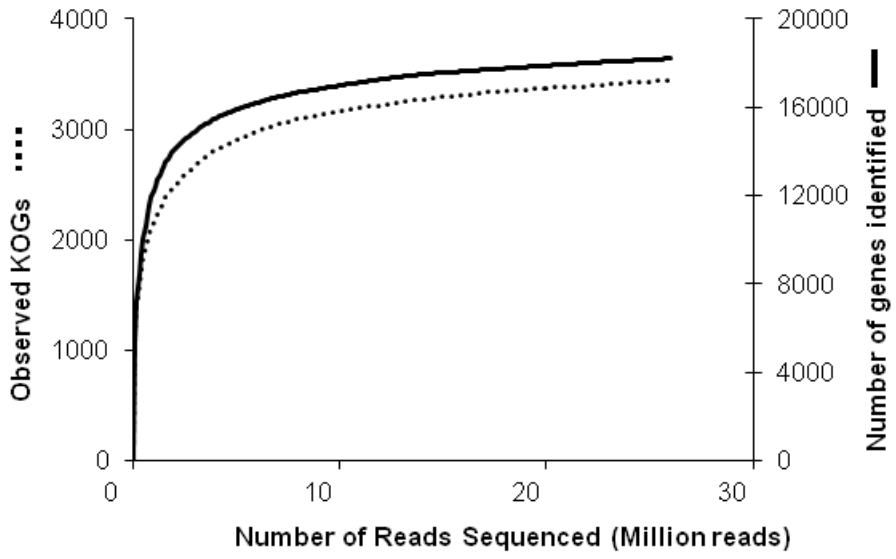

Figure 3.6 Collector's curve of richness as a function of reads analyzed.

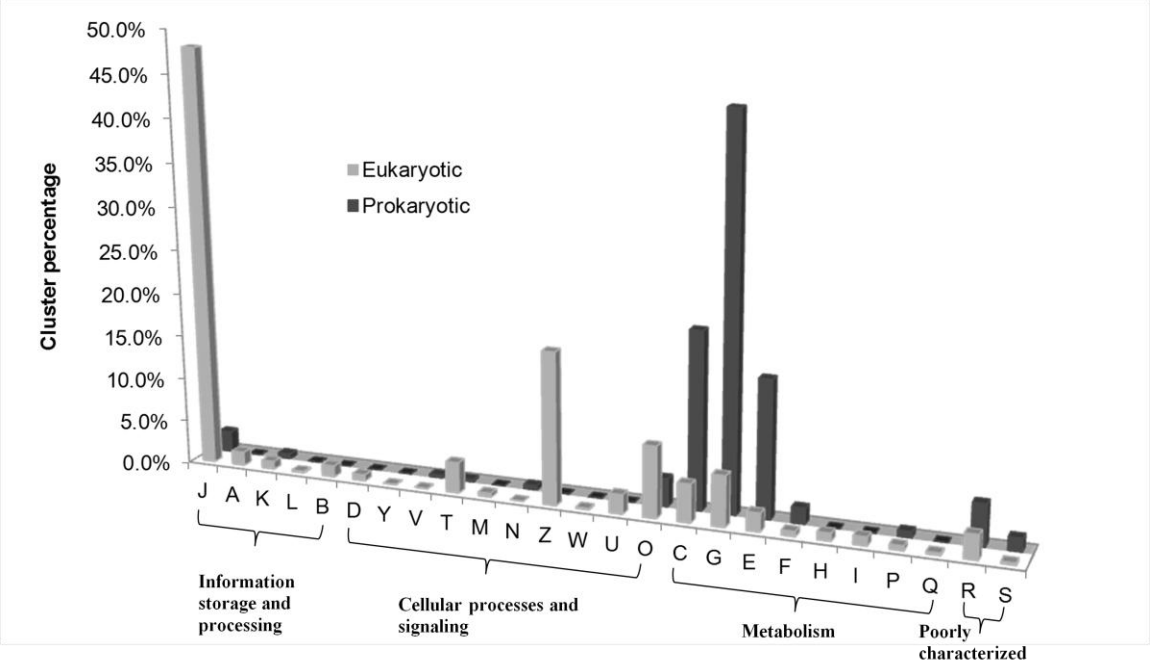

Figure 3.7 Category distribution of the muskoxen rumen metatranscriptome as annotated using Eukaryotic Orthologous Groups (KOGs, for reads showing top BLASTX match to eukaryotic genes; solid bar) and clusters of orthologous groups (COGs, reads showing top BLASTX match to bacterial genes; dotted bar).

The assigned letters are based on KOG/COG classifications (Tatusov et al., 2003).

A total of 5.7 million out of 21.1 million putative protein encoding sequences in the muskoxen rumen eukaryotic metatranscriptome were annotated to a KOG category or COG category. The percentage of annotated ORFs for each KOG/COG category is shown.

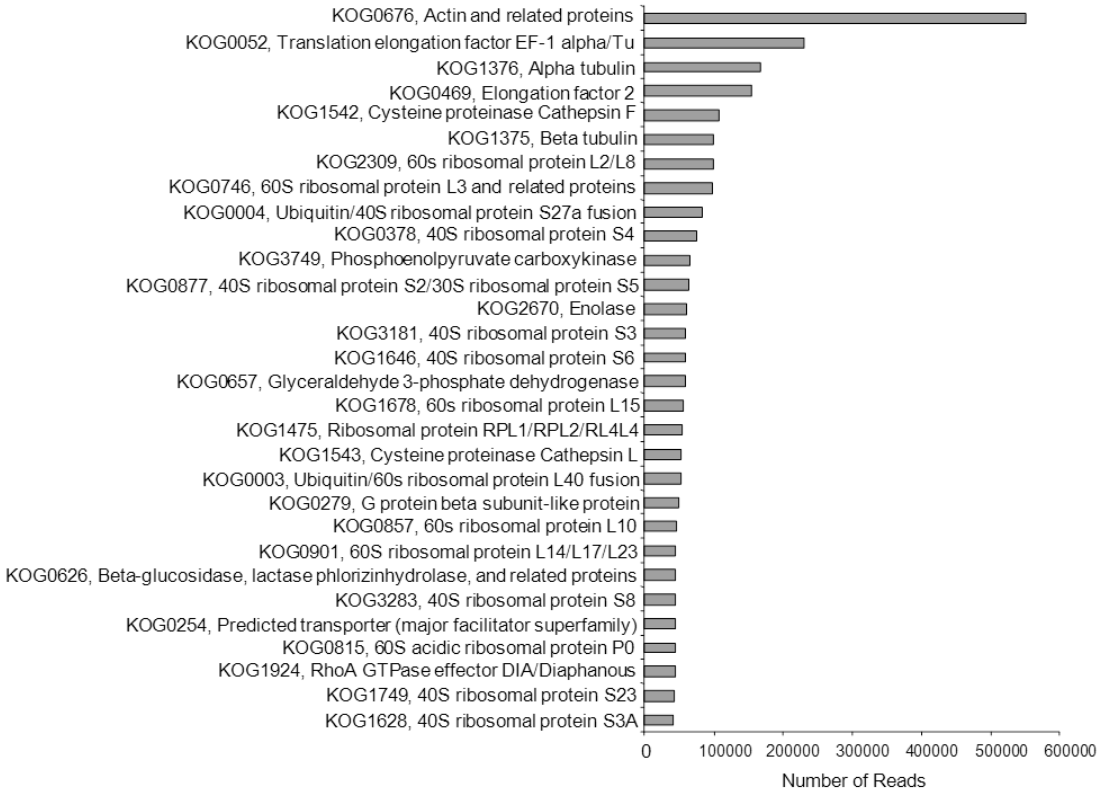

Figure 3.8 Top 30 KOG bins of the muskoxen rumen metatranscriptome as determined by comparison against KOG database.

Ranks are determined by the highest number of total reads for each KOG category.

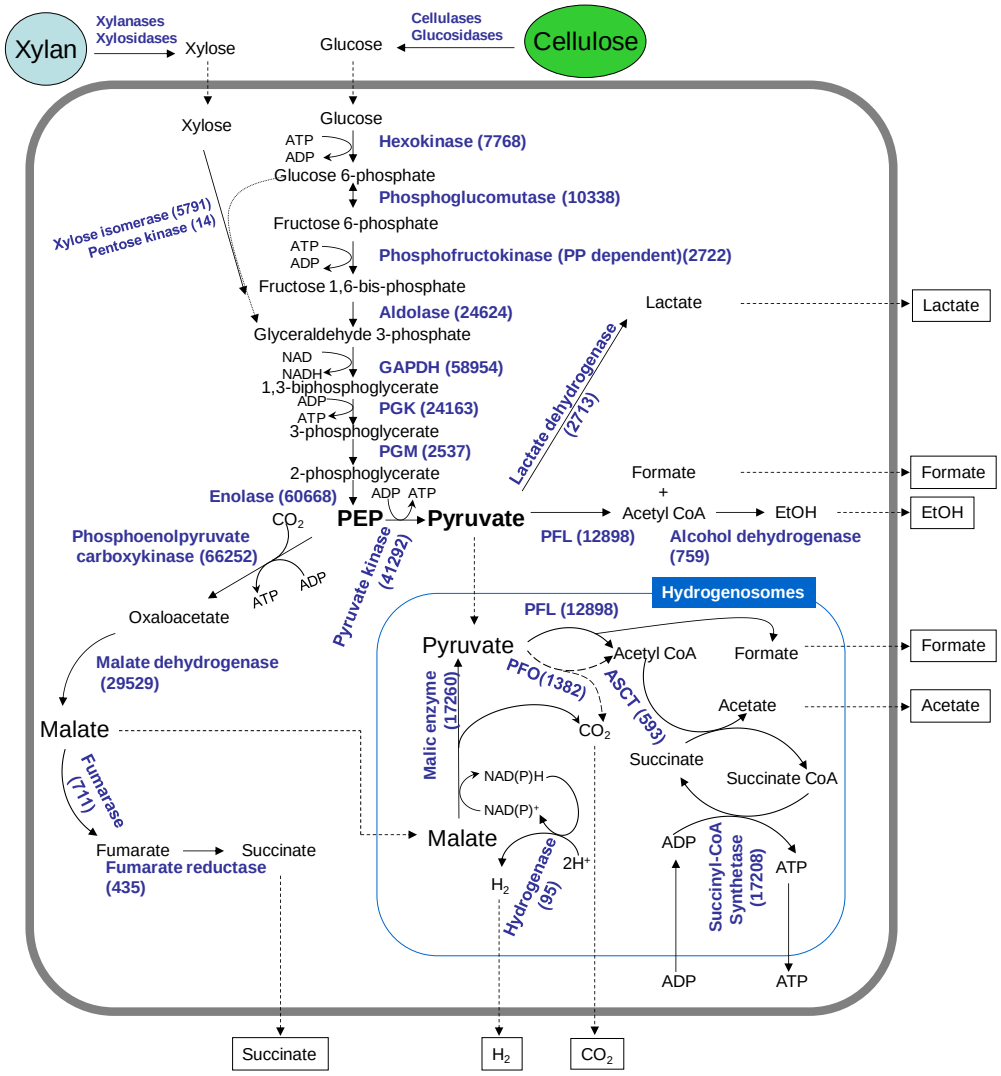

Figure 3.9 Schematic representation of plant cell wall polysaccharide and energy metabolism by the muskoxen rumen eukaryotic population.

The inner box represents the hydrogenosome present in anaerobic fungi and possibly the rumen protozoa. The number after each enzyme represents the read number identified by KOG/COG searches. Abbreviations: ASCT, Acetate: Succinate CoA-transferase; CAZY, carbohydrate active enzymes; GAPDH, glyceraldehydes-3-phosphate dehydrogenase; PEP, phosphoenolpyruvate; PFL, Pyruvate: Formate lyase; PFO, Pyruvate: ferredoxin oxidoreductase; PGK, Phosphoglycerate kinase; PGM, Phosphoglycerate mutase.

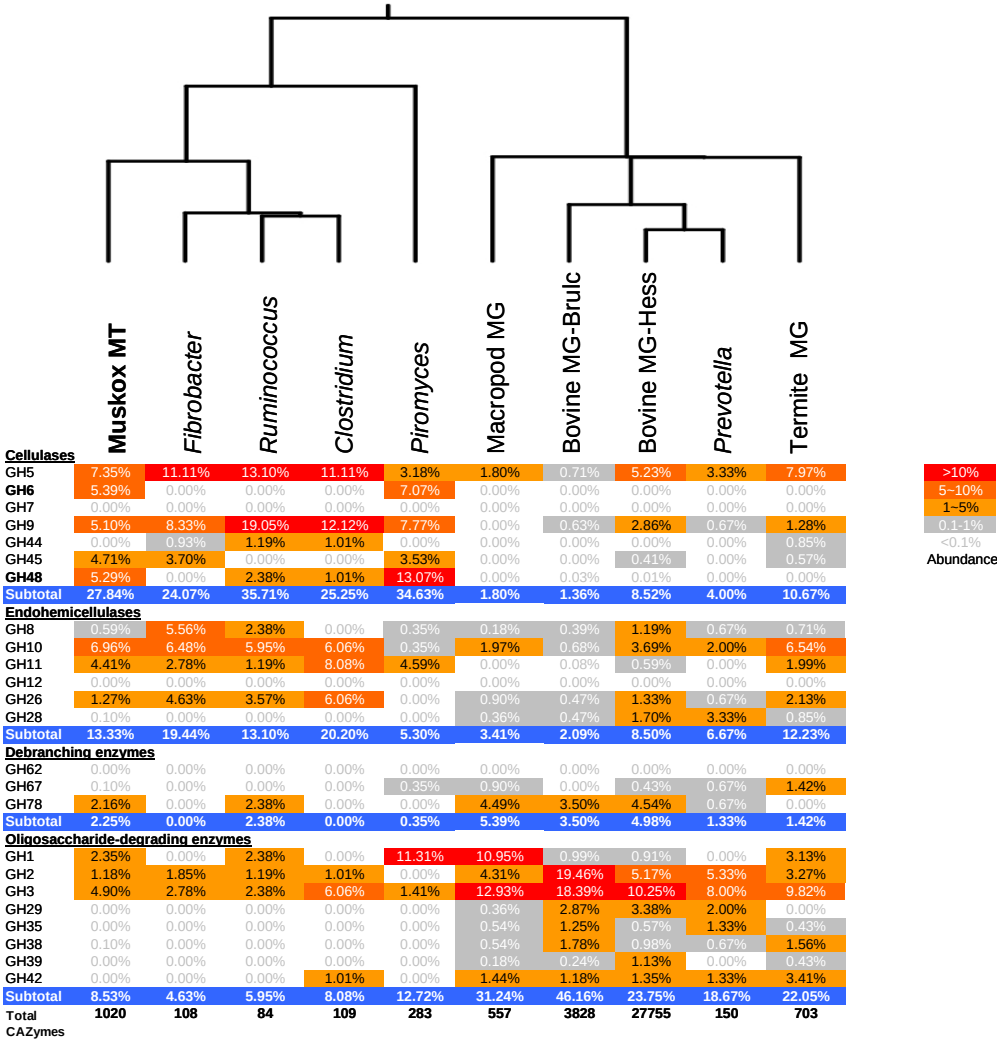

Figure 3.10 Comparison of the carbohydrate active enzymes identified from muskoxen rumen metatranscriptome (using all assembled contigs) with those of three other foregut metagenomes, the termite hindgut and five rumen/anaerobic microorganisms.

The percentages of each enzyme family were shown in the cells. Refer to Table 3.5 for a complete comparison. Dendrogram on the top indicates the relationship of the GHs identified based on similar percentage distribution. Muskox MT: Muskoxen rumen metatranscriptome; Fibrobacter: Genome of *Fibrobacter succinogenes* S85; Ruminococcus: Genome of *Ruminococcus flavefaciens*; Clostridium: Genome of

Clostridium thermocellum; Piromyces: EST sequence of *Piromyces* sp. E2; Macropod MG: Macropod foregut microbiome (Pope et al., 2010); Termite MG: Termite hindgut microbiome (Warnecke et al., 2007); Bovine MG-Hess: Bovine Rumen microbiome by Hess et al. (2011); Bovine MG-Brulc: Bovine Rumen microbiome by Brulc et al. (2009); and Prevotella: Genome of *Prevotella ruminicola*.

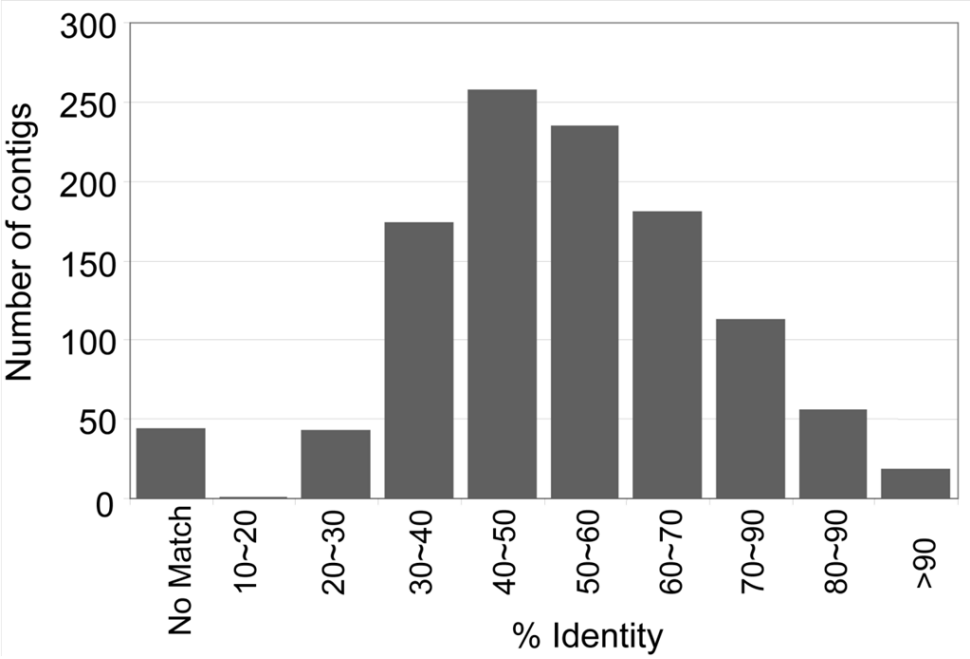

Figure 3.11 Amino acid sequence similarities of carbohydrate active enzymes identified from the muskoxen rumen metatranscriptome (using all assembled contigs) versus top BLASTX match to the Genbank nr database.

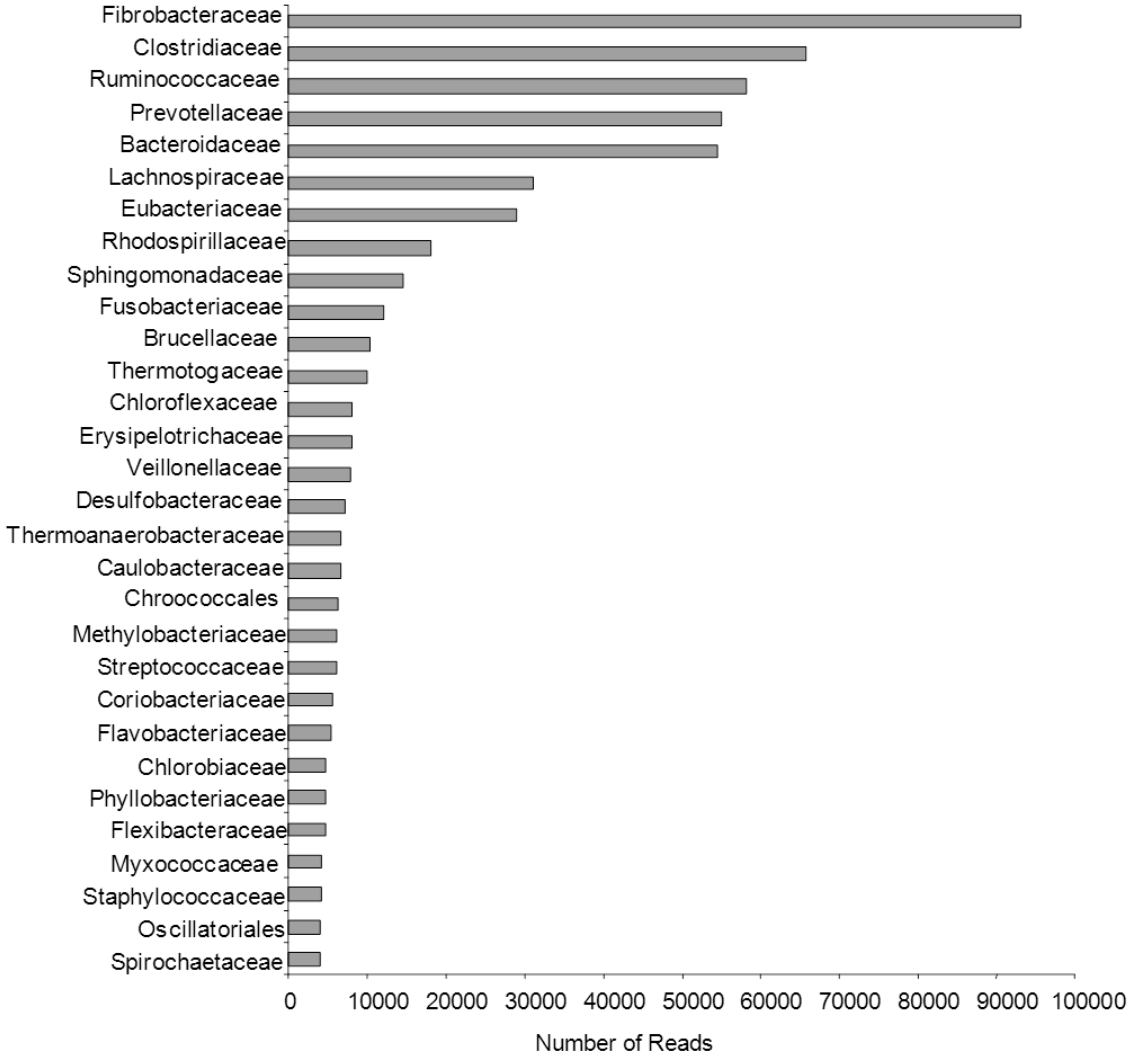

Figure 3.12 Top 30 phylogenetic bins of the bacterial reads of muskoxen rumen metatranscriptome as determined by comparison against NCBI's non-redundant amino acid (nr) database.

Ranks are determined by the highest number of total reads at the family level.

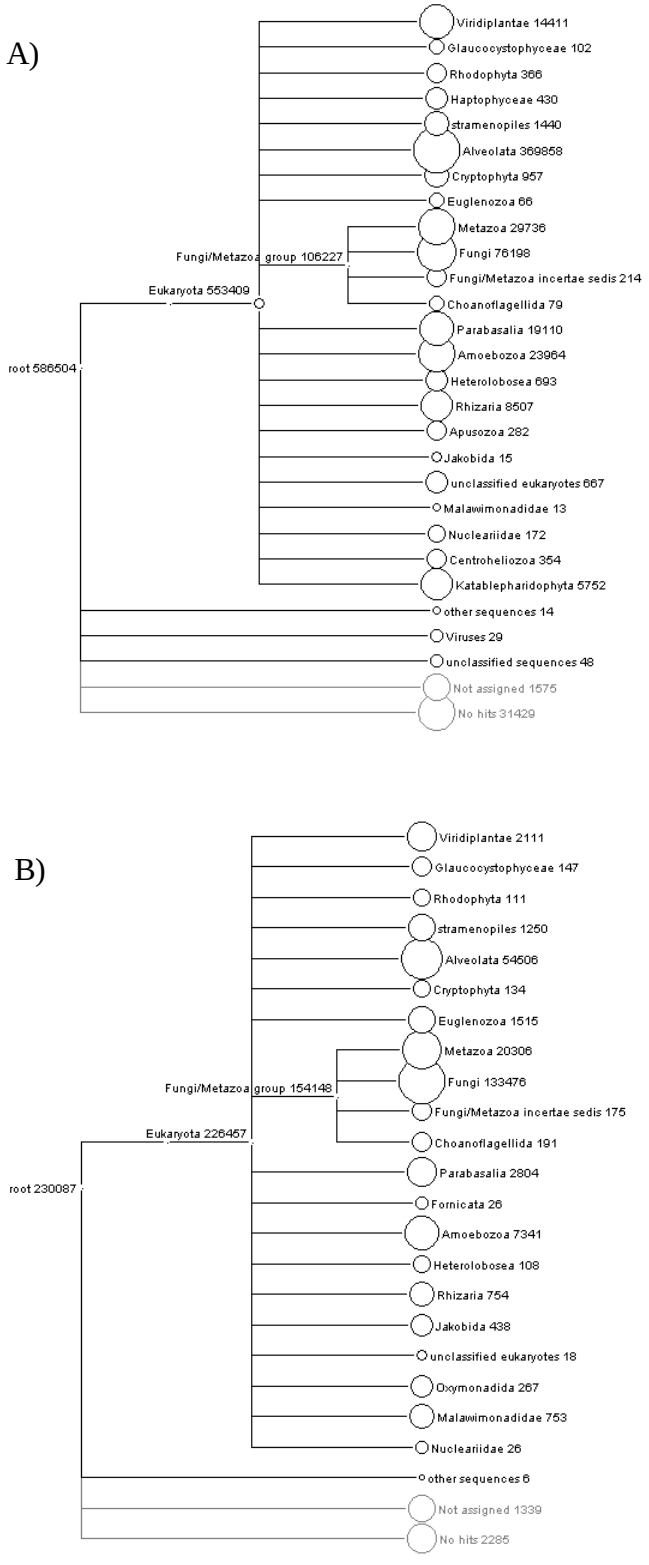

Figure 3.13 See page 122 for caption.

C)

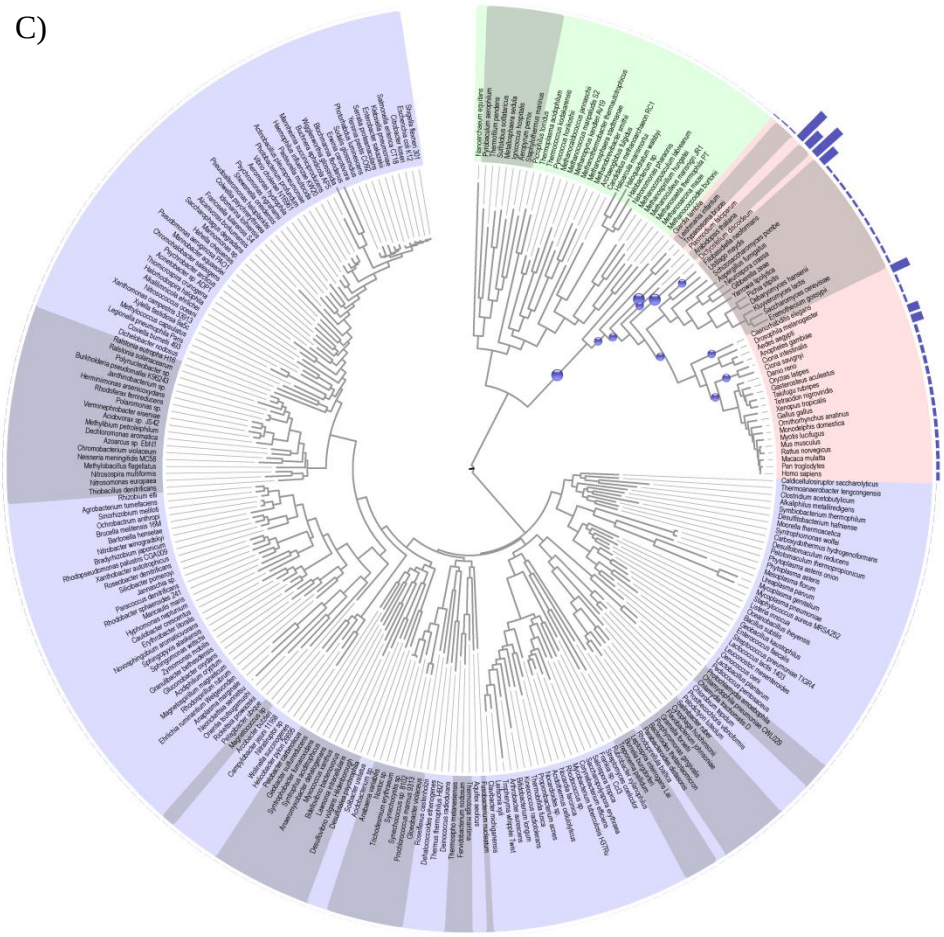

Figure 3.13 Evidence of eukaryotic origin of the metatranscriptome sequences based on BLASTX searches.

A): reads that were assigned to actin (KOG0676).

B): reads that were assigned to translation elongation factor EF1 (KOG0052).

C): MLTreeMap analysis of all the contigs. Eukaryotes are colored pink, archaea green and bacteria blue.

Number of reads that matched to each node are indicated in A and B.

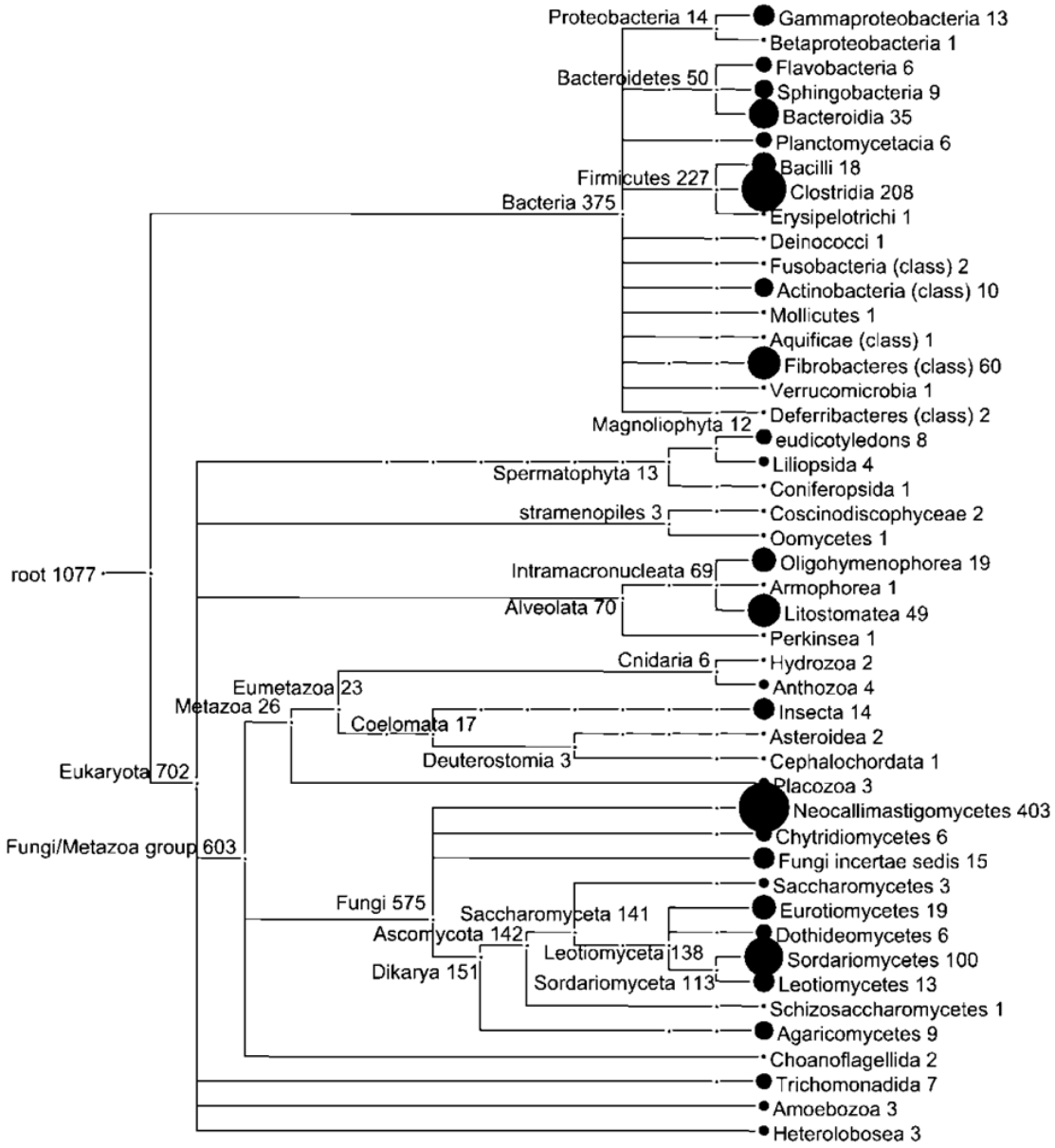

Figure 3.14 Phylogenetic distribution of muskoxen rumen metatranscriptome putative carbohydrate active enzymes based on MEGAN analysis of top BLASTX hits of the contigs against the Genbank non-redundant amino acid (nr) database.

The number of contigs (≥500 bp) that matched to each node is indicated.

Chapter 4 Comparative transcriptomic analysis of *Anaeromyces mucronatus* YE505

4.1 Introduction

Ruminants rely primarily on the microbial community in the rumen to break down plant polysaccharides and ferment released soluble sugars. Bacteria, protozoa and fungi are the key participants involved in cellulose degradation in the rumen. Within the past ten years, genomes of a number of rumen bacteria have been sequenced, and analyses of these genomes have indicated that rumen lignocellulolytic bacteria have many unique features in their cellulase system as compared to their aerobic counterparts (Berg Miller et al., 2009; Purushe et al., 2010; Suen et al., 2011a; Suen et al., 2011b). Many more previously unknown cellulolytic enzymes have been identified via bioinformatic analysis, highlighting the potential of this ecosystem to provide unique carbohydrate active enzymes for agricultural and commercial usage.

In contrast, although some enzymes have been detected or isolated from rumen anaerobic fungi, genomic information on these poorly researched microorganisms is lacking. Rumen fungi are strict anaerobes and widely distributed in the gastrointestinal tracts of many domestic and wild ruminant and non-ruminant herbivores (Gordon and Phillips, 1998; Liggenstoffer et al., 2010). Their high efficiency in plant cell wall degradation has been documented and it is undeniable that they play an important role in rumen function (Borneman and Akin, 1994). It is now generally known that the degradation of plant cell wall polysaccharides by rumen fungi accelerates digestion by decreasing the particle size of plant tissues, as not only do they produce highly active

fibrolytic enzymes, but are also capable of physically disrupting the plant cell walls including the cuticle via penetrating hyphae of vegetative thalli (Nagpal et al., 2009; Orpin and Joblin, 1988). Zoospores, the mobile phase of the fungal life cycle, also preferentially colonize lignin-rich regions of the plant cell wall, establish colonies localized on sclerenchyma and xylem cells and solubilize these regions to a greater extent than rumen bacteria (Akin and Borneman, 1990). As lignin is not degraded under anaerobic conditions, hydrolysis of both the ester linkages and sugar residue branches existing in hemicellulose plays a key role in the solublization of lignin and exposure of hemicellulose to microbial xylanases. Thus, fungi play a key role in facilitating plant cell wall degradation by other members of the rumen microbial community.

Anaeromyces is one of six genera of anaerobic fungi currently identified (Li et al., 1993; Liggenstoffer et al., 2010). It has two defined species originally isolated from the ovine rumen: *Anaeromyces mucronatus* (Breton et al., 1990) and *Anaeromyces elegans* (Ho et al., 1993). The genus *Anaeromyces* is characterized by a polycentric thallus, a polynuclear rhizomycelium and uniflagellated zoospores. *A. mucronatus* is known to produce a broad range of intracellular and extracellular enzymes involved in the degradation of plant structural and storage polysaccharides including cellulase, xylanase, β-glucosidase, mannosidase, amylase and chitinase (Novotná et al., 2010; Yang and Yue, 2012). Like other rumen fungi, the genome of *A. mucronatus* has an extremely rich AT content (GC% approx. 20%), and this property greatly increased the difficulties associated with molecular manipulation and sequence analysis (Chen et al., 2006; Nicholson et al., 2005). Based on molecular analysis, *Anaeromyces* clusters separately from *Orpinomyces*, *Neocallimastix*, and *Piromyces* (Li and Heath, 1992) suggesting that

it may possess some unique attributes. Presently very few studies have characterized the genomic nature of this unique microorganism (Qi et al., 2011).

In this chapter, the strain *Anaeromyces mucronatus* YE505 was investigated for its ability to utilize various carbon sources, and the transcriptome produced on each carbon source was characterized using an Illumina sequencing platform. Subsequently, the transcriptomes were analyzed comparatively through assembly of the RNA-Seq short reads into full length ORFs to generate a comprehensive view of the plant cell wall degrading enzymes and their regulation. By applying tandem mass spectrometry analysis a number of putative CAZy proteins in the extracellular culture fraction were also identified.

4.2 Material and methods

4.2.1 Chemicals

Unless otherwise noted, all the chemicals utilized were reagent grade or higher, purchased from either Sigma Aldrich (Oakville, Ontario, Canada) or Fisher Scientific (Ottawa, Ontario, Canada).

4.2.2 Fungal strain and culturing conditions

A. mucronatus YE505 was originally isolated from an elk (Hausner et al., 2000) and was grown anaerobically at 39 °C in modified semi defined Lowe's medium B (Lowe et al., 1985) with 0.67% (wt $\cdot$vol^{-1}) of one of the following carbon sources added: 1) Glucose, 2) Cellobiose, 3) Glucose-cellobiose-starch (GCS, weight ratio 1:1:1), 4) Avicel cellulose, 5) Oat spelt xylan, 6) Barley straw, or 7) Alfalfa hay.

For growth curve and extracellular enzyme activity assays, gas volumes produced were measured and liquid cultural samples were taken from three cultural replicates at the same time point every 24 hours for a period of 10 days. Samples were frozen immediately at -20 ℃ prior to further analysis.

4.2.3 Enzyme assays

Enzyme assays were carried out in 50 mmol L^{-1} sodium phosphate buffer, pH 6.5 at 37 ℃ unless otherwise stated. One unit (U) of enzyme activity was defined as one μmol of product released (glucose equivalent, *p*-nitrophenol or α-naphthol) per minute. Background corrections were made by subtracting the readings for assays conducted with heat-inactivated enzymes.

Glycoside hydrolase activities on polysaccharides were assayed by incubating appropriately diluted enzyme samples in an assay mixture containing one of the following substrates: 1% (wt·vol^{-1}) low viscosity carboxymethylcellulose (CMC), 1% Avicel cellulose, 1% oat spelt xylan, 1% starch or 1% lichenan. A standard incubation period of 1 h was used. Released reducing sugars were detected by the *p*-hydroxybenzoic acid hydrazide (PAHBAH) method as described by Lever (1972). The absorbance of each assay mixture was read at 420 nm using a Synergy – HT microtitre plate reader (BioTek, Winooski, VT). The amount of reducing sugar produced by the enzyme was calculated by reference to a glucose standard curve.

Glycoside hydrolase activities on arylglycosides were determined by incubation of appropriately diluted enzymes in an assay mixture containing 5 mmol L^{-1} of the substrate: *p*-nitrophenyl-*β*-D-cellobioside (pNPC) or *p*-nitrophenyl-*β*-D-glucoside (pNPG). The reaction was stopped by addition of an equal volume of 1 mol L^{-1} Na_2CO_3 and released

p-nitrophenol was determined by measuring the absorbance at 405 nm with a Synergy – HT microtitre plate reader using *p*-nitrophenol as the reference standard.

The acetyl esterase assay procedure described by Qi et. al. (2011) was used with the following minor modifications. In a 100 µl reaction mixture, 1 mmol L^{-1} α-naphthyl acetate (α-NA) was used as the substrate. After incubation at 37 °C for 30 min, 50 µl of 2.5 mg ml^{-1} (wt vol^{-1}) fast garnet GBC in 10% SDS was added. Absorbance at 560 nm was measured in a Synergy – HT microtitre plate reader with α-naphthol as the reference standard.

4.2.4 Isolation of total RNA

For isolation of mRNA for deep sequencing, *A.mucronatus* YE505 was grown anaerobically at 39 °C for 72 h in Lowe's medium B with 0.67% (wt vol^{-1}) of the various carbon sources described in section 4.2.2. The mycelia were harvested from the culture medium by vacuum filtration through four layers of cheesecloth and immediately frozen in liquid nitrogen. Total RNA was subsequently isolated from the mycelia by following the improved total RNA isolation procedure for solid rumen samples established in Chapter 2.

4.2.5 Sequencing and sequence assembly

Equal amounts of RNA extracted from fungal cultures grown on GCS, Avicel, xylan, barley straw and alfalfa hay were pooled together to constitute a mixed sample and sequenced. The RNA samples from fungal cultures grown on GCS, xylan, barley straw and alfalfa hay were also sequenced individually. Using the Illumina mRNA-Seq sample preparation kit according to the manufacturer's instructions (Illumina Inc, San Diego,

USA), mRNA libraries were constructed. High throughput sequencing was performed on an Illumina HiSeq 2000 sequencer system at the McGill University/Génome Québec Innovation Centre.

Obtained sequencing reads were assembled *de novo* with two assemblers: the Trinity assembler (Grabherr et al., 2011) using the "jellyfish" method fork-mer counting, and the Velvet assembler (Zerbino and Birney, 2008). The assembly with Velvet was done on a split dataset comprised of 45 sets of 2 million reads, then reassembled with CAP3 (Huang and Madan, 1999). Because the contigs from the Trinity and Velvet assembly are highly redundant, results from the two analyses were combined and a dataset with 95% sequence identity was obtained with the CD-HIT program (http://www.bioinformatics.org/cd-hit/) (Li, 2009). These contigs were translated in all six reading frames and those that comprised at least 150 amino acids and possessed a start and stop codon were considered to be a potential open reading frame (ORF).

4.2.6 Bioinformatic sequence analysis

All sequence analyses, unless otherwise specified were performed using the assembled full length ORFs. The databases employed for this analysis were the latest versions available during the analysis period (May 2012 to Sep 2012).

The predicted ORF sequences were searched using RPS-BLAST against both the KOG and the COG databases (Tatusov et al., 2003) and the GenBank non-redundant amino acid (nr) database. The functional roles of the sequences were assigned based on KOG and COG searches. Matches that had E-values less than or equal to 10^{-5} were considered significant. CAZy protein annotation was performed by both HMMER3 (Eddy, 2009) and BLASTX searches following the procedures described in Chapter 3.

Relative transcript expression levels were calculated by the FPKM (Fragments Per Kilobase of transcript per Million mapped) method using Cufflinks software (Trapnell et al., 2010). Expression data analysis was performed using Spotfire Software (Spotfire Inc., Somerville, MA, USA). Cluster analysis of the log_2-transformed transcript expression data obtained from CAZymes was carried out using UPGMA (Unweighted Pair Group Method with Arithmetic Mean) analysis, with an ordering function based on the input rank. The resulting dendrograms were generated with the closest branches of the tree representing samples with similar gene expression patterns.

4.2.7 Extracellular protein preparation

Extracellular proteins were collected at the same time as mycelia were collected for total RNA isolation by collecting the liquid fractions, which were filtered through cheesecloth as described in section 4.2.4. The liquid fraction was centrifuged at 10,000 x g for 30 min and supernatants were concentrated ~100-fold by ultra filtration using an Amicon concentrator with an Ultracel YM-10 membrane (Millipore, Billerica, MA, USA). The samples were further cleared of particulate matter by centrifugation at 3200 x g for 30 min at 4 ℃. Proteins were precipitated by adding 1 equal volume of precipitating solution (0.2% dithiothreitol, 20 % trichloroacetic acid in acetone) followed by incubation for 1 h on ice. The pellet was subsequently washed with prechilled 20 mmol L^{-1} DTT/80% acetone (-20 ℃).

Pellets were resuspended in 2 x Laemmli buffer (~ 50 µL) and quantified using the RCDC Protein Assay Kit (Bio-Rad, Mississauga Ontario, Canada). Isolated protein (15 µg) was loaded in each lane of a 10% SDS-PAGE gel. After electrophoresis and visualization by staining, the whole lane was excised, proteins were destained, reduced,

cysteine-alkylated and in-gel digested with trypsin over-night as previously described (Wasiak et al., 2002).

4.2.8 LC-MS/MS and peptide identification

The peptide extracts of each corresponding lane were subjected to LC-MS/MS on a Velos LTQ-Orbitrap (ThermoFisher Scientific, San Jose, CA, USA) instrument at McGill University and Génome Québec Innovation Centre.

Mass spectrometric data were acquired by employing the Data Dependent Scans from the Xcalibur 2.1 software (Thermo Fisher Scientific, San Jose, CA, USA). Raw data from LC-MS/MS were processed with the Proteome Discoverer 1.3 software (Thermo Fisher Scientific). The peaklist files were searched against the generated database of the *de novo* assembled RNA-Seq contigs. SEQUEST (Thermo Fisher Scientific) was used to carry out peptide identification using a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 10.0 ppm. An iodoacetamide derivative of cysteine was set as static modification and oxidation of methionine was specified as a variable modification. Peptide identifications were accepted if they could be established at high probability and with a SEQUEST score greater than 2.5. Protein identifications were accepted if they could be established at high probability and at least two unique peptides assigned.

4.3 Results

4.3.1 Growth of *Anaeromyces mucronatus* YE505 on different carbon sources and enzyme activity in extracellular culture fluid

The ability of *A. mucronatus* to grow on a number of carbon sources in a chemically semi-defined medium designed by Lowe (Lowe et al., 1985) was tested using

glucose, cellobiose, GCS, xylan, avicel, barley straw and alfalfa hay. Rumen fungi produce both hydrogen and carbon dioxide during fermentation so the volume of gas produced can be used as a crude indicator of active growth. As demonstrated by the gas production curves in Figure 4.1, YE505 yielded approximately 2 fold higher gas volume on GCS or xylan in comparison to alfalfa hay or barley straw. Growth on Avicel had an extended lag phase lasting nearly 96 h; however, culture yields approached those achieved on GCS and xylan after 8 – 10 d growth. Suprisingly, little growth was detected on media containing only glucose or cellobiose as the sole carbon source. On most substrates, the highest rate of gas production occurred between day 2 and day 5. This period of rapid growth was shifted by 3 d when YE505 was grown on Avicel (Figure 4.1). These results were also verified by visual observation of mycelial growth inside the culture container. Thus, thereafter only the five carbon sources (GCS, Avicel, xylan, barley straw and alfalfa hay) were selected as growth substrates for futher studies with *A. mucronatus* YE505.

As the first step to define the expression level of carbohydrate degrading enzymes under different culture conditions, the extracellular cultural fluid (ECCF) of *A. mucronatus* grown on five carbon sources was sampled over a period of 10 days, and the samples' activities toward eight enzyme substrates were measured as described in section 4.2.3 (Figure 4.2 - Figure 4.9).

Enzyme activities detected in the ECCF against the substrate Avicel were barely detectable regardless of the carbon source used to support fungal growth (Figure 4.2). The activities on CMC (Figure 4.3) - endoglucanase activities, on lichenan (Figure 4.4), on starch (Figure 4.5) and on xylan (Figure 4.6) were found to increase rapidly after 24 h,

and were generally higher when the fungus was grown on alfalfa hay and barley straw as compared to other carbon sources. Interestingly, *A. mucronatus* grown on Avicel after 8 days showed comparable CMCase, xylanase and lichenase activities as those grown on alfalfa hay and barley straw. This was consistent with the gas production (Figure 4.1), in that when *A. mucronatus* was grown on Avicel it required a longer time to reach the active growth phase. Amylase activities remained low in ECCF from Avicel even after 8 days. An interesting aspect was that when grown on GCS *A. mucronatus* did not exhibit high amylase activity in ECCF even though the medium contained starch. Growth on xylan generated xylanase activities that were higher than GCS, comparable to Avicel (for the first 6 days), but lower than alfalfa and barley straw. The xylanase activities in GCS medium were barely detectable over the 10 day experiment.

For the synthetic substrates tested, the same trends were observed: enzyme activities were observed sooner in the incubation period and were higher when *A. mucronatus* was grown on alfalfa hay or barley straw. Compared to other synthetic substrates, activities to pNPC remained low throughout 10 day incubation, regardless of substrate (Figure 4.7). β-glycosidase activities detected by the pNPG assay were comparably high in alfalfa and barley straw, with that from Avicel reaching similar levels after 7 d of incubation (Figure 4.8). *A. mucronatus* grown on alfalfa exhibited the highest esterase activity to α-NA, being two fold higher than that grown on barley straw (Figure 4.9), but growth as reflected by gas production was comparable on these two substrates (Figure 4.1).

Generally speaking, the enzyme activity levels started to increase on day 2 and reached the highest level after day 4 or day 5, which was in agreement with the active

growth period demonstrated by gas accumulation. *A. mucronatus* grown on alfalfa produced the highest detectable ECCF activities to five substrates (CMC, lichenan, starch, xylan and α-NA); whereas activity against pNPG was higher in barley straw.

Based on the growth curve and enzyme activity profiles, we chose 96 h for Avicel and 72 h for the rest of carbon sources as the time points to harvest mycelia for RNA extraction, as well as ECCF collection. By choosing these time points, we were able to obtain samples during the most active growing period that yielded sufficient levels of RNA and extracellular protein for analysis.

4.3.2 Sequencing summary, assembly and BLAST analysis

In the present study, the RNA-Seq approach was applied to identify ORFs coding for proteins and to define the level of transcript expression in *A. mucronatus* YE505 grown on different carbon sources. Total RNA was extracted from mycelia and after purification, the mRNA samples produced on four different carbon sources (i.e., GCS, xylan, alfalfa hay and barley straw) were sequenced using an Illumina HiSeq 2000 sequencer. Unfortunately, the mRNA sample obtained from *A. mucronatus* grown on Avicel was not sequenced individually; however, it contributed 20% of the RNA in the mixed samples, and consequently contributed to the assembled sequences from this sample.

A total of 89,241,369 paired end 100 x 100 base reads were obtained. The raw sequencing reads were assembled with both the Trinity and Velvet assemblers and the resultant contigs were combined. Potential full length ORFs were predicted as described in section 4.2.5. A total of 6,670 full length ORFs were obtained, with an average length of 1,427 bp, for a total of 9.52 Mb. Size distribution of the ORFs is illustrated in Figure

4.10 with the longest ORFs being 16.5 kb encoding for a hypothetical protein with approximately 5,500 amino acid residues. This protein showed high similarity to a hypothetical protein from another rumen fungi, *Orpinomyces* sp. OUS1 (Nicholson et al., 2005). A large hypothetical protein showing similarity to this protein was also detected within the muskoxen rumen eukaryotic metatranscriptome described in Chapter 3. The average GC content of all the ORFs was 28.8%, which was expected owing to the AT-rich genomes of rumen fungi.

A BLAST search was performed using all the potential ORF sequences against the nr database from GenBank. A total of 1,200 sequences from *A. mucronatus* were most similar to sequences from the chytrid *Batrachochytrium dendrobatidi*s. BLAST searches also identified *A. mucronatus* homologues to over 600 sequences from *Rhizopusoryzae* and ~160 from *Myceliophthora thermophila*. Interestingly, over 600 ORFs showed high similarity to bacterial derived genes as opposed to those originating from eukaryotic species.

4.3.3 Functional analyses of ORFs identified

Based on RPS-BLAST search results, 3,808 ORFs could be assigned to a cluster of the KOG database (Figure 4.11, pane A). From the rest non-KOG matched candidates, further RPS-BLAST identified 422 ORFs that matched to the bacterial genome derived COG database (Figure 4.11, pane B).

The KOG/COG databases classified each ORF into one of 20 groups (Tatusov et al., 2003). Proteins involved in “translation, ribosomal structure and biogenesis” had the highest total FPKM number (Figure 4.11), indicating that the fungal mycelia were harvested during active growth. Particularly, AmuTC4, an ORF of 1,359 bp encoding for

α unit of the elongation factor 1 that delivers aminoacyl-tRNAs to the ribosome during protein translation, was represented by a total FPKM of 136,764, or 5.3% of all the sequencing reads. Many putative genes coding for roles in "carbohydrate transport and metabolism", and "energy production and conversion" were also highly expressed.

Most of the Embden-Meyerhof-Parnas pathway enzyme sequences were recovered from our dataset. Some of these ORFs were transcribed at very high levels. Among these, ORFs coding for enolase, glyceraldehyde-3-phosphate dehydrogenase and aldolase were the most highly expressed. Genes involved in the partial TCA cycle were also highly expressed including a phosphoenolpyruvate carboxykinase and an NAD-dependent malate dehydrogenase.

Many ORFs with a very high FPKM were not members of the above-mentioned categories. One group included genes involved in rhizoidal growth. Among these, ORFs coding for actin and tubulin, important cytoskeleton components in actively growing fungi, were represented by FPKMs of 36,853 and 21,510, respectively. Several putative genes that function in amino acid transport and metabolism were also highly expressed, including those encoding ketol-acid reductoisomerase (which is known to be involved in the biosynthesis of leucine/isoleucine/valine), glycine/serine hydroxymethyltransferase, lysine-ketoglutarate reductase-saccharopine dehydrogenase, and argininosuccinate synthase.

The function of some highly transcribed ORFs, like AmuTC14, could not be accurately predicted by KOG/COG analysis. This putative gene was annotated as coding for a sugar transporter, but without biochemical characterization, its substrate preference could not be predicted.

4.3.4 CAZymes prediction

Like other rumen anaerobic fungi, *A. mucronatus* degrades a number of plant cell wall polysaccharides including cellulose, xylan, mannan and pectin (Novotná et al., 2010). From the RNA-Seq data, a total of 344 ORFs containing at least one CAZy domain were identified, including glycoside hydrolases (GHs), carbohydrate esterases (CEs), pectin lyases (PLs) and carbohydrate binding modules (CBMs) (Figure 4.12, Table S.3). The length of the ORFs for CAZy genes ranged from 491 bp to 5,916 bp, with an average coding sequence length of 1,592 bp.

Translated BLAST searches against the GenBank nr database showed that only 12% of the products coded by these ORFs were more than 70% identical to proteins in the nr database, while 59% of them showed an identity of less than 50%. Altogether 121 ORFs coded for two or more distinct CAZy domains, among which, CBM10 like domains were in 93 ORFs, accounting for 7% of over 6,000 total full length ORFs obtained.

4.3.4.1 Catalytic modules

Most of the putative cellulases identified from *A. mucronatus* transcriptomes were classified as members of families GH5, 6, 9, 45 and 48, representing a total of 35 ORFs. In addition, for the first time from an anaerobic fungus, five ORFs were predicted to code for swollenin, a protein without hydrolytic activity that has been reported to dissociate cellulose fibers (Brotman et al., 2008). Open reading frames matching GH families containing β-glycosidases, including GH1 and GH3, were also identified. The ORF AmuTC51 harbouring a potential GH1 module was highly expressed on alfalfa, xylan and barley straw. The encoded protein was 72% identical to a β-glycoside hydrolase from

Piromyces sp. E2 that was also highly expressed when this fungus was grown on fructose (Harhangi et al., 2002).

Xylan and other hemicellulose degrading enzymes were identified from 39 ORFs. They were mostly from GH families 10, 11, 26, 39 and 43. In this study, seven GH10 and six GH11 ORFs were identified. All the GH10 enzymes were associated with at least one CBM. Sixteen ORFs that contained GH43 domain were identified. Two ORFs encoding putative xylosidases from GH39 were identified, both of which contained a CBM13 and a CBM10 module. AmuVC11847, an ORF coding for a GH8 module coupled with two tandem CBM10 modules, was also identified. The predicted gene product was 74% identical to a putative xylanase in *Fibrobacter succinogenes* (Suen et al., 2011b). To our knowledge, this is the first GH8 family member reported for a rumen fungus.

There were 59 ORFs showing similarity to carbohydrate esterases in the CAZy database. The majority of these (20 ORFs) were associated with family CE4. The CE1 family, which contains feruloyl esterases, was represented by 10 ORFs. A total of 13 ORFs coded for family CE10 CAZymes, which includes pectin acetyl esterases. Two CE15 ORFs encoding for 4-*O*-methyl-glucuronoyl methylesterases were also identified. As with feruloyl esterases, CE15 family enzymes cleave cross linkages between lignin and the xylan backbone.

Other catalytic modules identified included pectin degrading enzymes from GH53 (endo-β-1,4-galactanase) as well as from the polysaccharide lyase (PL) family 1, 2, 4 and 9.

4.3.4.2 Accessory modules

The most prominent accessory module was CBM10. A total of 350 CBM10 modules were identified in 162 ORFs. Most of the ORFs have 2~3 tandem CBM10 domains. A total of 91 CBM18 domains were identified in 40 ORFs. The number of CBM18 coded by one ORF ranged from one to 15. Other major carbohydrate binding modules include CBM1 (35 ORFs), CBM13 (20 ORFs), CBM6 (6 ORFs) and CBM29 (2 ORFs).

4.3.4.3 Effect of carbon source on CAZyme gene transcription in *Anaeromyces mucronatus* YE505

The expression of plant cell wall degrading enzymes in *A. mucronatus* YE505 was influenced by carbon source (Figure 4.12). For all carbohydrate active enzyme transcripts identified, the average FPKMs ranged from 71 in GCS to 206 in barley straw, 209 in oat spelt xylan and 321 in alfalfa hay. Generally speaking, the majority of predicted CAZymes were most induced when *A. mucronatus* was grown on alfalfa hay, followed by barley straw and xylan, and the lowest when grown on GCS.

The transcript expression levels of individual ORFs differed considerably, regardless whether or not they belonged to the same CAZy family. For example, ORFs AmuTC352 and AmuVC1295 both belong to GH11. Expression of AmuTC352 was higher when *A. mucronatus* YE505 was grown on GCS or xylan, while expression of AmuVC1295 was 25 and 125 times higher on alfalfa hay and barley straw as compared to expression on GCS and xylan, respectively. This strongly suggests that there is no one-strategy-fits-all for regulation of gene expression within GH families. On the contrary, individual genes are likely subject to differential regulation individually. The same

situation was observed for CAZymes with the CBM10 modules, where the ratio of the FPKM for growth on alfalfa, straw and xylan as compared to GCS ranged broadly from 0.04 (repressed) to 1,260 (induced).

To better illustrate the complex expression patterns of different CAZymes, ORFs with similar expression patterns were grouped using the UPGMA method and normalized expression levels (i.e., $\log_2$FPKM; Figure 4.12). The gene transcription profile for *A. mucronatus* YE505 grown on GCS differed substantially from those grown on xylan, barley straw or alfalfa hay. Expression patterns were most similar for cultures grown on barley straw and alfalfa hay. The CAZyme transcripts grouped into seven clades according to expression levels and patterns when *A. mucronatus* YE505 was grown on different substrates.

Clade A, C and E generally showed less variation among the four carbon sources. In clade A, 10 ORFs were expressed at high levels regardless of substrate, with average FPKM ranging from 1,227 for GCS to 4,246 for alfalfa hay. This group included AmuTC99, an ORF encoding a putative GH43 β-xylosidase that cleaves xylo-oligosaccharides into xylose. The transcript expression level for this ORF was very high, especially when grown on xylan (Figure 4.12). Within this clade, there were two CAZymes belonging to GH6 and GH48 with putative exoglucanase activity. Intriguingly, ORF AmuTC72 contained two CBM29 and three CBM10 domains, but no catalytic module. Two CE4 esterases and two CAZymes harbouring a CBM13 and a CBM10 module were also identified within this group of 10 enzymes.

