Bioprospecting Carbohydrate-Active Enzymes in Lignocellulose-Degrading Microcosms Enriched from Pulp Mill Anaerobic Granules and Digestive Microbiomes of Canadian Beaver and North American Moose

by

Mabel Ting Wong

A thesis submitted in conformity with the requirements

for the degree of Doctor of Philosophy

Department of Chemical Engineering & Applied Chemistry

University of Toronto

Abstract

Bioprospecting Carbohydrate-Active Enzymes in Lignocellulose-Degrading Microcosms Enriched from Pulp Mill Anaerobic Granules and Digestive Microbiomes of Canadian Beaver and North American Moose

Mabel Ting Wong

Doctor of Philosophy

Department of Chemical Engineering & Applied Chemistry

University of Toronto

2018

Lignocellulosic biomass, including wood fibre, is an abundant resource for biorefining; however, the structural heterogeneity and occurrence of inhibitors present challenges to corresponding bioconversion technologies. In an effort to identify carbohydrate-active enzymes (CAZymes) with potential to transform wood-derived feedstocks, multiple lignocellulosic microcosms were established from pulp mill anaerobic granules, beaver droppings and moose rumen and cultivated for over 3 years on (i) cellulose, (ii) cellulose + lignosulphonate, (iii) cellulose + tannic acid, and (iv) pretreated poplar. Microbial community analysis of 16S rRNA amplicons revealed post-enrichment decrease in species richness and substrate-induced convergence. Amendment with added inhibitors (ii, iii) or pretreated poplar particularly enriched known biomass degraders along with microbes belonging to poorly-described lineages including *BSV26, SJA-28, TG3* classes, *OPB54* and *Cloacamonales* orders.

Comparative metagenomics targeting cellulose- and pretreated poplar-fed microcosms revealed CAZyme profiles that converged based on substrate for gut microbiomes, while a separate cluster for anaerobic granules revealed the impact of inoculum. Compared to the cellulose-fed microcosms, the pretreated poplar-fed counterparts were enriched with reported broad substrate families with cellulolytic and hemicellulolytic activities (GH2, GH3, GH5, and GH43), as well as putative β-L-arabinofuranosidases (GH127), glucuronoyl methylesterases (CE15), and pectinases (PL1, GH28, CE8, and GH105). While no major differences emerged from the profiles of proteins with unknown function within the 416 predicted polysaccharide utilization loci and catalytic cellulosomal subunits, 2 carbohydrate-binding proteins with domains of unknown function (DUF3459-CBM48-GH13_10 and DUF5011-CBM4) were consistently present in pretreated poplar-fed microcosms only.

Metasecretomes from the microcosms at ~50% and ~80% g COD substrate to biogas conversion were functionally characterized. Despite the similar saccharification capacities of different enrichments, protein profiles of the pretreated poplar-fed microcosms observed by SDS-PAGE were distinct from the cellulose-fed ones. Moreover, the emergence of new bands and greater activity on pectin, xylans and mannans from secretomes at the later stage implied a succession of enzymes that hydrolyze β-(1→4)-glycosidic bonds in cellulose before other polysaccharides.

Collectively, putative lignocellulolytic proteins specialized for processing forest fibre were revealed via comparative meta-analyses for downstream characterization. Also, the potential of the microcosms serving as a repertoire of biocatalysts for developing wood specific-enzymatic cocktails was confirmed.

This thesis is dedicated to my parents Janet and Kan,

as well as people who are not privileged to receive education

Acknowledgement

My tremendous thanks are dedicated to:

- Ph. D. supervisors: Prof. Emma R. Master and Prof. Elizabeth A. Edwards
- Supervisory committee: Prof. John Parkinson, Prof. Alexander Yakunin
- Collaborators: Dr. Marie Couturier, Dr. Weijun Wang, Fakhria M. Razeq, Prof. Nicolas Terrapon, Prof. Bernard Henrissat, Dr. Vincent Lombard, Dr. Pascal Lapebie
- Research funding agency: Natural Sciences and Engineering Research Council of Canada, Genome Canada, European Research Council
- Research support: Dr. Thu Vuong, Dr. Camilla Nesbø, Yang Xiu
- Support staff: Endang Susilawati, Dean Robson, Weijun Gao, Vinthiya Paramananthasivam, Dr. Sean Caffrey
- Members of the Master, Edwards, and BioZone lab
- My beloved parents Janet and Kan, siblings Jackson, Jennifer, Wendy and William, partner Danny and extended family
- My ever supportive friends Fakhria M. Razeq, Zahra Choolaei, Maryam Arefmanesh, Dr. Marie Couturier and too many to list

Declaration

An ethics approval from an Animal Care and Use Committee was not required by the Office of Research Ethics of the University of Toronto, as the moose rumen sample was collected from a dead moose that was hunted in the wild for meat by a registered hunter with a license authorized by the Ministry of Natural Resources and Forestry under Government of Ontario, Canada.

I declare that this research proposal represents my own work, except where due acknowledgement is made, and that it has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma or other qualifications.

Chapters 3 and 4 of the thesis are adapted from the following peer-reviewed publications:

Wong, M.T., Wang, W.J., Lacourt, M., Couturier, M., Edwards, E.A., and Master, E.R. (2016). Substrate-driven convergence of the microbial community in lignocellulose-amended enrichments of gut microflora from the Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*). *Frontiers in Microbiology* 7.

Wong, M.T., Wang, W., Couturier, M., Razeq, F.M., Lombard, V., Lapebie, P., Edwards, E.A., Terrapon, N., Henrissat, B., and Master, E.R. (2017). Comparative metagenomics of cellulose- and poplar hydrolysate-degrading microcosms from gut microflora of the Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*) after long-term enrichment. *Frontiers in Microbiology* 8.

Here are the roles of the researchers involved in the collaborative work presented in this thesis: Mike W. Lacourt collected the environmental samples and established the enrichment cultures in 2009; Dr. Weijun Wang maintained the enrichment cultures between 2011 and 2013, and prepared DNA samples for sequencing in 2013. From the 2013 summer onwards, I maintained and scaled-up the enrichment cultures, performed sequence and proteomic analyses, analysis and visualization of the metagenomic and metasecretomic data, and academic writing. Prof. Bernard Henrissat, Prof. Nicolas Terrapon, Dr. Vincent Lombard, and Dr. Pascal Lapebie from Aix-Marseille Université contributed to the annotation of carbohydrate active enzymes, carbohydrate binding modules, polysaccharide utilization loci, as well as data interpretation. Dr. Marie Couturier and Fakhria M. Razeq contributed to the data interpretation. In the metasecretomic analysis, Dr. Thu Vuong provided guidance on experimental design and Yang Xiu assisted in the concentration of secretomes. As part of the future work, I initiated the analysis of LC-MS/MS data of the metasecretomes and metagenome-assembled genomes based on the pipeline developed by Dr. Camilla Nesbø.

In my capacity as a member of the Edwards lab, I also contributed to the bioinformatic analysis for the following publication in collaboration with the Tang lab:

Zhai, W.W., Wong, M.T., Luo, F., Hashmi, M.Z., Liu, X.M., Edwards, E.A., Tang, X.J., and Xu, J.M. (2017). Arsenic methylation and its relationship to abundance and diversity of arsM genes in composting manure. *Scientific Reports* 7.

Table of Contents

List of Tables

Supplementary Tables

List of Figures

Supplementary Figures

List of Abbreviations

AA	auxiliary activity
Abyss	Assembly By Short Sequences
CAZyme	carbohydrate-active enzyme
CBM	carbohydrate-binding module
CE	carbohydrate esterase
COD	chemical oxygen demand
dNTP	deoxynucleoside triphosphate
DUF	domains of unknown function
ECF	extra-cytoplasmic function
G	coniferyl alcohol
GH	glycoside hydrolase
GT	glycosyltransferase
H	*p*-coumaryl alcohol
HPAEC	high-performance anion exchange chromatography
LC-MS/MS	liquid chromatography-mass spectrometry
NCBI	National Center for Biotechnology Information
NCBI-NR	National Center for Biotechnology Information non-redundant
ORF	open reading frame
OTU	operational taxonomic unit
PAHBAH	*para*-hydroxybenzoic acid hydrazide
PCA	principal component analysis
PCoA	principal coordinate analysis
PCR	Polymerase chain reaction
PL	polysaccharide lyase
*p*NP	4-*para*-nitrophenyl
PULs	polysaccharide utilization loci
QIIME	Quantitative Insights Into Microbial Ecology
S	sinapyl alcohol
sus	starch utilization system
ToF-SIMS	time-of-flight secondary ion mass spectrometry
UPGMA	Unweighted Pair Group Method with Arithmetic mean

Chapter 1 Introduction

1.1 Background

In face of the evanescent fossil resources and the related environmental pollution, the 21st century marks a transition from societies reliant on petroleum to ones increasingly sustained by renewable resources and bio-based economies. Lignocelluloses as agricultural, industrial, forest residues and energy crops represent the majority of global biomass, and is considered an important renewable resource for the production of bioenergy, liquid biofuels, and specialty chemicals (Kumar et al., 2008). In Canada, the vast forests represent a tremendously abundant source of biomass, which already contributes over 23 billion CAD to Canada's nominal GDP (http://cfs.nrcan.gc.ca/statsprofile) (Natural Resources Canada, 2018). The Bio-pathways Project published in 2011, along with several international assessments of bioeconomy development worldwide, revealed that technologies for high value-added bioproduct synthesis have the potential to significantly boost the economic sustainability of agricultural and forest sectors (Forest Products Association of Canada and FPInnovations, 2011). To foster the competitiveness in our bio-based industry, the Canadian National Research Council has launched the Bio-based Specialty Chemicals Program to accelerate the technology improvement and commercialization of bio-based specialty chemicals from the existing lignocellulosic biomass (Johnston, 2015).

In nature, the transformation of lignocellulose is catalysed by fungal and microbial enzymes, providing a rich resource for bioprospecting useful enzymes that can be applied in forestry for sustainable bio-economy. Lignocelluloses are largely comprised of cellulose (40-50%), hemicelluloses (25-35%) and lignin (20-30%). While the polysaccharide fraction is converted to simpler sugars by the synergistic action of myriad fungal and bacterial enzymes, microorganisms (particularly white-rot fungi from the phylum *Basidiomycota*) have evolved peroxidases and laccases to transform lignin (Kumar et al., 2008). Enzymes that selectively and reproducibly tailor the chemistry of lignocellulose components are recognized as being important to bioproduct diversification (Dale, 2003). Aside from the discovery of new enzyme activities for lignocellulose-transformation, reduction in enzyme costs is key to the growing bio-economy that generate renewable products through sustainable industrial processes. Herein, contributors to cost include high enzyme loading due to substrate heterogeneity and susceptibility to inhibitory compounds, that commonly accompany pretreated wood lignocellulose, such as lignosulphonates and tannic acid (Jonsson and Martin, 2016). While a detoxification process that can sufficiently remove the inhibitors from the pretreated biomass in an economically-viable manner remains under-development, it is essential to discover lignocellulose-active biocatalysts that are resistant to common inhibitory compounds. Over the past decade, several teams have sought new enzymes to improve the cost-effectiveness of lignocellulose bioprocessing; however, most of this research concentrated on transforming annual crops (e.g. corn) and grasses to biofuel (Brulc et al., 2009; Hess et al., 2011). By comparison, Canada is characterized by its vast boreal forest, which generates lignocellulose feedstocks compositionally divergent to agricultural fibres and grasses (Shiro, 2001). Microbial communities that grow on wood lignocellulose, including recalcitrant inhibitors from pretreatment and industrially-relevant pretreated wood, are hence more suitable

for mining relevant hydrolytic enzymes for processing biomass resources that dominate the northern hemisphere.

My PhD project explored the enzyme activities encoded and produced by the digestive microflora of iconic Canadian mammals, namely the Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*), as well as the anaerobic granules from a Canadian pulp mill. Specifically, the project began by enriching the most pertinent microbial biomass degraders from these sources through serial anaerobic cultivation on various lignocellulosic substrates, an endeavour initiated by former graduate student Mr. Mike Lacourt since 2009 (Lacourt, 2011). Lacourt (2011) established an array of biogas-producing microcosms enriched with i) cellulose only, ii) cellulose plus lignosulphonate, iii) cellulose plus tannic acid, and iv) pretreated poplar from the selected environmental sources. After ~4 years of serial cultivation by Lacourt and Dr. Weijun Wang, the project continued with a subsequent amplicon sequencing of the 16S rRNA genetic marker using 454 pyrosequencing to verify enrichment of each culture (Krober et al., 2009), and metagenomic shotgun sequencing on Illumina-based platform to evaluate the functional potential of cellulose- and pretreated poplar-fed microcosms. In particular, the metagenomes were comprehensively mined for proteins with putative functions in lignocellulose degradation. They included carbohydrate-active enzymes (CAZymes) that contain single or multiple catalytic domains that target lignocellulose, catalytic subunits of plant cell wall-degrading multi-enzyme complexes known as cellulosomes, polysaccharide utilization loci (PULs) that encode protein suites for hydrolysis of polysaccharides and uptake of resultant products (Martens et al., 2009; Grondin et al., 2017), as well as carbohydrate-binding proteins with domains of unknown function (DUFs) to uncover novel activities on lignocellulose. Concurrently, total secreted proteins (metasecretomes) from the scaled-up microbial enrichment

amended with cellulose and pretreated poplar were screened for activity on polysaccharides to assess their expressed lignocellulolytic capacities.

Altogether, the potential of lignocellulose-degrading microcosms serving as a repertoire of polysaccharide-degrading biocatalysts was confirmed. Comparative meta-analyses of the microcosms enriched with strategically chosen substrates revealed putative lignocellulolytic taxa and protein candidates specialized for processing forest fibre for downstream genomic and functional characterization. Aside from the known biomass degraders, several taxonomically and functionally less defined microbial lineages belonged to *BSV26, SJA-28, TG3* classes, *OPB54* and *Cloacamonales* orders were identified as dominant populations in the lignocellulose-degrading microcosms compared to those fed with cellulose only. Continuing with the comparison with cellulose-fed cultures, highlighted protein candidates particularly enriched in pretreated poplar-fed cultures included putative β-L-arabinofuranosidases (GH127), 4-O-methyl-glucuronoyl methylesterases (CE15), pectinases (PL1, GH28, CE8, and GH105), as well as two putative carbohydrate binding proteins with domains of unknown function (DUF3459-CBM48-GH13_10 and DUF5011-CBM4). On the hand, the anticipated substrate-induced differential membership of the proteins with unknown function within the predicted PULs, catalytic cellulosomal subunits, as well as the differential polysaccharide-degrading capacity were absent. Nonetheless, the functionally-active metasecretomes provide opportunities to identify the protein components to develop enzymatic cocktails specialized for pretreated wood fibres. Moreover, the developed secretome preparation method that retains enzyme activity of secretomes from mixed cultures is transferrable to the investigations of other mixed cultures.

1.2 Research hypotheses, objectives and thesis structure

The guiding hypotheses of my study were:

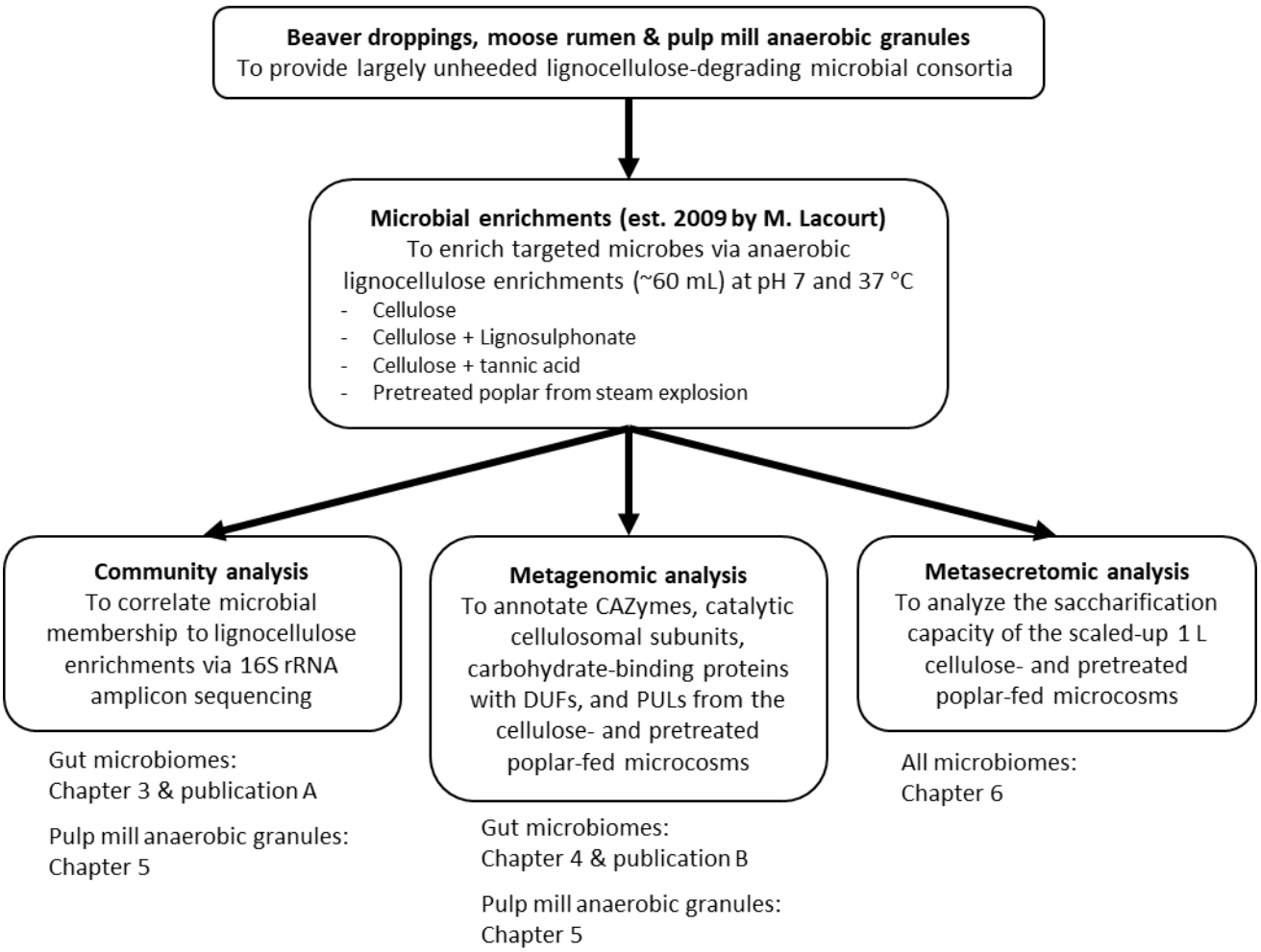
1. Microbial communities from wood-degrading niches will encode CAZymes that efficiently transform lignocellulose from wood resources;
2. Enrichment of microbial communities on lignocellulosic substrates will simplify community composition, thereby facilitating the discovery of relevant enzymes;
3. Comparative analysis of CAZymes, catalytic cellulosomal subunits, carbohydrate-binding proteins with DUFs, and PULs predicted from metagenome sequences will reveal enzymes and proteins with unknown function relevant to lignocellulose modification;
4. Comparative analysis of secretomes from enrichment cultures will reveal differential saccharification capacities based on the fed amendment.

Accordingly, my specific research aims were:

1. Correlation of microbial community membership of anaerobic microcosms from pulp mill anaerobic granules, beaver dropping and moose rumen microcosms to lignocellulose enrichments using 16S rRNA gene amplicon sequencing;
2. Annotation of CAZymes, catalytic cellulosomal subunits, carbohydrate-binding proteins with DUFs, and PULs from the cellulose- and pretreated poplar-fed anaerobic microcosms from pulp mill anaerobic granules, beaver dropping and moose rumen microcosms through shotgun metagenomic sequencing;

3. Analysis of the saccharification capacity of the cellulose- and pretreated poplar-degrading microcosms through metasecretomic analysis of secretomes harvested from each microbial enrichment.

In this thesis, **Chapter 2** serves to provide a summary of literature pertaining to lignocellulase bioprospecting. Structural features of lignocellulose constituents, common inhibitory compounds present in pretreated wood fibres, biological principle and recent discoveries from lignocellulose-degrading microbial communities are described to provide a background to the project. Under the objectives aforementioned, **Chapter 3** describes the correlation of the microflora of beaver dropping and moose rumen microbial cultures to lignocellulose enrichments using 16S rRNA pyrotag sequencing; **Chapter 4** describes the differential profiles of annotated CAZymes, PULs and other putative carbohydrate-active proteins from the cellulose- and pretreated poplar-fed microcosms from the gut microbiomes of beaver and moose. Using the same rationale and methodology, **Chapter 5** switches gear from the gut microbiomes and reports the pyrotag and metagenomic finding of the lignocellulose-degrading microcosms established from the pulp mill anaerobic granules. **Chapter 6** summarizes the polysaccharide-degrading capacities of the metasecretomes collected from the scaled-up cultures that were amended with cellulose and pretreated poplar. Finally, **Chapter 7** offers an overall summary of the thesis, the significance of this work to engineering, and the future directions based on the discovery from the current work. Figure 1.1 shows the outline of the project presented in this thesis.



Publication A: Wong, M.T., Wang, W.J., Lacourt, M., Couturier, M., Edwards, E.A., and Master, E.R. (2016). Substrate-driven convergence of the microbial community in lignocellulose-amended enrichments of gut microflora from the Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*). *Frontiers in Microbiology* 7.

Publication B: Wong, M.T., Wang, W., Couturier, M., Razeq, F.M., Lombard, V., Lapebie, P., Edwards, E.A., Terrapon, N., Henrissat, B., and Master, E.R. (2017). Comparative metagenomics of cellulose- and poplar hydrolysate-degrading microcosms from gut microflora of the Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*) after long-term enrichment. *Frontiers in Microbiology* 8.

Figure 1.1 Thesis project outline

Chapter 2 Literature review

2.1 Wood lignocellulose

2.1.1 Cellulose, hemicellulose, lignin, pectin and extractives

Wood lignocellulose is a composite structure of primarily cellulose (40-50%), hemicelluloses (25-35%), lignin (20-35%) (Table 2.1), as well as minor amounts of pectins and low molecular weight compounds collectively termed extractives (Kumar et al., 2008). Of these plant cell wall components, cellulose and lignin are often described as the most abundant biopolymers on earth. While distinguishing compositions in the structure and composition of hemicelluloses and lignin are observed between softwood, hardwood, and grasses, further variations based on the botanical origin, plant tissue and growth condition exist (Scheller and Ulvskov, 2010; Zhang et al., 2014a). As summarized below, cellulose microfibrils create a cell wall scaffold that imparts compression strength to plant fibre, whereas hemicelluloses are often thought of as bridging polymers between cellulose and lignin, which increases the tensile strength. Lignin can be considered as an encrusting material, which encases the cellulose-hemicellulose network thereby providing resistance to microbial attack and improved water transport. Pectin, on the other hand, is located in the middle lamella and primary plant cell wall.

Table 2.1 Composition of lignocellulose components in various biomass

Biomass	*Cellulose %*	*Hemicellulose %*	*Lignin %*	*Reference*
Hardwood	40–55	24–40	18–25	(Malherbe and Cloete, 2002)
Softwood	45-50	25–35	25–35	(Malherbe and Cloete, 2002)
Corn cobs	45	35	15	(Prasad et al., 2007)
Corn stover	38	26	19	(Zhu et al., 2005)
Rice straw	32	24	18	(Howard et al., 2004)
Wheat straw	29–35	26–32	16–21	(McKendry, 2002)
Switchgrass	35-45	25-30	15-20	(Howard et al., 2004)

Cellulose is the major polysaccharide in plant cell walls, and is composed of β-(1→4)-linked D-glucopyranose repeating units that are predicted to form linear macro-structures having between 7,000-15,000 residues (Figure 2.1) (Stone, 2001). During cellulose synthesis, the cellulose polymers assemble to form microfibrils that are stabilized through hydrogen bonds and hydrophobic interactions between neighbouring cellulose molecules (Stone, 2001). Resulting cellulose microfibrils comprise both amorphous and crystalline regions, fractions of which differ depending on the plant source and cell wall layer. As a raw material, cellulose is currently used to produce paper, packaging materials, and biofuels (Carroll and Somerville, 2009), as well as higher-value products such as cellophane films, textiles, food and cosmetic additives, and pharmaceutical excipients (Iqbal et al., 2013).

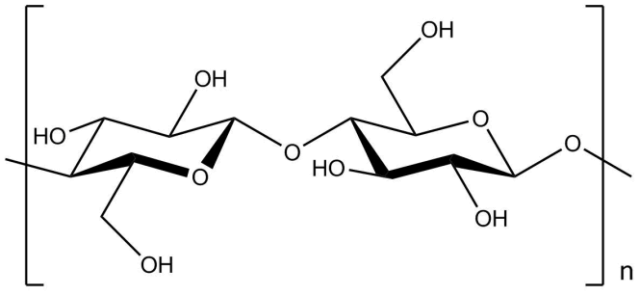


Figure 2.1 Structure of cellulose

Hemicelluloses represent a family of polysaccharides, which, like cellulose, are characterized by backbone structures comprised of β-(1→4) linked sugars (Scheller and Ulvskov, 2010; Rennie and Scheller, 2014). However, hemicelluloses are composed of different sugar constituents such as glucose, xylose, mannose and galactose (Figure 2.2), which can be substituted with branching sugars and acetylated sugars. The degree of polymerization of hemicelluloses is also generally lower than that of cellulose, with estimates ranging from 500-3,000 subunits (Sjöström, 1993). The particular sugar composition and abundance of corresponding hemicelluloses are governed by the plant species and region of the cell wall that it is extracted from (Figure 2.3). For instance, glucuronoarabinoxylan is typically produced by agricultural crops, whereas glucuronoxylan is the main xylan and hemicellulose fraction from hardwood (deciduous) trees (Peng et al., 2012). Comparing the wood hemicelluloses, hardwoods produce low amounts of glucomannan (2-5%), while the main hemicellulose produced by coniferous softwoods is (galacto)glucomannan (~25%) with comparatively low amounts of arabinoglucuronoxylan (5-10%) (Sjöström, 1993).

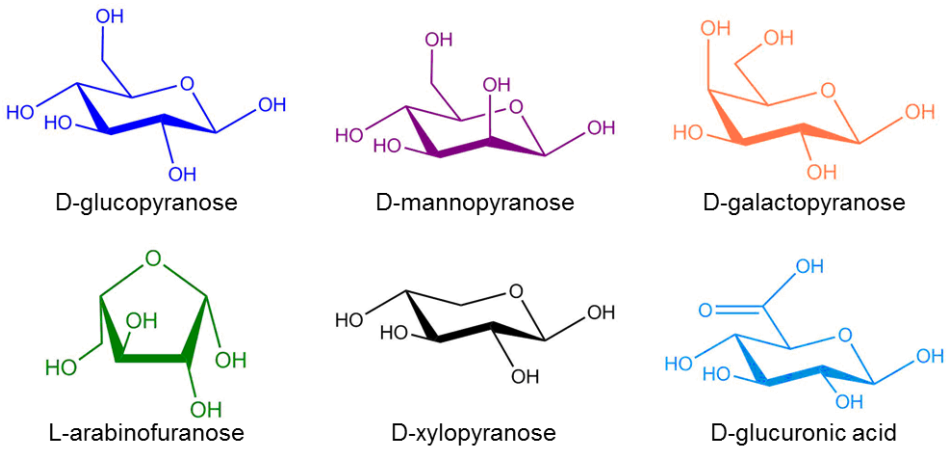


Figure 2.2 Main constituents of hemicellulose

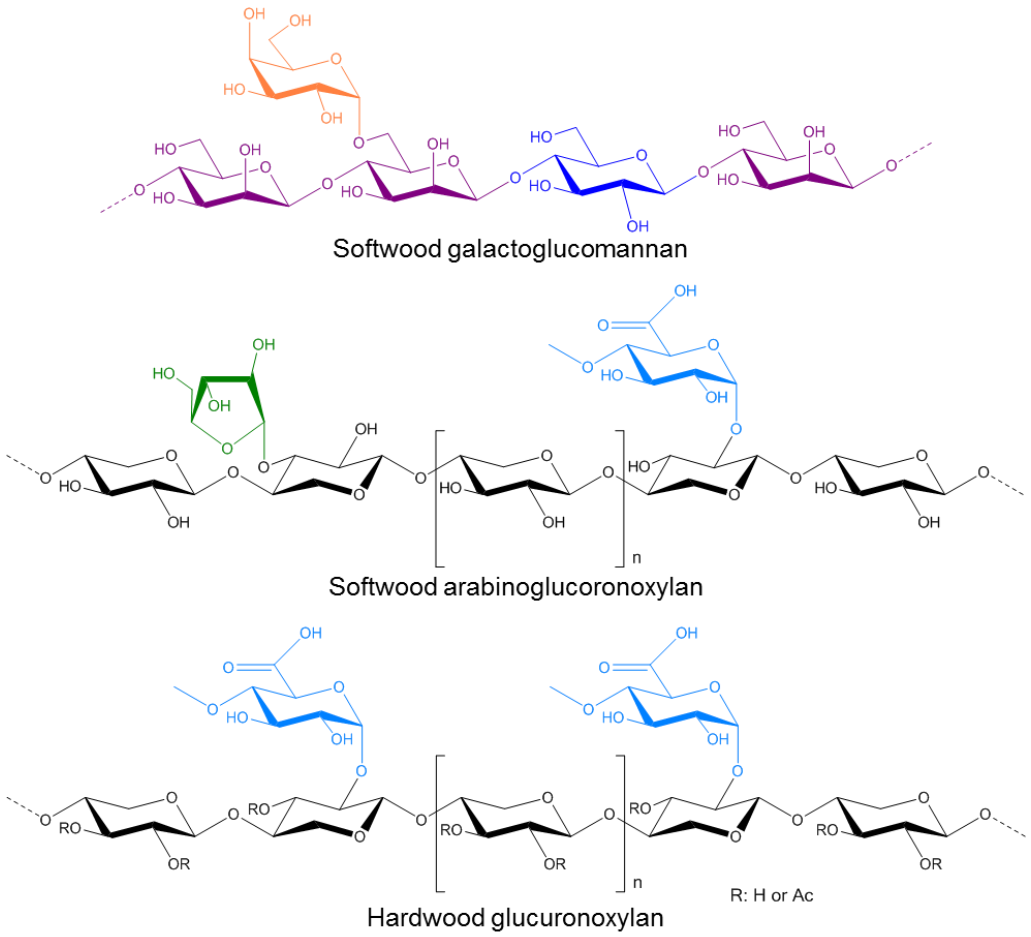


Figure 2.3 Major hemicelluloses in softwood and hardwood

Lignin is a polyaromatic macromolecule comprised of phenyl propanoid subunits, the chemistry of which depends on plant source and fibre type. The main phenyl propanoids present in lignin are *p*-coumaryl alcohol (H), coniferyl alcohol (G) and sinapyl alcohol (S) (Chakar and Ragauskas, 2004; Potters et al., 2010) (Figure 2.4). Lignin from softwood is typically termed guaiacyl lignin (or G-lignin) and is composed principally of coniferyl alcohol (G). By contrast, lignin from hardwood is termed guaiacyl-syringyl lignin (GS-lignin) and is composed of coniferyl (G) and sinapyl alcohol (S) (Chakar and Ragauskas, 2004). Meanwhile, lignin produced by grasses is enriched in *p*-coumaryl alcohol, and termed H-lignin. The phenyl propanoid subunits

are thought to mainly polymerize directly in the plant cell wall through a free-radical mediated process initiated by plant-encoded peroxidases and laccases (Vanholme et al., 2010), although oligomerization, including covalent linkages to hemicellulose, may begin intracellularly (Mottiar et al., 2016). In lignocellulose fermenters, lignin reduces the accessibility of cellulose and hemicellulose substrates through steric hindrance effects and enzyme adsorption, further reducing enzyme accessibility and the ability to recycle enzymes in biorefining (Tu et al., 2009; Rahikainen et al., 2011; Jorgensen and Pinelo, 2017).

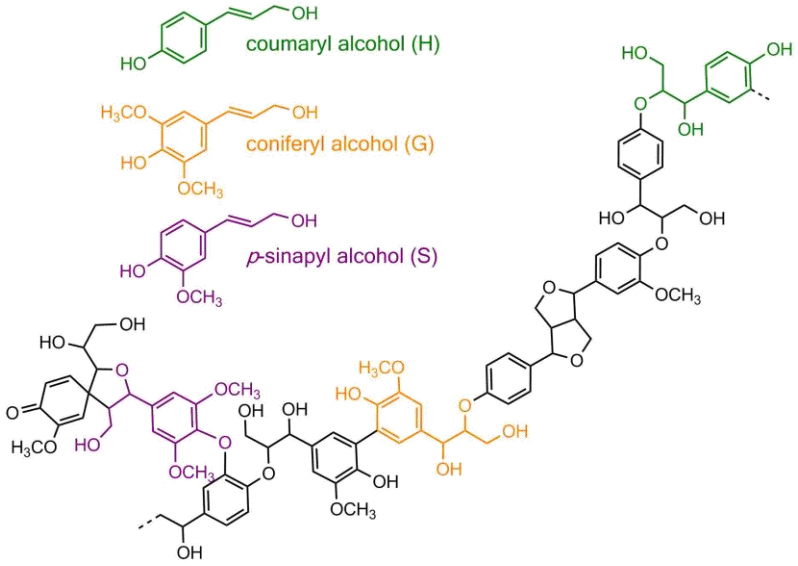


Figure 2.4 Schematic diagram of lignin structure (adapted from (Karkas et al., 2016))

Pectins are highly complex polysaccharides abundant in the primary cell walls of dicotyledonous plants, and in a smaller amounts in the secondary walls of dicots and both cell walls of monocotonous plants (Xiao and Anderson, 2013). The estimated fractions of pectins are 2-10% in grasses and 5% in wood tissue (Fry, 1988; Voragen et al., 2009). Structurally, the pectic macromolecules can be composed of 17 different monosaccharides connected through more than 20 different linkages (Voragen et al., 2009). There are four predominant pectic polysaccharides:

homogalacturonan (~65% of pectin), rhamnogalacturonan I (20-35%), rhamnogalacturonan II (10%), and xylogalacturonan (Mohnen, 2008). The α-(1→4)-linked galacturonan constitutes the backbones of homogalacturonan, rhamnogalacturonan II, and xylogalacturonan, wherein the C6 carboxyl group can be methyl-esterified and the O2 or O3 can be acetylated (O'Neill et al., 2004). On the other hand, the backbone of rhamnogalacturonan I is comprised alternating units of rhamnose and galacturonic acid. Considering the side groups, RG I is decorated with side chains consisting mainly of arabinose and galactose, rhamnogalacturonan II contains clusters of complex side chains comprising 12 different types of sugars in over 20 different linkages at the O2 or O3 position (O'Neill et al., 2004), and xylogalacturonan is partially substituted with β-(1→3)-xylosyl side groups at O3 positions (Mohnen, 2008). Three-dimensional pectic networks are made via further covalent and noncovalent cross-linking between the polysaccharides (O'Neill et al., 2004). Similar to lignin, pectin inhibits the accessibility of plant polysaccharides to enzymes via steric hindrance in the conversion of lignocellulose to biofuels.

Last but not the least, extractives are wide range of low molecular weight, non-structural secondary metabolites concentrated in bark, leaves and heartwood of trees to resist attacks from termite and fungus (Kirker et al., 2013). Named after the ease of extraction with organic or aqueous solvents, extractives vary widely in composition and total concentration among tree species, individual trees, tissue and growth conditions (Nascimento et al., 2013). These compounds are generally categorized into i) aliphatic and alicyclic compounds, ii) gums, iii) phenolic compounds such as tannin (Table 2.2), and iv) other compounds such as sugars, amino acids (Stenius et al., 2000; Nascimento et al., 2013). Among them, representative phenolic compounds such as tannins have been extracted from wood to develop commercial wood

treatments against wood decay fungi (Kirker et al., 2013; Nascimento et al., 2013). The mechanism behind the biological inhibition of hydrolysable tannins is described in 2.1.3.

Table 2.2 Tannin content in various trees (Fulling, 1956; Bakuzis and Hansen, 1965; Lacourt, 2011; Li et al., 2011)

	Wood	*Tissue*	*Tannin weight %*
Softwoods	Sumac	Leaves	31
	Sitka spruce	Bark	24.4
	Western hemlock	Bark	15
	Red spruce	Bark	13
	Eastern hemlock	Bark	9
	Balsam fir	Bark	3.4 - 4.1
	Western red cedar	Wood	0.2 – 0.4
	Douglas fir	Wood	0.27
	Lodgepole pine	Wood	0.061
Hardwoods	American chestnut	Wood	5 – 15
	Oak	Bark	9
	Poplar	Bark	2.5-3.5
	American white oak	Wood	0.2 – 0.4
	Sugar Maple	Wood	0.1
	White ash	Wood	0.063

2.1.2 Wood lignocellulose biorefining

The increasing global interest in biorefining in the 21st century is motivated by the incentives in mitigating climate change, volatile price and limited supply of fossil fuels, as well as the growing demand for bioproducts (Table 2.3). International Energy Agency Task 42 has defined biorefining as '*the sustainable processing of biomass into a spectrum of bio-based products (food, feed, chemicals, and materials) and bioenergy (biofuels, power and/or heat)*' (de

Jong et al., 2012). Under this notion, wood originated from sustainably-managed forests represent a renewable feedstock that offers opportunities for a comprehensive bioproduct portfolio without negatively impacting the food security compared to agricultural biomass and energy crops (Sheldon, 2014). Wood-based biorefining encompasses refining wood biomass into intermediate outputs (cellulose, hemicellulose, lignin, extractives) to be processed into a spectrum of bioproducts and bioenergy via chemical and/or biological conversions such as enzyme treatment and fermentation (Kobayashi and Fukuoka, 2013). While lignocellulose biorefineries were broadly divided into 'biofuel-driven' and 'bioproduct-driven' categories (Cherubini et al., 2009), it has become obvious that co-production would generate most economic benefit (Ed et al., 2012). According to last year's market research on biorefinery products, the global market for bioproducts is projected to reach US$714.6 billion by 2021 from US$466.6 billion in 2016 at a compound annual growth rate of 8.9% (BCC Research, 2017).

Table 2.3 Global market potential for bioproducts (extracted from (Natural Resources Canada, 2016))

Bioproduct	*Compound annual growth rate (%), 2009 to 2015*	*Global market potential (US$ billion) in 2015*
Green chemicals	5.3	62.3
of which Alcohols	5.3	62
Bioplastic and plastic resins	23.7	3.6
Platform chemicals (chemicals derived from biomass and used to make other chemicals)	12.6	4
Wood fibre composites	10	35
Glass fibre market	6.3[1]	8.4
Carbon fibre	9.5	18.6

[1] Compound annual growth rate for 2010–2015

Such initiative for wood lignocellulose biorefining is particularly relevant to Canada, whose 347 million hectares of forest amounts to 9% of the world's forests (Natural Resources Canada, 2017). In terms of tree species, there are about 1,200 that exist naturally in North America and ~100 are of commercial importance; meanwhile the corresponding estimated numbers in Europe are 100 and 20 (Yang and Jaakkola, 2011). Aside from the vast forest resource, residues and waste stream from existent pulp and paper mills represent available lignocellulosic feedstock to biorefining for value-added products, providing alternative revenue when the global demand for newsprint continues to decline (Resolute Forest Products, 2017). To capitalize on the vast wood resource and meet global demands, Canadian pulp and paper mills are undergoing a transformation into modern biorefineries over the past decade by integrating biomass conversion processes and equipment to allow the production of biofuels and a host of biomaterials and biochemicals (Turriff, 2015; Hunter, 2018). A Bio-based Specialty Chemicals Program was also launched by the Canadian National Research Council to accelerate the technology improvement and commercialization of bio-based specialty chemicals from the existing lignocellulosic biomass to foster the competitiveness in our bio-based industry (Johnston, 2015).