Clade C was a large group that contained 62 CAZymes, which were expressed at lower levels than clade A. Average FPKMs within this group ranged from 118 to 439 and were comparable across substrates.

Clade E also represented a large group containing 72 ORFs that were moderately expressed with average FPKMs ranging from 47 to 119. Interestingly, some of the ORFs (eg., AmuVC2993 and AmuVC4852) in this clade were down regulated when *A. mucronatus* was grown on alfalfa hay or barley straw.

Eight CAZymes belonging to Clade B were induced when *A. mucronatus* was grown on barley straw or alfalfa hay, as illustrated by the dramatic increase in average FPKMs 17 for GCS to 994 in straw and 2,030 in alfalfa. ORFs in this clade included a GH6, a GH10 and a GH11 member. Within this clade, two esterases from CE6 and CE15 were each linked to a CBM10 module. Three CBM10 harbouring ORFs without known catalytic domains were also members of this clade.

Similarly, the 11 CAZymes in clade D were expressed at low levels with GCS, compared to FPKM values 100 fold higher for barley straw and alfalfa hay. The ORFs in this clade were more difficult to annotate with only two catalytic domains (a GH8 and a CE4) being identified. The remainder of CAZy proteins in this clad appeared to be CBMs, which appeared to be upregulated when *A. mucronatus* was grown on complex plant fibers.

Similar to clades B and D, Clade F contained 56 ORFs that were induced by xylan, straw and alfalfa. Their expression levels as well as their regulation tended to be lower and less stringent with less variation between different carbon sources, indicating they may play a less important role in fiber degradation.

The remainders of the putative CAZy genes were expressed at low levels with FPKMs averaging from 8 to 35, and were grouped into clade G. Two of these CAZymes, a GH26 member and a CE15 member were expressed at much higher levels when *A. mucronatus* was grown on alfalfa hay than on other carbon sources.

4.3.5 Secretomic analysis of *Anaeromyces mucronatus* YE505 grown on different carbon sources

To further characterize the plant cell wall degrading enzyme system present in the ECCF, a secretomic study was performed to analyze the ECCF protein profiles produced on five different carbon sources. Together with the other four samples, the Avicel grown ECCF protein sample was also subjected to LC-MS/MS and secretomic analysis. In total, over 3,400 peptide spectra were identified, which matched to 341 ORFs, including 103 ORFs belonging to predicted CAZy proteins (Table 4.1). Transcripts with higher FPKM values as determined by RNA-Seq were more likely to have a protein counterpart identified by protein mass spectroscopy, such as AmuTC51, AmuTC352 and AmuTC99. Interestingly, alfalfa hay or barley straw grown samples generated the fewest number of transcript/peptide matches. Growth on Avicel yielded a peptide subset of CAZy proteins that were not detected in the other samples (Table 4.1).

4.3.6 ORFs without predicted function

After combined search of COG/KOG, pFAM, CAZy and nr databases, about 2,200 ORFs still could not be assigned a function. This accounts for approximately one third of the total ORFs obtained. BLAST searches against the nr database showed about 550 of these coded for conserved hypothetical proteins, which matched to proteins with

unknown function from other organisms. The remaining 1,650 ORFs showed no match to any reported sequence. As these ORFs were assembled from mRNA, they are likely to represent functional genes. Some of these ORFs, such as AmuVC11385 and AmuTC42 were expressed at higher levels when *A. mucronatus* was grown on alfalfa hay, barley straw and xylan. These genes may be involved in some unknown pathway of plant cell wall digestion and warrant future investigation.

4.4 Discussion

This study identified a number of CAZymes from different families with differing substrate specificity. CAZymes act synergistically to enable rumen fungi to degrade recalcitrant plant cell wall polysaccharides (Blum et al., 1999). Difference in expression levels with different carbon sources was clearly demonstrated by the grouped $\log_2$FPKM patterns in Figure 4.12. Logarithm scale ($\log_2$) analysis is an accepted and widely applied approach for normalizing relative expression levels obtained through microarray and RNA-Seq analysis of microorganisms grown on different substrates (Brooks et al., 2011; Marioni et al., 2008).

The extracellular enzyme activities were calculated based on the ECCF volume instead of the protein mass, since the protein concentrations in the ECCF were too low to be accurately measured, especially when components in the medium could interfere with the protein concentration assay. Apparently, growth rate would influence the amount of proteins secreted, and consequently enzyme activities detected. But on the contrary, despite the relatively moderate growth rates, the enzyme activities on alfalfa hay and barley straw were generally the highest, being more than 10 fold higher than those on GCS in the case of xylan and α-NA activities (Figure 4.6, Figure 4.9) during the active

growing period. Meanwhile, the esterase activities were ~ 2 fold higher with alfalfa hay than with barley straw while the gas production of *A. mucronatus* on these two substrates was comparable. This higher level of induction was probably because alfalfa cell wall contains more ester linked branches compared to barley straw (Varel et al., 1989). Therefore, when considering the growth rate influence, differences in enzyme activities between complex carbon sources and GCS would likely have been even greater than detected and were most probably due to differences in gene expression. This observation is in agreement with the expression patterns identified from RNA-Seq, as generally CAZymes were more highly expressed when grown on alfalfa hay or barley straw.

The LC-MS/MS procedure applied in this study is not considered quantitative, but still there was a weak correlation that the ORFs with high FPKMs were more likely detected by LC-MS/MS. Occassionally more peptide spectra from Avicel, xylan or GCS grown samples were matched to one ORF than those from alfalfa hay or barley straw, regardless whether it was up-regulated by the latter carbon sources (Table 4.1, Figure 4.12). This was probably because the protein samples grown on simple carbon sources (Avicel, xylan and GCS) contained fewer impurities and was more suitable for mass spectral detection in preparations.

It has been previously reported that xylanases were predominant among the glycoside hydrolases in *A. mucronatus* strain KF8 (Novotná et al., 2010). Our findings through activity assays and transcriptomic sequencing agree with this report. Among all the predicted CAZy families, GH43 enzymes coding mainly for xylaneses or xylosidases had the highest FPKMs, together with some GH10 and GH11 xylanase candidates. But contrary to enzyme assays, which showed that xylanase activities were highest with

alfalfa hay followed by barley straw (Figure 4.6), the RNA-Seq analysis showed that fungi grown on xylan generated the highest xylanase FPKM values (Figure 4.12). And very surprisingly, xylanase activity was almost undetectable in the ECCF from *A. mucronatus* grown on GCS (Figure 4.6), even though considerable xylanase expression was detected in GCS grown transcriptome. One possible explanation is that direct enzyme assays only reflected those xylanases that were secreted into the medium, wherease RNA-Seq detected the overall expression including those from intracellular, cell-associated and extracellular xylanases. This may well reflect the possibility that a large portion of xylanases are tightly associated with the fungal cell surface. Another aspect that is worth pointing out is that the existence of highly expressed enzymes does not always equal to high enzyme activities, since specific activities can vary considerably between different enzymes. It is also possible that genes may not be effectively translated to proteins even when they are highly transcribed. All these factors contributed to the difficulties for clearly elucidating and comparing all the details in transcriptomic studies.

Xylanases from GH10 and GH11 families have been previously cloned and characterized from rumen fungi (Black et al., 1994; Gilbert et al., 1992). But to date, other than a report from *Neocallimastix patriciarum* (Wang et al., 2011), no GH43 enzymes have been identified from rumen fungi. The enzyme assay also detected very high esterase activities when *A. mucronatus* YE505 was grown on alfalfa hay as well as on barley straw (Figure 4.9). Indeed, transcriptomic sequencing revealed that CE4 members were highly expressed, together with CE6 and CE15 members also contributing activities (Figure 4.12). These highly expressed ORFs such as AmuTC99 (GH43, with 38 Mass Spec peptide matches), AmuTC468 (CE4) and AmuVC9821 (CE15, which was

highly induced in alfalfa hay) suggest that they have an important function in plant cell wall degradation and are good candidates for future detailed biochemical analysis considering their uniqueness and high level of expression.

Other than the ordinary CAZy members involved in different stages of plant cell wall degradation, a particular group of hypothetical swollenins was of great interest to us. Swollenin was firstly identified from *Trichoderma reesei* (Saloheimo et al., 2002) and was subsequently identified in several other aerobic fungal species (Chen et al., 2010; Yao et al., 2008), and exhibits sequence similarity to plant expansins. Swollenin is believed to disrupt the structure of crystallized cellulose by breaking hydrogen bonds between cellulose fibers, without detectable formation of reducing sugars. Pretreatment of cellulose with swollenin decreased the particle size of cellulosic substrates as well as cellulose crystallinity and increased cellulose hydrolysis rates by cellulases (Jäger et al., 2011). In the present study, five putative swollenin genes were identified, all of which were also attached to CBM10-like modules. Four of these rumen fungal putative swollenin genes showed higher FPKM values when the fungus was grown on alfalfa hay, barley straw or xylan as compared to GCS medium. Swollenin like sequences were also identified from muskoxen rumen eukaryotes metatranscriptome as outlined in Chapter 3. All of these results suggest swollenin plays an important role in rumen fungal cellulose digestion and is worthy of further characterization.

A reason for anaerobic fungi's high cellulolytic capability may be attributed to the possible presence of cellulosomes. High-molecular-mass enzyme complexes (>700 kDa) have been described in *Piromyces*, *Orpinomyces*, and *Neocalimastix,* which contain as many as 15 protein components (Ali et al., 1995; Li et al., 1997; Wilson and Wood,

1992). This high-molecular-mass structure resembles the cellulosomes produced by several anaerobic bacteria such as *Clostridium* and *Ruminococcus* spp. in which various cellulases are attached to a protein called scaffoldin through a dockrin-cohesin interaction, resulting in a complex that is very efficient at cellulose degradation (Bayer et al., 2008). A CBM10 like, 40-amino-acid cysteine-rich, non-catalytic domain has been shown to be associated with many rumen fungal glycoside hydrolases, and was proposed to be a fungal dockerin by some researchers (Nagy et al., 2007; Raghothama et al., 2001; Steenbakkers et al., 2001). But to date, no cellulosome scaffoldin that mimics the bacterial counterpart has been identified from rumen fungi, although it has been shown that a CBM10 domain can interact with a GH3 enzyme in *Piromyces equi* (Nagy et al., 2007). This finding is surprising since unlike other scaffoldins in bacteria, this protein possesses only a catalytic domain and no identifiable cohesion domain. Although no promising candidate has been detected yet, transcriptome sequencing data generated from this study may facilitate the identification of the potential dockerin-interacting partners in the future as more evidence is accumulated.

Interestingly, it is still debatable whether or not the fungal CBM10 module is a true fungal dockerin. A number of previously identified anaerobic fungal genes have been shown to have one or more CBM10 like domains (Blum et al., 1999; Nagy et al., 2007; Steenbakkers et al., 2008; Steenbakkers et al., 2001). In the present study 350 CBM10 domains were identified in 162 ORFs (Figure 4.12). As expected, many of these ORFs had CAZy catalytic domains, including members with cellulase or xylanse activity from families GH1, 3, 5, 6, 9, 10, 11, 43, 45 and 48. Carbohydrate esterase family CE15 members were also associated with CBM10. But the fungal CBM10 module showed no

sequence homology to bacterial dockerins, and NMR structures of these modules from *P. equi* showed no structural similarity to bacterial dockerins (Nagy et al., 2007; Raghothama et al., 2001). In addition, the CBM10 module did not bind to either xylan or cellulose (Fanutti et al., 1995). Rather this module from *P. equi* recognized and bond to a glycosylated β-glucosidase via its oligosaccharide components (Nagy et al., 2007). Therefore the function of this module as dockerin remains to be further verified, with some researchers considering this module to be merely a fungal CBM with unknown carbohydrate preference (Peer et al., 2009).

A little surprisingly, no CAZy catalytic domains were identified from 78 of the CBM10 containing ORFs. The size of polypeptides encoded by these ORFs range from 161 to 1,750 residues, suggesting that it is likely that other non-CAZy catalytic functional domains may be present. In order to identify potential non-CAZy domains, BLASTP searches were performed using the 78 ORFs against the Genbank nr database. Over a half of them did not exhibit catalytic domains. Among them, the ORF AmuTC72 was highly expressed when YE505 was grown on xylan, barley straw or alfalfa hay and contained two CBM29 domains other than three CBM10 domains (Figure 4.12). The protein product of this transcript was also detected by LC-MS/MS (Table 4.1). Amino acid sequence of AmuTC72 showed 59% identity to NCP1 protein from *P. equi*, which was proposed to be a non catalytic protein that anchored the fungal enzyme complex onto the plant cell wall through CBM29-polysaccharide interaction (Freelove et al., 2001).

The other 31 ORFs possessed domains which shared similarities to serpin, CotH and Ser/Thr protein phosphatase, as well as swollenin as discussed above (Table 4.2). In seven ORFs, these CBM10-like domains were found to attach to a serpin-like protease

inhibitor. Serpins have been previously identified in the anaerobic fungus *Piromyces* sp. E2 (Steenbakkers et al., 2008). Fungal serpins are believed to be involved in protection of the cellulosome against proteases produced by plants or other microbes within the plant cell wall degrading community (Meguro et al., 2011; Steenbakkers et al., 2008). Similar protease inhibitors have also been identified in cellulosome producing bacteria such as *Clostridium thermocellum* (Zverlov et al., 2005), *Clostridium cellulolyticum* (Fendri et al., 2009) and *Clostridium cellulovorans* (Meguro et al., 2011). These proteins may play a critical role in conferring functionality to the cellulosome for the period of time that is required for plant cell wall degradation.

Another group of 13 ORFs contained a module loosely related to spore coat assembly protein (CotH) domain. CotH protein was first identified and characterized from *Bacillus subtilis* (Naclerio et al., 1996), which played a role in the assembly of at least nine other coat components into endospore protein shell called spore coat (Kim et al., 2006). A similar protein has also been identified in *Clostridium* sp. (Henriques and Moran, 2007). A CotH containing ORFs was also detected to be associated with a putative cellulase from *N. patriciarum* W5 (Wang et al., 2011). The role of the CotH-like domain in rumen fungi is unknown. It is logical to conjecture that it guides the formation of certain macromolecular protein structures. Our bold guess would be that the fungal CotH-like protein may function similar to the bacterial counterpart, directing the assembly of a subset of the zoospore's spore coat proteins that help to protect the zoospores in the resistant stage, and with a CBM10 modules attached, may also facilitate the zoospores to attach to fiber rich substrates prior to germination and growth. Or rather cautiously, the CotH-like protein may play a role in the assembly of the fungal

cellulosome through interaction with other CAZy proteins, in conjuction with CBM10 modules (Nagy et al., 2007). Thus, CotH like protein may lead to a greater understanding of the cellulosome-like complexes that may be produced by rumen fungi.

Another six of the CBM10 containing ORFs showed similarity to Ser/Thr protein phosphatases. It is commonly known that protein phosphatases are very important regulators of variety of physiological processes, such as cell cycle control, regulation of cell growth and division (Depaoli-Roach et al., 1994; Ingebritsen and Cohen, 1983). However, the role of Ser/Thr protein phosphatase domains associated with potential cellullosome CBM10 like domain will require further detailed biochemical analysis.

Although the exact function of the fungal CBM10 module may currently remain unclear, its high prevalence suggests that it plays an important role. Besides the possibility to bind to some unknown plant cell wall components, a reasonable hypothesis is that it probably ancher a large variety of proteins onto the fungal cell wall, functioning as the fungal counterpart of the bacterial SLH (Fontes and Gilbert, 2010). Detailed binding assay will be needed to determine the components that this module interacts with.

It is unlikely that the association of CBM10 with a variety of different domains could be an artifact of the *de novo* assembly process. As the assembly pipeline used in this chapter was also applied to the muskoxen metagenomic study reported in Chapter 3, and artifact assembly was demonstrated to be minimum, if any. It is also in agreement with other researchers who frequently found the CBM10 module associated with putative GH proteins in the *N. patriciarum* transcriptome (Wang et al., 2011). On the other hand, because of the short length of CBM10 (only ~100 bp), the Illumina Hi-Seq sequencing reads supported the abundant expression of CBM10 domain, and subsequently indicated

its importance in the metabolism, with minimum assembly and regardless of whether and/or which catalytic domains CBM10 was associated to.

The CBM18 domain was the second most abundant CAZy member predicted. From 40 ORFs, a total of 91 CBM18 domains were identified. CBM18 is known as a chitin binding domain, which has been identified exclusively in eukaryotes, including fungi, plants and arthropods (Suetake et al., 2000). The function of the CBM18 within rumen fungi is unknown, but it has been characterized in other organisms. The domain binds to fungal cell wall chitin, which in turn protects the fungi from the environmental chitinases. This may be important for rumen fungi to maintain a stable biomass within the rumen ecosystem, as several rumen microorganisms, including bacteria (Kopecny et al., 1996) and ciliate protozoa (Morgavi et al., 1994) produce chitinases. The expression of CBM18 modules by rumen fungi may serve to block chitinase activity in the rumen, through protective binding to fungal chitin. Multiple CBM18 modules existing in one ORF may enhance the binding affinity to chitin. Three CBM18 modules were associated with polysaccharide deacetylases (CE4). This enzyme removes the acetyl group from chitin and converts it to chitosan. This function may also serve as a protection for the fungi as chitosan is a poor substrate for chitinases (El Gueddari et al., 2002). Two ORFs harbouring the CBM18 modules were associated with one CBM13 module. Since the CBM13 can bind to xylan, these two ORFs may serve as a bridge that also promotes fungal attachment to xylan.

To my knowledge, the study on *N. patriciarum* was the first and only published transcriptomic study targeting an anaerobic fungus (Wang et al., 2011), displaying several similar findings with my research reported here. In the absence of genomic

information, we both combined NGS based transcriptomic and LC-MS/MS based secretomic studies, with a focus on plant cell wall degrading enzymes. The *Neocallimastix* study identified a similar number of total 219 putative GH proteins, including the major induced cellulases from GH families 1, 3, 6, 10, 11, 43 and 48 with most of them possessing a CBM10-like domain. However, unlike the study of Wang et al. (2011), GH9 and GH45 members were also detected to be highly expressed in my study, possibly a reflection of genetic differences between two fungal genera as well as the use of different carbon sources for growth. My study also provided gene expression comparison among several carbon sources, and besides GH modules, CE, PL and CBM modules were also predicted. Unfortunately as no sequence data from the other study is publicly available, no further sequence comparison could be made.

With the continued growth of the human population and the demand for animal products increasing, it is likely that livestock agriculture will have to shift from low fiber/high concentrate diets to a greater reliance on high fiber feedstocks. This could lead to an increased interest in defining the importance of anaerobic fungi in rumen function (Nagpal et al., 2011). The results reported here demonstrated many previously unknown features of the rumen fungus *A. mucronatus*, with many potential candidate genes predicted. Putative swollenins were identified from a rumen fungus for the first time. It may be a good candidate as a feed additive or feed pretreatment agent, aiding in improving the digestibility of poor quality fibrous feeds for livestock production. The high abundance of CBM10 modules in association with various functional domains raise interests for more detailed characterization, and may facilitate to elucidate the nature of rumen fungal cellulosome-like complex. Many CAZyme candidates, such as members

from GH43, CE4 and CE15 may possess potential properties suitable for cellulosic biofuel industry or agricultural application. All the information will open up brand new avenues to illustrate the full potential of the anaerobic fungi in the future.

4.5 Tables and Figures

Table 4.1 Matches of secretomic peptide detected by LC-MS/MS to predicted CAZy ORFs from RNA-Seq results when *Aneromyces mucronatus* YE505 grown on five different carbon sources.

ORF	CAZy Domains	LC-MS/MS peptide number					
		GCS	Avicel	Oat Spelt Xylan	Barley Straw	Alfalfa Hay	Total
AmuTC51	GH1	11	6	21	12	2	52
AmuVC306	CBM26/GH31	6	8	5		28	47
AmuTC352	GH11	12	4	16	8	3	43
AmuTC99	GH43	11	3	11	7	6	38
AmuTC2165	GH43	11	2	15	4	1	33
AmuTC686_seq3	CE4/CBM10	10	10	1		10	31
AmuVC11828	CBM10/GH3	13	6	5		6	30
AmuVC9826	CBM13	1	1	12	12	2	28
AmuTC72	CBM10/CBM29_Blast	5	3	12		6	26
AmuVC10375	GH11	12	7	2		4	25
AmuVC1910	CBM10	16	6	1			23
AmuVC2498	CBM18	12	7	2		1	22
AmuTC8	GH6	6	2	6	3	4	21
AmuTC858	CBM13/CBM10/GH39	8	5	7			20
AmuVC12166	CBM10/GH48	4	2	6	3	3	18
AmuTC149	GH117/CBM6/GH43	14	3				17
AmuTC290	CBM10/GH9	5	6	3		3	17
AmuVC1002	CBM10	6	7	3			16
AmuVC219	CBM10/GH45	5	1	4	3	2	15
AmuVC557	CBM13/CBM10	6	4	1		4	15
AmuVC12040	CBM10	9	3	1		2	15
AmuTC468	CE4	8	4	1		1	14
AmuVC1082	GH117/CBM6/GH43	13	1				14
AmuTC459	CE15/CBM10	2	2	3		6	13
AmuTC1982	GH115	3	1	4	4	1	13
AmuVC10744	CBM10/GH48	2	2	4		4	12
AmuTC585	GH3_C/CBM10/GH6/GH3	6	4	2			12
AmuVC2417	CBM10/GH26	2	4	1	3	2	12
AmuVC12088	CBM13/CBM10/GH39	6	5				11
AmuVC10088	CE1	2	2	6		1	11
AmuTC623	CE4	7	3				10
AmuVC732	CBM10/GH6	2	5	1		2	10

ORF	CAZy Domains	LC-MS/MS peptide number					
		GCS	Avicel	Oat Spelt Xylan	Barley Straw	Alfalfa Hay	Total
AmuTC169	CBM10	7	2	1			10
AmuVC58	CBM10/CBM1	5	2	3			10
AmuVC9724	CBM10/GH9	7	3				10
AmuTC5285	CE10	3	7				10
AmuVC2993	CBM10/GH11	6	3	1			10
AmuVC2262	CE6/CBM10	4	2	2		2	10
AmuVC12041	CE4		9				9
AmuVC1325	CBM10/GH11	4	3	2			9
AmuVC9760	CBM1/GH6	5	1	1		2	9
AmuVC9693	CBM13/CBM10	1	3	1		3	8
AmuTC3045	CE2 or CE3	4	1	3			8
AmuVC12184	CBM10	5	2	1			8
AmuVC1464	Swollenin/CBM10	6	2				8
AmuVC10470	CBM10/GH53	3	5				8
AmuVC9781	CBM26/GH31	3	2	3			8
AmuVC10313	CBM10	5	3				8
AmuVC11854	CE6/CBM10	3	4				7
AmuVC11861	CBM13/CBM10/CBM6/GH43/CBM36_Blast	2	2	3			7
AmuVC309	Swollenin/CBM10		7				7
AmuVC9956	Swollenin/CBM10	4	3				7
AmuTC2143	CBM10	6	1				7
AmuTC3998	CBM1	6	1				7
AmuVC3673	CBM10/CBM6/GH43	4	1	1		1	7
AmuVC10435	GH2/Bgal_small_N/CBM10	2	2	3			7
AmuVC375	CBM18	4	3				7
AmuVC10015	CE1	1	5				6
AmuVC1110	GH117/CBM6/GH43/CBM36_Blast	5	1				6
AmuVC12245	CBM10/GH48	3	3				6
AmuTC5124	CBM10/GH10	4	2				6
AmuVC3418	CBM10	2	2	1		1	6
AmuVC1304	CBM10/GH18	2	2	1		1	6
AmuVC1201	CBM6/GH43	2	3				5
AmuVC172	CE4		5				5
AmuTC2171	GH67	1	4				5
AmuVC11161	CBM13/CBM10		5				5
AmuVC3782	CBM10/GH5	4	1				5
AmuTC2294_seq2	CBM10/CBM6	2	1	2			5
AmuVC2158	CBM10	3	2				5

ORF	CAZy Domains	LC-MS/MS peptide number					
		GCS	Avicel	Oat Spelt Xylan	Barley Straw	Alfalfa Hay	Total
AmuTC832	CBM13/CBM10	2	2				4
AmuTC1644	CE4	1	1	2			4
AmuVC1308	CE4	1	3				4
AmuVC34	Swollenin/CBM10	2	1	1			4
AmuVC8468	CBM10/GH9	3	1				4
AmuVC11847	CBM10/GH8	1	3				4
AmuVC852	CBM10/CBM1		4				4
AmuVC10350	CE1		4				4
AmuVC12163	CBM10/GH5	2	2				4
AmuTC1512	GH3_C/GH3	1	2	1			4
AmuTC1541Seq2	PL1/CBM1	1	1	1		1	4
AmuVC10033	CBM10	2	1				3
AmuVC11850	GH1		3				3
AmuVC2325	CBM10/GH6	1	1	1			3
AmuVC8984	CBM10/CBM1		3				3
AmuTC4153	GH18	2	1				3
AmuVC1073	CBM18		3				3
AmuTC3711	GH5	1	1	1			3
AmuVC537	CBM10	1	2				3
AmuTC290	CBM10		2				2
AmuTC910	GH117/CBM10/GH43		2				2
AmuVC11071	CBM10		2				2
AmuVC2112	CBM10		2				2
AmuVC408	CE1		2				2
AmuVC6446	CBM10/CBM1		2				2
AmuVC923	CBM18/CBM13		2				2
AmuVC10050	CBM13/CBM10/GH10		2				2
AmuVC10153	CBM10		2				2
AmuVC10637	GH18		2				2
AmuVC1718	CBM10/GH10		2				2
AmuVC10672	CE1/CBM10		2				2
AmuVC2278	CBM10/CBM1		2				2
AmuTC2525	GH13/CBM26		2				2
Total peptide matched		**385**	**306**	**192**	**59**	**115**	**1057**

Table 4.2 ORFs containing CBM10-like domain in association with predicted non-CAZy functional domains.

Domain	ORF
CotH, Spore coat assembly protein	AmanuTC10428 AmanuTC20741 AmanuTC2143 AmanuTC4860 AmanuTC87421 AmanuVC10133 AmanuVC10313 AmanuVC1910 AmanuVC3496 AmanuVC2112 AmanuVN2600 AmanuVN13467 AmanuVC2278
Serpin	AmanuTC169 AmanuVC11071 AmanuVC12225 AmanuVC1590 AmanuVC3541 AmanuVC3669 AmanuVC523
Ser/Thr protein phosphatase	AmanuVC808 AmanuVC10249 AmanuVC10681 AmanuVC3357 AmanuVC6446 AmanuVC8984
Swollenin	AmanuTC4 AmanuVC1464 AmanuVC309 AmanuVC34 AmanuVC9956

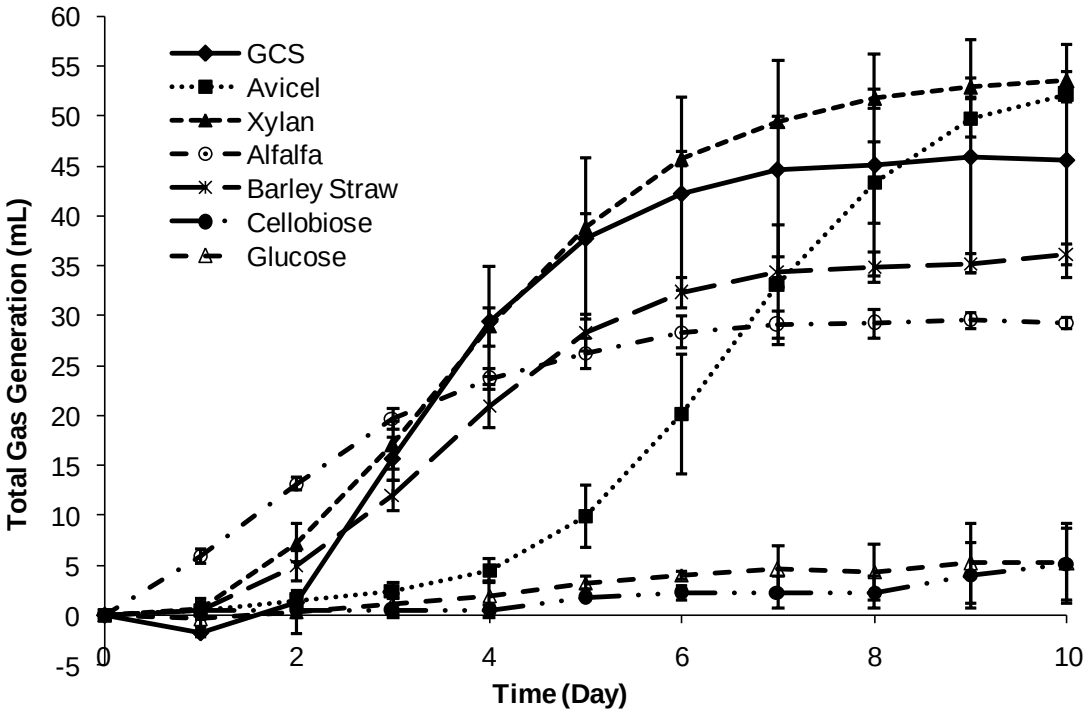

Figure 4.1 The total gas volume generated by *Aneromyces mucronatus* YE505 grown on different carbon sources.

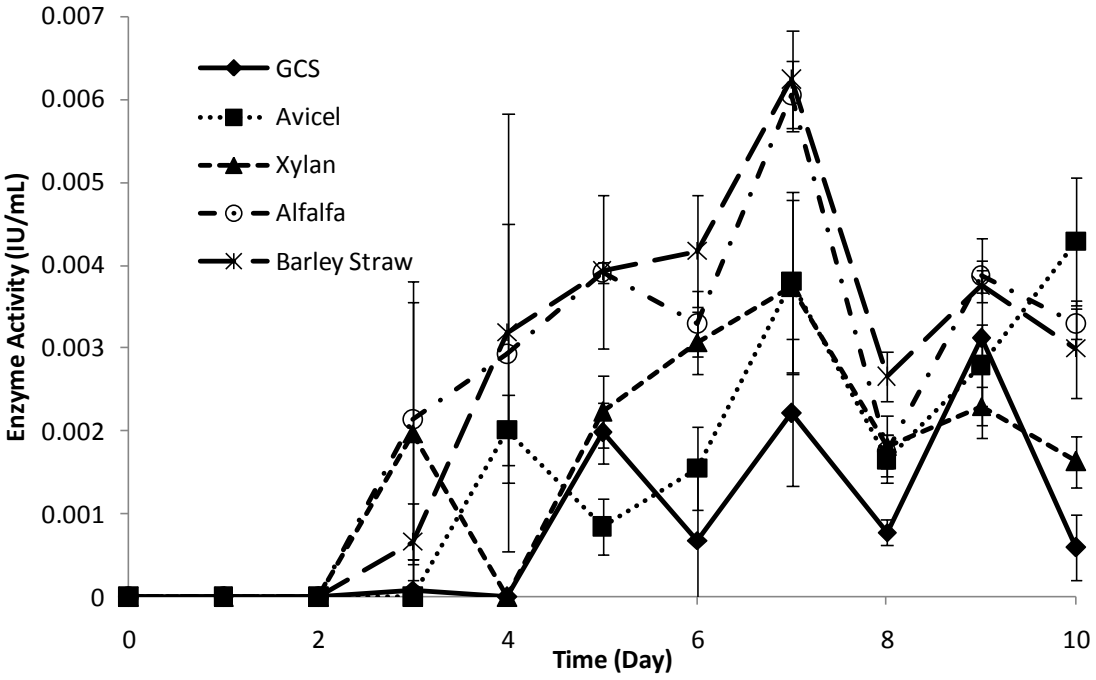

Figure 4.2 The extracellular crude Avicel degrading enzyme activities of *Aneromyces mucronatus* YE505 grown on different carbon sources.

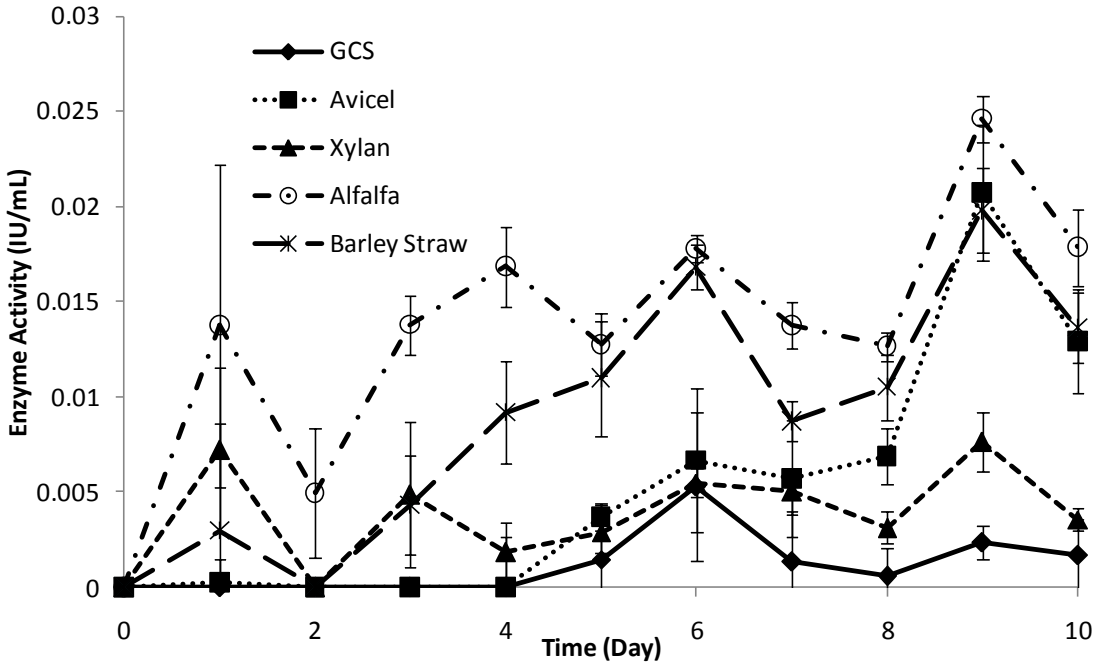

Figure 4.3 The extracellular crude carboxymethylcellulose (CMC) degrading enzyme activities of *Aneromyces mucronatus* YE505 grown on different carbon sources.

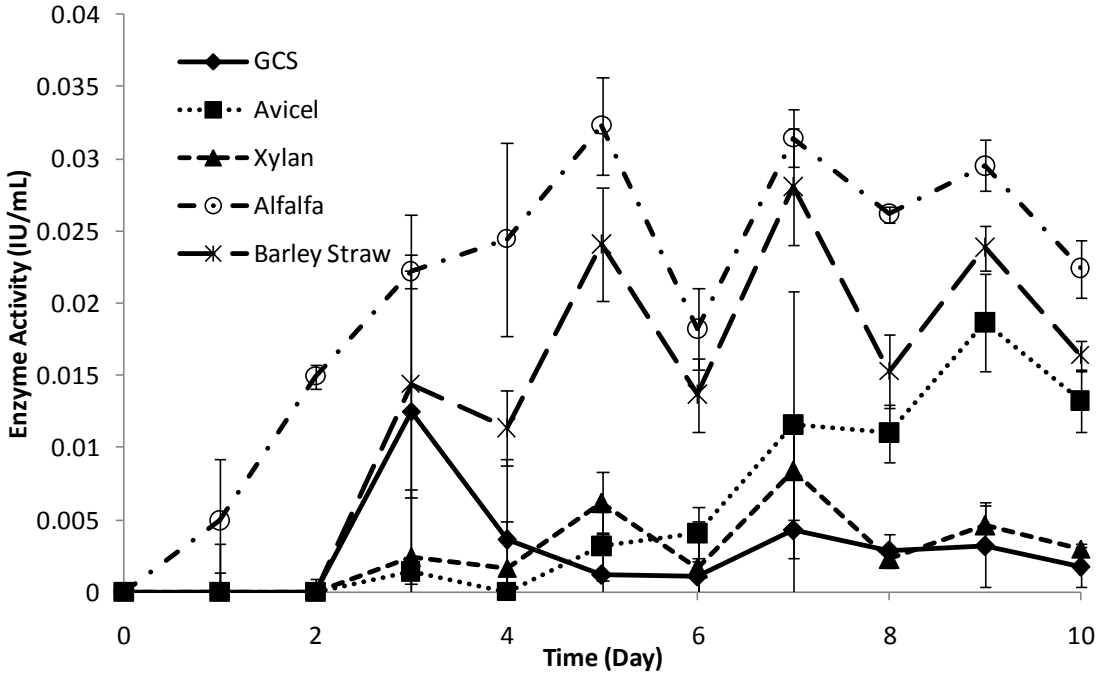

Figure 4.4 The extracellular crude lichenan degrading enzyme activities of *Aneromyces mucronatus* YE505 grown on different carbon sources.

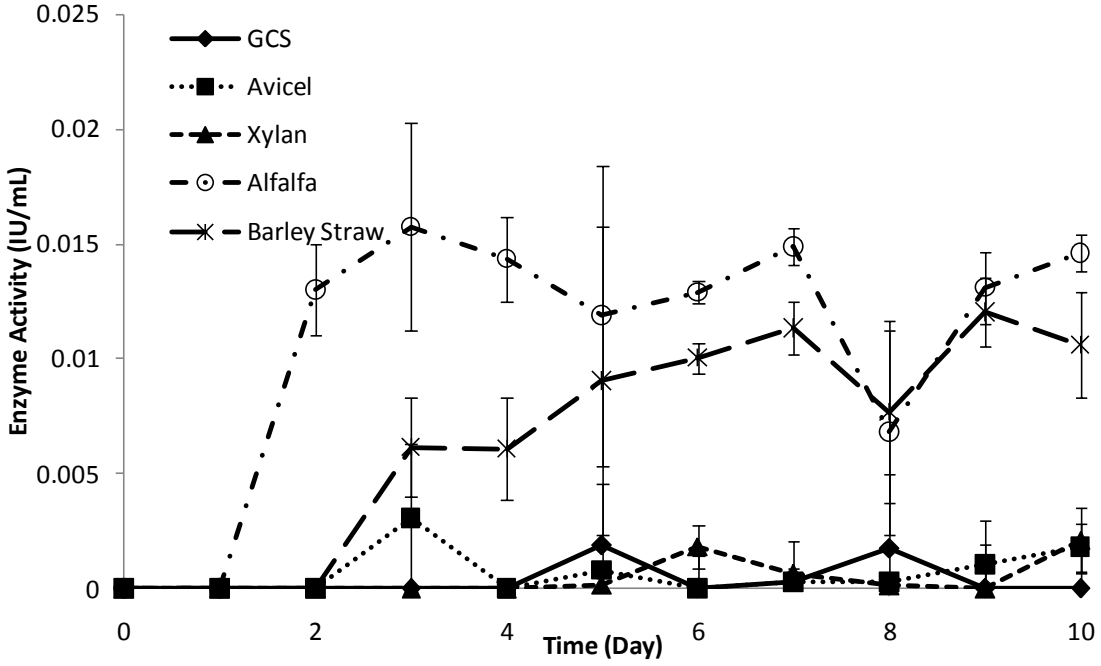

Figure 4.5 The extracellular crude starch degrading enzyme activities of *Aneromyces mucronatus* YE505 grown on different carbon sources.

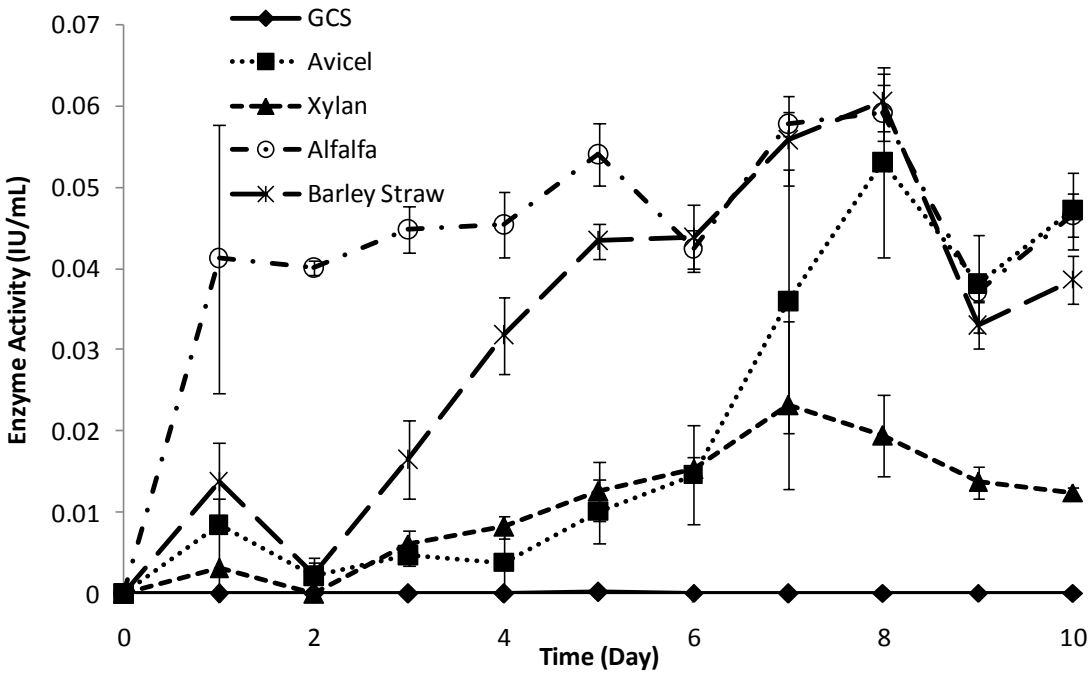

Figure 4.6 The extracellular crude oat spelt xylan degrading enzyme activities of *Aneromyces mucronatus* YE505 grown on different carbon sources.

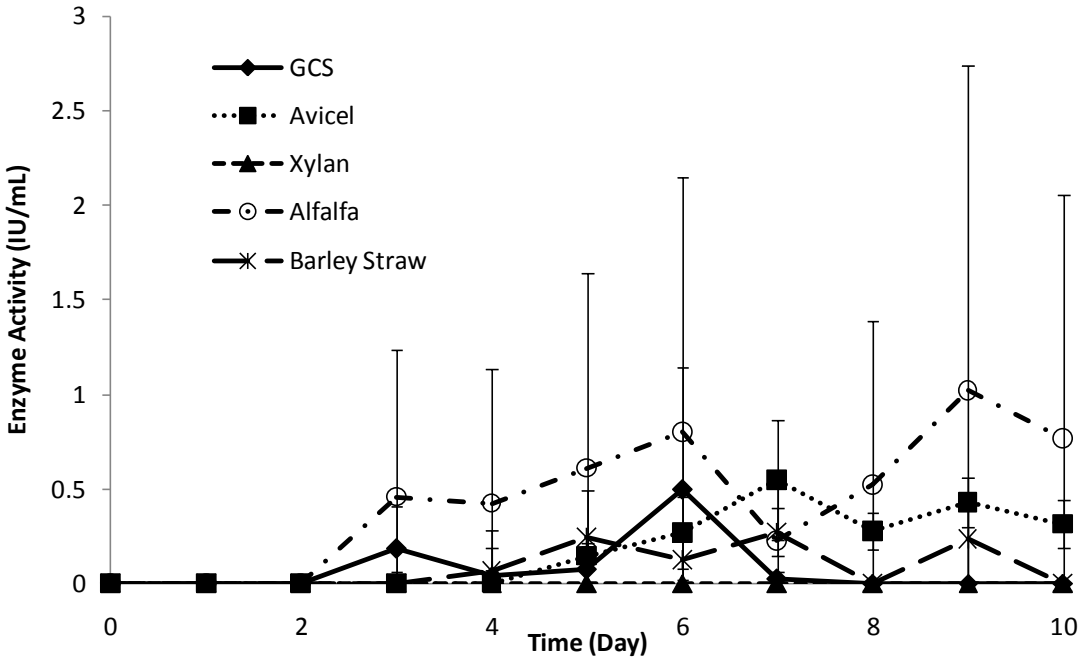

Figure 4.7 The extracellular crude *p*-nitrophenyl-β-d-cellobioside (pNPC) degrading enzyme activities of *Aneromyces mucronatus* YE505 grown on different carbon sources.

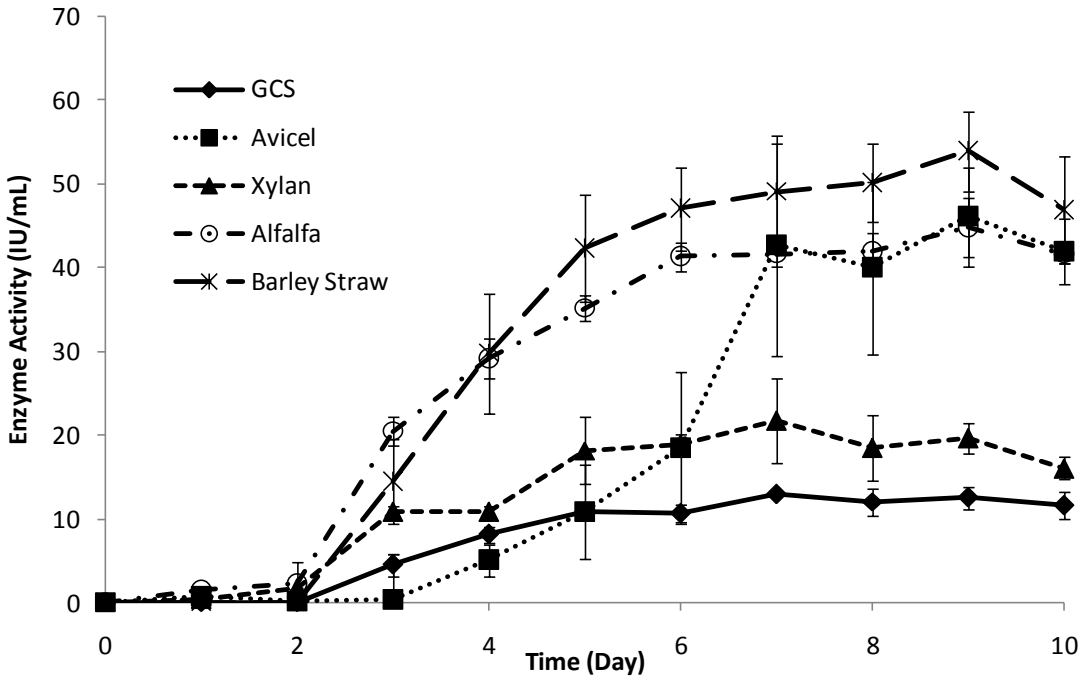

Figure 4.8 The extracellular crude *p*-nitrophenyl-β-d-glucoside (pNPG) degrading enzyme activities of *Aneromyces mucronatus* YE505 grown on different carbon sources.

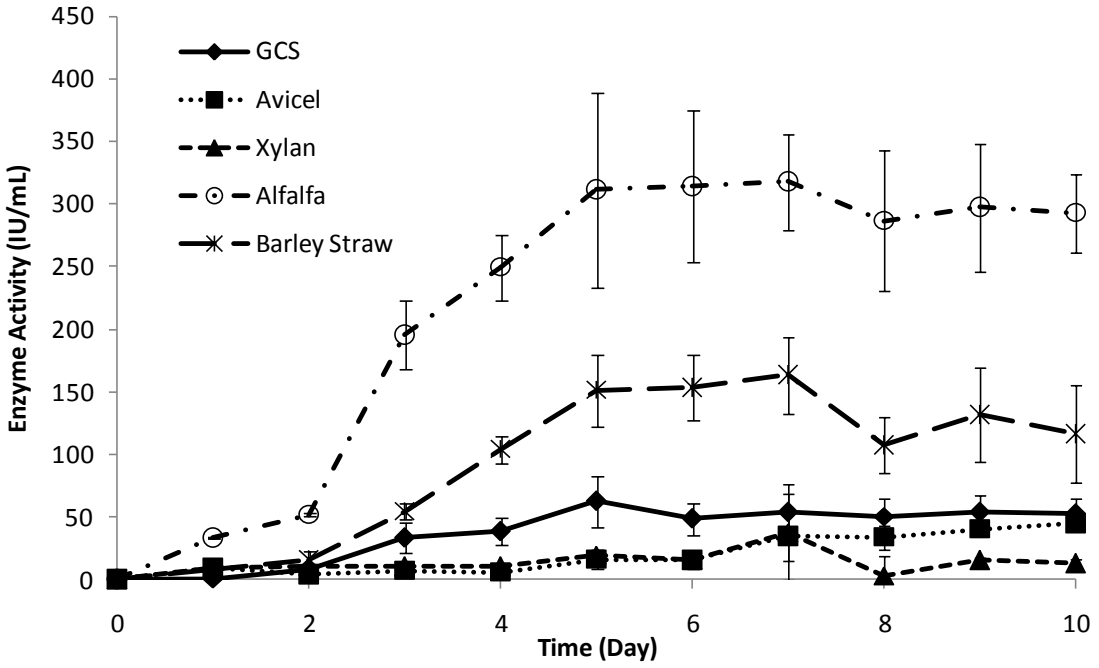

Figure 4.9 The extracellular crude carbohydrate esterase activities toward substrate α-Naphthyl acetate (α-NA) of *Aneromyces mucronatus* YE505 grown on different carbon sources.

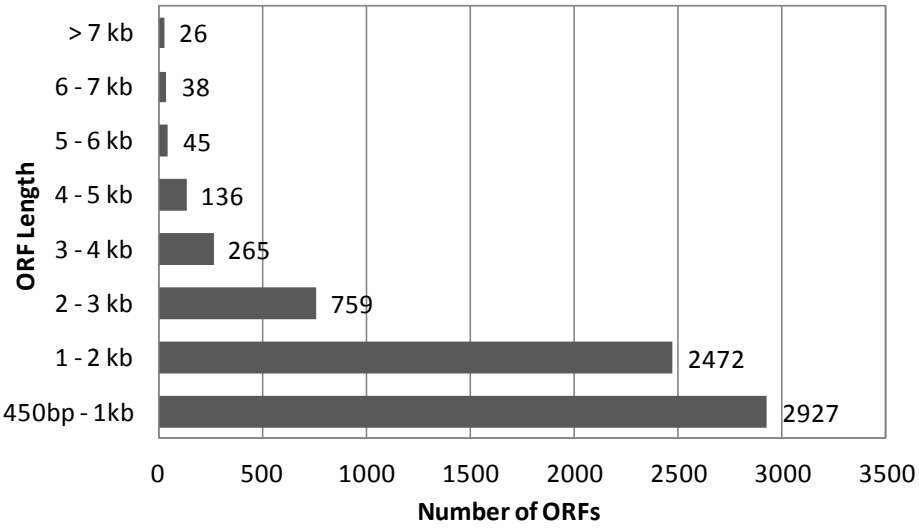

Figure 4.10 Size distribution of ORFs that were identified from *Aneromyces mucronatus* YE505.

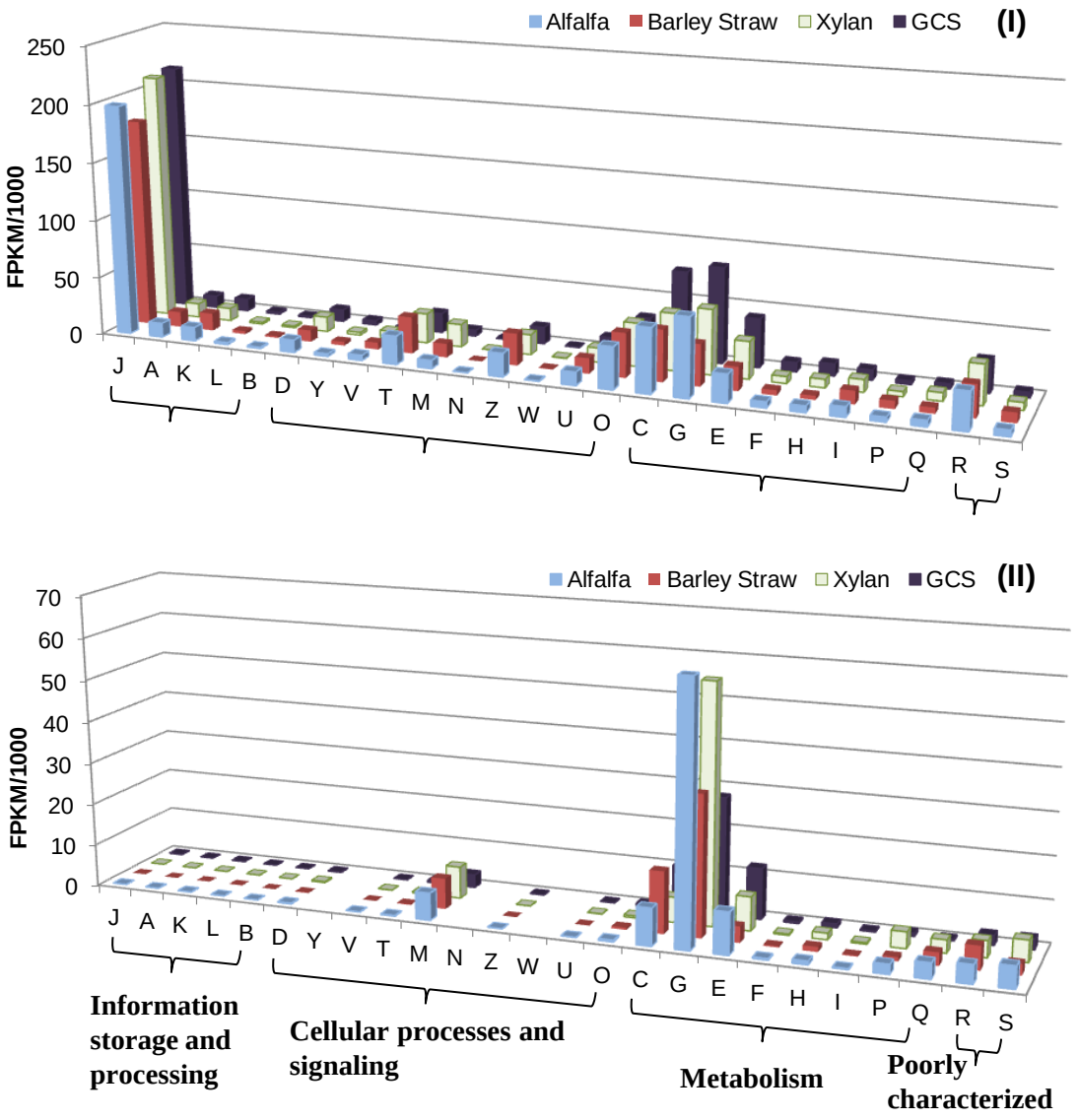

Figure 4.11 ORFs with predicted KOG/COG functions and their total expression in four different carbon sources.

The assigned letters are based on KOG/COG classifications (Tatusov et al., 2003).

(I): ORFs with predicted KOG functions;

(II): ORFs with predicted COG (but without KOG) functions.

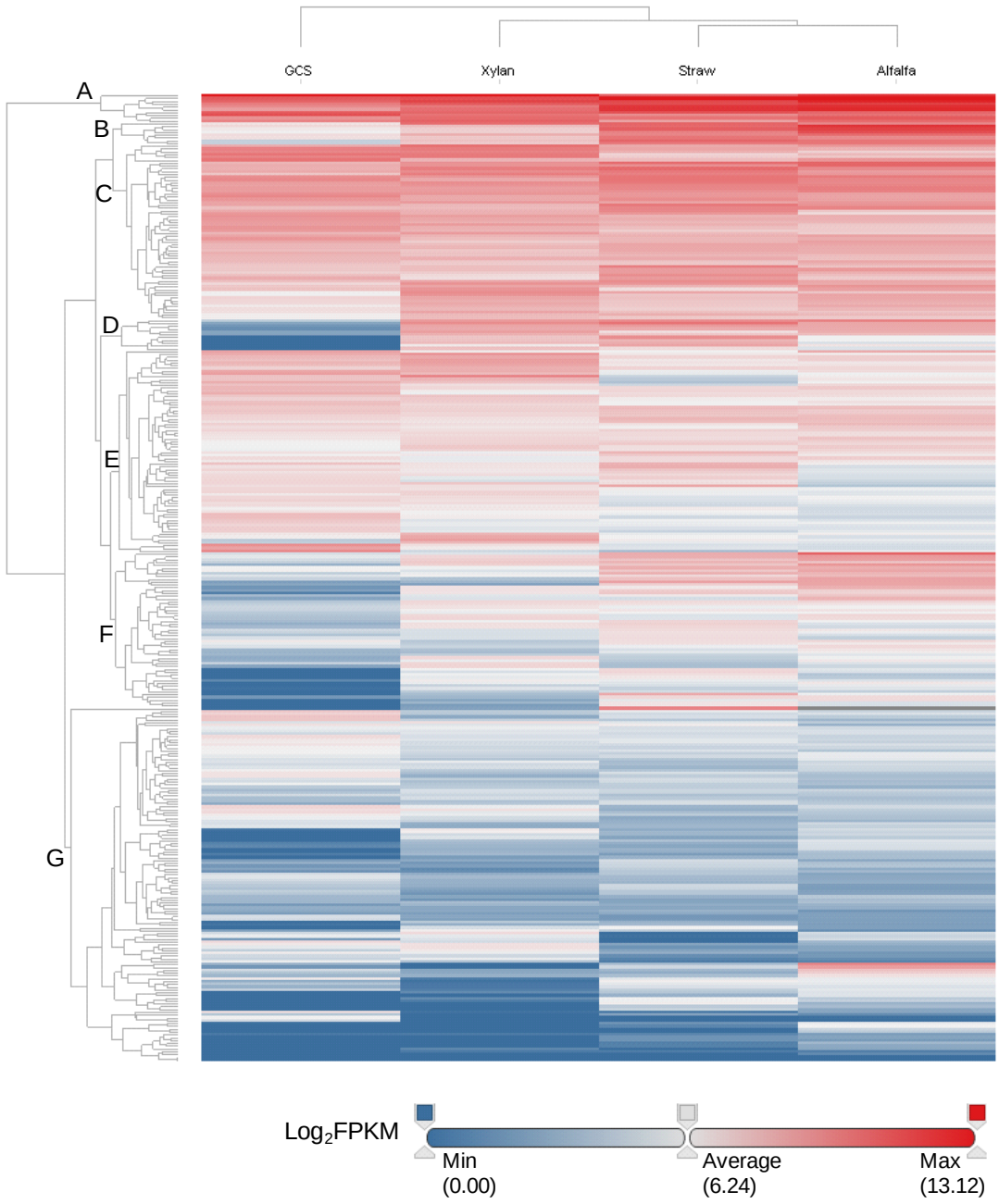

Figure 4.12 Comparison of expression level of predicted CAZymes from *Aneromyces mucronatus* YE505 grown on four different carbon sources.

The CAZymes were grouped into seven clades (A-G) according to their expression patterns and color illustrated based on log_2FPKM value. Refer to Table S.3 for complete list of the predicted CAZymes and their expression.

Chapter 5 Conclusions and future directions

5.1 Accomplishments

This dissertation established an improved RNA isolation procedure to extract excellent quality RNA from particle associated rumen contents, described the current advances of the eukaryotic metatranscriptome of the muskoxen rumen microbiome, as well as the comparative transcriptomic and secretomic analyses of the rumen anaerobic fungus, *Anaeromyces mucronatus* YE505 grown on different substrates.

(Meta)transcriptomics has distinct properties when compared to (meta)genomics. The (Meta)genome of an organism or ecosystem is relatively stable; in contrast (meta)transcriptomes are dynamic and in a continuous state of change with alterations in environmental conditions. While (meta)genomic sequencing identifies the most numerically dominant genes, metatranscriptomic analysis identifies those genes that are most extensively transcribed, and provides more direct and rational evidence for selecting active gene candidates for future studies. On the other hand, genome sequencing can provide useful information on the structure of gene clusters and possible regulatory mechanisms of gene expression, information that is difficult to obtain from mRNA sequencing.

Prior to my research project, several studies focused on the metagenomics of the plant fiber digesting gut microbiota from termite (Warnecke et al., 2007), wallaby (Pope et al., 2010) and the bovine rumen (Brulc et al., 2009; Hess et al., 2011). However no metatranscriptomic study of the mammalian digestive tract was described in the literature. Compared to the well established DNA isolation from gut samples, RNA isolation is

more challenging because it is dynamically changing and readily degraded by the myriad of RNases that are present in microbial dense environments. In order to remove this obstacle to the study of gene expression in the rumen ecosystem, I improved the RNA isolation method for rumen samples, and in particular developed a procedure that was optimal for rumen solid samples as described in Chapter 2. Subsequently, this method made it possible to isolate the RNA required for the metatranscriptome of the muskoxen rumen and the transcriptome of the anerobic fungus *Aneromyces mucronatus.*

In Chapter 3, a metatranscriptomic analysis focusing specifically on feed particle-associated rumen eukaryotic microorganisms was carried out. As particle-associated microorganisms represent the major proportion of total rumen microbes (McAllister et al., 1994), and account for up to 90% of the endoglucanase and xylanase activities in the rumen (Miron et al., 2001), I selected to study the solid phase microbial community as it was likely to yield the most information about the function of rumen microbes. Although rumen cellulolytic bacteria are generally believed to play a major role in ruminal plant cell wall biomass degradation, anaerobic fungi are thought to play a significant role in the degradation of low quality forages in part due to their ability to physically disrupt plant particles through mycelia growth (Nagpal et al., 2009). Furthermore, it was more practical to target expressed eukaryotic genes, since the nature of the polyadenylated mature eukaryotic mRNAs enables effective mRNA enrichment by hybridizing to immobilized oligo(dT) for subsequent sequencing, substantially increasing sensitivity through the removal of the most abundant non-coding RNA as well as bacterial mRNA. The resultant Illumina sequencing dataset was analyzed, with a focus on plant cell wall polysaccharide degrading enzymes. The putative genes were found mainly from rumen

eukaryotes especially anaerobic fungi. A total of over 1,000 CAZy proteins were identified in muskoxen rumen samples with the majority from rumen eukaryotes, including anaerobic fungi, protozoa and possibly transient yeast. Compared to the previous gut metagenomics studies which were based solely on DNA sequencing and the genes present (Brulc et al., 2009; Hess et al., 2011; Pope et al., 2010), my study directly elucidated the actual expressed eukaryotic genes of the rumen sample, and was the first report of the rumen eukaryotic metatranscriptome.

In the research described in Chapter 4, transcriptomic sequencing and secretomic analysis were executed upon the anaerobic fungus *A. mucronatus* YE505, one of the least characterized of the cultured anaerobic fungi. Prior to this study, a number of glycoside hydrolases and carbohydrate esterases have been isolated from rumen fungi. However, due to the limitation of classical methods based on activity, cDNA library screening, or microarray, genes that show low activities on substrates used or low sequence similarity to previously reported genes would have been overlooked. The genomic sequencing of *A. mucronatus* YE505 was attempted on various next generation sequencing platforms, including Illumina Hi-Seq and PacBio genetic analyzer (unpublished results). However, assembly of the derived sequence into a draft genome proved to be extremely difficult owing to the AT-rich nature of the genome and the relatively short reads and high error rates associated with NGS technologies. Here my research demonstrated that under such limitations, the transcriptomic studies served to be a practical approach to circumvent these obstacles to explore the interesting anaerobic fungi for novel and potential useful genes, by targeting transcribed sequences directly and assembling sequencing reads into full-length ORFs. Should improvements in the accuracy of NGS or the predictive power

of bioinformatic techniques occur in the future, the established transcriptomes may serve as a solid blueprint for facilitating future genomic assembly.

Thus similar to the experimental design described in Chapter 3, RNA-Seq was performed using an Illumina sequencing platform for *A. mucronatus* YE505. For the first time, a comprehensive insight to the physiological system of *A. mucronatus* was elucidated. Over 300 putative proteins containing CAZy modules were identified. By comparing transcriptomes from four culture conditions on different carbon sources, the actual gene transcription profiles were obtained, and the potentially important highly expressed enzymes were identified by comparing the FPKM values. A number of putative CAZyme genes were induced when *A. mucronatus* was grown on alfalfa hay, xylan or barley straw, suggesting their important roles in the degradation of respective substrates. According to the FPKM, members from GH43, GH6, GH1, GH48, GH45 and CE4 were the predominant CAZymes in *A. mucronatus*. Secretomic study by applying LC-MS/MS and subsequent analysis complemented and provided more proof of the genes detected from the above transcriptomic studies.

When the CAZy proteins detected from the muskoxen rumen and those from *A. mucronatus* were compared, they shared relatively similar CAZy sets responsible for plant fiber digestion. As expected, a larger CAZy set existed in the rumen sample with over 1,000 members (>500 bp) covering 92 CAZy families, compared to over 300 members in 75 families from *A. mucronatus*. A total of 67 families were shared by both datasets, including most CAZymes involved in cellulose and xylan degradations. Over 80% of CAZy members from each dataset shared sequence similarities with members from the other dataset. Unsurprisingly, there were several differences in the two characterized

datasets. For example, glycoside hydrolases from GH39 and GH74, which potentially encode xylosidase and endoglucanase activities respectively, were not identified in the *A. mucronatus* YE505 transcriptome, suggesting either these sequences were absent from strain YE505 or not sufficiently expressed to be detected. The presence of these hydrolases in the muskoxen metatranscriptome suggests that these GHs may exist in other fungal species or protozoa.

5.2 Future perspectives

Based on the proven effective strategies established in my studies, more experiments are currently underway. Comparative transcriptomic studies of three other rumen fungal species (*Neocallimastix patriciarum* 27, *Orpinomyces joyonii* SG4, *Piromyces rhizinflata* YM600) from the Agriculture and Agri-Food Canada anaerobic culture are currently being performed in a manner similar to that described in Chapter 4. Total RNA was isolated from the fungal strains grown on the same carbon sources as *A. mucronatus* YE505, and RNA-Seq sequencing is currently underway. As the sequence assembly and analysis pipeline have already been established for *A. mucronatus* YE505, we anticipate that this will expedite the analysis of these additional genera. The secreted proteins were collected in a similar way as described in Chapter 4, and will be subject to LC-MS/MS. Our primary study has already shown that transcriptome analysis enables full length cDNA assembly and the secretome can facilitate the identification of regions coding for proteins of interest.

However, bioinformatics alone is not able to verify the actual function of predicted proteins and classical biochemical based studies and protein structural studies will be necessary to assign functions to these predicted proteins. The work in this thesis has

identified several candidate proteins that are worthy of further study. Anaerobic fungi are believed to produce cellulosome-like (see Sections 1.4.3 and 4.4) structures, but to date the foundational scaffolding protein required for these structures has not been identified. It has been proposed that the anaerobic fungal CBM10 domain may act as a dockerin module that facilitates the binding of catalytic proteins to the cellulosome-like complex (Nagy et al., 2007; Raghothama et al., 2001; Steenbakkers et al., 2001). In the present study, many CBM10-like modules were detected, highly transcribed and associated with many sequences coding for protein modules from *A. mucronatus*, as well as the fiber degrading rumen eukaryotic consortia from muskoxen (see Chapters 3 and 4). On the other hand, all the CBM10 containing enzymes previously identified were involved in plant cell wall degradation, but there were exceptions to this pattern in my study. The CBM10 like module was found to be associated with four non-CAZy domains including swollenin, CotH, Serpin and phosphatase. This raised the possibility that the CBM10 domain may have a function other than those directly related to fiber degrdation, and could provide integral information towards defining the function of the rumen fungal cellulosome.

Wild type and truncated mutations from a GH43 gene associated with two tandem CBM10 modules from *A. mucronatus* YE505 have been introduced into expression vectors. However, no activity has been detected when these recombinant proteins were expressed in *Escherichia coli* (unpublished data). This may reflect that the *E. coli* expression system is not suitable to express these genes from anaerobic fungi as a result of different codon usage (unpublished data) or possibly the lack of appropriate post translational modification such as glycosylation. In the future, satisfactory expression

may be obtained using other expression systems such as *Saccharomyces cerevisiae*, *Pichia pastoris* or *Aspergillus niger*. By doing so, studies will be performed to investigate the function of CBM10, for example, whether it can bind certain types of polysaccharides, or whether the existence of CBM10 influences enzyme activities.

Swollenin is another module worthy of further study. This project identified mRNA coding for swollenin both within the muskoxen rumen and from pure cultures of rumen fungus (Chapters 3 and 4). This module was reported to destabilize the cellulose structure with no hydrolytic activity (Brotman et al., 2008). Most swollenin proteins detected in this study were associated with CBM10 modules, again raising interest in the function of this domain.

Rumen fungi are a novel group of microorganisms and their importance in fiber digestion within the rumen community has probably been long underestimated. In recent years, the desire to create a cellulosic biofuel industry has increased the demand for novel lignocellulolysic enzymes. The new trend to use more high fiber diets for agricultural livestock production is likely on the horizon, a development that will make it even more imperative that a better understanding of the process of ruminal fiber digestion be established. My studies described in this dissertation, and a recent study (Wang et al., 2011) have demonstrated the unique CAZyme gene pool harboured and actively utilized by some anaerobic fungi. Enzymatic functional analysis will further target the potential candidates for industrial or agricultural usage. Hypothetical xylanases from GH43, carbohydrate esterases from CE4 and CE15 are good candidates to start with, as there are few studies on these putative enzymes and these enzymes were found to be expressed at a high level when *A. mucronatus* was grown on alfalfa hay and barley straw (Chapter 4).

My thesis and the combined rumen microbiome and genome sequencing information across studies (Brulc et al., 2009; Dai et al., 2012; Hess et al., 2011; Pope et al., 2010; Pope et al., 2012) suggest that the rumen microbiota seems to lack cellulases from GH7 and GH12 families. So far, all members of GH7, a family of exoglucanases, have been isolated from aerobic fungi. Enzymes in family GH12 have been shown to have endoglucanase and xyloglucan hydrolase activities. Although functional aspects of these enzymes may be complemented by other enzymes in the rumen, it would be worth investigating if the addition of the fungal enzymes from GH7 and GH12 families results in significant improvements in rumen fiber digestion.

5.3 Conclusions

Although lots of effort has been made, many aspects of the complex rumen system remain in a black box, with many of the microbial species present and their interactions remaining undefined. My studies provided an overview of gene expression information pertaining to the active eukaryotic lignocellulolytic degradation system existing in rumen fungus *A. mucronatus* and the rumen of muskoxen, and elucidated the potential power of these poorly characterized rumen eukaryotic microorganisms. My dissertation from the perspective of transcriptomic and metatranscriptomic studies shed light on a corner of the black box, demonstrating that sequencing results obtained from high-throughput RNA-Seq and *de novo* assembly were able to provide excellent comprehensive overview of the metabolic activities of the rumen eukaryotic population, as well as rumen fungi in pure culture. By applying various bioinformatic tools, a unique set of hypothetical carbohydrate active enzymes and binding modules were identified. This provided a

powerful source for discovering enzymes that may have significance to both agricultural and biofuel industries.

References

Akhmanova, A., Voncken, F.G., Hosea, K.M., Harhangi, H., Keltjens, J.T., op den Camp, H.J., Vogels, G.D., and Hackstein, J.H. (1999). A hydrogenosome with pyruvate formate-lyase: anaerobic chytrid fungi use an alternative route for pyruvate catabolism. Mol. Microbiol. **32**(5): 1103-1114.

Akin, D.E., and Borneman, W.S. (1990). Role of rumen fungi in fiber degradation. J. Dairy Sci. **73**(10): 3023-3032.

Ali, B.R.S., Zhou, L., Graves, F.M., Freedman, R.B., Black, G.W., Gilbert, H.J., and Hazlewood, G.P. (1995). Cellulases and hemicellulases of the anaerobic fungus *Piromyces* constitute a multiprotein cellulose-binding complex and are encoded by multigene families. FEMS Microbiol. Lett. **125**(1): 15-21.

Allgaier, M., Reddy, A., Park, J.I., Ivanova, N., D'Haeseleer, P., Lowry, S., Sapra, R., Hazen, T.C., Simmons, B.A., VanderGheynst, J.S., and Hugenholtz, P. (2010). Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. PLoS One **5**(1): e8812.

Barboza, P.S., Peltier, T.C., and Forster, R.J. (2006). Ruminal fermentation and fill change with season in an arctic grazer: responses to hyperphagia and hypophagia in muskoxen (*Ovibos moschatus*). Physiol. Biochem. Zool. **79**(3): 497-513.

Bayer, E.A., Lamed, R., White, B.A., and Flint, H.J. (2008). From cellulosomes to cellulosomics. Chemical record (New York, N.Y.) **8**(6): 364-377.

Béra-Maillet, C., Devillard, E., Cezette, M., Jouany, J.P., and Forano, E. (2005). Xylanases and carboxymethylcellulases of the rumen protozoa *Polyplastron multivesiculatum*, *Eudiplodinium maggii* and *Entodinium* sp. FEMS Microbiol. Lett. **244**(1): 149-156.

Béra-Maillet, C., Mosoni, P., Kwasiborski, A., Suau, F., Ribot, Y., and Forano, E. (2009). Development of a RT-qPCR method for the quantification of *Fibrobacter succinogenes* S85 glycoside hydrolase transcripts in the rumen content of gnotobiotic and conventional sheep. J. Microbiol. Methods **77**(1): 8-16.

Béra-Maillet, C., Ribot, Y., and Forano, E. (2004). Fiber-degrading systems of different strains of the genus *Fibrobacter*. Appl. Environ. Microbiol. **70**(4): 2172-2179.

Berg Miller, M.E., Antonopoulos, D.A., Rincon, M.T., Band, M., Bari, A., Akraiko, T., Hernandez, A., Thimmapuram, J., Henrissat, B., Coutinho, P.M., Borovok, I., Jindou, S., Lamed, R., Flint, H.J., Bayer, E.A., and White, B.A. (2009). Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. PLoS One **4**(8): e6650.

Biely, P. (2012). Microbial carbohydrate esterases deacetylating plant polysaccharides. Biotechnol. Adv. **30**(6): 1575-1588.

Black, G.W., Hazlewood, G.P., Xue, G.P., Orpin, C.G., and Gilbert, H.J. (1994). Xylanase B from *Neocallimastix patriciarum* contains a non-catalytic 455-residue linker sequence comprised of 57 repeats of an octapeptide. Biochem. J. **299 (Pt 2)**: 381-387.

Blum, D.L., Li, X.L., Chen, H., and Ljungdahl, L.G. (1999). Characterization of an acetyl xylan esterase from the anaerobic fungus *Orpinomyces* sp. strain PC-2. Appl. Environ. Microbiol. **65**(9): 3990-3995.

Borneman, S., and Akin, D. (1994). The nature of anaerobic fungi and their polysaccharide degrading enzymes. Mycoscience **35**(2): 199-211.

Boxma, B., de Graaf, R.M., van der Staay, G.W., van Alen, T.A., Ricard, G., Gabaldon, T., van Hoek, A.H., Moon-van der Staay, S.Y., Koopman, W.J., van Hellemond, J.J., Tielens, A.G., Friedrich, T., Veenhuis, M., Huynen, M.A., and Hackstein, J.H. (2005). An anaerobic mitochondrion that produces hydrogen. Nature **434**(7029): 74-79.

Boxma, B., Voncken, F., Jannink, S., van Alen, T., Akhmanova, A., van Weelden, S.W., van Hellemond, J.J., Ricard, G., Huynen, M., Tielens, A.G., and Hackstein, J.H. (2004). The anaerobic chytridiomycete fungus *Piromyces* sp. E2 produces ethanol via pyruvate:formate lyase and an alcohol dehydrogenase E. Mol. Microbiol. **51**(5): 1389-1399.

Breton, A., Bernalier, A., Dusser, M., Fonty, G., Gaillard-Martinie, B., and Guillot, J. (1990). *Anaeromyces mucronatus* nov. gen., nov. sp. a new strictly anaerobic rumen fungus with polycentric thallus. FEMS Microbiol. Lett. **58**(2): 177-182.

Breton, A., Dusser, M., Gaillard-Martinie, B., Guillot, J., Millet, L., and Prensier, G. (1991). *Piromyces rhizinflata* nov. sp., a strictly anaerobic fungus from faeces of the Saharian ass: a morphological, metabolic and ultrastructural study. FEMS Microbiol. Lett. **66**(1): 1-8.

Brooks, M.J., Rajasimha, H.K., Roger, J.E., and Swaroop, A. (2011). Next-generation sequencing facilitates quantitative analysis of wild-type and Nrl(-/-) retinal transcriptomes. Mol. Vis. **17**: 3034-3054.

Brotman, Y., Briff, E., Viterbo, A., and Chet, I. (2008). Role of swollenin, an expansin-like protein from *Trichoderma*, in plant root colonization. Plant Physiol **147**(2): 779-789.

Brownlee, A.G. (1989). Remarkably AT-rich genomic DNA from the anaerobic fungus *Neocallimastix*. Nucleic Acids Res. **17**(4): 1327-1335.

Brulc, J.M., Antonopoulos, D.A., Miller, M.E., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A., Edwards, R.E., Frank, E.D., Emerson, J.B., Wacklin, P., Coutinho, P.M., Henrissat, B., Nelson, K.E., and White, B.A. (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. Proc. Natl. Acad. Sci. U. S. A. **106**(6): 1948-1953.

Cai, S., Li, J., Hu, F.Z., Zhang, K., Luo, Y., Janto, B., Boissy, R., Ehrlich, G., and Dong, X. (2010). *Cellulosilyticum ruminicola*, a newly described rumen bacterium that possesses redundant fibrolytic-protein-encoding genes and degrades lignocellulose with multiple carbohydrate- borne fibrolytic enzymes. Appl. Environ. Microbiol. **76**(12): 3818-3824.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. **37**(Database issue): D233-238.

CCAC. (1993). Canadian Council on Animal Care. Guide to the care and use of experimental animals. Ottawa, ON, Canada.

Chen, H., Hopper, S.L., Li, X.L., Ljungdahl, L.G., and Cerniglia, C.E. (2006). Isolation of extremely AT-rich genomic DNA and analysis of genes encoding carbohydrate-degrading enzymes from *Orpinomyces* sp. strain PC-2. Curr. Microbiol. **53**(5): 396-400.

Chen, J., Stevenson, D.M., and Weimer, P.J. (2004). Albusin B, a bacteriocin from the ruminal bacterium *Ruminococcus albus* 7 that inhibits growth of *Ruminococcus flavefaciens*. Appl. Environ. Microbiol. **70**(5): 3167-3170.

Chen, X.A., Ishida, N., Todaka, N., Nakamura, R., Maruyama, J., Takahashi, H., and Kitamoto, K. (2010). Promotion of efficient Saccharification of crystalline cellulose by *Aspergillus fumigatus* Swo1. Appl. Environ. Microbiol. **76**(8): 2556-2561.

Cheng, K.-J., and McAllister, T.A. (1997). Compartmentation in the rumen. *In* The Rumen microbial Ecosystem. *Edited by* P. N. Hobson, and C. S. Stewart. Blackie Academic and Professional. pp. 492-522.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**(1): 156-159.

Christov, L.P., and Prior, B.A. (1993). Esterases of xylan-degrading microorganisms: production, properties, and significance. Enzyme. Microb. Technol. **15**(6): 460-475.

Collins, T., Gerday, C., and Feller, G. (2005). Xylanases, xylanase families and extremophilic xylanases. FEMS Microbiol. Rev. **29**(1): 3-23.

Conway, T., and Schoolnik, G.K. (2003). Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. Mol. Microbiol. **47**(4): 879-889.

Costa, V., Angelini, C., De Feis, I., and Ciccodicola, A. (2010). Uncovering the complexity of transcriptomes with RNA-Seq. J. Biomed. Biotechnol. **2010**: 853916.

Costerton, J.W. (2007). The biofilm primer. First ed. Springer.

Costerton, J.W., Cheng, K.J., Geesey, G.G., Ladd, T.I., Nickel, J.C., Dasgupta, M., and Marrie, T.J. (1987). Bacterial biofilms in nature and disease. Annu. Rev. Microbiol. **41**: 435-464.

Crater, A.R., Barboza, P.S., and Forster, R.J. (2007). Regulation of rumen fermentation during seasonal fluctuations in food intake of muskoxen. Comp Biochem Physiol A Mol Integr Physiol **146**(2): 233-241.

Dai, X., Zhu, Y., Luo, Y., Song, L., Liu, D., Liu, L., Chen, F., Wang, M., Li, J., Zeng, X., Dong, Z., Hu, S., Li, L., Xu, J., Huang, L., and Dong, X. (2012). Metagenomic insights into the fibrolytic microbiome in yak rumen. PLoS One **7**(7): e40430.

Dashtban, M., Schraft, H., and Qin, W. (2009). Fungal bioconversion of lignocellulosic residues; opportunities & perspectives. Int J Biol Sci **5**(6): 578-595.

Datta, S., Kim, S., Chakraborty, S., and Gill, R.S. (2010). Statistical analyses of next generation sequence data: a partial overview. J Proteomics Bioinform **3**(6): 183-190.

Depaoli-Roach, A.A., Park, I.K., Cerovsky, V., Csortos, C., Durbin, S.D., Kuntz, M.J., Sitikov, A., Tang, P.M., Verin, A., and Zolnierowicz, S. (1994). Serine/threonine

protein phosphatases in the control of cell function. Adv. Enzyme Regul. **34**: 199-224.

Devillard, E., Newbold, C.J., Scott, K.P., Forano, E., Wallace, R.J., Jouany, J.P., and Flint, H.J. (1999). A xylanase produced by the rumen anaerobic protozoan *Polyplastron multivesiculatum* shows close sequence similarity to family 11 xylanases from gram-positive bacteria. FEMS Microbiol. Lett. **181**(1): 145-152.

Eddy, S.R. (2009). A new generation of homology search tools based on probabilistic inference. Genome Inform **23**(1): 205-211.

Edwards, J.E., Huws, S.A., Kim, E.J., and Kingston-Smith, A.H. (2007). Characterization of the dynamics of initial bacterial colonization of nonconserved forage in the bovine rumen. FEMS Microbiol. Ecol. **62**(3): 323-335.

Edwards, J.E., Huws, S.A., Kim, E.J., Lee, M.R.F., Kingston-Smith, A.H., and Scollan, N.D. (2008a). Advances in microbial ecosystem concepts and their consequences for ruminant agriculture. Animal **2**(05): 653-660.

Edwards, J.E., Kingston-Smith, A.H., Jimenez, H.R., Huws, S.A., Skot, K.P., Griffith, G.W., McEwan, N.R., and Theodorou, M.K. (2008b). Dynamics of initial colonization of nonconserved perennial ryegrass by anaerobic fungi in the bovine rumen. FEMS Microbiol. Ecol. **66**(3): 537-545.

El Gueddari, N.E., Rauchhaus, U., Moerschbacher, B.M., and Deising, H.B. (2002). Developmentally regulated conversion of surface-exposed chitin to chitosan in cell walls of plant pathogenic fungi. New Phytol. **156**(1): 103-112.

Fanutti, C., Ponyi, T., Black, G.W., Hazlewood, G.P., and Gilbert, H.J. (1995). The Conserved noncatalytic 40-residue sequence in cellulases and hemicellulases from anaerobic fungi functions as a protein docking domain. J. Biol. Chem. **270**(49): 29314-29322.

Fendri, I., Tardif, C., Fierobe, H.P., Lignon, S., Valette, O., Pages, S., and Perret, S. (2009). The cellulosomes from *Clostridium cellulolyticum*: identification of new components and synergies between complexes. FEBS J. **276**(11): 3076-3086.

Fields, M.W., Mallik, S., and Russell, J.B. (2000). *Fibrobacter succinogenes* S85 ferments ball-milled cellulose as fast as cellobiose until cellulose surface area is limiting. Appl. Microbiol. Biotechnol. **54**(4): 570-574.

Firkins, J.L., Yu, Z., and Morrison, M. (2007). Ruminal nitrogen metabolism: perspectives for integration of microbiology and nutrition for dairy. J. Dairy Sci. **90 Suppl 1**: E1-16.

Flemming, H.C., and Wingender, J. (2010). The biofilm matrix. Nat. Rev. Microbiol. **8**(9): 623-633.

Flint, H.J. (1997). The rumen microbial ecosystem--some recent developments. Trends Microbiol. **5**(12): 483-488.

Flint, H.J., Bayer, E.A., Rincon, M.T., Lamed, R., and White, B.A. (2008). Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nature reviews. Microbiology **6**(2): 121-131.

Fontes, C.M., and Gilbert, H.J. (2010). Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu. Rev. Biochem. **79**: 655-681.

Freelove, A.C., Bolam, D.N., White, P., Hazlewood, G.P., and Gilbert, H.J. (2001). A novel carbohydrate-binding protein is a component of the plant cell wall-degrading complex of *Piromyces equi*. J. Biol. Chem. **276**(46): 43010-43017.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., and Delong, E.F. (2008). Microbial community gene expression in ocean surface waters. Proc. Natl. Acad. Sci. U. S. A. **105**(10): 3805-3810.

Garcia-Vallve, S., Romeu, A., and Palau, J. (2000). Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. Mol. Biol. Evol. **17**(3): 352-361.

Gifford, S.M., Sharma, S., Rinta-Kanto, J.M., and Moran, M.A. (2011). Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. ISME J **5**(3): 461-472.

Gilbert, H.J., Hazlewood, G.P., Laurie, J.I., Orpin, C.G., and Xue, G.P. (1992). Homologous catalytic domains in a rumen fungal xylanase: evidence for gene duplication and prokaryotic origin. Mol. Microbiol. **6**(15): 2065-2072.

Gilbert, J.A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P., and Joint, I. (2008). Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS One **3**(8): e3042.

Gilbert, J.A., and Hughes, M. (2011). Gene expression profiling: metatranscriptomics. Methods Mol. Biol. **733**: 195-205.

Gloster, T.M., Ibatullin, F.M., Macauley, K., Eklof, J.M., Roberts, S., Turkenburg, J.P., Bjornvad, M.E., Jorgensen, P.L., Danielsen, S., Johansen, K.S., Borchert, T.V., Wilson, K.S., Brumer, H., and Davies, G.J. (2007). Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12. J. Biol. Chem. **282**(26): 19177-19189.

Gordon, G.L., and Phillips, M.W. (1998). The role of anaerobic gut fungi in ruminants. Nutr. Res. Rev. **11**(1): 133-168.

Gosalbes, M.J., Durban, A., Pignatelli, M., Abellan, J.J., Jimenez-Hernandez, N., Perez-Cobas, A.E., Latorre, A., and Moya, A. (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. PLoS One **6**(3): e17447.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. **29**(7): 644-652.

Harhangi, H.R., Steenbakkers, P.J., Akhmanova, A., Jetten, M.S., van der Drift, C., and Op den Camp, H.J. (2002). A highly expressed family 1 beta-glucosidase with transglycosylation capacity from the anaerobic fungus *Piromyces* sp. E2. Biochim. Biophys. Acta **1574**(3): 293-303.

Hausner, G., Inglis, G.D., Yanke, L.J., Kawchuk, L.M., and McAllister, T.A. (2000). Analysis of restriction fragment length polymorphisms in the ribosomal DNA of a selection of anaerobic chytrids. Can. J. Bot. **78**(7): 917-927.

Henriques, A.O., and Moran, C.P., Jr. (2007). Structure, assembly, and function of the spore surface layers. Annu. Rev. Microbiol. **61**: 555-588.

Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L., and Mornon, J.P. (1989). Cellulase families revealed by hydrophobic cluster analysis. Gene **81**(1): 83-95.

Herrera, P., Kwon, Y.M., and Ricke, S.C. (2009). Ecology and pathogenicity of gastrointestinal *Streptococcus bovis*. Anaerobe **15**(1-2): 44-54.

Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S., Clark, D.S., Chen, F., Zhang, T., Mackie, R.I., Pennacchio, L.A., Tringe, S.G., Visel, A., Woyke, T., Wang, Z., and Rubin, E.M. (2011). Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. Science **331**(6016): 463-467.

Hibbett, D.S., Binder, M., Bischoff, J.F., Blackwell, M., Cannon, P.F., Eriksson, O.E., Huhndorf, S., James, T., Kirk, P.M., Lucking, R., Thorsten Lumbsch, H., Lutzoni, F., Matheny, P.B., McLaughlin, D.J., Powell, M.J., Redhead, S., Schoch, C.L., Spatafora, J.W., Stalpers, J.A., Vilgalys, R., Aime, M.C., Aptroot, A., Bauer, R., Begerow, D., Benny, G.L., Castlebury, L.A., Crous, P.W., Dai, Y.C., Gams, W., Geiser, D.M., Griffith, G.W., Gueidan, C., Hawksworth, D.L., Hestmark, G., Hosaka, K., Humber, R.A., Hyde, K.D., Ironside, J.E., Koljalg, U., Kurtzman, C.P., Larsson, K.H., Lichtwardt, R., Longcore, J., Miadlikowska, J., Miller, A., Moncalvo, J.M., Mozley-Standridge, S., Oberwinkler, F., Parmasto, E., Reeb, V., Rogers, J.D., Roux, C., Ryvarden, L., Sampaio, J.P., Schussler, A., Sugiyama, J., Thorn, R.G., Tibell, L., Untereiner, W.A., Walker, C., Wang, Z., Weir, A., Weiss, M., White, M.M., Winka, K., Yao, Y.J., and Zhang, N. (2007). A higher-level phylogenetic classification of the *Fungi*. Mycol. Res. **111**(Pt 5): 509-547.

Ho, Y.W., Barr, D.J.S., Abdullah, N., and Jalaludin, S. (1993). *Anaeromyces*, an earlier name for *Ruminomyces*. Mycotaxon **47**: 283-284.

Hobson, P.N. (1997). Introduction. *In* The rumen microbial ecosystem. *Edited by* P. N. Hobson, and C. S. Stewart. Blackie Academic and Professional. pp. 1-9.

Huang, X., and Madan, A. (1999). CAP3: A DNA sequence assembly program. Genome Res. **9**(9): 868-877.

Huang, Y., Gilna, P., and Li, W. (2009). Identification of ribosomal RNA genes in metagenomic fragments. Bioinformatics **25**(10): 1338-1340.

Hungate, R.E. (1947). Studies on cellulose fermentation: III. the culture and isolation for cellulose-decomposing bacteria from the rumen of cattle. J. Bacteriol. **53**(5): 631-645.

Hungate, R.E., Bryant, M.P., and Mah, R.A. (1964). The rumen bacteria and protozoa. Annu. Rev. Microbiol. **18**: 131-166.

Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007). MEGAN analysis of metagenomic data. Genome Res. **17**(3): 377-386.

Ingebritsen, T.S., and Cohen, P. (1983). Protein phosphatases: properties and role in cellular regulation. Science **221**(4608): 331-338.

Irwin, D., Shin, D.H., Zhang, S., Barr, B.K., Sakon, J., Karplus, P.A., and Wilson, D.B. (1998). Roles of the catalytic domain and two cellulose binding domains of

Thermomonospora fusca E4 in cellulose hydrolysis. J. Bacteriol. **180**(7): 1709-1714.

Jäger , G., Girfoglio, M., Dollo, F., Rinaldi, R., Bongard, H., Commandeur, U., Fischer, R., Spiess, A.C., and Büchs, J. (2011). How recombinant swollenin from *Kluyveromyces lactis* affects cellulosic substrates and accelerates their hydrolysis. Biotechnol Biofuels **4**(1): 33.

Janssen, P.H., and Kirs, M. (2008). Structure of the archaeal community of the rumen. Appl. Environ. Microbiol. **74**(12): 3619-3625.

Joblin, K.N. (2005). Methanogenic archaea. *In* Methods in gut microbial ecology for ruminants. *Edited by* H. P. S. Makkar, and C. S. McSweeney. Springer, Dordrecht, The Netherlands. pp. 47-53.

Joblin, K.N., Matsui, H., Naylor, G.E., and Ushida, K. (2002). Degradation of fresh ryegrass by methanogenic co-cultures of ruminal fungi grown in the presence or absence of *Fibrobacter succinogenes*. Curr. Microbiol. **45**(1): 46-53.

Jun, H.S., Qi, M., Ha, J.K., and Forsberg, C.W. (2007). *Fibrobacter succinogenes*, a dominant fibrolytic ruminal bacterium: transition to the post genomic era. Asian-Australas. J. Anim. Sci. **20**(5): 802-810.

Kamra, D.N.K.D.N. (2005). Rumen microbial ecosystem. Current Science **89**(1): 124-135.

Kang, S., Denman, S.E., Morrison, M., Yu, Z., and McSweeney, C.S. (2009). An efficient RNA extraction method for estimating gut microbial diversity by polymerase chain reaction. Curr. Microbiol. **58**(5): 464-471.

Keeling, P.J., and Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. Nature reviews. Genetics **9**(8): 605-618.

Kelly, W.J., Leahy, S.C., Altermann, E., Yeoman, C.J., Dunne, J.C., Kong, Z., Pacheco, D.M., Li, D., Noel, S.J., Moon, C.D., Cookson, A.L., and Attwood, G.T. (2010). The glycobiome of the rumen bacterium *Butyrivibrio proteoclasticus* B316(T) highlights adaptation to a polysaccharide-rich environment. PLoS One **5**(8): e11942.

Kim, H., Hahn, M., Grabowski, P., McPherson, D.C., Otte, M.M., Wang, R., Ferguson, C.C., Eichenberger, P., and Driks, A. (2006). The *Bacillus subtilis* spore coat protein interaction network. Mol. Microbiol. **59**(2): 487-502.

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

Kircher, M., and Kelso, J. (2010). High-throughput DNA sequencing--concepts and limitations. BioEssays : news and reviews in molecular, cellular and developmental biology **32**(6): 524-536.

Kittelmann, S., Naylor, G.E., Koolaard, J.P., and Janssen, P.H. (2012). A proposed taxonomy of anaerobic fungi (class Neocallimastigomycetes) suitable for large-scale sequence-based community structure analysis. PLoS One **7**(5): e36866.

Klieve, A.V., Bain, P.A., Yokoyama, M.T., Ouwerkerk, D., Forster, R.J., and Turner, A.F. (2004). Bacteriophages that infect the cellulolytic ruminal bacterium Ruminococcus albus AR67. Lett. Appl. Microbiol. **38**(4): 333-338.

Kong, Y., Teather, R., and Forster, R. (2010). Composition, spatial distribution, and diversity of the bacterial communities in the rumen of cows fed different forages. FEMS Microbiol. Ecol. **74**(3): 612-622.

Kopecny, J., Hodrova, B., and Stewart, C.S. (1996). The effect of rumen chitinolytic bacteria on cellulolytic anaerobic fungi. Lett. Appl. Microbiol. **23**(3): 199-202.

Krause, D.O., Denman, S.E., Mackie, R.I., Morrison, M., Rae, A.L., Attwood, G.T., and McSweeney, C.S. (2003). Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics. FEMS Microbiol. Rev. **27**(5): 663-693.

Kudo, H., Cheng, K.J., and Costerton, J.W. (1987). Interactions between *Treponema bryantii* and cellulolytic bacteria in the in vitro degradation of straw cellulose. Can. J. Microbiol. **33**(3): 244-248.

Kumar, R., Singh, S., and Singh, O.V. (2008). Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. J. Ind. Microbiol. Biotechnol. **35**(5): 377-391.

Kurasin, M., and Valjamae, P. (2011). Processivity of cellobiohydrolases is limited by the substrate. J. Biol. Chem. **286**(1): 169-177.

Lever, M. (1972). A new reaction for colorimetric determination of carbohydrates. Anal. Biochem. **47**(1): 273-279.

Li, J., and Heath, I.B. (1992). The phylogenetic relationships of the anaerobic chytridiomycetous gut fungi (Neocallimasticaceae) and chytridiomycota: I Cladistic analysis of rRNA sequences. Can. J. Bot. **70**(9): 1738-1746.

Li, J., Heath, I.B., and Packer, L. (1993). The phylogenetic relationships of the anaerobic chytridiomycetous gut fungi (Neocallimasticaceae) and the Chytridiomycota: II Cladistic analysis of structural data and description of Neocallimasticales ord nov. Can. J. Bot. **71**(3): 393-407.

Li, W. (2009). Analysis and comparison of very large metagenomes with fast clustering and functional annotation. BMC Bioinformatics **10**: 359.

Li, X.L., Chen, H., and Ljungdahl, L.G. (1997). Two cellulases, CelA and CelC, from the polycentric anaerobic fungus *Orpinomyces* strain PC-2 contain N-terminal docking domains for a cellulase-hemicellulase complex. Appl. Environ. Microbiol. **63**(12): 4721-4728.

Li, X.L., Skory, C.D., Cotta, M.A., Puchart, V., and Biely, P. (2008). Novel family of carbohydrate esterases, based on identification of the *Hypocrea jecorina* acetyl esterase gene. Appl. Environ. Microbiol. **74**(24): 7482-7489.

Liggenstoffer, A.S., Youssef, N.H., Couger, M.B., and Elshahed, M.S. (2010). Phylogenetic diversity and community structure of anaerobic gut fungi (phylum Neocallimastigomycota) in ruminant and non-ruminant herbivores. The ISME journal **4**(10): 1225-1235.

Loman, N.J., Constantinidou, C., Chan, J.Z., Halachev, M., Sergeant, M., Penn, C.W., Robinson, E.R., and Pallen, M.J. (2012). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat. Rev. Microbiol. **10**(9): 599-606.

Lowe, S.E., Theodorou, M.K., Trinci, A.P.J., and Hespell, R.B. (1985). Growth of anaerobic rumen fungi on defined and semi-defined media lacking rumen fluid. Journal of General Microbiology **131**(9): 2225-2230.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**(5): 955-964.

Lynd, L.R., Weimer, P.J., van Zyl, W.H., and Pretorius, I.S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. Microbiol. Mol. Biol. Rev. **66**(3): 506-577.

Malone, J.H., and Oliver, B. (2011). Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol **9**: 34.

Mardis, E.R. (2008). The impact of next-generation sequencing technology on genetics. Trends in genetics : TIG **24**(3): 133-141.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. **18**(9): 1509-1517.

Martín-Orúe, S.M., Balcells, J., Zakraoui, F., and Castrillo, C. (1998). Quantification and chemical composition of mixed bacteria harvested from solid fractions of rumen digesta: effect of detachment procedure. Anim. Feed Sci. Technol. **71**(3-4): 269-282.

Martin, J.A., and Wang, Z. (2011). Next-generation transcriptome assembly. Nature reviews. Genetics **12**(10): 671-682.

Martínez, A.T., Ruiz-Dueñas, F.J., Martínez, M.J., Del Río, J.C., and Gutiérrez, A. (2009a). Enzymatic delignification of plant cell wall: from nature to mill. Curr. Opin. Biotechnol. **20**(3): 348-357.

Martínez, M.E., Ranilla, M.J., Ramos, S., Tejido, M.L., Saro, C., and Carro, M.D. (2009b). Evaluation of procedures for detaching particle-associated microbes from forage and concentrate incubated in Rusitec fermenters: efficiency of recovery and representativeness of microbial isolates. J. Anim. Sci. **87**(6): 2064-2072.

McAllister, T.A., Bae, H.D., Jones, G.A., and Cheng, K.J. (1994). Microbial attachment and feed digestion in the rumen. J. Anim. Sci. **72**(11): 3004-3018.

McAllister, T.A., and Newbold, C.J. (2008). Redirecting rumen fermentation to reduce methanogenesis. Aust. J. Exp. Agric. **48**(2): 7-13.

McGavin, M., and Forsberg, C.W. (1988). Isolation and characterization of endoglucanases 1 and 2 from *Bacteroides succinogenes* S85. J. Bacteriol. **170**(7): 2914-2922.

McGinn, S.M., Beauchemin, K.A., Coates, T., and Colombatto, D. (2004). Methane emissions from beef cattle: Effects of monensin, sunflower oil, enzymes, yeast, and fumaric acid. J. Anim. Sci. **82**(11): 3346-3356.

Meguro, H., Morisaka, H., Kuroda, K., Miyake, H., Tamaru, Y., and Ueda, M. (2011). Putative role of cellulosomal protease inhibitors in *Clostridium cellulovorans* based on gene expression and measurement of activities. J. Bacteriol. **193**(19): 5527-5530.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., Gnirke, A., Jaenisch, R., and Lander,

E.S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature **454**(7205): 766-770.

Metzker, M.L. (2010). Sequencing technologies - the next generation. Nature reviews. Genetics **11**(1): 31-46.

Miron, J., Ben-Ghedalia, D., and Morrison, M. (2001). Invited review: adhesion mechanisms of rumen cellulolytic bacteria. J. Dairy Sci. **84**(6): 1294-1309.

Mitsumori, M., and Minato, H. (1995). Distribution of cellulose-binding proteins among the representative strains of rumen bacteria. J. Gen. Appl. Microbiol. **41**(4): 297-306.

Monteiro, L., Gras, N., Vidal, R., Cabrita, J., and Megraud, F. (2001). Detection of *Helicobacter pylori* DNA in human feces by PCR: DNA stability and removal of inhibitors. J. Microbiol. Methods **45**(2): 89-94.

Morgavi, D.P., Forano, E., Martin, C., and Newbold, C.J. (2010). Microbial ecosystem and methanogenesis in ruminants. Animal **4**(7): 1024-1036.

Morgavi, D.P., Sakurada, M., Mizokami, M., Tomita, Y., and Onodera, R. (1994). Effects of ruminal protozoa on cellulose degradation and the growth of an anaerobic ruminal fungus, *Piromyces* sp. strain OTS1, *in vitro*. Appl. Environ. Microbiol. **60**(10): 3718-3723.

Morrison, M., Daugherty, S.C., Nelson, W.C., Davidsen, T., Nelson, K.E., and North American Consortium for Fibrolytic Ruminal, B. (2010). The FibRumBa database: a resource for biologists with interests in gastrointestinal microbial ecology, plant biomass degradation, and anaerobic microbiology. Microb. Ecol. **59**(2): 212-213.

Mosoni, P., Martin, C., Forano, E., and Morgavi, D.P. (2011). Long-term defaunation increases the abundance of cellulolytic ruminococci and methanogens but does not affect the bacterial and methanogen diversity in the rumen of sheep. J. Anim. Sci. **89**(3): 783-791.

Naclerio, G., Baccigalupi, L., Zilhao, R., De Felice, M., and Ricca, E. (1996). *Bacillus subtilis* spore coat assembly requires cotH gene expression. J. Bacteriol. **178**(15): 4375-4380.

Nagalakshmi, U., Waern, K., and Snyder, M. (2010). RNA-Seq: a method for comprehensive transcriptome analysis. Current protocols in molecular biology(89): 89:84.11.81–84.11.13.

Nagaraja, T.G., and Titgemeyer, E.C. (2007). Ruminal acidosis in beef cattle: the current microbiological and nutritional outlook. J. Dairy Sci. **90 Suppl 1**: E17-38.

Nagpal, R., Puniya, A., Sehgal, J., and Singh, K. (2011). In vitro fibrolytic potential of anaerobic rumen fungi from ruminants and non-ruminant herbivores. Mycoscience **52**(1): 31-38.

Nagpal, R., Puniya, A.K., Griffith, G.W., Goel, G., Puniya, M., Sehgal, J.P., and Singh, K. (2009). Anaerobic rumen fungi: potential and applications. *In* Agriculturally Important Microorganisms. *Edited by* G.G. Khachatourians, D.K. Arora, T.P. Rajendran, and A.K. Srivastava. Academic World International. pp. 375-393.

Nagy, T., Tunnicliffe, R.B., Higgins, L.D., Walters, C., Gilbert, H.J., and Williamson, M.P. (2007). Characterization of a double dockerin from the cellulosome of the anaerobic fungus *Piromyces equi*. J. Mol. Biol. **373**(3): 612-622.

Neidhardt, F.C., and Umbarger, H.E. (1996). Chemical composition of *Escherichia coli*. *In Escherichia coli* and *Salmonella*: Cellular and Molecular Biology *Edited by* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger. ASM Press, Washington, D.C. pp. 13-17.

Nicholson, M.J., Theodorou, M.K., and Brookman, J.L. (2005). Molecular analysis of the anaerobic rumen fungus *Orpinomyces* - insights into an AT-rich genome. Microbiology (Reading, England) **151**(Pt 1): 121-133.

Nikolaev, Y.A., and Plakunov, V.K. (2007). Biofilm - "City of microbes" or an analogue of multicellular organisms? Microbiology **76**(2): 125-138.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. **302**(1): 205-217.

Novotná, Z., Procházka, J., Simůnek, J., and Fliegerová, K. (2010). Xylanases of anaerobic fungus *Anaeromyces mucronatus*. Folia Microbiol. **55**(4): 363-367.

Odenyo, A.A., Mackie, R.I., Stahl, D.A., and White, B.A. (1994). The use of 16S rRNA-targeted oligonucleotide probes to study competition between ruminal fibrolytic bacteria: pure-culture studies with cellulose and alkaline peroxide-treated wheat straw. Appl. Environ. Microbiol. **60**(10): 3697-3703.

Orpin, C.G. (1975). Studies on the rumen flagellate *Neocallimastix frontalis*. Journal of general microbiology **91**(2): 249-262.

Orpin, C.G., and Joblin, K.N. (1988). The rumen anaerobic fungi. *In* The rumen microbial ecosystem. *Edited by* P. N. Hobson. Elsevier Appl Sci, London. pp. 129-150.

Orpin, C.G., and Joblin, K.N. (1997). The rumen anaerobic fungi. *In* The rumen microbial ecosystem. *Edited by* P. N. Hobson, and C. S. Stewart. Blackie Academic and Professional. pp. 140-195.

Ozkose, E., Thomas, B.J., Davies, D.R., Griffith, G.W., and Theodorou, M.K. (2001). *Cyllamyces aberensis* gen.nov. sp.nov., a new anaerobic gut fungus with branched sporangiophores isolated from cattle. Can. J. Bot. **79**(6): 666-673.

Ozsolak, F., and Milos, P.M. (2011). RNA sequencing: advances, challenges and opportunities. Nature reviews. Genetics **12**(2): 87-98.

Park, P.J. (2009). ChIP-seq: advantages and challenges of a maturing technology. Nat. Rev. Genet. **10**(10): 669-680.

Peer, A., Smith, S.P., Bayer, E.A., Lamed, R., and Borovok, I. (2009). Noncellulosomal cohesin- and dockerin-like modules in the three domains of life. FEMS Microbiol. Lett. **291**(1): 1-16.

Percival Zhang, Y.H., Himmel, M.E., and Mielenz, J.R. (2006). Outlook for cellulase improvement: screening and selection strategies. Biotechnol. Adv. **24**(5): 452-481.

Peršoh, D., Theuerl, S., Buscot, F., and Rambold, G. (2008). Towards a universally adaptable method for quantitative extraction of high-purity nucleic acids from soil. J. Microbiol. Methods **75**(1): 19-24.

Pop, M., and Salzberg, S.L. (2008). Bioinformatics challenges of new sequencing technology. Trends Genet. **24**(3): 142-149.

Pope, P.B., Denman, S.E., Jones, M., Tringe, S.G., Barry, K., Malfatti, S.A., McHardy, A.C., Cheng, J.F., Hugenholtz, P., McSweeney, C.S., and Morrison, M. (2010). Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. Proc. Natl. Acad. Sci. U. S. A. **107**(33): 14793-14798.

Pope, P.B., Mackenzie, A.K., Gregor, I., Smith, W., Sundset, M.A., McHardy, A.C., Morrison, M., and Eijsink, V.G. (2012). Metagenomics of the svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization Loci. PLoS One **7**(6): e38571.

Poretsky, R.S., Hewson, I., Sun, S., Allen, A.E., Zehr, J.P., and Moran, M.A. (2009). Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. Environmental microbiology **11**(6): 1358-1375.

Poroyko, V., White, J.R., Wang, M., Donovan, S., Alverdy, J., Liu, D.C., and Morowitz, M.J. (2010). Gut microbial gene expression in mother-fed and formula-fed piglets. PLoS One **5**(8): e12459.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glockner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. **35**(21): 7188-7196.

Purushe, J., Fouts, D.E., Morrison, M., White, B.A., Mackie, R.I., North American Consortium for Rumen, B., Coutinho, P.M., Henrissat, B., and Nelson, K.E. (2010). Comparative genome analysis of *Prevotella ruminicola* and *Prevotella bryantii*: insights into their environmental niche. Microb. Ecol. **60**(4): 721-729.

Qi, M., Nelson, K.E., Daugherty, S.C., Nelson, W.C., Hance, I.R., Morrison, M., and Forsberg, C.W. (2005). Novel molecular features of the fibrolytic intestinal bacterium *Fibrobacter intestinalis* not shared with *Fibrobacter succinogenes* as determined by suppressive subtractive hybridization. J. Bacteriol. **187**(11): 3739-3751.

Qi, M., Wang, P., Selinger, L.B., Yanke, L.J., Forster, R.J., and McAllister, T.A. (2011). Isolation and characterization of a ferulic acid esterase (Fae1A) from the rumen fungus *Anaeromyces mucronatus*. J. Appl. Microbiol. **110**(5): 1341-1350.

Raghothama, S., Eberhardt, R.Y., Simpson, P., Wigelsworth, D., White, P., Hazlewood, G.P., Nagy, T., Gilbert, H.J., and Williamson, M.P. (2001). Characterization of a cellulosome dockerin domain from the anaerobic fungus *Piromyces equi*. Nat. Struct. Biol. **8**(9): 775-778.

Ramos, S., Tejido, M.L., Ranilla, M.J., Mart ńez, M.E., Saro, C., and Carro, M.D. (2009). Influence of detachment procedure and diet on recovery of solid-associated bacteria from sheep ruminal digesta and representativeness of bacterial isolates as assessed

by automated ribosomal intergenic spacer analysis-polymerase chain reaction. J. Dairy Sci. **92**(11): 5659-5668.

Ranilla, M.J., and Carro, M.D. (2003). Diet and procedures used to detach particle-associated microbes from ruminal digesta influence chemical composition of microbes and estimation of microbial growth in Rusitec fermenters. J. Anim. Sci. **81**(2): 537-544.

Ransom-Jones, E., Jones, D.L., McCarthy, A.J., and McDonald, J.E. (2012). The Fibrobacteres: an important phylum of cellulose-degrading bacteria. Microb. Ecol. **63**(2): 267-281.

Reverbel-Leroy, C., Pages, S., Belaich, A., Belaich, J.P., and Tardif, C. (1997). The processive endocellulase CelF, a major component of the *Clostridium cellulolyticum* cellulosome: purification and characterization of the recombinant form. J. Bacteriol. **179**(1): 46-52.

Rezaeian, M., Beakes, G.W., and Parker, D.S. (2004). Distribution and estimation of anaerobic zoosporic fungi along the digestive tracts of sheep. Mycol. Res. **108**(Pt 10): 1227-1233.

Ricard, G., McEwan, N.R., Dutilh, B.E., Jouany, J.P., Macheboeuf, D., Mitsumori, M., McIntosh, F.M., Michalowski, T., Nagamine, T., Nelson, N., Newbold, C.J., Nsabimana, E., Takenaka, A., Thomas, N.A., Ushida, K., Hackstein, J.H., and Huynen, M.A. (2006). Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. BMC Genomics **7**: 22.

Rivas, L., Coffey, B., McAuliffe, O., McDonnell, M.J., Burgess, C.M., Coffey, A., Ross, R.P., and Duffy, G. (2010). In vivo and ex vivo evaluations of bacteriophages e11/2 and e4/1c for use in the control of *Escherichia coli* O157:H7. Appl. Environ. Microbiol. **76**(21): 7210-7216.

Russell, J.B., and Rychlik, J.L. (2001). Factors that alter rumen microbial ecology. Science (New York, N.Y.) **292**(5519): 1119-1122.

Sakon, J., Irwin, D., Wilson, D.B., and Karplus, P.A. (1997). Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. Nat. Struct. Biol. **4**(10): 810-818.

Saloheimo, M., Paloheimo, M., Hakola, S., Pere, J., Swanson, B., Nyyssonen, E., Bhatia, A., Ward, M., and Penttila, M. (2002). Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. Eur. J. Biochem. **269**(17): 4202-4211.

Sánchez, C. (2009). Lignocellulosic residues: biodegradation and bioconversion by fungi. Biotechnol. Adv. **27**(2): 185-194.

Schadt, E.E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. Hum. Mol. Genet. **19**(R2): R227-240.

Scheller, H.V., and Ulvskov, P. (2010). Hemicelluloses. Annu. Rev. Plant Biol. **61**: 263-289.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA

integrity number for assigning integrity values to RNA measurements. BMC Mol. Biol. **7**: 3.

Selinger, L.B., Forsberg, C.W., and Cheng, K.J. (1996). The rumen: a unique source of enzymes for enhancing livestock production. Anaerobe **2**(5): 263-284.

Sessitsch, A., Gyamfi, S., Stralis-Pavese, N., Weilharter, A., and Pfeifer, U. (2002). RNA isolation from soil for bacterial community and functional analysis: evaluation of different extraction and soil conservation protocols. J. Microbiol. Methods **51**(2): 171-179.

Sharma, R., John, S.J., Damgaard, M., and McAllister, T.A. (2003). Extraction of PCR-quality plant and microbial DNA from total rumen contents. BioTechniques **34**(1): 92-94, 96-97.

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. Nat. Biotechnol. **26**(10): 1135-1145.

Simon, C., and Daniel, R. (2011). Metagenomic analyses: past and future trends. Appl. Environ. Microbiol. **77**(4): 1153-1161.

Sorek, R., and Cossart, P. (2010). Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat. Rev. Genet. **11**(1): 9-16.

Stark, M., Berger, S.A., Stamatakis, A., and von Mering, C. (2010). MLTreeMap--accurate Maximum Likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. BMC Genomics **11**: 461.

Steenbakkers, P.J., Irving, J.A., Harhangi, H.R., Swinkels, W.J., Akhmanova, A., Dijkerman, R., Jetten, M.S., van der Drift, C., Whisstock, J.C., and Op den Camp, H.J. (2008). A serpin in the cellulosome of the anaerobic fungus *Piromyces* sp. strain E2. Mycol. Res. **112**(Pt 8): 999-1006.

Steenbakkers, P.J., Li, X.L., Ximenes, E.A., Arts, J.G., Chen, H., Ljungdahl, L.G., and Op Den Camp, H.J. (2001). Noncatalytic docking domains of cellulosomes of anaerobic fungi. J. Bacteriol. **183**(18): 5325-5333.

Stevenson, D.M., and Weimer, P.J. (2007). Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. Appl. Microbiol. Biotechnol. **75**(1): 165-174.

Stewart, C.S., Flint, H.J., and Bryant, M.P. (1997). The rumen bacteria. *In* The Rumen microbial Ecosystem. *Edited by* P. N. Hobson, and C. S. Stewart. Blackie Academic and Professional. pp. 10-72.

Su, C., Lei, L., Duan, Y., Zhang, K.Q., and Yang, J. (2012). Culture-independent methods for studying environmental microorganisms: methods, application, and perspective. Appl. Microbiol. Biotechnol. **93**(3): 993-1003.

Suen, G., Stevenson, D.M., Bruce, D.C., Chertkov, O., Copeland, A., Cheng, J.F., Detter, C., Detter, J.C., Goodwin, L.A., Han, C.S., Hauser, L.J., Ivanova, N.N., Kyrpides, N.C., Land, M.L., Lapidus, A., Lucas, S., Ovchinnikova, G., Pitluck, S., Tapia, R., Woyke, T., Boyum, J., Mead, D., and Weimer, P.J. (2011a). Complete genome of the cellulolytic ruminal bacterium *Ruminococcus albus* 7. J. Bacteriol. **193**(19): 5574-5575.

Suen, G., Weimer, P.J., Stevenson, D.M., Aylward, F.O., Boyum, J., Deneke, J., Drinkwater, C., Ivanova, N.N., Mikhailova, N., Chertkov, O., Goodwin, L.A., Currie, C.R., Mead, D., and Brumm, P.J. (2011b). The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. PLoS One **6**(4): e18814.

Suetake, T., Tsuda, S., Kawabata, S., Miura, K., Iwanaga, S., Hikichi, K., Nitta, K., and Kawano, K. (2000). Chitin-binding proteins in invertebrates and plants comprise a common chitin-binding structural motif. J. Biol. Chem. **275**(24): 17929-17932.

Sukharnikov, L.O., Cantwell, B.J., Podar, M., and Zhulin, I.B. (2011). Cellulases: ambiguous nonhomologous enzymes in a genomic perspective. Trends Biotechnol. **29**(10): 473-479.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., and Natale, D.A. (2003). The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4**: 41.

Tilman, D., Socolow, R., Foley, J.A., Hill, J., Larson, E., Lynd, L., Pacala, S., Reilly, J., Searchinger, T., Somerville, C., and Williams, R. (2009). Energy. Beneficial biofuels--the food, energy, and environment trilemma. Science **325**(5938): 270-271.

Trabalza-Marinucci, M., Poncet, C., Delval, E., and Fonty, G. (2006). Evaluation of techniques to detach particle-associated microorganisms from rumen contents. Anim. Feed Sci. Technol. **125**(1-2): 1-16.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. **28**(5): 511-515.

Treangen, T.J., and Salzberg, S.L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat. Rev. Genet. **13**(1): 36-46.

Tuckwell, D.S., Nicholson, M.J., McSweeney, C.S., Theodorou, M.K., and Brookman, J.L. (2005). The rapid assignment of ruminal fungi to presumptive genera using ITS1 and ITS2 RNA secondary structures to produce group-specific fingerprints. Microbiology (Reading, England) **151**(Pt 5): 1557-1567.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., Egholm, M., Henrissat, B., Heath, A.C., Knight, R., and Gordon, J.I. (2009). A core gut microbiome in obese and lean twins. Nature **457**(7228): 480-484.

Urich, T., Lanzen, A., Qi, J., Huson, D.H., Schleper, C., and Schuster, S.C. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PLoS One **3**(6): e2527.

van den Brink, J., and de Vries, R.P. (2011). Fungal enzyme sets for plant polysaccharide degradation. Appl. Microbiol. Biotechnol. **91**(6): 1477-1492.

Varel, V.H., Richardson, A.J., and Stewart, C.S. (1989). Degradation of barley straw, ryegrass, and alfalfa cell walls by Clostridium longisporum and Ruminococcus albus. Appl. Environ. Microbiol. **55**(12): 3080-3084.

Wallace, R.J. (1994). Ruminal microbiology, biotechnology, and ruminant nutrition: progress and problems. J. Anim. Sci. **72**(11): 2992-3003.

Wang, T.Y., Chen, H.L., Lu, M.Y., Chen, Y.C., Sung, H.M., Mao, C.T., Cho, H.Y., Ke, H.M., Hwa, T.Y., Ruan, S.K., Hung, K.Y., Chen, C.K., Li, J.Y., Wu, Y.C., Chen, Y.H., Chou, S.P., Tsai, Y.W., Chu, T.C., Shih, C.C., Li, W.H., and Shih, M.C. (2011). Functional characterization of cellulases identified from the cow rumen fungus *Neocallimastix patriciarum* W5 by transcriptomic and secretomic analyses. Biotechnology for biofuels **4**: 24.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. **10**(1): 57-63.