2.1.3 Lignocellulose-derived inhibitors in pretreated wood for biological processing

Wood pretreatments via physical and chemical methods are successfully developed to fractionate lignocellulose into its components and facilitate downstream processing. However, they also form a number of lignocellulosic by-products that hamper enzymatic hydrolysis and microbial fermentation. Indeed, such inhibition is considered a major bottleneck in the production of bioethanol from lignocelluloses (Kim, 2018).

Steam explosion represents a competitive physical pretreatment that unlocks the recalcitrance of lignocellulosic feedstock by reducing particle size, defibrating the substrate, disrupting the lignin network and hydrolysing hemicelluloses (Cara et al., 2006; Van Heiningen, 2006; Dererie et al., 2011). During this pretreatment, debarked wood chips are treated with hot steam (160 to 260 °C) under pressure (20 to 50 bar) briefly (30 seconds to 20 min), causing a rapid release of acids from the acetylated wood fibres, and in-turns an auto-hydrolysis of hemicelluloses and a decrease in the β-O-4 linkages in lignin (Robert et al., 1988; Al-Dajani et al., 2009). The following rapid decompression to atmospheric pressure in the reaction vessel opens up the wood structure as the steam trapped within is explosively released (Glasser and Wright, 1998; Martin-Sampedro et al., 2011). Chemical structure of the resultant pretreated wood can be altered by adjusting the temperature and treatment time (San Martin-Davison et al., 2015); for instance, the rise in temperature typically improves the removal of hemicellulose at the cost of sugar degradation. Compared to chemical pretreatments such as acid hydrolysis, steam explosion operates without the need for added stages for chemical addition and recycling, solubilizes hemicelluloses without excessive degradation, and generates exploded wood with softened lignin (Avellar and Glasser, 1998; Ahvazi et al., 2007). The enhanced susceptibility of the fibres by microbial enzymes facilitates the downstream enzymatic hydrolysis to generate fermentable sugars for ethanol production (Figure 2.5) (Olsson and Hahn-Hagerdal, 1996).

Figure 2.5 Steam explosion pretreatment of wood for biofuel production

Chemical pretreatments such as Kraft and sulphite pulping typically involves the extraction of cellulose from the woody feedstock via delignification. Kraft pulping accounts for > 90% of pulp production (Sixta, 2006). The process begins with cooking debarked wood chips in an aqueous solution containing sodium sulphide and sodium hydroxide, also known as white liquor, in a digester at 160-170 °C (Figure 2.6). Lignin is solubilized by sodium sulphide via sulphonation and cleavage of the ether linkages within the lignin network by sodium hydroxide; meanwhile, approximately 50% of hemicellulose is dissolved as saccharine acids in the liquor. On the other hand, sulphite pulping produces ~6% of the pulp market and involves treating the wood feedstock with sodium bisulphite and surplus sulphur dioxide at ~175 °C for 6-12 hours (Figure 2.7) (Holtzapple, 2003). Along with the delignified pulp, this process generates sugar-containing spent liquor that is typically fermented into bioethanol in North America and Europe (Roehr, 2001).

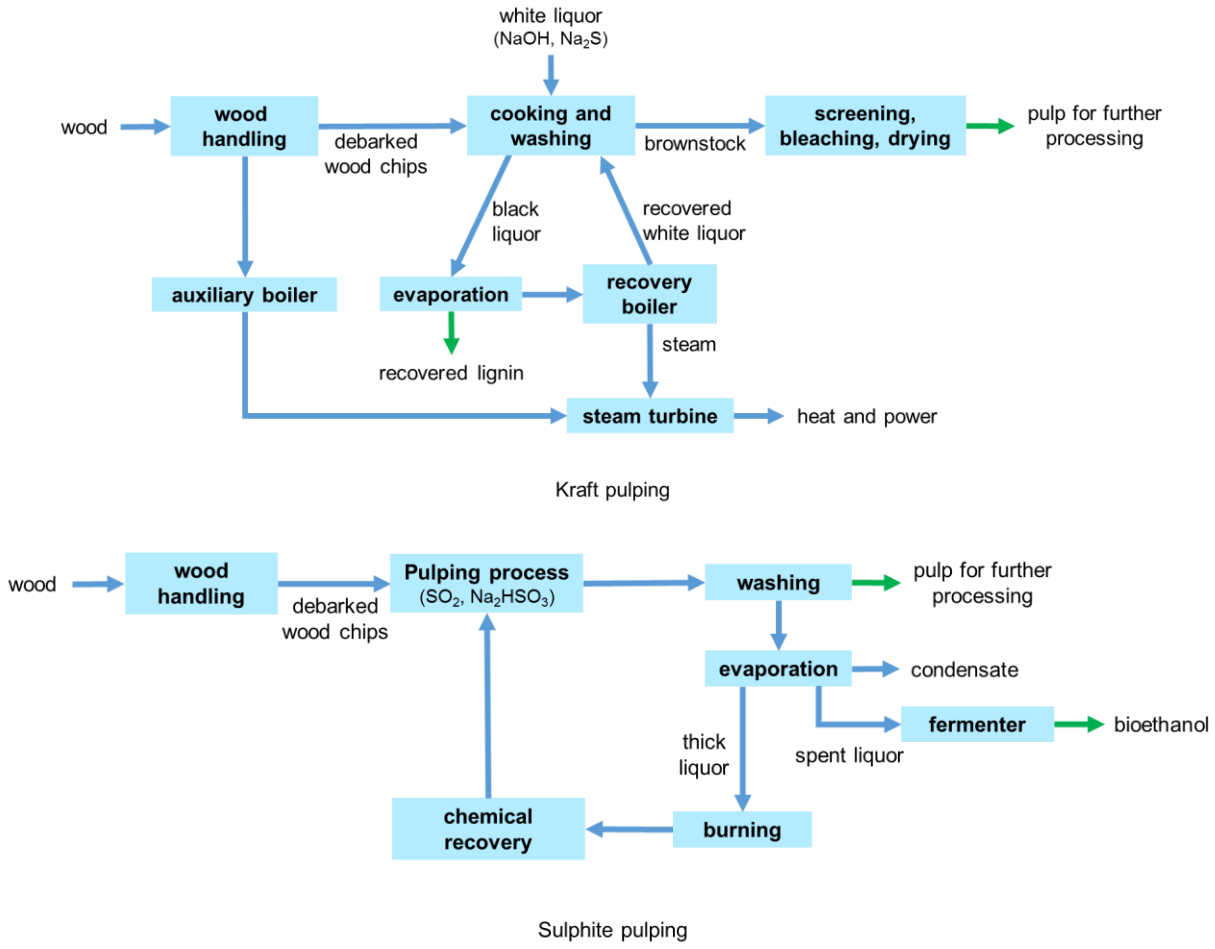


Figure 2.6 Kraft and sulphite pulping

The main lignocellulose-derived inhibitors in pretreated wood for downstream biological processing include phenolic compounds from lignin, extractives (such as tannic acid and resins), organic acids (such as acetic acid and formic acid), and furan derivatives (such as hydroxymethyl fulfural and furfural) (Luo et al., 2002; Kim, 2018). Amongst them, phenolic compounds derived from the lignin breakdown are considered potent inhibitors to the fermentation of pretreated wood, as they disrupt the microbial replication, sugar metabolism, cell membrane integrity (Palmqvist and Hahn-Hagerdal, 2000; Yu and Christopher, 2017), and suppress actions of enzymes like endo-

glucanase and β-glucosidase (Ximenes et al., 2010; Tejirian and Xu, 2011; Ximenes et al., 2011). As a lignin derivative, lignosulphonate (Figure 2.7) represents the most abundant organic fraction in the pulping liquor (~50% of liquor dry matter) (Marques et al., 2009). Next, the phenolic pyrogallol groups (on average ten) of tannic acid (Figure 2.8) form complexes with proteins, metal ions, amino acids and polysaccharides (Makkar, 2003). These interactions have been shown to inhibit enzyme action, prevent uptake of essential carbon sources and metal ions, and thereby inhibit microbial activity, including methanogens and other anaerobes in ruminant digestive systems (Asiegbu et al., 1995; Singh et al., 2001; Kamra et al., 2006; Bhatta et al., 2009) as well as anaerobic waste water treatment (Field and Lettinga, 1987; Field et al., 1988; Sierraalvarez et al., 1994; Vidal and Diez, 2003; Goel et al., 2005). In the context of wood biorefinery, pyrogallol and gallic acid from hydrolysable tannin (such as tannic acid) are major phenolic extractives in sulphite liquor from hardwood (Marques et al., 2009; Jonsson and Martin, 2016), and the presence of gallic acid is also consistently detected in pretreated lignocellulose (Du et al., 2010; Kim et al., 2013; Mitchell et al., 2014). Moving on, the metabolism of yeast suffers in the presence of acetic acid (as low as 6 g/L) as the cell pump out excess hydrogen ions that flow into the cytosol by expending ~ 1 mole ATP per mole proton (Pampulha and Loureiro-Dias, 2000; Nogue et al., 2013), despite an improvement in fermentative performance when exposed to concentrations lower than 100 mM (Larsson et al., 1999; Greetham, 2015). Such inhibition is generally more pronounced in hardwood (such as poplar, beech, eucalyptus) than softwood (such as pine, spruce, larch) due to the greater acetylated glucuronoxylan content. Furan compounds, on the other hand, prolong the lag phase of batch fermentation by inhibiting the glycolysis pathways and synthesis of RNA and protein (Banerjee et al., 1981). When compared, lignin degradation products impose

a greater toxicity to the microbes than furan compounds as low weight phenolics are highly toxic even at low doses (Hahn-Hagerdal et al., 2007; Bajwa et al., 2009).

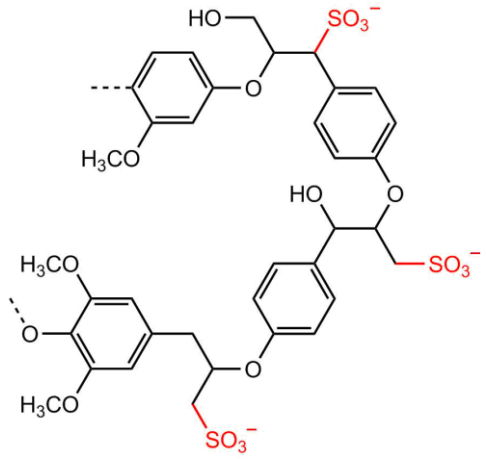


Figure 2.7 Schematic representation of lignosulphonate

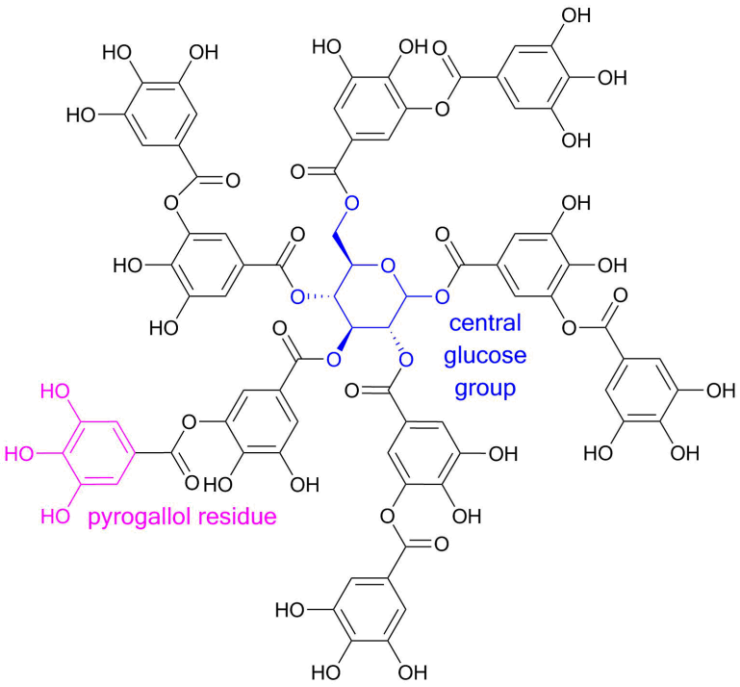


Figure 2.8 Example of molecule structure of tannic acid

As the removal of inhibitors prior to the downstream biological processing is operationally costly (Yu and Christopher, 2017), the economic sustainability of the emerging wood biorefining will depend on the discovery of microbial biocatalysts that could degrade wood polysaccharides while tolerate the inhibitory compounds present in the pretreated lignocellulosic feedstocks. Focusing on the recalcitrant organic inhibitors that are abundant in the pretreated wood biomass, the scope of the current project will be focus on lignosulphonate and tannic acid.

2.1.4 The demand for novel biocatalysts for sustainable wood lignocellulose biorefining

In view of a paradigm shift in the industry moving from petroleum resources towards renewable bioresources, there is a growing incentive to also adopt sustainable bioconversion technologies such as enzyme application to valorise lignocellulosic biomass (Boutros, 2012). In truth, the diversification of bioproducts from wood fibre via enzymes is considered as a viable opportunity to strengthen the bioeconomy of the forestry industry in North America and Europe with the growing competition from the southern hemisphere. Under this rationale, the development of industrial enzymes is a key strategic research area under the Bio-based Specialty Chemicals Program initiated by the Canadian National Research Council in 2015 (Johnston, 2015).

With a number of intrinsic benefits, enzyme application is a desired technology already integrated into the forestry industry today. By nature, enzymes are renewable and selective catalysts that typically require milder operational conditions compared to chemical processing (Schmid et al., 2001). Accordingly, they offer advantages in converting lignocellulose into fermentable intermediates, synthesizing value-added biochemicals from specific lignocellulose

components, as well as treating residual biomass with high water content, which is less amenable to thermo-chemical processing. Within the pulp and paper industry, lignin-active enzymes (laccases and manganese peroxidases), xylanases, mannanases, and pectinases have been used to increase pulp brightness by removing lignin and hemicelluloses (Kirk and Jeffries, 1996; Clarke et al., 2000; Kenealy and Jeffries, 2003; Bajpai, 2004; Maijala et al., 2008; Tringe and Hugenholtz, 2008). It is estimated that 10% of pulp and paper industries in North American have used microbial xylanases to boost fibre bleaching processes (Subramaniyan and Prema, 2002; Kenealy and Jeffries, 2003). Cellulases and xylanases have also been used to improve fibre drainage and promote fibre recycling, whereas pectinases have been used to remove pitch deposits on paper manufacturing equipment (Kenealy and Jeffries, 2003; Bajpai, 2004; Singh et al., 2012). Aside from the manufacture and recycling of paper, various glycosyl hydrolase mixtures have been developed for bio-pulping (Kong et al., 2017; Rocha-Martin et al., 2017), and conversion of pretreated fibre to fermentable sugars for bioethanol production (Kenealy and Jeffries, 2003; Amezcua-Allieri et al., 2017). More recently, researchers began to incorporate enzymes for regio-specific modification of complex oligosaccharides for novel biopolymers (MacCormick et al., 2018; Razeq et al., 2018), as traditional chemical conversions yield products with low regio-selectivity, require rare metal catalysts and multiple stages (Mirescu and Prusse, 2007; Wang et al., 2007). For applications of microbial lignocellulases in other industries, please refer to the reviews written by Singh et al. (2016) and Garg et al. (2016).

Despite the demonstrated utility of lignocellulose-active enzymes for fibre processing, cheaper and more efficient enzymes are needed to meet the economic and process contraints of agricultural and forest sector industries. With a compound annual growth rate of over 7.0%, the market for carbohydrate-active enzymes is projected to be the fastest growing product segment

compared to proteases and lipases (Singh et al., 2016). Nonetheless, the high cost of enzyme continues to impose an economic impediment to further applications in biorefineries. Factors contributing to the high costs are enzyme production, typically low enzyme accessibility to the insoluble heterogeneous plant fibres, susceptibility to deactivation by pH, heat or residual lignocellulose pretreatment chemicals (Bouws et al., 2008). Alas, both reduction in enzyme cost and discovery of new enzyme activities for lignocellulose modification are warranted. To reduce the cost of enzyme application, enzymes with competitive pH and temperature profiles, tolerance to pretreatment chemicals, and comparatively high catalytic efficiency are sought (Montella et al., 2016; Rosnow et al., 2016).

2.2 Microbial degradation of lignocellulose

2.2.1 Microbial carbohydrate-active enzymes

The efficient hydrolysis of lignocellulosic biomass into simpler sugars requires multiple actions of myriad ligninases and carbohydrate-active enzymes (Kumar et al., 2008). The catalytic breakdown of cellulose, hemicellulose, pectin and lignin are described below.

In the classic example of cellulose to glucose conversion, three major groups of hydrolytic enzymes are required to catalyse the breakdown of β-(1→4)-glycosidic bonds between the β-glucopyranose monomers (Kumar et al., 2008): endoglucanases (EC 3.2.1.4) that cleave internal β-(1→4)-glycosidic bonds, cellobiohydrolases (EC 3.2.1.91) that hydrolyse cellobiose from chain ends, and β-glucosidases (EC 3.2.1.21) that release glucose from cellobiose (Kumar et al., 2008) (Figure 2.9). Later, lytic polysaccharide monooxygenases that cleave internal glycosidic bond in cellulose by oxidation were discovered (Vaaje-Kolstad et al., 2010). Both de-crystallization caused by local disruption and release of C1/C4 oxidized oligosaccharides facilitate hydrolytic depolymerisation of cellulose (Langston et al., 2011; Agger et al., 2014).

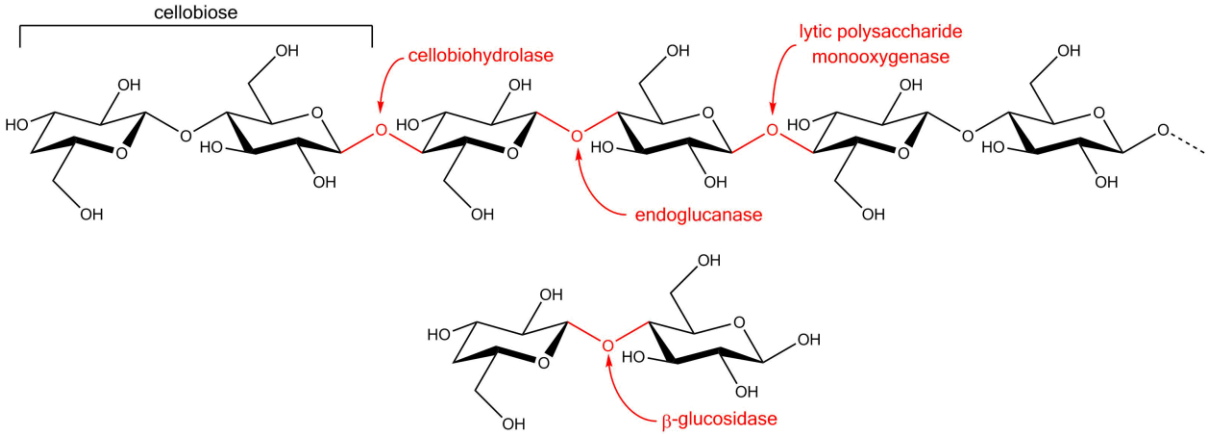


Figure 2.9 Enzymatic degradation of cellulose

Hemicellulases are a collective term used to describe enzymes that act on diverse hemicellulose substrates, and include xylanases, mannanases, as well as accessory enzymes that target branching sugars and acetyl groups (Liu et al., 2013). Xylan active enzymes include endo-β-(1→4)-xylanases (EC 3.2.1.8), β-xylosidases (EC 3.2.1.37), α-D-glucuronidases (EC 3.2.1.139), α-L-arabinofuranosidases (EC 3.2.1.55) and acetylxylan esterases (EC 3.1.1.72) (Figure 2.10).

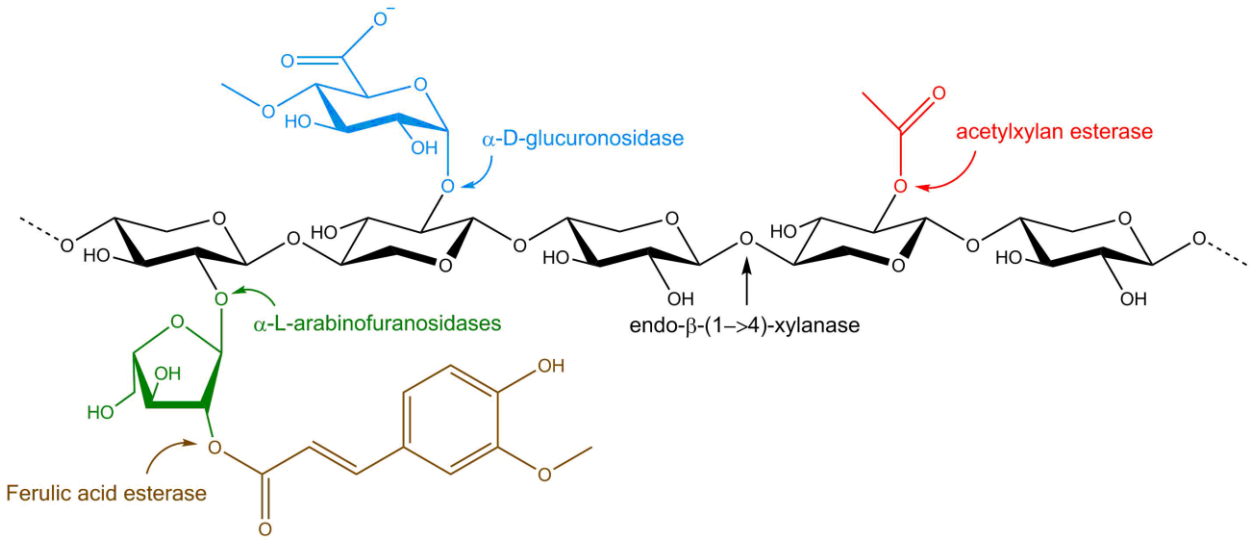


Figure 2.10 Xylan degradation

As pectin represents complex macrostructures, many enzymes are involved in its degradation. These enzymes are broadly grouped under pectin esterases, glucosidases and lyases based on their modes of action (Garg et al., 2016). Pectin esterases cleave the methyl ester group and acetyl ester group from the galacturonate units of homogalacturonan as methylesterases (EC 3.1.1.11) and acetylesterases (EC 3.1.1.6), respectively (Garg et al., 2016); they also remove the acetyl ester group and feruloyl ester group from rhamnogalacturonan as rhamnogalacturonan acetylesterase and feruloyl esterase, respectively (Ruiz et al., 2017). Meanwhile, glucosidases

catalyse the hydrolytic cleavage of α-(1→4)-glycosidic linkage in pectic acid as polygalacturonases (EC 3.2.1.15, 3.2.1.67, 3.2.1.82) and in pectin as polymethylgalacturonases (Garg et al., 2016). Next, lyases perform trans-elimination to cleave the α-(1→4)-glycosidic linkage in pectic acid as polygalacturonate lyase (EC 4.2.2.2, 4.2.2.6, 4.2.2.9), in pectin as polymethylgalacturonate lyases (EC 4.2.2.10), and in rhamnogalacturonan as rhamnogalacturonan lyase. Other pectinases include galacturonic oxidases that oxidize the galacturonic acid (Ruiz et al., 2017).

Lignin is transformed by lignin peroxidases (EC 1.11.1.14), manganese peroxidases (EC 1.11.1.13), versatile peroxidases (EC 1.11.1.16), dye-decolourizing peroxidases, and laccases (EC 1.10.3.2), based on free radical mechanisms. Bacterial β-etherases that selectively catalyse the reductive cleavage of β-O-4-aryl-ether bonds in model lignin compounds has also been reported (Figure 2.11) (Bugg and Rahmanpour, 2015; Picart et al., 2015). Whereas lignin peroxidases from fungi typically oxidize and cleave non-phenolic lignin units (Tien and Kirk, 1983; Hammel et al., 1993; Caramelo et al., 1999), manganese peroxidases oxidize Mn^{2+} to Mn^{3+} to degrade minor phenolic units in lignin (Gold and Alic, 1993), while versatile peroxidases), as implied by its name, are able to perform both catalytic activities. Found in both fungi and bacteria, laccases are multi-copper oxidases that can oxidize phenolic lignin compounds by reducing oxygen to water (Pollegioni et al., 2015), and dye-decolourizing peroxidase that can oxidize Mn^{2+} to Mn^{3+} as well as lignin sites as heme-containing peroxidases (Ahmad et al., 2011; Rahmanpour and Bugg, 2015).

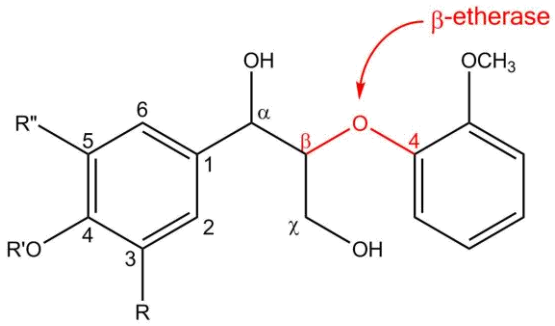


Figure 2.11 Cleavage of β-O-4-aryl-ether bond in lignin by bacterial β-etherase

Also known collectively as carbohydrate-active enzymes (CAZymes), enzymes that catalyse the breakdown, biosynthesis or modification of carbohydrates and glycolconjugates have been classified based on sequence and structure, by the carbohydrate-active enzyme database (CAZy database, https//www.cazy.org) (Lombard et al., 2014). The main classifications are: (1) Glycoside hydrolases (GHs) that hydrolyze glycosidic bonds through retaining and inverting mechanisms, (2) glycosyltransferases (GTs) that catalyse the formation of glycosidic bonds, (3) polysaccharide lyases (PLs) that catalyse the non-hydrolytic cleavage of glycosidic bonds, and (4) carbohydrate esterases (CEs) that hydrolyse carbohydrate esters. Most recently, the CAZy database was expanded to include auxiliary activities (AAs) that comprise redox enzymes, including carbohydrate oxidoreductases and lytic polysaccharide monooxygenases. Lignin peroxidases, manganese peroxidases, versatile peroxidases, and laccases are also classified within families AA1 and AA2 of the CAZy database. Lastly, CAZymes can be appended with non-catalytic carbohydrate-binding modules (CBMs) for substrate adhesion, which constitute their own CAZyme family. Based on the amino acid sequences, there are 152 families of GHs, 105 families of GTs, 28 families of PLs, 16 families of CEs, 15 families of AAs, 83 families of CBMs as of March 2018, and the numbers are expected to grow with the ever-expanding genomic and metagenomic sequence data. While the key active-site residues, catalytic mechanism and the

tertiary protein structures are generally conserved within a family, members exhibit distinct activity profiles on different substrates and sub-classification based on the catalytic mechanism and protein structure is required for some larger families (Stam et al., 2006; Aspeborg et al., 2012). The rich sets of manually curated information available in this comprehensive repository assist in the identification of CAZymes in newly sequenced genomes, metagenomes and metatranscriptomes.

2.2.2 Cellulosomes from anaerobic microbes

Anaerobic microbes produce self-assembled multi-enzyme complexes known as cellulosomes that are highly efficient in lignocellulose degradation (Bayer, 2017). First isolated from *Clostridium thermocellum* (Lamed et al., 1983), cellusomes are also produced by other *Clostridial* and non-*Clostridial* bacteria such as *Acetivibrio cellulolyticus* (Ding et al., 1999), *Bacteroides cellulosolvens* (Ding et al., 2000), *Ruminococcus albus* (Lamed et al., 1987), among others. Molecular arrangements with varying degrees of complexity are presented by these species. The architecture of cellulosomes consists of a scaffoldin subunit with cohesins that are linked to dockerin-containing CAZymes via cohesin-dockerin interaction and CBMs that bind to targeted substrates (Artzi et al., 2017) (Figure 2.12). When a cellulosome is not expressed as a cell-free entity, the dockerin on the scaffoldin mediate its attachment directly via the cohesin of the anchoring scaffoldins on the bacterial cell wall, or indirectly through an adaptor scaffoldin which comprises the corresponding cohesin as well as dockerins that connects to the cohesin on the anchoring scaffoldins (Bayer et al., 2004; Bayer, 2017). Overall, the spatial enzyme proximity and enhanced enzyme-substrate targeting are thought to potentiate synergistic degradation of

lignocelluloses. In recent years, the complexity and diversity of bacterial cellulosomes have been revealed through genome and metagenomic sequencing (Bayer et al., 2008; Bensoussan et al., 2017), and designer cellulosomes with targeted functions have become a key emerging field in the degradation of lignocellulosic biomass (Morais et al., 2012; Davidi et al., 2016; Bayer, 2017).

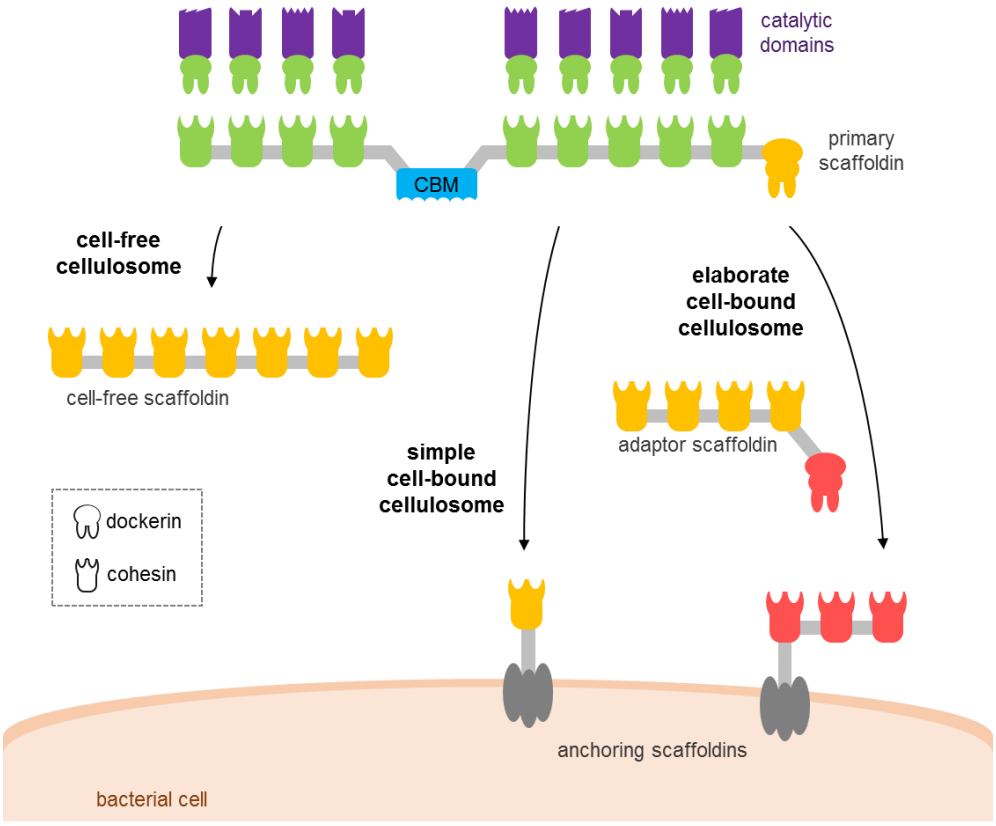


Figure 2.12 **Cellulosome** (adapted from (Artzi et al., 2017))

2.2.3 Polysaccharide utilization loci

Gram-negative *Bacteroidetes* possess sophisticated polysaccharide utilization loci (PULs) that encode various components for hydrolysis of polysaccharides and uptake of resulting products (Martens et al., 2009; Grondin et al., 2017). PULs characterized to date similarly encode

cell membrane-associated proteins that are homologous to SusC and SusD, which are involved in the uptake and utilization of hydrolysis products from starch (Figure 2.13). More specifically, SusD-like proteins are thought to bind and mediate the transfer of specific oligosaccharides to SusC-like proteins, which are oligosaccharide transport proteins. PULs typically also encode sensor-regular systems (e.g. hybrid two-component systems), which regulate the expression of associated hydrolytic enzymes. PULs representing the starch utilization system (sus) is the most thoroughly studied example to date (Martens et al., 2009). Briefly for this case, released sugars first make their way through the capsular polysaccharide layer to bind SusD, which facilitates further hydrolysis by the outer membrane-associated enzyme, SusG, and uptake of smaller oligosaccharides by SusC. These oligosaccharides are then further digested into mono- and disaccharides by periplasmic enzymes, and the released saccharides can function as transcription signals for PUL gene expression. A study by Larsbrink et al. (2014) uncovered an analogous xyloglucan PUL, and identified each step of the digestion process that was required for a complete hydrolysis of the polysaccharide.

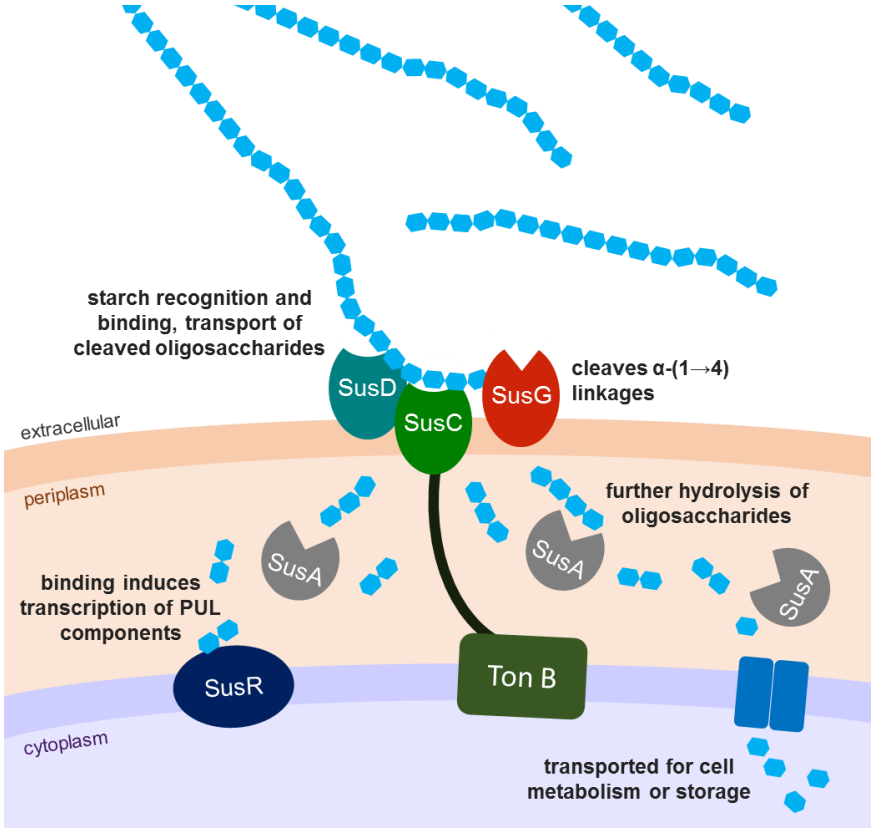


Figure 2.13 Polysaccharide utilization system

PULs are ubiquitous in genomes of *Bacteroidetes* and are believed to have expanded through duplication of genetic elements as well as horizontal gene transfer (Martens et al., 2009). In consequence, extant *Bacteroidetes* species encode an extensive suite of PULs that are thought to encode functionally related carbohydrate-active enzymes. By identifying PULs in bacterial and metagenome sequences, it might be possible to uncover new synergisms between known CAZymes, or even entirely new lignocellulose-active proteins. Given the knowledge likely gained through PUL characterization, Terrapon et al. (2014) developed an algorithm to predict PULs from bacterial genome and metagenome sequences. Briefly, the PUL predictions begin by identifying tandem *susCD*-like pairs, followed by the identification of nearby regulatory genes, the presence of CAZymes and of regulatory genes (e.g. hybrid two-component system protein,

extracytoplasmic function (ECF) σ/anti-σ factors, etc.) that mark the boundary of the PUL. Notably, gene models encoding proteins without known CBM or CAZyme modules are often identified within the PULs. These genes might encode entirely new enzyme activity relevant to lignocellulose conversion or else fine-tune the regulation of predicted CAZymes. Today, PUL predictions in *Bacteroidetes* species from human gastrointestinal tract, 64 rumen gut species, and soil or marine environments are now accessible online as a continually updated feature of the CAZy database (www.cazy.org/PULDB/index.php) (Terrapon et al., 2018).

2.3 Bioprospecting lignocellulases from lignocellulose-degrading microbiomes

2.3.1 Next generation sequencing of metagenomes and 16S rRNA gene markers

Next generation sequencing is commonly used in metagenomic investigations on complex microbial communities to discover useful lignocellulases for efficient breakdown of plant biomass (Batista-Garcia et al., 2016; Tiwari et al., 2018). These studies have been spurred by the decreasing DNA sequencing costs, along with the rich repertoire of CAZymes encoded by gut microflora, and potential high-throughput production of novel CAZymes. In the next paragraphs, overview of the metagenomic approaches, as well as the principles of Illumina sequencing technology and Roche 454 pyrosequencing are described.

Metagenomics refer to the study of genetic material recovered directly from environmental samples, commonly performed by sequencing the DNA fragments from the entire genomes (also known as shotgun metagenomics) or DNA amplicons of phylogenetic gene markers such as 16S ribosomal RNA (16S rRNA) gene for prokaryotes (Oulas et al., 2015; Quince et al., 2017). The former allows a comprehensive investigation of the rich diversity of genes present in a microbiome, whereas the latter enables a cost-effective phylogenetic classification of the microbial populations. Specifically, the ~1,500 bp long prokaryotic 16S rRNA gene contains nine hypervariable regions that could be targeted via polymerase chain reaction (PCR) for phylogenetic classification (Tringe and Hugenholtz, 2008). Furthermore, cost reductions via multiplexing multiple samples in a single sequencing run is possible by incorporating sample-specific nucleotide barcodes into the PCR primers (Quince et al., 2017). The output sequences from the shotgun approach will be quality screened and assembled into longer contigs or genomes for gene prediction and annotation using relevant database (Roumpeka et al., 2017), such as CAZy database for annotating CAZymes (Lombard et al., 2014). Meanwhile, quality-screened sequence

output from the 16S rRNA approach is clustered into phylotypes according to their similarity to the deposits from a 16S rRNA gene repositories such as Greengenes Database (DeSantis et al., 2006; Liu et al., 2008). Alternatively, *de novo* similarity-based clustering of the sequences into operational taxonomic units (OTUs) can be performed (Chen et al., 2013), followed by taxonomic assignment with the use of reference database. In either case, the discovery of taxonomically unknown phylotypes or OTUs is not uncommon due to the lack of similar taxonomically-resolved gene markers in the reference database.

Illumina sequencing technology and Roche 454 pyrosequencing are widely adopted next-generation sequencing platforms based on 'sequencing by synthesis' (Mardis, 2008). Illumina sequencing begins with ligating the DNA fragments with adapters that anneal onto a flow cell surface via complimentary base pairing, followed by bridge PCR that creates local clusters with approximately one million identical copies of each DNA template. In the subsequent sequencing cycles, reversible terminator bases known as deoxynucleoside triphosphates (dNTPs) are added to allow the addition of one base at a time with DNA polymerase. Since the dNTPs (A, T, C, G) are labelled with base-unique fluorescent labels that are cleaved after the incorporation of each base, the images captured are then used to inform the DNA sequences via a base-calling algorithm. After each imaging step, the 3' blocking groups of the incorporated bases are removed chemically for the next cycle. For 454 pyrosequencing, the adaptor-ligated DNA fragments are immobilized onto nanobeads in a 1:1 ratio and amplified via emulsion PCR (Rothberg and Leamon, 2008). Then, the beads are positioned within the DNA polymerase-containing wells on a sequencing plane at one bead per well. Over the next sequencing cycles, each of the four nucleotide base is added to the wells and cleared sequentially. The release of pyrophosphate is conveyed into luminescent signal in amounts proportional to the number of bases added in a series of enzymatic

reactions with luciferase. The intensity and locations of the signals captured by the images are then translated into DNA sequences by 454 base-calling software. While base insertion and deletion errors for longer stretches (> 6) of the same base are prone to occur due to improper base-calling, substitution errors are rare as the nucleotides are added one species at a time.

The implementation of these metagenomic approaches and advanced sequencing platforms on biomass degrading microbial communities for lignocellulolytic enzymes is reviewed in the next section.

2.3.2 Metagenomic approaches to discover lignocellulolytic enzymes from enrichment cultures

With the advancement in biotechnology tools, the efforts in characterizing microbial communities that degrades plant biomass have included the analysis of metagenomes obtained from digestive systems of animals that graze on straw, grasses and lichens, including the cow (Hess et al., 2011), reindeer (Pope et al., 2012), Tammar wallaby (Pope et al., 2010) and yak (Dai et al., 2012), to name a few. Overall, analyses of lignocellulose-degrading gut microflora using pyrosequencing of 16S rRNA gene markers revealed the prevalence of *Firmicutes* (*Bacilli*, *Clostridia*), *Bacteroidetes*, *Proteobacteria* (*Alpha-*, *Beta-*, *Gamma-proteobacteria*), *Chloroflexi*, and *Actinobacteria*. Herein, the presence of biomass degraders in the community can be interpreted as an indicator of microbial degradation, especially when these members proliferate after a biomass enrichment.

Metagenomic approaches to identify CAZymes relevant to the conversion of a given biomass feedstock have included i) collecting environmental samples persistently subjected to the

targeted feedstock, and ii) enrichment of environmental samples using the biomass feedstock as the sole carbon source. Compared to the investigation on the degradation of crops and grasses, the efforts to identify CAZymes pertinent to bioconversion of woody substrates have flourished in more recent years. They included the metagenomic analyses of forest soils (Damon et al., 2012; Pold et al., 2016), wood-feeding beetle (Scully et al., 2013) and wood-feeding molluscs (O'Connor et al., 2014). An exception is the termite gut microbiome, which has been regarded as a paradigmatic model for wood bioconversion (Warnecke et al., 2007; He et al., 2013; Brune, 2014; Rossmassler et al., 2015). However, others have noted that pre-digestion of wood fibre by the termite might generate rather simple substrates for further hydrolysis by the gut microbiota (Scharf and Tartar, 2008). Moreover, these enrichment microbiomes are not challenged by the lignocellulosic inhibitors liberated from pretreatments, hence the augmented microbes and expressed CAZymes may not be representative of those that can handle the pretreated wood lignocellulose.

Subsequent to the enrichment of a microbial community on selected lignocellulose substrates, comparative metagenomic analyses could highlight the impact of amended carbon source on the population and encoded CAZymes (Brulc et al., 2009; Colman et al., 2012; Pope et al., 2012; Scully et al., 2013; Jimenez et al., 2014b; Wang et al., 2016a). In addition to identifying microbes and CAZymes that most likely contribute to the conversion of selected feedstocks, microbial enrichment prior to sequencing typically improves sequence assembly (DeAngelis et al., 2010). These findings justify and motivate efforts to enrich key community members prior to metagenome sequencing. For example, metagenomic analysis of fibre-adherent microorganisms sampled from bovine rumens revealed that first colonizers likely target readily available side groups of complex polysaccharides (Brulc et al., 2009); enrichment of compost samples on

switchgrass identified several GHs implicated in cellulose degradation (Allgaier et al., 2010), whereas enrichment of a rice straw adapted community on manure compost uncovered the role of *Actinobacteria* in lignocellulose conversion (Wang et al., 2016a); and metatranscriptomic analysis of forest soil microbiomes revealed enrichment of CAZymes from families GH5, GH12 and GH16 following enrichment on cellulose and wheat bran (Takasaki et al., 2013). Furthermore, controlled enrichment of pre-existing microbial communities present on poplar chips identified CAZymes potentially relevant to wood degradation, including members from families GH2, GH3, and GH13, as well as activities likely contributing to lignin transformation (van der Lelie et al., 2012). Similarly, the microflora from sugarcane bagasse pile expressed a unique composition of lignocellulases with prevalent GH2, 3, 9, 10 and 43 after successive anoxic enrichment with rice straw (Wongwilaiwalin et al., 2013), and soil microbial community that was grown on cotton expressed high numbers of GH2, GH3 and GH43 (Zhang et al., 2016). Along with new CAZymes, the rapidly expanding set of bacterial genome and metagenome sequences has unveiled a plethora of PULs in herbivores gut environments such as reindeer rumen (Pope et al., 2012), tammar wallaby foregut (Pope et al., 2010), buffalo rumen (Brulc et al., 2009), and cow rumen (Hess et al., 2011). Overall, informative differential CAZyme profiles can be predicted through metagenomic analysis of enrichment cultures, yet whether the annotated CAZyme-encoding genes are translated or the CAZymes are indeed functional in biomass remains unknown.

2.3.3 Metasecretomic approach to discover functionally-active lignocellulolytic enzymes from enrichment cultures

To circumvent the limitation of metagenomics, functional metasecretomic analysis is an attractive yet underexplored strategy to directly characterize the lignocellulolytic capacity of the secreted proteomes in mixed cultures for novel carbohydrate-active enzymes and cocktails. Moreover, this direct assessment can also profile activities encoded by the less characterized CAZymes with amino acid sequences and catalytic domains, that may not be annotated by metagenomic analysis due to their low representation in the current reference database (Li et al., 2009). A number of insightful discoveries have been made with the application of secretomic analyses for the development of enzyme cocktails for heterogeneous lignocelluloses. This includes the advantage of analysing mixed culture, reflected in the significant improvement in enzymatic activities by the secretome of thermophilic bacteria when they were cultured together with corn stalk rather than those from pure cultures (Zhang et al., 2014b), as well as the unique enrichment of GH10, GH51 and GH95 CAZyme families in microcosms fed with wheat straw not observed in those fed with xylan or xylose (Jimenez et al., 2015).

Looking forward, optimization of the protocol is required for microbial metasecretomics due to a number of technical obstacles (Shah and Gharbia, 2010). Main challenges include the difficulty in recovering metabolically-active microbial enzymes that are typically secreted in very low concentrations, potential contamination with non-target intracellular proteins liberated from lysed cells, interfering residual salts and nutrient from the culture media for downstream proteomic analysis. As for the functional screening, saccharification assay with *para*-hydroxybenzoic acid hydrazide (PAHBAH) provide a sensitive quantification of the reducing

sugars released from the catalysis of plant polysaccharides among the developed assays (Moretti and Thorson, 2008). The reaction involves the formation of hydrazone at the C1 position, followed by the oxidation of C2, and a final formation of osazone (Figure 2.14) (Barry and Mitchell, 1955). Proportional to the extent of catalysis, the amount of yellow osazones generated is readily measured by a spectrophotometer at 410 nm and quantifiable with a standard curve. Compared to other model substrates such as 4-*para*-nitrophenyl (pNP) substrates, which inform release of specific sugars, complex polysaccharides generate a more realistic representation of enzyme activity on high molecular weigh substrates and in the presence of native side groups. Accordingly, the inclusion of a wide range of plant polysaccharides with varying chemical structure is essential to provide a comprehensive functional profile.

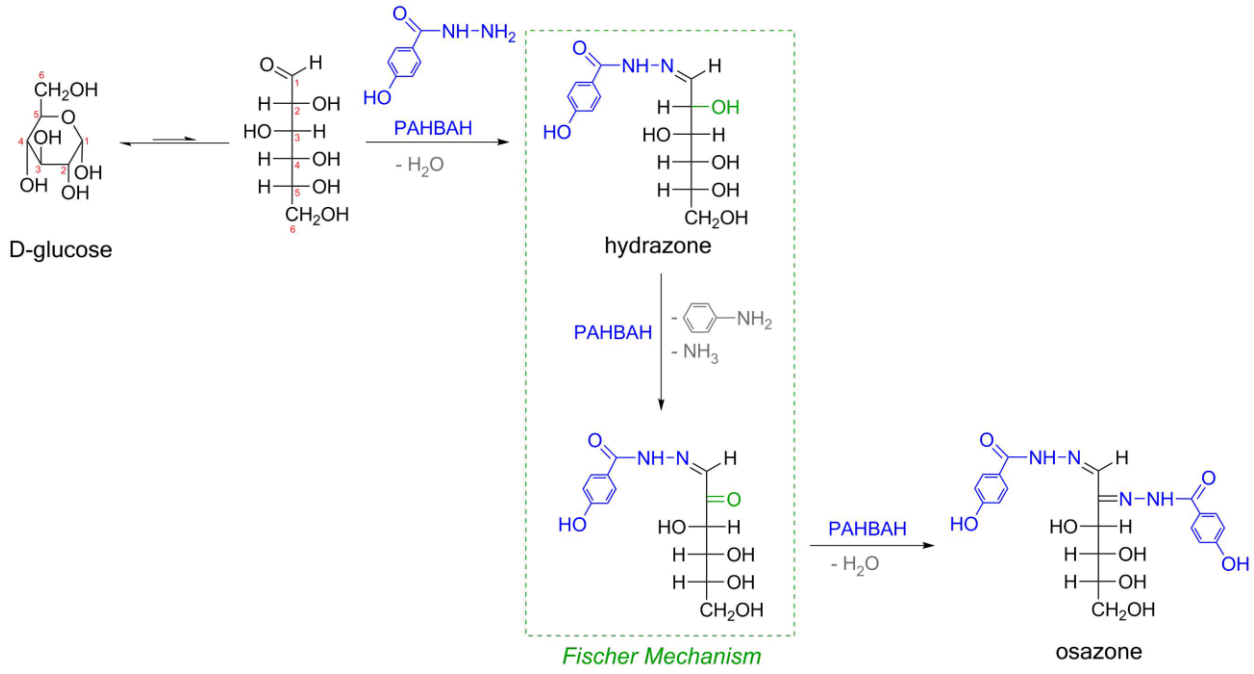


Figure 2.14 Principle of PAHBAH assay illustrated with glucose

2.3.4 A strategy to bioprospect lignocellulolytic enzymes—dual metagenomic and metasecretomic investigation on anaerobic microcosms from pulp mill anaerobic digesters and gut microbiomes from Canadian beaver and North American moose after long-term enrichment on wood lignocellulose

Summarized from previous sections (2.1.3, 2.2.2, 2.2.3, 2.3.2, 2.3.3), future approach to bioprospecting lignocellulolytic enzymes relevant to wood lignocellulosic feedstock in biorefinery needs to be reinvented to address some major limitations: i) concentrated focus on microbiomes that transform crops and grasses, ii) enrichment of microbiomes on either crops and grasses, or woody substrates in absence of inhibitors generated by pretreatment process, iii) gene prediction based on CAZy database only, and iv) reliance on metagenomic inference without metasecretomic analysis for functional characterization.

Shifting the focus away from microbiomes that process crops and grasses to those that degrade wood fibre, pulp mill anaerobic digesters, gut microbiomes of wood-feeding Canadian beaver and North American moose represent largely-unheeded potential sources of lignocellulolytic lineages and enzymes relevant to the wood biorefinery. These microbiomes would then be enriched with industrially relevant woody substrates and followed by a comprehensive metagenomic analysis, which encompass taxonomic assignment and annotation of CAZymes, and other lignocellulolytic candidates such as cellulosomal subunits, PULs and carbohydrate-active proteins with domains of unknown function. The microbial community analysis serves to indicate a population that is selected by the fed substrate, while the functional annotation provides the blue print on which to anchor the metasecretomic activity assays by means of identifying the proteins in the secretome and their host genomes.

The microbiomes were selected based on their capacities in wood degradation. Pulp mill anaerobic digesters are considered promising sites for enzyme discovery since the microbial community is highly adapted towards wood extractives and toxic inhibitors (Olsson and Hahn-Hagerdal, 1996). In fact, other anthropogenic sites such as anaerobic digesters in wastewater treatment plants and biofuel plants are already proven to harness novel CAZymes (Healy et al., 1995; Jiang et al., 2010). As for the Canadian herbivores, both beaver and moose rely on a wood-based diet especially during the winter months. The main diet of beaver consists of twigs, bark and tree trunks, while the ruminant forge on twigs, shrubs, and bark (Chaney, 2003; Hood and Bayley, 2009). Past literature has reported a wood-degrading microbial community in moose containing lineages resembling that in termites (Ishaq and Wright, 2012), an epitomized model insect for wood-degradation (Warnecke et al., 2007). Enzymes belonging to families GH2 and GH3 predicted to participate in plant cell wall deconstruction were also found to be abundant in the rumen microbiome of moose (Svartstrom et al., 2017). As for the beavers, preliminary data suggested cellulolytic and xylanolytic activities in their lower gut (Gogola et al., 2011). While these findings collectively provide stimulating insights into the microbial wood-degrading activities, further work on harvesting these enzymes remains largely unattempted.

Taken together, the discovery of novel enzymes for processing wood fibre requires a judicious selection of environmental sources and enrichment with wood lignocellulose. Systematic metagenomic (of 16S rRNA gene markers and whole genome shotgun sequences) and metasecretomic investigation of the lignocellulose amendments from the gut microbiomes of wood-feeding beaver and moose, and pulp mill anaerobic digesters are worthy pursuits to unveil microbes and enzymes with potentially new lignocellulolytic activities. The study of these wood-degrading microbial communities is grouped by the nature of the environmental sources (i.e. gut

microbiomes versus reactor sample) and reported in the remaining chapters of this thesis as outline in 1.2.

Chapter 3 Substrate-driven convergence of the microbial community in lignocellulose-amended enrichments of gut microflora from the Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*)

This chapter is adapted from the following peer-reviewed publication:

Wong, M.T., Wang, W.J., Lacourt, M., Couturier, M., Edwards, E.A., and Master, E.R. (2016). Substrate-driven convergence of the microbial community in lignocellulose-amended enrichments of gut microflora from the Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*). *Frontiers in Microbiology* 7.

Note that the enrichment substrate pretreated poplar was referred to as poplar hydrolysate in the publication according to the supplier SunOpta Inc., Canada.

3.1 Introduction

Lignocellulose in agricultural and forest residues, as well as energy crops, is considered an important renewable resource for the production of bioenergy, liquid biofuels, and specialty chemicals. As the main component of plant cell walls, lignocellulose is largely comprised of polysaccharides (cellulose, hemicellulose and pectin) and lignin, with varying chemical composition and structure depending on plant species, tissue, and cell type (Harris and Stone, 2009). Wood fibre typically contains a higher lignin content and hemicelluloses with chemical structures distinct from those found in grasses. Fungi and bacteria are the dominant organisms

responsible for lignocellulose biodegradation and encoded enzymes offer advantages in lignocellulose processing particularly when i) converting lignocellulose to fermentable intermediates, ii) synthesizing high-value chemicals from specific lignocellulose components, and iii) handling residual biomass with high water content, which is less amenable to processing through thermo-chemical options.

Metagenomic analysis of microbial communities that degrade lignocellulose has been motivated by decreasing DNA sequencing costs, along with the rich repertoire of CAZymes encoded by gut microflora. Such efforts have included the analysis of metagenomes obtained from foregut of Tammar wallaby (Pope et al., 2010), mid-gut of wood-feeding Asian longhorn beetles (Scully et al., 2013), hindgut of termite (Warnecke et al., 2007), as well as the rumen of ox (Brulc et al., 2009), cow (Hess et al., 2011), yak (Dai et al., 2012) and reindeer (Pope et al., 2012). Corresponding analyses have identified thousands of new genes predicted to encode enzymes relevant to lignocellulose conversion. For instance, metagenomic analysis of the cow rumen alone led to over 27,000 new candidate CAZymes (Hess et al., 2011). In an effort to identify genes likely to encode enzymes optimized for transforming wood fibre, Scully et al. (2013) applied hierarchical cluster analysis of Pfam abundances to compare the gut metagenome of a wood-boring pest, *Anoplophora galbripennis*, to 19 herbivore-related metagenomes (Scully et al., 2013). Distinct clusters representing different herbivore biome-types were identified, including herbivore gut communities, fungal gallery communities, and communities associated with insects that feed on heartwood.

In contrast to grass-fed mammalian herbivores, North American moose (*Alces americanus*) and Canadian beavers (*Castor canadensis*) are iconic Canadian foragers of coniferous and deciduous trees in riparian zones of the boreal mixed-wood forests (Hood and

Bayley, 2009). North American moose is the largest browsing ruminant of the deer family *Cervidae* (Ishaq and Wright, 2012), while Canadian beavers represent one of the largest and ecologically most distinct rodent species with a monogastric digestive system, whose dietary subscription shifted from omnivory to obligate herbivory (Horn et al., 2011). With wood biomass being a significant part of the diet, the microbial communities within the digestive system of these Canadian mammals are likely to include lignocellulose degrading bacteria. Recent studies report lignocellulose-degrading bacterial lineages among the gut microbes from moose (Ishaq and Wright, 2014), which resemble those residing in the termite hindgut (Ishaq and Wright, 2012). Preliminary data also suggest cellulolytic and xylanolytic activities in the lower gut of beavers (Gogola et al., 2011).

Enrichment of microbial communities on selected lignocellulose substrates could augment the fraction of most pertinent lignocellulose degraders and encoded activities. For example, feeding termites with grasses enriched *Clostridiales incertae sedis* and *Spirochaetaceae* lineages of *Firmicutes* in their hindgut populations, whereas feeding with wood fibre proliferated members across several phyla, including *Bacteroidetes*, *Elusimicrobia*, *Firmicutes*, *Plantomycetes*, *Proteobacteria*, *Spirochaetes*, and *Verrumicrobia* (Huang et al., 2013). Likewise, fecal microbiomes obtained from cattle fed with unprocessed grain were enriched with bacteria belonging to the *Ruminococcaceae* order, whereas those obtained from cattle fed with forage or processed grain were enriched in bacteria belonging to the *Prevotella* genus (order *Bacteroidales*) (Shanks et al., 2011). Notably, specific phyla were enriched in nearly all lignocellulose-degrading gut microflora analysed to date, including *Firmicutes* (*Bacilli*, *Clostridia*), *Proteobacteria*, *Bacteroidetes*, *Chloroflexi*, and *Actinobacteria.* However, most enrichment studies have been performed *in situ*, and so are confounded by the presence of additional glycan sources, including

mucin glycans produced by the host (Koropatkin et al., 2012; Tailford et al., 2015). Alternatively, *ex situ* enrichment of microbial communities on lignocellulosic carbon sources could uncover microbial lineages that are quintessential to degrading specific biomass components.

Here we directly compared shifts in microbial profiles that result from long-term enrichment (> 3 years) of digestive microflora from the Canadian beaver and North American moose, on four lignocellulosic carbon sources: cellulose, cellulose + tannic acid, cellulose + lignosulphonate, and pretreated poplar. These amendments represented increasingly heterogeneous and potentially inhibitory carbon sources. For example, inhibition of methanogenic activity by tannic acids is well-known (Bhatta et al., 2009), while pretreated poplar typically contains mixed wood extractives, organic acids, furan derivatives and lignin derivatives that can inhibit microbial activity, including methanogenesis (den Camp et al., 1988; Sierra-Alvarez and Lettinga, 1991; Mills et al., 2009). Aside from monitoring metabolic activities through biogas yield from each enrichment, amplicon sequencing of 16S rRNA genes was performed to characterize shifts in microbial communities that would suggest specialization and expression of distinct lignocellulolytic activities.

3.2 Materials and Methods

Collection of beaver droppings and moose rumen samples. Beaver droppings were collected in November 2009 with assistance from the Ontario Ministry of Natural Resources. Collection sites included mixed boreal forests near Parry Sound, Mount Albert, and Mississauga. Moose rumen fluid samples were collected by hunters on October 10th, 2009 at a location just west of Nabakwasi Lake (GPS: 47.5612 N, 81.4504 W), about 100 km north of Sudbury, Ontario. All samples were stored on ice during transport, and then transferred to a -20 °C freezer upon arrival at the University of Toronto.

Carbon sources and their chemical oxygen demand. Amended carbon sources included i) microcrystalline cellulose (Avicel PH101, Sigma-Aldrich, MO, USA), ii) cellulose (Avicel) + lignosulphonate (Tembec Industries Inc., QC, CA), iii) cellulose (Avicel) + tannic acid (Sigma-Aldrich, MO, USA), and iv) steam-exploded (200 ºC for 8 minutes) pretreated poplar (SunOpta Inc., Canada; October 2009). Previous investigations demonstrated that pretreated poplar yielded from these condition from the same facility contain approximately 50% glucan, 30% lignin (including acid soluble lignin and lignin ash), 14 % xylan, 2 % mannan and 0.6% ash in dry mass (Di Risio et al., 2011). To determine the stoichiometric maximal biogas yield (Symons and Buswell, 1933) of these lignocellulose substrates (in ml biogas/ mg COD), the chemical oxygen demand (COD) was determined as described by the APHA standard method 5220D using potassium dichromate as an oxidizing agent (Table S3.1) (Clesceri et al., 1998).

Set up and maintenance of anaerobic enrichments. Based on the protocol established previously by Edwards and Grbić-Galić (Grbić-Galić and Vogel, 1987; Edwards and Grbić-Galić, 1994), sulphide-reduced mineral medium (pH 7.0) was prepared, autoclaved at 121°C for 20 min and purged with 80 % N_2, 20 % CO_2 gas mixture to maintain methanogenic consortia. The

collected beaver droppings and moose rumen were thawed and mixed separately to ensure sample homogeneity. To begin each anaerobic cultivation, the prepared mineral medium, environmental inoculum, and autoclaved lignocellulosic carbon sources were introduced into an anaerobic glove-bag (medium tape seal AtmosBag, Aldrich) (Edwards and Grbić-Galić, 1994). Approximately 15 mL of beaver dropping or moose rumen inoculum was then transferred to separate 160 mL Wheaton glass serum bottles, which were subsequently amended with selected lignocellulosic substrates (average 36.1 mg COD equivalent) and 45 mL of mineral medium (Table S3.2). Corresponding cultures were prepared in triplicate and sealed with butyl rubber stoppers before being incubated anaerobically at 36°C; biogas production was then monitored at regular intervals using a pressure transducer (Omega PX725 Industrial Pressure Transmitter, Omega DP24-E Process Meter).

When biogas production ceased, each microbial community was transferred to a new bottle and amended with fresh anaerobic medium and lignocellulose carbon source (approximately 170 mg COD; Table S3.2). Transfer by dilution was performed in the first four phases, wherein approximately 1/3 of the entire enrichment was transferred to a new bottle and topped up with fresh carbon source and medium to the original volume. To subsequently increase the density of lignocellulose-active communities and remove accumulated metabolic wastes, subsequent enrichments were centrifuged at ~3,400 g for 15 min at 24 °C, and then the spent cultivation medium was replaced with fresh mineral medium and carbon source. The amounts of lignocellulose substrates added to the microcosm bottles and duration across the enrichment phases are summarized in Table S3.2 and Table S3.3.

DNA extraction from inocula and anaerobic enrichments. Following 3 years of cultivation, comprising over 10 enrichment phases (Table S3.2), 10 mL of each culture derived

from moose rumen and beaver droppings were harvested at early stationary phase of biogas production by centrifugation at ~9,500 g for 15 min at 4 °C. Total community DNA was extracted from corresponding cell pellets using the QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany). DNA was also extracted from ~1 g (wet weight) of original beaver dropping and moose rumen inocula that had been stored at -80°C. The manufacturer's instructions were followed, including 5 min incubation at 95 °C to improve DNA recovery. The extracted metagenomic DNA was then quantified using a Nanodrop 2000 spectrophotometer (Thermo Scientific, Massachusetts, USA) and stored at -80 ºC.

Multiplex-pyrosequencing of 16S rDNA. Polymerase chain reaction (PCR) was used to amplify the V6-8 hypervariable region of 16S rRNA genes for multiplex-pyrosequencing. A set of 10-nt barcodes designed by Roche was incorporated to the 926 Forward (5'-AAACTYAAAKGAATTGACGG) and 1392 Reverse (5'- ACGGGCGGTGTGTRC) primers for multiplexing (Table S3.4) (DeAngelis et al., 2012). PCR reactions were performed using *Taq* DNA polymerase in 2X PCR Master Mix (Thermo Scientific, Massachusetts, USA) with the following conditions: i) initial denaturation at 95 ºC for 3 min, and ii) 35 cycles of 95 ºC for 30 s, annealing at 55 ºC for 45 s, and elongation at 72 ºC for 90 s, followed by iii) a final extension at 72 ºC for 10 min. Amplicons were purified using the GeneJET PCR Purification Kit (Thermo Scientific, Massachusetts, USA) and quantified using agarose electrophoresis and a Nanodrop 2000 spectrophotometer (Table S3.4), before being sent to the Génome Québec Innovation Centre for precise quantification using the PicoGreen assay (Thermo Scientific, Massachusetts, USA) and sequencing using a 454 GS FLX platform (454 Life Sciences-a Roche Company, Branford, CT, USA).

Data analyses of pyrosequences. Pyrosequencing flowgrams were converted to sequence reads and quality scores using Roche 454 Life Science propriety software (http://www.454.com) and then analysed by Quantitative Insights Into Microbial Ecology (QIIME) 1.8.0 (Caporaso et al., 2010). Sequences were quality checked by filtering those with i) quality scores less than 25, ii) lengths less than 200bp, and iii) longer than 1000 bp. Uclust then clustered high quality 16S rRNA sequences from all samples into OTUs at the threshold of 97% sequence similarity (Edgar, 2010). After filtering OTUs that occurred fewer than twice (i.e. singletons) among the samples, representative sequences from each OTU were aligned to the 16S rRNA sequences archived in the Greengenes Database for taxonomic inference (DeSantis et al., 2006). Based on the OTU table, alpha diversity analysis was used to generate rarefaction plots in QIIME. Using the same program package, hierarchical clustering using Unweighted Pair Group Method with Arithmetic mean (UPGMA) was also performed to interpret the beta diversity distance matrix calculated by UniFrac. Finally, heatmap and Venn diagrams were generated using R scripts (http://www.r-project.org) to visualize the taxonomic affiliation and the distribution of core OTUs among the enrichments.

The raw sequence reads were submitted to National Center for Biotechnology Information (NCBI) under the BioprojectID PRJNA298982.

3.3 Results and Discussion

3.3.1 Establishment of biogas-producing microbial enrichments

Anaerobic enrichments were established and methanogenic activity was sustained over ten feedings for > 3 years on four different lignocellulosic substrate mixes (Table S3.2). Over the enrichment phases, the average volumes of biogas produced per amount of COD added decreased initially and then increased by the ninth growth phase (Figure 3.1; Table S3.3). The improved extent of conversion suggested gradual adaptation to each lignocellulosic carbon source by 154 to 171 weeks of enrichment. At the growth phase just prior to DNA extraction (i.e., phase 9), the extent of carbon source conversion for beaver dropping enrichments were: cellulose, pretreated poplar, cellulose + lignosulphonate, and cellulose + tannic acid. By comparison, the extent of carbon source conversion for moose rumen enrichments were: pretreated poplar, cellulose, cellulose + lignosulphonate, and cellulose + tannic acid (Figure 3.1). As reported by Lacourt (2011) in his thesis, tannic acid consistently inhibited biogas production. The impact of tannic acid is consistent with previous studies, which report that the phenolic hydroxyl groups of tannic acid can complex with proteins, metal ions, amino acids and polysaccharides (Makkar, 2003), thereby inhibiting enzyme action and uptake of essential carbon sources and metal ions (Asiegbu et al., 1995; Bhatta et al., 2009).

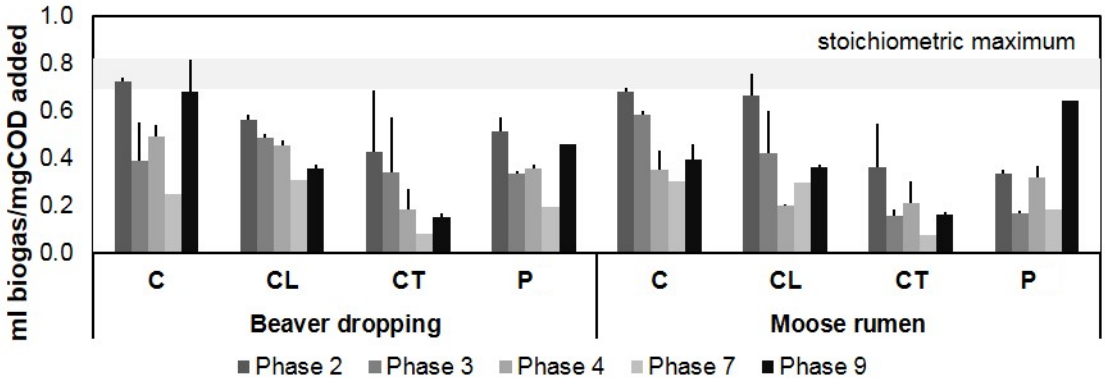


Figure 3.1 Biogas production profile of microcosms fed with various lignocellulosic carbon sources for over 3 years. The range of stoichiometric maximum biogas yield is shown in the grey band to provide a reference for the conversion extents of the fed substrates in the microcosms (see supplementary method for the calculation based on Buswell's formula (Symons and Buswell, 1933)); error bars indicate standard deviation; n=3. C: cellulose; CL: cellulose + lignosulphonate; CT: cellulose + tannic acid; P: pretreated poplar.

3.3.2 Biodiversity indices in enrichment cultures

Overall, 179,801 high quality reads were retained for downstream analyses of community structure, richness and diversity estimators (Table S3.5). Despite extraction of high quality DNA and successful PCR amplification, few reads (less than 15) were retrieved for beaver dropping cultivations enriched on cellulose + tannic acid (data not shown), and so this dataset was removed from downstream analyses. Altogether, quality-screened sequences pooled from the samples were clustered to 5,800 unique OTUs at 97% similarity threshold after the removal of singletons. In the moose rumen enrichments, the decrease in Chao I (richness) and Shannon (diversity) indices compared to the original inoculum were consistent with enrichment of microorganisms best suited

to transform amended carbon sources (Figure S3.1). A similar trend was reported recently in soil microbiota following enrichment with wheat straw (Jimenez et al., 2014a). In contrast, in the beaver dropping enrichments, the comparatively low Chao I index of the inoculum likely reflects the dominance of a soil species as described in the next section. Consistent with this interpretation of Chao I richness, the Simpson's diversity index was highest for the beaver dropping and moose rumen inoculum samples compared to corresponding enrichments. While the time span between defecation and sampling was unclear (as is the case for most fecal studies of wild animals), sample collection from beavers in the wilderness rather than captivity overcomes potential human interference to the gut microbiota. For example, loss in microbial diversity upon captivity has been reported in closely-related woodrats (Kohl et al., 2014). With greater loss in microbial diversity in the diet-specialist (e.g., Stephens' woodrat which consumes a diet of 60–95% juniper) than the diet-generalist (e.g., white-throated woodrat consumes actus, yucca, juniper, other shrubs and grasses), where the original microbiota could not be restored despite the provision of a natural diet.

A high level overview of the amplicon sequencing data revealed that samples from triplicate microcosms for a given enrichment condition clustered most closely, providing confidence in the reproducibility of the analysis (Figure 3.2A). Moreover, the inoculum samples were most divergent from subsequent enrichment cultures, clearly revealing substrate-driven convergence of microbial communities by both UPGMA and Unifrac clustering (Figure 3.2A-B). In the following sections, we first discuss community shifts in the beaver dropping samples, and then those in the moose rumen samples, and finally a holistic view of the emergent OTUs that are shared in the enrichment microcosms. Since the absolute microbial abundance varies across the samples, the following description was based on the relative abundance with arbitrary cut-off

applied to focus on the compositional variation based on enrichment condition. Accordingly, emergent major lineages in the respective community would then be key microbes adapted to the amendment.

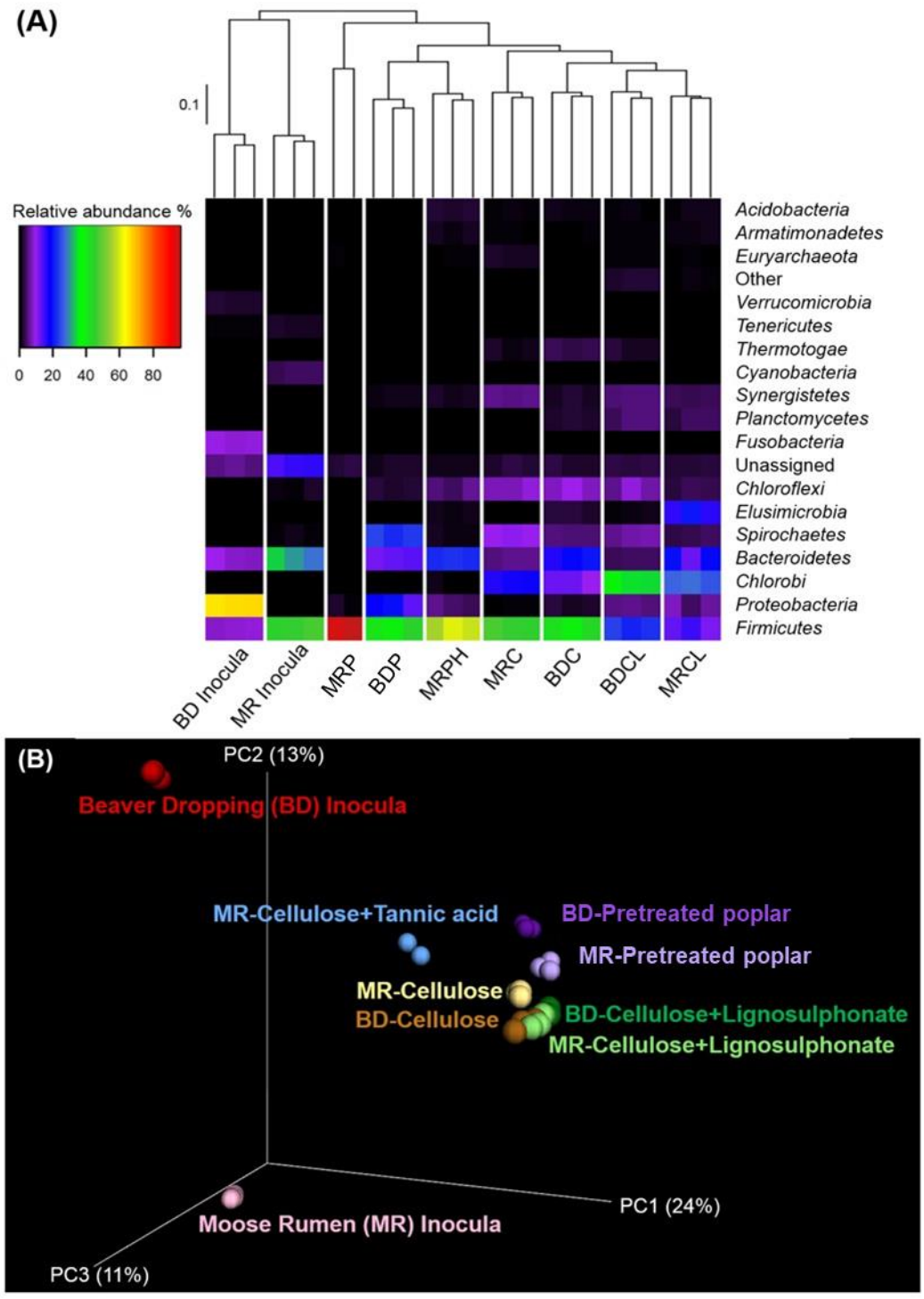


Figure 3.2 Substrate-based clustering of lignocellulose-active microbial communities in beaver dropping and moose rumen, and their corresponding enrichment cultures. (A) Beta diversity-based UPGMA dendrogram with relative abundances of microbial phyla (≥ 1.0% in at

least one sample) shown in a heatmap. (B) Three-dimensional Unifrac principal coordinate analysis (PCoA) plot. BD: beaver dropping; C: cellulose; CL: cellulose + lignosulphonate; CT: cellulose + tannic acid; MR: moose rumen; P: pretreated poplar.

3.3.3 Impact of lignocellulosic substrates on microbial communities originating from beaver droppings

Upon enrichment on lignocellulosic substrates under strictly anaerobic methanogenic conditions, we observed a dramatic decrease in the microbes belonging to the *Proteobacteria* (Figure 3.2A). Specifically, the relative abundance of *Proteobacteria* diminished from approximately 62% in beaver dropping inoculum (sum of all *Proteobacteria* shown in Figure 3.3A) to between 1-17% in corresponding enrichments (Figure 3.3B, C and E), where highest numbers remained in cultures enriched on pretreated poplar (Figure 3.3E). Most of the *Proteobacteria* present in the inoculum belonged to the genus *Pseudomonas* (approximately 30 %) (Figure 3.3A), which are ubiquitous soil facultative bacteria, and notably also comprise species with ability to detoxify pretreated wood (Lopez et al., 2004). Although enrichment on pretreated poplar retained a comparatively high fraction of *Proteobacteria,* the largest group in that enrichment were assigned to the genus *Gammaproteobacteria* (14 %) (Figure 3.3E). Notably, *Gammaproteobacteria* were previously identified in other biomass-degrading communities, including a cellulose-degrading marine biofilm and a wheat straw-degrading microbial consortia (Edwards et al., 2010; Jimenez et al., 2014a). Similarly, the *Fusobacteriaceae* family represented 8.5% of the beaver dropping inoculum (Figure 3.3A), but were not detected in any of corresponding enrichments (Figure 3.3B, C and E). This family comprises microaerophilic to

obligate anaerobes that can ferment carbohydrates and amino acids into various organic acids in anaerobic environments, including the oral and gastrointestinal lining of mammals and anaerobic sediments (Olsen, 2014).