Warnecke, F., and Hess, M. (2009). A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. J. Biotechnol. **142**(1): 91-95.

Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., Cayouette, M., McHardy, A.C., Djordjevic, G., Aboushadi, N., Sorek, R., Tringe, S.G., Podar, M., Martin, H.G., Kunin, V., Dalevi, D., Madejska, J., Kirton, E., Platt, D., Szeto, E., Salamov, A., Barry, K., Mikhailova, N., Kyrpides, N.C., Matson, E.G., Ottesen, E.A., Zhang, X., Hernandez, M., Murillo, C., Acosta, L.G., Rigoutsos, I., Tamayo, G., Green, B.D., Chang, C., Rubin, E.M., Mathur, E.J., Robertson, D.E., Hugenholtz, P., and Leadbetter, J.R. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature **450**(7169): 560-565.

Wasiak, S., Legendre-Guillemin, V., Puertollano, R., Blondeau, F., Girard, M., de Heuvel, E., Boismenu, D., Bell, A.W., Bonifacino, J.S., and McPherson, P.S. (2002). Enthoprotin: a novel clathrin-associated protein identified through subcellular proteomics. J. Cell Biol. **158**(5): 855-862.

Weimer, P.J., Russell, J.B., and Muck, R.E. (2009). Lessons from the cow: what the ruminant animal can teach us about consolidated bioprocessing of cellulosic biomass. Bioresour. Technol. **100**(21): 5323-5331.

Wells, J.E., and Russell, J.B. (1996). Why do many ruminal bacteria die and lyse so quickly? J. Dairy Sci. **79**(8): 1487-1495.

Whitehouse, N.L., Olson, V.M., Schwab, C.G., Chesbro, W.R., Cunningham, K.D., and Lykos, T. (1994). Improved techniques for dissociating particle-associated mixed ruminal microorganisms from ruminal digesta solids. J. Anim. Sci. **72**(5): 1335-1343.

Whitford, M.F., Teather, R.M., and Forster, R.J. (2001). Phylogenetic analysis of methanogens from the bovine rumen. BMC Microbiol. **1**: 5.

Williams, A.G. (1986). Rumen holotrich ciliate protozoa. Microbiol Rev **50**(1): 25-49.

Williams, A.G., and Coleman, G.S. (1997). The rumen protozoa. *In* The rumen microbial ecosystem. *Edited by* P. N. Hobson, and C. S. Stewart. Blackie Academic and Professional. pp. 73-139.

Williams, A.G., and Strachan, N.H. (1984). Polysaccharide degrading enzymes in microbial populations from the liquid and solid fractions of bovine rumen digesta. Can. J. Anim. Sci. **64**: 58-59.

Willner, D., Thurber, R.V., and Rohwer, F. (2009). Metagenomic signatures of 86 microbial and viral metagenomes. Environ Microbiol **11**(7): 1752-1766.

Wilson, C.A., and Wood, T.M. (1992). Studies on the cellulase of the rumen anaerobic fungus Neocallimastix frontalis, with special reference to the capacity of the enzyme to degrade crystalline cellulose. Enzyme. Microb. Technol. **14**(4): 258-264.

Wilson, D. (2009). Evidence for a novel mechanism of microbial cellulose degradation. Cellulose **16**(4): 723-727.

Wilson, D.B. (2008). Three microbial strategies for plant cell wall degradation. Ann. N. Y. Acad. Sci. **1125**: 289-297.

Wilson, D.B. (2012). Processive and nonprocessive cellulases for biofuel production--lessons from bacterial genomes and structural analysis. Appl. Microbiol. Biotechnol. **93**(2): 497-502.

Wolin, M.J., Miller, T.L., and Stewart, C.S. (1997). Microbe-microbe interactions. *In* The rumen microbial ecosystem. *Edited by* P. N. Hobson, and C. S. Stewart. Blackie Academic and Professional. pp. 467-491.

Xie, G., Bruce, D.C., Challacombe, J.F., Chertkov, O., Detter, J.C., Gilna, P., Han, C.S., Lucas, S., Misra, M., Myers, G.L., Richardson, P., Tapia, R., Thayer, N., Thompson, L.S., Brettin, T.S., Henrissat, B., Wilson, D.B., and McBride, M.J. (2007). Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. Appl. Environ. Microbiol. **73**(11): 3536-3546.

Yang, H.J., and Yue, Q. (2012). The modification of glucose levels and N source in the Hungate's medium to stimulate the production of fibrolytic enzymes of *Anaeromyces* sp. YQ3 grown on corn stalks. Anim. Feed Sci. Technol. **171**(2–4): 146-153.

Yao, Q., Sun, T.T., Liu, W.F., and Chen, G.J. (2008). Gene cloning and heterologous expression of a novel endoglucanase, swollenin, from *Trichoderma pseudokoningii* S38. Biosci. Biotechnol. Biochem. **72**(11): 2799-2805.

Yaoi, K., Kondo, H., Hiyoshi, A., Noro, N., Sugimoto, H., Tsuda, S., Mitsuishi, Y., and Miyazaki, K. (2007). The structural basis for the exo-mode of action in GH74 oligoxyloglucan reducing end-specific cellobiohydrolase. J. Mol. Biol. **370**(1): 53-62.

Yu, Z., and Forster, R.J. (2005). Nucleic acid extraction, oligonucleotide probes and PCR methods. *In* Methods in gut microbial ecology for ruminants. *Edited by* H. P. S. Makkar, and C. S. McSweeney. Springer. pp. 81-104.

Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**(5): 821-829.

Zoetendal, E.G., Booijink, C.C., Klaassens, E.S., Heilig, H.G., Kleerebezem, M., Smidt, H., and de Vos, W.M. (2006). Isolation of RNA from bacterial samples of the human gastrointestinal tract. Nature protocols **1**(2): 954-959.

Zverlov, V.V., Kellermann, J., and Schwarz, W.H. (2005). Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes. Proteomics **5**(14): 3646-3653.

Appendices

Supplementary Tables

Table S.1 Muskoxen rumen metatranscriptome contigs (≥500 bp) that have one putative CAZY module

Contig number	Domains	Length (bp)	Number of Reads	Accession	GI	E-value	HSP Length	HSP Id%	Hit Description
Contig26982	GH1	1879	11902	AAD45834	gi|5639612|gb|AAD45834.1|AF016864_1	0	598	82.3	beta-glucosidase [Orpinomyces sp. PC-2]
Contig27636	GH1	734	1195	AAD45834	gi|5639612|gb|AAD45834.1|AF016864_1	1.00E-104	194	85.6	beta-glucosidase [Orpinomyces sp. PC-2]
Contig30163	GH1	954	11088	AAD45834	gi|5639612|gb|AAD45834.1|AF016864_1	1.00E-146	328	74.1	beta-glucosidase [Orpinomyces sp. PC-2]
Contig22426	GH1	846	2011	AAD45834	gi|5639612|gb|AAD45834.1|AF016864_1	1.00E-142	261	89.7	beta-glucosidase [Orpinomyces sp. PC-2]
Contig22325	GH1	895	2868	AAD45834	gi|5639612|gb|AAD45834.1|AF016864_1	1.00E-103	270	67.0	beta-glucosidase [Orpinomyces sp. PC-2]
Contig3834	GH1	691	2289	AAP30745	gi|30315031|gb|AAP30745.1|	3.00E-66	175	73.7	beta-glucosidase Cel1C [Piromyces sp. E2]
Contig29365	GH1	683	3236	AAD45834	gi|5639612|gb|AAD45834.1|AF016864_1	5.00E-81	216	69.4	beta-glucosidase [Orpinomyces sp. PC-2]
Contig12311	GH1	1369	11613	CAC34952	gi|13445202|emb|CAC34952.1|	0	367	83.7	beta-glucosidase [Piromyces sp. E2]
Contig21506	GH1	821	14352	AAP30745	gi|30315031|gb|AAP30745.	1.00E-110	105	88.6	beta-glucosidase Cel1C [Piromyces sp. E2]

					1|			
Contig29533	GH1	819	14351	AAD45834	gi|5639612|gb|AAD45834.1|AF016864_1	1.00E-138	254	88.2	beta-glucosidase [Orpinomyces sp. PC-2]
Contig29741	GH1	586	13272	AAP30745	gi|30315031|gb|AAP30745.1|	7.00E-91	183	84.7	beta-glucosidase Cel1C [Piromyces sp. E2]
Contig29793	GH1	574	2659	AAD45834	gi|5639612|gb|AAD45834.1|AF016864_1	1.00E-101	179	96.1	beta-glucosidase [Orpinomyces sp. PC-2]
Contig26250	GH1	982	2876	CAC34952	gi|13445202|emb|CAC34952.1|	1.00E-169	326	84.7	beta-glucosidase [Piromyces sp. E2]
Contig30783	GH1	699	5751	AAD45834	gi|5639612|gb|AAD45834.1|AF016864_1	1.00E-126	232	90.1	beta-glucosidase [Orpinomyces sp. PC-2]
Contig21061	GH1	605	3796	AAP30745	gi|30315031|gb|AAP30745.1|	8.00E-94	200	79.0	beta-glucosidase Cel1C [Piromyces sp. E2]
Contig8436	GH1	541	193	CAC34952	gi|13445202|emb|CAC34952.1|	2.00E-86	180	78.3	beta-glucosidase [Piromyces sp. E2]
Contig20700	GH2C	776	254	ZP_06142336	gi|268608609|ref|ZP_06142336.1|	1.00E-127	259	80.7	beta-galactosidase [Ruminococcus flavefaciens FD-1]
Contig2681	GH2C	730	248	ZP_06142336	gi|268608609|ref|ZP_06142336.1|	1.00E-114	245	77.6	beta-galactosidase [Ruminococcus flavefaciens FD-1]
Contig5538	GH2C	663	168	ZP_06142336	gi|268608609|ref|ZP_06142336.1|	1.00E-105	223	78.9	beta-galactosidase [Ruminococcus flavefaciens FD-1]
Contig290*	GH2C	1528	89	YP_001560254	gi|160881286|ref|YP_001560254.1|	1.00E-160	507	53.3	glycoside hydrolase family protein [Clostridium phytofermentans ISDg]gi|160429952|gb|ABX43515.1| glycoside hydrolase family 2 sugar binding [Clostridium phytofermentans ISDg]
Contig1357	GH2C	1203	347	CBL34622	gi|291557505|emb|CBL34622.1|	1.00E-112	407	50.9	Beta-galactosidase/beta-glucuronidase [Eubacterium siraeum V10Sc8a]
Contig8041	GH2C	567	216	ADE82717	gi|294473328|gb|ADE82717	4.00E-89	189	80.4	glycosyl hydrolase, family 2 [Prevotella ruminicola 23]

					.1|			
Contig7865	GH2C	633	150	YP_001560254	gi|160881286|ref|YP_001560254.1|	5.00E-78	211	62.1	glycoside hydrolase family protein [Clostridium phytofermentans ISDg]gi|160429952|gb|ABX43515.1| glycoside hydrolase family 2 sugar binding [Clostridium phytofermentans ISDg]
Contig801	GH2C	844	496	ZP_03477286	gi|218263041|ref|ZP_03477286.1|	1.00E-88	278	54.0	hypothetical protein PRABACTJOHN_02967 [Parabacteroides johnsonii DSM 18315]gi|218222974|gb|EEC95624.1| hypothetical protein PRABACTJOHN_02967 [Parabacteroides johnsonii DSM 18315]
Contig28997	GH2C	1736	3250	ZP_03477286	gi|218263041|ref|ZP_03477286.1|	1.00E-156	570	48.8	hypothetical protein PRABACTJOHN_02967 [Parabacteroides johnsonii DSM 18315]gi|218222974|gb|EEC95624.1| hypothetical protein PRABACTJOHN_02967 [Parabacteroides johnsonii DSM 18315]
Contig671	GH2C	2005	37	ZP_06617065	gi|293370513|ref|ZP_06617065.1|	1.00E-167	589	49.6	glycosyl hydrolase family 2, sugar binding domain protein [Bacteroides ovatus SD CMC 3f]gi|292634247|gb|EFF52784.1| glycosyl hydrolase family 2, sugar binding domain protein [Bacteroides ovatus SD CMC 3f]
Contig21714	GH3	2343	1925	AAO41704	gi|28557461|gb|AAO41704.1|	0	700	57.9	beta-glucosidase precursor [Piromyces sp. E2]
Contig9134	GH3	1201	746	AAO41704	gi|28557461|gb|AAO41704.1|	4.00E-89	342	49.1	beta-glucosidase precursor [Piromyces sp. E2]
Contig23162	GH3	918	678	AAO41704	gi|28557461|gb|AAO41704.1|	1.00E-80	305	51.5	beta-glucosidase precursor [Piromyces sp. E2]
Contig4265	GH3	2203	1437	AAO41704	gi|28557461|gb|AAO41704.1|	0	709	49.4	beta-glucosidase precursor [Piromyces sp. E2]
Contig5433*	GH3	2446	4116	AAO41704	gi|28557461|gb|AAO41704.1|	1.00E-170	713	47.0	beta-glucosidase precursor [Piromyces sp. E2]
Contig24191	GH3	957	857	AAO41704	gi|28557461|gb|AAO41704.1|	8.00E-66	301	43.2	beta-glucosidase precursor [Piromyces sp. E2]
Contig29507*	GH3	3032	4266	ZP_04744384	gi|240145783|ref|ZP_04744	1.00E-133	800	37.1	beta-glucosidase A [Roseburia intestinalis L1-82]gi|257202114|gb|EEV00399.1| beta-

					384.1|			glucosidase A [Roseburia intestinalis L1-82]
Contig463*	GH3	2676	360	CBL11427	gi|291538316|emb|CBL11427.1|	0	805	50.8	Beta-glucosidase-related glycosidases [Roseburia intestinalis XB6B4]
Contig14426	GH3	2061	1055	CBL11427	gi|291538316|emb|CBL11427.1|	0	690	49.7	Beta-glucosidase-related glycosidases [Roseburia intestinalis XB6B4]
Contig3004*	GH3	1834	2108	CBL11427	gi|291538316|emb|CBL11427.1|	1.00E-163	608	50.0	Beta-glucosidase-related glycosidases [Roseburia intestinalis XB6B4]
Contig1098	GH3	1034	567	ACZ98612	gi|280977789|gb|ACZ98612.1|	2.00E-76	328	49.1	glucosidase [Cellulosilyticum ruminicola]
Contig23195	GH3	2015	949	ZP_04746179	gi|240147578|ref|ZP_04746179.1|	1.00E-169	659	46.6	beta-glucosidase [Roseburia intestinalis L1-82]gi|257200210|gb|EEU98494.1| beta-glucosidase [Roseburia intestinalis L1-82]
Contig3679	GH3	549	196	AAO41704	gi|28557461|gb|AAO41704.1|	4.00E-52	185	54.1	beta-glucosidase precursor [Piromyces sp. E2]
Contig13027	GH3	664	457	AAO41704	gi|28557461|gb|AAO41704.1|	5.00E-42	224	41.5	beta-glucosidase precursor [Piromyces sp. E2]
Contig25386	GH3	974	369	CBL11427	gi|291538316|emb|CBL11427.1|	2.00E-76	328	48.5	Beta-glucosidase-related glycosidases [Roseburia intestinalis XB6B4]
Contig6513	GH3	950	441	ZP_06646492	gi|293402355|ref|ZP_06646492.1|	2.00E-62	257	49.0	beta-N-acetylhexosaminidase [Erysipelotrichaceae bacterium 5_2_54FAA]gi|291304202|gb|EFE45454.1| beta-N-acetylhexosaminidase [Erysipelotrichaceae bacterium 5_2_54FAA]
Contig21869	GH3	1046	795	ZP_04746179	gi|240147578|ref|ZP_04746179.1|	1.00E-71	356	44.4	beta-glucosidase [Roseburia intestinalis L1-82]gi|257200210|gb|EEU98494.1| beta-glucosidase [Roseburia intestinalis L1-82]
Contig5802	GH3	524	321	AAO41704	gi|28557461|gb|AAO41704.1|	7.00E-54	168	58.3	beta-glucosidase precursor [Piromyces sp. E2]
Contig22281	GH3	860	1140	AAO41704	gi|28557461|gb|AAO41704.1|	3.00E-60	288	44.1	beta-glucosidase precursor [Piromyces sp. E2]
Contig21835	Cellulase	2428	4234	BAC57896	gi|51090374|dbj|BAC57896.	0	746	86.5	cellulase celA [Polyplastron multivesiculatum]

					2|			
Contig21651	Cellulase	2258	4303	BAC57896	gi|51090374|dbj|BAC57896.2|	0	639	69.2	cellulase celA [Polyplastron multivesiculatum]
Contig29732	Cellulase	1569	1435	AAC06321	gi|2981484|gb|AAC06321.1|	0	513	76.0	cellulase CelD [Neocallimastix patriciarum]
Contig792	Cellulase	1400	145	AAC06321	gi|2981484|gb|AAC06321.1|	0	472	76.5	cellulase CelD [Neocallimastix patriciarum]
Contig21888*	Cellulase	1388	2071	ZP_06720041	gi|294642164|ref|ZP_06720041.1|	1.00E-148	441	58.5	cellulase (glycosyl hydrolase family 5) [Ruminococcus albus 8]gi|291503294|gb|EFF16053.1| cellulase (glycosyl hydrolase family 5) [Ruminococcus albus 8]
Contig4250	Cellulase	1297	2251	CAL91968	gi|218081332|emb|CAL91968.1|	0	480	67.9	cellulase [Epidinium ecaudatum]
Contig22088*	Cellulase	1229	2727	CAL91974	gi|218081346|emb|CAL91974.1|	1.00E-166	354	75.4	cellulase [Epidinium ecaudatum]
Contig28894	Cellulase	1256	1373	CAL91974	gi|218081346|emb|CAL91974.1|	1.00E-136	356	64.6	cellulase [Epidinium ecaudatum]
Contig98	Cellulase	2087	64	BAC57896	gi|51090374|dbj|BAC57896.2|	1.00E-142	397	59.4	cellulase celA [Polyplastron multivesiculatum]
Contig1999	Cellulase	1192	63	BAC57893	gi|28569970|dbj|BAC57893.1|	1.00E-150	323	69.0	endoglucanase epi2 [Epidinium caudatum]
Contig4398	Cellulase	1097	353	CAH69214	gi|59932919|emb|CAH69214.1|	0	364	94.0	cellulase family 5 protein [Epidinium ecaudatum]gi|218081349|emb|CAL91975.1| cellulase [Epidinium ecaudatum]
Contig28882	Cellulase	1250	4749	AAD30364	gi|4836168|gb|AAD30364.1|AF078739_2	1.00E-141	302	60.6	CelB [Caldicellulosiruptor sp. Tok7B.1]
Contig28881	Cellulase	1118	951	BAC57896	gi|51090374|dbj|BAC57896.2|	1.00E-122	375	56.0	cellulase celA [Polyplastron multivesiculatum]
Contig21436*	Cellulase	1585	4195	CAH69214	gi|59932919|emb|CAH69214.1|	0	485	84.1	cellulase family 5 protein [Epidinium ecaudatum]gi|218081349|emb|CAL91975.1| cellulase [Epidinium ecaudatum]
Contig1110	Cellulase	1120	76	CBK96759	gi|291531174|	1.00E	359	58.2	Endoglucanase [Eubacterium siraeum 70/3]

					emb|CBK96759.1|	-115			
Contig29277	Cellulase	1165	2707	BAC57896	gi|51090374|dbj|BAC57896.2|	1.00E-158	392	64.0	cellulase celA [Polyplastron multivesiculatum]
Contig2730	Cellulase	597	78	AAC06321	gi|2981484|gb|AAC06321.1|	1.00E-115	198	100.0	cellulase CelD [Neocallimastix patriciarum]
Contig6521	Cellulase	874	441	BAA76394	gi|4586414|dbj|BAA76394.1|	1.00E-155	256	100.0	endo-1,4-beta-glucanase [Epidinium caudatum]
Contig2292	Cellulase	922	231	CAL91969	gi|218081334|emb|CAL91969.1|	1.00E-104	301	61.5	cellulase [Epidinium ecaudatum]
Contig21902	Cellulase	827	301	ZP_06145755	gi|268612028|ref|ZP_06145755.1|	5.00E-89	240	65.8	cellulose 1,4-beta-cellobiosidase [Ruminococcus flavefaciens FD-1]
Contig26471	Cellulase	660	185	1ECE	gi|1827681|pdb|1ECE|A	1.00E-60	222	51.4	Chain A, Acidothermus Cellulolyticus Endocellulase E1 Catalytic Domain In Complex With A Cellotetraosegi|1827682|pdb|1ECE|B Chain B, Acidothermus Cellulolyticus Endocellulase E1 Catalytic Domain In Complex With A Cellotetraose
Contig5076	Cellulase	790	334	BAC57896	gi|51090374|dbj|BAC57896.2|	1.00E-77	258	51.9	cellulase celA [Polyplastron multivesiculatum]
Contig22970	Cellulase	800	379	BAC57896	gi|51090374|dbj|BAC57896.2|	2.00E-71	272	47.4	cellulase celA [Polyplastron multivesiculatum]
Contig22908	Cellulase	927	540	ZP_06142925	gi|268609198|ref|ZP_06142925.1|	9.00E-72	289	46.0	mannan endo-1,4-beta-mannosidase [Ruminococcus flavefaciens FD-1]
NODE_26010_length_577_cov_5.098787	Cellulase	613	236	ZP_06145755	gi|268612028|ref|ZP_06145755.1|	2.00E-66	210	59.0	cellulose 1,4-beta-cellobiosidase [Ruminococcus flavefaciens FD-1]
Contig519	Cellulase	2045	127	ZP_06720388	gi|294642514|ref|ZP_06720388.1|	5.00E-64	300	42.3	cellulase (glycosyl hydrolase family 5) [Ruminococcus albus 8]gi|291502786|gb|EFF15548.1| cellulase (glycosyl hydrolase family 5) [Ruminococcus albus 8]
Contig29835	Cellulase	504	582	YP_003490	gi|290959511|	1.00E	148	54.7	putative cellulase [Streptomyces scabiei

				693	ref|YP_00349 0693.1|	-41			87.22]gi|260649037|emb|CBG72151.1| putative secreted cellulase [Streptomyces scabiei 87.22]
Contig15712	Cellulase	1864	7153	ZP_020251 68	gi|154482720| ref|ZP_02025 168.1|	1.00E -175	502	58.4	hypothetical protein EUBVEN_00397 [Eubacterium ventriosum ATCC 27560]gi|149736496|gb|EDM52382.1| hypothetical protein EUBVEN_00397 [Eubacterium ventriosum ATCC 27560]
Contig5107	Cellulase	1307	463	ZP_020251 68	gi|154482720| ref|ZP_02025 168.1|	1.00E -145	409	58.7	hypothetical protein EUBVEN_00397 [Eubacterium ventriosum ATCC 27560]gi|149736496|gb|EDM52382.1| hypothetical protein EUBVEN_00397 [Eubacterium ventriosum ATCC 27560]
Contig15879	Cellulase	691	452	BAC57895	gi|28569974|d bj|BAC57895. 1|	4.00E -66	213	53.1	cellulase celA [Epidinium caudatum]
Contig30151	Cellulase	563	485	BAC57896	gi|51090374|d bj|BAC57896. 2|	7.00E -94	188	79.8	cellulase celA [Polyplastron multivesiculatum]
Contig3875	Cellulase	518	142	CAL91968	gi|218081332| emb|CAL919 68.1|	8.00E -60	170	62.4	cellulase [Epidinium ecaudatum]
Contig9951	Cellulase	846	405	XP_002679 377	gi|290993512| ref|XP_00267 9377.1|	2.00E -33	280	32.1	predicted protein [Naegleria gruberi]gi|284092993|gb|EFC46633.1| predicted protein [Naegleria gruberi]
Contig4151	Cellulase	719	276	XP_002679 377	gi|290993512| ref|XP_00267 9377.1|	3.00E -23	228	32.5	predicted protein [Naegleria gruberi]gi|284092993|gb|EFC46633.1| predicted protein [Naegleria gruberi]
Contig8299	Cellulase	501	220	XP_001430 218	gi|145488428| ref|XP_00143 0218.1|	1.00E -14	166	32.5	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124397314|emb|CAK62820.1| unnamed protein product [Paramecium tetraurelia]
Contig27963	Cellulase	511	180	ACU30843	gi|255710036| gb|ACU30843 .1|	8.00E -30	168	41.7	beta-mannanase [Paenibacillus sp. A1]
Contig21599	Cellulase	537	325	Q12647	gi|2494328|sp| Q12647.1|GU NB_NEOPA	2.00E -26	98	54.1	RecName: Full=Endoglucanase B; AltName: Full=Endo-1,4-beta-glucanase B; AltName: Full=Cellulase B; Flags: Precursorgi|467687|emb|CAA83238.1| endoglucanase B [Neocallimastix patriciarum]
NODE_18303	GH6	1445	737	ABY52798	gi|164375385|	1.00E	349	72.8	1,4-beta-D-glucan-cellobiohydrolase

_length_1409_cov_4.853087					gb|ABY52798.1|	-148			[Piromyces rhizinflatus]
Contig23192	GH6	1085	823	ACX32999	gi|260169862|gb|ACX32999.1|	1.00E-124	294	72.8	1,4-beta-glucanase [Piromyces sp. BTrP1]
Contig29702	GH6	676	518	ABY52799	gi|164375387|gb|ABY52799.1|	2.00E-95	214	76.2	1,4-beta-D-glucan-cellobiohydrolase [Piromyces rhizinflatus]
Contig1051	GH6	775	598	AAP33843	gi|32395719|gb|AAP33843.1|	2.00E-75	201	71.1	hybrid 1,4-beta-glucanase [synthetic construct]
Contig346	GH6	863	51	AAL01211	gi|15529294|gb|AAL01211.1|AF177204_1	2.00E-76	238	59.2	cellobiohydrolase II-like cellulase CelH [Orpinomyces sp. PC-2]
Contig31943	GH8	1253	901	YP_003248565	gi|261414882|ref|YP_003248565.1|	0	379	84.4	glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261371338|gb|ACX74083.1| glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig3115	GH8	904	182	YP_003248565	gi|261414882|ref|YP_003248565.1|	1.00E-109	214	82.2	glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261371338|gb|ACX74083.1| glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig925	GH9	1836	139	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	609	58.9	cellulase Cel9A precursor [Piromyces sp. E2]
Contig31316	GH9	1868	2473	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	590	61.2	cellulase Cel9A precursor [Piromyces sp. E2]
Contig21367	GH9	1881	2250	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	609	60.9	cellulase Cel9A precursor [Piromyces sp. E2]
Contig29001	GH9	1810	1149	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	604	67.2	cellulase Cel9A precursor [Piromyces sp. E2]

Contig2248*	GH9	1955	4945	AAM81967	gi|21929669|gb|AAM81967.1|AF459453_1	0	628	57.6	cellulase Cel9A precursor [Piromyces sp. E2]
Contig595	GH9	1813	1215	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	606	57.9	cellulase Cel9A precursor [Piromyces sp. E2]
Contig1790*	GH9	1984	4432	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	608	52.1	cellulase Cel9A precursor [Piromyces sp. E2]
Contig30085	GH9	1824	1505	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	607	84.7	cellulase Cel9A precursor [Piromyces sp. E2]
Contig17971	GH9	1900	1159	AAM81967	gi|21929669|gb|AAM81967.1|AF459453_1	0	624	53.7	cellulase Cel9A precursor [Piromyces sp. E2]
Contig29004*	GH9	1944	2673	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	608	54.6	cellulase Cel9A precursor [Piromyces sp. E2]
Contig28899	GH9	1861	4329	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	611	62.4	cellulase Cel9A precursor [Piromyces sp. E2]
Contig9942	GH9	1538	1128	XP_002603873	gi|260818001|ref|XP_002603873.1|	2.00E-73	464	38.1	hypothetical protein BRAFLDRAFT_119431 [Branchiostoma floridae]gi|229289197|gb|EEN59884.1| hypothetical protein BRAFLDRAFT_119431 [Branchiostoma floridae]
Contig29228	GH9	1164	1002	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	1.00E-153	388	66.0	cellulase Cel9A precursor [Piromyces sp. E2]
Contig1989	GH9	1667	44	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	542	57.0	cellulase Cel9A precursor [Piromyces sp. E2]

Contig4681	GH9	1091	600	AAP30753	gi|30315047|gb|AAP30753.1|	1.00E-147	354	68.6	cellulosomal glycoside hydrolase family 9 endoglucanase Cel9B [Piromyces sp. E2]
Contig26087	GH9	755	410	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	1.00E-131	251	86.5	cellulase Cel9A precursor [Piromyces sp. E2]
Contig265	GH9	807	118	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	1.00E-105	265	68.3	cellulase Cel9A precursor [Piromyces sp. E2]
Contig29128	GH9	560	549	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	2.00E-71	174	70.7	cellulase Cel9A precursor [Piromyces sp. E2]
Contig13687	GH9	971	473	CAL91976	gi|218081351|emb|CAL91976.1|	1.00E-165	319	86.8	cellulase [Epidinium ecaudatum]
Contig29291	GH9	889	1049	CAL91976	gi|218081351|emb|CAL91976.1|	1.00E-144	296	80.4	cellulase [Epidinium ecaudatum]
Contig4260	GH9	515	153	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	9.00E-59	146	71.2	cellulase Cel9A precursor [Piromyces sp. E2]
Contig25347	GH9	639	628	AAP30753	gi|30315047|gb|AAP30753.1|	2.00E-96	212	77.4	cellulosomal glycoside hydrolase family 9 endoglucanase Cel9B [Piromyces sp. E2]
Contig30906	GH9	923	553	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	1.00E-151	310	81.9	cellulase Cel9A precursor [Piromyces sp. E2]
Contig31652	GH9	642	209	AAP30753	gi|30315047|gb|AAP30753.1|	7.00E-38	195	45.6	cellulosomal glycoside hydrolase family 9 endoglucanase Cel9B [Piromyces sp. E2]
Contig9838	GH9	564	228	EFA81213	gi|281207029|gb|EFA81213.1|	5.00E-08	40	65.0	cellulase 270-6 [Polysphondylium pallidum PN500]
Contig806	GH10	1258	688	AAB30669	gi|560649|gb|AAB30669.1|	4.00E-84	287	54.4	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig405	GH10	1102	497	CAL91982	gi|218081365|	1.00E	329	77.8	xylanase [Eudiplodinium maggii]

					emb|CAL91982.1|	-153			
Contig306*	GH10	1000	979	CAL91981	gi|218081363|emb|CAL91981.1|	1.00E-160	326	81.6	xylanase [Epidinium ecaudatum]
Contig21293	GH10	1111	3483	CAL91982	gi|218081365|emb|CAL91982.1|	1.00E-125	282	74.8	xylanase [Eudiplodinium maggii]
Contig196	GH10	1246	782	AAB30669	gi|560649|gb|AAB30669.1|	1.00E-125	326	64.4	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig28861	GH10	1016	617	AAB30669	gi|560649|gb|AAB30669.1|	0	346	97.1	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig6672	GH10	816	277	AAB30669	gi|560649|gb|AAB30669.1|	4.00E-71	271	48.7	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig13683	GH10	866	454	AAB30669	gi|560649|gb|AAB30669.1|	3.00E-59	239	48.5	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig6829	GH10	825	342	AAB30669	gi|560649|gb|AAB30669.1|	5.00E-44	271	35.4	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig5473	GH10	651	209	CAL91979	gi|218081359|emb|CAL91979.1|	1.00E-118	216	95.8	xylanase [Epidinium ecaudatum]
Contig5361	GH10	807	460	AAB30669	gi|560649|gb|AAB30669.1|	4.00E-94	201	80.1	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig20523	GH10	745	545	AAB30669	gi|560649|gb|AAB30669.1|	2.00E-80	200	67.5	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig9398	GH10	547	139	YP_003506085	gi|291294687|ref|YP_003506085.1|	4.00E-33	163	42.3	Endo-1,4-beta-xylanase [Meiothermus ruber DSM 1279]gi|290469646|gb|ADD27065.1| Endo-1,4-beta-xylanase [Meiothermus ruber DSM 1279]
Contig9325	GH10	707	380	P26223	gi|139879|sp|P26223.1|XYNB_BUTFI	4.00E-30	216	36.1	RecName: Full=Endo-1,4-beta-xylanase B; Short=Xylanase B; AltName: Full=1,4-beta-D-xylan xylanohydrolase Bgi|48963|emb|CAA43712.1| beta-1,4-D-xylanase [Butyrivibrio fibrisolvens]
Contig4858	GH10	691	199	AAB30669	gi|560649|gb|AAB30669.1|	3.00E-24	127	38.6	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig10190	GH10	516	167	AAB30669	gi|560649|gb|AAB30669.1|	4.00E-44	116	69.8	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig3258*	GH11	805	4643	CAD56867	gi|38343952|emb|CAD56867.1|	1.00E-113	209	92.3	xylanase 11D [Polyplastron multivesiculatum]

Contig31636	GH11	2035	3284	YP_003248456	gi|261414773|ref|YP_003248456.1|	1.00E-106	465	43.9	Endo-1,4-beta-xylanase [Fibrobacter succinogenes subsp. succinogenes S85]gi|284018150|sp|P35811.2|XYNC_FIBSS RecName: Full=Endo-1,4-beta-xylanase C; Short=Xylanase C; AltName: Full=1,4-beta-D-xylan xylanohydrolase C; Flags: Precursorgi|261371229|gb|ACX73974.1| Endo-1,4-beta-xylanase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig16126*	GH11	615	150	CAD56867	gi|38343952|emb|CAD56867.1|	2.00E-87	201	74.1	xylanase 11D [Polyplastron multivesiculatum]
Contig3557	GH11	706	358	CAD56867	gi|38343952|emb|CAD56867.1|	2.00E-90	211	73.9	xylanase 11D [Polyplastron multivesiculatum]
Contig21286*	GH11	782	5803	CAD56867	gi|38343952|emb|CAD56867.1|	2.00E-86	213	70.9	xylanase 11D [Polyplastron multivesiculatum]
Contig28863	GH11	1148	44380	AAT99015	gi|51236716|gb|AAT99015.1|	0	370	86.8	xylanase [Neocallimastix frontalis]
Contig30226	GH11	919	1309	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	1.00E-103	210	81.9	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags: Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig13018	GH11	871	6499	ACL68347	gi|219964511|gb|ACL68347.1|	1.00E-101	224	79.0	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig1400	GH11	814	743	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	1.00E-102	221	75.1	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags: Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig12222*	GH11	1624	1671	AAG18439	gi|10505338|gb|AAG18439.1|	1.00E-104	223	78.9	xylanase [Piromyces communis]
Contig26517	GH11	684	396	CAA57820	gi|565626|emb|CAA57820.	2.00E-90	189	78.3	endoxylanase [Neocallimastix frontalis]

					1|			
Contig1997	GH11	1089	87	ABW04217	gi|157930095|gb|ABW04217.1|	1.00E-113	229	83.0	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig23170	GH11	1106	732	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	3.00E-75	159	80.5	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags: Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig5033	GH11	972	662	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	4.00E-34	146	48.6	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags: Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig8774	GH11	802	379	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	4.00E-54	140	72.9	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags: Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig4808	GH11	639	388	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	5.00E-45	111	71.2	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags: Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig29382	GH11	612	715	YP_003248456	gi|261414773|ref|YP_003248456.1|	1.00E-44	117	69.2	Endo-1,4-beta-xylanase [Fibrobacter succinogenes subsp. succinogenes S85]gi|284018150|sp|P35811.2|XYNC_FIBSS RecName: Full=Endo-1,4-beta-xylanase C; Short=Xylanase C; AltName: Full=1,4-beta-D-xylan xylanohydrolase C; Flags: Precursorgi|261371229|gb|ACX73974.1| Endo-1,4-beta-xylanase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig23899	Alpha-amylase	1319	1344	XP_001462315	gi|145553281|ref|XP_001462315.1|	1.00E-120	450	46.7	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124430154|emb|CAK94942.1| unnamed protein product [Paramecium tetraurelia]

Contig22533	Alpha-amylase	1325	1567	XP_640516	gi|66812674|ref|XP_640516.1|	9.00E-96	425	43.1	hypothetical protein DDB_G0281547 [Dictyostelium discoideum AX4]gi|60468532|gb|EAL66535.1| hypothetical protein DDB_G0281547 [Dictyostelium discoideum AX4]
Contig2108	Alpha-amylase	1065	294	XP_002114911	gi|196010093|ref|XP_002114911.1|	5.00E-98	359	48.7	hypothetical protein TRIADDRAFT_58902 [Trichoplax adhaerens]gi|190582294|gb|EDV22367.1| hypothetical protein TRIADDRAFT_58902 [Trichoplax adhaerens]
Contig30641	Alpha-amylase	1246	995	XP_001441176	gi|145510486|ref|XP_001441176.1|	1.00E-107	449	44.8	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124408415|emb|CAK73779.1| unnamed protein product [Paramecium tetraurelia]
Contig28650	Alpha-amylase	1004	747	XP_001441508	gi|145511161|ref|XP_001441508.1|	5.00E-88	350	47.4	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124408758|emb|CAK74111.1| unnamed protein product [Paramecium tetraurelia]
Contig22809	Alpha-amylase	1294	1251	XP_001462315	gi|145553281|ref|XP_001462315.1|	1.00E-112	443	45.1	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124430154|emb|CAK94942.1| unnamed protein product [Paramecium tetraurelia]
NODE_11739_length_1341_cov_10.972408	Alpha-amylase	1385	4489	XP_001441508	gi|145511161|ref|XP_001441508.1|	1.00E-110	449	44.1	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124408758|emb|CAK74111.1| unnamed protein product [Paramecium tetraurelia]
Contig310	Alpha-amylase	1417	259	XP_001462315	gi|145553281|ref|XP_001462315.1|	1.00E-105	440	43.9	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124430154|emb|CAK94942.1| unnamed protein product [Paramecium tetraurelia]
Contig6668	Alpha-amylase	1381	3445	XP_001462315	gi|145553281|ref|XP_001462315.1|	1.00E-106	438	43.4	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124430154|emb|CAK94942.1| unnamed protein product [Paramecium tetraurelia]
Contig5037	Alpha-amylase	1088	526	XP_001462315	gi|145553281|ref|XP_001462315.1|	1.00E-92	370	43.2	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124430154|emb|CAK94942.1| unnamed protein product [Paramecium tetraurelia]
Contig21244	Alpha-amylase	1326	3086	XP_001462315	gi|145553281|ref|XP_001462315.1|	1.00E-111	455	43.7	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124430154|emb|CAK94942.1| unnamed protein product [Paramecium

									tetraurelia]
Contig3803*	Alpha-amylase	1382	1860	XP_001462315	gi|145553281|ref|XP_001462315.1|	1.00E-105	449	43.2	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124430154|emb|CAK94942.1| unnamed protein product [Paramecium tetraurelia]
Contig3478	Alpha-amylase	1799	2970	XP_001462315	gi|145553281|ref|XP_001462315.1|	1.00E-101	421	44.7	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124430154|emb|CAK94942.1| unnamed protein product [Paramecium tetraurelia]
Contig29368	Alpha-amylase	746	2874	XP_001441176	gi|145510486|ref|XP_001441176.1|	1.00E-71	230	55.2	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124408415|emb|CAK73779.1| unnamed protein product [Paramecium tetraurelia]
Contig22555	Alpha-amylase	889	357	ZP_05758482	gi|260172070|ref|ZP_05758482.1|	2.00E-59	303	43.2	alpha amylase catalytic region [Bacteroides sp. D2]
Contig448	Alpha-amylase	625	108	XP_001441176	gi|145510486|ref|XP_001441176.1|	1.00E-69	212	56.6	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124408415|emb|CAK73779.1| unnamed protein product [Paramecium tetraurelia]
Contig6348	Alpha-amylase	630	441	XP_001460601	gi|145549844|ref|XP_001460601.1|	1.00E-66	208	52.9	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124428431|emb|CAK93204.1| unnamed protein product [Paramecium tetraurelia]
Contig30877	Alpha-amylase	705	1199	YP_001193561	gi|146298970|ref|YP_001193561.1|	1.00E-38	215	38.1	alpha amylase, catalytic region [Flavobacterium johnsoniae UW101]gi|146153388|gb|ABQ04242.1| Candidate alpha glycosidase; Glycoside hydrolase family 13 [Flavobacterium johnsoniae UW101]
Contig6120	Alpha-amylase	1267	814	ACD93218	gi|288915565|gb|ACD93218.3|	1.00E-108	386	49.0	alpha-amylase [Bacillus sp. KR-8104]
Contig30966	Alpha-amylase	549	469	XP_002114911	gi|196010093|ref|XP_002114911.1|	2.00E-56	182	52.2	hypothetical protein TRIADDRAFT_58902 [Trichoplax adhaerens]gi|190582294|gb|EDV22367.1| hypothetical protein TRIADDRAFT_58902 [Trichoplax adhaerens]
Contig17090	Alpha-amylase	588	139	CAI59813	gi|60417372|emb|CAI59813.1|	6.00E-59	169	62.7	alpha-glucosidase [Nyctotherus ovalis]

Contig7190	Alpha-amylase	1154	486	ACD93218	gi|288915565|gb|ACD93218.3|	1.00E-105	385	47.0	alpha-amylase [Bacillus sp. KR-8104]
Contig25421	Alpha-amylase	696	1511	ZP_05758482	gi|260172070|ref|ZP_05758482.1|	9.00E-40	235	40.0	alpha amylase catalytic region [Bacteroides sp. D2]
Contig21016	Alpha-amylase	520	132	XP_002114911	gi|196010093|ref|XP_002114911.1|	8.00E-45	151	53.6	hypothetical protein TRIADDRAFT_58902 [Trichoplax adhaerens]gi|190582294|gb|EDV22367.1| hypothetical protein TRIADDRAFT_58902 [Trichoplax adhaerens]
Contig4922	Alpha-amylase	1483	734	XP_001584271	gi|154422518|ref|XP_001584271.1|	1.00E-153	500	53.6	Alpha amylase, catalytic domain containing protein [Trichomonas vaginalis G3]gi|121918517|gb|EAY23285.1| Alpha amylase, catalytic domain containing protein [Trichomonas vaginalis G3]
Contig24372	Alpha-amylase	1918	3528	CAL92192	gi|218411097|emb|CAL92192.1|	0	494	74.7	amylase [Eudiplodinium maggii]
Contig4791	Alpha-amylase	763	1298	XP_640516	gi|66812674|ref|XP_640516.1|	1.00E-40	224	39.7	hypothetical protein DDB_G0281547 [Dictyostelium discoideum AX4]gi|60468532|gb|EAL66535.1| hypothetical protein DDB_G0281547 [Dictyostelium discoideum AX4]
Contig22216*	Alpha-amylase	1896	8564	CAL92192	gi|218411097|emb|CAL92192.1|	0	506	75.5	amylase [Eudiplodinium maggii]
Contig16777*	Alpha-amylase	1958	1352	CAL92192	gi|218411097|emb|CAL92192.1|	0	496	75.2	amylase [Eudiplodinium maggii]
Contig2327	Alpha-amylase	1466	115	CAL92191	gi|218411095|emb|CAL92191.1|	0	419	86.4	amylase [Eudiplodinium maggii]
Contig29562	Alpha-amylase	1523	2226	CAL92192	gi|218411097|emb|CAL92192.1|	0	500	76.0	amylase [Eudiplodinium maggii]
Contig21845	Alpha-amylase	1156	936	CAL92191	gi|218411095|emb|CAL92191.1|	1.00E-118	385	55.1	amylase [Eudiplodinium maggii]
Contig2471	Alpha-amylase	1857	149	XP_001584271	gi|154422518|ref|XP_00158	1.00E-173	619	48.9	Alpha amylase, catalytic domain containing protein [Trichomonas vaginalis

					4271.1|			G3]gi|121918517|gb|EAY23285.1| Alpha amylase, catalytic domain containing protein [Trichomonas vaginalis G3]
Contig19126*	Alpha-amylase	1992	4174	CAL92192	gi|218411097|emb|CAL92192.1|	0	504	82.7	amylase [Eudiplodinium maggii]
Contig7354	Alpha-amylase	1112	756	CAL92192	gi|218411097|emb|CAL92192.1|	1.00E-160	371	71.4	amylase [Eudiplodinium maggii]
Contig640	Alpha-amylase	883	1084	CAL92191	gi|218411095|emb|CAL92191.1|	1.00E-124	265	77.7	amylase [Eudiplodinium maggii]
Contig17739	Alpha-amylase	745	311	CAL92191	gi|218411095|emb|CAL92191.1|	1.00E-112	247	76.5	amylase [Eudiplodinium maggii]
Contig29090*	Alpha-amylase	1979	1857	CAL92192	gi|218411097|emb|CAL92192.1|	0	506	69.6	amylase [Eudiplodinium maggii]
Contig8432	Alpha-amylase	601	204	XP_001462315	gi|145553281|ref|XP_001462315.1|	6.00E-33	198	39.4	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124430154|emb|CAK94942.1| unnamed protein product [Paramecium tetraurelia]
Contig29102	Alpha-amylase	774	89	CAL92192	gi|218411097|emb|CAL92192.1|	1.00E-131	258	85.3	amylase [Eudiplodinium maggii]
Contig24480	Alpha-amylase	653	287	XP_001298742	gi|123382892|ref|XP_001298742.1|	9.00E-73	223	61.0	Alpha amylase, catalytic domain containing protein [Trichomonas vaginalis G3]gi|121879396|gb|EAX85812.1| Alpha amylase, catalytic domain containing protein [Trichomonas vaginalis G3]
Contig17242	Alpha-amylase	1023	2315	CAL92192	gi|218411097|emb|CAL92192.1|	1.00E-106	214	80.8	amylase [Eudiplodinium maggii]
Contig376*	GH16	804	208	YP_003250162	gi|261416479|ref|YP_003250162.1|	8.00E-79	233	58.4	Licheninase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372935|gb|ACX75680.1| Licheninase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig27493	GH16	878	19963	YP_003250162	gi|261416479|ref|YP_003250162.1|	9.00E-74	233	57.1	Licheninase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372935|gb|ACX75680.1|

									Licheninase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig22042	GH16	761	2340	YP_003250162	gi|261416479|ref|YP_003250162.1|	7.00E-84	233	61.4	Licheninase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372935|gb|ACX75680.1| Licheninase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig24068	GH16	744	202	YP_003250162	gi|261416479|ref|YP_003250162.1|	3.00E-74	233	59.2	Licheninase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372935|gb|ACX75680.1| Licheninase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig12909	GH16	802	240	YP_003250162	gi|261416479|ref|YP_003250162.1|	6.00E-55	178	57.9	Licheninase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372935|gb|ACX75680.1| Licheninase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig16779	GH16	797	558	AAC60453	gi|452882|gb|AAC60453.1|	4.00E-50	255	42.4	beta-1,3-glucanase [Bacillus circulans]gi|601878|dbj|BAA04469.1| beta-1,3-glucanase bglH [Bacillus circulans]
Contig4908	GH16	808	295	ZP_05512236	gi|256773773|ref|ZP_05512236.1|	4.00E-85	233	63.1	glycoside hydrolase family 16 [Streptomyces hygroscopicus ATCC 53653]
Contig30582	GH18	1709	2452	ZP_02027367	gi|154484919|ref|ZP_02027367.1|	1.00E-140	546	47.4	hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]gi|149733872|gb|EDM49991.1| hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]
Contig253*	GH18	1819	2051	ZP_02027367	gi|154484919|ref|ZP_02027367.1|	1.00E-158	561	52.4	hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]gi|149733872|gb|EDM49991.1| hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]
Contig385	GH18	1246	74	ZP_02027367	gi|154484919|ref|ZP_02027367.1|	1.00E-134	421	58.4	hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]gi|149733872|gb|EDM49991.1| hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]
Contig21608	GH18	1598	3847	ZP_02027367	gi|154484919|ref|ZP_02027367.1|	1.00E-130	540	47.6	hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]gi|149733872|gb|EDM49991.1|

									hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]
Contig21754	GH18	1313	734	ZP_02027367	gi|154484919|ref|ZP_02027367.1|	1.00E-131	426	56.3	hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]gi|149733872|gb|EDM49991.1| hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]
Contig20742	GH18	1217	565	ZP_02027367	gi|154484919|ref|ZP_02027367.1|	1.00E-110	406	50.2	hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]gi|149733872|gb|EDM49991.1| hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]
Contig22956	GH18	1236	722	ZP_04260281	gi|229131382|ref|ZP_04260281.1|	6.00E-44	302	35.4	Chitinase C [Bacillus cereus BDRD-ST196]gi|228652073|gb|EEL08011.1| Chitinase C [Bacillus cereus BDRD-ST196]
Contig5831	GH18	1008	371	CAC35202	gi|13508934|emb|CAC35202.1|	6.00E-43	292	37.7	endochitinase [Amanita muscaria]
Contig7444	GH18	1105	388	ZP_04850994	gi|253573651|ref|ZP_04850994.1|	1.00E-35	283	35.0	chitinase A1 [Paenibacillus sp. oral taxon 786 str. D14]gi|251847179|gb|EES75184.1| chitinase A1 [Paenibacillus sp. oral taxon 786 str. D14]
Contig23448	GH18	780	358	YP_001643260	gi|163938376|ref|YP_001643260.1|	6.00E-39	257	38.5	glycoside hydrolase family protein [Bacillus weihenstephanensis KBAB4]gi|163860573|gb|ABY41632.1| glycoside hydrolase family 18 [Bacillus weihenstephanensis KBAB4]
Contig5897	GH18	780	173	ZP_02027367	gi|154484919|ref|ZP_02027367.1|	1.00E-67	234	56.0	hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]gi|149733872|gb|EDM49991.1| hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]
Contig13273	GH18	1120	686	ZP_02027367	gi|154484919|ref|ZP_02027367.1|	4.00E-81	383	45.4	hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]gi|149733872|gb|EDM49991.1| hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]
Contig3673	GH18	642	176	YP_001643260	gi|163938376|ref|YP_001643260.1|	2.00E-30	212	39.2	glycoside hydrolase family protein [Bacillus weihenstephanensis KBAB4]gi|163860573|gb|ABY41632.1| glycoside hydrolase family 18 [Bacillus

									weihenstephanensis KBAB4]
Contig2109	GH18	981	1036	ZP_02027367	gi|154484919|ref|ZP_02027367.1|	1.00E-106	304	62.5	hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]gi|149733872|gb|EDM49991.1| hypothetical protein EUBVEN_02637 [Eubacterium ventriosum ATCC 27560]
Contig6553*	SLT	638	981	AAB61345	gi|2198832|gb|AAB61345.1|	4.00E-05	78	37.2	lysozyme [Anopheles darlingi]
Contig1850	Phage_lysozyme	837	1065	XP_001840847	gi|169868552|ref|XP_001840847.1|	9.00E-25	233	34.3	hypothetical protein CC1G_03076 [Coprinopsis cinerea okayama7#130]gi|116498005|gb|EAU80900.1| hypothetical protein CC1G_03076 [Coprinopsis cinerea okayama7#130]
Contig2228	Phage_lysozyme	965	556	ZP_04656141	gi|238912304|ref|ZP_04656141.1|	7.00E-17	147	35.4	hypothetical protein SentesTe_14381 [Salmonella enterica subsp. enterica serovar Tennessee str. CDC07-0191]
Contig2496*	GH25	758	926	ZP_06472267	gi|289640042|ref|ZP_06472267.1|	4.00E-57	192	53.1	glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]gi|289518501|gb|EFD38839.1| glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]
Contig29697	GH25	626	423	ZP_06718232	gi|294640234|ref|ZP_06718232.1|	9.00E-56	198	51.5	glycosyl hydrolase family 25 [Ruminococcus albus 8]gi|291504931|gb|EFF17569.1| glycosyl hydrolase family 25 [Ruminococcus albus 8]
Contig21944	GH25	844	516	ZP_06472267	gi|289640042|ref|ZP_06472267.1|	2.00E-55	237	49.4	glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]gi|289518501|gb|EFD38839.1| glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]
Contig22378	GH25	1512	969	ZP_06144424	gi|268610697|ref|ZP_06144424.1|	2.00E-58	204	56.4	Lyzozyme M1 (1,4-beta-N-acetylmuramidase) [Ruminococcus flavefaciens FD-1]
Contig1638	GH25	628	10	ZP_06472267	gi|289640042|ref|ZP_06472267.1|	2.00E-39	193	44.6	glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]gi|289518501|gb|EFD38839.1| glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]
Contig1351	GH25	618	170	ZP_06472267	gi|289640042|ref|ZP_06472267.1|	2.00E-44	194	48.5	glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]gi|289518501|gb|EFD38839.1| glycoside

									hydrolase family 25 [Ethanoligenens harbinense YUAN-3]
NODE_22897_length_576_cov_5.784722	GH25	620	396	ZP_06472267	gi|289640042|ref|ZP_06472267.1|	9.00E-45	196	47.4	glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]gi|289518501|gb|EFD38839.1| glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]
Contig699	GH25	1021	172	ZP_02092212	gi|160944986|ref|ZP_02092212.1|	4.00E-29	204	38.7	hypothetical protein FAEPRAM212_02501 [Faecalibacterium prausnitzii M21/2]gi|158442717|gb|EDP19722.1| hypothetical protein FAEPRAM212_02501 [Faecalibacterium prausnitzii M21/2]
Contig5156	GH25	627	243	ZP_06472267	gi|289640042|ref|ZP_06472267.1|	1.00E-36	189	43.9	glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]gi|289518501|gb|EFD38839.1| glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]
Contig24106	GH25	646	273	ZP_06472267	gi|289640042|ref|ZP_06472267.1|	1.00E-47	203	48.8	glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]gi|289518501|gb|EFD38839.1| glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]
Contig22298	GH25	699	545	ZP_06472267	gi|289640042|ref|ZP_06472267.1|	1.00E-46	203	47.3	glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]gi|289518501|gb|EFD38839.1| glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]
Contig3961*	GH25	1118	447	ZP_03463625	gi|218134821|ref|ZP_03463625.1|	4.00E-25	203	39.4	hypothetical protein BACPEC_02724 [Bacteroides pectinophilus ATCC 43243]gi|217990206|gb|EEC56217.1| hypothetical protein BACPEC_02724 [Bacteroides pectinophilus ATCC 43243]
Contig4078	GH25	628	291	ZP_06142561	gi|268608834|ref|ZP_06142561.1|	2.00E-36	200	41.5	Lyzozyme M1 (1,4-beta-N-acetylmuramidase) [Ruminococcus flavefaciens FD-1]
Contig27175	GH25	661	1598	ZP_06472267	gi|289640042|ref|ZP_06472267.1|	3.00E-39	191	45.0	glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]gi|289518501|gb|EFD38839.1| glycoside hydrolase family 25 [Ethanoligenens harbinense YUAN-3]
Contig4426*	GH25	889	735	XP_001749	gi|167536282|	3.00E	201	43.3	hypothetical protein [Monosiga brevicollis

				813	ref|XP_001749813.1|	-49			MX1]gi|163771740|gb|EDQ85402.1| predicted protein [Monosiga brevicollis MX1]
Contig12360	GH25	658	469	XP_001749813	gi|167536282|ref|XP_001749813.1|	2.00E-46	193	45.1	hypothetical protein [Monosiga brevicollis MX1]gi|163771740|gb|EDQ85402.1| predicted protein [Monosiga brevicollis MX1]
Contig31052	GH26	1689	2993	ZP_06144005	gi|268610278|ref|ZP_06144005.1|	1.00E-156	493	55.8	mannan endo-1,4-beta-mannosidase [Ruminococcus flavefaciens FD-1]
Contig23792	GH26	895	1346	YP_001559233	gi|160880265|ref|YP_001559233.1|	1.00E-106	297	60.3	mannan endo-1,4-beta-mannosidase [Clostridium phytofermentans ISDg]gi|160428931|gb|ABX42494.1| Mannan endo-1,4-beta-mannosidase, Cellulose 1,4-beta-cellobiosidase [Clostridium phytofermentans ISDg]
Contig1939	GH26	1054	44	ADE83298	gi|294473909|gb|ADE83298.1|	1.00E-131	333	64.3	carbohydrate esterase, family 7/glycosyl hydrolase, family 26 [Prevotella ruminicola 23]
Contig29593	GH26	511	4477	ZP_06144005	gi|268610278|ref|ZP_06144005.1|	7.00E-59	174	60.9	mannan endo-1,4-beta-mannosidase [Ruminococcus flavefaciens FD-1]
Contig1519	Melibiase	1066	68	XP_002513158	gi|255544191|ref|XP_002513158.1|	1.00E-122	355	58.0	alpha-galactosidase/alpha-n-acetylgalactosaminidase, putative [Ricinus communis]gi|223548169|gb|EEF49661.1| alpha-galactosidase/alpha-n-acetylgalactosaminidase, putative [Ricinus communis]
Contig31797	Melibiase	1026	509	XP_001014570	gi|118362944|ref|XP_001014570.1|	1.00E-104	298	59.4	alpha-galactosidase, putative [Tetrahymena thermophila]gi|89296464|gb|EAR94452.1| alpha-galactosidase, putative [Tetrahymena thermophila SB210]
Contig19135	Melibiase	619	274	AAA73964	gi|927577|gb|AAA73964.1|	3.00E-75	207	61.4	alpha-galactosidase [Phaseolus vulgaris]
Contig4109	Melibiase	686	466	XP_002325481	gi|224144974|ref|XP_002325481.1|	2.00E-81	228	61.4	predicted protein [Populus trichocarpa]gi|222862356|gb|EEE99862.1| predicted protein [Populus trichocarpa]
Contig21614	GH30	1217	4832	YP_001039401	gi|125975491|ref|YP_001039401.1|	1.00E-123	392	55.9	carbohydrate-binding family 6 protein [Clostridium thermocellum ATCC 27405]gi|256004221|ref|ZP_05429204.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM 2360]gi|281419476|ref|ZP_06250490.1|