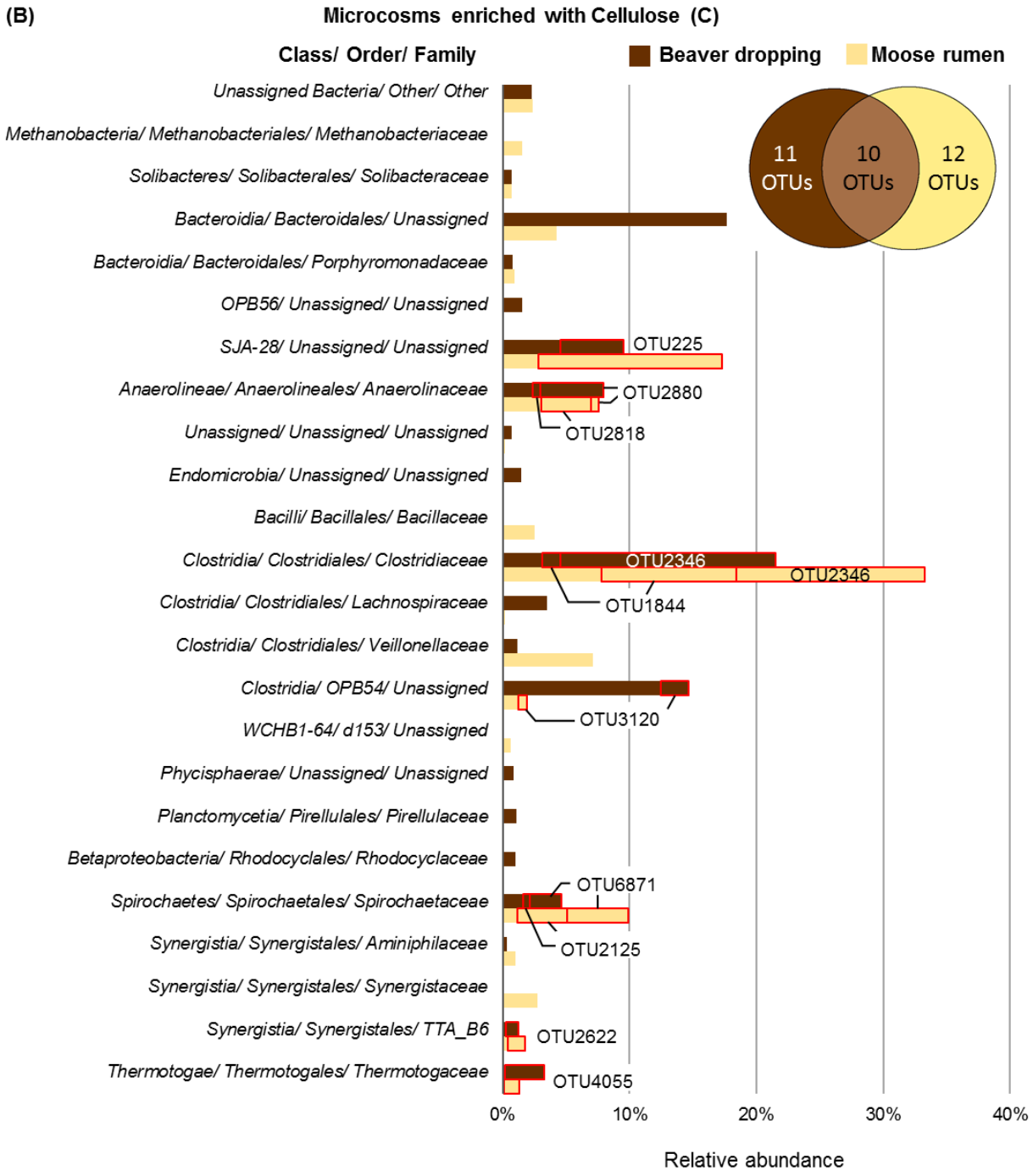
The impact of microbial enrichment was further illustrated by the detection of microbial phyla in enrichments that were not detected in the beaver dropping inoculum given their low abundance. For example, *Chlorobi* in the cellulose and cellulose + lignosulphonate enrichments represented over 10% and 35% of corresponding communities (Figure 3.3B-C), even though this phylum was not detected in the original inoculum. An unassigned member of the uncultured class *SJA-28* constituted nearly 10% of cellulose enrichments and 35% cellulose + lignosulphonate enrichments (Figure 3.3B-C), while comprising 83% and 95% of the *Chlorobi* phylum in these respective cultures. Enrichment on lignocellulosic substrates also led to the detection of *Spirochaetes* and *Chloroflexi* in enrichments originating from beaver droppings. Most notably among the *Spirochetaceae*, bacteria belonging to the genus *Treponema* contributed approximately 5% of all enrichments (Figure 3.3B, C and E), whereas the genus *W22* from the *Cloacamonaceae* family comprised over 17% of the microbial community enriched on pretreated poplar (Figure 3.3E). *Treponema* acetogens were previously identified in the termite gut microbiome, and were predicted to encode glycoside hydrolases targeting cellulose and xylan (Warnecke et al., 2007).

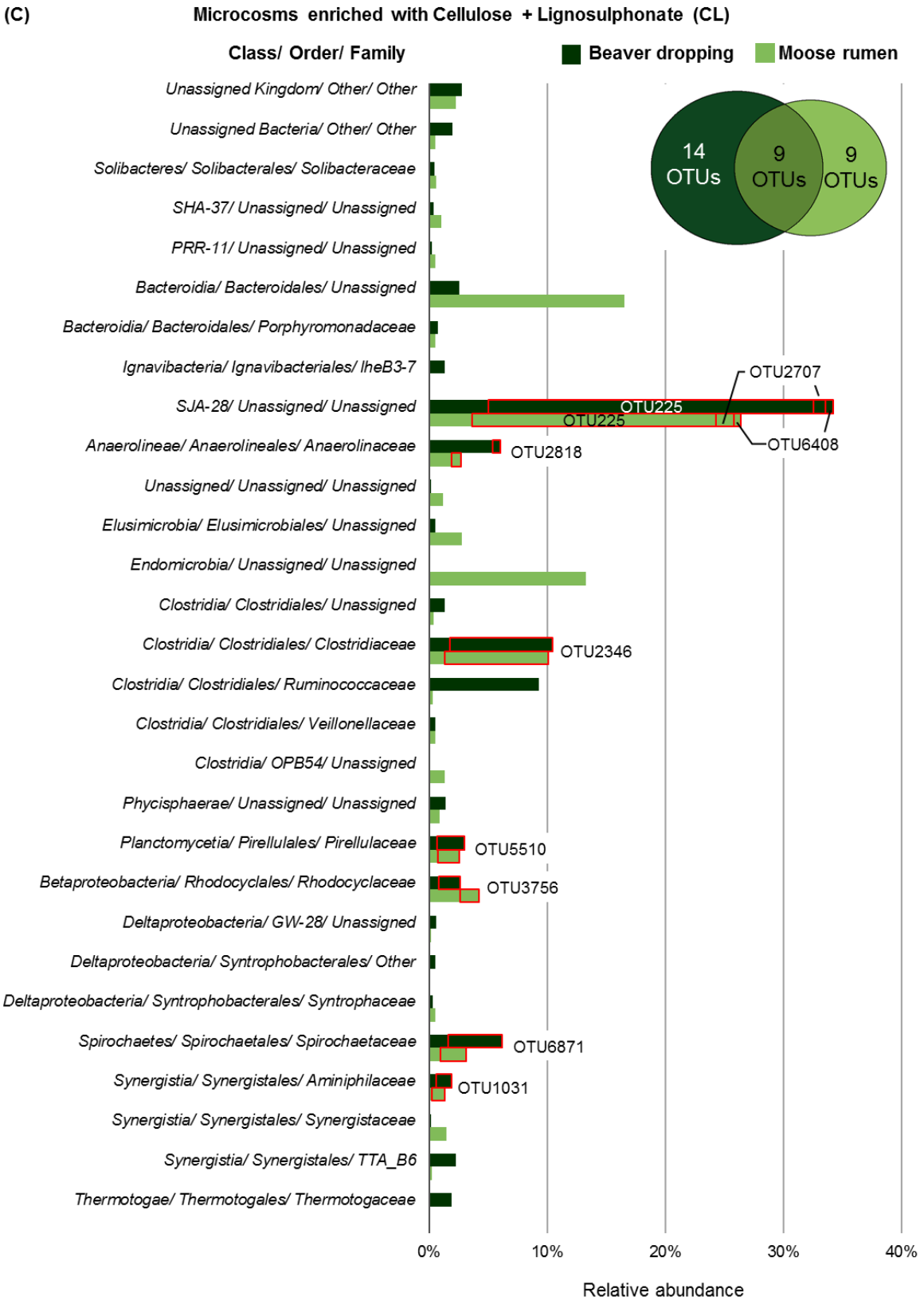
Among the members of *Bacteroidetes*, family S24-7 made up 4 % of the community in the beaver droppings inoculum (Figure 3.3A), while an unassigned lineage contributed up to 18% in enrichments established on cellulose (Figure 3.3B), and 11 % of enrichments established on pretreated poplar (Figure 3.3E). Notably, uncultured *Bacteroidetes* lineages dominate numerous lignocellulose-degrading communities. For example, metagenomic studies of microbiomes from human gut and reindeer rumen revealed high abundance of PUL-like systems originating from

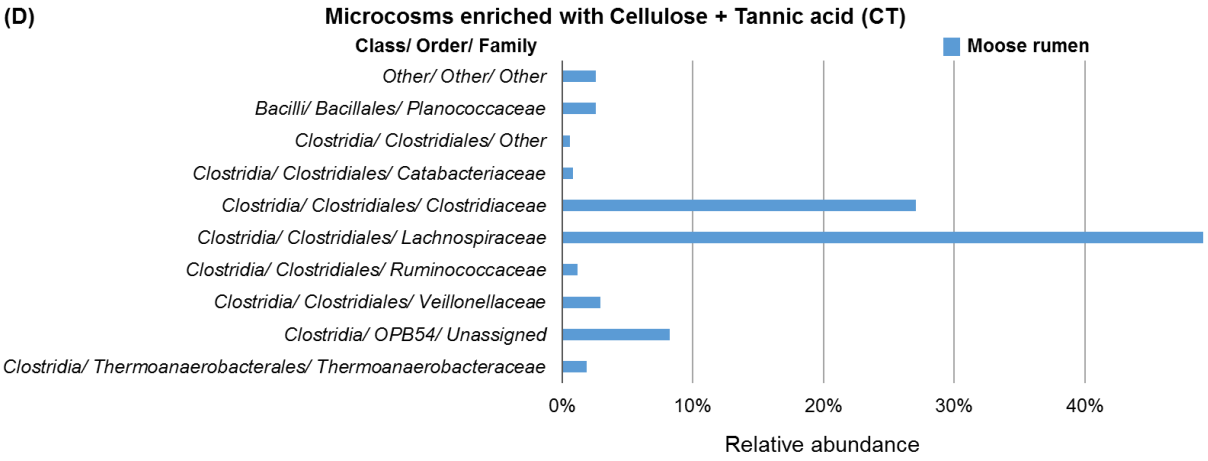
Bacteroidetes (Martens et al., 2009; Pope et al., 2012). These gene clusters encode carbohydrate-active enzymes as well as transport proteins for glycan hydrolysis and uptake and represent a rich reservoir of new lignocellulolytic activities (Terrapon et al., 2015).

Several members of the *Firmicutes* have been implicated as key cellulose degraders. Consistent with this pattern, enrichment of beaver droppings on lignocellulosic substrates led to a four- to eight-fold increase in the relative abundance of members from this phylum (Figure 3.2A). Most significantly, microbes belonging to the genus *Clostridium* and *Ruminococcus* were particularly enriched (Figure 3.3B, C and E), which is consistent with the importance of corresponding species to polysaccharide degradation (Tracy et al., 2012). Moreover, an uncultured lineage in order *OPB54* made up to 15 % of the beaver dropping enrichments amended with cellulose (Figure 3.3B). Notably, *OPB54* was previously identified in low abundance in stillage biogas reactors that operated in high temperatures (Roske et al., 2014).


(A)
Inocula (IN)
Class/ Order/ Family
Beaver dropping
Moose rumen
22 OTUs
29 OTUs
Unassigned Bacteria/ Other/ Other
Bacteroidia/ Bacteroidales/ Unassigned
Bacteroidia/ Bacteroidales/ BS11
Bacteroidia/ Bacteroidales/ Bacteroidaceae
Bacteroidia/ Bacteroidales/ Prevotellaceae
Bacteroidia/ Bacteroidales/ S24-7
Bacteroidia/ Bacteroidales/ Paraprevotellaceae
Flavobacteriia/ Flavobacteriales/ Flavobacteriaceae
Anaerolineae/ Anaerolineales/ Anaerolinaceae
Chloroplast/ Streptophyta/ Unassigned
Bacilli/ Lactobacillales/ Carnobacteriaceae
Clostridia/ Clostridiales/ Other
Clostridia/ Clostridiales/ Clostridiaceae
Clostridia/ Clostridiales/ Lachnospiraceae
Clostridia/ Clostridiales/ Peptostreptococcaceae
Clostridia/ Clostridiales/ Ruminococcaceae
Clostridia/ Coriobacteriales/ Unassigned
Clostridia/ Coriobacteriales/ Coriobacteriaceae
Erysipelotrichi/ Erysipelotrichales/ Erysipelotrichaceae
Fusobacteria/ Fusobacteriales/ Fusobacteriaceae
Alphaproteobacteria/ Sphingomonadales/ Sphingomonadaceae
Betaproteobacteria/ Burkholderiales/ Burkholderiaceae
Betaproteobacteria/ Burkholderiales/ Comamonadaceae
Betaproteobacteria/ Burkholderiales/ Oxalobacteraceae
Gammaproteobacteria/ Other/ Other
Gammaproteobacteria/ Enterobacteriales/ Enterobacteriaceae
Gammaproteobacteria/ Pseudomonadales/ Moraxellaceae
Gammaproteobacteria/ Pseudomonadales/ Pseudomonadaceae
Spirochaetes/ Spirochaetales/ Spirochaetaceae
Mollicutes/ RF39/ Unassigned
Verrucomicrobiae/ Verrucomicrobiales/ Verrucomicrobiaceae
0%
10%
20%
30%
40%
Relative abundance


(B)
Microcosms enriched with Cellulose (C)
Class/ Order/ Family
Beaver dropping
Moose rumen
Unassigned Bacteria/ Other/ Other
Methanobacteria/ Methanobacteriales/ Methanobacteriaceae
Solibacteres/ Solibacterales/ Solibacteraceae
Bacteroidia/ Bacteroidales/ Unassigned
Bacteroidia/ Bacteroidales/ Porphyromonadaceae
OPB56/ Unassigned/ Unassigned
SJA-28/ Unassigned/ Unassigned
Anaerolineae/ Anaerolineales/ Anaerolinaceae
Unassigned/ Unassigned/ Unassigned
Endomicrobia/ Unassigned/ Unassigned
Bacilli/ Bacillales/ Bacillaceae
Clostridia/ Clostridiales/ Clostridiaceae
Clostridia/ Clostridiales/ Lachnospiraceae
Clostridia/ Clostridiales/ Veillonellaceae
Clostridia/ OPB54/ Unassigned
WCHB1-64/ d153/ Unassigned
Phycisphaerae/ Unassigned/ Unassigned
Planctomycetia/ Pirellulales/ Pirellulaceae
Betaproteobacteria/ Rhodocyclales/ Rhodocyclaceae
Spirochaetes/ Spirochaetales/ Spirochaetaceae
Synergistia/ Synergistales/ Aminiphilaceae
Synergistia/ Synergistales/ Synergistaceae
Synergistia/ Synergistales/ TTA_B6
Thermotogae/ Thermotogales/ Thermotogaceae
11 OTUs
10 OTUs
12 OTUs
OTU225
OTU2880
OTU2818
OTU2346
OTU2346
OTU1844
OTU3120
OTU6871
OTU2125
OTU2622
OTU4055
0%
10%
20%
30%
40%
Relative abundance


(C)
Microcosms enriched with Cellulose + Lignosulphonate (CL)
Class/ Order/ Family
Beaver dropping
Moose rumen
Unassigned Kingdom/ Other/ Other
Unassigned Bacteria/ Other/ Other
Solibacteres/ Solibacterales/ Solibacteraceae
SHA-37/ Unassigned/ Unassigned
PRR-11/ Unassigned/ Unassigned
Bacteroidia/ Bacteroidales/ Unassigned
Bacteroidia/ Bacteroidales/ Porphyromonadaceae
Ignavibacteria/ Ignavibacteriales/ IheB3-7
SJA-28/ Unassigned/ Unassigned
Anaerolineae/ Anaerolineales/ Anaerolinaceae
Unassigned/ Unassigned/ Unassigned
Elusimicrobia/ Elusimicrobiales/ Unassigned
Endomicrobia/ Unassigned/ Unassigned
Clostridia/ Clostridiales/ Unassigned
Clostridia/ Clostridiales/ Clostridiaceae
Clostridia/ Clostridiales/ Ruminococcaceae
Clostridia/ Clostridiales/ Veillonellaceae
Clostridia/ OPB54/ Unassigned
Phycisphaerae/ Unassigned/ Unassigned
Planctomycetia/ Pirellulales/ Pirellulaceae
Betaproteobacteria/ Rhodocyclales/ Rhodocyclaceae
Deltaproteobacteria/ GW-28/ Unassigned
Deltaproteobacteria/ Syntrophobacterales/ Other
Deltaproteobacteria/ Syntrophobacterales/ Syntrophaceae
Spirochaetes/ Spirochaetales/ Spirochaetaceae
Synergistia/ Synergistales/ Aminiphilaceae
Synergistia/ Synergistales/ Synergistaceae
Synergistia/ Synergistales/ TTA_B6
Thermotogae/ Thermotogales/ Thermotogaceae
14 OTUs
9 OTUs
9 OTUs
OTU2707
OTU225
OTU225
OTU6408
OTU2818
OTU2346
OTU5510
OTU3756
OTU6871
OTU1031
0%
10%
20%
30%
40%
Relative abundance


(D)
Microcosms enriched with Cellulose + Tannic acid (CT)
Class/ Order/ Family
Moose rumen
Other/ Other/ Other
Bacilli/ Bacillales/ Planococcaceae
Clostridia/ Clostridiales/ Other
Clostridia/ Clostridiales/ Catabacteriaceae
Clostridia/ Clostridiales/ Clostridiaceae
Clostridia/ Clostridiales/ Lachnospiraceae
Clostridia/ Clostridiales/ Ruminococcaceae
Clostridia/ Clostridiales/ Veillonellaceae
Clostridia/ OPB54/ Unassigned
Clostridia/ Thermoanaerobacterales/ Thermoanaerobacteraceae
0%
10%
20%
30%
40%
Relative abundance

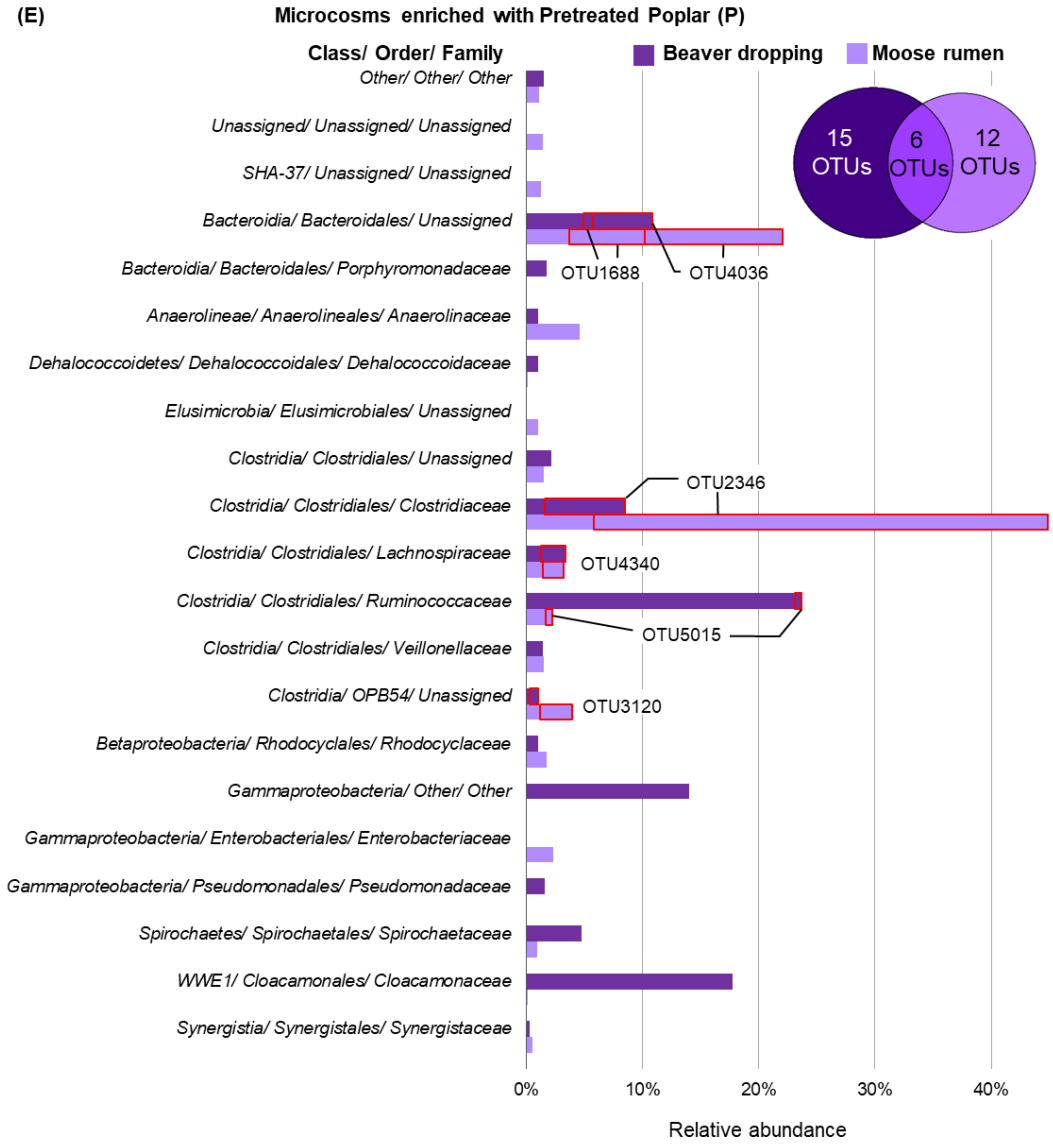

(E)
Microcosms enriched with Pretreated Poplar (P)
Class/ Order/ Family
Beaver dropping
Moose rumen
Other/ Other/ Other
Unassigned/ Unassigned/ Unassigned
SHA-37/ Unassigned/ Unassigned
Bacteroidia/ Bacteroidales/ Unassigned
Bacteroidia/ Bacteroidales/ Porphyromonadaceae
Anaerolineae/ Anaerolineales/ Anaerolinaceae
Dehalococcoidetes/ Dehalococcoidales/ Dehalococcoidaceae
Elusimicrobia/ Elusimicrobiales/ Unassigned
Clostridia/ Clostridiales/ Unassigned
Clostridia/ Clostridiales/ Clostridiaceae
Clostridia/ Clostridiales/ Lachnospiraceae
Clostridia/ Clostridiales/ Ruminococcaceae
Clostridia/ Clostridiales/ Veillonellaceae
Clostridia/ OPB54/ Unassigned
Betaproteobacteria/ Rhodocyclales/ Rhodocyclaceae
Gammaproteobacteria/ Other/ Other
Gammaproteobacteria/ Enterobacteriales/ Enterobacteriaceae
Gammaproteobacteria/ Pseudomonadales/ Pseudomonadaceae
Spirochaetes/ Spirochaetales/ Spirochaetaceae
WWE1/ Cloacamonales/ Cloacamonaceae
Synergistia/ Synergistales/ Synergistaceae
15 OTUs
6 OTUs
12 OTUs
OTU1688
OTU4036
OTU2346
OTU4340
OTU5015
OTU3120
0%
10%
20%
30%
40%
Relative abundance

Figure 3.3 Relative abundances of microbial families (≥ 1.0% in at least one sample) in beaver dropping and moose rumen, and their corresponding enrichment cultures. (A) Inocula and microcosms fed with (B) cellulose, (C) cellulose + lignosulphonate, (D) cellulose + tannic acid, and (E) pretreated poplar. The distribution of OTUs (≥ 0.5% in at least one sample) are shown in the Venn diagram; shared OTUs are highlighted with red.

3.3.4 Impact of lignocellulosic substrates on microbial communities originating from moose rumen

Similar to the beaver droppings enrichments, phyla that were common to all enrichments derived from moose rumen samples included *Firmicutes*, *Bacteroidetes*, *Chlorobi*, *Elusimicrobia*, and *Spirochaetes* (Figure 3.2A). In the case of samples from the moose rumen enrichment cultures, all samples had a high relative abundance of *Firmicutes* that was comparable between the inoculum (42%) (Figure 3.3A) and enrichments amended with cellulose (45%) and pretreated poplar (58%) (Figure 3.3B and E), but lower in enrichments on cellulose + lignosulphonate (13%) and higher in the enrichments on cellulose + tannic acid (93%) (Figure 3.3C-D). Most dramatically, the fraction of *Clostridium* species increased from 3% in the inoculum to 33% and 45% in cellulose and pretreated poplar enrichments, respectively (Figure 3.3A, B and E). By comparison, enrichments amended with cellulose + tannic acid were distinguished by over 45% of bacteria belonging to the *Lachnospiraceae* family (Figure 3.3D). The representation of this family decreased from 12% of the inoculum (Figure 3.3A) to less than 5% of other enrichments (Figure 3.3B-E). *Lachnospiraceae* members were previously identified in moose rumen and foreguts of dromedary camels (Samsudin et al., 2011; Meehan and Beiko, 2014), but were not

reported in reindeer gut (Pope et al., 2012). Cellulose + tannic acid cultivations were further distinguished by an increase in the fraction of uncultured bacteria within the *Clostridia* class (Figure 3.3D), particularly from order *OPB54* as was observed for beaver dropping cultures enriched on cellulose (increase from 0.1% in the inoculum to 6.3% in the enrichment) as well as unassigned genera within the *Lachnospiraceae* family (from 3% to 48% in the enrichment) (Figure 3.3A).

In addition to the *Firmicutes*, members of the phylum *Chlorobi* increased from non-detectable levels in the inoculum to 17% and 27% of cellulose and cellulose + lignosulphonate cultures (Figure 3.3B-C), respectively. As observed for corresponding enrichments of beaver droppings, this increase was mainly attributed to enrichment of bacteria belonging to class *SJA-28*. Members of the phylum *Elusimicrobia* (formerly *Termite Group 1*; *TG1*) were also enriched through growth on cellulose + lignosulphonate, from less than 1% in the inoculum to over 17% in the enrichment culture (Figure 3.3C). Notably, growth on other lignocellulosic amendments did not increase levels of *Elusimicrobia* members.

Whereas the relative abundances of *Firmicutes*, *Chlorobi*, and *Elusimicrobia* increased upon various lignocellulosic enrichments, the total fraction and species diversity of *Bacteroidetes* decreased from 23% in the inoculum to lower levels in the enrichment cultures (Figure 3.2A). Specifically, *BS11* (11%) and *Prevotella* (7%) in the moose rumen sample became non-detectable after the enrichment process (Figure 3.3A), whereas an unassigned lineage under *Bacteroidales* was maintained after enrichment with cellulose (4%), and increased upon cellulose + lignosulphonate (17%) and pretreated poplar (22%) (Figure 3.3B, C and E). By contrast, none of the *Bacteroidetes* species were detected in enrichment cultures amended with cellulose + tannic acid.

3.3.5 Comparative analysis of all microbial enrichments

As explained above, an underlying hypothesis of the enrichment study was that the relative abundance of microbes most relevant to lignocellulose conversion would increase by culturing the selected inocula on lignocellulosic substrates. Consistent with our hypothesis, abundances of known lignocellulose degraders increased following amendment with selected lignocellulosic carbon sources and included microbial lineages previously identified in the termite hindgut, such as *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Spirochaetes*, and *Elusimicrobia*, or the bovine rumen microbiome, such as *Chlorobi*, *Chloroflexi*, and *Fusobacteria* (Brulc et al., 2009). Moreover, UPGMA clustering of OTU sequences revealed convergence of microbial communities enriched with the same carbon source (Figure 3.2A); convergence of community composition was also revealed through UniFrac analysis (Figure 3.2B). Overall, microbial community compositions could be grouped into three main sub-clusters (Figure 3.2A), namely, i) original inoculum, ii) enrichment on cellulose or cellulose + lignosulphonate, and iii) enrichment on pretreated poplar; the moose rumen enrichment on cellulose + tannic acid formed a fourth, unique branch. Furthermore, the majority of OTUs present in beaver dropping and moose rumen inocula were not detected in the corresponding enrichments (Figure 3.4A), this trend was even more obvious when a threshold of 0.5% abundance is applied (Figure 3.4B). Indeed, the most abundant OTUs in enrichments represented organisms that comprised less than 0.5% of all OTUs in the original inoculum. This demonstrates strong selection of microbial members for a given lignocellulose amendment, which was underscored by the few overlapping OTUs between enrichments on different lignocellulose amendments (Figure 3.4B). One notable exception was

otu6272 (assigned to class *Gammaproteobacteria*), which represented nearly 8% of all OTUs in the beaver dropping inoculum and 14% of OTUs after enrichment on pretreated poplar.

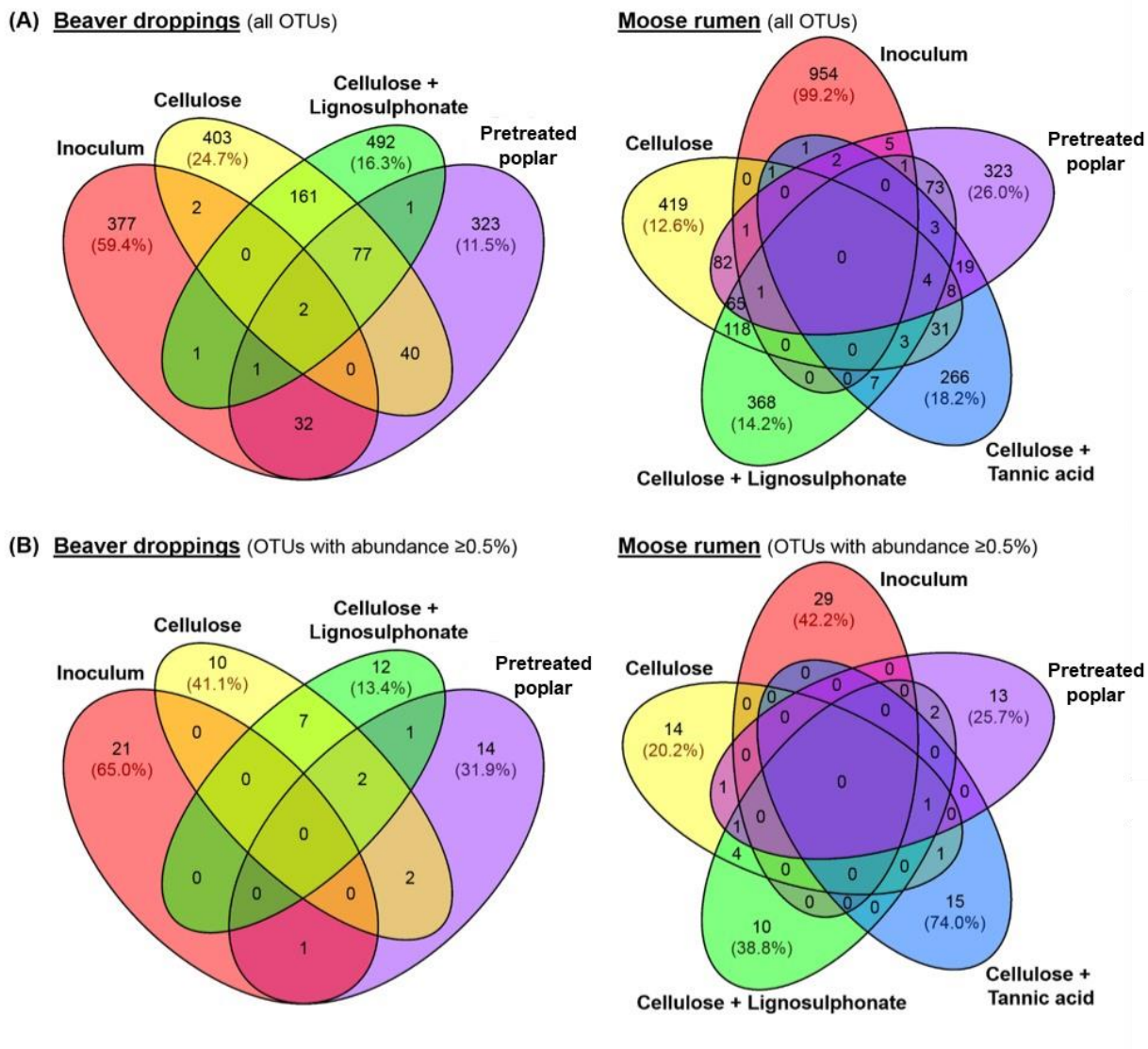


Figure 3.4 Distribution of (A) all OTUs and (B) OTUs with relative abundances ≥ 0.5% in beaver dropping and moose rumen, and their corresponding enrichment cultures.

Abundances of represented OTUs are shown in brackets.

Core species were not identified among enrichments originating from the same inoculum (Figure 3.4A-B). Moreover, despite no shared OTUs between beaver droppings and moose rumen inocula at a 0.5% abundance cut-off (Figure 3.3A), a high number of overlapping members constituting the shared dominant taxonomic lineages was observed between enrichments fed with the same lignocellulose carbon source (Figure 3.3B-E). For example, otu2346 assigned to the *Clostridium* genus was not detected in beaver dropping or moose rumen inocula, but comprised a significant fraction of cellulose enrichments (15-17%), cellulose + lignosulphonate enrichments (9%), and pretreated poplar enrichments (7-49%) (Figure 3.3B, C and E). Similarly, an OTU assigned to class *SJA-28* (otu225) was not observed in either inoculum, but represented 5-15% of both cellulose enrichments and 20-28% of both cellulose + lignosulphonate enrichments (Figure 3.3B-C). Finally, otu4036 belonging to the order *Bacteroidales* was uniquely detected in enrichments established on pretreated poplar (Figure 3.3E), where it comprised 5-10% of the bacterial community even though it was not detected in either inoculum or any other enrichment condition.

Despite the convergence of microbial communities enriched on the same lignocellulosic carbon source, unique lineages were also observed that reflect the impact of starting inocula. After enrichment of beaver droppings on cellulose + lignosulphonate and pretreated poplar, a *Ruminococcus* OTU (otu2378) represented 6% and 20% of corresponding microbial communities (Figure 3.3C and E). By contrast, otu2378 was not found in any moose rumen enrichments. Similarly, the *Gammaproteobacteria* OTU (otu6272) and *W22* OTU (otu3890) identified in beaver droppings (Figure 3.3A) and corresponding enrichments on pretreated poplar (Figure 3.3E) was not detected in moose rumen samples or any of the derived enrichments.

Correlations between microbial membership and lignocellulosic substrate also emerged by identifying key differences between communities resulting from the different amendments. For example, the abundance of class *SJA-28* in enrichments amended with cellulose + lignosulphonate was double that of enrichments amended with cellulose. By contrast, the abundance of orders *OPB54* and *Clostridiales* consistently decreased upon amendment with cellulose + lignosulphonate compared to addition of cellulose alone. Figure 3.5 summarizes the specific OTUs that were enriched upon lignocellulosic amendment, as well as reported habitats of the understudied lineages. In addition to these genera, class *Endomicrobia*, order *Bacteroidales*, family *Lachnospiraceae*, and genus *W22* represent additional sources of understudied microorganisms that could comprise unique enzymes and biochemical pathways relevant to lignocellulose conversion.

OTU	Class/ Order/ Family/ Genus	Beaver dropping	Moose rumen	Habitat
Inocula (IN)				
7271	*Fusobacteria/ Fusobacteriales/ Fusobacteriaceae/ Unassigned*	8.5%	0.0%	
2645	*Betaproteobacteria/ Burkholderiales/ Oxalobacteraceae/ Janthinobacterium*	6.5%	0.0%	
6272	***Gammaproteobacteria/ Unassigned/ Unassigned/ Unassigned***	7.9%	0.0%	
1063	*Gammaproteobacteria/ Pseudomonadales/ Moraxellaceae/ Acinetobacter*	5.6%	0.0%	
3776	*Gammaproteobacteria/ Pseudomonadales/ Pseudomonadaceae/ Pseudomonas*	16.9%	0.0%	
3137	Unassigned/Unassigned/ Unassigned/ Unassigned	0.0%	13.1%	
	Total represented abundance	45.4%	13.1%	
Cellulose (C)				
8016	***Bacteroidia/ Bacteroidales/ Unassigned/ Unassigned***	17.2%	0.0%	
225	***SJA-28/ Unassigned/ Unassigned/ Unassigned***	5.0%	14.5%	
2880	*Anaerolineae/ Anaerolineales/ Anaerolinaceae/ Unassigned*	5.0%	0.6%	
1844	*Clostridia/ Clostridiales/ Clostridiaceae/ Clostridium*	1.4%	10.6%	
2346	***Clostridia/ Clostridiales/ Clostridiaceae/ Clostridium***	17.0%	14.9%	
332	*Clostridia/ OPB54/ Unassigned/ Unassigned*	9.7%	0.0%	Thermophilic stillage bioreactor (Roske et al., 2014)
6871	***Spirochaetes/ Spirochaetales/ Spirochaetaceae/ Treponema***	2.5%	4.8%	Cellulose- and xylan-degrading hindgut microbiota of termite (Warnecke et al., 2007)
	Total represented abundance	57.8%	45.4%	
Cellulose + Lignosulphonate (CL)				
8016	***Bacteroidia/ Bacteroidales/ Unassigned/ Unassigned***	0.0%	15.8%	
225	***SJA-28/ Unassigned/ Unassigned/ Unassigned***	27.5%	20.6%	
712	*Endomicrobia/ Unassigned/ Unassigned/ Unassigned*	0.0%	12.1%	
2346	***Clostridia/ Clostridiales/ Clostridiaceae/ Clostridium***	8.7%	8.8%	
2378	***Clostridia/ Clostridiales/ Ruminococcaceae/ Ruminococcus***	5.8%	0.0%	
6871	***Spirochaetes/ Spirochaetales/ Spirochaetaceae/ Treponema***	4.6%	2.2%	Cellulose- and xylan-degrading hindgut microbiota of termite (Warnecke et al., 2007)
	Total represented abundance	46.6%	59.5%	
Cellulose + Tannic acid (CT)				
6522	*Clostridia/ Clostridiales/ Clostridiaceae/ Clostridium*	-	17.1%	
212	*Clostridia/ Clostridiales/ Lachnospiraceae/ Unassigned*	-	27.0%	Moose rumen and forequts of dromedary camels (Samsudin et al., 2011; Meehan and Beiko, 2014)
1272	*Clostridia/ Clostridiales/ Lachnospiraceae/ Unassigned*	-	6.3%	Moose rumen and forequts of dromedary camels (Samsudin et al., 2011; Meehan and Beiko, 2014)
1878	*Clostridia/ Clostridiales/ Lachnospiraceae/ Unassigned*	-	8.3%	Moose rumen and forequts of dromedary camels (Samsudin et al., 2011; Meehan and Beiko, 2014)
3120	*Clostridia/ OPB54/ Unassigned/ Unassigned*	-	4.6%	Thermophilic stillage bioreactor (Roske et al., 2014)
	Total represented abundance	-	63.3%	
Pretreated poplar (P)				
1688	*Bacteroidia/ Bacteroidales/ Unassigned/ Unassigned*	0.8%	6.5%	
4036	*Bacteroidia/ Bacteroidales/ Unassigned/ Unassigned*	5.1%	11.9%	
2346	***Clostridia/ Clostridiales/ Clostridiaceae/ Clostridium***	6.9%	39.0%	
2378	***Clostridia/ Clostridiales/ Ruminococcaceae/ Ruminococcus***	19.3%	0.0%	
6272	***Gammaproteobacteria/ Unassigned/ Unassigned/ Unassigned***	14.0%	0.0%	
3890	*WWE1/ Cloacamonales/ Cloacamonaceae/ W22*	15.7%	0.0%	
	Total represented abundance	61.8%	57.4%	

Figure 3.5 Abundant OTUs from enrichment microcosms fed with various lignocellulosic carbon sources. Abundances (≥ 4% in at least one sample) are indicated by the

relative length of data bars, which are colour-coded to represent the inocula (red), and enrichment microcosms fed with cellulose (yellow), cellulose + lignosulphonate (green), cellulose + tannic acid (blue), and pretreated poplar (purple).

3.4 Conclusion

Overall, enrichment of beaver droppings and moose rumen on multiple lignocellulosic substrates led to the proliferation of recognized cellulolytic bacteria as well as unique lineages that were in low or undetectable abundances in corresponding inocula. These unassigned lineages were grouped in classes *SJA-28*, *Endomicrobia*, orders *Bacteroidales*, *OPB54* and family *Lachnospiraceae*, and comprised up to half of corresponding communities, warranting future investigation on their potential in lignocellulose-degradation. At the same time, microorganisms that were unique to the presence of pretreatment inhibitors, such as *SJA-28* in enrichments amended with cellulose + lignosulphonate and bacteria from order *OPB54* or *Lachnospiraceae* family in enrichments amended with cellulose + tannic acid, may comprise specialized catabolic activities resistant to the inhibition or relevant to the detoxification of pretreated wood (Lopez et al., 2004). In the context of this thesis, the substrate-based convergence of microbial community compositions originating from beaver droppings and moose rumen suggests that the resulting communities have specialized to the enrichment condition, and that corresponding microorganisms may encode distinct carbohydrate-active enzymes that are particularly effective towards the given lignocellulosic carbon source and despite the inhibition from the added lignosulphonate or tannic acid. Accordingly, the following Chapter 4 explores the genetic capacity of the resulting microbiomes in the degradation of plant polysaccharide via metagenomic analysis.

Chapter 4 Comparative metagenomics of cellulose- and pretreated poplar-degrading microcosms from gut microflora of the Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*) after long-term enrichment

---

This chapter is adapted from the following peer-reviewed publication:

Wong, M.T., Wang, W., Couturier, M., Razeq, F.M., Lombard, V., Lapebie, P., Edwards, E.A., Terrapon, N., Henrissat, B., and Master, E.R. (2017). Comparative metagenomics of cellulose- and poplar hydrolysate-degrading microcosms from gut microflora of the Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*) after long-term enrichment. *Frontiers in Microbiology* 8.

Note that the enrichment substrate pretreated poplar was referred to as poplar hydrolysate in the publication as it was the name provided by the supplier SunOpta Inc., Canada.

4.1 Introduction

Lignocellulose comprises the non-edible fraction of plant biomass and as such is a recognized resource for the production of renewable energy, chemicals, and materials. The main components of lignocellulose are cellulose, hemicellulose, pectin and lignin, the proportions and particular chemistries of which depend largely on the plant type and fraction (Kumar et al., 2008). For instance, glucuronoarabinoxylan is a typical hemicellulose found in agricultural crops, whereas glucuronoxylan is the predominant hemicellulose in wood tissue of deciduous trees,

including poplar (Peng et al., 2012). Bioconversion pathways that convert various lignocellulose sources to targeted end products require the concerted action of CAZymes, which include GHs, CEs, PLs, AAs, and of CBMs that are classified into families according to amino acid sequence similarity in the CAZy database (www.cazy.org) (Lombard et al., 2014). Various pretreatment methods, including steam explosion, have been developed to maximize enzymatic conversion of lignocellulosic resources (Excoffier et al., 1991).

Metagenomic approaches to identify CAZymes relevant to the conversion of a given biomass feedstock have considered environmental samples persistently subjected to the targeted feedstock. For example, metagenomic analyses aimed at identifying CAZymes most relevant to bioconversion of non-woody biomass have sampled digestive systems of animals that graze on straw, grasses and lichens (Pope et al., 2010). Parallel metagenomic analyses to identify CAZymes contributing to bioconversion of woody biomass have included samples ranging from forest soils (Damon et al., 2012; Pold et al., 2016), to insects (Warnecke et al., 2007; He et al., 2013; Rossmassler et al., 2015), and wood-feeding mollusks (O'Connor et al., 2014). Alternatively, enrichment of environmental samples on specific biomass feedstocks prior to metagenome sequencing can improve sequence assembly (DeAngelis et al., 2010), while facilitating the identification of most pertinent CAZymes (van der Lelie et al., 2012). In a few cases, direct comparison of CAZyme profiles has also been performed for enrichment cultures originating from the same source. Examples include soil-derived microbial communities enriched with wheat straw, switchgrass and corn stover (Jimenéz et al., 2016), and those digesting mixed lignocellulosic substrates in stationary versus submerged and agitated conditions (Heiss-Blanquet et al., 2016; Wang et al., 2016b). In addition to the influence of enrichment condition, such metagenomic studies highlight the increase in number of CAZyme sequences from families

associated with hydrolysis of oligosaccharides and side groups of hemicelluloses and/or pectins (e.g., GH3, GH43).

Aside from identifying CAZyme families most pertinent to conversion of specific biomass feedstocks, metagenomic analysis of biomass-degrading communities has uncovered a diverse array of encoded PULs and multi-modular proteins. Briefly, PULs comprise physically-linked genes that encode CAZymes and other proteins that work in concert to degrade specific glycans. Accordingly, PULs have emerged as especially fruitful regions within metagenome sequences for enzyme discovery (Larsbrink et al., 2016; Patrascu et al., 2017). For instance, in the past year alone, detailed biochemical characterization of PULs with different selectivity has uncovered novel activities that contribute to the degradation of pectin (Ndeh et al., 2017), xylan (Wang et al., 2016b), and galactomannan (Bagenholm et al., 2017), as well as fungal cell wall components including chitin (Larsbrink et al., 2016) and β-glucans (Temple et al., 2017). Likewise, multi-modular proteins and cellulosomal subunits identified from metagenome sequences and bacterial isolates constitute an additional reservoir for CAZyme discovery (Zhang et al., 2014c). Most recently, CAZyme-linked dockerins were reported in PULs in a moose rumen microbiome (Svartström et al., 2017).

With the aim to identify CAZymes and novel proteins that target woody biomass, here we applied a comparative metagenomics approach to identify microbial enzymes encoded by the gut digestive microflora of wood-feeding Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*) that are likely to promote the conversion of pretreated wood chips. Briefly, in winter months especially, beaver apply an obligate woody diet consisting of twigs, bark and tree trunks; such seasonal confinement to a wood-based diet is also common to moose, ungulates who consume twigs, shrubs, and bark during the winter (Chaney, 2003; Hood and

Bayley, 2009). In Chapter 3, we confirmed the existence of biomass-degrading microorganisms in gut digestive microflora of beaver and moose, and their ability to transform lignocellulosic substrates under anaerobic conditions (Wong et al., 2016). Herein, we report the metagenomes of corresponding gut digestive microflora enriched for over 3 years on either microcrystalline cellulose or pretreated poplar. In particular, metagenome sequences were compared to reveal CAZyme families that are consistently enriched following growth on pretreated poplar compared to growth on cellulose. The four metagenomes were also mined in an effort to identify novel candidate enzymes for future characterization. Two approaches were devised to facilitate this analysis: (1) prediction of bacterial PULs and analysis of encoded proteins with unknown function within, and (2) prediction of multi-modular proteins that comprise both modules recognized to contribute to polysaccharide conversion (e.g., a carbohydrate binding domain) and domains with unknown function.

4.2 Materials and Methods

Setup and maintenance of lignocellulose active enrichment cultures. As previously described (Wong et al., 2016), lignocellulose-degrading microorganisms from the digestive systems of Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*) were sampled and enriched under anaerobic conditions at 36 °C. Briefly, approximately 15 mL of the beaver dropping and moose rumen inocula were transferred to separate 160 mM Wheaton glass serum bottles, which were amended with 45 mL of sulphide-reduced mineral medium and 36 mg COD equivalents of microcrystalline cellulose (Avicel PH101; purchased from Sigma Aldrich) or steam-exploded poplar (provided by SunOpta Inc., Canada) (Wong et al., 2016). Biogas production by resulting cultures was carefully monitored to track metabolic activity.

Metagenomic DNA extraction and sequencing. Following 3 years of cultivation and 10 enrichment phases, 10 mL of each enrichment culture were harvested at early stationary phase of biogas production. Samples were centrifuged at 15,000 g for 15 min at 4°C, and total community DNA was extracted using the QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany) (Table S4.1). The concentration and quality of the extracted metagenomic DNA were assessed by measuring the 260/280 absorbance ratio using a Nanodrop 2000 spectrophotometer (Thermo Scientific, MA, USA), and then stored at −80 °C. A TruSeq library was constructed for the each DNA sample. Illumina paired-end sequencing was chosen and performed with Illumina HiSeq 2000 (Illumina Inc., San Diego, CA, USA) at Génome Québec Innovation Centre.

Metagenome assembly. The output reads were processed by Trimmomatic 0.32 for the removal of adapters and quality filtering (Bolger et al., 2014). The quality-trimmed metagenomic

reads were assembled using Assembly By Short Sequences (Abyss) with minimum coverage of 20, and minimum kmer length of 96 nucleotides (Simpson et al., 2009).

Annotation of CAZyme families, multi-modular sequences, and PULs. Assembled contigs were subjected to open reading frame prediction using Prodigal (Hyatt et al., 2010); predicted proteins were then assigned to CAZyme families using a combination of BLAST and HMM searches against CAZy reference sequences and families as already described (Al-Masaudi et al., 2017). Counts of CAZyme sequences were normalized to compare the diversity of CAZyme sequences identified within each enrichment culture. Specifically, since the number of predicted open reading frames (ORFs) was highest for the metagenome of beaver droppings enriched on pretreated poplar (BD-P, Table 4.1) then:

$$Normalized\ count = \frac{count\ of\ identified\ sequences\ in\ a\ given\ CAZyme\ family\ within\ a\ metagenome\ of\ interest\ \times\ number\ of\ ORFs\ of\ BD-P\ metagenome}{number\ of\ ORFs\ of\ a\ metagenome\ of\ interest}$$

Relative abundance for a given CAZyme family in a metagenome was calculated by

$$Relative\ abundance = \frac{identified\ counts\ for\ a\ given\ CAZyme\ family\ in\ a\ metagenome}{total\ identified\ counts\ of\ CAZyme\ in\ a\ metagenome} \times 100\%$$

Public metagenomes from the digestive systems of cow (Hess et al., 2011), moose (Svartström et al., 2017), panda (Zhu et al., 2011), reindeer (Pope et al., 2012), Saudi sheep (Al-Masaudi et al., 2017), termite (Warnecke et al., 2007), and wallaby (Pope et al., 2010) were also reannotated based on the latest version of the CAZy database and included in this calculation. Relative abundances of predicted plant (poly)saccharide-active CAZyme families (see Figure 4.4A) were then extracted for hierarchical clustering (correlation clustering and average linkage)

and principal component analysis (PCA) using R statistics in ClustVis (Metsalu and Vilo, 2015). Taxonomic assignment of predicted CAZymes from each metagenome was determined using protein sequences belonging to archaea, bacteria, fungi, other microbial eukaryotes and viruses reported in the National Center for Biotechnology Information non-redundant (NCBI-NR) database (downloaded May 16, 2017) using Kaiju in greedy mode with default settings (Menzel et al., 2016). The phylogenetic distributions of the top ten identified organisms were visualized at phylum and class levels in a chord diagram using Circos (Krzywinski et al., 2009).

Cellulosomal modules (dockerin and cohesin domains) were identified using reference sequences and models built from the literature (Mesnage et al., 2000; Venditto et al., 2016; Artzi et al., 2017). Proteins with CBMs appended to > 200 amino acids not covered by any CAZyme module were subjected to Pfam domain annotation (Finn et al., 2016) using Interproscan (Jones et al., 2014) to identify conserved domains of unknown function (DUF, Goodacre et al., 2014). PULs were predicted around *susCD*-like genes, and boundaries were extended based on intergenic distances, the presence of CAZymes and of regulatory genes (e.g. hybrid two-component system protein, ECF σ/anti-σ factors, etc.) following the automatic method used in PULDB (Terrapon et al., 2015). The proteins with unknown functions from the PULs reported here and in PULDB (http://www.cazy.org/PULDB/) were pooled and submitted to CD-HIT web server (Huang et al., 2010) to identify proteins that meet a similarity threshold, which was defined by being ≥ 70% identical to the representative sequence, and having ≥ 70% alignment coverage.

The sequence reads were submitted to NCBI under the BioprojectID PRJNA298982.

4.3 Results and Discussion

4.3.1 Comparison of predicted CAZyme sequences with existing datasets

Each metagenomic DNA sample (Table S4.1) yielded 71 to 88 Mb high quality reads (150 bp long each), which were assembled into assemblies between the size of 58.3 to 81.5 Mbp per metagenome (Table 4.1). Correlated with the assembly size, which is a reflection of both the number and lengths of distinct contigs, the number of ORFs ranged between 56 K and 81 K in the following order: pretreated poplar-fed microcosm from beaver droppings, cellulose-fed microcosm from beaver dropping, pretreated poplar-fed microcosm from moose rumen, and cellulose-fed microcosm from moose rumen. In total, 9,386 genes encoding carbohydrate-active enzymes and CBMs were predicted from the four metagenomes (Table S4.2). These sequences were assigned to 100 distinct families of GHs, 13 families of CEs, 15 families of PLs and 39 families of GTs, as well as 43 families of associated CBMs. As observed for other anaerobic microbial communities (Al-Masaudi et al., 2017; Svartström et al., 2017), no auxiliary redox enzymes were identified.

Table 4.1 Statistics of the sequencing and assembly of the metagenomes of cellulose- and pretreated poplar-fed enrichment cultures from beaver dropping and moose rumen

Inocula	*Beaver droppings*		*Moose rumen*	
Enrichment	Cellulose	Pretreated poplar	Cellulose	Pretreated poplar
Abbreviation	BD-C	BD-P	MR-C	MR-P
Number of quality trimmed reads (% passed)	74,999,337 (99.7%)	78,144,385 (99.7%)	71,980,296 (99.8%)	88,305,224 (99.8%)
Total Mbp assembled (Mbp)	78	81.5	58.3	67.6
Number of contigs	5,010	10,553	5,705	6,941
Longest contig (bp)	1,266,318	674,797	1,299,156	1,131,879
N50 (bp)	92,758	68,167	71,246	106,046
Number of ORFs	71,348	81,969	56,127	66,970
Normalized counts[1] of plant polysaccharide-active CAZymes (% of all CAZymes)	709 (23.4%)	830 (32.3%)	525 (22.4%)	834 (32.7%)
PUL	38	171	47	42
type 1 dockerin	95	116	68	94
type 1 cohesin	42	50	35	41

[1] Counts were normalized by the number of ORFs

On average, 13% (up to 17% in cellulose-fed moose rumen culture) of the annotated CAZymes were taxonomically unassigned or assigned to unknown species. The phylogenetic origins of the remaining varied among the metagenomes, with most of them derived from *Firmicutes, Bacteroidetes, Chloroflexi* and *Proteobacteria* phyla (Figure 4.1). Up to 14 % of predicted CAZyme sequences from cellulose-fed enrichments were assigned to class *Anaerolineae*, whereas CAZyme assignments to class *Gammaproteobacteria* were unique to cultures fed on pretreated poplar (Figure 4.1; Figure S4.1). By contrast, members from *Clostridia* and *Bacteroidia* classes contributed to 23% to 52% of the annotated CAZymes across the metagenomes, where representation by these classes was between two and five times higher in pretreated poplar-fed cultures than in cellulose-fed ones. Moreover, CAZyme families that comprise plant polysaccharide-active enzymes (i.e., families GH2, GH3, GH5, GH9, GH43, GH51, and CE1) were most frequently assigned to either *Clostridia* or *Bacteroidia* (Figure S4.1). In particular, GH5 sequences were most frequently assigned to *Clostridia* in pretreated poplar-fed cultures, whereas GH2, GH3 and GH43 were most frequently assigned to class *Bacteroidia* in pretreated poplar-fed cultures.

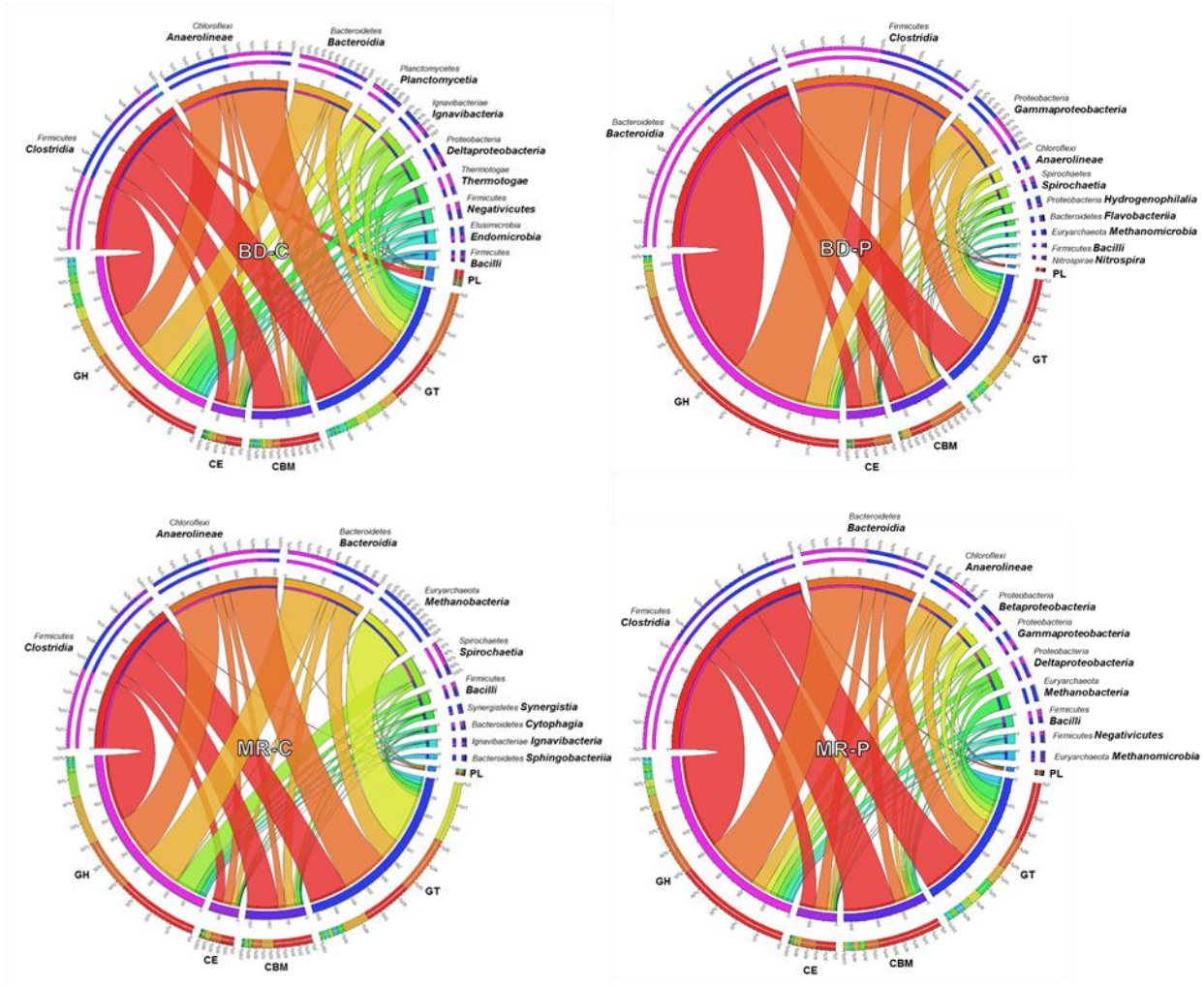


Figure 4.1 Phylogenetic distribution of CAZyme sequences assigned to the top ten identified classes

Considering all plant polysaccharides-degrading CAZymes predicted in each metagenome, more than half were less than 60% identical to the CAZyme amino acid sequences reported in the CAZy database (Figure 4.2). While percent identities varied depending on CAZyme family (Figure S4.2), the most divergent sequences belong to families GH113 (on average 34% identical to their closest blast hits), PL22 (35%), GH5 (40%), GH74 (40%), and PL9 (43%).

Hierarchical clustering analysis of plant polysaccharide-active CAZyme families distinguished those reported in this study from those previously predicted from grass-feeding mammals or mixed plants foragers (Figure 4.3A). In particular, the distribution of the CAZyme families predicted in moose rumen enrichments differed from that recently reported for the moose rumen metagenome (Svartström et al., 2017), where highest contributing factors were attributed to relatively high abundance of CE4, GH94 and GH78 families, and low abundance of GH2, GH43 in the moose rumen enrichments (Table S4.3). Consistent with substrate-induced convergence reported in Chapter 3 (Wong et al., 2016), long-term *ex situ* enrichment prior to metagenome sequencing also led to higher similarity of CAZyme distributions for cultures fed with the same carbon source (i.e., pretreated poplar or cellulose) as opposed to originating from the same environmental source (i.e., moose rumen or beaver droppings) (Figure 4.3A). The observed substrate-driven convergence of metagenomes was mostly attributed to higher relative abundances of GH2, GH3, GH43, CE1, CE4 in cultures enriched on pretreated poplar (Table S4.3). At the same time, a greater overlap of unique CAZyme sequences was observed between cultures fed with the same substrate than those with the same inoculum (Figure 4.3B). It is also worthwhile to note that the plant polysaccharide-active CAZyme families from termite gut, albeit wood-feeding, do not cluster closely with those from the pretreated poplar enrichments due to the latter's lower relative abundances of GH5, GH10 and GH94 (Table S4.3). This likely reflects differences in the wood substrates consumed, as well as intrinsic differences in the gut microbiome of mammals and insects. Nonetheless, along PC2 where the metagenomes diverge the most, the PCA plot depicted a closer resemblance of CAZyme profiles microbiomes from termite gut and moose rumen samples enriched on pretreated poplar, than that between the former and the non-enriched moose rumen.

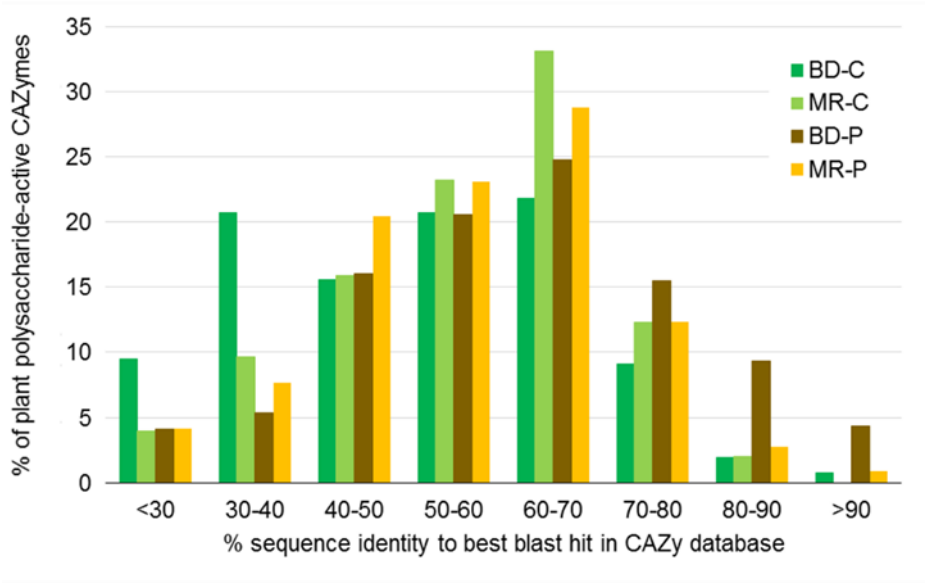
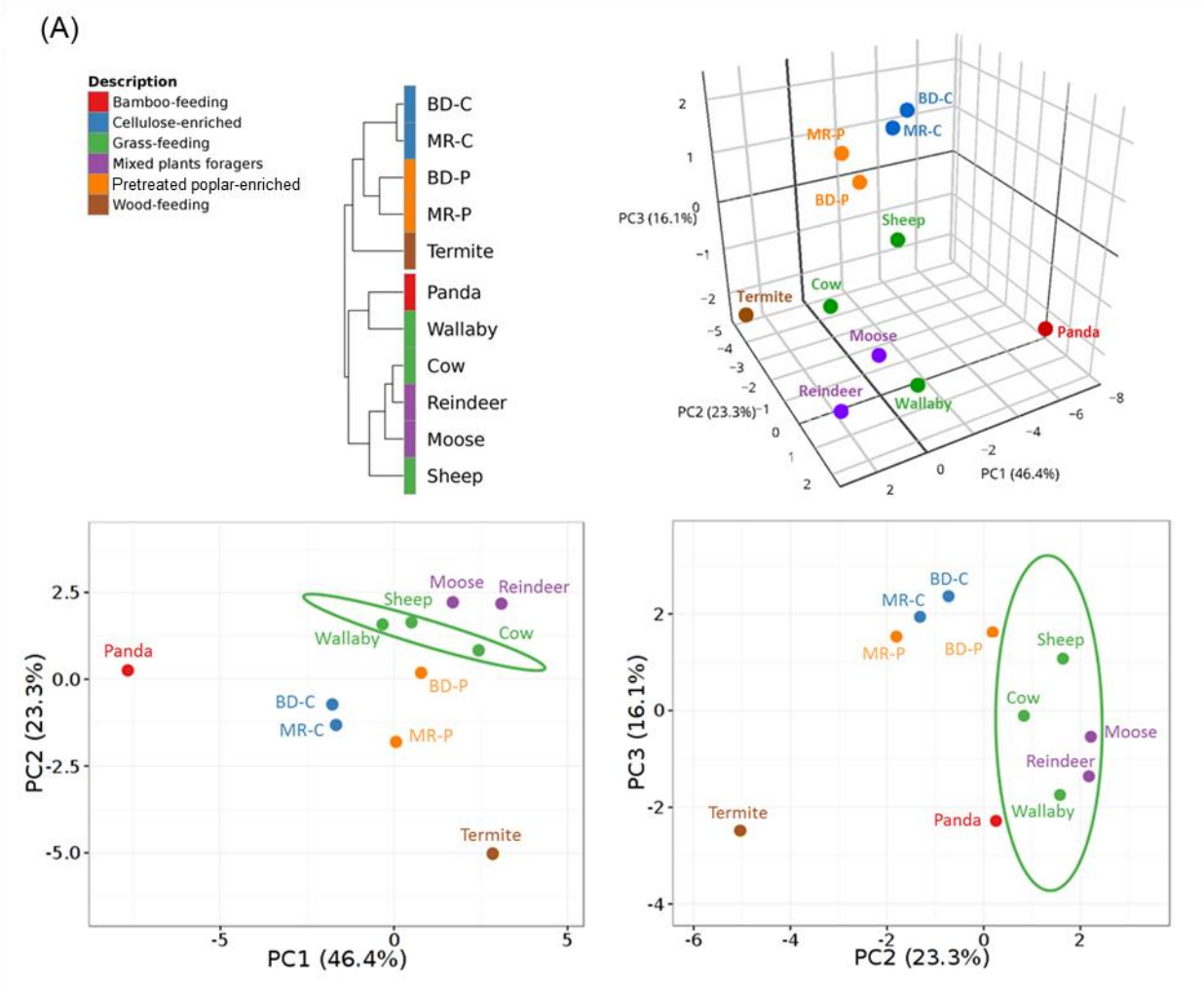


Figure 4.2 Percent identity between amino acid sequences in the CAZy database and CAZyme sequences predicted in beaver dropping (BD) and moose rumen (MR) microcosms enriched with cellulose (C) and pretreated poplar (P). Percent identities correspond to best blast hits in the CAZy database, and were obtained for CAZyme sequences belonging to CAZyme families known to contain enzymes that act on plant cell wall carbohydrates (see Figure 4.4A).


(A)
Description
Bamboo-feeding
Cellulose-enriched
Grass-feeding
Mixed plants foragers
Pretreated poplar-enriched
Wood-feeding
BD-C
MR-C
BD-P
MR-P
Termite
Panda
Wallaby
Cow
Reindeer
Moose
Sheep
PC3 (16.1%)
PC2 (23.3%)
PC1 (46.4%)
PC1 (46.4%)
PC2 (23.3%)
PC3 (16.1%)
PC2 (23.3%)

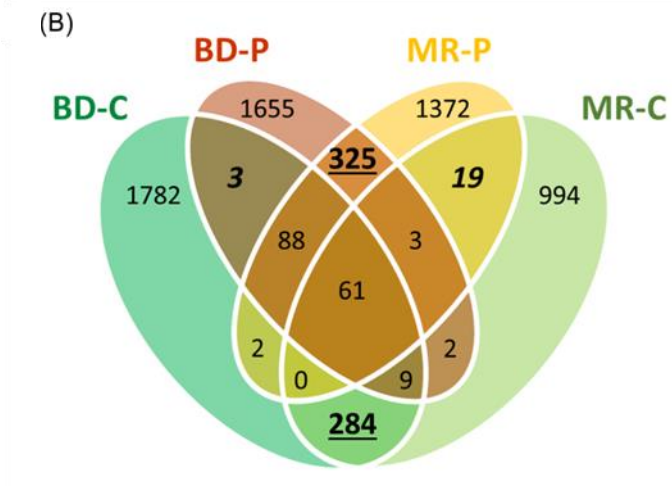
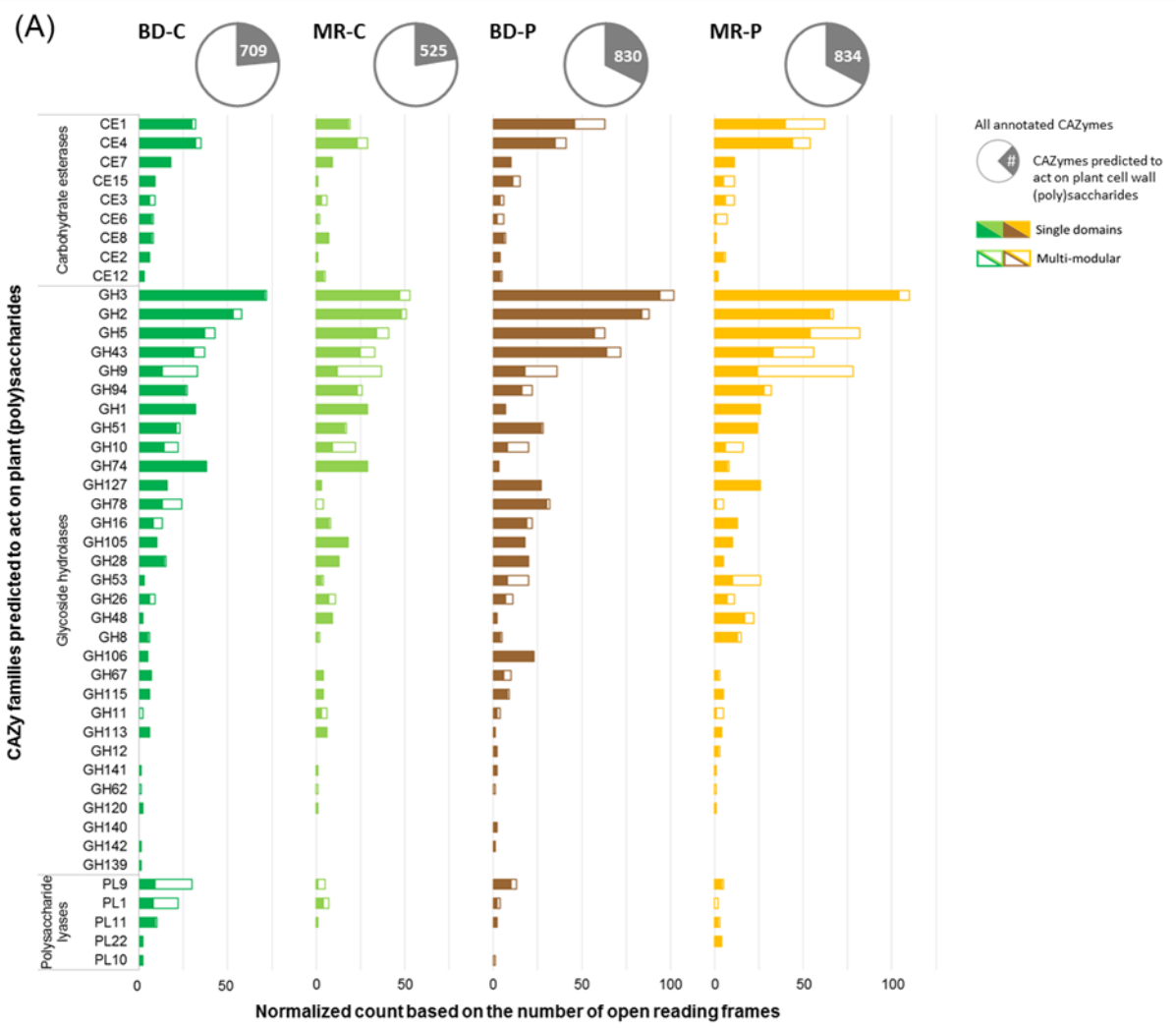

(B)
BD-P
MR-P
BD-C
MR-C
1655
1372
1782
994
325
3
19
88
3
61
2
2
0
9
284

Figure 4.3 (A) Correlation clustering and PCA plots of CAZyme profiles encoded by metagenomes from lignocellulose degrading microbial communities. CAZyme families known to contain enzymes that act on plant cell wall carbohydrates were considered in the analysis (see Figure 4.4A). Public datasets included cow (Hess et al., 2011), moose (Svartström et al., 2017), panda (Zhu et al., 2011), reindeer (Pope et al., 2012), Saudi sheep (Al-Masaudi et al., 2017), termite (Warnecke et al., 2007), and wallaby (Pope et al., 2010). A 3D PCA plot is shown on the top right corner with the corresponding 2D PCA plots shown at the bottom; confidence intervals (95%) are indicated by the ellipses. **(B) Venn diagram showing a greater overlap of unique CAZyme sequences in cultures fed with the same substrates (numbers underlined) than those that originate from the same inocula (numbers in italics)**

4.3.2 Impact of enrichment substrate on profiles of predicted plant-polysaccharide degrading CAZyme sequences

Metagenomes of cultures enriched on pretreated poplar yielded a higher proportion of predicted plant polysaccharides-active CAZymes (~33%) compared to metagenomes of cultures enriched on cellulose (~23%) (Table 4.1, Figure 4.4A). In particular, pretreated poplar-degrading communities were enriched in sequence counts from families GH3, GH5, GH43, CE1 and GH53 (Figure 4.4A). Additional substrate-induced differences were noted when considering moose rumen and beaver dropping enrichments separately. Beaver dropping samples enriched on pretreated poplar encoded higher counts of GH2 (1.5 times higher) and GH106 (4.6 times) than corresponding samples enriched on cellulose (Figure 4.4B). Meanwhile, moose rumen samples enriched on pretreated poplar encoded higher counts of GH9 (2 times higher), CE4 (1.8 times),

GH127 (9 times), and CE15 (11 times) compared to corresponding samples enriched on cellulose (Figure 4.4B).



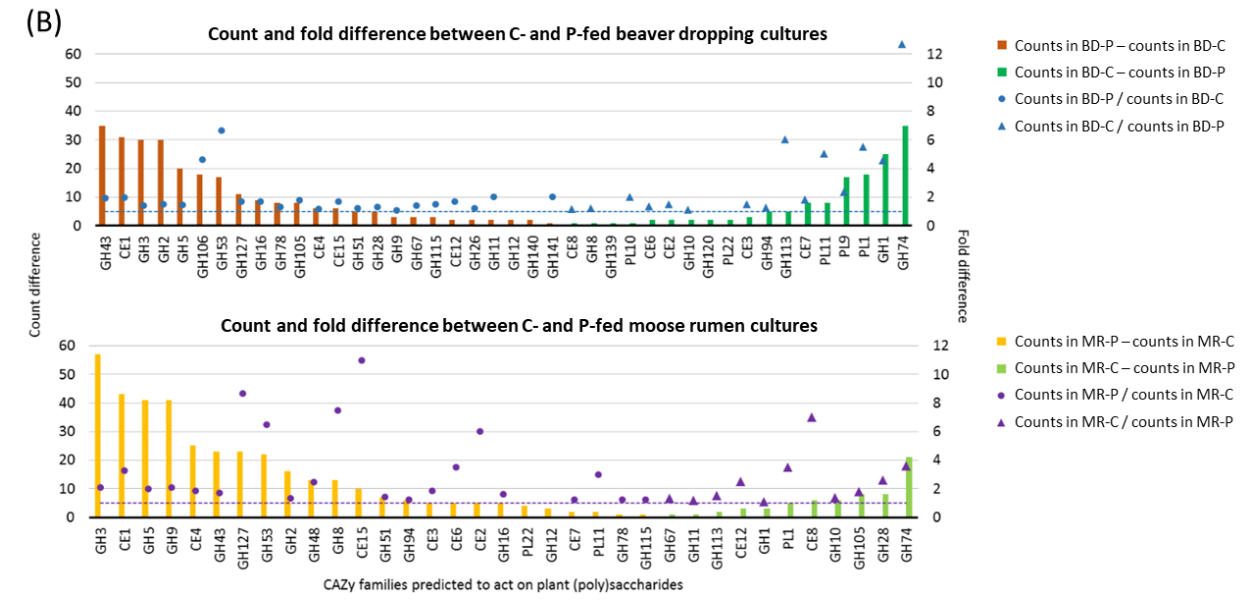


Figure 4.4 (A) Distribution of plant (poly)saccharide degrading-CAZyme families as single and multi-modular domains. (B) Normalized count and fold difference of CAZyme families predicted to act on plant polysaccharides between pretreated poplar (P)- and cellulose (C)-fed cultures. Fold difference was only calculated for non-zero counts.

CAZyme families that were enriched through growth on pretreated poplar included those that comprise enzymes involved in plant polysaccharide deconstruction. For example, family GH43 includes enzymes that target arabinoxylan (Borsenberger et al., 2014; Mewis et al., 2016), family CE1 members were shown to deacetylate polymeric xylans (Neumuller et al., 2015; Mai-Gisondi et al., 2017), and family GH5 members include endoxylanases that targets xylans with or without methyl-glucuronic acid side chain (Gallardo et al., 2010), as well as enzymes that target cellulose and mannans (Aspeborg et al., 2012). Notably, enzymes belonging to families GH2 and GH3 were also abundant in the moose rumen microbiome, and predicted to participate in plant cell wall deconstruction (Svartström et al., 2017). Characterized CE15 members display 4-*O*-methyl-glucuronoyl methylesterase activity, which is thought to hydrolyse ester linkages that may

form between hydroxyl groups in lignin and 4-*O*-methyl-D-glucuronic acid residues in glucuronoxylans that dominate in hardwood fibre (Biely et al., 2015; Biely, 2016). Recently, a marine bacterial CE15 enzyme predicted to act on alginates was also reported, suggesting a broader substrate range for this CE family (De Santi et al., 2016; Agger et al., 2017). On the other hand, GH127 enzymes typically contain β-L-arabinofuranosidases that have been shown to target plant cell wall glycoproteins, such as extensin (Fujita et al., 2011). By contrast, between 3 and 13 times more GH74 sequences were identified in cellulose-fed enrichments compared to corresponding cultures enriched on pretreated poplar, fitting with the endoglucanase activity reported for this CAZyme family (Song et al., 2017). Similarly, 4.6 times more GH1 sequences were identified in the beaver dropping culture enriched on cellulose than that on pretreated poplar; characterized bacteria members from this family large act as β-glucosidases that hydrolyze cellobiose and soluble cellodextrins to glucose (Singhania et al., 2013). Interestingly, families capable of pectin degradation (PL1, PL9) were also found at higher abundances in cellulose-fed beaver dropping culture than that fed with pretreated poplar.