									Carbohydrate binding family 6 [Clostridium thermocellum JW20]gi|125715716|gb|ABN54208.1| Carbohydrate binding family 6 [Clostridium thermocellum ATCC 27405]gi|255991811|gb|EEU01910.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM 2360]gi|281406882|gb|EFB37146.1| Carbohydrate binding family 6 [Clostridium thermocellum JW20]
Contig28893	GH30	1167	1561	CBL17903	gi|291544794|emb|CBL17903.1|	1.00E-120	366	59.6	O-Glycosyl hydrolase [Ruminococcus sp. 18P13]
Contig29246	GH30	1282	2110	CBL17903	gi|291544794|emb|CBL17903.1|	1.00E-127	388	58.2	O-Glycosyl hydrolase [Ruminococcus sp. 18P13]
Contig23726	GH31	815	422	ZP_03009674	gi|189460889|ref|ZP_03009674.1|	1.00E-104	267	64.8	hypothetical protein BACCOP_01536 [Bacteroides coprocola DSM 17136]gi|189432463|gb|EDV01448.1| hypothetical protein BACCOP_01536 [Bacteroides coprocola DSM 17136]
Contig1948	GH31	681	286	ZP_03852821	gi|227369311|ref|ZP_03852821.1|	2.00E-51	240	42.1	alpha-glycosidase [Chryseobacterium gleum ATCC 35910]gi|227107742|gb|EEI42742.1| alpha-glycosidase [Chryseobacterium gleum ATCC 35910]
Contig4038	GH31	975	464	ZP_02065660	gi|160884657|ref|ZP_02065660.1|	1.00E-121	307	62.5	hypothetical protein BACOVA_02646 [Bacteroides ovatus ATCC 8483]gi|156109692|gb|EDO11437.1| hypothetical protein BACOVA_02646 [Bacteroides ovatus ATCC 8483]
Contig21780	GH31	635	347	XP_002676411	gi|290987401|ref|XP_002676411.1|	1.00E-28	209	37.3	glycoside hydrolase [Naegleria gruberi]gi|284090013|gb|EFC43667.1| glycoside hydrolase [Naegleria gruberi]
Contig12512	GH32N	563	185	ZP_03644168	gi|224025802|ref|ZP_03644168.1|	3.00E-30	108	58.3	hypothetical protein BACCOPRO_02544 [Bacteroides coprophilus DSM 18228]gi|224019038|gb|EEF77036.1| hypothetical protein BACCOPRO_02544 [Bacteroides coprophilus DSM 18228]
Contig16703	GH32N	643	175	YP_003180274	gi|257785057|ref|YP_00318	1.00E-93	214	72.4	LPXTG-motif cell wall anchor domain protein [Atopobium parvulum DSM

					0274.1|			20469]gi|257473564|gb|ACV51683.1| LPXTG-motif cell wall anchor domain protein [Atopobium parvulum DSM 20469]
Contig1404	GH32N	610	518	ZP_06255644	gi|281424731|ref|ZP_06255644.1|	6.00E-38	157	49.7	glycoside Hydrolase Family 32 [Prevotella oris F0302]gi|281401101|gb|EFB31932.1| glycoside Hydrolase Family 32 [Prevotella oris F0302]
Contig29808	GH43	972	1181	ZP_03013720	gi|189464935|ref|ZP_03013720.1|	9.00E-91	307	55.4	hypothetical protein BACINT_01279 [Bacteroides intestinalis DSM 17393]gi|189437209|gb|EDV06194.1| hypothetical protein BACINT_01279 [Bacteroides intestinalis DSM 17393]
Contig29797	GH43	993	2602	ZP_03013720	gi|189464935|ref|ZP_03013720.1|	7.00E-89	307	52.8	hypothetical protein BACINT_01279 [Bacteroides intestinalis DSM 17393]gi|189437209|gb|EDV06194.1| hypothetical protein BACINT_01279 [Bacteroides intestinalis DSM 17393]
Contig4855	GH43	1011	334	YP_001038591	gi|125974681|ref|YP_001038591.1|	3.00E-92	292	55.5	carbohydrate-binding family 6 protein [Clostridium thermocellum ATCC 27405]gi|256004120|ref|ZP_05429104.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM 2360]gi|281418847|ref|ZP_06249866.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]gi|125714906|gb|ABN53398.1| Carbohydrate binding family 6 [Clostridium thermocellum ATCC 27405]gi|255991868|gb|EEU01966.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM 2360]gi|281407931|gb|EFB38190.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]
Contig2691	GH43	1633	677	ZP_06141916	gi|268608189|ref|ZP_06141916.1|	0	517	66.2	glycoside hydrolase family 43 [Ruminococcus flavefaciens FD-1]
Contig10710	GH43	687	506	YP_001038591	gi|125974681|ref|YP_001038591.1|	1.00E-79	219	66.2	carbohydrate-binding family 6 protein [Clostridium thermocellum ATCC 27405]gi|256004120|ref|ZP_05429104.1| Carbohydrate binding family 6 [Clostridium

									thermocellum DSM 2360]gi|281418847|ref|ZP_06249866.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]gi|125714906|gb|ABN53398.1| Carbohydrate binding family 6 [Clostridium thermocellum ATCC 27405]gi|255991868|gb|EEU01966.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM 2360]gi|281407931|gb|EFB38190.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]
Contig1312	GH43	565	42	ZP_02065668	gi|160884665|ref|ZP_02065668.1|	1.00E-64	194	57.7	hypothetical protein BACOVA_02654 [Bacteroides ovatus ATCC 8483]gi|156109700|gb|EDO11445.1| hypothetical protein BACOVA_02654 [Bacteroides ovatus ATCC 8483]
Contig31679	GH43	1635	2419	ACZ98594	gi|280977753|gb|ACZ98594.1|	1.00E-122	493	46.7	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig8161	GH43	1227	533	CBL17231	gi|291544122|emb|CBL17231.1|	1.00E-176	396	71.7	Beta-xylosidase [Ruminococcus sp. 18P13]
Contig27001*	GH43	1127	10074	YP_001981807	gi|192360510|ref|YP_001981807.1|	1.00E-115	311	60.1	beta-xylosidase/alpha-L-arabinfuranosidase, putative, gly43F [Cellvibrio japonicus Ueda107]gi|190686675|gb|ACE84353.1| beta-xylosidase/alpha-L-arabinfuranosidase, putative, gly43F [Cellvibrio japonicus Ueda107]
Contig4795	GH43	600	299	ADE82665	gi|294473276|gb|ADE82665.1|	5.00E-74	202	67.8	glycosyl hydrolase, family 43 [Prevotella ruminicola 23]
Contig22716	GH43	686	518	ACZ98594	gi|280977753|gb|ACZ98594.1|	7.00E-85	223	67.7	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig29325	GH43	1409	12451	YP_001981807	gi|192360510|ref|YP_001981807.1|	1.00E-113	315	59.0	beta-xylosidase/alpha-L-arabinfuranosidase, putative, gly43F [Cellvibrio japonicus Ueda107]gi|190686675|gb|ACE84353.1| beta-xylosidase/alpha-L-arabinfuranosidase, putative, gly43F [Cellvibrio japonicus

									Ueda107]
Contig21412	GH43	760	389	CBK75020	gi|291519799|emb|CBK75020.1|	4.00E-69	230	57.4	Cellobiohydrolase A (1,4-beta-cellobiosidase A) [Butyrivibrio fibrisolvens 16/4]
Contig22944	GH43	713	323	YP_001981807	gi|192360510|ref|YP_001981807.1|	6.00E-84	237	57.0	beta-xylosidase/alpha-L-arabinfuranosidase, putative, gly43F [Cellvibrio japonicus Ueda107]gi|190686675|gb|ACE84353.1| beta-xylosidase/alpha-L-arabinfuranosidase, putative, gly43F [Cellvibrio japonicus Ueda107]
Contig30609	GH43	574	3277	YP_001981807	gi|192360510|ref|YP_001981807.1|	1.00E-64	188	56.9	beta-xylosidase/alpha-L-arabinfuranosidase, putative, gly43F [Cellvibrio japonicus Ueda107]gi|190686675|gb|ACE84353.1| beta-xylosidase/alpha-L-arabinfuranosidase, putative, gly43F [Cellvibrio japonicus Ueda107]
Contig24702	GH43	824	352	YP_001981807	gi|192360510|ref|YP_001981807.1|	2.00E-95	274	55.5	beta-xylosidase/alpha-L-arabinfuranosidase, putative, gly43F [Cellvibrio japonicus Ueda107]gi|190686675|gb|ACE84353.1| beta-xylosidase/alpha-L-arabinfuranosidase, putative, gly43F [Cellvibrio japonicus Ueda107]
Contig21677	GH43	1649	989	ADE83588	gi|294474199|gb|ADE83588.1|	3.00E-69	520	36.3	glycosyl hydrolase, family 43 [Prevotella ruminicola 23]
Contig13452	GH43	668	495	CBK75020	gi|291519799|emb|CBK75020.1|	1.00E-71	184	66.8	Cellobiohydrolase A (1,4-beta-cellobiosidase A) [Butyrivibrio fibrisolvens 16/4]
Contig988	GH43	1584	417	ADE83588	gi|294474199|gb|ADE83588.1|	3.00E-75	523	36.7	glycosyl hydrolase, family 43 [Prevotella ruminicola 23]
Contig29906	GH43	1383	877	ADE83588	gi|294474199|gb|ADE83588.1|	4.00E-68	516	35.7	glycosyl hydrolase, family 43 [Prevotella ruminicola 23]
Contig13831	GH43	1430	690	ADE83588	gi|294474199|gb|ADE83588.1|	2.00E-69	363	42.7	glycosyl hydrolase, family 43 [Prevotella ruminicola 23]
Contig28788	GH43	791	458	ACZ98594	gi|280977753|gb|ACZ98594.1|	7.00E-64	241	53.1	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig13688	GH43	763	199	ACZ98594	gi|280977753|	1.00E	258	57.8	endo-1,4-beta-xylanase [Cellulosilyticum

					gb|ACZ98594.1|	-80			ruminicola]
Contig202	GH43	980	43	CBK73885	gi|291518664|emb|CBK73885.1|	1.00E-126	337	64.4	Beta-xylosidase [Butyrivibrio fibrisolvens 16/4]
Contig23129	GH43	572	603	CBL17682	gi|291544573|emb|CBL17682.1|	2.00E-52	170	60.6	Beta-1,4-xylanase [Ruminococcus sp. 18P13]
Contig3262	GH43	1176	631	ZP_05921816	gi|261207127|ref|ZP_05921816.1|	9.00E-92	388	48.5	predicted protein [Enterococcus faecium TC 6]gi|289565250|ref|ZP_06445701.1| predicted protein [Enterococcus faecium D344SRF]gi|294615113|ref|ZP_06694999.1| glycoside hydrolase, family 43 [Enterococcus faecium E1636]gi|260078755|gb|EEW66457.1| predicted protein [Enterococcus faecium TC 6]gi|289162906|gb|EFD10755.1| predicted protein [Enterococcus faecium D344SRF]gi|291592055|gb|EFF23678.1| glycoside hydrolase, family 43 [Enterococcus faecium E1636]
Contig869*	GH43	1222	317	CBK73885	gi|291518664|emb|CBK73885.1|	1.00E-140	392	62.5	Beta-xylosidase [Butyrivibrio fibrisolvens 16/4]
Contig25381	GH43	518	178	ACZ98594	gi|280977753|gb|ACZ98594.1|	7.00E-43	146	58.2	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig26483	GH43	695	353	CBK73780	gi|291518559|emb|CBK73780.1|	1.00E-105	221	80.5	Predicted beta-xylosidase [Butyrivibrio fibrisolvens 16/4]
Contig11742	GH43	507	159	ZP_06142338	gi|268608611|ref|ZP_06142338.1|	7.00E-63	168	69.0	Alpha-N-arabinofuranosidase [Ruminococcus flavefaciens FD-1]
Contig14307	GH43	640	166	ZP_05591660	gi|257413461|ref|ZP_05591660.1|	3.00E-70	173	68.8	xylosidase/arabinosidase [Roseburia intestinalis L1-82]gi|257203448|gb|EEV01733.1| xylosidase/arabinosidase [Roseburia intestinalis L1-82]
Contig4054	GH43	1207	538	ZP_04743177	gi|240144576|ref|ZP_04743177.1|	1.00E-155	189	72.0	xylosidase/arabinosidase [Roseburia intestinalis L1-82]gi|257203391|gb|EEV01676.1|

									xylosidase/arabinosidase [Roseburia intestinalis L1-82]
Contig6586	GH43	783	409	CBL17103	gi|291543994|emb|CBL17103.1|	1.00E-109	260	69.2	Glycosyl hydrolases family 43./Dockerin type I repeat. [Ruminococcus sp. 18P13]
Contig1414	GH43	1051	455	ZP_05921816	gi|261207127|ref|ZP_05921816.1|	7.00E-98	348	52.3	predicted protein [Enterococcus faecium TC 6]gi|289565250|ref|ZP_06445701.1| predicted protein [Enterococcus faecium D344SRF]gi|294615113|ref|ZP_06694999.1| glycoside hydrolase, family 43 [Enterococcus faecium E1636]gi|260078755|gb|EEW66457.1| predicted protein [Enterococcus faecium TC 6]gi|289162906|gb|EFD10755.1| predicted protein [Enterococcus faecium D344SRF]gi|291592055|gb|EFF23678.1| glycoside hydrolase, family 43 [Enterococcus faecium E1636]
Contig6617	GH43	1763	846	YP_003486881	gi|290955699|ref|YP_003486881.1|	6.00E-51	307	38.4	putative hydrolase [Streptomyces scabiei 87.22]gi|260645225|emb|CBG68311.1| putative secreted hydrolase [Streptomyces scabiei 87.22]
Contig28666	GH45	1199	1728	BAC53956	gi|27530542|dbj|BAC53956.1|	2.00E-74	200	62.0	endo-beta-1,4-D-glucanase [Rhizopus oryzae]
NODE_15203_length_752_cov_3.756649	GH45	788	280	BAC53956	gi|27530542|dbj|BAC53956.1|	7.00E-80	205	63.9	endo-beta-1,4-D-glucanase [Rhizopus oryzae]
Contig5478	GH45	651	173	CAB92325	gi|8052314|emb|CAB92325.1|	1.00E-76	220	60.0	endoglucanase 45A [Piromyces equi]
Contig31867	GH45	561	341	BAC53956	gi|27530542|dbj|BAC53956.1|	3.00E-66	190	58.9	endo-beta-1,4-D-glucanase [Rhizopus oryzae]
Contig29608	GH45	1062	513	ABU49185	gi|158138919|gb|ABU49185.2|	4.00E-55	152	61.2	endoglucanase [Syncephalastrum racemosum]
NODE_50541_length_652_cov_1.337423	GH45	688	141	BAC53987	gi|27530615|dbj|BAC53987.1|	7.00E-42	109	66.1	endo-glucanase RCE2 [Rhizopus oryzae]
Contig22615*	GH45	775	530	YP_003249	gi|261415823|	5.00E	229	56.8	Cellulase [Fibrobacter succinogenes subsp.

				506	ref|YP_003249506.1|	-67			succinogenes S85]gi|261372279|gb|ACX75024.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig16557	GH45	577	247	BAA98045	gi|8926983|dbj|BAA98045.1|	1.00E-36	127	58.3	family 45 cellulase homologue [Reticulitermes speratus hindgut protist 130484]
NODE_9170_length_673_cov_10.176820	GH45	717	1764	YP_003249506	gi|261415823|ref|YP_003249506.1|	1.00E-76	232	62.1	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372279|gb|ACX75024.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig28909	GH45	713	1341	YP_003249506	gi|261415823|ref|YP_003249506.1|	1.00E-69	225	57.8	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372279|gb|ACX75024.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig30798	GH45	523	10896	YP_003249506	gi|261415823|ref|YP_003249506.1|	3.00E-48	151	63.6	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372279|gb|ACX75024.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig28966	GH45	784	1076	YP_003249506	gi|261415823|ref|YP_003249506.1|	6.00E-66	231	57.1	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372279|gb|ACX75024.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig29058	GH45	765	2269	YP_003249505	gi|261415822|ref|YP_003249505.1|	3.00E-60	232	50.4	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372278|gb|ACX75023.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig914	GH45	719	100	YP_003249506	gi|261415823|ref|YP_003249506.1|	2.00E-67	236	57.2	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372279|gb|ACX75024.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig21206*	GH45	925	15508	YP_003249505	gi|261415822|ref|YP_003249505.1|	1.00E-68	248	52.4	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372278|gb|ACX75023.1| Cellulase [Fibrobacter succinogenes subsp.

									succinogenes S85]
Contig21772	GH45	836	607	YP_003249505	gi|261415822|ref|YP_003249505.1|	6.00E-64	249	49.8	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372278|gb|ACX75023.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig24962	GH45	830	8742	YP_003249505	gi|261415822|ref|YP_003249505.1|	1.00E-62	247	50.2	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372278|gb|ACX75023.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig916	GH45	782	48	YP_003249505	gi|261415822|ref|YP_003249505.1|	2.00E-58	242	47.1	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372278|gb|ACX75023.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig28726	GH45	711	1452	YP_003249506	gi|261415823|ref|YP_003249506.1|	6.00E-48	182	53.8	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372279|gb|ACX75024.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig1946	GH45	546	240	YP_003249506	gi|261415823|ref|YP_003249506.1|	4.00E-37	132	56.8	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372279|gb|ACX75024.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig22616	GH45	840	492	YP_003249505	gi|261415822|ref|YP_003249505.1|	1.00E-44	246	41.1	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372278|gb|ACX75023.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig21346	GH45	725	5791	YP_003249506	gi|261415823|ref|YP_003249506.1|	7.00E-46	200	51.5	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372279|gb|ACX75024.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig764	GH45	630	188	YP_003249505	gi|261415822|ref|YP_003249505.1|	1.00E-54	116	52.6	Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372278|gb|ACX75023.1| Cellulase [Fibrobacter succinogenes subsp. succinogenes S85]

Contig30005	GH48	1203	11562	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	0	372	83.1	cellulase Cel48A precursor [Piromyces sp. E2]
Contig753	GH48	1134	225	AAN76735	gi|25990957|gb|AAN76735.1|AF449413_1	0	374	81.8	cellulase Cel48A precursor [Piromyces equi]
Contig28684	GH48	898	33408	AAN76735	gi|25990957|gb|AAN76735.1|AF449413_1	1.00E-154	302	81.1	cellulase Cel48A precursor [Piromyces equi]
Contig2716	GH48	773	418	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	1.00E-120	229	86.0	cellulase Cel48A precursor [Piromyces sp. E2]
Contig213*	GH48	671	7160	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	2.00E-98	214	76.2	cellulase Cel48A precursor [Piromyces sp. E2]
Contig29348	GH48	550	4168	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	2.00E-75	169	78.1	cellulase Cel48A precursor [Piromyces sp. E2]
Contig28870	GH48	652	396	AAN76735	gi|25990957|gb|AAN76735.1|AF449413_1	2.00E-75	190	68.9	cellulase Cel48A precursor [Piromyces equi]
Contig30900	GH48	586	3032	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	3.00E-73	168	76.8	cellulase Cel48A precursor [Piromyces sp. E2]
Contig31226	GH48	549	436	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	2.00E-80	182	74.2	cellulase Cel48A precursor [Piromyces sp. E2]
Contig264*	GH48	559	7121	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	3.00E-64	124	81.5	cellulase Cel48A precursor [Piromyces sp. E2]
Contig28942	GH67M	1745	2054	YP_003096	gi|255535865|	1.00E	533	57.0	Alpha-glucuronidase [Flavobacteriaceae

				236	ref|YP_00309 6236.1|	-179			bacterium 3519-10]gi|255342061|gb|ACU08174.1| Alpha-glucuronidase [Flavobacteriaceae bacterium 3519-10]
Contig9529	Glucosa minidase	831	250	CBK90584	gi|291524997| emb|CBK905 84.1|	6.00E -37	131	61.8	Muramidase (flagellum-specific) [Eubacterium rectale A1-86 (DSM 17629)]
Contig3078	Glucosa minidase	601	13803	CBK90584	gi|291524997| emb|CBK905 84.1|	1.00E -26	80	72.5	Muramidase (flagellum-specific) [Eubacterium rectale A1-86 (DSM 17629)]
Contig21748	GH76	1093	2902	ZP_036320 62	gi|223940201| ref|ZP_03632 062.1|	9.00E -53	312	39.7	glycoside hydrolase family 76 [bacterium Ellin514]gi|223891146|gb|EEF57646.1| glycoside hydrolase family 76 [bacterium Ellin514]
Contig1443	GH77	1591	2031	ZP_059004 31	gi|260889168| ref|ZP_05900 431.1|	1.00E -159	495	53.1	4-alpha-glucanotransferase [Leptotrichia hofstadii F0254]gi|260861228|gb|EEX75728.1| 4-alpha-glucanotransferase [Leptotrichia hofstadii F0254]
Contig30043*	GH77	1574	6379	ZP_059004 31	gi|260889168| ref|ZP_05900 431.1|	1.00E -158	495	52.7	4-alpha-glucanotransferase [Leptotrichia hofstadii F0254]gi|260861228|gb|EEX75728.1| 4-alpha-glucanotransferase [Leptotrichia hofstadii F0254]
Contig21437	GH77	1010	662	ZP_048620 75	gi|253681277| ref|ZP_04862 075.1|	1.00E -103	329	55.0	4-alpha-glucanotransferase [Clostridium botulinum D str. 1873]gi|253562515|gb|EES91966.1| 4-alpha-glucanotransferase [Clostridium botulinum D str. 1873]
Contig21520	GH77	1129	1200	ZP_048620 75	gi|253681277| ref|ZP_04862 075.1|	3.00E -91	352	47.2	4-alpha-glucanotransferase [Clostridium botulinum D str. 1873]gi|253562515|gb|EES91966.1| 4-alpha-glucanotransferase [Clostridium botulinum D str. 1873]
Contig23542	GH77	919	376	ZP_026216 23	gi|168186988| ref|ZP_02621 623.1|	2.00E -93	308	53.9	4-alpha-glucanotransferase [Clostridium botulinum C str. Eklund]gi|169295028|gb|EDS77161.1| 4-alpha-glucanotransferase [Clostridium botulinum C str. Eklund]
Contig30017	GH77	978	955	CBL39922	gi|291561123| emb|CBL3992	1.00E -89	316	49.7	4-alpha-glucanotransferase [Clostridiales sp. SS3/4]

					2.1|			
Contig22344	GH77	907	620	ZP_04862075	gi|253681277|ref|ZP_04862075.1|	3.00E-92	289	56.1	4-alpha-glucanotransferase [Clostridium botulinum D str. 1873]gi|253562515|gb|EES91966.1| 4-alpha-glucanotransferase [Clostridium botulinum D str. 1873]
Contig9224	GH77	595	334	ZP_03706825	gi|225017633|ref|ZP_03706825.1|	2.00E-52	170	57.1	hypothetical protein CLOSTMETH_01562 [Clostridium methylpentosum DSM 5476]gi|224949598|gb|EEG30807.1| hypothetical protein CLOSTMETH_01562 [Clostridium methylpentosum DSM 5476]
Contig2949	NAGidase	1209	1199	ZP_05403793	gi|260881178|ref|ZP_05403793.2|	8.00E-93	380	46.8	O-GlcNAcase NagJ [Mitsuokella multacida DSM 20544]gi|260849716|gb|EEX69723.1| O-GlcNAcase NagJ [Mitsuokella multacida DSM 20544]
Contig5988	GH88	620	285	YP_001197274	gi|146302683|ref|YP_001197274.1|	1.00E-61	206	53.4	N-acylglucosamine 2-epimerase [Flavobacterium johnsoniae UW101]gi|146157101|gb|ABQ07955.1| N-acylglucosamine 2-epimerase [Flavobacterium johnsoniae UW101]
Contig8574	NAGLU	1069	579	XP_001638539	gi|156399499|ref|XP_001638539.1|	1.00E-103	333	51.4	predicted protein [Nematostella vectensis]gi|156225660|gb|EDO46476.1| predicted protein [Nematostella vectensis]
Contig2412	NAGLU	788	178	XP_001638539	gi|156399499|ref|XP_001638539.1|	3.00E-66	251	47.8	predicted protein [Nematostella vectensis]gi|156225660|gb|EDO46476.1| predicted protein [Nematostella vectensis]
Contig18975	NAGLU	723	569	XP_001638539	gi|156399499|ref|XP_001638539.1|	9.00E-70	241	51.5	predicted protein [Nematostella vectensis]gi|156225660|gb|EDO46476.1| predicted protein [Nematostella vectensis]
Contig7502	NAGLU	527	187	XP_002156234	gi|221122271|ref|XP_002156234.1|	4.00E-26	125	44.0	PREDICTED: similar to predicted protein, partial [Hydra magnipapillata]
Contig7175	NAGLU	534	224	YP_003122794	gi|256422141|ref|YP_003122794.1|	2.00E-35	153	42.5	Alpha-N-acetylglucosaminidase [Chitinophaga pinensis DSM 2588]gi|256037049|gb|ACU60593.1| Alpha-N-acetylglucosaminidase [Chitinophaga pinensis DSM 2588]
Contig8333	NAGLU	525	228	XP_002156234	gi|221122271|ref|XP_002156234.1|	5.00E-20	143	38.5	PREDICTED: similar to predicted protein, partial [Hydra magnipapillata]
Contig15910	GH97	1464	671	ADE82389	gi|294473000|	0	472	72.5	alpha-glucosidase family protein [Prevotella

					gb|ADE82389.1|			ruminicola 23]
Contig4297	GH97	853	207	ADE82389	gi|294473000|gb|ADE82389.1|	1.00E-116	282	68.4	alpha-glucosidase family protein [Prevotella ruminicola 23]
Contig2772	CBM1	2512	2253	XP_001400266	gi|145233787|ref|XP_001400266.1|	3.00E-66	613	30.7	hypothetical protein An02g11390 [Aspergillus niger]gi|134057200|emb|CAK44468.1| unnamed protein product [Aspergillus niger]
Contig30158	CBM1	936	556	EDP52952	gi|159127837|gb|EDP52952.1|	1.00E-14	143	33.6	extracellular serine-rich protein, putative [Aspergillus fumigatus A1163]
Contig12942	CBM1	567	407	AAB30669	gi|560649|gb|AAB30669.1|	2.00E-26	174	47.7	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig29986	CBM1	854	13133	AAB30669	gi|560649|gb|AAB30669.1|	6.00E-33	270	37.8	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig30586	CBM1	588	373	AAF14365	gi|6502585|gb|AAF14365.1|AF123252_1	3.00E-13	40	72.5	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig1663	CBM1	986	189	AAB30669	gi|560649|gb|AAB30669.1|	2.00E-50	330	41.5	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig2703	CBM1	1150	465	XP_002565826	gi|255950118|ref|XP_002565826.1|	0.68	43	46.5	Pc22g19230 [Penicillium chrysogenum Wisconsin 54-1255]gi|211592843|emb|CAP99211.1| Pc22g19230 [Penicillium chrysogenum Wisconsin 54-1255]
Contig27058	CBM1	539	313	XP_001618689	gi|156327341|ref|XP_001618689.1|	5.00E-15	122	44.3	hypothetical protein NEMVEDRAFT_v1g5061 [Nematostella vectensis]gi|156199910|gb|EDO26589.1| predicted protein [Nematostella vectensis]
Contig22877	CBM1	857	648	CBI51366	gi|289622188|emb|CBI51366.1|	1.00E-07	37	56.8	unnamed protein product [Sordaria macrospora]
Contig23143	CBM1	558	458	AAA34208	gi|506848|gb|AAA34208.1|	0.004	25	60.0	beta-mannase [Hypocrea jecorina]
Contig3846	CBM1	1367	566	AAK20910	gi|13446353|gb|AAK20910.1|	1.00E-06	313	19.2	non-catalytic protein 1 [Piromyces equi]
Contig25213	CBM4_9	586	204	CAL91979	gi|218081359|emb|CAL91979.1|	1.00E-103	194	91.8	xylanase [Epidinium ecaudatum]

Contig1687	CBM4_9	572	41	CAL91978	gi|218081356|emb|CAL91978.1|	2.00E-76	144	95.8	xylanase [Polyplastron multivesiculatum]
Contig5630	CBM6	586	146	ACZ98594	gi|280977753|gb|ACZ98594.1|	2.00E-40	195	41.5	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig29224	CBM6	830	483	ACZ98594	gi|280977753|gb|ACZ98594.1|	4.00E-47	228	42.1	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig8109	CBM6	590	174	ADE82665	gi|294473276|gb|ADE82665.1|	2.00E-40	162	56.2	glycosyl hydrolase, family 43 [Prevotella ruminicola 23]
Contig22052	CBM6	824	287	ACZ98594	gi|280977753|gb|ACZ98594.1|	2.00E-56	278	42.8	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig21969	CBM10	659	248	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	1.00E-25	79	59.5	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig22570	CBM10	554	671	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	1.00E-47	132	66.7	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig31340	CBM10	571	535	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	7.00E-48	132	67.4	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig21298	CBM10	522	175	AAK20910	gi|13446353|gb|AAK20910.1|	3.00E-32	135	47.4	non-catalytic protein 1 [Piromyces equi]
Contig29014	CBM10	1048	1072	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	2.00E-26	132	45.5	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig6731*	CBM10	557	215	CAB92325	gi|8052314|emb|CAB92325.1|	2.00E-31	139	44.6	endoglucanase 45A [Piromyces equi]
Contig30127	CBM10	643	401	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	7.00E-62	201	54.2	cellulase Cel9A precursor [Piromyces sp. E2]

Contig23396	CBM10	619	193	AAO41704	gi|28557461|gb|AAO41704.1|	6.00E-89	201	73.6	beta-glucosidase precursor [Piromyces sp. E2]
NODE_4_length_806_cov_7.171216	CBM10	842	533	P55296	gi|1708917|sp|P55296.1|MANA_PIRSP	4.00E-28	123	50.4	RecName: Full=Mannan endo-1,4-beta-mannosidase A; AltName: Full=Beta-mannanase A; AltName: Full=1,4-beta-D-mannan mannanohydrolase A; Flags: Precursorgi|1197368|emb|CAA62968.1| mannan endo-1,4-beta-mannosidase; mannanase A [Piromyces sp.]
Contig92	CBM10	605	43	AAR97891	gi|40950523|gb|AAR97891.1|	2.00E-41	219	44.3	cellulosomal serpin precursor [Piromyces sp. E2]
Contig1167*	CBM10	844	1037	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	6.00E-27	144	40.3	cellulase Cel9A precursor [Piromyces sp. E2]
NODE_51456_length_483_cov_2.997930	CBM10	519	161	AAK20910	gi|13446353|gb|AAK20910.1|	1.00E-44	143	55.9	non-catalytic protein 1 [Piromyces equi]
Contig29345	CBM10	537	370	ACL68347	gi|219964511|gb|ACL68347.1|	2.00E-30	91	62.6	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig11735	CBM10	554	146	ACL68347	gi|219964511|gb|ACL68347.1|	2.00E-44	143	54.5	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig5673*	CBM10	2133	1569	ZP_06145332	gi|268611605|ref|ZP_06145332.1|	0	557	55.7	hypothetical protein RflaF_19133 [Ruminococcus flavefaciens FD-1]
Contig3456	CBM10	1100	394	ZP_06145404	gi|268611677|ref|ZP_06145404.1|	9.00E-38	183	46.4	cellulosome enzyme, dockerin type I [Ruminococcus flavefaciens FD-1]
Contig29963	CBM10	864	805	CAB92325	gi|8052314|emb|CAB92325.1|	3.00E-28	177	38.4	endoglucanase 45A [Piromyces equi]
Contig4815	CBM10	549	259	AAL01214	gi|15529300|gb|AAL01214.1|AF177207_1	2.00E-51	183	50.8	endo-glucanase CelJ [Orpinomyces sp. PC-2]
Contig8213	CBM10	608	325	AAK20910	gi|13446353|gb|AAK20910.	7.00E-29	131	46.6	non-catalytic protein 1 [Piromyces equi]

					1|			
Contig644*	CBM10	635	449	CAB92325	gi|8052314|emb|CAB92325.1|	1.00E-33	141	47.5	endoglucanase 45A [Piromyces equi]
NODE_4349_length_668_cov_6.729042	CBM10	704	336	CAB92325	gi|8052314|emb|CAB92325.1|	2.00E-28	145	42.1	endoglucanase 45A [Piromyces equi]
Contig3770	CBM10	3281	1617	ABY52795	gi|164375379|gb|ABY52795.1|	3.00E-20	105	44.8	endo-1,4-beta-xylanase [Piromyces communis]
Contig15680	CBM10	650	305	AAK20910	gi|13446353|gb|AAK20910.1|	2.00E-26	168	35.1	non-catalytic protein 1 [Piromyces equi]
NODE_26024_length_595_cov_7.152941	CBM10	631	381	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	9.00E-75	213	59.6	cellulase Cel9A precursor [Piromyces sp. E2]
Contig20211	CBM10	665	264	CAB92325	gi|8052314|emb|CAB92325.1|	9.00E-29	147	42.2	endoglucanase 45A [Piromyces equi]
Contig23474	CBM10	542	239	AAL01214	gi|15529300|gb|AAL01214.1|AF177207_1	5.00E-55	181	56.9	endo-glucanase CelJ [Orpinomyces sp. PC-2]
Contig22135	CBM10	520	175	CAB92325	gi|8052314|emb|CAB92325.1|	4.00E-32	149	41.6	endoglucanase 45A [Piromyces equi]
Contig9042	CBM10	679	191	P55296	gi|1708917|sp|P55296.1|MANA_PIRSP	1.00E-33	131	53.4	RecName: Full=Mannan endo-1,4-beta-mannosidase A; AltName: Full=Beta-mannanase A; AltName: Full=1,4-beta-D-mannan mannanohydrolase A; Flags: Precursorgi|1197368|emb|CAA62968.1| mannan endo-1,4-beta-mannosidase; mannanase A [Piromyces sp.]
Contig6361	CBM10	543	219	AAL01214	gi|15529300|gb|AAL01214.1|AF177207_1	4.00E-27	110	51.8	endo-glucanase CelJ [Orpinomyces sp. PC-2]
Contig31078	CBM10	1075	551	AAO41704	gi|28557461|gb|AAO41704.1|	1.00E-149	345	71.9	beta-glucosidase precursor [Piromyces sp. E2]

Contig26995	CBM10	645	270	CAB92326	gi|8052316|emb|CAB92326.1|	5.00E-23	85	56.5	endoglucanase 5A [Piromyces equi]
Contig750	CBM10	1467	121	ZP_06145332	gi|268611605|ref|ZP_06145332.1|	1.00E-111	302	61.3	hypothetical protein RflaF_19133 [Ruminococcus flavefaciens FD-1]
Contig2191	CBM10	1573	192	CAB92325	gi|8052314|emb|CAB92325.1|	1.00E-27	155	40.6	endoglucanase 45A [Piromyces equi]
Contig13568	CBM10	621	199	ACL68347	gi|219964511|gb|ACL68347.1|	9.00E-53	202	51.5	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig3684	CBM10	608	263	P55296	gi|1708917|sp|P55296.1|MANA_PIRSP	5.00E-30	123	49.6	RecName: Full=Mannan endo-1,4-beta-mannosidase A; AltName: Full=Beta-mannanase A; AltName: Full=1,4-beta-D-mannan mannanohydrolase A; Flags: Precursorgi|1197368|emb|CAA62968.1| mannan endo-1,4-beta-mannosidase; mannanase A [Piromyces sp.]
Contig6963	CBM10	525	211	AAL01213	gi|15529298|gb|AAL01213.1|AF177206_1	5.00E-26	95	52.6	mannanase ManA [Orpinomyces sp. PC-2]
Contig7410	CBM10	802	514	YP_003250020	gi|261416337|ref|YP_003250020.1|	4.00E-45	139	64.7	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig21510	CBM10	522	1608	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	1.00E-54	135	70.4	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig5819	CBM10	757	565	Q12647	gi|2494328|sp|Q12647.1|GUNB_NEOPA	9.00E-29	82	63.4	RecName: Full=Endoglucanase B; AltName: Full=Endo-1,4-beta-glucanase B; AltName: Full=Cellulase B; Flags: Precursorgi|467687|emb|CAA83238.1| endoglucanase B [Neocallimastix patriciarum]
Contig29862	CBM10	526	903	AAL01211	gi|15529294|gb|AAL01211.1|AF177204_1	4.00E-56	176	63.1	cellobiohydrolase II-like cellulase CelH [Orpinomyces sp. PC-2]

Contig13907	CBM10	1148	930	YP_003250020	gi|261416337|ref|YP_003250020.1|	1.00E-109	269	68.8	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig31197	CBM10	596	197	Q9Y871	gi|23821548|sp|Q9Y871.1|FAEB_PIREQ	1.00E-37	83	91.6	RecName: Full=Feruloyl esterase B; AltName: Full=Ferulic acid esterase B; AltName: Full=Esterase A; Short=EstA; AltName: Full=Cinnamoyl ester hydrolase; Flags: Precursorgi|5566342|gb|AAD45376.1|AF164516_1 cinnamoyl ester hydrolase EstA [Piromyces equi]
NODE_4207_length_571_cov_10.173380	CBM10	623	1016	AAP30750	gi|30315041|gb|AAP30750.1|	3.00E-29	88	62.5	cellobiohydrolase Cel6C [Piromyces sp. E2]
Contig3862	CBM10	534	6927	AAL01211	gi|15529294|gb|AAL01211.1|AF177204_1	3.00E-69	176	71.0	cellobiohydrolase II-like cellulase CelH [Orpinomyces sp. PC-2]
Contig29489	CBM10	1181	624	CAB92325	gi|8052314|emb|CAB92325.1|	2.00E-25	149	36.9	endoglucanase 45A [Piromyces equi]
Contig31049	CBM10	730	431	Q9Y871	gi|23821548|sp|Q9Y871.1|FAEB_PIREQ	8.00E-64	127	92.9	RecName: Full=Feruloyl esterase B; AltName: Full=Ferulic acid esterase B; AltName: Full=Esterase A; Short=EstA; AltName: Full=Cinnamoyl ester hydrolase; Flags: Precursorgi|5566342|gb|AAD45376.1|AF164516_1 cinnamoyl ester hydrolase EstA [Piromyces equi]
Contig23411	CBM10	564	324	AAL92497	gi|29465670|gb|AAL92497.1|	5.00E-49	169	56.2	exoglucanase Cel6A [Piromyces sp. E2]
Contig6783*	CBM10	590	396	AAL01211	gi|15529294|gb|AAL01211.1|AF177204_1	1.00E-17	84	47.6	cellobiohydrolase II-like cellulase CelH [Orpinomyces sp. PC-2]
Contig22741	CBM10	583	326	AAL92497	gi|29465670|gb|AAL92497.1|	7.00E-51	175	56.0	exoglucanase Cel6A [Piromyces sp. E2]

Contig894	CBM10	928	98	P55296	gi|1708917|sp|P55296.1|MANA_PIRSP	3.00E-24	132	43.2	RecName: Full=Mannan endo-1,4-beta-mannosidase A; AltName: Full=Beta-mannanase A; AltName: Full=1,4-beta-D-mannan mannanohydrolase A; Flags: Precursorgi|1197368|emb|CAA62968.1| mannan endo-1,4-beta-mannosidase; mannanase A [Piromyces sp.]
Contig28640	CBM10	617	1105	AAM94167	gi|33620325|gb|AAM94167.1|	5.00E-28	83	63.9	cellulosomal glycoside hydrolase family 6 exoglucanase Cel6A [Piromyces equi]
NODE_8335_length_1156_cov_10.518167	CBM10	1208	3559	Q9Y871	gi|23821548|sp|Q9Y871.1|FAEB_PIREQ	1.00E-136	252	90.5	RecName: Full=Feruloyl esterase B; AltName: Full=Ferulic acid esterase B; AltName: Full=Esterase A; Short=EstA; AltName: Full=Cinnamoyl ester hydrolase; Flags: Precursorgi|5566342|gb|AAD45376.1|AF164516_1 cinnamoyl ester hydrolase EstA [Piromyces equi]
Contig2286	CBM10	500	414	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	9.00E-34	139	51.8	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig23820	CBM10	644	344	Q9Y871	gi|23821548|sp|Q9Y871.1|FAEB_PIREQ	4.00E-36	86	84.9	RecName: Full=Feruloyl esterase B; AltName: Full=Ferulic acid esterase B; AltName: Full=Esterase A; Short=EstA; AltName: Full=Cinnamoyl ester hydrolase; Flags: Precursorgi|5566342|gb|AAD45376.1|AF164516_1 cinnamoyl ester hydrolase EstA [Piromyces equi]
Contig26006*	CBM10	1246	793	YP_003250020	gi|261416337|ref|YP_003250020.1|	1.00E-108	264	70.1	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig21644	CBM10	534	12	AAP30750	gi|30315041|gb|AAP30750.1|	2.00E-29	88	62.5	cellobiohydrolase Cel6C [Piromyces sp. E2]
Contig17552	CBM10	524	263	AAL92497	gi|29465670|gb|AAL92497.1|	1.00E-28	81	64.2	exoglucanase Cel6A [Piromyces sp. E2]

Contig1447	CBM10	1026	859	AAL01214	gi|15529300|gb|AAL01214.1|AF177207_1	8.00E-32	149	42.3	endo-glucanase CelJ [Orpinomyces sp. PC-2]
Contig28743	CBM10	666	475	AAP30750	gi|30315041|gb|AAP30750.1|	4.00E-33	94	62.8	cellobiohydrolase Cel6C [Piromyces sp. E2]
NODE_37676_length_519_cov_4.001927	CBM10	555	178	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	2.00E-59	159	66.7	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig29725	CBM10	1174	2896	YP_003250020	gi|261416337|ref|YP_003250020.1|	3.00E-86	252	59.9	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig29013	CBM10	940	707	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	1.00E-10	107	34.6	cellulase Cel9A precursor [Piromyces sp. E2]
Contig26433*	CBM10	654	831	AAP30750	gi|30315041|gb|AAP30750.1|	1.00E-25	82	59.8	cellobiohydrolase Cel6C [Piromyces sp. E2]
Contig27699	CBM10	663	230	AAB69090	gi|2231243|gb|AAB69090.1|	5.00E-88	225	68.4	acetylxylan esterase [Neocallimastix patriciarum]
Contig22604*	CBM10	1261	3053	AAM94167	gi|33620325|gb|AAM94167.1|	2.00E-27	82	63.4	cellulosomal glycoside hydrolase family 6 exoglucanase Cel6A [Piromyces equi]
Contig2685*	CBM10	1520	1526	ZP_06145404	gi|268611677|ref|ZP_06145404.1|	5.00E-51	251	45.4	cellulosome enzyme, dockerin type I [Ruminococcus flavefaciens FD-1]
Contig13044*	CBM10	1277	2431	Q9Y871	gi|23821548|sp|Q9Y871.1|FAEB_PIREQ	1.00E-127	262	82.8	RecName: Full=Feruloyl esterase B; AltName: Full=Ferulic acid esterase B; AltName: Full=Esterase A; Short=EstA; AltName: Full=Cinnamoyl ester hydrolase; Flags: Precursorgi|5566342|gb|AAD45376.1|AF164516_1 cinnamoyl ester hydrolase EstA [Piromyces equi]
Contig28799*	CBM10	848	5681	AAM94167	gi|33620325|gb|AAM94167.	3.00E-24	81	60.5	cellulosomal glycoside hydrolase family 6 exoglucanase Cel6A [Piromyces equi]

					1|			
Contig8961	CBM10	879	232	P55296	gi|1708917|sp|P55296.1|MANA_PIRSP	3.00E-22	127	40.9	RecName: Full=Mannan endo-1,4-beta-mannosidase A; AltName: Full=Beta-mannanase A; AltName: Full=1,4-beta-D-mannan mannanohydrolase A; Flags: Precursorgi|1197368|emb|CAA62968.1| mannan endo-1,4-beta-mannosidase; mannanase A [Piromyces sp.]
Contig28756	CBM10	720	4490	AAL92497	gi|29465670|gb|AAL92497.1|	6.00E-52	179	56.4	exoglucanase Cel6A [Piromyces sp. E2]
Contig15299	CBM10	1098	298	ZP_06145332	gi|268611605|ref|ZP_06145332.1|	1.00E-67	230	55.7	hypothetical protein RflaF_19133 [Ruminococcus flavefaciens FD-1]
NODE_1804_length_2041_cov_4.901029*	CBM10	2085	1159	AAP30750	gi|30315041|gb|AAP30750.1|	3.00E-24	87	55.2	cellobiohydrolase Cel6C [Piromyces sp. E2]
Contig2507	CBM10	717	768	CAB92325	gi|8052314|emb|CAB92325.1|	7.00E-33	214	36.4	endoglucanase 45A [Piromyces equi]
Contig668	CBM10	656	78	AAM94167	gi|33620325|gb|AAM94167.1|	2.00E-26	99	55.6	cellulosomal glycoside hydrolase family 6 exoglucanase Cel6A [Piromyces equi]
Contig31387	CBM10	530	253	Q12647	gi|2494328|sp|Q12647.1|GUNB_NEOPA	1.00E-24	97	53.6	RecName: Full=Endoglucanase B; AltName: Full=Endo-1,4-beta-glucanase B; AltName: Full=Cellulase B; Flags: Precursorgi|467687|emb|CAA83238.1| endoglucanase B [Neocallimastix patriciarum]
Contig2880*	CBM10	704	432	AAL01213	gi|15529298|gb|AAL01213.1|AF177206_1	9.00E-30	106	55.7	mannanase ManA [Orpinomyces sp. PC-2]
Contig30437	CBM10	738	460	AAM94167	gi|33620325|gb|AAM94167.1|	2.00E-25	99	54.5	cellulosomal glycoside hydrolase family 6 exoglucanase Cel6A [Piromyces equi]
Contig24987	CBM10	757	377	AAP30750	gi|30315041|gb|AAP30750.1|	6.00E-22	86	51.2	cellobiohydrolase Cel6C [Piromyces sp. E2]
Contig29774*	CBM10	954	780	AAP30750	gi|30315041|gb|AAP30750.1|	5.00E-25	82	57.3	cellobiohydrolase Cel6C [Piromyces sp. E2]

Contig1262	CBM10	2048	87	AAP30750	gi|30315041|gb|AAP30750.1|	1.00E-24	82	57.3	cellobiohydrolase Cel6C [Piromyces sp. E2]
Contig28841	CBM10	540	6915	AAD04194	gi|1655815|gb|AAD04194.1|	9.00E-52	147	64.6	xylanase [Orpinomyces sp. PC-2]
Contig29794	CBM10	568	430	AAL01213	gi|15529298|gb|AAL01213.1|AF177206_1	5.00E-25	88	56.8	mannanase ManA [Orpinomyces sp. PC-2]
Contig29726*	CBM10	574	1415	AAP30750	gi|30315041|gb|AAP30750.1|	2.00E-25	83	55.4	cellobiohydrolase Cel6C [Piromyces sp. E2]
Contig204*	CBM10	842	4025	AAB92679	gi|1813486|gb|AAB92679.1|	2.00E-27	83	60.2	cellulase C [Orpinomyces sp. PC-2]
Contig29063	CBM10	1299	957	ZP_06145332	gi|268611605|ref|ZP_06145332.1|	1.00E-113	302	61.9	hypothetical protein RflaF_19133 [Ruminococcus flavefaciens FD-1]
Contig28951	CBM10	534	1177	AAB69092	gi|2231247|gb|AAB69092.1|	2.00E-23	80	62.5	acetylxylan esterase [Neocallimastix patriciarum]
Contig21348*	CBM10	787	4478	AAP30750	gi|30315041|gb|AAP30750.1|	1.00E-25	82	53.7	cellobiohydrolase Cel6C [Piromyces sp. E2]
Contig539*	CBM10	1386	5913	AAL01214	gi|15529300|gb|AAL01214.1|AF177207_1	2.00E-27	136	40.4	endo-glucanase CelJ [Orpinomyces sp. PC-2]
Contig29505	CBM10	2076	2721	ZP_04668137	gi|239625106|ref|ZP_04668137.1|	4.00E-52	244	45.5	methionine adenosyltransferase [Clostridiales bacterium 1_7_47_FAA]gi|239519336|gb|EEQ59202.1| methionine adenosyltransferase [Clostridiales bacterium 1_7_47_FAA]
Contig29865	CBM10	604	210	AAL01214	gi|15529300|gb|AAL01214.1|AF177207_1	4.00E-26	101	50.5	endo-glucanase CelJ [Orpinomyces sp. PC-2]
Contig21376*	CBM10	1283	1725	AAB69092	gi|2231247|gb|AAB69092.1|	0	389	86.1	acetylxylan esterase [Neocallimastix patriciarum]
Contig3265	CBM10	618	512	AAM94167	gi|33620325|gb|AAM94167.1|	3.00E-23	98	51.0	cellulosomal glycoside hydrolase family 6 exoglucanase Cel6A [Piromyces equi]
Contig8692*	CBM10	860	866	AAF70241	gi|7839348|gb|	1.00E	182	62.1	feruloyl esterase A [Orpinomyces sp. PC-2]

					AAF70241.1| AF164351_1	-63			
Contig23421*	CBM10	1355	2867	AAB69092	gi|2231247|gb| AAB69092.1|	1.00E -131	351	67.2	acetylxylan esterase [Neocallimastix patriciarum]
Contig24133	CBM10	895	500	AAB69090	gi|2231243|gb| AAB69090.1|	5.00E -24	80	57.5	acetylxylan esterase [Neocallimastix patriciarum]
Contig36	CBM10	663	1752	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	1.00E -23	91	54.9	cellulase Cel48A precursor [Piromyces sp. E2]
Contig17825*	CBM10	2062	1161	ZP_06145332	gi|268611605|ref|ZP_06145332.1|	1.00E -175	588	51.0	hypothetical protein RflaF_19133 [Ruminococcus flavefaciens FD-1]
Contig325*	CBM10	791	60	YP_003250020	gi|261416337|ref|YP_003250020.1|	3.00E -44	167	56.3	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig4100	CBM10	679	174	CAA57717	gi|561530|emb|CAA57717.1|	4.00E -28	92	64.1	endoxylanase [Neocallimastix frontalis]
Contig4551	CBM10	1122	613	AAQ10006	gi|33329212|gb|AAQ10006.1|	1.00E -114	301	68.1	acetylxylan esterase [Neocallimastix frontalis]
Contig2401	CBM10	1237	488	YP_003250957	gi|261417274|ref|YP_003250957.1|	3.00E -95	265	67.2	hypothetical protein Fisuc_2894 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261373730|gb|ACX76475.1| hypothetical protein Fisuc_2894 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig29858	CBM10	1137	989	AAQ10006	gi|33329212|gb|AAQ10006.1|	1.00E -117	291	71.1	acetylxylan esterase [Neocallimastix frontalis]
Contig1365	CBM10	1603	107	ZP_05496436	gi|256755664|ref|ZP_05496436.1|	4.00E -78	376	44.7	lipolytic protein G-D-S-L family [Clostridium papyrosolvens DSM 2782]gi|256745498|gb|EEU58630.1| lipolytic protein G-D-S-L family [Clostridium papyrosolvens DSM 2782]
Contig7493*	CBM10	977	382	CBK75021	gi|291519800|emb|CBK75021.1|	1.00E -21	224	30.4	Beta-1,4-xylanase [Butyrivibrio fibrisolvens 16/4]
Contig1319	CBM10	766	91	AAM81966	gi|21929667|g	7.00E	100	46.0	cellulase Cel9A precursor [Piromyces sp. E2]

					b|AAM81966.1|AF459452_1	-20			
NODE_21249_length_531_cov_5.269303	CBM10	567	231	AAD43818	gi|5457159|gb|AAD43818.1|AF165266_1	5.00E-78	161	78.3	endoglucanase precursor [Piromyces rhizinflatus]
Contig6232	CBM10	540	218	AAM94167	gi|33620325|gb|AAM94167.1|	8.00E-21	75	57.3	cellulosomal glycoside hydrolase family 6 exoglucanase Cel6A [Piromyces equi]
Contig28921	CBM10	1122	1289	ACL68347	gi|219964511|gb|ACL68347.1|	1.00E-137	380	62.4	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig24540	CBM10	741	232	AAB69090	gi|2231243|gb|AAB69090.1|	2.00E-24	78	59.0	acetylxylan esterase [Neocallimastix patriciarum]
Contig3018*	CBM10	2298	470	AAL01211	gi|15529294|gb|AAL01211.1|AF177204_1	1.00E-38	80	56.3	cellobiohydrolase II-like cellulase CelH [Orpinomyces sp. PC-2]
Contig30041	CBM10	1158	886	ZP_06142046	gi|268608319|ref|ZP_06142046.1|	1.00E-50	178	49.4	hypothetical protein RflaF_02313 [Ruminococcus flavefaciens FD-1]
Contig28851	CBM10	863	388	YP_003250020	gi|261416337|ref|YP_003250020.1|	9.00E-49	145	64.1	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig5636	CBM10	2418	2391	ZP_04668137	gi|239625106|ref|ZP_04668137.1|	8.00E-54	288	40.6	methionine adenosyltransferase [Clostridiales bacterium 1_7_47_FAA]gi|239519336|gb|EEQ59202.1| methionine adenosyltransferase [Clostridiales bacterium 1_7_47_FAA]
Contig1477	CBM10	1160	667	ACL68347	gi|219964511|gb|ACL68347.1|	1.00E-146	378	63.0	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig28836	CBM10	994	874	ACL68347	gi|219964511|gb|ACL68347.1|	4.00E-35	86	69.8	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig28653	CBM10	554	758	AAB69092	gi|2231247|gb|AAB69092.1|	2.00E-20	86	57.0	acetylxylan esterase [Neocallimastix patriciarum]
Contig582	CBM10	1150	495	AAQ10005	gi|33329210|gb|AAQ10005.	1.00E-163	282	99.3	acetylxylan esterase [Neocallimastix patriciarum]

					1|			
Contig14329	CBM10	1270	683	AAP30750	gi|30315041|gb|AAP30750.1|	2.00E-25	108	50.0	cellobiohydrolase Cel6C [Piromyces sp. E2]
Contig225	CBM10	1135	300	ZP_06142046	gi|268608319|ref|ZP_06142046.1|	4.00E-47	225	44.0	hypothetical protein RflaF_02313 [Ruminococcus flavefaciens FD-1]
Contig31183	CBM10	644	241	YP_003250045	gi|261416362|ref|YP_003250045.1|	3.00E-33	160	51.3	polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372818|gb|ACX75563.1| polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig772	CBM10	988	325	ABY52795	gi|164375379|gb|ABY52795.1|	1.00E-14	83	44.6	endo-1,4-beta-xylanase [Piromyces communis]
Contig7593	CBM10	1084	1070	YP_003250020	gi|261416337|ref|YP_003250020.1|	1.00E-106	265	69.4	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig7347	CBM10	2022	1576	ZP_04668137	gi|239625106|ref|ZP_04668137.1|	5.00E-54	284	42.6	methionine adenosyltransferase [Clostridiales bacterium 1_7_47_FAA]gi|239519336|gb|EEQ59202.1| methionine adenosyltransferase [Clostridiales bacterium 1_7_47_FAA]
Contig31237	CBM10	876	470	ACL68347	gi|219964511|gb|ACL68347.1|	1.00E-32	95	57.9	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig30406	CBM10	753	533	AAB69092	gi|2231247|gb|AAB69092.1|	1.00E-70	249	50.6	acetylxylan esterase [Neocallimastix patriciarum]
Contig29362	CBM10	566	681	AAP30750	gi|30315041|gb|AAP30750.1|	4.00E-20	83	57.8	cellobiohydrolase Cel6C [Piromyces sp. E2]
Contig1116	CBM10	507	109	AAP30747	gi|30315035|gb|AAP30747.1|	1.00E-23	80	53.8	mannanase ManA [Piromyces sp. E2]
Contig5325	CBM10	1015	424	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	3.00E-91	338	50.0	cellulase Cel9A precursor [Piromyces sp. E2]
Contig22393	CBM10	687	601	AAP30750	gi|30315041|g	1.00E	90	58.9	cellobiohydrolase Cel6C [Piromyces sp. E2]

					b|AAP30750.1|	-36			
Contig8606	CBM10	790	234	Q9Y871	gi|23821548|sp|Q9Y871.1|FAEB_PIREQ	4.00E-56	121	88.4	RecName: Full=Feruloyl esterase B; AltName: Full=Ferulic acid esterase B; AltName: Full=Esterase A; Short=EstA; AltName: Full=Cinnamoyl ester hydrolase; Flags: Precursorgi|5566342|gb|AAD45376.1|AF164516_1 cinnamoyl ester hydrolase EstA [Piromyces equi]
NODE_43467_length_672_cov_3.133929	CBM10	708	256	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	2.00E-47	237	40.9	cellulase Cel9A precursor [Piromyces sp. E2]
Contig21916	CBM10	730	257	AAP30753	gi|30315047|gb|AAP30753.1|	5.00E-26	183	35.0	cellulosomal glycoside hydrolase family 9 endoglucanase Cel9B [Piromyces sp. E2]
Contig1411	CBM10	1515	831	ZP_06145404	gi|268611677|ref|ZP_06145404.1|	2.00E-50	251	43.4	cellulosome enzyme, dockerin type I [Ruminococcus flavefaciens FD-1]
Contig12252	CBM10	572	153	AAP30750	gi|30315041|gb|AAP30750.1|	1.00E-21	85	55.3	cellobiohydrolase Cel6C [Piromyces sp. E2]
Contig26764	CBM10	1454	594	ZP_05496436	gi|256755664|ref|ZP_05496436.1|	1.00E-73	361	43.5	lipolytic protein G-D-S-L family [Clostridium papyrosolvens DSM 2782]gi|256745498|gb|EEU58630.1| lipolytic protein G-D-S-L family [Clostridium papyrosolvens DSM 2782]
Contig30162	CBM10	785	513	YP_003248565	gi|261414882|ref|YP_003248565.1|	6.00E-66	173	69.4	glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261371338|gb|ACX74083.1| glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig29219	CBM10	1180	4114	YP_003250045	gi|261416362|ref|YP_003250045.1|	1.00E-116	331	64.4	polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372818|gb|ACX75563.1| polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig11932	CBM10	1037	1156	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_	2.00E-23	82	58.5	cellulase Cel9A precursor [Piromyces sp. E2]

					1				
Contig4621	CBM10	901	355	BAB39493	gi|13383322|dbj|BAB39493.1|	2.00E-49	145	66.9	xylanase B [Ruminococcus albus]
Contig25171	CBM10	1883	2210	ZP_05400407	gi|255654998|ref|ZP_05400407.1|	2.00E-24	405	26.2	hypothetical protein CdifQCD-2_04729 [Clostridium difficile QCD-23m63]
Contig28821*	CBM10	1843	931	ZP_05400407	gi|255654998|ref|ZP_05400407.1|	3.00E-20	390	23.6	hypothetical protein CdifQCD-2_04729 [Clostridium difficile QCD-23m63]
Contig6286	CBM10	1015	642	Q12647	gi|2494328|sp|Q12647.1|GUNB_NEOPA	1.00E-21	87	51.7	RecName: Full=Endoglucanase B; AltName: Full=Endo-1,4-beta-glucanase B; AltName: Full=Cellulase B; Flags: Precursorgi|467687|emb|CAA83238.1| endoglucanase B [Neocallimastix patriciarum]
Contig10643	CBM10	731	270	NP_149281	gi|15004821|ref|NP_149281.1|	1.00E-21	115	45.2	xylan degradation protein [Clostridium acetobutylicum ATCC 824]gi|14994433|gb|AAK76863.1|AE001438_116 Possible xylan degradation enzyme (glycosyl hydrolase family 30-like domain and Ricin B-like domain) [Clostridium acetobutylicum ATCC 824]
Contig4497	CBM10	825	327	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	2.00E-20	79	58.2	cellulase Cel48A precursor [Piromyces sp. E2]
Contig22537	CBM10	1628	977	ZP_06142046	gi|268608319|ref|ZP_06142046.1|	4.00E-52	243	44.9	hypothetical protein RflaF_02313 [Ruminococcus flavefaciens FD-1]
Contig1836	CBM10	1232	175	YP_003250045	gi|261416362|ref|YP_003250045.1|	1.00E-121	266	77.1	polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372818|gb|ACX75563.1| polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig3112	CBM10	630	368	ACL68347	gi|219964511|gb|ACL68347.1|	6.00E-50	163	57.7	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig269	CBM10	582	2554	AAB92679	gi|1813486|gb|AAB92679.1|	4.00E-18	64	59.4	cellulase C [Orpinomyces sp. PC-2]
Contig2420	CBM10	1091	420	CAB92325	gi|8052314|emb|CAB9232	1.00E-20	132	38.6	endoglucanase 45A [Piromyces equi]