CBMs can impact enzyme performance through targeting catalytic modules to polysaccharide substrates, and in some cases promote non-hydrolytic fibre disruption (Boraston et al., 2004; Gourlay et al., 2012); accordingly, CAZymes with cognate CBMs were also predicted from each metagenome sequence. About 20% of the sequences predicted to encode plant polysaccharides degrading enzymes (i.e., 669 sequences) were predicted to form multi-domain proteins (Table S4.4). Most frequent domain organizations included CBMs, such as CBM48-GH13_9 (7-8% in cellulose enrichments), GH9-CBM3-CBM3 (~6% in moose rumen samples enriched on pretreated poplar), and CBM50-CBM50-GH18 (~6% in beaver dropping and moose rumen samples enriched on pretreated poplar and cellulose, respectively). While CBM48-GH13

is a documented architecture for starch-degrading enzymes (Machovic and Janecek, 2008), the modular architecture GH9-CBM3-CBM3 was previously only reported as a non-cellulosomal enzyme encoded by *Clostridium thermocellum* (Anitori, 2012). CBM50-CBM50-GH18, like other GH18 chitinases with multiple CBM50 domains, was predicted to bind peptidoglycan-like and chitin-derived oligosaccharides (Bateman and Bycroft, 2000). Contrary to findings recently reported for the moose rumen microbiome (Svartström et al., 2017), the multi-modular enzymes comprising CBM50 and GH23 or GH73 were identified at low abundance (<3.5%) in the enriched moose rumen metagenomes reported herein.

In addition to multi-modular CAZymes comprising CBMs, those comprising potential cellulosomal subunits were also predicted. Cellulosomes are cell-associated multi-enzyme complexes that are produced by certain anaerobic bacteria to promote polysaccharide degradation (White et al., 2014; Artzi et al., 2017; Smith et al., 2017). For all enrichments, approximately twice the number of dockerins compared to cohesins were predicted (Table 4.1), and 56% of dockerins were appended to CAZyme sequences (Table S4.5). The most frequently occurring CAZyme module was family GH9 (~29%), followed by GH5 (~11%), CE3 (~8%), GH43 (~7%), and GH3 (~6%) (Figure 4.5, Table S4.4). And unlike the recent study on moose rumen metagenome, the recurrent GH13 appended-dockerins (Svartström et al., 2017) were not identified in the current moose rumen metagenomes, likely due to their long enrichment on cellulosic carbon sources. Other common components of cellulosome systems (Artzi et al., 2017), such as GH10 (~4%), GH11 (~4%) and GH48 (~4%) were also identified in the metagenomes of both moose rumen and beaver dropping enrichments, albeit at a lower abundance. Notably, the identification of few sequences containing a GH48 module and the ~9-fold higher number of those containing a GH9 module is consistent with earlier analyses of cellulose-degrading anaerobic

bacteria that generate high levels of a single GH48 and diverse GH9 enzymes with potential synergistic action (Morag et al., 1991; Ravachol et al., 2014; Artzi et al., 2015).

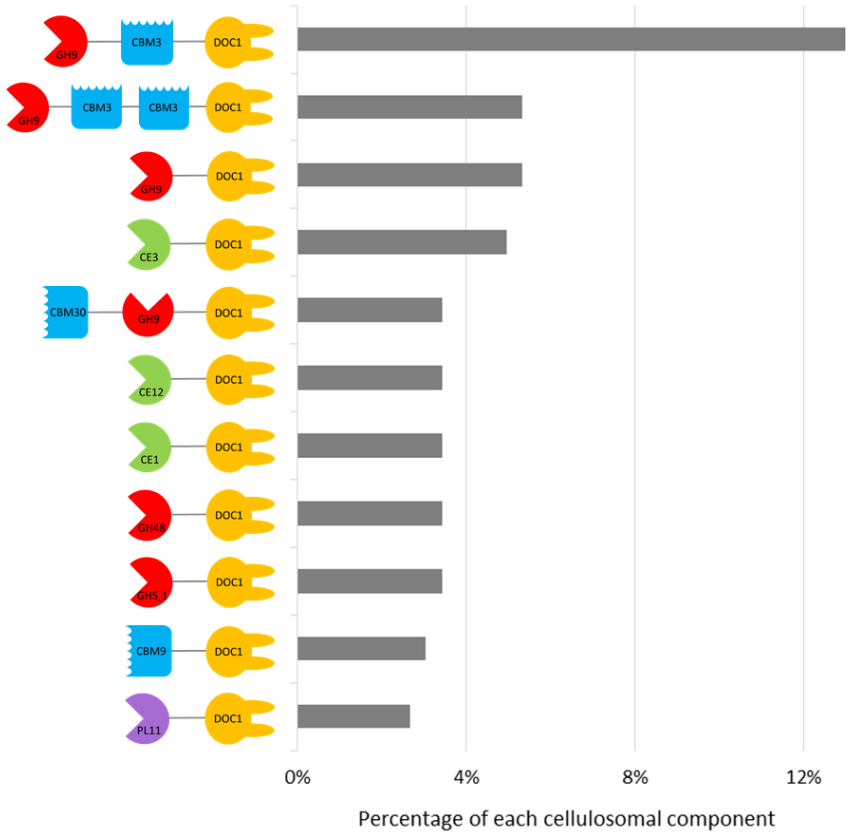
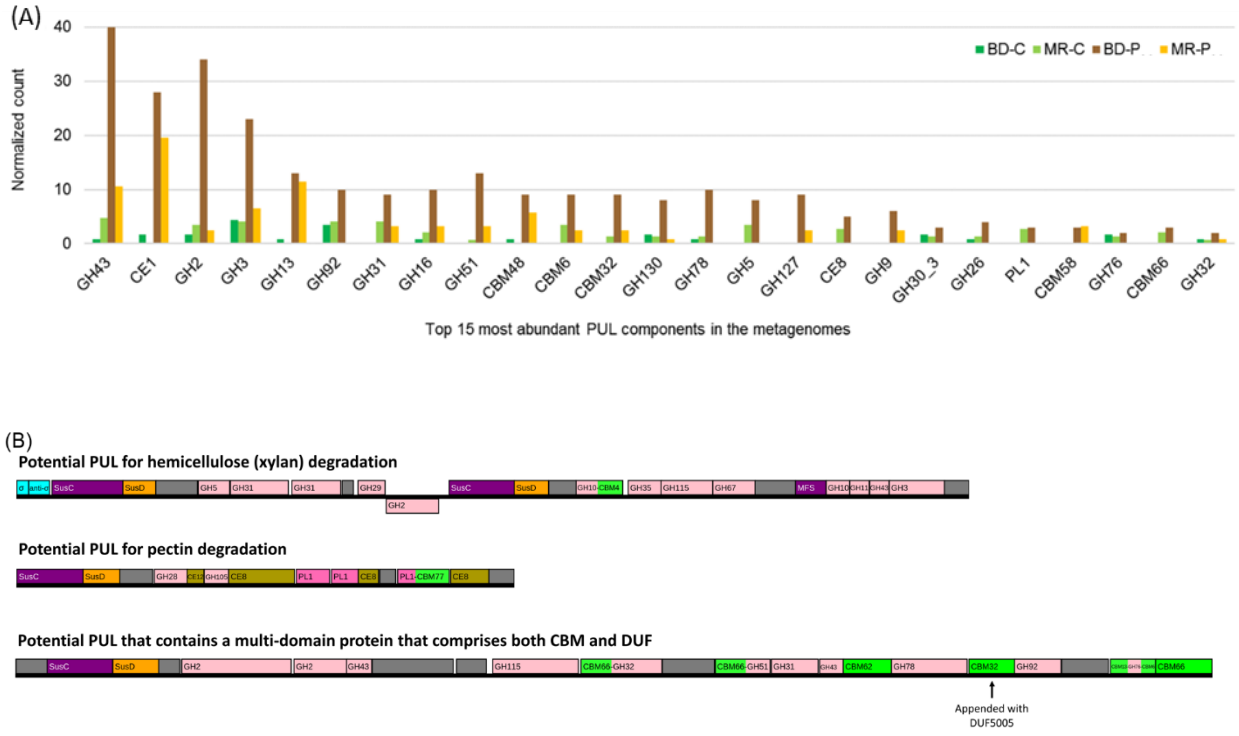


Figure 4.5 Catalogue of domain architectures of top 11 abundant CAZy-dockerins in the cellulose- and pretreated poplar-fed microbial enrichments from beaver dropping and moose rumen. Type 1 dockerin: DOC1.

4.3.3 Predicted polysaccharide utilization loci (PULs)

As summarized above, PULs have emerged as especially fruitful regions within genomic sequences for enzyme discovery (Larsbrink et al., 2016). Herein, 416 PULs were predicted

(Figure S4.3), where the normalized number predicted from pretreated poplar-fed microcosms of beaver droppings was 4.5 times that predicted from cellulose-fed microcosms. Consistent with the overall distribution of predicted CAZyme sequences, those belonging to families CE1, GH3, and GH43, were most frequently identified in the predicted PULs (Figure 4.6A, Table S4.6). Moreover, PULs comprising members of families GH127 and GH9 were exclusively identified in metagenomes from cultures enriched on pretreated poplar. Based on the family composition of a given PUL, substrate category of the PUL-encoded enzymes can be inferred. For example, Figure 4.6B illustrates PULs that potentially target xylan and pectin based on the established activities of the CAZyme families. Sequences annotated as unknown may include novel enzyme functions, for instance as shown recently in the case of the type II rhamnogalacturonan PUL of *B. thetaiotaomicron* (Ndeh et al., 2017).



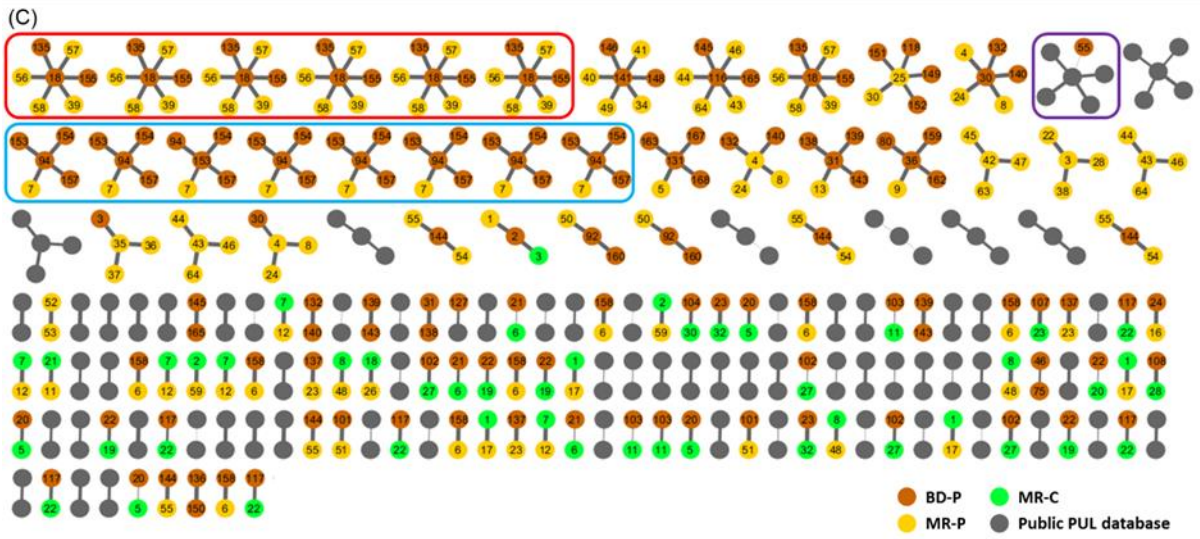


Figure 4.6 (A) Top 15 most abundant CAZyme families identified in predicted PULs from cellulose (C)- and pretreated poplar (P)-fed microbial enrichments of beaver droppings (BD) and moose rumen (MR). (B) Examples of predicted polysaccharide utilization loci (PULs) from beaver droppings enriched on pretreated poplar. (C) Similarity-based clustering (≥ 70%) of proteins with unknown function positioned in PULs identified herein and listed in the PUL database (http://www.cazy.org/PULDB/). Each cluster contains a central node that denotes the representative protein with unknown function (defined by the longest length) and connected nodes that represent a protein with unknown function that is ≥ 70% identical to the representative sequence. PUL identifiers are shown on each node; the thickness of the edges correlates to percent identity between sequences. Circled in red and blue are proteins with unknown functions that are ≥ 95% identical to one another; the architecture of PULs circled in red is identical, whereas those circled in blue share a common central architecture but differ at flanking regions. Circled in purple is the only cluster that comprises proteins with unknown function from both PULs predicted herein and those from the public PUL database.

Of note, 612 sequences annotated as non-CAZyme proteins with unknown function (with lengths ranging from 32 to 1320 amino acids) were identified in all candidate PULs (Figure S4.4A). In an effort to prioritize additional sequences for future characterization, a clustering network diagram was generated to uncover protein sequences with unknown function that reoccurred in the predicted PULs. However, little similarity was revealed between such sequences from PULs predicted herein and those reported in the public PULDB (Figure S4.4B). In fact, only one such sequence from beaver droppings enriched on pretreated poplar was ≥ 70% identical to those annotated in the PULDB, and the few sequences with unknown function that did cluster typically originated from PULs with similar architecture (Figure 4.6C; Figure S4.3). Further work pertaining the bioinformatics analysis and biochemical characterization of these proteins with unknown function within the PULs is reported in the thesis by Razeq (2017).

4.3.4 Predicted multi-modular proteins – an additional source of yet unknown carbohydrate-active proteins

A second approach to assist the discovery of potentially new CAZyme families considered multi-modular proteins predicted to comprise a DUF, appended to a CBM or to dockerins (cellulosomal subunit).

Considering all four metagenome sequences reported herein, 62 DUFs were identified that co-occurred with a CBM (Figure 4.7; Figure S4.5). The most frequent organizations were: DUF3794-CBM50 (10 identified), DUF362-CBM9-DOC1 (4 identified), DUF4366-CBM16 (4 identified), which were identified in all four metagenomes; five DUF3459-CBM48-GH13_10

sequences were also identified in the metagenome of beaver droppings enriched on pretreated poplar (Figure S4.5). As shown herein and also described in the Pfam database (Finn et al., 2016), DUF3794 was often found in association with CBM50. On the other hand, DUF362 is often present in proteins with domains that bind to iron-sulphur clusters, and its coexistence with CBM9_1 in an uncharacterized protein from soil bacteria *Sorangium cellulosum* was previously observed (UniProt entry S4XJL8). The structure of DUF3459 has been determined (UniProt entries B2IUW9, Q9RX51, Q8ZPF0, Q8P5I2, Q2PS28, M1E1F6, M1E1F3, H3K096), and as observed here, was previously shown to be part of multi-modular proteins comprising GH13 and CBM48 domains (UniProt entries W6LS46, R4KHQ4, C7RTS8). Although not frequently observed, one PUL identified in the pretreated poplar-fed beaver dropping culture contained DUF5005 appended to a predicted CBM32 (Figure 4.6B).

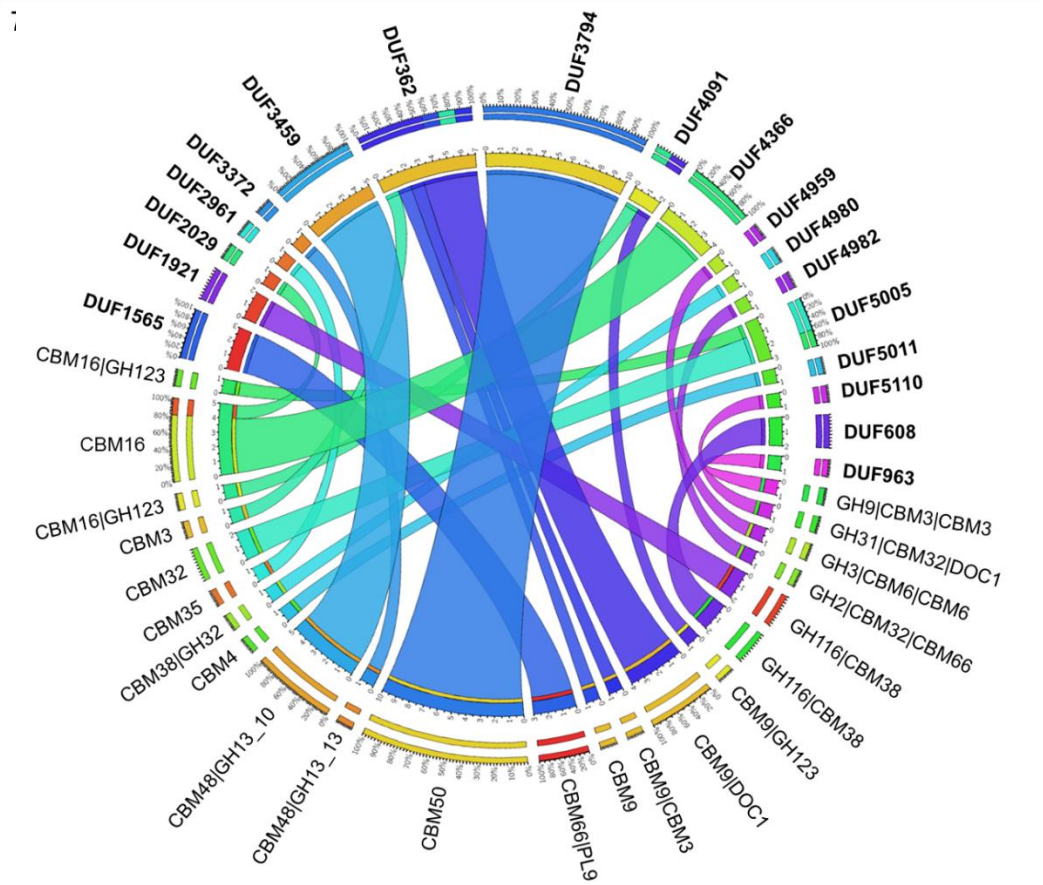


Figure 4.7 Carbohydrate-active proteins with domains of unknown functions identified in the metagenomes. In bold are the DUFs that are identified with various CBMs and CAZymes as shown via the connections colour-coded in accordance to the DUFs. The number and corresponding percentage of each combination of modules are shown in the outer scale

Of the predicted dockerin sequences, 44% lacked known appended CAZyme modules. Similar to previous reports (Finn et al., 2016), a few dockerin sequences were predicted to have appended domains of unknown function (i.e., DUF362, DUF1533 and DUF3237); however, the majority were not annotated as containing modules or domains functionally attributed to cellulosomes. In many cases, this could reflect sequence gaps due to incomplete metagenome assembly; however, it is also conceivable that dockerin-cohesin proteins might in fact participate

in other biological functions, as suggested by a phylogenetically distinct group of cohesins discovered in the cow rumen metagenome (Bensoussan et al., 2017).

4.4 Conclusion

Given their natural dietary habits, the Canadian beaver and North American moose have likely evolved digestive microbiomes with the ability to degrade diverse wood polysaccharides. To identify enzymes that may be most relevant to wood fibre bioprocessing, corresponding microbial communities were enriched for approximately 3 years on comparatively complex (pretreated poplar) and defined (microcrystalline cellulose) carbon sources. The enrichment led to a substrate-induced convergence of CAZyme profiles that are also distinguished from those previously predicted from grass-feeding mammals (including cow rumen) or mixed plants foragers; when compared, the pretreated poplar enrichments were particularly enriched in GH2, GH3, GH43, CE1, CE4. On the other hand, the CAZyme profile of the pretreated poplar enrichments do not cluster closely with that of termite gut due to a lower relative abundances of GH5, GH10 and GH94. Moving on, pairwise-comparison of CAZyme profiles between microcosms fed with different substrates significantly narrowed the number of CAZyme families and corresponding members that could be targeted for improved enzymatic conversion of wood fibre. For example, in addition to families GH2, GH3, GH5, and GH43 which previous reports have also identified, GH127 and CE15 may be especially relevant for the anaerobic conversion of pretreated wood fibre. Protein sequences containing both a CBM and a DUF, as well as proteins with unknown function positioned within PULs, were also identified and may facilitate the discovery of new CAZyme activities. Proteomic analysis of secretomes from the enrichment cultures prepared herein will provide an additional filter for protein selection and characterization. In the interim, however, enrichment followed by comparative metagenomics sufficiently narrowed protein lists of primary interest, enabling direct recombinant production and characterization.

Chapter 5 Uncovering the microbial dynamic and carbohydrate-active enzymes in the lignocellulose-degrading microbial community from pulp mill anaerobic granules over long-term enrichment

---

This chapter is prepared for submission as a journal article:

Wong, M.T., Wang, W., Lacourt, M., Couturier, M., Lombard, V., Lapebie, P., Edwards, E.A., Terrapon, N., Henrissat, B., and Master, E.R. Uncovering the microbial dynamic and carbohydrate-active enzymes in the lignocellulose-degrading microbial community from pulp mill anaerobic granules over long-term enrichment.

5.1 Introduction

Metagenomic analysis of lignocellulose-degrading microbial communities has dramatically increased the repository of CAZymes, which are considered a critical resource for cost-effective bioconversion of lignocellulosic biomass to valuable chemicals and energy. To date, attempts to identify lignocellulose-degrading enzymes through metagenomics analyses have focused predominantly on the grass-feeding gut microbiota (Hess et al., 2011). By comparison, few metagenomics analyses of wood-degrading microbial communities have been reported. Metagenomic studies that focus on wood transformation include analysis of the gut microbiome of termites (Warnecke et al., 2007), wood-feeding beetle (Scully et al., 2013), wood wasp (Adams et al., 2011) and moose (Svartstrom et al., 2017; Svartström et al., 2017; Wong et al., 2017), which identified abundant CAZymes belonging to GH1, GH3, GH5, GH13, GH43, and more.

Aside from the wood-feeding gut microbiomes, engineered bioreactors designed for lignocellulose degradation such as anaerobic digesters at pulp and paper mills represent promising sites for novel wood-degrading CAZymes and other carbohydrate-active proteins. The pulp and paper industry is amongst the world's biggest industrial wastewater producers, generating 42% of the industrial wastewater in the United States and Europe combined and amounting for a total of 3 billion m^3 wastewater globally (Ashrafi et al., 2015; Toczylowska-Maminska, 2017). Conventionally, the wastewater is first treated with physiochemical methods to remove organic matter and suspended solids, and then biologically treated through methods such as the aerobic activated sludge process (Thompson et al., 2001). The generated biosludge (a.k.a. secondary sludge) is then dewatered and incinerated for power generation or else landfilled. Increasingly, anaerobic bioconversion of pulp mill effluent prior to aerobic treatment is considered a means to reduce the accumulation of secondary biosludge while also generating biogas as a local and renewable source of energy (Meyer and Edwards, 2015). Today, about 10% of pulp and paper mills worldwide have installed anaerobic treatment technologies (Meyer and Edwards, 2015), and most of these systems were established since 2000 (Haddaway, 2014). Among the anaerobic bioreactors applied for treating pulp and paper wastewaters, internal circulation reactors are widely employed due to the high removal efficiencies at low operational cost (Meyer and Edwards, 2015). These high-rate reactors feature a two-stage separation of biogas with an internal circulation of effluent (Tauseef et al., 2013), thereby enabling effective bioconversion of organic compounds into biogas via the sequential activity of hydrolytic bacteria, acetogenic bacteria, and methanogenic archaea (Kamali et al., 2016). Given these required activities, pulp mill anaerobic bioreactors likely establish microbial communities encoding enzymes that transform wood-derived organic compounds (e.g., cellulose, hemicelluloses and lignin) while withstanding

common inhibitors to anaerobic digestion (e.g., resin acids, sulphur and organochlorine compounds) (Hagelqvist, 2013; Meyer and Edwards, 2015).

Driven by the interest to discover enzymes with novel activities and environmental resistance for industrial applications, researchers have begun to study the microbiomes associated with pulp and paper mills. The production of extracellular cellulases and xylanases by fungal isolates from paper mill sludge was reported as early as 1987 (Royer and Nakas, 1987); more recent analyses have isolated strains belonging to phyla *Firmicutes, Proteobacteria* and *Actinobacteria* with lignocellulolytic activities from different paper mill sludges, organic deposits around paper mill machines, as well as pulp paper mill effluent polluted site (Maki et al., 2011; Chandra and Singh, 2012; Ghribi et al., 2016). In addition, *Methanosarcina* species were found to be dominant methanogens of batch anaerobic digesters that processed pulp and paper waste (Walter et al., 2016). Functional screening of metagenomic clones also identified active GH5 β-1,4-endoglucanases from a *Bacillus* isolate from a field that was irrigated with pulp and paper mill effluents (Pandey et al., 2014; Pandey et al., 2016).

On top of screening microbial consortia from lignocellulose-degrading habitats, substrate enrichment is also applied as an effective strategy to promote the prevalence of biomass degraders and genes encoding lignocellulose-degrading activities (Cretoiu et al., 2012; Jimenez et al., 2014b). For instance, successive and anoxic sub-cultivation of pulp mill activated sludge on rice straw enriched genera *Clostridiales*, *Bacillales*, and *Bacteroidales* (Wongwilaiwalin et al., 2013). Nonetheless, the extent of the impact depends on the microbial source. For example, when establishing enrichment cultures on switchgrass and eucalyptus, Simmons et al. (2014) concluded that the effect of different compost inocula on the community structure and microbial activity was insignificant compared to other culture variables (e.g. feedstock, thermophilic, high-solids

condition). Meanwhile, distinctive microbial consortia with similar lignocellulose degradation capacity were reported in microbial consortia generated from forest soil, canal sediment and decaying wood (Cortes-Tolalpa et al., 2016). In our previous study, wherein digestive microbiomes from Canadian beaver and North American moose were enriched for 3 years upon a variety of wood lignocelluloses, the substrate-induced convergence of taxonomic and functional (in terms of plant polysaccharide-degrading CAZymes) profiles were observed (Wong et al., 2016; Wong et al., 2017).

In this study, cellulose- and lignocellulose-fed microbial cultures from pulp mill anaerobic granules were established as described in Wong et al. (2016) to unveil the microbial members that were preferentially proliferated under various lignocellulose amendments. The profile of microbial community members encoded CAZymes, multi-domain proteins and PULs of cellulose- and pretreated poplar-fed microcosms were selectively explored using whole genome shotgun sequencing. Reminiscent of our previously established microcosms from digestive microbiomes of beaver and moose, an overall substrate-induced convergence of microbial communities from the microcosms was identified. When the microbial communities and (poly)saccharide-active CAZyme profiles of the cellulose- and pretreated poplar-fed cultures were compared together, those of the wood-consuming herbivores clustered in accordance to the enrichment while those from the anaerobic granules form distinguished clusters. Moreover, contrary to the microcosms from the gut microbiomes, the pretreated poplar-fed microcosm from anaerobic granules presented a higher abundance (compared to the cellulose-fed counterpart) in a broader range of CAZyme families predicted to act on pectin. Aside from providing more protein candidates for potentially new catalytic functions, this study highlighted the importance of exploring more microbiomes for novel lignocellulases.

5.2 Materials and Methods

Collection of pulp mill anaerobic granules. Pulp mill anaerobic granules were collected from an internal circulation anaerobic wastewater treatment reactor at a pulp and paper complex in Témiscaming, Québec. The reactor typically received 15000 m^3/day of mixed wastewater including acid condensate from the evaporator system and bleached chemi-thermomechanical pulp effluent.

Set up and maintenance of lignocellulose-degrading enrichment cultures from pulp mill anaerobic granules. Lignocellulose-degrading microorganisms enriched on i) microcrystalline cellulose (Avicel PH101, Sigma-Aldrich, MO, USA), ii) cellulose (Avicel) + lignosulphonate (Tembec Industries Inc., QC, CA), iii) cellulose (Avicel) + tannic acid (Sigma-Aldrich, MO, USA), iv) steam-exploded (200 ºC for 8 minutes) pretreated poplar (SunOpta Inc., Canada; October 2009) from pulp mill anaerobic granules were established and maintained via serial cultivation under anaerobic conditions at 36 °C as described in section 3.2 (Wong et al., 2016). Table S3.2 provides the details regarding the transfer methods, the amounts of lignocellulose substrates added to the microcosm bottles and duration across the enrichment phases. The transfer dates and length of the enrichment phases were summarized in table S5.1.

DNA extraction from pulp mill anaerobic granules and lignocellulose-degrading enrichments. Following 2 years of cultivation, comprising over six enrichment phases (Table S5.1), DNA was extracted from each culture at the early stationary phase of biogas production for 16S rRNA gene amplicon sequencing as described in section 3.2 and our publication (Wong et al., 2016). A similar round of DNA extraction was conducted after the eighth enrichment phase in the following year (Table S5.1); the DNA extracted was used for both 16S rRNA gene amplicon sequencing and metagenomic shotgun sequencing.

Pyrosequencing of 16S rRNA genes and metagenomic shotgun sequencing. The V6-8 hypervariable region of 16S rRNA genes was amplified with 926 Forward (5'-AAACTYAAAKGAATTGACGG) and 1392 Reverse (5'- ACGGGCGGTGTGTRC) primers and multiplexed with 10-nt Roche barcodes by polymerase chain reaction (Table S5.2) (DeAngelis et al., 2012). The reaction condition, as well as the purification and quantification of amplicons, were conducted on a 454 GS FLX platform (454 Life Sciences-a Roche Company, Branford, CT, USA) at Génome Québec Innovation Centre as described previously in section 3.2 and our publication (Wong et al., 2016). Multiplex-pyrosequencing was performed on a 454 GS FLX platform (454 Life Sciences-a Roche Company, Branford, CT, USA) at Génome Québec Innovation Centre. Illumina paired-end sequencing with TruSeq library was performed on Illumina HiSeq 2000 (Illumina Inc., San Diego, CA, USA) at Génome Québec Innovation Centre.

Analyses of 16S rRNA gene pyrosequences. Pyrosequencing output was converted to sequence reads and quality scores using Roche 454 Life Science propriety software (http://www.454.com) and then analyzed by QIIME 1.8.0 (Caporaso et al., 2010) as described previously in section 3.2. Combining the taxonomic profiles from our previous analysis on lignocellulose-degrading microcosms from beaver dropping and moose rumen (Wong et al., 2017), relative abundances of genera (that were ≥ 1% for at least one of the samples) were extracted for hierarchical clustering (correlation clustering and average linkage) and PCA using R statistics in ClustVis (Metsalu and Vilo, 2015). Non-parametric Kruskal-Wallis test was conducted using R script (R Development Core Team, 2010).

Metagenome assembly, annotation of CAZyme families, multi-modular sequences, and PULs. Quality trimming and assembly of metagenomics shotgun sequences, prediction of open reading frames, CAZyme families, catalytic cellulosomal subunits, proteins with CBMs and

DUFs, PULs as well as taxonomic assignment of the predicted CAZymes were performed and visualized in accordance to the methodology described earlier in section 4.2 and the corresponding publication (Wong et al., 2017). Since the number of predicted ORFs was higher for the metagenome of the pretreated poplar-fed enrichment (AG-P) than that of the cellulose-fed counterpart (Table S5.3), then:

$$Normalized\ count = \frac{count\ of\ identified\ sequences\ in\ a\ given\ CAZyme\ family\ within\ a\ metagenome\ of\ interest\ \times\ number\ of\ ORFs\ of\ AG-P\ metagenome}{number\ of\ predicted\ ORFs\ of\ a\ metagenome\ of\ interest}$$

Normalized count of predicted plant (poly)saccharide-active CAZyme families from our cellulose- and pretreated poplar-degrading microcosms from beaver dropping and moose rumen (Wong et al., 2017) were combined with the current dataset for hierarchical clustering (correlation clustering and average linkage) and PCA using R statistics in ClustVis (Metsalu and Vilo, 2015). Relative abundance for a given plant (poly)saccharide-active CAZyme family in a metagenome was calculated by

$$Relative\ abundance = \frac{identified\ counts\ for\ a\ given\ CAZYme\ family\ in\ a\ metagenome}{total\ identified\ counts\ of\ CAZymes\ in\ a\ metagenome} \times 100\%$$

Non-CAZyme sequences within PULs marked as having an unknown function that were identified herein and in our previous analysis of lignocellulose-degrading microcosms from beaver dropping and moose rumen (Wong et al., 2017) were pooled and submitted to CD-HIT web server (Huang et al., 2010). Proteins that are ≥ 70% identical and the alignment covers ≥ 70% of the entire amino acid chain were identified and the relationships were portrayed in a network visualized by Cytoscape 3.2.1 (Shannon et al., 2003).

5.3 Results and Discussion

5.3.1 Establishment of Biogas-Producing Microbial Enrichments

Anaerobic enrichments were established and biogas production was sustained over eight enrichment phases for 3 years on four different lignocellulosic substrate mixtures (Table S5.1). As observed in the beaver dropping and moose rumen, the biogas yield produced per amount of COD added dropped over the early enrichment phases (Figure 5.1). This trend could be explained by the presence of COD present in the inocula, which then was gradually reduced through microbial degradation and dilution over the transfers. The first set of DNA extraction occurred following the sixth enrichment phase at the second year, the extent of substrate conversion for the enrichments were: cellulose + tannic acid (0.46 ml biogas/mg COD added), cellulose (0.40 ml biogas/mg COD added), cellulose + lignosulphonate (0.34 ml biogas/mg COD added), and pretreated poplar (0.31 ml biogas/mg COD added). The subsequent increase in biogas yield in the later enrichment phase suggested a microbial acclimatization to the fed lignocellulosic carbon source (Figure 5.1). Accordingly, the extent of biogas yield was switched to the following order: cellulose (0.69 ml biogas/mg COD added), pretreated poplar (0.58 ml biogas/mg COD added), cellulose + lignosulphonate (0.41 ml biogas/mg COD added), and cellulose + tannic acid (0.15 ml biogas/mg COD added). The apparent inhibition of biogas production through amendment with tannic acid was also observed in the enrichment cultures from beaver droppings and moose rumen (Wong et al., 2016), which could be caused by the formation of complexes between phenolic components of tannic acid and enzymes, polysaccharides and metal ions (Goldstein and Swain, 1965; Makkar, 2003).

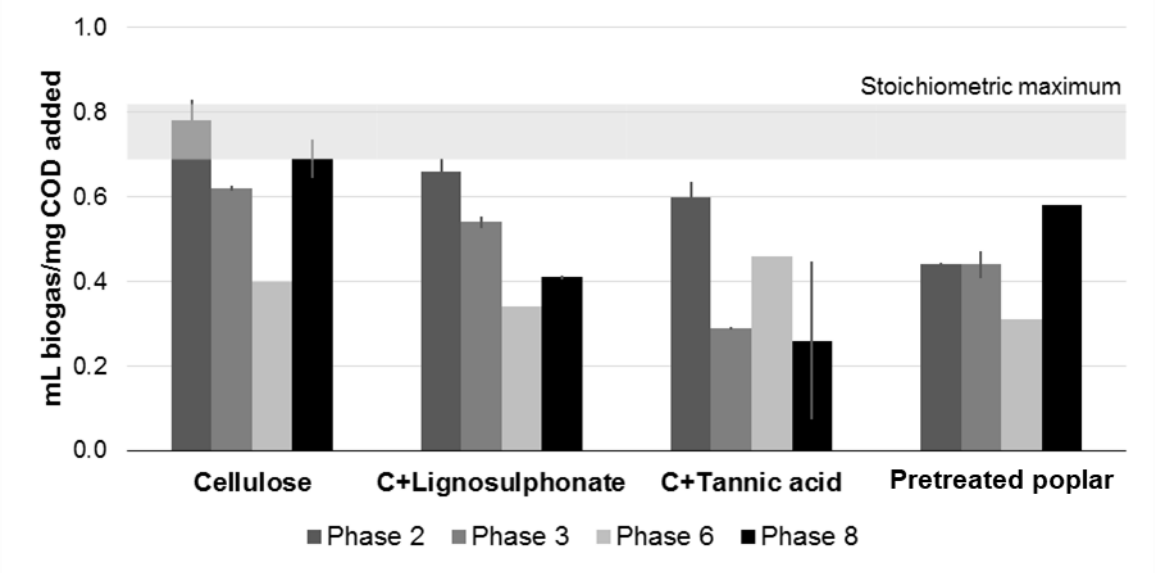


Figure 5.1 Biogas production profile of microcosms fed with various lignocellulosic carbon sources for 3 years. The range of stoichiometric maximum biogas yield is shown in the gray band to provide a reference for the conversion extents of the fed substrates in the microcosms (see supplemental method for the calculation based on Buswell's formula (Symons and Buswell, 1933; Wong et al., 2016)); error bars indicate standard deviation; $n = 3$.

5.3.2 Microbial diversity in lignocellulose-degrading microcosms enriched from pulp mill anaerobic granules

Overall, 144,347 high-quality 16S rRNA gene pyrosequences were assigned to 6,354 OTUs at 97% similarity threshold. (Table S5.2). Taxonomic assignment of OTUs was conducted to reveal the impact of the lignocellulose enrichment on the microbial composition of the pulp mill anaerobic granules and cultures. UPGMA clustering of the microbial communities revealed three main clades: i) inoculum, ii) microcosms fed with cellulose + tannic acid, and iii) remaining

microcosms that clustered based on the fed substrates and secondarily by the length of enrichment (except for those enriched with cellulose) (Figure 5.2). The clustering of biological replicates for a given enrichment condition reflected the reproducibility of the enrichment set up and 16S rRNA gene analysis. Moreover, the separation of inocula and microcosms fed with cellulose + tannic acid from all other microcosms was also observed in earlier analyses of gut microbiomes from beaver and moose (Wong et al., 2016). Similar to Chapter 3, the following description focuses on the microbial composition established within each sample and relative abundances are compared to infer key members that were adaptive towards the enrichment substrate.