					5.1|			
Contig29549	CBM10	525	494	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	3.00E-17	81	51.9	cellulase Cel9A precursor [Piromyces sp. E2]
Contig1228	CBM10	978	45	ZP_06142857	gi|268609130|ref|ZP_06142857.1|	1.00E-88	227	68.7	glycoside hydrolase family protein [Ruminococcus flavefaciens FD-1]
Contig688	CBM10	1403	642	ZP_05496436	gi|256755664|ref|ZP_05496436.1|	5.00E-70	348	43.4	lipolytic protein G-D-S-L family [Clostridium papyrosolvens DSM 2782]gi|256745498|gb|EEU58630.1| lipolytic protein G-D-S-L family [Clostridium papyrosolvens DSM 2782]
Contig1503*	CBM10	2230	2105	ZP_06142046	gi|268608319|ref|ZP_06142046.1|	3.00E-69	243	48.1	hypothetical protein RflaF_02313 [Ruminococcus flavefaciens FD-1]
Contig16443	CBM10	775	629	P55296	gi|1708917|sp|P55296.1|MANA_PIRSP	9.00E-16	171	30.4	RecName: Full=Mannan endo-1,4-beta-mannosidase A; AltName: Full=Beta-mannanase A; AltName: Full=1,4-beta-D-mannan mannanohydrolase A; Flags: Precursorgi|1197368|emb|CAA62968.1| mannan endo-1,4-beta-mannosidase; mannanase A [Piromyces sp.]
Contig6381	CBM10	701	232	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	1.00E-20	81	45.7	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig22989	CBM10	663	473	YP_001038591	gi|125974681|ref|YP_001038591.1|	1.00E-18	71	54.9	carbohydrate-binding family 6 protein [Clostridium thermocellum ATCC 27405]gi|256004120|ref|ZP_05429104.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM 2360]gi|281418847|ref|ZP_06249866.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]gi|125714906|gb|ABN53398.1| Carbohydrate binding family 6 [Clostridium thermocellum ATCC 27405]gi|255991868|gb|EEU01966.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM

									2360]gi|281407931|gb|EFB38190.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]
Contig30121*	CBM10	712	588	AAF70241	gi|7839348|gb|AAF70241.1|AF164351_1	4.00E-28	143	45.5	feruloyl esterase A [Orpinomyces sp. PC-2]
Contig22174	CBM10	1175	718	AAD43818	gi|5457159|gb|AAD43818.1|AF165266_1	9.00E-17	87	49.4	endoglucanase precursor [Piromyces rhizinflatus]
Contig29181*	CBM10	972	1684	AAD43818	gi|5457159|gb|AAD43818.1|AF165266_1	7.00E-14	86	44.2	endoglucanase precursor [Piromyces rhizinflatus]
Contig22301	CBM10	1693	1197	AAF70241	gi|7839348|gb|AAF70241.1|AF164351_1	1.00E-107	433	47.3	feruloyl esterase A [Orpinomyces sp. PC-2]
Contig2744	CBM10	841	740	P55296	gi|1708917|sp|P55296.1|MANA_PIRSP	1.00E-15	165	31.5	RecName: Full=Mannan endo-1,4-beta-mannosidase A; AltName: Full=Beta-mannanase A; AltName: Full=1,4-beta-D-mannan mannanohydrolase A; Flags: Precursorgi|1197368|emb|CAA62968.1| mannan endo-1,4-beta-mannosidase; mannanase A [Piromyces sp.]
Contig30404	CBM10	516	287	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	1.00E-29	179	38.5	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags: Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig21698	CBM10	556	261	P55297	gi|1708920|sp|P55297.1|MANB_PIRSP	1.00E-18	86	54.7	RecName: Full=Mannan endo-1,4-beta-mannosidase B; AltName: Full=Beta-mannanase B; AltName: Full=1,4-beta-D-mannan mannanohydrolase B; Flags: Precursorgi|1279643|emb|CAA66061.1| endo-1,4 beta-mannanase [Piromyces sp.]
Contig18696	CBM10	930	585	AAB69092	gi|2231247|gb|AAB69092.1|	1.00E-116	312	67.6	acetylxylan esterase [Neocallimastix patriciarum]
Contig29445	CBM10	1015	671	ZP_05496436	gi|256755664|ref|ZP_05496436.1|	2.00E-21	163	39.9	lipolytic protein G-D-S-L family [Clostridium papyrosolvens DSM 2782]gi|256745498|gb|EEU58630.1| lipolytic protein G-D-S-L family [Clostridium papyrosolvens DSM 2782]

NODE_9175_length_815_cov_8.579141*	CBM10	859	657	YP_003099042	gi|256375382|ref|YP_003099042.1|	6.00E-21	141	43.3	lipolytic protein G-D-S-L family [Actinosynnema mirum DSM 43827]gi|255919685|gb|ACU35196.1| lipolytic protein G-D-S-L family [Actinosynnema mirum DSM 43827]
Contig22783*	CBM10	1101	3835	AAD43818	gi|5457159|gb|AAD43818.1|AF165266_1	1.00E-13	88	44.3	endoglucanase precursor [Piromyces rhizinflatus]
Contig22403	CBM10	678	774	AAB92679	gi|1813486|gb|AAB92679.1|	1.00E-25	182	35.2	cellulase C [Orpinomyces sp. PC-2]
Contig9577	CBM10	598	259	CBL17363	gi|291544254|emb|CBL17363.1|	3.00E-19	90	51.1	Endoglucanase [Ruminococcus sp. 18P13]
Contig21635	CBM10	1309	761	YP_002504781	gi|220927872|ref|YP_002504781.1|	6.00E-98	359	52.6	cellulosome protein dockerin type I [Clostridium cellulolyticum H10]gi|219998200|gb|ACL74801.1| cellulosome protein dockerin type I [Clostridium cellulolyticum H10]
NODE_74290_length_492_cov_4.481707	CBM10	528	205	AAD43818	gi|5457159|gb|AAD43818.1|AF165266_1	1.00E-15	85	47.1	endoglucanase precursor [Piromyces rhizinflatus]
Contig24467	CBM10	676	365	AAC06321	gi|2981484|gb|AAC06321.1|	1.00E-16	89	43.8	cellulase CelD [Neocallimastix patriciarum]
Contig8953	CBM10	742	479	CAB92326	gi|8052316|emb|CAB92326.1|	5.00E-16	104	39.4	endoglucanase 5A [Piromyces equi]
Contig5038	CBM10	1395	938	ZP_04668137	gi|239625106|ref|ZP_04668137.1|	3.00E-18	114	43.9	methionine adenosyltransferase [Clostridiales bacterium 1_7_47_FAA]gi|239519336|gb|EEQ59202.1| methionine adenosyltransferase [Clostridiales bacterium 1_7_47_FAA]
Contig23376	CBM10	501	165	AAC05164	gi|2935581|gb|AAC05164.1|	8.00E-17	66	60.6	1,4-beta-D-glucan-4-glucanohydrolase [Orpinomyces sp. PC-2]
Contig19*	CBM10	572	24	AAM94167	gi|33620325|gb|AAM94167.1|	3.00E-17	98	46.9	cellulosomal glycoside hydrolase family 6 exoglucanase Cel6A [Piromyces equi]
Contig29925	CBM10	944	573	P55296	gi|1708917|sp|P55296.1|MANA_PIRSP	2.00E-17	116	40.5	RecName: Full=Mannan endo-1,4-beta-mannosidase A; AltName: Full=Beta-mannanase A; AltName: Full=1,4-beta-D-mannan mannanohydrolase A; Flags: Precursorgi|1197368|emb|CAA62968.1|

									mannan endo-1,4-beta-mannosidase; mannanase A [Piromyces sp.]
Contig23136	CBM10	521	243	ACL68347	gi|219964511|gb|ACL68347.1|	4.00E-22	63	73.0	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig19516*	CBM10	1437	3285	YP_003250020	gi|261416337|ref|YP_003250020.1|	1.00E-107	273	67.0	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig1217	CBM10	573	252	AAL01211	gi|15529294|gb|AAL01211.1|AF177204_1	2.00E-11	105	36.2	cellobiohydrolase II-like cellulase CelH [Orpinomyces sp. PC-2]
Contig29965	CBM10	1143	1000	YP_003250020	gi|261416337|ref|YP_003250020.1|	1.00E-114	266	72.6	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig23266*	CBM10	1413	1003	YP_003250020	gi|261416337|ref|YP_003250020.1|	1.00E-113	274	69.3	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig30390	CBM10	1022	5033	YP_003250020	gi|261416337|ref|YP_003250020.1|	1.00E-44	144	59.0	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig29158	CBM10	1144	1521	YP_003250020	gi|261416337|ref|YP_003250020.1|	2.00E-35	129	55.8	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig23209	CBM10	943	686	Q9Y871	gi|23821548|sp|Q9Y871.1|FAEB_PIREQ	2.00E-28	109	65.1	RecName: Full=Feruloyl esterase B; AltName: Full=Ferulic acid esterase B; AltName: Full=Esterase A; Short=EstA; AltName: Full=Cinnamoyl ester hydrolase; Flags: Precursorgi|5566342|gb|AAD45376.1|AF164516_1 cinnamoyl ester hydrolase EstA

									[Piromyces equi]
NODE_48706_length_531_cov_3.696799	CBM10	567	156	2J4M	gi|158428896|pdb|2J4M|A	6.00E-13	97	39.2	Chain A, Double Dockerin From Piromyces Equi Cel45agi|158428897|pdb|2J4N|A Chain A, Double Dockerin From Piromyces Equi Cel45a
Contig29630	CBM10	605	394	AAB69090	gi|2231243|gb|AAB69090.1|	4.00E-62	199	61.3	acetylxylan esterase [Neocallimastix patriciarum]
Contig4303	CBM10	1438	872	AAP30747	gi|30315035|gb|AAP30747.1|	2.00E-16	87	47.1	mannanase ManA [Piromyces sp. E2]
Contig2424*	CBM10	1033	1017	AAB69092	gi|2231247|gb|AAB69092.1|	1.00E-122	295	71.5	acetylxylan esterase [Neocallimastix patriciarum]
Contig30327*	CBM10	1485	689	Q9Y871	gi|23821548|sp|Q9Y871.1|FAEB_PIREQ	1.00E-134	255	87.8	RecName: Full=Feruloyl esterase B; AltName: Full=Ferulic acid esterase B; AltName: Full=Esterase A; Short=EstA; AltName: Full=Cinnamoyl ester hydrolase; Flags: Precursorgi|5566342|gb|AAD45376.1|AF164516_1 cinnamoyl ester hydrolase EstA [Piromyces equi]
Contig23562*	CBM10	599	1200	AAB69348	gi|2353007|gb|AAB69348.1|	2.00E-10	47	51.1	cellulase [Orpinomyces joyonii]
Contig10196	CBM10	657	272	ACL68347	gi|219964511|gb|ACL68347.1|	4.00E-71	205	62.4	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig8050	CBM10	517	186	ACL68347	gi|219964511|gb|ACL68347.1|	2.00E-14	70	47.1	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig21712	CBM10	2367	1470	ZP_04217292	gi|229085040|ref|ZP_04217292.1|	7.00E-16	295	22.0	Spore coat protein H [Bacillus cereus Rock3-44]gi|228698356|gb|EEL51089.1| Spore coat protein H [Bacillus cereus Rock3-44]
NODE_1486_length_882_cov_10.526077	CBM10	934	1699	ACL68347	gi|219964511|gb|ACL68347.1|	1.00E-128	302	72.8	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig22573	CBM10	1373	507	ZP_05496436	gi|256755664|ref|ZP_05496436.1|	1.00E-69	345	44.6	lipolytic protein G-D-S-L family [Clostridium papyrosolvens DSM 2782]gi|256745498|gb|EEU58630.1| lipolytic protein G-D-S-L family [Clostridium papyrosolvens DSM 2782]
Contig10776	CBM10	2066	1173	ACY02060	gi|261868885|gb|ACY02060	4.00E-14	306	22.9	spore coat protein H [Flammeovirga yaeyamensis]

					.1|			
NODE_39361_length_469_cov_2.656716	CBM10	505	135	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	8.00E-14	77	45.5	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig3198	CBM10	728	293	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	2.00E-22	147	37.4	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags: Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig21960	CBM10	1037	580	AAD43818	gi|5457159|gb|AAD43818.1|AF165266_1	2.00E-07	40	62.5	endoglucanase precursor [Piromyces rhizinflatus]
Contig20358	CBM10	673	554	AAD43818	gi|5457159|gb|AAD43818.1|AF165266_1	1.00E-11	63	46.0	endoglucanase precursor [Piromyces rhizinflatus]
Contig25183	CBM10	581	228	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	2.00E-08	84	39.3	cellulase Cel48A precursor [Piromyces sp. E2]
Contig24413	CBM10	582	168	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	2.00E-08	84	38.1	cellulase Cel48A precursor [Piromyces sp. E2]
Contig8015	CBM10	636	315	AAB92678	gi|1813484|gb|AAB92678.1|	9.00E-11	80	41.3	cellulase A [Orpinomyces sp. PC-2]
Contig6151	CBM10	546	222	XP_002378971	gi|238495470|ref|XP_002378971.1|	0.46	100	28.0	beta-galactosidase, putative [Aspergillus flavus NRRL3357]gi|220695621|gb|EED51964.1| beta-galactosidase, putative [Aspergillus flavus NRRL3357]
Contig1878	Ricin_B_lectin	1344	181	CAJ81035	gi|89241796|emb|CAJ81035.1|	4.00E-11	112	35.7	putative galactocerebrosidase [Actinoplanes sp. SE50/110]
Contig24564	Ricin_B_lectin	1005	552	ZP_04805535	gi|242260820|ref|ZP_04805535.1|	4.00E-84	281	54.4	Ricin B lectin [Clostridium cellulovorans 743B]gi|242226911|gb|EES30200.1| Ricin B lectin [Clostridium cellulovorans 743B]
Contig30729	Ricin_B_lectin	1230	512	CBL16867	gi|291543758|emb|CBL16867.1|	1.00E-66	422	37.0	Pectate lyase [Ruminococcus sp. 18P13]

Contig3006	Ricin_B_lectin	1660	298	ZP_04805535	gi|242260820|ref|ZP_04805535.1|	1.00E-139	371	63.1	Ricin B lectin [Clostridium cellulovorans 743B]gi|242226911|gb|EES30200.1| Ricin B lectin [Clostridium cellulovorans 743B]
Contig12042	Ricin_B_lectin	1096	1009	YP_003115585	gi|256394021|ref|YP_003115585.1|	2.00E-08	114	34.2	glycoside hydrolase family 62 [Catenulispora acidiphila DSM 44928]gi|256360247|gb|ACU73744.1| glycoside hydrolase family 62 [Catenulispora acidiphila DSM 44928]
Contig2473	Ricin_B_lectin	1163	414	YP_001195010	gi|146300419|ref|YP_001195010.1|	0.31	214	24.3	hypothetical protein Fjoh_2668 [Flavobacterium johnsoniae UW101]gi|146154837|gb|ABQ05691.1| hypothetical protein Fjoh_2668 [Flavobacterium johnsoniae UW101]
Contig2316	Ricin_B_lectin	894	388	XP_002520852	gi|255559665|ref|XP_002520852.1|	4.00E-81	213	68.5	alpha-galactosidase/alpha-n-acetylgalactosaminidase, putative [Ricinus communis]gi|223539983|gb|EEF41561.1| alpha-galactosidase/alpha-n-acetylgalactosaminidase, putative [Ricinus communis]
Contig7830	Ricin_B_lectin	811	286	XP_002372370	gi|238482263|ref|XP_002372370.1|	0.02	72	29.2	alpha-galactosidase, putative [Aspergillus flavus NRRL3357]gi|292495588|sp|B8MWJ5.1|AGALA_ASPFN RecName: Full=Probable alpha-galactosidase A; AltName: Full=Melibiase A; Flags: Precursorgi|220700420|gb|EED56758.1| alpha-galactosidase, putative [Aspergillus flavus NRRL3357]
Contig399	Ricin_B_lectin	1350	1239	YP_003250020	gi|261416337|ref|YP_003250020.1|	1.00E-112	261	71.6	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig13671	Ricin_B_lectin	638	176	CBL16867	gi|291543758|emb|CBL16867.1|	3.00E-47	204	45.1	Pectate lyase [Ruminococcus sp. 18P13]
Contig24952	Ricin_B_lectin	576	424	YP_590752	gi|94968704|ref|YP_590752.1|	2.00E-09	75	42.7	Alpha-galactosidase [Candidatus Koribacter versatilis Ellin345]gi|94550754|gb|ABF40678.1| Alpha-galactosidase [Candidatus Koribacter versatilis Ellin345]

Contig31172	Ricin_B_lectin	552	306	XP_001458591	gi|145545815|ref|XP_001458591.1|	0.025	176	25.0	hypothetical protein [Paramecium tetraurelia strain d4-2]gi|124426412|emb|CAK91194.1| unnamed protein product [Paramecium tetraurelia]
Contig4385	Ricin_B_lectin	537	177	XP_002372370	gi|238482263|ref|XP_002372370.1|	0.73	68	29.4	alpha-galactosidase, putative [Aspergillus flavus NRRL3357]gi|292495588|sp|B8MWJ5.1|AGALA_ASPFN RecName: Full=Probable alpha-galactosidase A; AltName: Full=Melibiase A; Flags: Precursorgi|220700420|gb|EED56758.1| alpha-galactosidase, putative [Aspergillus flavus NRRL3357]
Contig23124	Chitin_bind_1	1060	918	Q9AVB0	gi|56404659|sp|Q9AVB0.1|LECB_PHYAM	4.00E-68	313	44.4	RecName: Full=Lectin-B; AltName: Full=PL-B; Flags: Precursorgi|13537555|dbj|BAB40792.1| mitogen PL-B [Phytolacca americana]
Contig5823	Chitin_bind_1	1002	402	EEY22269	gi|261359841|gb|EEY22269.1|	6.00E-61	298	40.6	lectin [Verticillium albo-atrum VaMs.102]
Contig27417	Chitin_bind_1	857	474	XP_384927	gi|46119101|ref|XP_384927.1|	5.00E-31	296	32.4	hypothetical protein FG04751.1 [Gibberella zeae PH-1]
Contig4939	Chitin_bind_1	893	509	Q9AVB0	gi|56404659|sp|Q9AVB0.1|LECB_PHYAM	7.00E-50	309	38.2	RecName: Full=Lectin-B; AltName: Full=PL-B; Flags: Precursorgi|13537555|dbj|BAB40792.1| mitogen PL-B [Phytolacca americana]
Contig7224	Chitin_bind_1	781	322	EEY22269	gi|261359841|gb|EEY22269.1|	1.00E-47	334	33.8	lectin [Verticillium albo-atrum VaMs.102]
Contig27332	Chitin_bind_1	716	606	EEY22269	gi|261359841|gb|EEY22269.1|	6.00E-44	264	34.5	lectin [Verticillium albo-atrum VaMs.102]
Contig21647	Chitin_bind_1	799	150	XP_384927	gi|46119101|ref|XP_384927.1|	1.00E-38	255	39.6	hypothetical protein FG04751.1 [Gibberella zeae PH-1]
Contig590	Chitin_bind_1	1113	780	AAR97890	gi|40950521|gb|AAR97890.1|	3.00E-37	142	57.0	cellulosomal serpin precursor [Piromyces sp. E2]
Contig1173	Chitin_bind_1	988	33	EEY22269	gi|261359841|gb|EEY22269	3.00E-49	290	36.9	lectin [Verticillium albo-atrum VaMs.102]

					.1|			
Contig28903	Chitin_bind_1	944	1849	EEY22269	gi|261359841|gb|EEY22269.1|	1.00E-41	275	37.1	lectin [Verticillium albo-atrum VaMs.102]
Contig29161	Chitin_bind_1	829	742	XP_384927	gi|46119101|ref|XP_384927.1|	4.00E-27	281	33.5	hypothetical protein FG04751.1 [Gibberella zeae PH-1]
NODE_6978_length_664_cov_3.314759	Chitin_bind_1	700	124	XP_001905632	gi|171681377|ref|XP_001905632.1|	3.00E-30	218	34.9	unnamed protein product [Podospora anserina]gi|170940647|emb|CAP65875.1| unnamed protein product [Podospora anserina]
Contig241*	Chitin_bind_1	1220	843	EEY22269	gi|261359841|gb|EEY22269.1|	6.00E-41	269	33.1	lectin [Verticillium albo-atrum VaMs.102]
Contig10585	Chitin_bind_1	2014	963	EEY22269	gi|261359841|gb|EEY22269.1|	4.00E-35	393	26.0	lectin [Verticillium albo-atrum VaMs.102]
Contig23027	Chitin_bind_1	632	429	EEY22269	gi|261359841|gb|EEY22269.1|	6.00E-34	213	34.3	lectin [Verticillium albo-atrum VaMs.102]
Contig23154	Chitin_bind_1	707	385	EEY22269	gi|261359841|gb|EEY22269.1|	4.00E-33	191	39.8	lectin [Verticillium albo-atrum VaMs.102]
Contig10002	Chitin_bind_1	809	245	BAI44118	gi|260279095|dbj|BAI44118.1|	7.00E-22	211	36.0	chitin binding protein 4 [Magnaporthe oryzae]
Contig6491	Chitin_bind_1	1998	2815	EEY22269	gi|261359841|gb|EEY22269.1|	2.00E-34	229	35.4	lectin [Verticillium albo-atrum VaMs.102]
Contig20734	Chitin_bind_1	938	620	XP_384927	gi|46119101|ref|XP_384927.1|	2.00E-25	236	37.7	hypothetical protein FG04751.1 [Gibberella zeae PH-1]
Contig24365	Chitin_bind_1	1037	780	XP_384927	gi|46119101|ref|XP_384927.1|	2.00E-25	204	35.8	hypothetical protein FG04751.1 [Gibberella zeae PH-1]
Contig29503	Chitin_bind_1	630	664	BAI44115	gi|260279089|dbj|BAI44115.1|	9.00E-09	152	25.7	chitin binding protein 1 [Magnaporthe oryzae]
Contig15249	Chitin_bind_1	512	518	EEY22269	gi|261359841|gb|EEY22269.1|	8.00E-31	170	39.4	lectin [Verticillium albo-atrum VaMs.102]

Contig5044	Chitin_bind_1	875	533	XP_384927	gi|46119101|ref|XP_384927.1|	2.00E-25	272	32.0	hypothetical protein FG04751.1 [Gibberella zeae PH-1]
Contig360	Chitin_bind_1	567	1752	EEY22269	gi|261359841|gb|EEY22269.1|	3.00E-34	178	41.6	lectin [Verticillium albo-atrum VaMs.102]
Contig5276	Chitin_bind_1	657	399	BAI44126	gi|260279111|dbj|BAI44126.1|	8.00E-19	167	31.7	chitin binding 13 [Magnaporthe oryzae]
Contig14080	Chitin_bind_1	653	562	BAI44118	gi|260279095|dbj|BAI44118.1|	3.00E-24	187	33.7	chitin binding protein 4 [Magnaporthe oryzae]
Contig22883*	Chitin_bind_1	1040	724	XP_001905632	gi|171681377|ref|XP_001905632.1|	4.00E-27	220	33.2	unnamed protein product [Podospora anserina]gi|170940647|emb|CAP65875.1| unnamed protein product [Podospora anserina]
Contig26354	Chitin_bind_1	613	768	BAI44115	gi|260279089|dbj|BAI44115.1|	7.00E-24	139	41.7	chitin binding protein 1 [Magnaporthe oryzae]
Contig5345	Chitin_bind_1	650	452	BAI44118	gi|260279095|dbj|BAI44118.1|	3.00E-26	178	39.9	chitin binding protein 4 [Magnaporthe oryzae]
Contig3168	Chitin_bind_1	530	188	BAI44118	gi|260279095|dbj|BAI44118.1|	5.00E-20	129	37.2	chitin binding protein 4 [Magnaporthe oryzae]
Contig4429	Chitin_bind_1	630	1156	XP_384927	gi|46119101|ref|XP_384927.1|	7.00E-33	216	38.9	hypothetical protein FG04751.1 [Gibberella zeae PH-1]
Contig8326	Chitin_bind_1	686	343	EEY22269	gi|261359841|gb|EEY22269.1|	8.00E-25	110	49.1	lectin [Verticillium albo-atrum VaMs.102]
Contig2280*	Chitin_bind_1	2155	2020	EEY22269	gi|261359841|gb|EEY22269.1|	8.00E-26	201	33.3	lectin [Verticillium albo-atrum VaMs.102]
Contig12064	Chitin_bind_1	506	243	XP_384927	gi|46119101|ref|XP_384927.1|	1.00E-29	157	43.9	hypothetical protein FG04751.1 [Gibberella zeae PH-1]
Contig29245	Chitin_bind_1	1200	1544	Q9AVB0	gi|56404659|sp|Q9AVB0.1|LECB_PHYAM	5.00E-23	175	34.9	RecName: Full=Lectin-B; AltName: Full=PL-B; Flags: Precursorgi|13537555|dbj|BAB40792.1| mitogen PL-B [Phytolacca americana]

Contig8652	Chitin_bind_1	709	256	BAI44126	gi|260279111|dbj|BAI44126.1|	2.00E-17	174	33.3	chitin binding 13 [Magnaporthe oryzae]
Contig8855	Chitin_bind_1	597	213	EEY22269	gi|261359841|gb|EEY22269.1|	2.00E-27	226	34.5	lectin [Verticillium albo-atrum VaMs.102]
Contig21263	Chitin_bind_1	746	813	BAI44115	gi|260279089|dbj|BAI44115.1|	3.00E-22	202	35.6	chitin binding protein 1 [Magnaporthe oryzae]
Contig23746	Chitin_bind_1	919	1194	BAI44118	gi|260279095|dbj|BAI44118.1|	2.00E-17	117	44.4	chitin binding protein 4 [Magnaporthe oryzae]
Contig23001	Chitin_bind_1	993	1104	CAP65615	gi|170940388|emb|CAP65615.1|	1.00E-25	267	28.8	unnamed protein product [Podospora anserina]
Contig907*	Chitin_bind_1	2282	3332	EEY22269	gi|261359841|gb|EEY22269.1|	9.00E-21	113	46.9	lectin [Verticillium albo-atrum VaMs.102]
Contig5885	Chitin_bind_1	639	151	XP_001558442	gi|154318247|ref|XP_001558442.1|	8.00E-10	93	35.5	hypothetical protein BC1G_03291 [Botryotinia fuckeliana B05.10]gi|150842814|gb|EDN18007.1| hypothetical protein BC1G_03291 [Botryotinia fuckeliana B05.10]
Contig628	Chitin_bind_1	644	568	XP_001904882	gi|171679872|ref|XP_001904882.1|	8.00E-26	177	38.4	unnamed protein product [Podospora anserina]gi|170939562|emb|CAP64789.1| unnamed protein product [Podospora anserina]
Contig234*	Chitin_bind_1	2145	2223	EEY21089	gi|261358661|gb|EEY21089.1|	1.00E-19	124	44.4	chitinase [Verticillium albo-atrum VaMs.102]
Contig26500	Chitin_bind_1	551	253	XP_001560867	gi|154323105|ref|XP_001560867.1|	2.00E-14	86	45.3	hypothetical protein BC1G_00895 [Botryotinia fuckeliana B05.10]gi|150848229|gb|EDN23422.1| hypothetical protein BC1G_00895 [Botryotinia fuckeliana B05.10]
Contig13714	Chitin_bind_1	789	286	ABK58015	gi|117937826|gb|ABK58015.1|	2.00E-19	133	39.8	putative chitin deacetylase [Batrachochytrium dendrobatidis]
Contig22529	Chitin_bind_1	1224	701	BAI44126	gi|260279111|dbj|BAI44126.1|	3.00E-25	204	30.4	chitin binding 13 [Magnaporthe oryzae]

Contig21374	Chitin_bind_1	622	234	XP_001904882	gi|171679872|ref|XP_001904882.1|	2.00E-14	202	30.7	unnamed protein product [Podospora anserina]gi|170939562|emb|CAP64789.1| unnamed protein product [Podospora anserina]
Contig2590	Chitin_bind_1	712	556	EEY22269	gi|261359841|gb|EEY22269.1|	4.00E-26	188	35.6	lectin [Verticillium albo-atrum VaMs.102]
Contig2837*	Chitin_bind_1	968	102	XP_001598553	gi|156065263|ref|XP_001598553.1|	5.00E-38	347	32.0	hypothetical protein SS1G_00642 [Sclerotinia sclerotiorum 1980]gi|154691501|gb|EDN91239.1| hypothetical protein SS1G_00642 [Sclerotinia sclerotiorum 1980]
Contig1924	Chitin_bind_1	505	143	EEY21089	gi|261358661|gb|EEY21089.1|	5.00E-18	114	43.0	chitinase [Verticillium albo-atrum VaMs.102]
Contig13893*	Chitin_bind_1	594	279	XP_384927	gi|46119101|ref|XP_384927.1|	1.00E-33	184	42.4	hypothetical protein FG04751.1 [Gibberella zeae PH-1]
Contig11339	Chitin_bind_1	572	211	EEY22269	gi|261359841|gb|EEY22269.1|	2.00E-16	143	35.0	lectin [Verticillium albo-atrum VaMs.102]
Contig9219	Chitin_bind_1	968	277	ABE77384	gi|92430355|gb|ABE77384.1|	3.00E-25	189	35.4	HFR-3 [Triticum aestivum]
Contig5477	Chitin_bind_1	826	659	XP_001933394	gi|189194111|ref|XP_001933394.1|	3.00E-20	118	45.8	chitin binding protein [Pyrenophora tritici-repentis Pt-1C-BFP]gi|187978958|gb|EDU45584.1| chitin binding protein [Pyrenophora tritici-repentis Pt-1C-BFP]
Contig4943	Chitin_bind_1	640	220	XP_384927	gi|46119101|ref|XP_384927.1|	1.00E-28	165	41.8	hypothetical protein FG04751.1 [Gibberella zeae PH-1]
Contig22543	Chitin_bind_1	2138	1547	ZP_02181401	gi|163786953|ref|ZP_02181401.1|	4.00E-17	432	22.9	hypothetical protein FBALC1_17247 [Flavobacteriales bacterium ALC-1]gi|159878813|gb|EDP72869.1| hypothetical protein FBALC1_17247 [Flavobacteriales bacterium ALC-1]
Contig29996	Chitin_bind_1	720	318	EEY22269	gi|261359841|gb|EEY22269.1|	2.00E-15	113	33.6	lectin [Verticillium albo-atrum VaMs.102]
Contig6430	Chitin_bi	555	288	EEY21089	gi|261358661|	1.00E	125	40.8	chitinase [Verticillium albo-atrum VaMs.102]

	nd_1				gb|EEY21089.1|	-20			
Contig30348	Chitin_bind_1	521	295	XP_384927	gi|46119101|ref|XP_384927.1|	9.00E-14	176	29.5	hypothetical protein FG04751.1 [Gibberella zeae PH-1]
Contig1125	Chitin_bind_1	707	765	ABK58015	gi|117937826|gb|ABK58015.1|	2.00E-20	103	42.7	putative chitin deacetylase [Batrachochytrium dendrobatidis]
Contig2064	Chitin_bind_1	1619	32	EEY22269	gi|261359841|gb|EEY22269.1|	3.00E-34	218	37.6	lectin [Verticillium albo-atrum VaMs.102]
Contig21567	Chitin_bind_1	504	220	XP_001912680	gi|171695512|ref|XP_001912680.1|	4.00E-10	37	70.3	unnamed protein product [Podospora anserina]gi|170947998|emb|CAP60162.1| unnamed protein product [Podospora anserina]
NODE_7530_length_476_cov_9.714286	Chitin_bind_1	520	447	EEU39765	gi|256726404|gb|EEU39765.1|	4.00E-18	114	38.6	hypothetical protein NECHADRAFT_93052 [Nectria haematococca mpVI 77-13-4]
Contig6241	Chitin_bind_1	739	306	BAI44126	gi|260279111|dbj|BAI44126.1|	5.00E-17	123	39.8	chitin binding 13 [Magnaporthe oryzae]
Contig3097	Chitin_bind_1	893	7205	EEY21089	gi|261358661|gb|EEY21089.1|	3.00E-17	117	35.0	chitinase [Verticillium albo-atrum VaMs.102]
Contig3790	Chitin_bind_1	581	590	1ULK	gi|40889776|pdb|1ULK|A	9.00E-15	113	39.8	Chain A, Crystal Structure Of Pokeweed Lectin-Cgi|40889777|pdb|1ULK|B Chain B, Crystal Structure Of Pokeweed Lectin-Cgi|1110548|gb|AAB35257.1| lectin-C, PL-C [Phytolacca americana=pokeweeds, roots, Peptide, 126 aa]
Contig23623	Chitin_bind_1	539	426	XP_002384073	gi|238505728|ref|XP_002384073.1|	4.00E-09	34	67.6	class V chitinase, putative [Aspergillus flavus NRRL3357]gi|220690187|gb|EED46537.1| class V chitinase, putative [Aspergillus flavus NRRL3357]
Contig23088	Chitin_bind_1	616	503	EEY21089	gi|261358661|gb|EEY21089.1|	6.00E-15	107	41.1	chitinase [Verticillium albo-atrum VaMs.102]
Contig4227	Chitin_bind_1	665	263	XP_361310	gi|39943546|ref|XP_361310.1|	2.00E-12	109	36.7	hypothetical protein MGG_03784 [Magnaporthe grisea 70-15]gi|145014465|gb|EDJ99033.1| hypothetical protein MGG_03784 [Magnaporthe grisea 70-

									15]
Contig29832	Chitin_bind_1	534	867	EEY22269	gi|261359841|gb|EEY22269.1|	3.00E-16	121	40.5	lectin [Verticillium albo-atrum VaMs.102]
Contig17079	Chitin_bind_1	523	332	EEY22269	gi|261359841|gb|EEY22269.1|	4.00E-19	171	30.4	lectin [Verticillium albo-atrum VaMs.102]
Contig7906	Chitin_bind_1	521	234	XP_001403530	gi|145602203|ref|XP_001403530.1|	4.00E-08	42	54.8	hypothetical protein MGG_12939 [Magnaporthe grisea 70-15]gi|145010634|gb|EDJ95290.1| hypothetical protein MGG_12939 [Magnaporthe grisea 70-15]
Contig31423	Chitin_bind_1	1042	651	XP_359976	gi|145615059|ref|XP_359976.2|	3.00E-16	191	30.4	hypothetical protein MGG_05351 [Magnaporthe grisea 70-15]gi|145021980|gb|EDK06000.1| hypothetical protein MGG_05351 [Magnaporthe grisea 70-15]
NODE_12804_length_1122_cov_11.425134*	Chitin_bind_1	1166	3175	ABK58015	gi|117937826|gb|ABK58015.1|	1.00E-13	119	36.1	putative chitin deacetylase [Batrachochytrium dendrobatidis]
NODE_4383_length_1229_cov_11.374288*	Chitin_bind_1	1281	2636	ABK58015	gi|117937826|gb|ABK58015.1|	3.00E-12	119	34.5	putative chitin deacetylase [Batrachochytrium dendrobatidis]
Contig1224	Chitin_bind_1	537	251	XP_001797954	gi|169609070|ref|XP_001797954.1|	3.00E-11	137	30.7	hypothetical protein SNOG_07620 [Phaeosphaeria nodorum SN15]gi|111063966|gb|EAT85086.1| hypothetical protein SNOG_07620 [Phaeosphaeria nodorum SN15]
Contig21417	Chitin_bind_1	540	676	BAI44118	gi|260279095|dbj|BAI44118.1|	6.00E-20	171	31.0	chitin binding protein 4 [Magnaporthe oryzae]
Contig1644	Chitin_bind_1	1065	64	NP_982610	gi|45184892|ref|NP_982610.1|	1.00E-10	277	23.8	AAR069Wp [Ashbya gossypii ATCC 10895]gi|44980501|gb|AAS50434.1| AAR069Wp [Ashbya gossypii ATCC 10895]
Contig6589	Chitin_bind_1	581	308	XP_001912680	gi|171695512|ref|XP_001912680.1|	9.00E-12	40	72.5	unnamed protein product [Podospora anserina]gi|170947998|emb|CAP60162.1| unnamed protein product [Podospora anserina]
Contig29186	Chitin_bind_1	1991	2982	ZP_04563476	gi|237732995|ref|ZP_04563	2.00E-21	431	27.1	predicted protein [Mollicutes bacterium D7]gi|229383899|gb|EEO33990.1| predicted

					476.1|			protein [Mollicutes bacterium D7]
Contig21811	Chitin_bind_1	506	166	Q01MB6	gi|152013345|sp|Q01MB6.2|AGI_ORYSI	6.00E-20	132	37.9	RecName: Full=Lectin; AltName: Full=Agglutinin; Contains: RecName: Full=Lectin 10 kDa peptide; Contains: RecName: Full=Lectin 8 kDa peptide; Flags: Precursorgi|218194445|gb|EEC76872.1| hypothetical protein OsI_15064 [Oryza sativa Indica Group]
Contig9244	Chitin_bind_1	641	355	ABK58015	gi|117937826|gb|ABK58015.1|	4.00E-20	106	43.4	putative chitin deacetylase [Batrachochytrium dendrobatidis]
Contig20861	Chitin_bind_1	652	424	XP_001792460	gi|169598073|ref|XP_001792460.1|	2.00E-14	213	27.7	hypothetical protein SNOG_01835 [Phaeosphaeria nodorum SN15]gi|111070364|gb|EAT91484.1| hypothetical protein SNOG_01835 [Phaeosphaeria nodorum SN15]
Contig25359	Chitin_bind_1	525	487	XP_361056	gi|39943038|ref|XP_361056.1|	8.00E-07	47	48.9	hypothetical protein MGG_03599 [Magnaporthe grisea 70-15]gi|145009830|gb|EDJ94486.1| hypothetical protein MGG_03599 [Magnaporthe grisea 70-15]gi|260279099|dbj|BAI44120.1| putative chitinase [Magnaporthe oryzae]
Contig2203	Chitin_bind_1	559	147	BAI44115	gi|260279089|dbj|BAI44115.1|	5.00E-12	77	39.0	chitin binding protein 1 [Magnaporthe oryzae]
Contig5517	Chitin_bind_1	539	353	XP_002384073	gi|238505728|ref|XP_002384073.1|	2.00E-10	57	50.9	class V chitinase, putative [Aspergillus flavus NRRL3357]gi|220690187|gb|EED46537.1| class V chitinase, putative [Aspergillus flavus NRRL3357]
Contig29243	Chitin_bind_1	2286	2254	EDP52952	gi|159127837|gb|EDP52952.1|	1.00E-76	670	32.1	extracellular serine-rich protein, putative [Aspergillus fumigatus A1163]
Contig2638	Chitin_bind_1	904	72	XP_001912680	gi|171695512|ref|XP_001912680.1|	2.00E-09	39	66.7	unnamed protein product [Podospora anserina]gi|170947998|emb|CAP60162.1| unnamed protein product [Podospora anserina]
Contig29854	CBM20	1425	1653	ZP_01852968	gi|149174341|ref|ZP_01852968.1|	4.00E-09	342	24.0	Parallel beta-helix repeat protein [Planctomyces maris DSM 8797]gi|148846886|gb|EDL61222.1| Parallel beta-helix repeat protein [Planctomyces maris DSM 8797]

Contig4424	F5_F8_type_C	595	660	YP_003123468	gi|256422815|ref|YP_003123468.1|	6.00E-17	117	41.9	glycoside hydrolase family 18 [Chitinophaga pinensis DSM 2588]gi|256037723|gb|ACU61267.1| glycoside hydrolase family 18 [Chitinophaga pinensis DSM 2588]
Contig9433	CBM48	550	267	EEQ33998	gi|238844336|gb|EEQ33998.1|	1.00E-39	142	54.9	1,4-alpha-glucan branching enzyme [Microsporum canis CBS 113480]
Contig8017	LysM	762	301	ZP_06250496	gi|281419482|ref|ZP_06250496.1|	0.001	175	29.1	ErfK/YbiS/YcfS/YnhG family protein [Clostridium thermocellum JW20]gi|281406888|gb|EFB37152.1| ErfK/YbiS/YcfS/YnhG family protein [Clostridium thermocellum JW20]
Contig731*	LysM	827	286	ZP_05853124	gi|260587211|ref|ZP_05853124.1|	7.00E-33	165	44.8	N-acetylmuramoyl-L-alanine amidase [Blautia hansenii DSM 20583]gi|260542406|gb|EEX22975.1| N-acetylmuramoyl-L-alanine amidase [Blautia hansenii DSM 20583]
Contig29178	LysM	1118	2115	YP_003432889	gi|288818541|ref|YP_003432889.1|	0.026	112	26.8	lipoprotein [Hydrogenobacter thermophilus TK-6]gi|288787941|dbj|BAI69688.1| lipoprotein [Hydrogenobacter thermophilus TK-6]
Contig23026	LysM	712	538	XP_001310466	gi|123439388|ref|XP_001310466.1|	1.00E-63	168	66.7	hypothetical protein [Trichomonas vaginalis G3]gi|121892237|gb|EAX97536.1| conserved hypothetical protein [Trichomonas vaginalis G3]
Contig29982*	LysM	805	2273	ZP_05853124	gi|260587211|ref|ZP_05853124.1|	2.00E-33	166	45.2	N-acetylmuramoyl-L-alanine amidase [Blautia hansenii DSM 20583]gi|260542406|gb|EEX22975.1| N-acetylmuramoyl-L-alanine amidase [Blautia hansenii DSM 20583]
Contig21424*	LysM	672	6275	ZP_05853124	gi|260587211|ref|ZP_05853124.1|	7.00E-23	139	45.3	N-acetylmuramoyl-L-alanine amidase [Blautia hansenii DSM 20583]gi|260542406|gb|EEX22975.1| N-acetylmuramoyl-L-alanine amidase [Blautia hansenii DSM 20583]
Contig2890	LysM	752	1330	ZP_05853124	gi|260587211|ref|ZP_05853124.1|	1.00E-35	165	49.7	N-acetylmuramoyl-L-alanine amidase [Blautia hansenii DSM 20583]gi|260542406|gb|EEX22975.1| N-acetylmuramoyl-L-alanine amidase [Blautia hansenii DSM 20583]

Contig30200	LysM	567	694	EFE40484	gi|291184982|gb|EFE40484.1|	6.00E-04	99	28.3	class V chitinase Chi100 [Trichophyton verrucosum HKI 0517]
Contig1569	LysM	751	1094	ZP_05853124	gi|260587211|ref|ZP_05853124.1|	1.00E-34	164	47.0	N-acetylmuramoyl-L-alanine amidase [Blautia hansenii DSM 20583]gi|260542406|gb|EEX22975.1| N-acetylmuramoyl-L-alanine amidase [Blautia hansenii DSM 20583]
Contig30786	LysM	621	3701	ZP_03716306	gi|225027114|ref|ZP_03716306.1|	0.058	73	28.8	hypothetical protein EUBHAL_01370 [Eubacterium hallii DSM 3353]gi|224955578|gb|EEG36787.1| hypothetical protein EUBHAL_01370 [Eubacterium hallii DSM 3353]
Contig8327	LysM	516	245	XP_001883065	gi|170103701|ref|XP_001883065.1|	3.00E-04	97	32.0	predicted protein [Laccaria bicolor S238N-H82]gi|164641946|gb|EDR06204.1| predicted protein [Laccaria bicolor S238N-H82]
NODE_10482_length_716_cov_4.349162	Polysacc_deac_1	752	371	ZP_05349062	gi|255284507|ref|ZP_05349062.1|	5.00E-27	195	34.4	putative secreted protein [Bryantella formatexigens DSM 14469]gi|255264943|gb|EET58148.1| putative secreted protein [Bryantella formatexigens DSM 14469]
Contig30430	Polysacc_deac_1	628	412	EFE44609	gi|291189315|gb|EFE44609.1|	1.00E-24	212	34.0	hypothetical protein TRV_00605 [Trichophyton verrucosum HKI 0517]
Contig14819	Polysacc_deac_1	641	273	ZP_06347996	gi|283798843|ref|ZP_06347996.1|	9.00E-22	192	35.4	peptidoglycan N-acetylglucosamine deacetylase A [Clostridium sp. M62/1]gi|291073530|gb|EFE10894.1| peptidoglycan N-acetylglucosamine deacetylase A [Clostridium sp. M62/1]
Contig21423	Polysacc_deac_1	1445	2355	XP_359754	gi|39940434|ref|XP_359754.1|	4.00E-26	218	35.8	hypothetical protein MGG_05023 [Magnaporthe grisea 70-15]gi|145010731|gb|EDJ95387.1| hypothetical protein MGG_05023 [Magnaporthe grisea 70-15]gi|260279109|dbj|BAI44125.1| putative chitin deacetylase [Magnaporthe oryzae]
Contig15114	Polysacc_deac_1	988	632	ACF22100	gi|193804915|gb|ACF22100.1|	2.00E-28	222	37.4	chitin deacetylase [Emericella nidulans]
Contig31473	Polysacc_deac_1	651	452	CAQ16203	gi|209570274|emb|CAQ16203.1|	5.00E-26	203	36.0	hypothetical protein [Glomerella graminicola]

Contig7842	Polysacc_deac_1	570	167	EEY20068	gi|261357640|gb|EEY20068.1|	5.00E-26	181	40.3	chitin deacetylase [Verticillium albo-atrum VaMs.102]
Contig23048	Polysacc_deac_1	955	893	EEY21363	gi|261358935|gb|EEY21363.1|	9.00E-35	230	39.1	chitin deacetylase [Verticillium albo-atrum VaMs.102]
Contig22709	Polysacc_deac_1	946	830	XP_001905400	gi|171680912|ref|XP_001905400.1|	1.00E-32	221	36.2	unnamed protein product [Podospora anserina]gi|170940414|emb|CAP65640.1| unnamed protein product [Podospora anserina]
NODE_45052_length_553_cov_5.039783	Polysacc_deac_1	589	217	CAQ16203	gi|209570274|emb|CAQ16203.1|	4.00E-22	111	47.7	hypothetical protein [Glomerella graminicola]
NODE_7121_length_901_cov_10.762486*	Polysacc_deac_1	945	1563	XP_386725	gi|46124343|ref|XP_386725.1|	2.00E-36	251	38.6	hypothetical protein FG06549.1 [Gibberella zeae PH-1]
Contig17344	Polysacc_deac_1	908	314	ABK58012	gi|117937820|gb|ABK58012.1|	3.00E-16	228	30.3	putative chitin deacetylase [Batrachochytrium dendrobatidis]
Contig2348	Polysacc_deac_1	900	1189	CBI58524	gi|289614687|emb|CBI58524.1|	4.00E-31	239	35.6	unnamed protein product [Sordaria macrospora]
Contig1097	Polysacc_deac_1	668	250	2IW0	gi|112491417|pdb|2IW0|A	5.00E-27	218	34.4	Chain A, Structure Of The Chitin Deacetylase From The Fungal Pathogen Colletotrichum Lindemuthianum
Contig993	Polysacc_deac_1	851	467	XP_362692	gi|145612898|ref|XP_362692.2|	1.00E-25	217	33.6	hypothetical protein MGG_08356 [Magnaporthe grisea 70-15]gi|145019991|gb|EDK04219.1| hypothetical protein MGG_08356 [Magnaporthe grisea 70-15]
Contig22718	Polysacc_deac_1	941	330	XP_001831922	gi|169850453|ref|XP_001831922.1|	7.00E-19	266	29.3	hypothetical protein CC1G_12897 [Coprinopsis cinerea okayama7#130]gi|116506998|gb|EAU89893.1| hypothetical protein CC1G_12897 [Coprinopsis cinerea okayama7#130]
Contig28611	Polysacc_deac_1	1069	619	XP_001840108	gi|169867054|ref|XP_001840108.1|	3.00E-26	230	35.7	predicted protein [Coprinopsis cinerea okayama7#130]gi|116498660|gb|EAU81555.1| predicted protein [Coprinopsis cinerea okayama7#130]
Contig9520	Polysacc_deac_1	847	257	XP_001933998	gi|189195320|ref|XP_00193	7.00E-20	253	28.9	chitin binding protein [Pyrenophora tritici-repentis Pt-1C-

					3998.1|			BFP]gi|187979877|gb|EDU46503.1| chitin binding protein [Pyrenophora tritici-repentis Pt-1C-BFP]
Contig25338	Polysacc_deac_1	761	469	EEY20068	gi|261357640|gb|EEY20068.1|	7.00E-26	228	35.1	chitin deacetylase [Verticillium albo-atrum VaMs.102]
Contig25066	Polysacc_deac_1	804	379	XP_363190	gi|39949397|ref|XP_363190.1|	5.00E-20	211	31.8	hypothetical protein MGG_08774 [Magnaporthe grisea 70-15]gi|145009326|gb|EDJ94037.1| hypothetical protein MGG_08774 [Magnaporthe grisea 70-15]
Contig22541	Polysacc_deac_1	836	596	EEU39765	gi|256726404|gb|EEU39765.1|	4.00E-17	248	32.3	hypothetical protein NECHADRAFT_93052 [Nectria haematococca mpVI 77-13-4]
Contig29997	Polysacc_deac_1	977	587	XP_001268141	gi|121699752|ref|XP_001268141.1|	2.00E-17	165	34.5	chitin deacetylase, putative [Aspergillus clavatus NRRL 1]gi|119396283|gb|EAW06715.1| chitin deacetylase, putative [Aspergillus clavatus NRRL 1]
Contig28783	Polysacc_deac_1	1109	1716	XP_958437	gi|164422949|ref|XP_958437.2|	8.00E-25	232	36.2	hypothetical protein NCU09508 [Neurospora crassa OR74A]gi|157069884|gb|EAA29201.2| conserved hypothetical protein [Neurospora crassa OR74A]
Contig23626	Polysacc_deac_1	889	813	EEY20068	gi|261357640|gb|EEY20068.1|	4.00E-24	220	35.0	chitin deacetylase [Verticillium albo-atrum VaMs.102]
Contig931	Polysacc_deac_1	805	830	EEY19684	gi|261357256|gb|EEY19684.1|	5.00E-17	238	29.8	polysaccharide deacetylase family protein [Verticillium albo-atrum VaMs.102]
Contig7835*	Polysacc_deac_1	1253	522	XP_002374100	gi|238485724|ref|XP_002374100.1|	3.00E-18	253	29.2	chitin deacetylase, putative [Aspergillus flavus NRRL3357]gi|220698979|gb|EED55318.1| chitin deacetylase, putative [Aspergillus flavus NRRL3357]
Contig21498*	Polysacc_deac_1	981	2424	YP_677852	gi|110637645|ref|YP_677852.1|	3.00E-55	313	42.2	bifunctional xylanase/esterase; CBM9 module, glycoside hydrolase family 8 protein and carbohydrate esterase family 4 protein [Cytophaga hutchinsonii ATCC 33406]gi|110280326|gb|ABG58512.1| CHU large protein; candidate bifunctional xylanase/esterase; CBM9 module, Glycoside

									Hydrolase Family 8 protein and Carbohydrate Esterase Family 4 protein [Cytophaga hutchinsonii ATCC 33406]
Contig8991	Polysacc_deac_1	940	1944	EEY20068	gi|261357640|gb|EEY20068.1|	2.00E-24	227	33.0	chitin deacetylase [Verticillium albo-atrum VaMs.102]
Contig1781	Polysacc_deac_1	990	1108	YP_677852	gi|110637645|ref|YP_677852.1|	1.00E-52	312	40.1	bifunctional xylanase/esterase; CBM9 module, glycoside hydrolase family 8 protein and carbohydrate esterase family 4 protein [Cytophaga hutchinsonii ATCC 33406]gi|110280326|gb|ABG58512.1| CHU large protein; candidate bifunctional xylanase/esterase; CBM9 module, Glycoside Hydrolase Family 8 protein and Carbohydrate Esterase Family 4 protein [Cytophaga hutchinsonii ATCC 33406]
Contig4101	Polysacc_deac_1	800	337	XP_959693	gi|164422983|ref|XP_959693.2|	2.00E-18	160	32.5	hypothetical protein NCU09582 [Neurospora crassa OR74A]gi|157069898|gb|EAA30457.2| conserved hypothetical protein [Neurospora crassa OR74A]
Contig10063	Polysacc_deac_1	679	324	EEY20068	gi|261357640|gb|EEY20068.1|	1.00E-12	230	27.0	chitin deacetylase [Verticillium albo-atrum VaMs.102]
Contig24262	Polysacc_deac_1	536	186	YP_003412566	gi|284800701|ref|YP_003412566.1|	1.00E-08	133	29.3	hypothetical protein LM5578_0448 [Listeria monocytogenes 08-5578]gi|284993887|ref|YP_003415655.1| hypothetical protein LM5923_0447 [Listeria monocytogenes 08-5923]gi|284056263|gb|ADB67204.1| hypothetical protein LM5578_0448 [Listeria monocytogenes 08-5578]gi|284059354|gb|ADB70293.1| hypothetical protein LM5923_0447 [Listeria monocytogenes 08-5923]
Contig4916	Polysacc_deac_1	967	612	YP_003250045	gi|261416362|ref|YP_003250045.1|	1.00E-155	327	79.5	polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372818|gb|ACX75563.1| polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig22653	Polysacc_deac_1	643	386	XP_001802089	gi|169617349|ref|XP_00180	1.00E-24	225	32.0	hypothetical protein SNOG_11852 [Phaeosphaeria nodorum

					2089.1|			SN15]gi|160703387|gb|EAT80896.2| hypothetical protein SNOG_11852 [Phaeosphaeria nodorum SN15]
Contig426	Pectinesterase	1124	153	ZP_06718852	gi|294640900|ref|ZP_06718852.1|	1.00E-129	367	63.2	pectinesterase [Ruminococcus albus 8]gi|291504166|gb|EFF16850.1| pectinesterase [Ruminococcus albus 8]
Contig21578	Pectinesterase	1218	624	ZP_06143861	gi|268610134|ref|ZP_06143861.1|	1.00E-133	372	63.2	Pectate lyase/Amb allergen [Ruminococcus flavefaciens FD-1]
Contig7215	Pectinesterase	1074	480	ZP_06143861	gi|268610134|ref|ZP_06143861.1|	1.00E-133	355	65.4	Pectate lyase/Amb allergen [Ruminococcus flavefaciens FD-1]
Contig2434	Amidohydro_1	1371	197	XP_001580549	gi|154415047|ref|XP_001580549.1|	0	451	70.7	Amidohydrolase family protein [Trichomonas vaginalis G3]gi|121914768|gb|EAY19563.1| Amidohydrolase family protein [Trichomonas vaginalis G3]
Contig25221	Amidohydro_1	777	548	XP_001580549	gi|154415047|ref|XP_001580549.1|	1.00E-67	255	51.8	Amidohydrolase family protein [Trichomonas vaginalis G3]gi|121914768|gb|EAY19563.1| Amidohydrolase family protein [Trichomonas vaginalis G3]
Contig9584	Amidohydro_1	501	155	XP_001580549	gi|154415047|ref|XP_001580549.1|	4.00E-73	165	81.2	Amidohydrolase family protein [Trichomonas vaginalis G3]gi|121914768|gb|EAY19563.1| Amidohydrolase family protein [Trichomonas vaginalis G3]
Contig22339*	Pec_lyase_C	1515	687	ZP_06142260	gi|268608533|ref|ZP_06142260.1|	1.00E-120	404	51.0	Pectate lyase/Amb allergen [Ruminococcus flavefaciens FD-1]
Contig828	Pec_lyase_C	1837	2147	ZP_06717864	gi|294639839|ref|ZP_06717864.1|	1.00E-156	613	51.9	pectate lyase [Ruminococcus albus 8]gi|291505629|gb|EFF18240.1| pectate lyase [Ruminococcus albus 8]
Contig3060	Pec_lyase_C	671	979	YP_002487258	gi|220911949|ref|YP_002487258.1|	2.00E-26	197	37.6	Pectate lyase/Amb allergen [Arthrobacter chlorophenolicus A6]gi|219858827|gb|ACL39169.1| Pectate lyase/Amb allergen [Arthrobacter chlorophenolicus A6]
Contig16158	Pec_lyase_C	547	204	YP_003243122	gi|261406881|ref|YP_003243122.1|	6.00E-39	178	48.9	Pectate lyase/Amb allergen [Geobacillus sp. Y412MC10]gi|261283344|gb|ACX65315.1| Pectate lyase/Amb allergen [Geobacillus sp. Y412MC10]
Contig6104	Pec_lyase_C	604	265	CBL17651	gi|291544542|emb|CBL1765	4.00E-29	191	37.2	Pectate lyase [Ruminococcus sp. 18P13]

					1.1|			
Contig30471	Pec_lyase_C	518	285	YP_003243122	gi|261406881|ref|YP_003243122.1|	1.00E-14	110	36.4	Pectate lyase/Amb allergen [Geobacillus sp. Y412MC10]gi|261283344|gb|ACX65315.1| Pectate lyase/Amb allergen [Geobacillus sp. Y412MC10]
Contig2859*	Pectate_lyase	795	534	XP_001554464	gi|154310264|ref|XP_001554464.1|	8.00E-52	219	51.1	hypothetical protein BC1G_07052 [Botryotinia fuckeliana B05.10]gi|150851613|gb|EDN26806.1| hypothetical protein BC1G_07052 [Botryotinia fuckeliana B05.10]
Contig28811*	Pectate_lyase	828	4010	XP_001554464	gi|154310264|ref|XP_001554464.1|	3.00E-48	233	48.1	hypothetical protein BC1G_07052 [Botryotinia fuckeliana B05.10]gi|150851613|gb|EDN26806.1| hypothetical protein BC1G_07052 [Botryotinia fuckeliana B05.10]
Contig20357	Pectate_lyase	800	5711	XP_001554464	gi|154310264|ref|XP_001554464.1|	3.00E-50	223	50.2	hypothetical protein BC1G_07052 [Botryotinia fuckeliana B05.10]gi|150851613|gb|EDN26806.1| hypothetical protein BC1G_07052 [Botryotinia fuckeliana B05.10]
Contig2492	Pectate_lyase	526	585	XP_001215463	gi|115399748|ref|XP_001215463.1|	9.00E-30	146	50.0	hypothetical protein ATEG_06285 [Aspergillus terreus NIH2624]gi|114191129|gb|EAU32829.1| hypothetical protein ATEG_06285 [Aspergillus terreus NIH2624]
Contig21247	Alpha-L-AF_C	1382	873	CBL09057	gi|291535945|emb|CBL09057.1|	1.00E-127	466	49.1	Alpha-L-arabinofuranosidase [Roseburia intestinalis M50/1]gi|291538437|emb|CBL11548.1| Alpha-L-arabinofuranosidase [Roseburia intestinalis XB6B4]
Contig8415	Bgal_small_N	721	219	ZP_06142336	gi|268608609|ref|ZP_06142336.1|	1.00E-91	236	68.2	beta-galactosidase [Ruminococcus flavefaciens FD-1]
Contig3630	Bgal_small_N	832	356	ZP_06142336	gi|268608609|ref|ZP_06142336.1|	1.00E-96	264	64.4	beta-galactosidase [Ruminococcus flavefaciens FD-1]
Contig10124	Bgal_small_N	703	261	ZP_06142336	gi|268608609|ref|ZP_06142336.1|	2.00E-70	236	55.9	beta-galactosidase [Ruminococcus flavefaciens FD-1]
Contig10613	GDE_C	3641	3096	XP_001912837	gi|171695826|ref|XP_00191	0	1238	42.5	unnamed protein product [Podospora anserina]gi|170948155|emb|CAP60319.1|