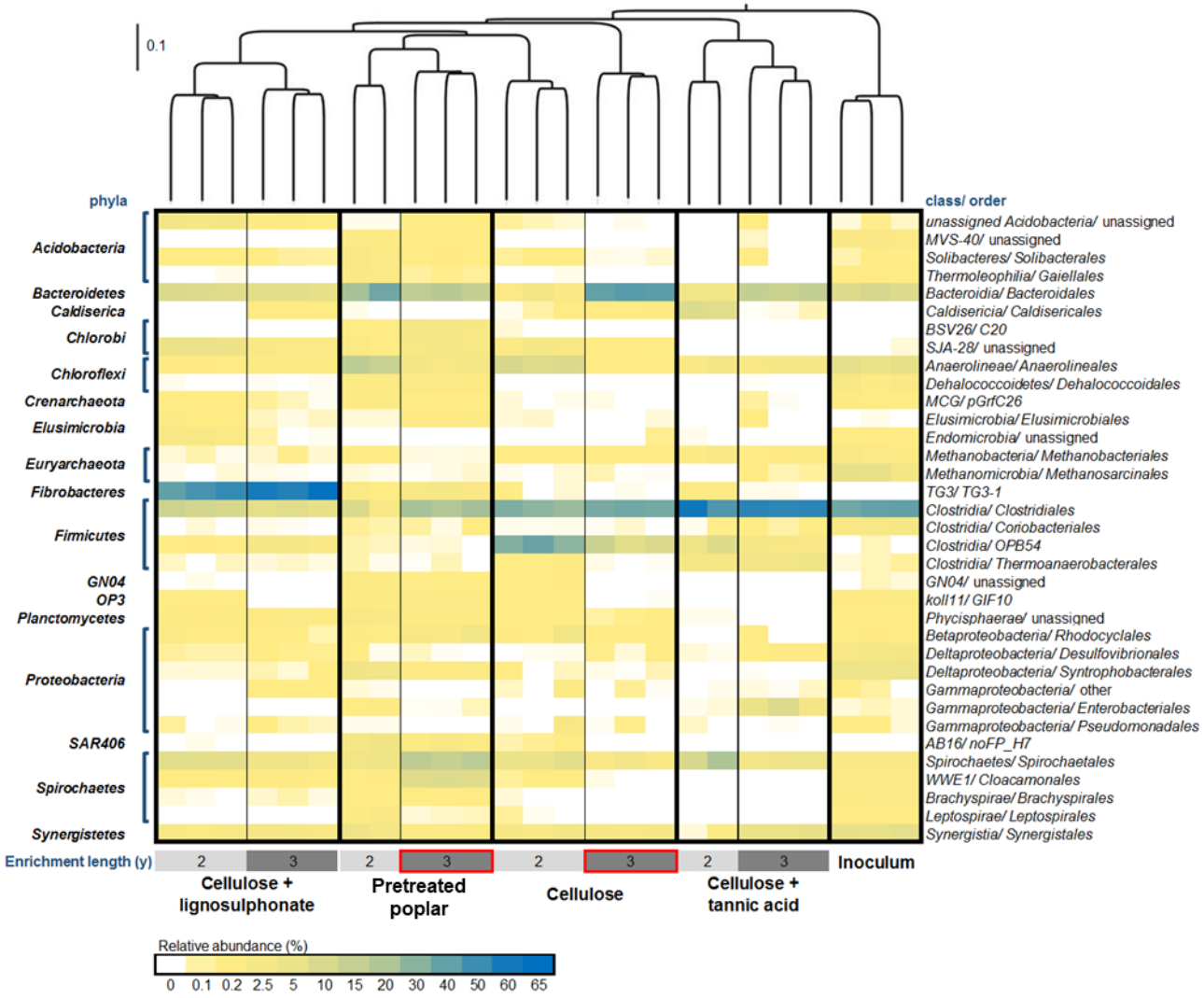


Figure 5.2 Heatmap of the microbial orders (≥ 1% in at least one sample) in lignocellulose-active microbial communities from pulp mill anaerobic granules and their corresponding enrichment cultures with UPGMA clustering. Bracketed in red were the samples selected for the following metagenomic analysis.

Prior to the enrichment condition, the pulp mill anaerobic granules were dominated by *Clostridiales* (on average 39%), *Bacteroidales* (10%), *Synergistales* (6%), *Anaerolineales* (6%), *Syntrophobacterales* (5%) bacterial orders, as well as methanogens belonging to *Methanosarcinales* order (5%) (Figure 5.2). Upon enrichment on lignocellulosic substrates under strictly anaerobic methanogenic conditions, a dramatic decrease in the microbial fraction belonging to the *Methanosarcinales* order was observed, to less than 0.1% in the enrichment cultures (Figure 5.2). Similarly, the population belonging to the *Syntrophobacterales* order diminished to below 0.1% in the enrichment cultures except for the pretreated poplar-fed cultures. The observed stochastic population dynamic over time could be an indicator of the ongoing niche differentiation and adaptation (Vanwonterghem et al., 2014). Indeed, there was an overall decrease in Chao I (richness) in the microcosms compared to the inoculum, reflecting the selective pressure imposed by the amended substrates (Figure S5.1). A similar trend was reported in our moose rumen microcosms after lignocellulose amendment (Wong et al., 2016), soil microbiota following enrichment with wheat straw (Jimenez et al., 2014a), as well as mesophilic anaerobic digesters that were fed with cellulose (Vanwonterghem et al., 2014).

After 3 years of enrichment, a few lineages that were implicated as lignocellulose degraders were present at high abundances in all the cultures despite the different enrichments.

The populations belonging to *Clostridiales* and *Bacteroidales* orders together constituted in similar abundances (± 6%) more than 75% of the microbial community in the cellulose-fed microcosms, ~40% in those fed with pretreated poplar, and equably (± 0.1%) ~14% of those fed with cellulose + lignosulphonate. As for the microcosms that were fed with cellulose + tannic acid, up to 54% of the community belonged to *Clostridiales* order, followed by ~15% in the *Bacteroidales* order. The following paragraph highlighted lineages that were uniquely abundant in specific enrichment microcosms.

As a reflection of the enrichment conditions, certain lineages were uniquely abundant in certain microcosms (summarized in Table 5.1). To begin with, *OPB54* was a dominant order that made up to 10% of the microflora in the cellulose-fed microcosms, while this order only occupied less than 3% of the cellulose + lignosulphonate and cellulose + tannic acid microcosms. Despite lack of detection in our anaerobic granule inoculum, members of *OPB54* were reported in diverse ecologies including forest soil (Kanokratana et al., 2011), rice paddy soil (Ishii et al., 2009), aquatic moss pillars (Nakai et al., 2012), bioreactors for landfill leachate (Burrell et al., 2004), solid waste (Tang et al., 2004; Muller et al., 2016), seaweed (FitzGerald et al., 2015), as well as our formerly established lignocellulose-degrading cultures from gut microbiomes of beaver and moose (Wong et al., 2017). Earlier, an *OPB54* isolate recovered from anaerobic sludge was experimentally determined to be an anaerobic bacterium that ferments a wide range of carbohydrates into acetate, ethanol and hydrogen (Liu et al., 2014). Next, the *TG3* (abbreviated from ‘termite group 3’) class under *Fibrobacteres* phylum represented a distinctively high abundant (60%) lineage unique to the cellulose + lignosulphonate enriched microcosms, attributed by a few unique OTUs only. As implied by the name, this class was initially detected in termite guts (Hongoh et al., 2006) and later surveyed in diverse habitats, such as anaerobic digesters

(Rahman et al., 2016), sheep rumen (bioproject PRJNA214227), hypersaline soda lakes (Sorokin et al., 2012), cockroach gut, lake sediment, rice paddy soil, and deep-sea sediments (Hongoh et al., 2006). It is also worthwhile to note that *TG3* was not detected in the anaerobic granules collected or in microcosms from beaver and moose (Wong et al., 2016), and only found at an insignificant abundance (<3%) in the microcosms fed with pretreated poplar. Moving on, the *Enterobacteriales* order under *Proteobacteria* phylum occupied ~6% of the microflora in the cellulose + tannic acid cultures. Members of this facultative anaerobic order were identified in the gut microflora of an invasive wood-boring beetle (Rizzi et al., 2013), and reported to promote synergism in lignocellulose biomass in mixed cultures via secretary products (Cortes-Tolalpa et al., 2017). Last but not least, the pretreated poplar-fed microcosms were especially enriched with the *Spirochaetales* order up to 16% and *Cloacamonales* order up to 8%. Ubiquitous in invertebrates as endosymbionts, *Spirochaetes* class is known for its role in the degradation of lignocellulose (Brune, 2014) and nitrogen fixation (Lilburn et al., 2001) in termite guts. Meanwhile, the ability to degrade cellulose was implicated for the candidate order *Cloacamonales* due to its presence in anaerobic reactors digesting energy crops (Lucas et al., 2015) and municipal sludge (Chouari et al., 2005; Limam et al., 2014). More recently, the enrichment of *Cloacamonales* in oil reservoirs (Shelton et al., 2016) and dechlorination cultures inspired the speculation that this microbial order could degrade hydrocarbon and uptake fermentation products syntrophically with the dechlorination of pentachlorophenol (Tong et al., 2017). The prevalence of lineages associated with dechlorination could be linked to the origin of the inocula, as low molecular weight chloro-organics are present in the effluents fed to the anaerobic wastewater treatment plant of pulp and paper mills (Freitas et al., 2009).

Table 5.1 Abundant orders (with relative abundances ≥ 5%) from the enrichments from pulp mill anaerobic granules

Enrichment	Abundant taxa (phylum/ class/ order)	Relative abundance	Reported habitat and functions
Cellulose	***Firmicutes*** *Clostridia/ OPB54*	10%	Forest soil (Kanokratana et al., 2011), rice paddy soil (Ishii et al., 2009), aquatic moss pillars (Nakai et al., 2012), bioreactors for landfill leachate (Burrell et al., 2004), solid waste (Tang et al., 2004; Muller et al., 2016), seaweed (FitzGerald et al., 2015), lignocellulose-degrading cultures from gut microbiomes of beaver and moose (Wong et al., 2017); ferments a wide range of carbohydrates to into acetate, ethanol and hydrogen (Liu et al., 2014)
Cellulose + lignosulphonate	***Fibrobacteres*** *TG3/ TG3-1*	60%	Termite guts (Hongoh et al., 2006), anaerobic digester (Rahman et al., 2016), sheep rumen (bioproject PRJNA214227), hypersaline soda lakes (Sorokin et al., 2012), cockroach gut, lake sediment, rice paddy soil, and deep-sea sediments (Hongoh et al., 2006)
Cellulose + tannic acid	***Proteobacteria*** *Gammaproteobacteria/ Enterobacteriales*	6%	gut microflora of an invasive wood-boring beetle (Rizzi et al., 2013); promotes synergism in lignocellulose biomass in mixed cultures via secretary products (Cortes-Tolalpa et al., 2017)
Pretreated poplar	***Spirochaetes*** *Spirochaetes/ Spirochaetales*	16%	Degradation of lignocellulose (Brune, 2014) and nitrogen fixation (Lilburn et al., 2001) in termite guts
	Spirochaetes *WWE1/ Cloacamonales*	8%	Anaerobic reactors for energy crops (Lucas et al., 2015) and municipal sludge (Chouari et al., 2005; Limam et al., 2014), oil reservoirs (Shelton et al., 2016) and dechlorinating cultures (Tong et al., 2017); potentially degrades hydrocarbon and uptakes fermentation products syntrophically with the dechlorination of pentachlorophenol (Tong et al., 2017)

To recognize key microbes that transformed the substrates prescribed, we also distinguished microbial OTUs that sustained through the enrichment microcosms. As a sign of substrate adaptation, the number of OTUs shared between the microflora in the inoculum and the microcosms reduced over time consistently for all enrichment conditions (Figure 5.3). Correspondingly, the relative abundances contributed by these shared OTUs decreased with time except for the microcosms fed with pretreated poplar. In fact, the microflora in the pretreated poplar-fed microcosms had the highest number of shared OTUs, contributing to ~60% of the microflora, a fraction that was at least twice as much as that in other enrichment microcosms. Certain OTUs were detected at high abundances in the cultures at both sampling points despite their low abundances in the inoculum (~0.1%), they included the *Clostridium* OTU2617 that was on average ~15% of the cellulose-fed cultures and ~7% of the cultures fed with cellulose + tannic acid. In the cellulose + tannic acid-fed cultures, the *Treponema* OTU1900 also represented ~5% of the microflora at both time points. Under the *Spirochaetales* order, the *Treponema* order was identified in landfill sites and the termite gut as symbionts for lignocellulose degradation (Warnecke et al., 2007; Rosenthal et al., 2011; Ransom-Jones et al., 2017). As defined by van der Gast et al. (2011), these OTUs could be considered core species by their persistent presence in the community at a high abundance (van der Gast et al., 2010). Contrarily, the transient presence of abundant OTUs was observed in the pretreated poplar-fed cultures. The *Bacteroidales* OTU2690 (19%) and *Anaerolinaceae* OTU351 (11%) were only transiently enriched after 2 years, whereas the *Treponema* OTU4415 (11%), *Clostridium* OTU2617 (10%), *W22* OTU3947 (7%) and *Bacteroidales* OTU 6109 (5%) were enriched after 3 years. While the ecological role of *W22* genus remains undefined, its presence is noted in the lignocellulose-degrading microcosms from beaver and moose (Wong et al., 2016), as well as dechlorinating cultures fed with

pentachlorophenol (Tong et al., 2017). In all, it was interesting to identify different dominant OTUs in the pretreated poplar-fed cultures at the sampling points since their taxonomic profiles clustered together regardless (Figure 5.2). A potential account for this would be that different microbial strains of the same genera proliferated over the long-term enrichment due to greater substrate heterogeneity or accumulation of metabolic waste.

Figure 5.3 Venn diagrams of the distribution of the overlapping OTUs between the inocula and microcosm over 2 and 3 years of substrate enrichment; bar charts show the sum of relative abundances contributed by the corresponding OTUs. Fractions contributed by major overlapping OTUs (those that occupy ≥ 3%) of the microflora are bracketed and labelled.

5.3.3 Comparing the abundance and taxonomic distribution of CAZymes in the cellulose- and pretreated poplar-fed microcosms

The metagenome DNA collected from cellulose- and pretreated poplar-fed cultures after 3 years of enrichment were selected for whole genome shotgun sequencing for the following reasons: i) contrasting selective pressures on the microbes ascribed by the structurally defined cellulose and heterogeneous wood polysaccharides found in pretreated poplar; ii) corresponding microcosms exhibited the greatest microbial acclimatization as exhibited in the improvement in biogas production with time (Figure 5.1).

Almost all (> 99%) the input read pairs were retained after quality trimming, and assembled into contigs to screen for genes encoding CAZymes and CBMs (Table S5.3). The annotation revealed in total 3,576 genes encoding 90 families of GHs, 12 families of CEs, 10 families of PLs and 33 families of GTs, as well as 33 families of associated CBMs (Table S5.4 A-E). No auxiliary redox enzymes were identified as expected for anaerobic microbial communities (Al-Masaudi et al., 2017; Svartström et al., 2017; Wong et al., 2017). Greater than half of the predicted plant polysaccharides-degrading CAZymes in the cellulose- (51%) and pretreated poplar-fed microcosms (70%) were less than 60% identical to the CAZyme amino acid sequences reported (Figure 5.4). Compared to the pretreated poplar-fed microcosms from beaver

dropping and moose rumen, the extent of plant polysaccharide-active CAZymes with low (less than 60%) sequence identity to those achieved in the CAZy database herein is greater by ~10%. Meanwhile, approximately 10% of the predicted plant polysaccharide-active CAZymes in both cellulose-fed cultures from beaver dropping and anaerobic granules were less than 30% identical to those reported previously. Across the CAZyme families, the percent identities varied (Figure S5.2), and the most divergent sequences belong to families GH113 (on average 26% identical to their closest blast hits), GH74 (34%), GH12 (32%), and PL9 (39%). The predicted sequences from families GH113, GH74, and PL9 were noted as highly divergent in microcosms from beaver dropping and moose rumen as well. In fact, only a small fraction (1 in 344 for GH113, 17 in 370 for GH74, and 10 in 960 for PL9) of the archived sequences in these families have been characterized, and so the added sequences from our metagenomes further expand the repository of CAZymes that require dedicated functional characterization.

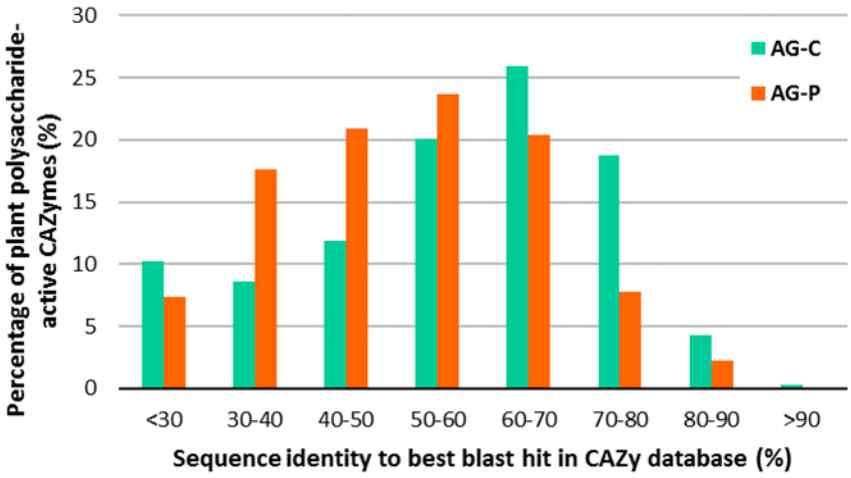


Figure 5.4 Percent identity between sequences in the CAZy database and CAZyme sequences predicted in pulp mill anaerobic granule cultures (AG) fed with cellulose (C) and pretreated poplar (P)

On average, 11% of the annotated CAZymes and CBMs were taxonomically unassigned or assigned to unknown species on the phyla level. The percentage increased to 16% for cellulose-fed culture and 43% for pretreated poplar-fed culture on the class level. Such high percentage could be owing to the limited CAZyme and CBM sequences with phylogenetic details deposited in the reference NCBI-NR database. For those that were phylogenetically assigned, most of them were derived from *Clostridia* (14% to 32%) and *Bacteroidia* classes (18% to 24%) as predicted by the 16S rRNA gene survey (Figure 5.5). Indeed, CAZyme families that comprise plant polysaccharide-active enzymes (i.e., families CE1, CE4, GH2, GH3, GH5 and GH43) were most frequently assigned to either *Clostridia* or *Bacteroidia* (Figure S5.3). Other abundant lineages found to be key contributors of CAZymes and CBMs included *Spirochaetes* class, which constituted ~17% of the microbial community of the pretreated poplar culture and encoded 3% of the CAZymes and CBMs (Figure 5.5). Furthermore, dominant microbial classes that contributed to the annotated CAZymes and CBMs including *Caldisericia* from *Caldiserica* phyla, *Chitinispirillia* from *Fibrobacteres* phyla, *Cytophagia* and *Sphingobacteriia* from *Bacteroidetes* phyla were distinctive to the microcosms from anaerobic granules and not those from the gut microbiomes of beaver and moose (Wong et al., 2017).

Interestingly, many CAZyme and CBM sequences were phylogenetically affiliated to microbes that were not necessarily abundant in the corresponding population based on the 16S rRNA gene amplicon analysis. Examples from the cellulose-fed microcosms included *Anaerolineae* class that was found to be 0.5% of the community, but contributed up to 7% of genes encoding CAZymes and CBMs; as well as undetected *Bacilli* class that contributed to a 3% of the annotated genes (Figure 5.5). Herein, the small fraction (3%) of CAZymes and CBMs linked to *Bacilli* class was surprising, since the *Bacillus* genus consistently represented the

dominant genus and encoded enzymes with a broad range of lignocellulolytic activities in other pulp and paper mills (Desjardins and Beaulieu, 2003; Maki et al., 2011; Ghribi et al., 2016). Similar observations were found in pretreated poplar-fed microcosms, where the 5% of the predicted CAZymes and CBMs belonged to the undetected *Ignavibacteria* class. The observed discrepancy could be a reflection of the fundamental limitations of the gene marker- and protein-based approaches for taxonomic assignment, including the different binding affinities of the primers to the corresponding conserved regions of the 16S rRNA gene (Jovel et al., 2016), the varying taxonomic resolution of the hypervariable regions (Soergel et al., 2012), and the use of taxonomically-biased protein database for the metagenomic taxonomic assignment. Nonetheless, it is also worthwhile to consider that the microbial memberships could be confounded by the cooperative interspecies interactions (Kato et al., 2005; Ju and Zhang, 2015). For instance, mutualistic co-culturing of ruminal bacteria with *Treponema byrantii* enhanced the degradation of straw despite the latter's lack of cellulose-degrading capacity (Kudo et al., 1987). In our case, non-hydrolytic fermenters such as *OPB54* were abundant in the microcosms and essential for the conversion of the fed substrates into biogas. Thus, it is important to characterize a lignocellulose-degrading microbiome both taxonomically and functionally to infer the genetic capacity for CAZymes and the potential CAZyme-encoding microbes.

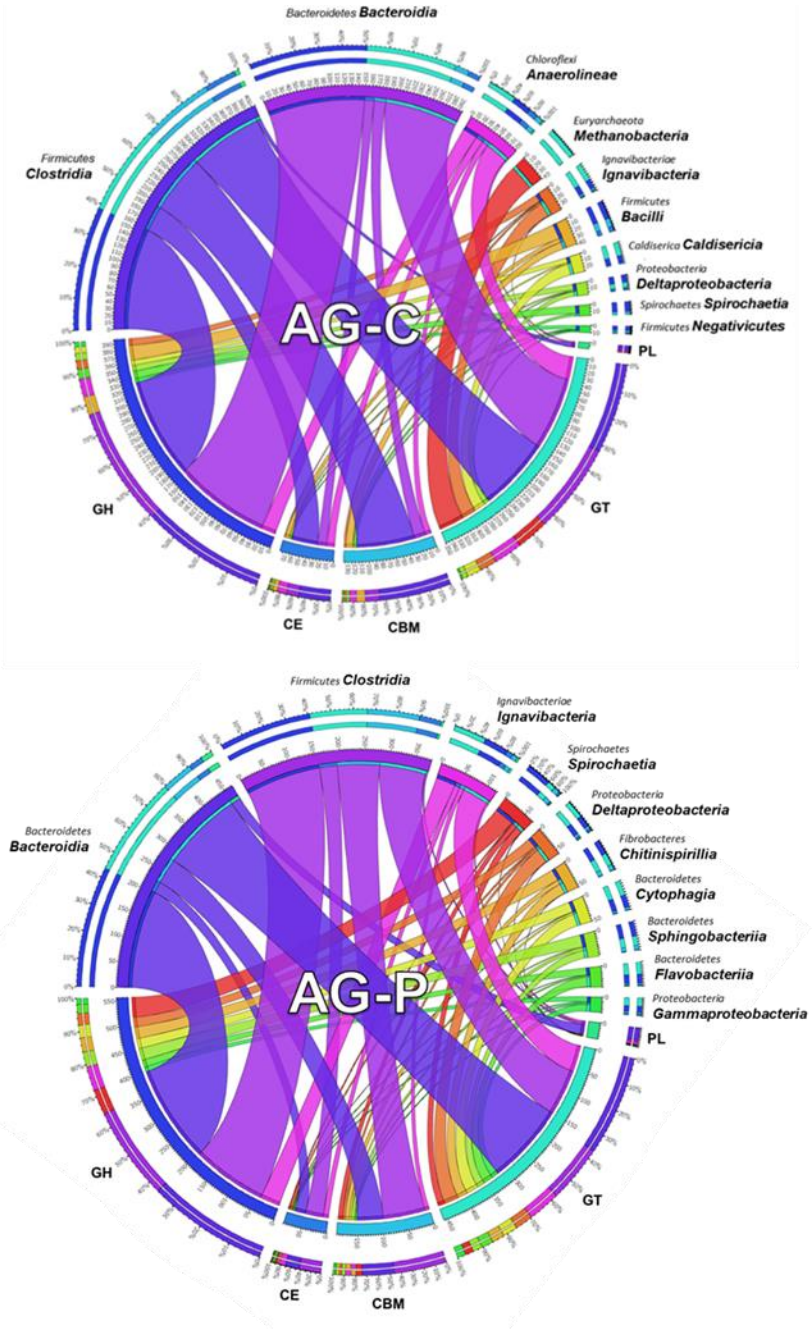


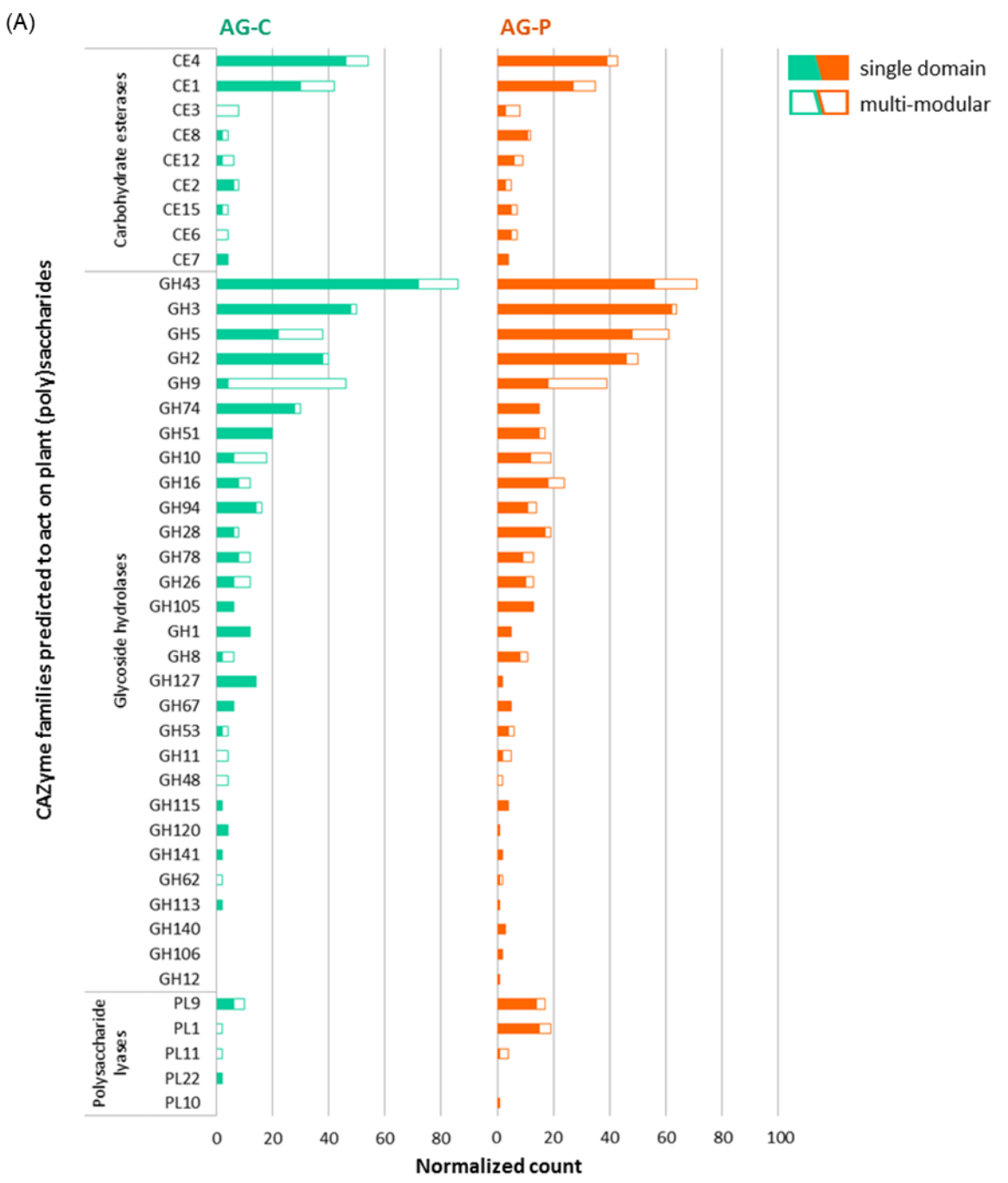
Figure 5.5 Phylogenetic distribution of CAZyme and CBM sequences assigned to the top ten identified classes in pulp mill anaerobic granule cultures (AG) fed with cellulose (C) and pretreated poplar (P) shown via the connections colour-coded in accordance to the classes. The number and corresponding percentage are shown on the outer scale.

Metagenomes of cultures enriched on cellulose and pretreated poplar yielded a similar proportion of predicted plant polysaccharides-active CAZymes (~28%) primarily belonging to families CE4, CE1, GH43, GH3, GH5, GH2, and GH9 (Figure 5.6A). While the profile of predominant families predicted to act on plant polysaccharides resemble that of the cultures from beaver dropping and moose rumen (Wong et al., 2017), the substrate-induced differences in the microcosms from anaerobic granules were distinguished from those reported previously. Comparing the two metagenomes, pretreated poplar-fed cultures encoded higher counts of GH5 (1.6 times higher), PL1 (9.5 times), GH16 (2 times), GH28 (2.4 times), CE8 (3 times), GH105 (2.2 times) (Figure 5.6B); none except GH5 were reported to be enriched in the pretreated poplar-degrading communities from the beaver droppings and moose rumen (Wong et al., 2017). Characterized members of the family GH5 exhibit a divergent range of target substrates, including cellulose, mannans, and xylans (Gallardo et al., 2010; Aspeborg et al., 2012). Similarly, GH16 members were known to act as endoglucanases, endogalactanases, as well as hydrolases and endo-transglycosylases on xyloglucans (Baumann et al., 2007; McGregor et al., 2017).

Contrary to our microcosms from the gut microbiomes, the pretreated poplar-fed microcosm from anaerobic granules presented a higher abundance (compared to the cellulose-fed counterpart) in a broader range of CAZyme families that are predicted to act on pectins. Structurally speaking, pectins are a family of galacturonic acid-rich polysaccharides including galacturonan, rhamnogalacturonan I, substituted homogalacturonan rhamnogalacturonan II and xylogalacturonan (Mohnen, 2008). Family PL1 harnesses pectate lyases that cleave α-1,4-linked d-galacturonans (Charnock et al., 2002); family GH28 primarily include polygalacturonases that hydrolyze the α-1,4-linkages between galacturonate backbone residues (Abbott and Boraston, 2007), rhamnogalaturonases that hydrolyze α-1,2 linkages between rhamnose residues and

galacturonic acid, and rhamnohydrolases that target the non-reducing end of rhamnogalacturonan (Schols et al., 2009). To date, all characterized members of CE8 were reported as pectin methylesterases (Rajulapati and Goyal, 2017), and GH105 as unsaturated rhamnogalacturonyl hydrolases (Itoh et al., 2006). The greater representation of pectin-active CAZymes from the microcosms from pulp mill anaerobic granules could be attributed to the extensive dissolution of pectin in the process water that is fed to the anaerobic wastewater treatment reactor in vast volumes (Konn et al., 2002); in comparison, the gut microbiomes are exposed to less pectins in their diets on a daily basis. Interestingly, notable increases in CAZyme families known to degrade hemicellulose (seven times in GH127, four times in GH120) were also identified in cellulose-fed microcosms.

Of the sequences predicted to encode plant polysaccharide-degrading enzymes (i.e., 669 sequences) in cellulose- and pretreated poplar-fed cultures, 29% and 22% were predicted to form multi-domain proteins (Figure 5.6A), respectively. Most frequent domain organizations included CBMs, such as CBM48-GH13_9 (3% of the predicted multi-modular CAZymes in both cultures) and GH23-CBM50-CBM50-CBM50 (4% in pretreated poplar-fed culture) (Table S5.5). Similar to the microcosms from beaver dropping and moose rumen, potential cellulosomal subunits characterized by the presence of CAZyme appended-dockerins were identified in the metagenomes from the anaerobic granules cultures. These CAZyme appended-dockerins were found at comparable frequencies in both cultures (Table S5.5), and the more commonly identified organizations included GH9-CBM3-DOC1 (4-6% of the predicted multi-modular CAZymes in both cultures), GH9-DOC1 (2-3%), GH9-CBM3-CBM3-DOC1 (2-3%) and CE3-DOC1 (2%). Xylanase cellulosomal genes with organizations akin to the first two were found in *Clostridium thermocellum* (Mudgal et al., 2015).


(A)
AG-C
AG-P
single domain
multi-modular
CAZyme families predicted to act on plant (poly)saccharides
Carbohydrate esterases
CE4
CE1
CE3
CE8
CE12
CE2
CE15
CE6
CE7
Glycoside hydrolases
GH43
GH3
GH5
GH2
GH9
GH74
GH51
GH10
GH16
GH94
GH28
GH78
GH26
GH105
GH1
GH8
GH127
GH67
GH53
GH11
GH48
GH115
GH120
GH141
GH62
GH113
GH140
GH106
GH12
Polysaccharide lyases
PL9
PL1
PL11
PL22
PL10
0
20
40
60
80
0
20
40
60
80
100
Normalized count

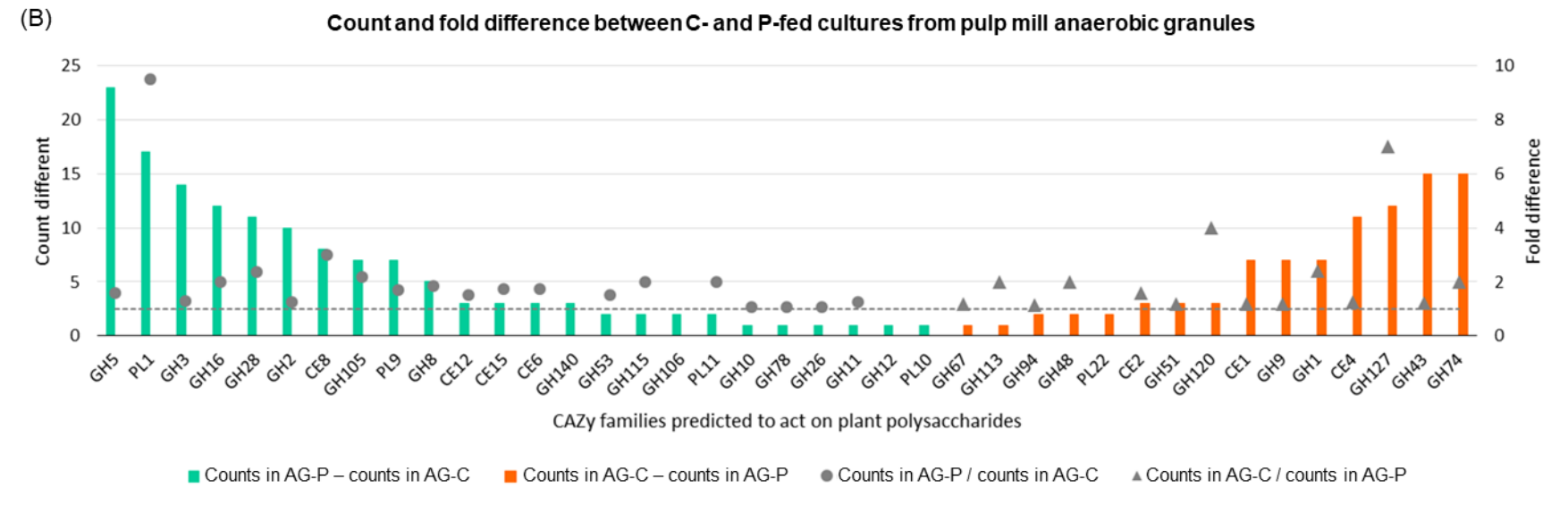


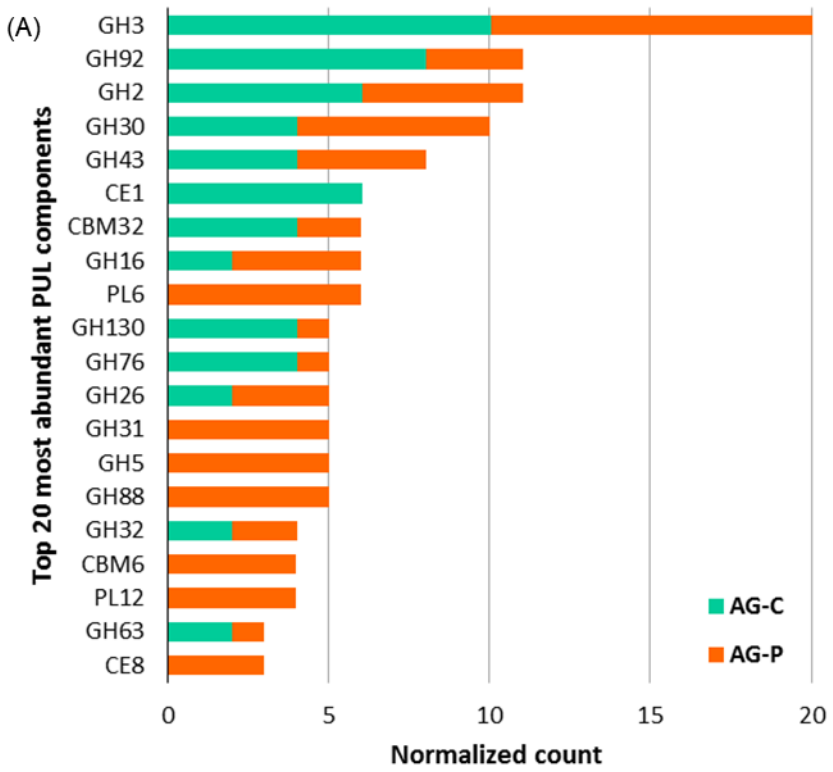
Figure 5.6 (A) Distribution of plant (poly)saccharide degrading-CAZyme families as single and multi-modular domains in pulp mill anaerobic granule cultures (AG) fed with cellulose (C) and pretreated poplar (P). (B) Normalized count and fold difference of CAZyme families predicted to act on plant polysaccharides between pretreated poplar- and cellulose-fed pulp mill anaerobic granule cultures. Fold difference was only calculated for non-zero counts.