					2837.1|			unnamed protein product [Podospora anserina]
Contig24123	GDE_C	1425	896	XP_001646218	gi|156846663|ref|XP_001646218.1|	1.00E-126	463	48.6	hypothetical protein Kpol_1013p31 [Vanderwaltozyma polyspora DSM 70294]gi|156116892|gb|EDO18360.1| hypothetical protein Kpol_1013p31 [Vanderwaltozyma polyspora DSM 70294]
Contig22750	GDE_C	612	352	XP_567389	gi|58259958|ref|XP_567389.1|	2.00E-49	203	48.8	Glycogen debranching enzyme [Cryptococcus neoformans var. neoformans JEC21]gi|134116190|ref|XP_773266.1| hypothetical protein CNBJ0440 [Cryptococcus neoformans var. neoformans B-3501A]gi|50255888|gb|EAL18619.1| hypothetical protein CNBJ0440 [Cryptococcus neoformans var. neoformans B-3501A]gi|57229439|gb|AAW45872.1| Glycogen debranching enzyme, putative [Cryptococcus neoformans var. neoformans JEC21]
Contig864	GH2N	656	88	ZP_06142336	gi|268608609|ref|ZP_06142336.1|	3.00E-91	216	72.7	beta-galactosidase [Ruminococcus flavefaciens FD-1]
Contig29590	GH2N	781	435	ZP_06721189	gi|294643369|ref|ZP_06721189.1|	4.00E-70	216	57.9	glycosyl hydrolase family 2, sugar binding domain protein [Bacteroides ovatus SD CC 2a]gi|292641306|gb|EFF59504.1| glycosyl hydrolase family 2, sugar binding domain protein [Bacteroides ovatus SD CC 2a]gi|294442313|gb|EFG11125.1| glycosyl hydrolase family 2, sugar binding domain protein [Bacteroides xylanisolvens SD CC 1b]
Contig9801	GH2N	689	218	ZP_06142336	gi|268608609|ref|ZP_06142336.1|	4.00E-90	228	69.7	beta-galactosidase [Ruminococcus flavefaciens FD-1]
Contig4594	GH2N	1710	1033	YP_206064	gi|59713289|ref|YP_206064.1|	1.00E-123	558	43.2	beta-mannosidase [Vibrio fischeri ES114]gi|59481537|gb|AAW87176.1| beta-mannosidase [Vibrio fischeri ES114]
Contig23341	GH3C	892	555	AAO41704	gi|28557461|gb|AAO41704.1|	1.00E-90	302	56.0	beta-glucosidase precursor [Piromyces sp. E2]
Contig1978	GH3C	739	337	AAO41704	gi|28557461|gb|AAO41704.	5.00E-50	251	42.6	beta-glucosidase precursor [Piromyces sp. E2]

					1|			
Contig18184	GH3C	712	465	AAO41704	gi|28557461|gb|AAO41704.1|	4.00E-41	239	40.2	beta-glucosidase precursor [Piromyces sp. E2]
Contig22297	GH3C	652	247	AAO41704	gi|28557461|gb|AAO41704.1|	7.00E-25	195	36.4	beta-glucosidase precursor [Piromyces sp. E2]
Contig29961	GH3C	741	973	AAO41704	gi|28557461|gb|AAO41704.1|	8.00E-42	247	39.7	beta-glucosidase precursor [Piromyces sp. E2]
Contig4420	GH3C	796	577	AAO41704	gi|28557461|gb|AAO41704.1|	4.00E-38	234	43.2	beta-glucosidase precursor [Piromyces sp. E2]
Contig18099	GH3C	1060	474	AAO41704	gi|28557461|gb|AAO41704.1|	3.00E-24	110	47.3	beta-glucosidase precursor [Piromyces sp. E2]
Contig8004	GH3C	940	455	AAO41704	gi|28557461|gb|AAO41704.1|	5.00E-70	294	48.6	beta-glucosidase precursor [Piromyces sp. E2]
Contig4387	GH3C	707	208	CBL11427	gi|291538316|emb|CBL11427.1|	7.00E-72	234	54.7	Beta-glucosidase-related glycosidases [Roseburia intestinalis XB6B4]
Contig2300	GH3C	572	204	ABA42186	gi|76365700|gb|ABA42186.1|	6.00E-13	207	29.5	beta-glucosidase [uncultured bacterium]
Contig2789	GH3C	582	101	CBL11427	gi|291538316|emb|CBL11427.1|	6.00E-40	167	47.9	Beta-glucosidase-related glycosidases [Roseburia intestinalis XB6B4]
Contig3755	TIG	5069	3424	XP_002293728	gi|224009540|ref|XP_002293728.1|	1.00E-30	554	28.3	predicted protein [Thalassiosira pseudonana CCMP1335]gi|220970400|gb|EED88737.1| predicted protein [Thalassiosira pseudonana CCMP1335]
Contig7630	TIG	1527	508	XP_002293728	gi|224009540|ref|XP_002293728.1|	3.00E-12	498	22.3	predicted protein [Thalassiosira pseudonana CCMP1335]gi|220970400|gb|EED88737.1| predicted protein [Thalassiosira pseudonana CCMP1335]
Contig5985*	TIG	1948	2301	EEY69233	gi|262111181|gb|EEY69233.1|	9.00E-13	436	22.0	conserved hypothetical protein [Phytophthora infestans T30-4]
Contig13151	TIG	559	213	EER11220	gi|239888533|gb|EER11220.	7.00E-21	175	32.0	dynein-1-alpha heavy chain, flagellar inner arm I1 complex, putative [Perkinsus marinus

					1|			ATCC 50983]
Contig226*	DUF297	667	51	EDK40790	gi|190348350|gb|EDK40790.2|	3.00E-39	207	38.6	hypothetical protein PGUG_04888 [Pichia guilliermondii ATCC 6260]
Contig11687	DUF297	627	243	EDK40790	gi|190348350|gb|EDK40790.2|	9.00E-35	193	39.4	hypothetical protein PGUG_04888 [Pichia guilliermondii ATCC 6260]
Contig21765	DUF297	1250	511	ZP_01113306	gi|88797718|ref|ZP_01113306.1|	1.00E-39	247	40.9	putative endo alpha-1,4 polygalactosaminidase precusor [Reinekea sp. MED297]gi|88779395|gb|EAR10582.1| putative endo alpha-1,4 polygalactosaminidase precusor [Reinekea sp. MED297]
Contig260*	DUF297	1229	562	ZP_01216573	gi|90408411|ref|ZP_01216573.1|	1.00E-34	250	37.6	putative endo alpha-1,4 polygalactosaminidase [Psychromonas sp. CNPT3]gi|90310504|gb|EAS38627.1| putative endo alpha-1,4 polygalactosaminidase [Psychromonas sp. CNPT3]
Contig5652	DUF297	783	342	YP_003503987	gi|291287171|ref|YP_003503987.1|	5.00E-40	188	48.9	hypothetical protein Dacet_1259 [Denitrovibrio acetiphilus DSM 12809]gi|290884331|gb|ADD68031.1| TM1410 hypothetical-related protein [Denitrovibrio acetiphilus DSM 12809]
Contig3727	DUF297	513	186	YP_340939	gi|77361364|ref|YP_340939.1|	9.00E-24	172	35.5	endo alpha-1,4 polygalactosaminidase precusor [Pseudoalteromonas haloplanktis TAC125]gi|76876275|emb|CAI87497.1| putative endo alpha-1,4 polygalactosaminidase precusor [Pseudoalteromonas haloplanktis TAC125]
Contig8184	DUF297	938	361	YP_156978	gi|56461697|ref|YP_156978.1|	2.00E-26	221	34.8	polysaccharide hydrolase related to endo alpha-1,4 polygalactosaminidase [Idiomarina loihiensis L2TR]gi|56180707|gb|AAV83429.1| Predicted polysaccharide hydrolase related to endo alpha-1,4 polygalactosaminidase [Idiomarina loihiensis L2TR]
Contig21383	DUF297	517	169	ZP_01216573	gi|90408411|ref|ZP_01216573.1|	3.00E-29	153	47.1	putative endo alpha-1,4 polygalactosaminidase [Psychromonas sp. CNPT3]gi|90310504|gb|EAS38627.1| putative endo alpha-1,4 polygalactosaminidase [Psychromonas sp. CNPT3]

Contig12037	DUF297	673	570	YP_156978	gi|56461697|ref|YP_156978.1|	3.00E-19	212	32.1	polysaccharide hydrolase related to endo alpha-1,4 polygalactosaminidase [Idiomarina loihiensis L2TR]gi|56180707|gb|AAV83429.1| Predicted polysaccharide hydrolase related to endo alpha-1,4 polygalactosaminidase [Idiomarina loihiensis L2TR]
Contig16432	DUF297	543	199	YP_003503987	gi|291287171|ref|YP_003503987.1|	8.00E-27	159	40.3	hypothetical protein Dacet_1259 [Denitrovibrio acetiphilus DSM 12809]gi|290884331|gb|ADD68031.1| TM1410 hypothetical-related protein [Denitrovibrio acetiphilus DSM 12809]
Contig17499	Esterase	1143	674	ACZ98648	gi|280977861|gb|ACZ98648.1|	1.00E-60	277	45.8	esterase [Cellulosilyticum ruminicola]
Contig1579	Esterase	906	27	ZP_06142343	gi|268608616|ref|ZP_06142343.1|	4.00E-73	276	52.2	glycoside hydrolase family 11 [Ruminococcus flavefaciens FD-1]
Contig8173	Esterase	695	317	AAF70241	gi|7839348|gb|AAF70241.1|AF164351_1	3.00E-80	231	62.3	feruloyl esterase A [Orpinomyces sp. PC-2]
Contig18676	Esterase	1525	1044	AAF70241	gi|7839348|gb|AAF70241.1|AF164351_1	1.00E-173	461	64.2	feruloyl esterase A [Orpinomyces sp. PC-2]
Contig9000	Esterase	1201	727	AAF70241	gi|7839348|gb|AAF70241.1|AF164351_1	1.00E-143	360	68.3	feruloyl esterase A [Orpinomyces sp. PC-2]
Contig2667	Esterase	1461	213	AAF70241	gi|7839348|gb|AAF70241.1|AF164351_1	1.00E-175	452	64.8	feruloyl esterase A [Orpinomyces sp. PC-2]
Contig8700	Esterase	571	322	ZP_04540353	gi|237709872|ref|ZP_04540353.1|	2.00E-74	179	72.6	prolyl oligopeptidase [Bacteroides sp. 9_1_42FAA]gi|237725458|ref|ZP_04555939.1| prolyl oligopeptidase [Bacteroides sp. D4]gi|265753524|ref|ZP_06088879.1| prolyl oligopeptidase [Bacteroides sp. 3_1_33FAA]gi|229436145|gb|EEO46222.1| prolyl oligopeptidase [Bacteroides sp. D4]gi|229455965|gb|EEO61686.1| prolyl oligopeptidase [Bacteroides sp. 9_1_42FAA]gi|263235238|gb|EEZ20762.1| prolyl oligopeptidase [Bacteroides sp.

									3_1_33FAA]
Contig3971	Esterase	622	356	ACZ98648	gi|280977861|gb|ACZ98648.1|	3.00E-24	140	42.9	esterase [Cellulosilyticum ruminicola]
Contig21493	DUF303	1232	2694	ZP_03458534	gi|218129730|ref|ZP_03458534.1|	3.00E-69	235	51.5	hypothetical protein BACEGG_01309 [Bacteroides eggerthii DSM 20697]gi|217988142|gb|EEC54466.1| hypothetical protein BACEGG_01309 [Bacteroides eggerthii DSM 20697]
Contig22188	DUF303	592	631	ZP_04545878	gi|237715397|ref|ZP_04545878.1|	4.00E-60	194	58.2	glycoside hydrolase family 43 [Bacteroides sp. D1]gi|262405241|ref|ZP_06081791.1| conserved hypothetical protein [Bacteroides sp. 2_1_22]gi|294646993|ref|ZP_06724610.1| conserved hypothetical protein [Bacteroides ovatus SD CC 2a]gi|229444706|gb|EEO50497.1| glycoside hydrolase family 43 [Bacteroides sp. D1]gi|262356116|gb|EEZ05206.1| conserved hypothetical protein [Bacteroides sp. 2_1_22]gi|292637664|gb|EFF56065.1| conserved hypothetical protein [Bacteroides ovatus SD CC 2a]gi|294446393|gb|EFG15017.1| conserved hypothetical protein [Bacteroides xylanisolvens SD CC 1b]
Contig21894	DUF303	708	164	AAB69090	gi|2231243|gb|AAB69090.1|	1.00E-96	235	74.0	acetylxylan esterase [Neocallimastix patriciarum]
Contig541	DUF303	867	900	ZP_03016604	gi|189467819|ref|ZP_03016604.1|	1.00E-90	257	62.6	hypothetical protein BACINT_04211 [Bacteroides intestinalis DSM 17393]gi|189436083|gb|EDV05068.1| hypothetical protein BACINT_04211 [Bacteroides intestinalis DSM 17393]
Contig25491	DUF303	1118	560	AAB69090	gi|2231243|gb|AAB69090.1|	1.00E-98	249	70.3	acetylxylan esterase [Neocallimastix patriciarum]
Contig685	DUF303	801	568	ZP_04545878	gi|237715397|ref|ZP_04545878.1|	6.00E-84	252	60.7	glycoside hydrolase family 43 [Bacteroides sp. D1]gi|262405241|ref|ZP_06081791.1| conserved hypothetical protein [Bacteroides sp. 2_1_22]gi|294646993|ref|ZP_06724610.1| conserved hypothetical protein [Bacteroides ovatus SD CC 2a]gi|229444706|gb|EEO50497.1| glycoside

									hydrolase family 43 [Bacteroides sp. D1]gi|262356116|gb|EEZ05206.1| conserved hypothetical protein [Bacteroides sp. 2_1_22]gi|292637664|gb|EFF56065.1| conserved hypothetical protein [Bacteroides ovatus SD CC 2a]gi|294446393|gb|EFG15017.1| conserved hypothetical protein [Bacteroides xylanisolvens SD CC 1b]
Contig30565	DUF303	865	316	AAC14690	gi|3080749|gb|AAC14690.1|	1.00E-102	245	73.1	acetyl xylan esterase A [Orpinomyces sp. PC-2]
Contig12482	DUF303	1063	754	AAC14690	gi|3080749|gb|AAC14690.1|	3.00E-88	276	58.0	acetyl xylan esterase A [Orpinomyces sp. PC-2]
Contig29194	DUF303	906	569	AAC14690	gi|3080749|gb|AAC14690.1|	1.00E-116	252	79.4	acetyl xylan esterase A [Orpinomyces sp. PC-2]
Contig22752	DUF303	972	534	AAB69090	gi|2231243|gb|AAB69090.1|	1.00E-75	324	46.0	acetylxylan esterase [Neocallimastix patriciarum]
Contig3819	DUF303	605	455	AAB69090	gi|2231243|gb|AAB69090.1|	3.00E-65	165	71.5	acetylxylan esterase [Neocallimastix patriciarum]
Contig29175	DUF303	544	425	AAB69090	gi|2231243|gb|AAB69090.1|	2.00E-67	175	70.9	acetylxylan esterase [Neocallimastix patriciarum]
Contig840	COesterase	1670	1250	CBK73419	gi|291518198|emb|CBK73419.1|	1.00E-158	531	53.3	Carboxylesterase type B [Butyrivibrio fibrisolvens 16/4]
Contig1868	COesterase	1029	35	NP_149214	gi|15004754|ref|NP_149214.1|	1.00E-66	343	43.1	Para-nitrobenzyl esterase, a/b hydrolase [Clostridium acetobutylicum ATCC 824]gi|14994366|gb|AAK76796.1|AE001438_49 Para-nitrobenzyl esterase, a/b hydrolase [Clostridium acetobutylicum ATCC 824]
Contig1174	COesterase	557	85	ZP_05497811	gi|256757073|ref|ZP_05497811.1|	8.00E-22	139	41.7	Alpha/beta hydrolase fold-3 domain protein [Clostridium papyrosolvens DSM 2782]gi|256744033|gb|EEU57199.1| Alpha/beta hydrolase fold-3 domain protein [Clostridium papyrosolvens DSM 2782]
Contig257	COesterase	743	130	YP_079964	gi|52081173|ref|YP_079964.1|	2.00E-17	118	40.7	para-nitrobenzyl esterase (intracellular esterase B) [Bacillus licheniformis ATCC 14580]gi|52786556|ref|YP_092385.1| hypothetical protein BLi02821 [Bacillus licheniformis ATCC 14580]gi|52004384|gb|AAU24326.1| para-nitrobenzyl esterase (intracellular esterase B)

									[Bacillus licheniformis ATCC 14580]gi|52349058|gb|AAU41692.1| putative protein [Bacillus licheniformis ATCC 14580]
Contig28300	Pfam-B_184	2733	1434	ZP_04741973	gi|240143372|ref|ZP_04741973.1|	0	915	44.2	conserved hypothetical protein [Roseburia intestinalis L1-82]gi|257204641|gb|EEV02926.1| conserved hypothetical protein [Roseburia intestinalis L1-82]
Contig7752*	Pfam-B_184	2512	2505	ZP_05498108	gi|256757388|ref|ZP_05498108.1|	1.00E-112	718	34.8	conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]gi|256743722|gb|EEU56906.1| conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]
Contig29412	Pfam-B_184	1336	1514	ZP_05498109	gi|256757389|ref|ZP_05498109.1|	6.00E-74	286	46.5	conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]gi|256743723|gb|EEU56907.1| conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]
Contig19696	Pfam-B_184	1048	476	ZP_05498109	gi|256757389|ref|ZP_05498109.1|	4.00E-72	286	47.6	conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]gi|256743723|gb|EEU56907.1| conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]
Contig1848	Pfam-B_184	1400	62	ZP_05498108	gi|256757388|ref|ZP_05498108.1|	9.00E-65	469	34.8	conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]gi|256743722|gb|EEU56906.1| conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]
Contig21105	Pfam-B_184	1501	635	ZP_05498108	gi|256757388|ref|ZP_05498108.1|	2.00E-59	499	31.9	conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]gi|256743722|gb|EEU56906.1| conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]
Contig7431	Pfam-B_184	524	245	ZP_05498109	gi|256757389|ref|ZP_05498109.1|	2.00E-37	174	47.1	conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]gi|256743723|gb|EEU56907.1| conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]
Contig31355	Pfam-B_184	541	392	ZP_05498109	gi|256757389|ref|ZP_05498109.1|	5.00E-37	174	44.3	conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]gi|256743723|gb|EEU56907.1|

									conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]
Contig25356	Pfam-B_184	516	177	ZP_05498109	gi|256757389|ref|ZP_05498109.1|	1.00E-33	170	41.2	conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]gi|256743723|gb|EEU56907.1| conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]
Contig21626*	Pfam-B_1434	1213	5338	ZP_02024895	gi|154482447|ref|ZP_02024895.1|	1.00E-124	374	58.3	hypothetical protein EUBVEN_00114 [Eubacterium ventriosum ATCC 27560]gi|149736696|gb|EDM52582.1| hypothetical protein EUBVEN_00114 [Eubacterium ventriosum ATCC 27560]
Contig10848	Pfam-B_1434	1065	356	ZP_02024895	gi|154482447|ref|ZP_02024895.1|	1.00E-116	355	57.5	hypothetical protein EUBVEN_00114 [Eubacterium ventriosum ATCC 27560]gi|149736696|gb|EDM52582.1| hypothetical protein EUBVEN_00114 [Eubacterium ventriosum ATCC 27560]
Contig5554	Pfam-B_2673	2133	873	AAB69091	gi|2231245|gb|AAB69091.1|	2.00E-84	361	47.4	acetylxylan esterase [Neocallimastix patriciarum]
Contig1179	Pfam-B_2673	1395	64	ZP_06142663	gi|268608936|ref|ZP_06142663.1|	1.00E-124	462	52.4	GDSL family lipase [Ruminococcus flavefaciens FD-1]
Contig21582	Pfam-B_2673	817	503	AAB69091	gi|2231245|gb|AAB69091.1|	2.00E-55	262	44.3	acetylxylan esterase [Neocallimastix patriciarum]
Contig9227	Pfam-B_2673	1032	371	AAB69091	gi|2231245|gb|AAB69091.1|	1.00E-178	344	91.9	acetylxylan esterase [Neocallimastix patriciarum]
Contig23673	Pfam-B_2673	1259	734	AAB69091	gi|2231245|gb|AAB69091.1|	5.00E-80	357	45.9	acetylxylan esterase [Neocallimastix patriciarum]
Contig25553	Pfam-B_2673	770	359	AAB69091	gi|2231245|gb|AAB69091.1|	3.00E-53	250	44.8	acetylxylan esterase [Neocallimastix patriciarum]
Contig22864	Pfam-B_8046	761	712	ADE81126	gi|294471737|gb|ADE81126.1|	3.00E-58	254	48.8	lipase/acylhydrolase [Prevotella ruminicola 23]
Contig11936	Pfam-B_8046	508	237	ADE81126	gi|294471737|gb|ADE81126.1|	5.00E-40	194	49.0	lipase/acylhydrolase [Prevotella ruminicola 23]
Contig26846	Pfam-B_8046	700	554	XP_001829175	gi|169844909|ref|XP_001829175.1|	2.00E-36	214	37.4	hypothetical protein CC1G_01855 [Coprinopsis cinerea okayama7#130]gi|116509915|gb|EAU92810.1| hypothetical protein CC1G_01855 [Coprinopsis cinerea okayama7#130]

Contig285*	Pfam-B_8046	522	166	XP_002394825	gi|238599237|ref|XP_002394825.1|	1.00E-28	175	41.1	hypothetical protein MPER_05224 [Moniliophthora perniciosa FA553]gi|215464484|gb|EEB95755.1| hypothetical protein MPER_05224 [Moniliophthora perniciosa FA553]
Contig7303	CE15	781	252	YP_677850	gi|110637643|ref|YP_677850.1|	2.00E-86	259	60.6	beta-xylosidase/endoglucanase [Cytophaga hutchinsonii ATCC 33406]gi|110280324|gb|ABG58510.1| CHU large protein; beta-xylosidase/endoglucanase [Cytophaga hutchinsonii ATCC 33406]
Contig11541	CE15	755	277	YP_677850	gi|110637643|ref|YP_677850.1|	2.00E-79	252	56.7	beta-xylosidase/endoglucanase [Cytophaga hutchinsonii ATCC 33406]gi|110280324|gb|ABG58510.1| CHU large protein; beta-xylosidase/endoglucanase [Cytophaga hutchinsonii ATCC 33406]
Contig29100	PL6	632	805	ZP_02862960	gi|169335767|ref|ZP_02862960.1|	3.00E-08	194	27.8	hypothetical protein ANASTE_02192 [Anaerofustis stercorihominis DSM 17244]gi|169258505|gb|EDS72471.1| hypothetical protein ANASTE_02192 [Anaerofustis stercorihominis DSM 17244]
NODE_5487_length_1436_cov_7.160863	PL6	1472	1045	ZP_01852968	gi|149174341|ref|ZP_01852968.1|	3.00E-11	419	24.3	Parallel beta-helix repeat protein [Planctomyces maris DSM 8797]gi|148846886|gb|EDL61222.1| Parallel beta-helix repeat protein [Planctomyces maris DSM 8797]
Contig22996	PL11_C	795	425	ZP_05497496	gi|256756746|ref|ZP_05497496.1|	4.00E-62	252	49.2	cellulosome protein dockerin type I [Clostridium papyrosolvens DSM 2782]gi|256744398|gb|EEU57552.1| cellulosome protein dockerin type I [Clostridium papyrosolvens DSM 2782]
Contig650	PL09	836	851	ZP_04803861	gi|242259131|ref|ZP_04803861.1|	4.00E-43	251	38.6	pectate lyase [Clostridium cellulovorans 743B]gi|242228465|gb|EES31701.1| pectate lyase [Clostridium cellulovorans 743B]
Contig7694	PL09	1435	2737	YP_003430993	gi|288905771|ref|YP_003430993.1|	1.00E-71	416	39.2	Putative pectate lyase related protein, polysaccharide lyase family [Streptococcus gallolyticus UCN34]gi|288732497|emb|CBI14069.1| Putative pectate lyase related protein, polysaccharide lyase family [Streptococcus gallolyticus UCN34]
Contig2460	PL09	1392	83	YP_003430	gi|288905771|	6.00E	440	39.1	Putative pectate lyase related protein,

				993	ref|YP_00343 0993.1|	-76			polysaccharide lyase family [Streptococcus gallolyticus UCN34]gi|288732497|emb|CBI14069.1| Putative pectate lyase related protein, polysaccharide lyase family [Streptococcus gallolyticus UCN34]
Contig8180	PL09	1249	514	YP_003430 993	gi|288905771| ref|YP_00343 0993.1|	2.00E -77	416	40.6	Putative pectate lyase related protein, polysaccharide lyase family [Streptococcus gallolyticus UCN34]gi|288732497|emb|CBI14069.1| Putative pectate lyase related protein, polysaccharide lyase family [Streptococcus gallolyticus UCN34]
Contig6102*	PL09	1275	723	YP_003430 993	gi|288905771| ref|YP_00343 0993.1|	8.00E -75	400	39.8	Putative pectate lyase related protein, polysaccharide lyase family [Streptococcus gallolyticus UCN34]gi|288732497|emb|CBI14069.1| Putative pectate lyase related protein, polysaccharide lyase family [Streptococcus gallolyticus UCN34]
Contig13767	PL09	702	338	YP_003430 993	gi|288905771| ref|YP_00343 0993.1|	3.00E -48	232	47.0	Putative pectate lyase related protein, polysaccharide lyase family [Streptococcus gallolyticus UCN34]gi|288732497|emb|CBI14069.1| Putative pectate lyase related protein, polysaccharide lyase family [Streptococcus gallolyticus UCN34]
Contig22095	PL09	732	456	YP_003430 993	gi|288905771| ref|YP_00343 0993.1|	1.00E -60	256	48.8	Putative pectate lyase related protein, polysaccharide lyase family [Streptococcus gallolyticus UCN34]gi|288732497|emb|CBI14069.1| Putative pectate lyase related protein, polysaccharide lyase family [Streptococcus gallolyticus UCN34]
Contig18967	PL09	523	174	YP_002505 335	gi|220928426| ref|YP_00250 5335.1|	2.00E -30	136	55.9	Carbohydrate-binding family 9 [Clostridium cellulolyticum H10]gi|219998754|gb|ACL75355.1| Carbohydrate-binding family 9 [Clostridium cellulolyticum H10]
Contig29157	Swollenin	1221	1032	CAB92328	gi|8052455|emb|CAB9232	1.00E -131	388	56.7	swollenin [Hypocrea jecorina]

					8.1|			
Contig4246	Swollenin	722	641	BAI83433	gi|291461597|dbj|BAI83433.1|	3.00E-77	231	59.3	swollenin like protein [Aspergillus fumigatus]
Contig31130	Swollenin	689	371	ABV57767	gi|157488002|gb|ABV57767.1|	2.00E-73	218	57.8	swollenin [Hypocrea pseudokoningii]
Contig5172	Swollenin	595	225	ABV57767	gi|157488002|gb|ABV57767.1|	4.00E-59	183	59.0	swollenin [Hypocrea pseudokoningii]
Contig22476	Swollenin	509	365	ABV57767	gi|157488002|gb|ABV57767.1|	6.00E-33	160	47.5	swollenin [Hypocrea pseudokoningii]
Contig14372	GH95	817	162	ZP_03016604	gi|189467819|ref|ZP_03016604.1|	6.00E-97	263	63.9	hypothetical protein BACINT_04211 [Bacteroides intestinalis DSM 17393]gi|189436083|gb|EDV05068.1| hypothetical protein BACINT_04211 [Bacteroides intestinalis DSM 17393]
Contig10206	GH95	702	116	ZP_03458534	gi|218129730|ref|ZP_03458534.1|	1.00E-82	229	60.7	hypothetical protein BACEGG_01309 [Bacteroides eggerthii DSM 20697]gi|217988142|gb|EEC54466.1| hypothetical protein BACEGG_01309 [Bacteroides eggerthii DSM 20697]
Contig23822*	GH87	2301	2654	YP_001558424	gi|160879456|ref|YP_001558424.1|	0	593	60.9	APHP domain-containing protein [Clostridium phytofermentans ISDg]gi|160428122|gb|ABX41685.1| APHP domain protein [Clostridium phytofermentans ISDg]
Contig6769	GH87	891	255	YP_001558424	gi|160879456|ref|YP_001558424.1|	1.00E-102	180	70.6	APHP domain-containing protein [Clostridium phytofermentans ISDg]gi|160428122|gb|ABX41685.1| APHP domain protein [Clostridium phytofermentans ISDg]
Contig13875	GH87	1351	1222	YP_001558424	gi|160879456|ref|YP_001558424.1|	1.00E-130	356	65.4	APHP domain-containing protein [Clostridium phytofermentans ISDg]gi|160428122|gb|ABX41685.1| APHP domain protein [Clostridium phytofermentans ISDg]
Contig2118*	GH74	2396	3004	ZP_06145570	gi|268611843|ref|ZP_06145570.1|	0	763	55.0	endo-1,4-beta-glucanase/xyloglucanase, putative, gly74A [Ruminococcus flavefaciens FD-1]

Contig28721	GH74	1428	894	ZP_06145570	gi|268611843|ref|ZP_06145570.1|	1.00E-128	520	47.3	endo-1,4-beta-glucanase/xyloglucanase, putative, gly74A [Ruminococcus flavefaciens FD-1]
Contig24305	GH74	1399	232	ZP_06145570	gi|268611843|ref|ZP_06145570.1|	1.00E-146	486	55.1	endo-1,4-beta-glucanase/xyloglucanase, putative, gly74A [Ruminococcus flavefaciens FD-1]
Contig3464	GH74	769	380	ZP_06145570	gi|268611843|ref|ZP_06145570.1|	2.00E-64	237	54.4	endo-1,4-beta-glucanase/xyloglucanase, putative, gly74A [Ruminococcus flavefaciens FD-1]
Contig23418	GH74	538	108	ZP_06145570	gi|268611843|ref|ZP_06145570.1|	1.00E-40	185	48.1	endo-1,4-beta-glucanase/xyloglucanase, putative, gly74A [Ruminococcus flavefaciens FD-1]
Contig5455	GH74	597	197	ZP_06145570	gi|268611843|ref|ZP_06145570.1|	2.00E-45	203	45.8	endo-1,4-beta-glucanase/xyloglucanase, putative, gly74A [Ruminococcus flavefaciens FD-1]
Contig22447	CBM29	1580	368	AAK20910	gi|13446353|gb|AAK20910.1|	4.00E-08	347	24.5	non-catalytic protein 1 [Piromyces equi]
Contig29519	CBM29	553	857	AAK20910	gi|13446353|gb|AAK20910.1|	2.00E-57	185	60.0	non-catalytic protein 1 [Piromyces equi]
Contig7989	CBM29	549	77	AAK20910	gi|13446353|gb|AAK20910.1|	4.00E-46	190	48.9	non-catalytic protein 1 [Piromyces equi]
Contig2640	CBM36	1288	497	YP_001038591	gi|125974681|ref|YP_001038591.1|	1.00E-156	425	62.1	carbohydrate-binding family 6 protein [Clostridium thermocellum ATCC 27405]gi|256004120|ref|ZP_05429104.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM 2360]gi|281418847|ref|ZP_06249866.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]gi|125714906|gb|ABN53398.1| Carbohydrate binding family 6 [Clostridium thermocellum ATCC 27405]gi|255991868|gb|EEU01966.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM 2360]gi|281407931|gb|EFB38190.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]

Contig22009	CBM36	1443	1537	ACZ98594	gi|280977753|gb|ACZ98594.1|	1.00E-136	462	51.9	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]

Note: * Possible Full Length Gene

Table S.2 Muskoxen rumen metatranscriptome contigs (≥500 bp) that have two or more distinct putative CAZy modules.

Contig number	Domains	Length	Number of Reads	Accession	GI	E-value	HSP Length	HSP Id%	Hit Description
Contig3930	GH3; GH6; CBM10	3927	6647	AAD02028	gi|4104400|gb|AAD02028.1|	0	468	82.7	exocellobiohydrolase precursor [Piromyces rhizinflatus]
Contig22419*	GH3; CBM10	2647	2360	AAO41704	gi|28557461|gb|AAO41704.1|	0	875	73.8	beta-glucosidase precursor [Piromyces sp. E2]
Contig28965	GH3; CBM10	2585	3385	AAO41704	gi|28557461|gb|AAO41704.1|	0	863	80.4	beta-glucosidase precursor [Piromyces sp. E2]
Contig1220	Cellulase; CBM10	2239	417	CAB92326	gi|8052316|emb|CAB92326.1|	0	565	67.6	endoglucanase 5A [Piromyces equi]
Contig1475	Cellulase; CBM10	1612	1800	AAL83749	gi|19070186|gb|AAL83749.1|	1.00E-107	347	51.6	endo-beta-1,4-glucanase [Paenibacillus sp. KCTC8848P]
Contig5338*	Cellulase; CBM10	1737	2114	AAL83749	gi|19070186|gb|AAL83749.1|	1.00E-106	345	51.3	endo-beta-1,4-glucanase [Paenibacillus sp. KCTC8848P]
Contig23559	Cellulase; CBM10	1658	1110	NP_521723	gi|17548383|ref|NP_521723.1|	2.00E-66	312	42.3	endoglucanase precursor (endo-1,4-BETA-glucanase) protein [Ralstonia solanacearum GMI1000]gi|17430629|emb|CAD17313.1| endoglucanase precursor (endo-1,4-beta-glucanase)(cellulase) protein [Ralstonia solanacearum GMI1000]
Contig31560	Cellulase; CBM10	1206	881	AAL83749	gi|19070186|gb|AAL83749.1|	2.00E-59	204	51.5	endo-beta-1,4-glucanase [Paenibacillus sp. KCTC8848P]
Contig23055	Cellulase; CBM10	787	422	CBL17363	gi|291544254|emb|CBL17363.1|	7.00E-45	148	58.1	Endoglucanase [Ruminococcus sp. 18P13]

Contig30802	Cellulase; CBM10	1038	910	P23548	gi|1346224|sp|P23548.2|GUN_PAEPO	2.00E-28	137	43.1	RecName: Full=Endoglucanase; AltName: Full=Endo-1,4-beta-glucanase; AltName: Full=Cellulasegi|143271|gb|AAA22631.1| endo-beta-1,4-glucanase [Paenibacillus polymyxa]
Contig30795	Cellulase; CBM10	612	1023	CAB92326	gi|8052316|emb|CAB92326.1|	3.00E-77	203	62.6	endoglucanase 5A [Piromyces equi]
Contig22047	GH6; CBM10	1921	44342	AAL01211	gi|15529294|gb|AAL01211.1|AF177204_1	0	480	83.1	cellobiohydrolase II-like cellulase CelH [Orpinomyces sp. PC-2]
Contig111*	GH6; CBM10	1476	1725	AAB92678	gi|1813484|gb|AAB92678.1|	0	463	76.2	cellulase A [Orpinomyces sp. PC-2]
Contig1523	GH6; CBM1	1532	63	AAR08200	gi|38018276|gb|AAR08200.1|	1.00E-178	319	96.6	CelA [Neocallimastix frontalis]
Contig29791	GH6; CBM10	1420	2244	ABY52799	gi|164375387|gb|ABY52799.1|	1.00E-125	287	74.9	1,4-beta-D-glucan-cellobiohydrolase [Piromyces rhizinflatus]
Contig3027	GH6; CBM10	1312	233	AAD51054	gi|9943835|gb|AAD51054.2|AF174361_1	1.00E-174	433	69.1	exocellobiohydrolase Cbh120 [Piromyces rhizinflatus]
Contig1794	GH6; CBM10	1107	4033	ACX32999	gi|260169862|gb|ACX32999.1|	1.00E-115	372	58.6	1,4-beta-glucanase [Piromyces sp. BTrP1]
Contig7415	GH6; CBM10	919	388	AAB92678	gi|1813484|gb|AAB92678.1|	1.00E-120	308	70.1	cellulase A [Orpinomyces sp. PC-2]
Contig28654	GH6; GH3C	2530	7263	AAD02028	gi|4104400|gb|AAD02028.1|	1.00E-113	252	78.2	exocellobiohydrolase precursor [Piromyces rhizinflatus]
Contig1597	GH6; CBM10	1020	10347	AAL92497	gi|29465670|gb|AAL92497.1|	1.00E-117	309	65.4	exoglucanase Cel6A [Piromyces sp. E2]
Contig30772	GH6; CBM10	1059	784	AAD51054	gi|9943835|gb|AAD51054.2|AF174361_1	1.00E-129	350	66.3	exocellobiohydrolase Cbh120 [Piromyces rhizinflatus]
Contig2418*	GH6; CBM10	998	1144	AAD02028	gi|4104400|gb|AAD02028.1|	1.00E-125	335	66.0	exocellobiohydrolase precursor [Piromyces rhizinflatus]
Contig9323	GH6; CBM10	656	152	AAP33843	gi|32395719|gb|AAP33843.	1.00E-52	125	80.0	hybrid 1,4-beta-glucanase [synthetic construct]

					1|			
Contig339	GH6; CBM10	784	519	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	1.00E-83	253	60.1	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig21750	GH6; CBM10	641	264	AAM94167	gi|33620325|gb|AAM94167.1|	3.00E-87	199	78.4	cellulosomal glycoside hydrolase family 6 exoglucanase Cel6A [Piromyces equi]
Contig22729	GH6; CBM10	632	1925	AAM94167	gi|33620325|gb|AAM94167.1|	2.00E-60	181	65.2	cellulosomal glycoside hydrolase family 6 exoglucanase Cel6A [Piromyces equi]
Contig29107	GH6; CBM10	719	2591	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	3.00E-76	224	64.3	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig6887	GH6; CBM10	657	542	AAL01211	gi|15529294|gb|AAL01211.1|AF177204_1	1.00E-68	202	63.9	cellobiohydrolase II-like cellulase CelH [Orpinomyces sp. PC-2]
Contig18293	GH6; CBM10	714	360	AAD02028	gi|4104400|gb|AAD02028.1|	1.00E-33	81	87.7	exocellobiohydrolase precursor [Piromyces rhizinflatus]
Contig9*	GH6; CBM10	1423	159	AAL92497	gi|29465670|gb|AAL92497.1|	4.00E-69	201	63.2	exoglucanase Cel6A [Piromyces sp. E2]
Contig24462	GH6; CBM10	532	240	AAL92497	gi|29465670|gb|AAL92497.1|	1.00E-58	179	63.1	exoglucanase Cel6A [Piromyces sp. E2]
Contig29682	GH6; CBM10	695	821	AAM94167	gi|33620325|gb|AAM94167.1|	4.00E-67	198	66.2	cellulosomal glycoside hydrolase family 6 exoglucanase Cel6A [Piromyces equi]
Contig30051	GH6; CBM10	573	1216	AAD02028	gi|4104400|gb|AAD02028.1|	3.00E-59	228	50.4	exocellobiohydrolase precursor [Piromyces rhizinflatus]
Contig31138	GH6; CBM10	637	333	AAL01211	gi|15529294|gb|AAL01211.1|AF177204_1	2.00E-71	214	64.5	cellobiohydrolase II-like cellulase CelH [Orpinomyces sp. PC-2]
Contig29495	GH6; CBM10	728	3299	AAD02028	gi|4104400|gb|AAD02028.1|	3.00E-63	229	55.5	exocellobiohydrolase precursor [Piromyces rhizinflatus]
Contig21518	GH8; CBM10	1799	4366	YP_003248565	gi|261414882|ref|YP_003248565.1|	0	475	73.5	glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes

									S85]gi|261371338|gb|ACX74083.1| glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig22627*	GH8; GH11; CBM10	3510	5526	YP_003248565	gi|261414882|ref|YP_003248565.1|	0	432	74.5	glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261371338|gb|ACX74083.1| glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig29700*	GH8; CBM10	1784	3464	YP_003248565	gi|261414882|ref|YP_003248565.1|	0	429	76.5	glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261371338|gb|ACX74083.1| glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig24688*	GH8; CBM10	2005	6086	YP_003248565	gi|261414882|ref|YP_003248565.1|	0	434	75.3	glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261371338|gb|ACX74083.1| glycoside hydrolase family 8 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig21257	GH9; CBM10	2365	1114	AAM81967	gi|21929669|gb|AAM81967.1|AF459453_1	0	783	63.7	cellulase Cel9A precursor [Piromyces sp. E2]
Contig21102	GH9; CBM10	2268	956	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	765	62.6	cellulase Cel9A precursor [Piromyces sp. E2]
Contig21532	GH9; CBM10	2325	2223	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	771	63.0	cellulase Cel9A precursor [Piromyces sp. E2]
Contig29363	GH9; CBM10	2220	2267	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	749	78.0	cellulase Cel9A precursor [Piromyces sp. E2]

Contig1941*	GH9; CBM10	2386	2616	AAP30753	gi|30315047|gb|AAP30753.1|	0	663	81.1	cellulosomal glycoside hydrolase family 9 endoglucanase Cel9B [Piromyces sp. E2]
Contig3953	GH9; CBM10	1867	1195	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	1.00E-112	475	45.9	cellulase Cel9A precursor [Piromyces sp. E2]
Contig8450	GH9; CBM10	2358	1807	AAP30753	gi|30315047|gb|AAP30753.1|	0	666	59.5	cellulosomal glycoside hydrolase family 9 endoglucanase Cel9B [Piromyces sp. E2]
Contig22760	GH9; CBM10	1655	724	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	5.00E-94	425	44.0	cellulase Cel9A precursor [Piromyces sp. E2]
Contig4358	GH9; CBM10	2239	1281	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	8.00E-88	507	39.8	cellulase Cel9A precursor [Piromyces sp. E2]
Contig436*	GH9; CBM3	3099	161	CAL91976	gi|218081351|emb|CAL91976.1|	0	621	75.2	cellulase [Epidinium ecaudatum]
Contig3929	GH9; CBM10	2170	1186	AAP30753	gi|30315047|gb|AAP30753.1|	0	607	52.9	cellulosomal glycoside hydrolase family 9 endoglucanase Cel9B [Piromyces sp. E2]
Contig21739	GH9; CBM10	2073	1407	AAM81966	gi|21929667|gb|AAM81966.1|AF459452_1	0	685	73.9	cellulase Cel9A precursor [Piromyces sp. E2]
Contig7660	GH9; CBM10	1425	1524	AAP30753	gi|30315047|gb|AAP30753.1|	1.00E-169	373	75.1	cellulosomal glycoside hydrolase family 9 endoglucanase Cel9B [Piromyces sp. E2]
Contig8733	GH9; CBM3	855	898	CAL91976	gi|218081351|emb|CAL91976.1|	1.00E-154	285	88.1	cellulase [Epidinium ecaudatum]
Contig1701	GH9; CBM3	797	274	CAL91976	gi|218081351|emb|CAL91976.1|	1.00E-152	264	95.8	cellulase [Epidinium ecaudatum]
Contig30777	GH9; CBM10	989	566	AAP30753	gi|30315047|gb|AAP30753.1|	2.00E-57	289	41.5	cellulosomal glycoside hydrolase family 9 endoglucanase Cel9B [Piromyces sp. E2]
Contig240	GH10; CBM4_9	1618	248	CAB65753	gi|6692066|e	0	514	94.9	xylanase 10B precursor

					mb|CAB65753.1|			[Polyplastron multivesiculatum]
Contig28860	GH10; CBM1	3168	2175	AAB30669	gi|560649|gb|AAB30669.1|	1.00E-157	318	82.7	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig21350	GH10; Ricin_B_lectin	1439	2576	BAC57894	gi|28569972|dbj|BAC57894.1|	0	478	94.4	xylanase xynA [Epidinium caudatum]
Contig29712	GH10; CBM1	1493	1531	AAB30669	gi|560649|gb|AAB30669.1|	1.00E-129	309	70.2	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig24573	GH10; CBM10	1138	548	AAB30669	gi|560649|gb|AAB30669.1|	1.00E-120	268	74.6	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig85	GH10; CBM10	1125	484	AAB30669	gi|560649|gb|AAB30669.1|	7.00E-96	246	65.4	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig29849	GH10; CBM29	2295		AAB30669	gi|560649|gb|AAB30669.1|	6.00E-43	318	31.8	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig28853	GH10; Ricin_B_lectin	1071	2539	BAC57894	gi|28569972|dbj|BAC57894.1|	1.00E-171	341	85.9	xylanase xynA [Epidinium caudatum]
Contig22204*	GH10; CBM10	653	352	AAB30669	gi|560649|gb|AAB30669.1|	1.00E-31	102	63.7	Xylanase B; XYLB [Neocallimastix patriciarum]
Contig28800	GH10; CBM10	751	622	AAB69092	gi|2231247|gb|AAB69092.1|	3.00E-28	83	61.4	acetylxylan esterase [Neocallimastix patriciarum]
Contig3245	GH10; CBM10	703	579	AAB69092	gi|2231247|gb|AAB69092.1|	2.00E-28	79	63.3	acetylxylan esterase [Neocallimastix patriciarum]
Contig11078	GH10; CBM10	1305	562	AAQ10005	gi|33329210|gb|AAQ10005.1|	1.00E-115	291	70.1	acetylxylan esterase [Neocallimastix patriciarum]
Contig19309	GH11; CBM10	1830	1583	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	0	619	58.3	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags: Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig22239	GH11; CBM10	2106	6698	ACL68347	gi|219964511|gb|ACL68347.1|	1.00E-107	203	88.2	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig22400	GH11; Polysacc_deac_1	1388	1774	AAF14365	gi|6502585|gb|AAF14365.1|AF123252_1	1.00E-115	224	89.3	endo-1,4-beta-xylanase [Neocallimastix patriciarum]

Contig30387	GH11; CBM10	2084	7333	ACL68347	gi|219964511|gb|ACL68347.1|	1.00E-139	360	68.9	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig30545*	GH11; CBM1	1032	1115	ABW04217	gi|157930095|gb|ABW04217.1|	1.00E-118	246	87.0	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig6117*	GH11; Polysacc_deac_1	1744	1843	ABW04217	gi|157930095|gb|ABW04217.1|	1.00E-108	231	78.4	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig23659	GH11; CBM10	1804	9175	ACL68347	gi|219964511|gb|ACL68347.1|	0	606	79.2	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig22195	GH11; CBM10	1036	4354	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	1.00E-112	315	61.6	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags: Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig31461	GH11; CBM1	548	267	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	4.00E-46	115	74.8	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags: Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig21648	GH11; CBM10	1187	576	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	2.00E-90	237	68.8	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags: Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig30887*	GH11; CBM10	1082	811	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	1.00E-86	192	77.1	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags:

									Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig2264	GH11; CBM10	916	266	Q12667	gi|2494337|sp|Q12667.1|XYNA_PIRSP	6.00E-87	306	52.0	RecName: Full=Endo-1,4-beta-xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase; AltName: Full=Xylanase A; Short=XYLA; Flags: Precursorgi|1197372|emb|CAA62969.1| endo-1,4-beta-xylanase; xylanase A [Piromyces sp.]
Contig29248	GH11; CBM10	541	2541	AAD04194	gi|1655815|gb|AAD04194.1|	6.00E-84	166	84.9	xylanase [Orpinomyces sp. PC-2]
Contig28636	GH11; CBM1	658	1139	AAF14365	gi|6502585|gb|AAF14365.1|AF123252_1	7.00E-26	65	89.2	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig28877	GH11; CBM10	895	8096	ACL68347	gi|219964511|gb|ACL68347.1|	7.00E-68	205	61.0	endo-1,4-beta-xylanase [Neocallimastix patriciarum]
Contig9426	Alpha-amylase; CBM48	2139	2329	XP_001029557	gi|118394367|ref|XP_001029557.1|	0	675	56.4	Alpha amylase, catalytic domain containing protein [Tetrahymena thermophila]gi|89283797|gb|EAR81894.1| Alpha amylase, catalytic domain containing protein [Tetrahymena thermophila SB210]
Contig22817*	GH16; CBM10	1146	1105	AAB69347	gi|2353005|gb|AAB69347.1|	1.00E-18	90	52.2	cellulase [Orpinomyces joyonii]
Contig29583	GH16; CBM10	1039	820	Q12647	gi|2494328|sp|Q12647.1|GUNB_NEOPA	5.00E-19	90	52.2	RecName: Full=Endoglucanase B; AltName: Full=Endo-1,4-beta-glucanase B; AltName: Full=Cellulase B; Flags: Precursorgi|467687|emb|CAA83238.1| endoglucanase B [Neocallimastix patriciarum]
Contig25795	GH16; CBM10	1080	911	ZP_05512236	gi|256773773|ref|ZP_05512236.1|	4.00E-90	232	64.7	glycoside hydrolase family 16 [Streptomyces hygroscopicus ATCC 53653]
Contig1050*	GH18; Chitin_bind_1	2417	3248	YP_001643260	gi|163938376|ref|YP_001643260.1|	2.00E-56	427	34.2	glycoside hydrolase family protein [Bacillus weihenstephanensis KBAB4]gi|163860573|gb|ABY41632.1| glycoside hydrolase family 18

									[Bacillus weihenstephanensis KBAB4]
Contig9218	GH18; Chitin_bind_1	2273	1527	YP_001643260	gi|163938376|ref|YP_001643260.1|	3.00E-57	427	34.4	glycoside hydrolase family protein [Bacillus weihenstephanensis KBAB4]gi|163860573|gb|ABY41632.1| glycoside hydrolase family 18 [Bacillus weihenstephanensis KBAB4]
Contig745	GH18; Chitin_bind_1	1796	598	ZP_04850994	gi|253573651|ref|ZP_04850994.1|	3.00E-43	400	31.3	chitinase A1 [Paenibacillus sp. oral taxon 786 str. D14]gi|251847179|gb|EES75184.1| chitinase A1 [Paenibacillus sp. oral taxon 786 str. D14]
Contig8720	GH18; Chitin_bind_1	1439	567	Q01MB6	gi|152013345|sp|Q01MB6.2|AGI_ORYSI	6.00E-24	227	32.2	RecName: Full=Lectin; AltName: Full=Agglutinin; Contains: RecName: Full=Lectin 10 kDa peptide; Contains: RecName: Full=Lectin 8 kDa peptide; Flags: Precursorgi|218194445|gb|EEC76872.1| hypothetical protein OsI_15064 [Oryza sativa Indica Group]
Contig12835	GH18; Chitin_bind_1	1118	480	Q01MB6	gi|152013345|sp|Q01MB6.2|AGI_ORYSI	7.00E-27	206	34.5	RecName: Full=Lectin; AltName: Full=Agglutinin; Contains: RecName: Full=Lectin 10 kDa peptide; Contains: RecName: Full=Lectin 8 kDa peptide; Flags: Precursorgi|218194445|gb|EEC76872.1| hypothetical protein OsI_15064 [Oryza sativa Indica Group]
Contig30315	GH26; CBM10	1355	1527	YP_003249349	gi|261415666|ref|YP_003249349.1|	2.00E-91	282	55.7	Beta-mannanase-like protein [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372122|gb|ACX74867.1| Beta-mannanase-like protein [Fibrobacter succinogenes subsp. succinogenes S85]
Contig30263	GH26; CBM10	1272	781	YP_003249349	gi|261415666|ref|YP_003249349.1|	1.00E-103	278	58.3	Beta-mannanase-like protein [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372122|gb|ACX74867.1| Beta-mannanase-like protein

									[Fibrobacter succinogenes subsp. succinogenes S85]
Contig22949*	GH26; CBM10	1278	865	YP_003249349	gi|261415666|ref|YP_003249349.1|	4.00E-96	281	57.3	Beta-mannanase-like protein [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372122|gb|ACX74867.1| Beta-mannanase-like protein [Fibrobacter succinogenes subsp. succinogenes S85]
Contig21972	GH26; CBM35	688		ZP_06144005	gi|268610278|ref|ZP_06144005.1|	3.00E-51	205	52.2	mannan endo-1,4-beta-mannosidase [Ruminococcus flavefaciens FD-1]
Contig4027	Melibiase; Ricin_B_lectin	1309	1006	BAC66445	gi|29335747|dbj|BAC66445.1|	1.00E-112	281	64.1	alpha-galactosidase [Helianthus annuus]
Contig2025	Melibiase; Ricin_B_lectin	1082	31	ABK25009	gi|116788809|gb|ABK25009.1|	1.00E-102	264	63.6	unknown [Picea sitchensis]
Contig7911	GH43; CBM10	1247	1162	YP_001038591	gi|125974681|ref|YP_001038591.1|	1.00E-101	297	59.6	carbohydrate-binding family 6 protein [Clostridium thermocellum ATCC 27405]gi|256004120|ref|ZP_05429104.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM 2360]gi|281418847|ref|ZP_06249866.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]gi|125714906|gb|ABN53398.1| Carbohydrate binding family 6 [Clostridium thermocellum ATCC 27405]gi|255991868|gb|EEU01966.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM 2360]gi|281407931|gb|EFB38190.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]
Contig2640	GH43; CBM6	1288	705	YP_001038591	gi|125974681|ref|YP_001038591.1|	1.00E-156	425	62.1	carbohydrate-binding family 6 protein [Clostridium thermocellum ATCC 27405]gi|256004120|ref|ZP_05429104.1| Carbohydrate binding family 6

									[Clostridium thermocellum DSM 2360]gi|281418847|ref|ZP_06249866.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]gi|125714906|gb|ABN53398.1| Carbohydrate binding family 6 [Clostridium thermocellum ATCC 27405]gi|255991868|gb|EEU01966.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM 2360]gi|281407931|gb|EFB38190.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]
Contig2588	GH43; CBM10	1215	1515	YP_001038591	gi|125974681|ref|YP_001038591.1|	2.00E-98	304	57.9	carbohydrate-binding family 6 protein [Clostridium thermocellum ATCC 27405]gi|256004120|ref|ZP_05429104.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM 2360]gi|281418847|ref|ZP_06249866.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]gi|125714906|gb|ABN53398.1| Carbohydrate binding family 6 [Clostridium thermocellum ATCC 27405]gi|255991868|gb|EEU01966.1| Carbohydrate binding family 6 [Clostridium thermocellum DSM 2360]gi|281407931|gb|EFB38190.1| glycoside hydrolase family 43 [Clostridium thermocellum JW20]
Contig22546	GH43; CBM6; CBM10	1686	1303	ACZ98594	gi|280977753|gb|ACZ98594.1|	1.00E-140	482	54.6	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig717	GH43; CBM1; CBM6	2028	183	ACZ98594	gi|280977753|gb|ACZ98594.1|	1.00E-125	449	51.2	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig22009	GH43; CBM6	1443	2151	ACZ98594	gi|280977753|gb|ACZ98594.1|	1.00E-136	462	51.9	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig110	GH43; CBM6	1448	3325	ACZ98594	gi|280977753|	1.00E-	460	52.6	endo-1,4-beta-xylanase

					gb|ACZ98594.1|	134			[Cellulosilyticum ruminicola]
Contig12288	GH43; CBM6	1266	761	ACZ98594	gi|280977753|gb|ACZ98594.1|	1.00E-124	416	54.3	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig23444	GH43; CBM6	1567	552	ACZ98594	gi|280977753|gb|ACZ98594.1|	1.00E-123	470	48.1	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig29841	GH43; CBM6	2231	2528	ACZ98594	gi|280977753|gb|ACZ98594.1|	1.00E-129	442	52.7	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig30950*	GH43; CBM6	1662	5701	ACZ98594	gi|280977753|gb|ACZ98594.1|	1.00E-114	476	46.0	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
NODE_1061_length_1544_cov_11.881476*	GH43; CBM6	1596	4307	ACZ98594	gi|280977753|gb|ACZ98594.1|	1.00E-121	474	48.7	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig3221	CBM10/CBM6/GH43/Ricin_B_lectin	2333	1609	CBK75020	gi|291519799|emb|CBK75020.1|	1.00E-162	661	46.9	Cellobiohydrolase A (1,4-beta-cellobiosidase A) [Butyrivibrio fibrisolvens 16/4]
Contig880	GH43; CBM6	1573	680	ACZ98594	gi|280977753|gb|ACZ98594.1|	1.00E-115	430	50.7	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig14064	GH43; Ricin_B_lectin	1463	646	ZP_06142338	gi|268608611|ref|ZP_06142338.1|	1.00E-129	558	47.3	Alpha-N-arabinofuranosidase [Ruminococcus flavefaciens FD-1]
Contig9591	GH43; CBM6	929	559	ADE82665	gi|294473276|gb|ADE82665.1|	1.00E-108	291	64.6	glycosyl hydrolase, family 43 [Prevotella ruminicola 23]
Contig30960	GH43; CBM10; Ricin_B_lectin	1712	377	ACZ98594	gi|280977753|gb|ACZ98594.1|	4.00E-91	477	39.8	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig1720	GH43; CBM6; CBM10	1206	390	ACZ98594	gi|280977753|gb|ACZ98594.1|	5.00E-77	298	49.7	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig1733*	GH45; CBM10	1399	1630	BAC53956	gi|27530542|dbj|BAC53956.1|	1.00E-79	325	46.8	endo-beta-1,4-D-glucanase [Rhizopus oryzae]
Contig22152	GH45; CBM10	1130	1019	CAB92325	gi|8052314|emb|CAB9232	3.00E-91	402	43.8	endoglucanase 45A [Piromyces equi]

					5.1|			
Contig6437	GH45; CBM10	1094	900	CAB92325	gi|8052314|emb|CAB92325.1|	1.00E-121	419	52.5	endoglucanase 45A [Piromyces equi]
Contig1766*	GH45; CBM10	1202	979	CAB92325	gi|8052314|emb|CAB92325.1|	1.00E-103	397	47.1	endoglucanase 45A [Piromyces equi]
Contig31575	GH45; CBM10	1425	1422	BAD95809	gi|62821724|dbj|BAD95809.1|	4.00E-78	331	46.5	endo-beta-D-1,4-glucanase [Mucor circinelloides]
Contig929	GH45; CBM10	1296	116	ABU49185	gi|158138919|gb|ABU49185.2|	3.00E-73	321	43.0	endoglucanase [Syncephalastrum racemosum]
Contig10492	GH45; CBM10	1102	984	CAB92325	gi|8052314|emb|CAB92325.1|	1.00E-101	421	45.1	endoglucanase 45A [Piromyces equi]
Contig30272	GH45; CBM10	861	474	CAB92325	gi|8052314|emb|CAB92325.1|	1.00E-79	282	52.1	endoglucanase 45A [Piromyces equi]
Contig32*	GH45; CBM10	1284	214	CAB92325	gi|8052314|emb|CAB92325.1|	0	407	77.4	endoglucanase 45A [Piromyces equi]
Contig22584	GH45; CBM10	1142	4309	CAB92325	gi|8052314|emb|CAB92325.1|	1.00E-102	420	46.4	endoglucanase 45A [Piromyces equi]
Contig3753*	GH45; CBM10	1440	1409	CAB92325	gi|8052314|emb|CAB92325.1|	1.00E-103	421	47.7	endoglucanase 45A [Piromyces equi]
Contig17541	GH45; CBM10	1420	433	ABU49185	gi|158138919|gb|ABU49185.2|	8.00E-71	298	43.3	endoglucanase [Syncephalastrum racemosum]
Contig1445*	GH45; CBM1; CBM10	1513	3881	BAC53988	gi|27530617|dbj|BAC53988.1|	5.00E-74	348	44.5	endo-glucanase RCE3 [Rhizopus oryzae]
Contig30384*	GH45; CBM10	1251	620	CAB92325	gi|8052314|emb|CAB92325.1|	1.00E-177	416	69.7	endoglucanase 45A [Piromyces equi]
Contig22532	GH45; CBM10	1275	1091	ABU49185	gi|158138919|gb|ABU49185.2|	5.00E-53	249	42.2	endoglucanase [Syncephalastrum racemosum]
Contig25134	GH45; CBM10	765	240	BAC53956	gi|27530542|d	7.00E-	128	64.8	endo-beta-1,4-D-glucanase

					bj|BAC53956.1|	45			[Rhizopus oryzae]
Contig6490	GH45; CBM1; CBM10	1252	805	BAC53956	gi|27530542|dbj|BAC53956.1|	7.00E-45	243	42.0	endo-beta-1,4-D-glucanase [Rhizopus oryzae]
Contig29633	GH45; CBM10	859	900	CAB92325	gi|8052314|emb|CAB92325.1|	2.00E-58	254	46.1	endoglucanase 45A [Piromyces equi]
Contig22037	GH45; CBM10	1010	640	BAC53988	gi|27530617|dbj|BAC53988.1|	2.00E-33	201	40.3	endo-glucanase RCE3 [Rhizopus oryzae]
Contig23383	GH45; CBM1; CBM10	1094	1485	BAC53988	gi|27530617|dbj|BAC53988.1|	1.00E-34	206	39.8	endo-glucanase RCE3 [Rhizopus oryzae]
Contig29736	GH45; CBM10	706	794	CAB92325	gi|8052314|emb|CAB92325.1|	3.00E-35	246	35.4	endoglucanase 45A [Piromyces equi]
Contig6087	GH45; CBM10	685	208	CAB92325	gi|8052314|emb|CAB92325.1|	4.00E-83	220	63.6	endoglucanase 45A [Piromyces equi]
Contig29807	GH45; CBM10	774	573	CAB92325	gi|8052314|emb|CAB92325.1|	7.00E-45	259	38.6	endoglucanase 45A [Piromyces equi]
Contig21589	GH48; CBM10	2292	52778	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	0	746	78.2	cellulase Cel48A precursor [Piromyces sp. E2]
NODE_3576_length_2015_cov_12.189578	GH48; CBM10	2067	11488	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	0	673	79.8	cellulase Cel48A precursor [Piromyces sp. E2]
Contig22697	GH48; CBM10	1232	5166	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	0	409	75.8	cellulase Cel48A precursor [Piromyces sp. E2]
Contig29557	GH48; CBM10	954	833	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	1.00E-137	318	71.1	cellulase Cel48A precursor [Piromyces sp. E2]
Contig29850	GH48; CBM10	1092	11311	AAN76735	gi|25990957|gb|AAN76735.	1.00E-152	340	73.8	cellulase Cel48A precursor [Piromyces equi]

					1|AF449413_1				
Contig22790	GH48; CBM10	1029	641	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	1.00E-137	324	68.5	cellulase Cel48A precursor [Piromyces sp. E2]
Contig29259	GH48; CBM10	1037	4832	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	1.00E-147	337	72.1	cellulase Cel48A precursor [Piromyces sp. E2]
Contig274	GH48; CBM10	1054	8504	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	1.00E-155	331	75.2	cellulase Cel48A precursor [Piromyces sp. E2]
Contig30113	GH48; CBM10	1009	1234	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	1.00E-138	328	70.1	cellulase Cel48A precursor [Piromyces sp. E2]
Contig28658	GH48; CBM10	1093	16001	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	1.00E-145	319	74.3	cellulase Cel48A precursor [Piromyces sp. E2]
Contig28843	GH48; CBM10	980	2910	AAN76735	gi|25990957|gb|AAN76735.1|AF449413_1	1.00E-138	319	69.6	cellulase Cel48A precursor [Piromyces equi]
Contig22913	GH48; CBM10	1134	496	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	1.00E-145	313	71.2	cellulase Cel48A precursor [Piromyces sp. E2]
Contig29098	GH48; CBM10	1484	10522	AAN76735	gi|25990957|gb|AAN76735.1|AF449413_1	1.00E-145	327	71.3	cellulase Cel48A precursor [Piromyces equi]
Contig19530	GH48; CBM10	939	1085	AAN76735	gi|25990957|gb|AAN76735.1|AF449413_1	1.00E-131	317	69.7	cellulase Cel48A precursor [Piromyces equi]
Contig29792	GH48; CBM10	993	721	AAN76735	gi|25990957|gb|AAN76735.1|AF449413_	1.00E-120	322	61.5	cellulase Cel48A precursor [Piromyces equi]

					1				
Contig22124	GH48; CBM10	975	2451	AAN76735	gi|25990957|gb|AAN76735.1|AF449413_1	1.00E-129	324	67.3	cellulase Cel48A precursor [Piromyces equi]
Contig26288	GH48; CBM10	1029	351	AAN76735	gi|25990957|gb|AAN76735.1|AF449413_1	1.00E-116	306	63.7	cellulase Cel48A precursor [Piromyces equi]
NODE_3719_length_675_cov_5.463704*	GH48; CBM10	727	503	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	3.00E-91	225	67.1	cellulase Cel48A precursor [Piromyces sp. E2]
Contig24405	GH48; CBM10	580	245	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	2.00E-71	192	66.7	cellulase Cel48A precursor [Piromyces sp. E2]
Contig29641	GH48; CBM10	505	1065	AAN76734	gi|25990955|gb|AAN76734.1|AF449412_1	5.00E-56	164	64.6	cellulase Cel48A precursor [Piromyces sp. E2]
Contig28825	GH48; CBM10	518	2213	AAN76735	gi|25990957|gb|AAN76735.1|AF449413_1	3.00E-47	172	51.7	cellulase Cel48A precursor [Piromyces equi]
Contig28681	GH48; CBM10	595	6599	AAN76735	gi|25990957|gb|AAN76735.1|AF449413_1	2.00E-63	165	64.8	cellulase Cel48A precursor [Piromyces equi]
Contig713*	CBM1; CBM10	1882	73	NP_593986	gi|19114898|ref|NP_593986.1|	1.00E-08	41	58.5	fungal cellulose binding domain protein [Schizosaccharomyces pombe 972h-]gi|74625381|sp|Q9P7F1.1|YKK5_SCHPO RecName: Full=Carbohydrate-binding domain-containing protein C2E1P3.05c; Flags: Precursorgi|7340823|emb|CAB83009.1| fungal cellulose binding domain protein [Schizosaccharomyces pombe]

Contig56*	CBM1; CBM10	2097	149	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	2.00E-10	97	41.2	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig10481	CBM1; CBM10	529	158	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	2.00E-24	85	56.5	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig1742	CBM1; CBM10	2042	102	CAB92325	gi|8052314|emb|CAB92325.1|	2.00E-09	115	34.8	endoglucanase 45A [Piromyces equi]
Contig30371	CBM1; Chitin_bind_1	829	633	ABY52793	gi|164375375|gb|ABY52793.1|	3.00E-08	36	61.1	cellobiohydrolase [Neocallimastix patriciarum]
Contig2052	CBM1; Pfam-B_8046	994	517	XP_001829175	gi|169844909|ref|XP_001829175.1|	1.00E-37	200	41.0	hypothetical protein CC1G_01855 [Coprinopsis cinerea okayama7#130]gi|116509915|gb|EAU92810.1| hypothetical protein CC1G_01855 [Coprinopsis cinerea okayama7#130]
Contig12713	CBM1; CBM10	1642	209	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	3.00E-10	90	45.6	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig17415*	CBM1; CBM10	2081	830	AAL01213	gi|15529298|gb|AAL01213.1|AF177206_1	1.00E-09	91	40.7	mannanase ManA [Orpinomyces sp. PC-2]
Contig1019	CBM1; CBM10	1686	1186	AAD43818	gi|5457159|gb|AAD43818.1|AF165266_1	1.00E-10	90	41.1	endoglucanase precursor [Piromyces rhizinflatus]
Contig29787	CBM6; Ricin_B_lectin	1281	757	NP_149283	gi|15004823|ref|NP_149283.1|	5.00E-85	417	40.5	xylan degradation protein [Clostridium acetobutylicum ATCC 824]gi|14994435|gb|AAK76865.1|AE001438_118 Possible xylan degradation enzyme (glycosyl hydrolase family 43-like domain, cellulose-binding domain and Ricin B-like domain) [Clostridium acetobutylicum ATCC 824]
Contig6460	CBM6; CBM36	520		ACZ98594	gi|280977753|	1.00E-	172	45.9	endo-1,4-beta-xylanase

					gb|ACZ98594.1|	38			[Cellulosilyticum ruminicola]
Contig5055	CBM6; CBM10	724	284	ACZ98594	gi|280977753|gb|ACZ98594.1|	2.00E-25	147	40.8	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
NODE_4815_length_842_cov_8.675772	CBM6; CBM10	878	528	ACZ98594	gi|280977753|gb|ACZ98594.1|	9.00E-34	178	42.1	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig29852	CBM6; CBM10	873	687	ACZ98594	gi|280977753|gb|ACZ98594.1|	2.00E-35	186	39.8	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig30353*	CBM6; CBM10; Ricin_B_lectin	1427	1252	ACZ98594	gi|280977753|gb|ACZ98594.1|	3.00E-55	346	37.0	endo-1,4-beta-xylanase [Cellulosilyticum ruminicola]
Contig24142	CBM10; Rhamnogal_lyase	2025	1180	NP_195516	gi|186517294|ref|NP_195516.2|	3.00E-69	531	33.1	lyase [Arabidopsis thaliana]
Contig345*	CBM10; Swollenin	1813	305	EDP47653	gi|159122532|gb|EDP47653.1|	1.00E-127	432	55.1	swollenin, putative [Aspergillus fumigatus A1163]
Contig28884	CBM10; CBM29	994		AAK20910	gi|13446353|gb|AAK20910.1|	2.00E-70	314	44.9	non-catalytic protein 1 [Piromyces equi]
Contig536	CBM10; CBM29	1467		AAK20910	gi|13446353|gb|AAK20910.1|	0	478	63.0	non-catalytic protein 1 [Piromyces equi]
Contig30769	CBM10; Swollenin	834	338	P55296	gi|1708917|sp|P55296.1|MANA_PIRSP	4.00E-32	131	52.7	RecName: Full=Mannan endo-1,4-beta-mannosidase A; AltName: Full=Beta-mannanase A; AltName: Full=1,4-beta-D-mannan mannanohydrolase A; Flags: Precursorgi|1197368|emb|CAA62968.1| mannan endo-1,4-beta-mannosidase; mannanase A [Piromyces sp.]
Contig3979	CBM10; Swollenin	1735	1487	ABV57767	gi|157488002|gb|ABV57767.1|	1.00E-119	392	56.6	swollenin [Hypocrea pseudokoningii]
Contig28713	CBM10; CBM29	1520		AAK20910	gi|13446353|gb|AAK20910.1|	1.00E-131	490	49.2	non-catalytic protein 1 [Piromyces equi]

Contig11987	CBM10; Ricin_B_lectin	885	378	CBK75021	gi|291519800|emb|CBK75021.1|	1.00E-31	151	44.4	Beta-1,4-xylanase [Butyrivibrio fibrisolvens 16/4]
Contig28612	CBM10; CBM29	1462		AAK20910	gi|13446353|gb|AAK20910.1|	1.00E-121	474	46.6	non-catalytic protein 1 [Piromyces equi]
Contig22156	CBM10; Swollenin	1753	3682	ABV57767	gi|157488002|gb|ABV57767.1|	1.00E-131	485	49.1	swollenin [Hypocrea pseudokoningii]
Contig2177	CBM10; DUF303	999	2652	AAB69090	gi|2231243|gb|AAB69090.1|	1.00E-138	347	66.6	acetylxylan esterase [Neocallimastix patriciarum]
Contig29571*	CBM10; DUF303	1198	1223	AAB69090	gi|2231243|gb|AAB69090.1|	0	394	86.0	acetylxylan esterase [Neocallimastix patriciarum]
Contig9091	CBM10; Swollenin	1685	1866	ACB05430	gi|169893727|gb|ACB05430.1|	1.00E-128	385	57.7	swollenin [Trichoderma asperellum]
Contig900	CBM10; CE15	1425	715	YP_677850	gi|110637643|ref|YP_677850.1|	1.00E-140	380	61.3	beta-xylosidase/endoglucanase [Cytophaga hutchinsonii ATCC 33406]gi|110280324|gb|ABG58510.1| CHU large protein; beta-xylosidase/endoglucanase [Cytophaga hutchinsonii ATCC 33406]
Contig3933	CBM10; Swollenin	994	543	CAB92328	gi|8052455|emb|CAB92328.1|	1.00E-21	166	39.2	swollenin [Hypocrea jecorina]
Contig5849	CBM10; Esterase	804	289	AAL01212	gi|15529296|gb|AAL01212.1|AF177205_1	6.00E-26	104	54.8	cellobiohydrolase II-like cellulase CelI [Orpinomyces sp. PC-2]
Contig29074*	CBM10; DUF303	1251	1488	AAB69090	gi|2231243|gb|AAB69090.1|	1.00E-174	393	75.8	acetylxylan esterase [Neocallimastix patriciarum]
Contig3715	CBM10; Swollenin	1450	912	CAB92328	gi|8052455|emb|CAB92328.1|	2.00E-81	303	52.1	swollenin [Hypocrea jecorina]
Contig25059	CBM10; CE15	1096	3051	YP_003250416	gi|261416733|ref|YP_003250416.1|	1.00E-83	227	68.7	Cip2 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261373189|gb|ACX75934.1| Cip2 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig3876*	CBM10;	1790	4205	ACB05430	gi|169893727|	1.00E-	400	52.0	swollenin [Trichoderma asperellum]

	Swollenin				gb|ACB05430.1|	107			
Contig30515*	CBM10; DUF303	1129	614	AAB69090	gi|2231243|gb|AAB69090.1|	4.00E-94	377	48.0	acetylxylan esterase [Neocallimastix patriciarum]
Contig5283	CBM10; CE15	815	322	YP_677850	gi|110637643|ref|YP_677850.1|	2.00E-60	189	58.2	beta-xylosidase/endoglucanase [Cytophaga hutchinsonii ATCC 33406]gi|110280324|gb|ABG58510.1| CHU large protein; beta-xylosidase/endoglucanase [Cytophaga hutchinsonii ATCC 33406]
Contig21613*	CBM10; Swollenin	1773	2760	CAB92328	gi|8052455|emb|CAB92328.1|	1.00E-109	451	48.1	swollenin [Hypocrea jecorina]
Contig26668	CBM10; Ricin_B_lectin	1134	607	CBK75021	gi|291519800|emb|CBK75021.1|	1.00E-39	151	51.7	Beta-1,4-xylanase [Butyrivibrio fibrisolvens 16/4]
Contig23671	CBM10; Swollenin	1094	549	CAB92328	gi|8052455|emb|CAB92328.1|	3.00E-39	218	45.9	swollenin [Hypocrea jecorina]
Contig11249*	CBM10; Polysacc_deac_1	1357	686	YP_003250045	gi|261416362|ref|YP_003250045.1|	1.00E-142	332	73.5	polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372818|gb|ACX75563.1| polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig14319	CBM10; Ricin_B_lectin	791	299	CBK74914	gi|291519693|emb|CBK74914.1|	5.00E-37	154	53.9	Enterochelin esterase and related enzymes [Butyrivibrio fibrisolvens 16/4]
Contig29530	CBM10; DUF303	1201	1925	AAB69090	gi|2231243|gb|AAB69090.1|	1.00E-141	374	64.4	acetylxylan esterase [Neocallimastix patriciarum]
Contig11158	CBM10; Ricin_B_lectin	683	161	ABY52795	gi|164375379|gb|ABY52795.1|	1.00E-28	88	61.4	endo-1,4-beta-xylanase [Piromyces communis]
Contig22082	CBM10; DUF303	1055	758	AAB69090	gi|2231243|gb|AAB69090.1|	1.00E-111	373	53.6	acetylxylan esterase [Neocallimastix patriciarum]
Contig3201	CBM10; Pfam-B_184	2196	1316	ZP_05498108	gi|256757388|ref|ZP_05498108.1|	1.00E-84	502	37.1	conserved hypothetical protein [Clostridium papyrosolvens DSM 2782]gi|256743722|gb|EEU56906.1| conserved hypothetical protein

									[Clostridium papyrosolvens DSM 2782]
Contig1549	CBM10; Esterase	1738	35	AAF70241	gi|7839348|gb|AAF70241.1|AF164351_1	1.00E-175	450	64.2	feruloyl esterase A [Orpinomyces sp. PC-2]
Contig29391	CBM10; CE15	1623	6258	YP_003250416	gi|261416733|ref|YP_003250416.1|	1.00E-170	381	76.6	Cip2 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261373189|gb|ACX75934.1| Cip2 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig15160*	CBM10; Polysacc_deac_1	1409	2791	YP_003250045	gi|261416362|ref|YP_003250045.1|	1.00E-152	401	67.3	polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372818|gb|ACX75563.1| polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig7471	CBM10; CE15	1565	3404	YP_003250416	gi|261416733|ref|YP_003250416.1|	1.00E-168	381	75.9	Cip2 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261373189|gb|ACX75934.1| Cip2 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig23065*	CBM10; Polysacc_deac_1	1544	999	YP_003250045	gi|261416362|ref|YP_003250045.1|	1.00E-141	333	73.6	polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372818|gb|ACX75563.1| polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig3750*	CBM10; Polysacc_deac_1	1487	9636	YP_003250045	gi|261416362|ref|YP_003250045.1|	1.00E-152	331	76.7	polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372818|gb|ACX75563.1| polysaccharide deacetylase [Fibrobacter succinogenes subsp. succinogenes S85]
Contig28807	CBM10; CE15	1617	4605	YP_003250416	gi|261416733|ref|YP_003250416.1|	1.00E-156	380	71.6	Cip2 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261373189|gb|ACX75934.1| Cip2 [Fibrobacter succinogenes subsp. succinogenes S85]

Contig5391	CBM10; Ricin_B_lectin	1086	580	YP_003250020	gi|261416337|ref|YP_003250020.1|	1.00E-48	137	65.7	hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]gi|261372793|gb|ACX75538.1| hypothetical protein Fisuc_1948 [Fibrobacter succinogenes subsp. succinogenes S85]
Contig21252	CBM10; DUF303	1456	1459	AAC14690	gi|3080749|gb|AAC14690.1|	1.00E-114	273	72.9	acetyl xylan esterase A [Orpinomyces sp. PC-2]
Contig24180*	CBM10; Chitin_bind_1	1682	1144	XP_001560867	gi|154323105|ref|XP_001560867.1|	7.00E-19	102	47.1	hypothetical protein BC1G_00895 [Botryotinia fuckeliana B05.10]gi|150848229|gb|EDN23422.1| hypothetical protein BC1G_00895 [Botryotinia fuckeliana B05.10]
Contig12787	CBM10; Esterase	1804	3399	AAF70241	gi|7839348|gb|AAF70241.1|AF164351_1	1.00E-165	455	61.3	feruloyl esterase A [Orpinomyces sp. PC-2]
Contig4845	CBM10; Ricin_B_lectin	613	187	NP_149235	gi|15004775|ref|NP_149235.1|	3.00E-31	142	47.9	xylan degradation protein [Clostridium acetobutylicum ATCC 824]gi|14994387|gb|AAK76817.1|AE001438_70 Possible xylan degradation enzyme (alpha/beta hydrolase domain and ricin-B-like domain) [Clostridium acetobutylicum ATCC 824]
Contig8656	CBM10; Ricin_B_lectin	680	186	NP_149235	gi|15004775|ref|NP_149235.1|	9.00E-40	148	55.4	xylan degradation protein [Clostridium acetobutylicum ATCC 824]gi|14994387|gb|AAK76817.1|AE001438_70 Possible xylan degradation enzyme (alpha/beta hydrolase domain and ricin-B-like domain) [Clostridium acetobutylicum ATCC 824]
Contig156	CBM10; Ricin_B_lectin	612	182	CBK75021	gi|291519800|emb|CBK75021.1|	4.00E-34	154	47.4	Beta-1,4-xylanase [Butyrivibrio fibrisolvens 16/4]
Contig4849	CBM10; Swollenin	605	340	CAB92328	gi|8052455|emb|CAB92328.1|	7.00E-18	116	42.2	swollenin [Hypocrea jecorina]
Contig1517	CBM10; Esterase	1037	83	ACZ98648	gi|280977861|	3.00E-	276	47.8	esterase [Cellulosilyticum

					gb|ACZ98648.1|	65			ruminicola]
Contig771*	CBM10; Esterase	1202	610	ACZ98648	gi|280977861|gb|ACZ98648.1|	6.00E-59	273	45.4	esterase [Cellulosilyticum ruminicola]
Contig21459*	Chitin_bind_1; Polysacc_deac_1	2241	2370	XP_001560867	gi|154323105|ref|XP_001560867.1|	3.00E-40	371	29.9	hypothetical protein BC1G_00895 [Botryotinia fuckeliana B05.10]gi|150848229|gb|EDN23422.1| hypothetical protein BC1G_00895 [Botryotinia fuckeliana B05.10]
Contig27125	Chitin_bind_1; Polysacc_deac_1	1195	807	XP_001598179	gi|156064515|ref|XP_001598179.1|	7.00E-47	378	32.8	hypothetical protein SS1G_00265 [Sclerotinia sclerotiorum 1980]gi|154691127|gb|EDN90865.1| hypothetical protein SS1G_00265 [Sclerotinia sclerotiorum 1980]
Contig6919	Chitin_bind_1; Polysacc_deac_1	1280	815	XP_001560867	gi|154323105|ref|XP_001560867.1|	6.00E-44	312	35.3	hypothetical protein BC1G_00895 [Botryotinia fuckeliana B05.10]gi|150848229|gb|EDN23422.1| hypothetical protein BC1G_00895 [Botryotinia fuckeliana B05.10]
Contig442*	Chitin_bind_1; Polysacc_deac_1	1446	577	XP_001598179	gi|156064515|ref|XP_001598179.1|	3.00E-45	486	29.2	hypothetical protein SS1G_00265 [Sclerotinia sclerotiorum 1980]gi|154691127|gb|EDN90865.1| hypothetical protein SS1G_00265 [Sclerotinia sclerotiorum 1980]
Contig1675*	Chitin_bind_1; Polysacc_deac_1	1496	54	XP_384322	gi|46116608|ref|XP_384322.1|	3.00E-51	372	33.9	hypothetical protein FG04146.1 [Gibberella zeae PH-1]
Contig28694	Chitin_bind_1; Polysacc_deac_1	1220	4311	XP_001598179	gi|156064515|ref|XP_001598179.1|	5.00E-39	397	29.2	hypothetical protein SS1G_00265 [Sclerotinia sclerotiorum 1980]gi|154691127|gb|EDN90865.1| hypothetical protein SS1G_00265 [Sclerotinia sclerotiorum 1980]
Contig29559	Chitin_bind_1; Polysacc_deac_1	1282	2767	XP_001560867	gi|154323105|ref|XP_001560867.1|	5.00E-41	376	30.1	hypothetical protein BC1G_00895 [Botryotinia fuckeliana B05.10]gi|150848229|gb|EDN23422.1| hypothetical protein BC1G_00895 [Botryotinia fuckeliana B05.10]

Contig7781	Chitin_bind_1; Polysacc_deac_1	921	348	XP_001912680	gi|171695512|ref|XP_001912680.1|	2.00E-36	224	37.5	unnamed protein product [Podospora anserina]gi|170947998|emb|CAP60162.1| unnamed protein product [Podospora anserina]
Contig5140	Chitin_bind_1; Polysacc_deac_1	585	158	XP_001394100	gi|145243114|ref|XP_001394100.1|	1.00E-23	173	36.4	hypothetical protein An11g00920 [Aspergillus niger]gi|134078770|emb|CAK96883.1| unnamed protein product [Aspergillus niger]
Contig28933*	CBM20; PL6	1772	1446	ZP_01852968	gi|149174341|ref|ZP_01852968.1|	2.00E-17	418	24.9	Parallel beta-helix repeat protein [Planctomyces maris DSM 8797]gi|148846886|gb|EDL61222.1| Parallel beta-helix repeat protein [Planctomyces maris DSM 8797]
Contig3095*	CBM20; PL6	1733	5701	ZP_01852968	gi|149174341|ref|ZP_01852968.1|	7.00E-14	414	23.2	Parallel beta-helix repeat protein [Planctomyces maris DSM 8797]gi|148846886|gb|EDL61222.1| Parallel beta-helix repeat protein [Planctomyces maris DSM 8797]
Contig627	CBM20; PL6	1777	20162	ZP_01852968	gi|149174341|ref|ZP_01852968.1|	7.00E-14	417	24.5	Parallel beta-helix repeat protein [Planctomyces maris DSM 8797]gi|148846886|gb|EDL61222.1| Parallel beta-helix repeat protein [Planctomyces maris DSM 8797]
Contig22058	CBM20; PL6	1676	5416	ZP_01852968	gi|149174341|ref|ZP_01852968.1|	2.00E-10	426	24.4	Parallel beta-helix repeat protein [Planctomyces maris DSM 8797]gi|148846886|gb|EDL61222.1| Parallel beta-helix repeat protein [Planctomyces maris DSM 8797]
Contig21224*	CBM48; Alpha-amylase_C	2217	2906	XP_002470098	gi|242208495|ref|XP_002470098.1|	0	675	67.0	candidate 1,4-alpha-glucan branching enzyme from glycoside hydrolase family GH13 [Postia placenta Mad-698-R]gi|220730850|gb|EED84701.1| candidate 1,4-alpha-glucan branching enzyme from glycoside hydrolase family GH13 [Postia placenta Mad-698-R]
Contig21432*	Carb_bind; Pfam-B_1434	1589	1775	ZP_02024895	gi|154482447|ref|ZP_02024	1.00E-117	386	54.1	hypothetical protein EUBVEN_00114 [Eubacterium

					895.1|			ventriosum ATCC 27560]gi|149736696|gb|EDM52582.1| hypothetical protein EUBVEN_00114 [Eubacterium ventriosum ATCC 27560]
Contig21311	GH94; GT36_AF	2521	15361	ADE82788	gi|294473399|gb|ADE82788.1|	0	807	72.5	cellobiose phosphorylase [Prevotella ruminicola 23]
Contig21420*	CBM29; CBM10	1338	1764	P55296	gi|1708917|sp|P55296.1|MANA_PIRSP	1.00E-25	120	47.5	RecName: Full=Mannan endo-1,4-beta-mannosidase A; AltName: Full=Beta-mannanase A; AltName: Full=1,4-beta-D-mannan mannanohydrolase A; Flags: Precursorgi|1197368|emb|CAA62968.1| mannan endo-1,4-beta-mannosidase; mannanase A [Piromyces sp.]
Contig2330	CBM29; CBM10	969	784	AAK20910	gi|13446353|gb|AAK20910.1|	1.00E-78	316	46.5	non-catalytic protein 1 [Piromyces equi]
Contig2646	CBM29; CBM1	1014	576	1W90	gi|62738220|pdb|1W90|A	8.00E-05	139	25.2	Chain A, Cbm29-2 Mutant D114a: Probing The Mechanism Of Ligand Recognition By Family 29 Carbohydrate Binding Modulesgi|62738221|pdb|1W90|B Chain B, Cbm29-2 Mutant D114a: Probing The Mechanism Of Ligand Recognition By Family 29 Carbohydrate Binding Modules

Note: * Possible Full Length ORFs.

Table S.3 ORFs with predicted CAZy modules and their expression level represented by FPKMs for *Aneromyces mucronatus* grown on four different carbon sources and the clade they were grouped into as illustrated in Figure 4.12**.**

ORF	Length (bp)	CAZy Domains	Clade	FPKM			
				GCS	Oat Spelt Xylan	Barley Straw	Alfalfa Hay
AmuTC468	1134	CE4	A	973.0	1663.1	1983.2	1541.4
AmuTC51	1971	GH1	A	796.8	4314.9	2461.7	6980.4
AmuTC623	957	CE4	A	1412.9	1157.3	514.9	1230.9
AmuTC72	1428	CBM10/CBM29_Blast	A	141.7	1315.1	3797.9	6228.9
AmuTC8	1020	GH6	A	1035.6	3052.5	7188.6	8891.3
AmuTC832	1638	CBM13/CBM10	A	337.0	1867.4	292.6	1448.7
AmuTC99	984	GH43	A	6738.8	13185.3	1385.5	5002.7
AmuVC10744	708	CBM10/GH48	A	259.9	982.7	3603.8	5376.4
AmuVC219	1101	CBM10/GH45	A	375.5	2358.7	2111.1	3690.0
AmuVC9693	1629	CBM13/CBM10	A	199.3	1481.2	875.6	2065.2
AmuTC459	1539	CE15/CBM10	B	13.1	101.4	634.2	2674.0
AmuTC686_seq4	1182	CBM10	B	19.6	1359.2	1294.3	758.6
AmuVC11851	1314	CBM10/GH10	B	17.7	108.6	1439.8	3573.1
AmuVC11854	1164	CE6/CBM10	B	26.5	167.2	1140.9	4806.0
AmuVC1295	1215	CBM10/GH11	B	6.9	141.9	1215.8	1160.7
AmuVC1489	1185	CBM10	B	7.0	112.6	794.1	1122.8
AmuVC732	1377	CBM10/GH6	B	23.7	253.7	583.7	1428.7
AmuVN325	1164	CBM10	B	21.8	124.1	845.6	718.3
AmuTC1006	1059	CBM10	C	14.9	167.6	129.4	177.9
AmuTC149	1635	GH117/CBM6/GH43	C	330.5	929.7	292.7	301.2
AmuTC1541Seq1	5034	GDE_C	C	95.8	70.0	365.5	174.1
AmuTC1644	1113	CE4	C	272.2	280.1	385.6	411.2
AmuTC169	1506	CBM10	C	48.1	870.0	662.2	530.0
AmuTC169_seq3	1521	CBM10	C	25.8	523.6	407.9	352.6
AmuTC1949	3165	CBM13/CBM10/GH95_Blast	C	12.8	504.1	137.0	413.7
AmuTC2165	1671	GH43	C	161.0	210.9	114.5	178.5
AmuTC223	1476	GH13	C	168.1	227.3	400.3	120.1
AmuTC290	5250	CBM10	C	45.7	138.6	88.7	116.5
AmuTC3141	666	CBM13/CBM10	C	51.1	475.3	80.8	440.8
AmuTC352	822	GH11	C	473.9	981.4	99.0	258.1
AmuTC4	1926	Swollenin/CBM13/CBM10	C	123.6	426.1	284.0	527.9
AmuTC533	2943	CBM10/GH10	C	75.1	138.7	194.2	239.6

ORF	Length (bp)	CAZy Domains	Clade	FPKM			
				GCS	Oat Spelt Xylan	Barley Straw	Alfalfa Hay
AmuTC585	3891	GH3_C/CBM10/GH6/GH3	C	121.2	184.4	833.0	587.8
AmuTC66	1701	CBM10/GH5	C	148.1	590.3	788.2	733.7
AmuTC686_seq3	1338	CE4/CBM10	C	73.6	514.5	1034.8	1624.2
AmuTC858	2106	CBM13/CBM10/GH39	C	26.0	283.3	134.7	303.7
AmuTC910	1209	GH117/CBM10/GH43	C	98.0	1398.5	250.6	538.6
AmuVC10013	840	CBM10	C	52.7	325.6	520.6	377.8
AmuVC10015	1692	CE1	C	173.4	166.7	189.5	341.6
AmuVC10033	1557	CBM10	C	66.5	232.0	252.6	222.0
AmuVC10141	1014	CBM10	C	48.2	111.8	132.0	296.2
AmuVC10249	2664	CBM10/CBM1	C	184.1	233.3	162.1	176.7
AmuVC10358	945	CBM10	C	23.2	189.3	111.8	269.0
AmuVC11828	2589	CBM10/GH3	C	149.1	371.2	576.8	943.8
AmuVC11850	885	GH1	C	483.8	208.2	166.1	542.4
AmuVC11861	2121	CBM13/CBM10/CBM6/GH43/CBM36_Blast	C	93.0	493.4	457.3	1635.4
AmuVC11927	936	CE6	C	21.8	323.3	156.5	206.2
AmuVC1201	1596	CBM6/GH43	C	215.1	468.6	382.2	424.6
AmuVC12062	1590	GH114	C	150.8	242.3	187.8	278.3
AmuVC12088	2103	CBM13/CBM10/GH39	C	30.3	304.3	150.8	383.2
AmuVC12166	1035	CBM10/GH48	C	56.9	173.7	588.8	845.5
AmuVC126	1248	GH117/CBM10/GH43	C	43.3	313.2	204.8	573.9
AmuVC1308	876	CE4	C	175.6	278.3	493.0	428.6
AmuVC1394	1269	CE4/CBM18	C	73.0	404.3	175.4	163.8
AmuVC1590	1464	CBM10	C	14.6	611.0	277.2	198.7
AmuVC172	939	CE4	C	103.0	219.5	567.1	196.7
AmuVC1910	1899	CBM10	C	293.0	1012.3	117.0	185.1
AmuVC2077	1275	CBM10	C	68.8	201.3	273.8	348.7
AmuVC226	645	CBM10	C	267.6	278.0	863.9	780.6
AmuVC2304	1443	CBM10/CBM1/GH45	C	72.1	162.0	289.3	397.6
AmuVC2325	1380	CBM10/GH6	C	89.6	142.1	233.8	329.1
AmuVC2417	1263	CBM10/GH26	C	44.7	159.4	343.1	523.4
AmuVC306	3843	CBM26/GH31	C	53.6	78.9	492.0	236.1
AmuVC309	1725	Swollenin/CBM10	C	167.7	294.8	239.2	305.4
AmuVC3374	1527	CBM1	C	136.4	408.9	221.1	191.2
AmuVC34	1734	Swollenin/CBM10	C	145.9	405.1	452.7	904.0
AmuVC427	1059	CBM10/GH45	C	131.2	507.9	625.7	813.1
AmuVC4389	1122	CBM10/GH11	C	14.8	179.0	236.2	250.9

ORF	Length (bp)	CAZy Domains	Clade	FPKM			
				GCS	Oat Spelt Xylan	Barley Straw	Alfalfa Hay
AmuVC557	1614	CBM13/CBM10	C	28.7	242.9	135.5	342.4
AmuVC567	1470	CBM10/GH45	C	57.9	126.4	270.8	329.3
AmuVC58	2079	CBM10/CBM1	C	202.7	377.5	201.4	325.4
AmuVC8269	1494	CBM10/CBM1/GH45	C	33.5	111.6	351.3	314.9
AmuVC8468	2355	CBM10/GH9	C	114.7	209.5	216.8	424.5
AmuVC88	594	CBM13/CBM10	C	16.3	233.2	128.5	366.5
AmuVC9698	1263	GH18	C	280.0	780.4	227.4	152.5
AmuVC9715	2697	CBM10/GH9	C	102.7	595.1	762.3	463.8
AmuVC9724	2337	CBM10/GH9	C	105.3	107.1	197.9	441.5
AmuVC9753	2094	CBM10	C	43.0	161.9	691.3	209.7
AmuVC9826	1599	CBM13	C	84.3	881.0	1123.7	737.7
AmuVC9956	1698	Swollenin/CBM10	C	64.1	125.5	143.0	380.0
AmuTC1887	3573	CBM13	D	0.3	153.0	84.6	46.0
AmuTC991	525	CE4	D	2.8	237.6	133.2	241.6
AmuVC1058	555	CBM1	D	6.3	419.0	522.7	953.0
AmuVC11071	1398	CBM10	D	0.9	174.2	241.2	41.5
AmuVC11847	1791	CBM10/GH8	D	2.9	586.0	876.4	387.7
AmuVC12024	1170	CBM18	D	0.6	134.9	491.4	71.5
AmuVC12049	1209	CBM10	D	0.3	54.0	303.1	102.0
AmuVC3541	1485	CBM10	D	2.2	319.9	373.7	372.3
AmuVC3669	1575	CBM10	D	0.3	142.7	200.1	78.1
AmuVC523	1530	CBM10	D	1.9	274.7	256.1	356.2
AmuVC852	1428	CBM10/CBM1	D	2.9	379.3	73.2	329.3
AmuTC11219	492	CBM50	E	6.7	191.0	41.3	48.6
AmuTC11272	954	CE10	E	22.9	75.6	26.8	55.7
AmuTC1198	1434	CBM10/CBM1/GH45	E	41.4	93.1	175.8	230.2
AmuTC1213	3591	CBM21	E	30.7	55.7	140.3	49.2
AmuTC1303	654	CE10	E	32.2	237.4	101.3	131.6
AmuTC17890	744	CE4	E	44.8	27.8	48.7	44.7
AmuTC2143	1830	CBM10	E	74.7	359.5	72.4	96.1
AmuTC2171	1761	GH67	E	125.8	49.2	50.6	361.6
AmuTC290	2376	CBM10/GH9	E	220.6	463.5	841.8	1037.5
AmuTC2911	1452	GH43	E	24.6	53.8	24.1	42.2
AmuTC3045	1464	CE2 or CE3	E	14.5	91.2	105.9	157.6
AmuTC3201	1011	CE12/CBM1	E	19.1	86.6	81.3	205.7
AmuTC3825	957	CE4	E	15.1	57.2	62.1	35.1
AmuTC3998	456	CBM1	E	92.0	147.8	12.3	73.4
AmuTC408	681	CBM10	E	20.3	74.2	87.7	117.0
AmuTC408_seq2	642	CBM10	E	22.1	81.5	85.3	116.5
AmuTC4473_seq2	771	GH16	E	20.7	59.8	168.3	221.5

ORF	Length (bp)	CAZy Domains	Clade	FPKM			
				GCS	Oat Spelt Xylan	Barley Straw	Alfalfa Hay
AmuTC4587	1311	CBM10	E	43.9	83.0	146.0	174.9
AmuTC471	2940	GH31	E	21.5	37.3	200.2	116.4
AmuTC4860	2175	CBM10	E	32.6	65.2	31.2	64.6
AmuTC5285	963	CE10	E	61.0	310.0	63.4	104.3
AmuTC7281	2238	PL3	E	29.6	24.4	72.4	30.6
AmuTC9476	1122	CE4	E	37.1	29.2	31.2	45.6
AmuVC10002	2334	CBM10/GH9	E	12.9	123.1	78.6	175.8
AmuVC1002	2454	CBM10	E	53.0	108.8	53.8	112.7
AmuVC10088	1611	CE1	E	15.4	293.6	51.5	40.9
AmuVC10228	1452	CBM10/GH6	E	23.1	68.3	285.3	17.1
AmuVC10296	1068	GH114	E	18.6	200.5	41.0	83.2
AmuVC1032	1407	CBM10/GH5	E	34.0	69.6	74.6	249.6
AmuVC10350	1548	CE1	E	29.5	131.7	80.5	25.1
AmuVC10375	1506	GH11	E	17.6	34.3	83.7	102.0
AmuVC10459	3141	CBM21	E	229.0	68.7	30.0	38.9
AmuVC10543	1065	CBM10	E	51.8	69.9	22.3	130.4
AmuVC10681	2484	CBM10/CBM1	E	20.6	57.3	112.8	50.1
AmuVC1082	1596	GH117/CBM6/GH43	E	100.2	522.2	64.5	83.2
AmuVC1106	1341	GH114	E	114.1	251.3	44.9	119.0
AmuVC1110	1467	GH117/CBM6/GH43/CBM36_Blast	E	52.7	376.0	66.4	155.5
AmuVC11161	1641	CBM13/CBM10	E	49.0	817.6	21.2	84.3
AmuVC11366	1974	CBM18	E	293.7	20.0	10.6	61.6
AmuVC1148	2760	CBM21	E	28.0	31.9	113.0	40.3
AmuVC12040	2004	CBM10	E	14.1	55.7	128.5	92.5
AmuVC12041	933	CE4	E	66.1	289.2	96.9	165.0
AmuVC12163	1587	CBM10/GH5	E	90.4	38.5	81.5	127.6
AmuVC12184	1548	CBM10	E	47.3	99.4	152.3	171.1
AmuVC12245	1407	CBM10/GH48	E	29.4	66.5	192.0	239.7
AmuVC1268	1041	CBM18	E	21.4	44.3	51.0	37.2
AmuVC1464	1689	Swollenin/CBM10	E	29.5	34.9	41.7	48.0
AmuVC1601	1485	CE4/CBM18	E	56.5	82.8	53.2	93.3
AmuVC2112	1911	CBM10	E	141.7	173.9	41.0	71.0
AmuVC2440	1869	CBM10	E	25.2	71.5	88.1	41.8
AmuVC2668	1488	GH114	E	17.3	93.0	45.5	50.0
AmuVC2684	2055	GH13_C/CBM48/GH13	E	59.5	45.5	44.3	40.0
AmuVC279	1401	CBM10/CBM1	E	30.1	69.1	106.9	163.6
AmuVC2993	1554	CBM10/GH11	E	128.0	51.2	34.4	37.5
AmuVC338	1128	CBM10	E	63.1	49.1	99.1	125.6
AmuVC3496	1938	CBM10	E	30.4	94.0	44.6	48.8

ORF	Length (bp)	CAZy Domains	Clade	FPKM			
				GCS	Oat Spelt Xylan	Barley Straw	Alfalfa Hay
AmuVC3673	1767	CBM10/CBM6/GH43	E	38.2	33.7	43.2	68.2
AmuVC408	1563	CE1	E	6.1	387.4	61.3	62.6
AmuVC420	1743	CBM10	E	28.3	72.4	33.0	76.5
AmuVC4510	744	CBM18	E	12.4	95.0	75.9	144.2
AmuVC4852	1095	CBM10/CBM1	E	48.9	207.7	15.9	100.2
AmuVC514	5130	GH33	E	21.1	59.4	234.8	39.6
AmuVC6446	2439	CBM10/CBM1	E	57.4	48.8	24.4	29.4
AmuVC677	2511	CBM21	E	50.2	63.5	213.2	84.1
AmuVC7056	1437	CBM18	E	69.2	82.7	90.9	184.0
AmuVC7831	2091	CBM18	E	13.0	94.3	128.9	129.7
AmuVC784	1731	CBM10/GH5	E	23.8	58.3	56.7	179.6
AmuVC8984	4731	CBM10/CBM1	E	28.5	64.5	59.3	67.8
AmuVC923	3573	CBM18/CBM13	E	32.1	118.7	59.4	86.1
AmuVC9648	1083	CBM18	E	24.6	34.0	85.9	166.6
AmuVN2600	1938	CBM10	E	35.9	31.7	21.0	56.8
AmuVN3221	1365	CBM10/GH45	E	31.2	87.5	136.0	227.6
AmuTC10428	1947	CBM10	F	0.4	24.5	96.1	43.8
AmuTC1512	1935	GH3_C/GH3	F	14.6	27.5	132.2	393.7
AmuTC17	2439	CBM10/CBM4_9/CBM4_9/GH43/CBM37_Blast	F	5.6	82.7	84.9	141.4
AmuTC1982	3372	GH115	F	6.4	36.3	233.9	276.0
AmuTC26300	1185	GH5	F	1.1	28.7	19.2	50.1
AmuTC3377	1269	CBM18	F	0.6	39.7	78.1	17.4
AmuTC3752	2307	PL4/CBM10	F	4.5	30.0	64.3	195.7
AmuTC3767_seq2	1014	CE12/CBM1	F	6.0	35.4	52.4	77.5
AmuTC4153	1017	GH18	F	2.9	87.8	11.0	31.5
AmuTC5124	1926	CBM10/GH10	F	5.1	86.9	34.2	122.8
AmuTC5594	2247	CBM48/GH13	F	5.4	24.2	79.7	41.9
AmuTC6167	1008	CBM10/CBM1	F	6.7	74.5	75.5	103.3
AmuTC9091	2622	GH27	F	9.6	25.1	81.1	24.3
AmuVC10042	1098	CE6/CBM10	F	11.6	14.5	76.4	90.3
AmuVC10050	1821	CBM13/CBM10/GH10	F	13.3	39.1	232.6	323.6
AmuVC10133	1974	CBM10	F	1.7	49.9	43.2	46.5
AmuVC10153	1890	CBM10	F	8.6	25.4	81.9	98.0
AmuVC10238	1635	CE6/CBM10	F	10.1	43.9	91.3	247.9
AmuVC10423	1794	CBM1/GH10	F	18.1	12.6	79.8	134.8
AmuVC10435	4830	GH2/Bgal_small_N/CBM10/	F	3.8	98.2	137.3	550.0
AmuVC10470	1626	CBM10/GH53	F	4.2	15.2	21.1	50.7
AmuVC105	1857	CBM10	F	4.1	10.5	29.3	99.5

ORF	Length (bp)	CAZy Domains	Clade	FPKM			
				GCS	Oat Spelt Xylan	Barley Straw	Alfalfa Hay
AmuVC10545	1860	CE1	F	0.9	33.2	66.3	20.3
AmuVC10637	1755	GH18	F	4.5	63.3	91.7	35.8
AmuVC1073	1323	CBM18	F	2.5	7.8	318.1	418.9
AmuVC1090	1839	CBM10	F	0.6	4.6	38.6	124.3
AmuVC1177	1905	CBM18	F	0.2	3.8	68.2	41.7
AmuVC11891	1878	CBM10	F	2.1	60.9	63.6	311.0
AmuVC12009	2175	CBM10	F	7.3	11.0	152.1	423.4
AmuVC12225	1488	CBM10	F	0.4	41.8	87.9	51.9
AmuVC1325	1125	CBM10/GH11	F	7.0	119.7	212.6	200.8
AmuVC1405	2478	GH3_C/GH3	F	3.6	66.4	105.0	45.4
AmuVC1718	1800	CBM10/GH10	F	11.0	87.5	234.4	493.8
AmuVC1852	1320	CBM18	F	0.3	9.4	252.9	68.7
AmuVC2037	1125	CBM10	F	1.3	59.2	135.3	144.0
AmuVC2117	1104	GH45	F	3.0	4.2	114.3	403.8
AmuVC2420	1401	CBM10/GH10	F	9.0	28.3	245.8	1721.4
AmuVC2424	1626	CBM1/GH43	F	0.8	36.1	35.4	96.5
AmuVC2605	1134	CBM18	F	7.4	86.4	11.6	46.8
AmuVC3357	3558	CBM10/CBM1	F	7.7	61.2	44.8	78.7
AmuVC3418	1389	CBM10	F	7.3	25.2	76.1	66.5
AmuVC3661	1707	CE1/CBM10	F	2.3	8.0	92.9	122.9
AmuVC3782	1668	CBM10/GH5	F	2.9	27.4	28.3	289.3
AmuVC4517	1116	CBM10	F	0.7	7.4	69.0	42.9
AmuVC4623	993	CBM10	F	0.1	13.9	27.4	23.9
AmuVC479	1554	CBM18	F	0.2	16.7	29.8	60.1
AmuVC486	2772	CBM10/GH5	F	3.7	36.2	15.3	96.1
AmuVC686	1716	GH117/CBM13/CBM10/GH43	F	3.9	9.0	39.1	85.3
AmuVC808	2184	CBM10	F	4.5	39.6	98.5	71.2
AmuVC900	1008	CBM10/GH16	F	8.8	102.7	143.3	444.6
AmuVC9708	2307	CBM10/GH9	F	2.7	62.3	118.2	154.1
AmuVC9725	963	CE12/CBM1	F	5.5	79.5	32.6	136.2
AmuVC9760	1428	CBM1/GH6	F	9.9	18.0	45.9	126.4
AmuVC9781	4044	CBM26/GH31	F	7.3	45.5	68.9	105.2
AmuVN3685	1536	CBM18	F	3.4	82.5	17.5	69.6
AmuTC10803	1098	CE10	G	0.2	6.7	8.0	37.8
AmuTC1101	1980	CBM18/CBM10/GH18	G	47.6	5.4	29.5	21.6
AmuTC110785	1323	GH64	G	0.5	0.1	3.1	5.8
AmuTC134187	681	GH37	G	8.5	0.5	0.1	0.8
AmuTC14421	942	CE10	G	0.7	0.0	0.6	30.2
AmuTC144673	906	GH24or104	G	0.2	1.4	1.5	1.7
AmuTC14860	2031	CBM1	G	6.1	14.0	17.3	9.2

ORF	Length (bp)	CAZy Domains	Clade	FPKM			
				GCS	Oat Spelt Xylan	Barley Straw	Alfalfa Hay
AmuTC1541Seq2	1320	PL1/CBM1	G	1.0	0.2	3.2	60.1
AmuTC17634	501	CE4	G	0.2	1.0	22.5	6.2
AmuTC17666	804	CBM18	G	6.7	53.2	4.9	6.3
AmuTC17966	1299	CBM13	G	7.4	35.9	1.9	2.8
AmuTC18125	957	CE10	G	1.0	23.7	51.4	5.1
AmuTC19625	735	GH16	G	0.3	10.0	12.7	50.8
AmuTC196606	549	CBM1/GH5	G	0.1	0.1	0.3	2.6
AmuTC20089	498	CE1	G	4.2	3.0	4.7	13.3
AmuTC20259	1569	CBM1	G	2.2	4.0	21.9	17.5
AmuTC20360_seq2	954	CBM18	G	3.2	5.3	7.8	3.7
AmuTC20682	1317	CE10	G	9.5	15.2	27.4	15.6
AmuTC20741	1722	CBM10	G	8.3	18.7	11.8	18.5
AmuTC20928	921	CE10	G	10.0	34.5	10.0	31.0
AmuTC21521	1101	GH109	G	12.6	12.1	17.8	14.6
AmuTC21906	1392	CE4	G	4.7	10.8	27.3	17.4
AmuTC22233	1491	CE10	G	9.8	17.3	7.1	10.6
AmuTC2294	1605	CBM10/GH9	G	1.5	8.5	6.8	43.2
AmuTC2294_seq2	1266	CBM10/CBM6	G	19.8	23.7	27.0	42.5
AmuTC2294_seq4	1437	CBM10/GH9	G	1.7	7.5	5.3	40.8
AmuTC23754	846	GH31	G	8.4	6.8	9.3	7.4
AmuTC24443	1440	CBM1	G	0.1	0.9	56.0	9.7
AmuTC24689	963	GH3	G	5.1	4.7	22.0	4.6
AmuTC24777	582	CE10	G	22.5	1.2	1.4	6.3
AmuTC25985	2688	GH47	G	10.4	8.6	9.7	6.8
AmuTC28640	498	CE4	G	0.4	0.0	0.2	21.2
AmuTC28884	483	CBM10	G	1.1	6.2	9.6	4.9
AmuTC29801	1026	CE10	G	11.9	32.4	1.1	28.7
AmuTC29844	1110	GH25	G	20.1	10.2	28.9	17.3
AmuTC30267	1617	GH37	G	13.6	0.6	0.1	1.1
AmuTC30342	729	CBM21	G	3.6	3.2	24.1	14.8
AmuTC30990	2040	GH20	G	6.9	6.8	8.0	8.0
AmuTC31964	915	CE4	G	6.7	3.8	13.9	5.4
AmuTC33256	1164	GH16	G	0.0	0.6	0.5	0.0
AmuTC34112	1047	GH109	G	2.8	2.2	17.2	7.8
AmuTC34284	459	CBM13	G	4.3	4.9	7.2	19.6
AmuTC3711	1098	GH5	G	8.4	4.8	9.1	214.3
AmuTC37304	780	CBM1	G	0.0	3.8	7.7	13.8
AmuTC3759	927	GH24or104	G	2.5	27.0	0.8	37.7
AmuTC37632	927	CE10	G	10.0	23.5	8.4	32.3
AmuTC3767	1077	CE12/CBM1	G	11.0	16.4	26.0	60.1
AmuTC39485	1464	GH117/CBM10/GH	G	0.4	7.3	6.9	16.2

ORF	Length (bp)	CAZy Domains	Clade	FPKM			
				GCS	Oat Spelt Xylan	Barley Straw	Alfalfa Hay
		43					
AmuTC41693	954	CE16	G	3.5	5.0	5.4	8.5
AmuTC4220_seq2	777	GH114	G	12.2	22.2	6.0	15.1
AmuTC42946	1365	CBM18	G	8.2	20.2	2.6	4.2
AmuTC45468	843	CBM10	G	0.3	13.4	7.8	5.3
AmuTC45752	795	CE12	G	1.3	8.4	4.1	13.4
AmuTC46314	2097	CBM13	G	12.1	10.7	1.6	2.0
AmuTC55544	1242	GH109	G	5.7	3.0	14.0	5.7
AmuTC61070	1956	CBM26/CBM25/GH13	G	0.3	0.4	15.8	12.8
AmuTC62238	1449	CE10	G	8.6	4.1	3.7	5.1
AmuTC62584	831	GH114	G	0.3	0.2	1.7	11.3
AmuTC64800	462	CBM50	G	2.6	4.0	6.0	5.2
AmuTC7633	612	CE4	G	0.1	2.3	49.8	20.4
AmuTC7767	987	GH114	G	44.3	14.3	42.6	28.6
AmuTC79685	453	CE2 or CE3	G	5.6	0.8	4.8	2.7
AmuTC79765	870	CE1	G	9.2	5.0	3.5	5.3
AmuTC8236	1500	GH114	G	3.8	1.4	31.0	45.6
AmuTC84759	1116	GH76	G	0.0	0.0	0.0	0.0
AmuTC8654	3639	TIG	G	21.2	28.9	57.8	9.1
AmuTC87312	1188	CBM18	G	0.1	0.2	3.2	4.3
AmuTC87421	825	CBM10	G	1.2	1.5	2.0	4.6
AmuTC8966	1626	CBM1	G	2.5	5.6	27.7	6.7
AmuTC8982	984	CE7	G	9.7	68.7	35.9	22.0
AmuTC915	648	CBM10/CBM1	G	2.3	7.9	6.6	9.5
AmuTC9150	2064	PL09	G	10.7	28.6	31.6	43.5
AmuVC10313	1920	CBM10	G	17.7	63.9	4.9	8.9
AmuVC10379	1305	CBM1	G	8.0	29.6	25.8	42.5
AmuVC10419	1263	CE4/CBM18	G	17.0	1.5	36.8	60.3
AmuVC10672	1293	CE1/CBM10	G	25.3	22.3	1.0	11.2
AmuVC1070	1245	CBM18	G	8.5	20.8	11.2	37.9
AmuVC10773	1560	CBM10	G	0.2	46.8	8.8	29.0
AmuVC10781	1299	CBM10	G	26.0	30.9	11.2	18.7
AmuVC10945	1236	CBM10	G	4.1	9.2	5.0	15.0
AmuVC12239	1146	CBM10	G	0.5	2.3	14.4	47.5
AmuVC12453	2154	CBM10/GH9	G	7.2	1.6	17.7	54.8
AmuVC1304	1608	CBM10/GH18	G	20.3	6.5	17.4	13.5
AmuVC1484	1329	CE4/CBM50	G	25.0	24.5	29.6	27.8
AmuVC1613	1686	CBM18	G	9.0	6.1	14.4	5.9
AmuVC1727	1065	CE1/CBM10	G	10.5	5.7	9.0	25.8
AmuVC1775	1716	CBM18	G	16.2	35.9	16.5	16.9

ORF	Length (bp)	CAZy Domains	Clade	FPKM			
				GCS	Oat Spelt Xylan	Barley Straw	Alfalfa Hay
AmuVC1871	1404	GH3_C	G	6.0	6.3	24.2	4.9
AmuVC2008	930	CBM10	G	46.7	16.2	15.2	10.8
AmuVC21	2562	GH3_C/GH3	G	5.4	5.1	6.4	166.1
AmuVC2158	1263	CBM10	G	14.5	20.0	21.0	32.3
AmuVC2212	1110	CBM18	G	1.1	53.1	6.1	38.2
AmuVC2262	1578	CE6/CBM10	G	31.1	13.2	25.5	45.2
AmuVC2278	2982	CBM10/CBM1	G	4.8	23.8	15.2	14.1
AmuVC2376	825	CBM10	G	5.5	1.3	39.1	57.2
AmuVC2498	1344	CBM18	G	16.3	22.5	38.1	26.1
AmuVC2604	891	GH3	G	0.5	0.2	2.1	73.9
AmuVC262	2775	CBM18/CBM10	G	13.0	22.9	31.8	17.0
AmuVC2712	1350	CBM52/GH16	G	13.8	19.4	19.1	62.6
AmuVC2926	957	CBM18	G	0.3	1.5	46.1	26.3
AmuVC3592	1779	CBM10	G	5.5	23.5	21.3	21.3
AmuVC3612	2025	CBM18	G	0.6	21.6	8.5	32.7
AmuVC363	1833	CBM10/GH43	G	9.3	11.4	13.2	56.5
AmuVC3671	2172	CBM48/GH13	G	12.0	37.1	28.5	35.0
AmuVC375	5916	CBM18	G	8.5	76.2	0.9	38.1
AmuVC3846	2172	CBM18	G	29.7	36.1	8.1	35.1
AmuVC4088	1647	GH5	G	14.4	13.3	13.3	29.6
AmuVC4244	1227	CBM18	G	2.0	2.4	13.9	11.9
AmuVC4325	1641	CBM18	G	3.4	1.1	56.0	63.9
AmuVC4403	1389	CBM18	G	0.1	28.0	4.9	21.8
AmuVC4589	1296	CBM10	G	3.2	12.4	29.0	20.4
AmuVC508	1683	CBM10/CBM6/GH 26/CBM35	G	3.9	0.9	10.9	841.0
AmuVC5227	2697	GH32C/GH32N	G	4.2	6.6	11.0	109.7
AmuVC537	1617	CBM10	G	18.2	29.1	37.7	31.4
AmuVC568	1038	CBM10	G	0.2	0.1	7.7	48.4
AmuVC5814	1011	GH114	G	23.0	69.3	6.9	22.6
AmuVC6325	981	CBM1	G	2.7	46.2	16.2	2.9
AmuVC728	1953	CBM18	G	11.2	67.4	3.1	6.0
AmuVC893	1350	PL3/CBM1	G	9.2	6.2	7.8	33.3
AmuVC9126	1119	GH114	G	0.2	0.1	1.8	6.9
AmuVC9538	3726	CBM18/CBM13	G	0.5	2.2	0.5	6.3
AmuVC9660	5346	CBM18	G	15.5	36.9	6.8	3.8
AmuVC9821	1539	CE15/CBM10	G	2.2	0.4	23.9	466.1
AmuVN11467	1647	CBM18	G	8.8	9.9	26.2	12.6
AmuVN13467	2139	CBM10	G	4.5	25.7	22.8	30.4
AmuTC2525	2178	GH13/CBM26	NA	0.0	4.3	646.1	0.0