5.3.4 Uncovering potential novel CAZymes through analysis of PULs and CAZyme sequences comprising domains of unknown function

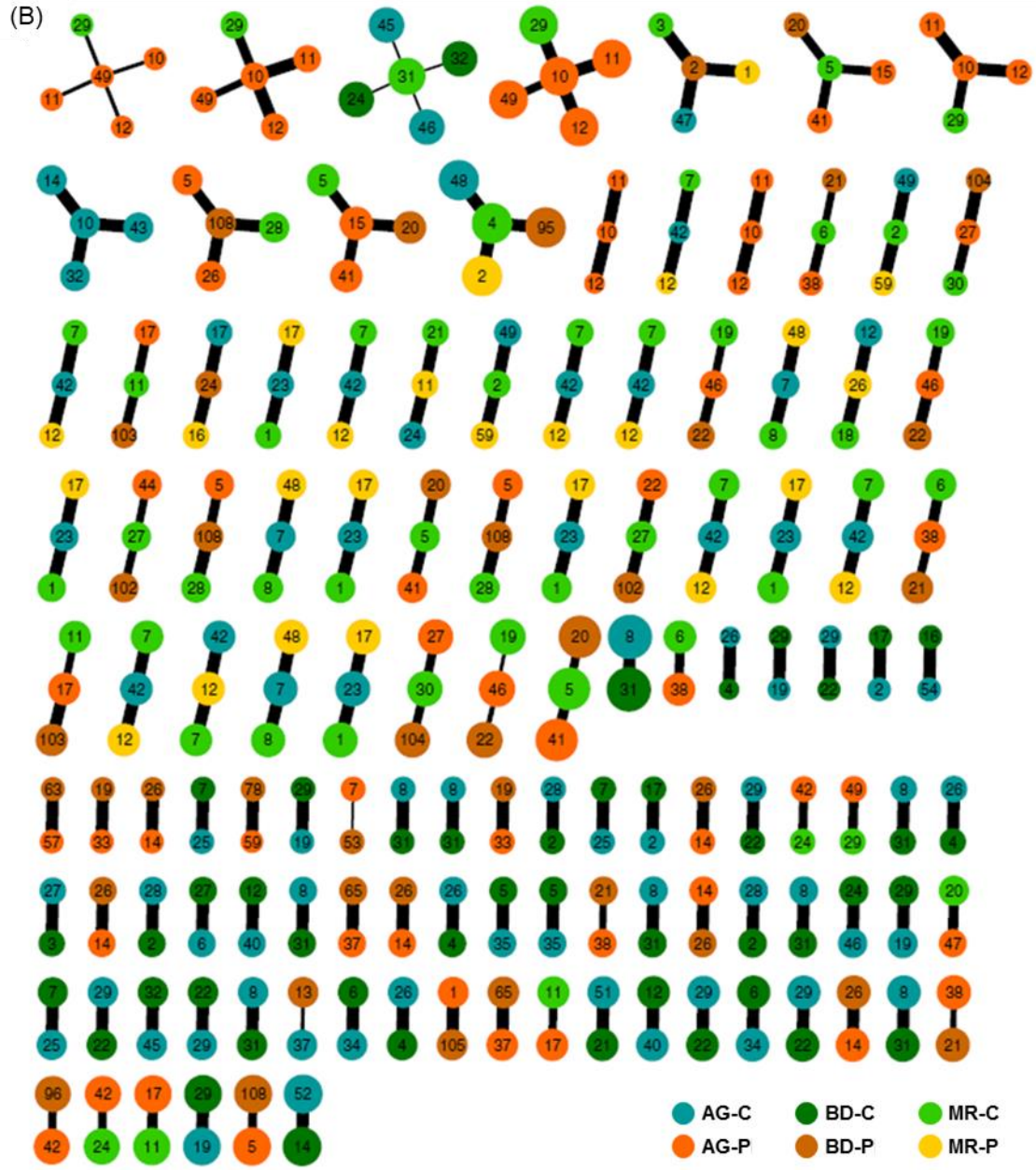
Recent efforts have focused on characterizing PULs for the discovery of novel enzymes and synergism within the PUL components (Despres et al., 2016; Larsbrink et al., 2016). A total of 116 PULs were predicted herein, where the number predicted in the pretreated poplar-fed microcosm was almost twice that predicted in the cellulose-fed one (Figure S5.4). Consistent with the overall distribution of predicted CAZyme sequences, those belonging to abundant families GH3, GH2, and GH43 were frequently identified in the predicted PULs (Figure 5.7A). In addition to that, sequences belonging to families GH92 and GH30 were frequently identified as components of the PULs. Comparing the PULs between the two metagenomes, PULs comprising members of families GH31, GH5, GH88, CBM6, and PL12 were exclusively identified in the metagenome from the culture enriched on pretreated poplar. In the context of lignocellulose conversion, characterized bacterial members of GH31 include glucosidases, galactosidases and xylosidases (Lovering et al., 2005; Larsbrink et al., 2011). And CBM6 modules, first identified as a xylan-binding domain from a multi-modular xylanase (Fernandes et al., 1999), later revealed other substrate specificities in accordance to their appended catalytic domains via co-evolution (Michel et al., 2009). Meanwhile, glucuronyl hydrolases in the family GH88 degrade bacterial biofilms and mammalian glycosaminoglycans. The characterized members of family PL12, in a mere total of three to date, are heparin-sulfate lyases active on glycosaminoglycans (Garron and Cygler, 2010).

Under the notion that novel lignocellulose-active functions may be presented by the proteins marked as having unknown function found within the PULs (Ndeh et al., 2017), a total

of 249 proteins with unknown functions (with lengths ranging from 38 to 1,075 amino acids) were identified in the current analysis (Figure S5.4). In an effort to prioritize additional sequences for future characterization, a sequence identity-based clustering network diagram was constructed to uncover protein sequences with unknown function that reoccurred in the enrichment cultures. More than half (60%) of these sequences clustered with those found in beaver and moose rumen cultures (Figure 5.7B), and those that clustered together typically originated from PULs with similar architecture (Figure S5.4). Since substrate category of the PUL-encoded enzymes can be inferred based on the established activities of the CAZyme families of a given PUL (Terrapon et al., 2017), we could extract insights for the experimental design to characterize the proteins of unknown function. For instance, Figure 5.7C illustrates PULs that potentially target cellulose, hemicelluloses and pectins based on the established activities of the CAZyme families.



(B)



(C)

Potential PUL for cellulose or xylan main chain degradation from cellulose-fed microcosm

SusC ▸ SusD ▸ GH30_3 ▸ unk ▸ GH30_3 ▸ GH3 ▸

Potential PUL for pectin degradation from poplar hydrolysate-fed microcosm

SusC ▸ SusD ▸ unk ▸ GH28 ▸ CE12 ▸ GH105 ▸ CE8 ▸ PL1 ▸ PL1 ▸ CE8 ▸ unk ▸ PL1|CBM77 ▸ CE8 ▸ unk ▸

Figure 5.7 (A) Top 20 most abundant CAZyme families identified in predicted PULs from cellulose (C)- and pretreated poplar (P)-fed pulp mill anaerobic granules (AG) microcosms. (B) Similarity-based clustering (≥ 70%) of proteins with unknown function positioned in PULs identified herein with those in the lignocellulose-degrading microcosms from beaver dropping and moose rumen (Wong et al., 2017). Each cluster contains a central node that denotes the representative protein with unknown function (defined by the longest length) and connected nodes that represent a protein with unknown function that is ≥ 70% identical to the representative sequence. PUL identifiers are shown on each node; the size of the nodes correlates to the amino acid length of the protein; the thickness of the edges correlates to percent identity between sequences. **(C) Examples of predicted polysaccharide utilization loci (PULs) that potentially target cellulose, xylan, or other hemicelluloses based on the membership of the components**.

Another avenue to discover potentially new CAZyme families looked at lignocellulose-active multi-modular proteins that comprise a domain of unknown function (DUF). Interestingly, the multi-modular proteins with CBMs and DUFs identified herein (Table 5.2) had the same organizations and almost identical sequences (percent identity ≥ 99%) as those reported previously (Wong et al., 2017) except for DUF3459-CBM48-GH13_10 and DUF3794-CBM50. Despite the same module organization as those from the previous analysis, the identity within the unique amino acid sequences of DUF3459-CBM48-GH13_10 and DUF3794-CBM50 were as low as 33% and 27%, respectively. As described in chapter 4, the structure of DUF3459 as part of the multi-modular proteins with GH13 and CBM48 domains were determined, yet its functional role in the presumed α-amylases remained uncharacterized (UniProt entries W6LS46, R4KHQ4, C7RTS8). Similarly, limited knowledge is known about the DUF3794 aside from its usual

association with CBM50 (Finn et al., 2016), which are known to bind to various types of chitin and peptidoglycan (Bateman and Bycroft, 2000; Buist et al., 2008). Since little information is known about the DUFs to date (Mudgal et al., 2015), the occurrences of these multi-modular proteins in the cultures were observed to glean potential functional insights. For instance, while DUF362-CBM9-DOC1, DUF3794-CBM50 and DUF4366-CBM16 were present in all microcosms, some were consistently found in microcosms enriched with specific substrates. For instance, DUF5110-GH31-CBM32-DOC1 were consistently found in cellulose-fed cultures from anaerobic granules and beaver droppings, whereas DUF3459-CBM48-GH13_10 and DUF5011-CBM4 were found in the pretreated poplar-fed counterparts. As mentioned earlier, characterized bacterial members of GH31 include glucosidases, galactosidases and xylosidases (Lovering et al., 2005; Larsbrink et al., 2011), while CBM32 are known to bind to a broad substrate range including mannan (Mizutani et al., 2012), galactan (Fujita et al., 2014) and pectin (Abbott et al., 2007). As for CBM4, modules in this family primarily target xylan (Abou Hachem et al., 2000), β-1,3-glucan (Boraston et al., 2001) and β-1,4-glucan (Tomme et al., 1996), β-1,6-glucan (Zverlov et al., 2001) and amorphous cellulose (Coutinho et al., 1992).

Table 5.2 Carbohydrate-active proteins with DUFs identified in the metagenomes. C, cellulose; P, pretreated poplar.

		Anaerobic granules		*Beaver droppings*		*Moose rumen*	
DUF	CAZyme / CBM	C	P	C	P	C	P
DUF362	CBM9-DOC1	1	1	1	1	1	1
DUF3459	CBM48-GH13_10		1		5		
DUF3794	CBM50	2	1	3	2	2	3
DUF4366	CBM16	1	1	1	1	1	1
DUF5011	CBM4		1		1		
DUF5110	GH31-CBM32-DOC1	1		1			

5.3.5 Impact of inoculum source on the microbial composition and CAZymes predicted to degrade plant polysaccharides

The microbial composition and predicted plant polysaccharide-degrading CAZymes of the cellulose- and pretreated poplar-fed cultures from the pulp mill anaerobic granules and the gut microbiomes of beaver and moose (Wong et al., 2017) were compared to gain insights regarding the impact of the seeding inoculum.

Hierarchical clustering of the community composition showed a clear separation depending on the fed lignocellulose substrates among the microcosms developed from the gut microbiomes of beaver and moose (Figure 5.8); microcosms from pulp mill anaerobic granules, on the other hand, formed a unique cluster that was further distinguished based on the enrichment substrate. The two groups showed a partial overlap when visualized in a PCA plot (that explained 49.5% variance), which was reasonable given the presence of common lineages for the degradation of cellulose that were present in pretreated poplar as well. Within the cellulose-fed microcosms, a significant difference was found in the abundance of *Bacteroidales* order, the relative abundance in the microcosms from the anaerobic granules was around two and seven times of that in the microcosms from beaver and moose (Table 5.3). Aside from that, the cellulose-fed microcosms from anaerobic granules also had a significantly lower prevalence of an unassigned order under *SJA-28* class (0.5%), *Anaerolineales* (0.5%) and *Spirochaetales* (2.3%) orders than their counterparts which ranged from 5% to 17% in the microcosms from the gut microbiomes. As for the pretreated poplar-fed microcosms, those originating from the anaerobic granules have a distinctly higher abundance of *Spirochaetales* (16%) and a lower abundance of *Clostridiales* (22%), which were less than ¼ and more than twice of those from the gut

microbiomes, respectively. On top of that, lesser-known lineages such as *SJA-28* class, *C20* and *TG3-1* orders were found at > 2% in the pretreated poplar-fed microcosms from the anaerobic granules, while the occurrence in the microcosms from the gut microbiomes were ≤ 0.4%. The *SJA-28* class under *Chloribi* phylum was also identified in other anaerobic digesters, termite gut, and more recently in a photosynthetic bioreactor inoculated with different activated sludges (Oyserman et al., 2017), as well as a dominant lineage in the gut microflora enriched with cellulose and cellulose + lignosulphonate (Wong et al., 2016). Similarly, the *C20* order remains to be functionally unknown, and *BSV26* class it belongs to occurred in mesothermal microbial mats near hot springs (Ross et al., 2012) and shallow sea hydrothermal vents (Huang, 2012). These lesser-known lineages, along with their occurrences according to other literature, are summarized in Table 5.3.

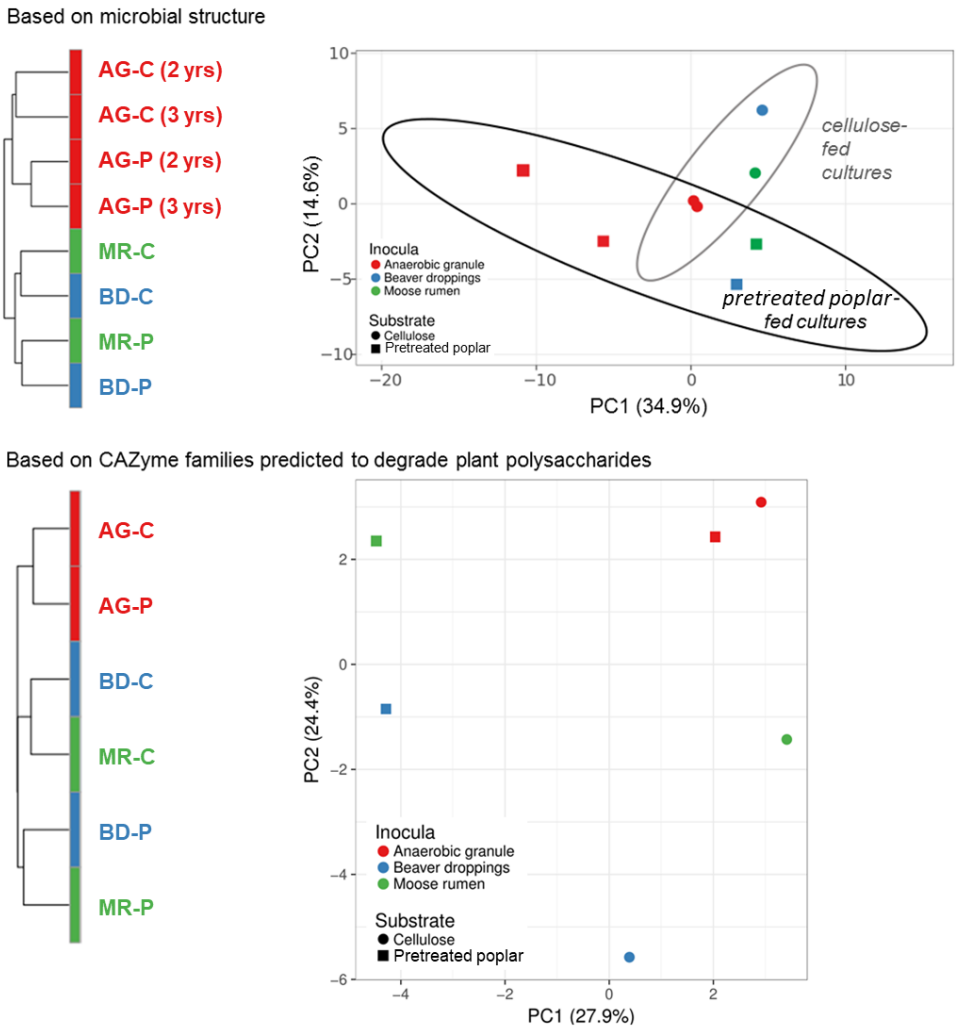


Figure 5.8 Correlation clustering and PCA plots based on community 16S rRNA gene sequencing and CAZyme families predicted to degrade plant polysaccharides. Ellipses showed 95% confidence. AG, pulp mill anaerobic granule; BD, beaver dropping; C, cellulose; MR, moose rumen; P, pretreated poplar.

Table 5.3 Orders of the anaerobic granule microcosms with abundances consistently significantly different (p-value < 0.05) from the microcosms from the gut microbiomes after 3 years of cellulose and pretreated poplar enrichment. Reported habitats of the relatively uncharacterized lineages are listed.

Enrichment	Taxa (phylum/ class/ order)	*Relative abundance*			Reported habitat
		Anaerobic granule	Beaver dropping	Moose rumen	
Cellulose	***Bacteroidetes*** *Bacteroidia/ Bacteroidales*	41.3%	18.4%	5.6%	
	Chlorobi *SJA-28/* unassigned order	0.5%	9.3%	17.1%	Anaerobic digesters, termite gut, photosynthetic bioreactor inoculated with activated sludges (Oyserman et al., 2017)
	Chloroflexi *Anaerolineae/ Anaerolineales*	0.5%	7.9%	7.7%	
	Spirochaetes *Spirochaetes/ Spirochaetales*	2.3%	4.6%	9.8%	
Pretreated poplar	***Chlorobi*** *SJA-28/* unassigned order	2.3%	0.0%	0.4%	Anaerobic digesters, termite gut, photosynthetic bioreactor inoculated with activated sludges (Oyserman et al., 2017)
	Fibrobacteres *TG3/ TG3-1*	2.2%	0.0%	0.1%	Termite guts (Hongoh et al., 2006), anaerobic digester (Rahman et al., 2016), sheep rumen (bioproject PRJNA214227), hypersaline soda lakes (Sorokin et al., 2012), cockroach gut, lake sediment, rice paddy soil, and deep-sea sediments (Hongoh et al., 2006)
	Spirochaetes *Spirochaetes/ Spirochaetales*	16.4%	4.8%	0.9%	
	Chlorobi *BSV26/ C20*	2.5%	0.0%	0.0%	*BSV26* class occurred in mesothermal microbial mats near hot springs (Ross et al., 2012), shallow sea hydrothermal vents (Huang, 2012)
	Firmicutes *Clostridia/ Clostridiales*	22.0%	39.9%	53.6%	

Hierarchical clustering analysis of the predicted plant polysaccharide-active CAZyme families distinguished those reported in this study from those from the gut microbiomes of beaver and moose, which were clustered based on the substrates fed (Figure 5.8). Based on the PCA loadings, both cellulose- and pretreated poplar-fed microcosms from pulp mill anaerobic granules were distant from the cellulose-fed microcosms from the gut microbiomes due to the higher abundances of GH26, CE4 and GH43 in the former (Figure S5.5). When compared to the pretreated poplar-fed microcosms from gut microbiome, both cellulose- and pretreated poplar-fed microcosms from the anaerobic granules had a higher frequency of genes encoding GH3, GH53, CE15, GH127, CE1 and GH12. These enriched carbohydrate-active enzymes families included those that comprise enzymes involved in plant polysaccharide deconstruction. Characterized bacterial members of family GH26 are predominantly endo- and exo-acting mannanases (Cartmell et al., 2008), β-1,3-xylanase and xyloglucanase (Araki et al., 2000). As for CE4, characterized bacterial enzymes included endo-1,4-β-xylanase and acetyl xylan esterase (Caufrier et al., 2003). The major activities reported for GH43, on the other hand, are α-arabinofuranosidases (Flipphi et al., 1993), endo-α-arabinanases (McKie et al., 1997) and β-xylosidases (Shallom et al., 2005). As a family with broad substrate specificities, GH3 family groups together enzymes that possess β-D-glucosidase, β-D-xylosidase, α-L-arabinofuranosidase, and N-acetyl-β-D-glucosaminidase activities (Faure, 2002). GH53, on the other hand, is known for its specificity to act as β-1,4-galactanases in the microbial degradation of galactans and arabinogalactans in pectins (Sakamoto et al., 2013). Next, characterized CE15 members display 4-O-methyl-glucuronoyl methylesterase activity that hydrolyzes ester linkages between hydroxyl groups in lignin and 4-O-methyl-D-glucuronic acid residues in glucuronoxylans that dominate in hardwood fiber (Biely et al., 2015). Family GH127 typically contain β-L-arabinofuranosidases that were shown to target plant cell

wall glycoproteins such as extensin (Fujita et al., 2011), while family CE1 members were shown to deacetylate polymeric xylans (Mai-Gisondi et al., 2017). Last but not the least, xyloglucan endo-hydrolase activities were reported for GH12 (Gloster et al., 2007).

Taken together, it was corroborated that both inocula and fed substrate were definite influencers on the microbial enrichment taxonomically and functionally. In the case of microcosms from pulp mill anaerobic granules, the inocula overrode the substrate-induced convergence observed in the microcosms from the gut microbiomes, and yielded taxonomically and functionally distinct microbial consortia after substrate enrichment. The overarching greater abundance of CAZymes families with reported activities on hemicelluloses and pectins could be a functional vestige from the anaerobic digestion of heterogeneous woody biomass by pulp mill anaerobic granules prior to our enrichment. On that note, surveying microbial enrichments developed from divergent environmental samples is hence a recommended strategy to screen for novel microbes and enzymes for the degradation of the fed substrate.

5.4 Conclusion

To explore the biomass-degrading microbiomes beyond the grass-degrading gut microbiota and to compare the impact of inocula versus substrate on the enrichment microcosm, various lignocellulosic substrates were fed to the microcosms from pulp mill anaerobic granules to enrich the lignocellulolytic microbes for taxonomic and functional characterization. Temporal tracking of the microbial communities showed a greater similarity and OTU overlap between the microcosms fed with the same substrates than they were to the original inocula, reflecting the impact of substrate conditioning on the community structure. Absent in the microcosms from the beaver dropping and moose rumen, a poorly defined termite gut-associated lineage known as *TG3* class under *Fibrobacteres* phylum represented the majority (60%) of the microflora in the anaerobic granules fed with cellulose + lignosulphonate. Among the top microbial classes that encoded the CAZymes and CBMs, *Caldisericia* from *Caldiserica* phyla, *Chitinispirillia* from *Fibrobacteres* phyla, *Cytophagia* and *Sphingobacteriia* from *Bacteroidetes* phyla were distinctive to the microcosms from anaerobic granules and not those from the gut microbiomes. Surprisingly, the *Bacilli* class occurred in a low abundance in the community and only contributed to a small fraction of CAZymes (3%) and CBMs despite being reported as a dominant lignocellulolytic lineage in other pulp and paper mills. Furthermore, the pretreated poplar-fed microcosm from the anaerobic granules uniquely exhibited a broader range of dominant CAZyme families that are predicted to act on pectins, including PL1, GH28, CE8, and GH105, when compared to the cellulose-fed counterpart. Altogether, we settled the debate and cogitated that both the inocula and substrate enrichment were definite drivers of the taxonomic and functional composition of the resultant microbiomes. Substrate-induced convergence was more pronounced among enrichments with similar inocula such as gut microbiomes, whilst distinctive inocula from pulp

mill anaerobic digester left a discernible imprint on the developed microcosm taxonomically and functionally. The greater presence of poorly characterized *SJA-28*, *TG3* and *BSV26* classes, as well as the higher counts of presumably pectinolytic GH53 and hemicellulolytic GH53, CE15, CE1 and GH12 members distinguished the pretreated poplar-fed microcosm from anaerobic granules from those enriched from the gut microbiomes. To capitalize on the unique microbiome from the anaerobic granules, predicted pectin- and hemicellulose-degrading CAZymes from the less prominent lineages in the microcosms from the gut microbiomes shall be mined for potentially new catalytic activities. On this note, developing microbial enrichments from divergent microbiomes is emphasized to discover novel lignocellulolytic enzymes.

Chapter 6 Saccharification capacity of the secretomes from cellulose- and pretreated poplar-degrading microcosms from pulp mill anaerobic granules and the digestive microflora from Canadian beaver and North American moose

6.1 Introduction

The environmental impact and dwindling reserve of fossil fuels have fostered a global consensus that lignocellulose, the most abundant non-edible plant biomass on earth, should be functionalized as renewable chemical precursors to provide viable alternatives for petrochemicals (Kobayashi and Fukuoka, 2013). Unfortunately, the recalcitrance of plant cell wall lignocelluloses to enzymatic hydrolysis prevents an efficient bioconversion into simpler units, and thereby imposes a major impediment to the economic sustainability of renewable biochemicals. In nature, species-rich fungal and microbial communities break down lignocellulose via synergistic actions of degradative CAZymes (Gao et al., 2010), for instance, cellulases, cellobiohydrolases, laccases, xylanases, side chain active enzymes such as arabinofuranosidases, and bacterial cellulosomes as multi-enzymatic complexes. In today's biorefinery, most commercial enzyme cocktails applied are mixtures of fungal cellulases and hemicellulases (King et al., 2009), although some can become deactivated by heat or residual lignocellulose pretreatment chemicals (Bouws et al., 2008). To uncover genes that encode CAZymes with greater robustness and novel activities for industrial application, homology-based approaches such as metagenomics are applied to various biomass-degrading habitats ranging from forest soil (Lopez-Mondejar et al., 2016), gut

microbiome of termites and herbivorous mammals (Warnecke et al., 2007; Hess et al., 2011; Svartström et al., 2017), to anaerobic digesters for wood chips (van der Lelie et al., 2012) and wastewater sludge (Wilkens et al., 2017). Moreover, environmental conditioning via substrate enrichment, exposure to specific temperature, pH, oxygen content, and fermentation inhibitors are often applied prior to the sequencing in order to preferentially augment the community members that encode CAZymes that could withstand austere industrial conditions (Gao et al., 2010). Nonetheless, metagenomics alone cannot decipher whether the annotated CAZymes are translated or functional in biomass degradation.

The direct proteomic analysis of the secreted proteins of microbial and fungal isolates, also known as secretomics, compliments the metabolic insights inferred by metagenomics and enables a direct assessment of the lignocellulolytic capacity. Furthermore, assaying the secretomes directly captures the lignocellulolytic activities of the less characterized CAZymes with amino acid sequences and catalytic domains may not be well-established in existing reference database for the homology-based search (Li et al., 2009). More recently, the function-based screening is applied in mixed cultures as metasecretomics to capitalize the synergism between the secreted proteins for developing enzyme cocktails. Indeed, significant improvement in enzymatic activities was reported in the secretome of thermophilic bacteria when they were cultivated together with corn stalk than those from their pure cultures (Zhang et al., 2014b). Under the notion of developing robust enzymatic cocktails for lignocellulose biorefinery, the secreted proteins from switchgrass-fed compost microbial consortia demonstrated high levels of xylanase and endoglucanase activities, as well as a greater tolerance to heat and ionic liquid than commercial enzymes (Gladden et al., 2011). Another thermophilic ionic liquid-tolerant cellulase cocktail was also developed by Park et al. (2012) by incorporating secreted endoglucanases from

a cellulose-degrading microbial consortium to a mix of recombinant hydrolytic enzymes. Additional insights into the design of enzyme cocktails to saccharify different lignocelluloses could also be inferred by secretomics. For example, identified key secreted enzymes belonging to CAZyme family GH 10, GH51 and GH95 that were specifically abundant in microcosms fed with wheat straw and not in those fed with xylan or xylose (Jimenez et al., 2015).

The preparation of secretomes from lignocellulose-degrading microcosms should aim to comprehensively recover active enzymes that are typically secreted in very low concentrations (as described in 2.3.3) and membrane-bound cellulosomes that are attached via cohesin-dockerin interaction. Useful strategies for sample preparation, such as cell/substrate debris removal via filtration and/or centrifugation, protein concentration via ultrafiltration, are transferrable from other metasecretomic analyses (Zhang et al., 2014b; Cortes-Tolalpa et al., 2017). Additional to that, the use of eluent with high salt concentration recommended for dissociating peripheral membrane proteins (Lin and Guidotti, 2009) followed by buffer exchange represents a worthy strategy to improve the recovery of cell-bound cellulosomes along with other secreted enzymes.

Previously, we established lignocellulose-degrading microcosms from pulp mill anaerobic granules as well as digestive microbiomes of Canadian beaver and North American moose (Wong et al., 2016); pyrotag and whole genome shotgun sequencing were also applied to unveil the membership, plant polysaccharide-active CAZymes, polysaccharide utilization loci, as well as other predicted proteins that potentially degrade lignocellulose (Wong et al., 2017). In this study, secretomes from the microcosms fed with microcrystalline cellulose (Avicel) and pretreated poplar were prepared with high salt buffer to enhance the recovery of membrane-bound cellulosomes. Following that, the total polysaccharide-degrading capacity of the secretomes were

evaluated directly using enzymatic assays, with the contention that the differential saccharification activities displayed by microcosms fed with substrates with distinct complexity would provide guidance for selecting polysaccharide-degrading enzymes that are relevant to the pretreated heterogeneous wood biomass in biorefining.

6.2 Materials and Methods

Scale up of culture. The established 60 mL lignocellulose-degrading anaerobic microcosms from pulp mill anaerobic granules, beaver dropping, and moose rumen (Wong et al., 2016) were successively scaled up to 1 L cultures over 2 years in mesophilic (37 °C) conditions at pH 7 for secretome analysis. As described, biogas production was monitored at regular intervals using a pressure transducer (Omega PX725 Industrial Pressure Transmitter, Omega DP24-E Process Meter) and the cultures were transferred with fresh substrates (Table S6.1) and sulphide-reduced mineral medium (pH 7.0) (Grbić-Galić and Vogel, 1987) upon cessation of biogas production. In triplicates, the volume of the microcosms was doubled to 120 mL by the addition of 60 mL of fresh medium and lignocellulose substrates in the first phase of scaling up. In the following phase, the triplicate microcosms (120 mL each) were first combined, lignocellulose substrates and further mineral medium were then added to yield 1 L volume for each culture. In the subsequent phases, 10-20% of the culture was wasted at the end of each phase, and an equal volume of mineral medium was added along with lignocellulose substrates to prevent the accumulation of metabolic waste and maintain the culture volume.

Preparation of secretomes. After comparing different protocols for culture collection, concentration, and temperature control, the following method was used to recover secreted proteins from the microcosms within the shortest extraction time. To start, ~200 mL of cultures were collected from the scaled-up 1 L cultures when the biogas production reached early (~50% of theoretical biogas yield) and late stationary phase (~64% yield, equivalent to 80% yield by the remaining 800 mL culture after the first sampling). If the conversion plateaued before reaching 64%, the second sampling occurred at the end when the maximum conversion was achieved. After

a 30-minute centrifugation at 5,820 g at 4 °C, the culture pellets were resuspended and washed twice with ~45 mL of 50 mM HEPES buffer (pH 7.0) with 0.5 M NaCl to also recover proteins that might be bound to the residual substrate. Specifically, the pellets were gently stirred in the wash solution in a falcon tube on a rotary mixer at 4 °C for 45 minutes and the mixtures were centrifuged at 7,500 g at 4 °C for 15 minutes to allow the dissociation of proteins from the substrates into the supernatants. Supernatants recovered from the same microcosm were then combined and filtered using Nalgene rapid-flow sterile disposable bottle top filters (Nalge Nunc International, NY, USA) with 0.2 µm PES membranes on ice. The filtered supernatant was then concentrated to a final volume of ~3 mL using Jumbosep centrifugal devices (Pall Corporation, NY, USA) with 3 kDa PES membranes according to the manufacturer's instruction at 4 °C for 10 to 12 hours. The long duration required was due to membrane clogging; membranes were replaced maximally once if the flow stopped due to a complete clog. To remove the non-target solutes such as NaCl, the concentrated secretomes were buffer exchanged with 50 mM HEPES buffer (pH 7.0) using Microsep advance centrifugal device (Pall Corporation, NY, USA) with 3 k molecular weight cutoff. The resultant samples were flash-frozen in liquid N_2 and stored at –80 °C for later analyses.

Quantification of protein content. Protein contents of the secretomes were quantified based on both Bradford method (Bradford, 1976) using Bio-Rad Protein Assay (Bio-Rad Laboratories, CA, USA) and bicinchoninic acid (BCA) method using Thermo Scientific Pierce BCA Protein Assay Kit (Thermo Scientific, MA, USA) according to the manufacturer's instructions with bovine serum albumin (BSA; Thermo Scientific, MA, USA) as standard. Due to the high pigmentation of the concentrated secretomes, the secretomes were diluted to 20% to 2% in milli-Q water (MilliporeSigma, MA, USA) for a total volume of 10 µL for the assays; wells

with the equivalent diluted secretomes were included as absorbance blanks. In the second round of protein quantification, the secretome samples were treated with ethanol precipitation to remove the alcohol-soluble interfering chemicals and background buffer for the Bradford assay; other purification methods such as trichloriacetic acid/acetone extraction optimized for purifying fungal secretomic proteins (Bianco and Perrotta, 2015) was considered but not applied due to the intension of preserving the folding of the proteins for other enzyme activity assays (Koontz, 2014). In brief, 900 µL of cold anhydrous ethanol was added to 100 µL of the secretome. The mixtures were stored at -20 °C for 2 hours, followed by centrifugation at 16,000 g for 10 minutes. The supernatant was discarded and the pellet was vortexed with 400 µL of cold anhydrous ethanol. After centrifugation with the aforementioned settings, the supernatant was discarded and the pellets were dried under dry air and re-suspended in 100 µL of milli-Q water. To evaluate the presence of interfering substances in the secretomes, mixtures containing equal parts (5 µL each) of the original or ethanol-precipitated secretomes and 0.5 µg/µL BSA standard were prepared for protein quantification using the Bradford assay. Based on the protein quantification by Bradford assay, the theoretical protein concentrations of the secretomes could be calculated by

$$C_{theoretical} = (C_{secretome} + C_{BSA}) \times \frac{1}{2}$$

where

$C_{theoretical}$ is the theoretical protein concentration of the mixture containing equal parts of secretomes and BSA standard;

$C_{secretome}$ is the protein concentration of the secretomes measured by Bradford assay;

C_{BSA} is the protein concentration of the BSA standard, which was 0.5 µg/µL.

SDS-PAGE electrophoresis of secretomic proteins. Secretomic proteins (20 µL each) were separated by SDS-PAGE using 1.5 mm 15% polyacrylamide gels at 100 V for ~90 minutes; 5 µL of 10 to 170 kDa PageRuler Pre-stained Protein Ladder (Thermo Scientific, MA, USA) and 0.5 to 8 µg of BSA were also loaded. The gels were then fixed with a fixing solution (40% ethanol, 10% acetic acid) for 15 minutes, stained with QC Colloidal Coomassie stain (Bio-Rad Laboratories, CA, USA) with gentle agitation for ~16 hours, and de-stained in milli-Q water with gentle agitation for 24 hours with 3 changes of water within the first 3 hours for optimal signal to background level. Notably, the concentration of the polyacrylamide gels was decidedly increased to 15% to enhance the resolution of the protein bands between 25 and 70 kDA, as an earlier SDS-PAGE analysis (not shown) with a subset of the samples using 12% polyacrylamide gels showed a number of bands positioned between those sizes densely.

Saccharification assays. Overall hydrolytic activity (combined exo- and endo-activity) was tested on various polysaccharides (including avicel, beechwood xylan (glucuronoxylan), oat spelt xylan (arabinoglucuronoxylan), guar galactomannan, and pectin from apple purchased from Sigma–Aldrich (MO, USA), and carboxymethylcellulose, xyloglucan, wheat arabinoxylan, konjac glucomannan, and larch arabinogalactan from Megazyme (Ireland)) and assayed by the PAHBAH reducing-end assay in duplicates. The schematic representations of the polysaccharides are shown in Figure S6.1. Reaction mixtures containing 0.1% w/v polysaccharide in sodium phosphate buffer (pH 7.0) and 12.5 µL of secretomes were dispensed into polystyrene 96-well plates at 125 µL per well. The plates were then re-sealed and incubated at 37 °C with shaking at 1000 rpm for 16 hours, and the release of reducing sugars was quantified using PAHBAH assay

(Moretti and Thorson, 2008). To remove the insoluble substrates, the plates were first centrifuged at 2000 g for 5 minutes and the supernatants were transferred to new 96-well plates to remove the insoluble substrates. A working PAHBAH solution was then prepared by mixing one part of para-hydroxybenzoic acid hydrazide (5% w/w) in 0.5 M HCl with four parts of 0.5 M NaOH. Equal volumes of working solution (125 µL) were added to each reaction mix, the plates were then resealed and heated at 70 °C for 30 minutes. After cooling, absorbance was measured at 410 nm. Background absorbance due to the pigmentation of the secretomes was accounted for by including wells with 12.5 µL of diluted secretomes (usually ½) and 112.5 µL of milli-Q water as blanks. The reducing-end concentrations from xylans and non-xylan substrates were quantified using xylose and glucose calibration curves, respectively. After z-score standardization per culture, the activity data were visualized on a PCA plot using 'prcomp' and 'ggbiplot' packages based on R scripts.

6.3 Results and Discussion

6.3.1 Establishment of scaled-up cultures

The volumes of microbial enrichments were increased to ~1 L over two enrichment phases to enable direct biochemical and proteomic characterization of corresponding secretomes (Figure 6.1). Except for the cellulose-fed culture from moose rumen, the extent of COD conversion increased by up to 50% when the triplicate 120 mL microcosms were pooled to generate the 1 L cultures. Meanwhile, the length of the enrichment increased during initial phases of scaling up and then decreased after the culture volume was maintained at 1 L for all cultures excluding the pretreated poplar-fed culture from beaver droppings. Shown in Figure 6.2 was the biogas profile of the enrichment phase in which secretomes were collected from the scaled-up cultures twice. The first sampling of secretomes occurred at ~50% COD conversion as intended, while the second sampling aimed at ~80% COD conversion was not applied to all the microcosms, such as the pretreated poplar-fed ones from anaerobic granules and moose rumen, as the biogas production gradually ceased before attaining that level. Instead, secretome samples were harvested at ~60% conversion, which is equivalent to ~75% conversion of the remaining substrates. Regardless, the secretome samples still represented the repertoire of secreted proteins at midway and towards the end of the substrate conversion by the respective microcosms. Aside from sampling the secretomes on the basis of COD conversion, it could be performed when the cultures attain a certain cell density (Cortes-Tolalpa et al., 2017). While the latter would enable more direct comparison between cultures and eliminate the potential error due to the leftover COD from previous phases, the aim of this study is to capture the range of saccharification activities expressed at different extents of substrate degradation. Hence, the secretomes were sampled strategically at based on COD conversion as described.

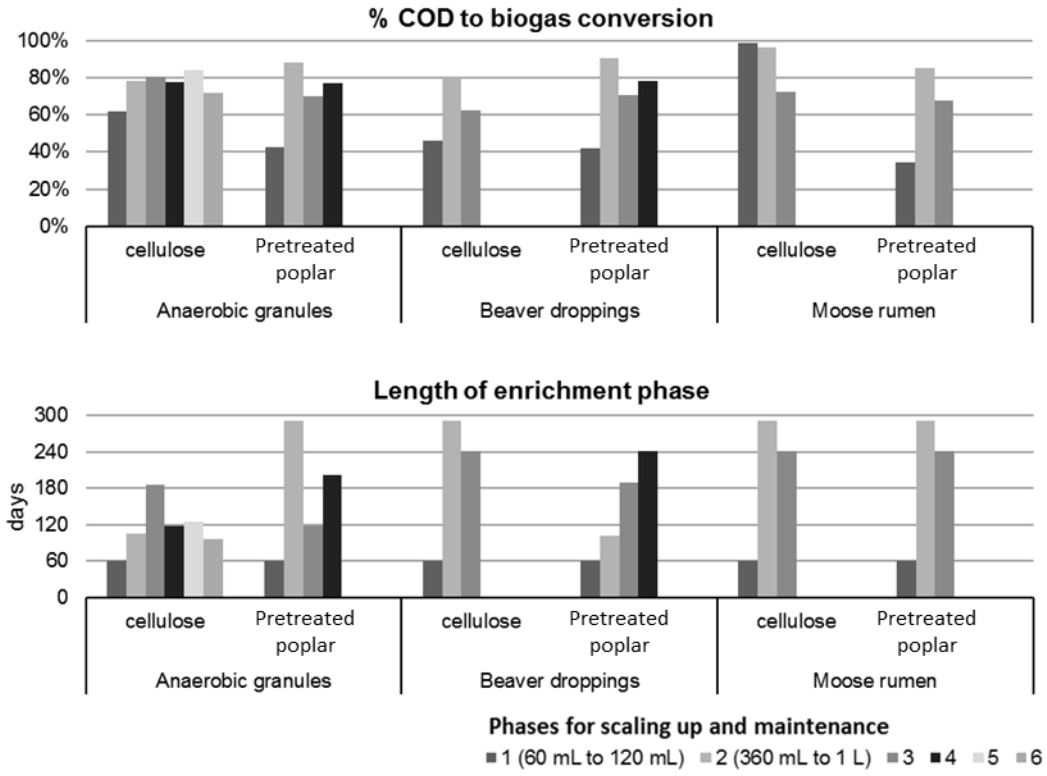


Figure 6.1 Performances of the cellulose- and pretreated poplar-degrading microcosms prior to secretomes sampling

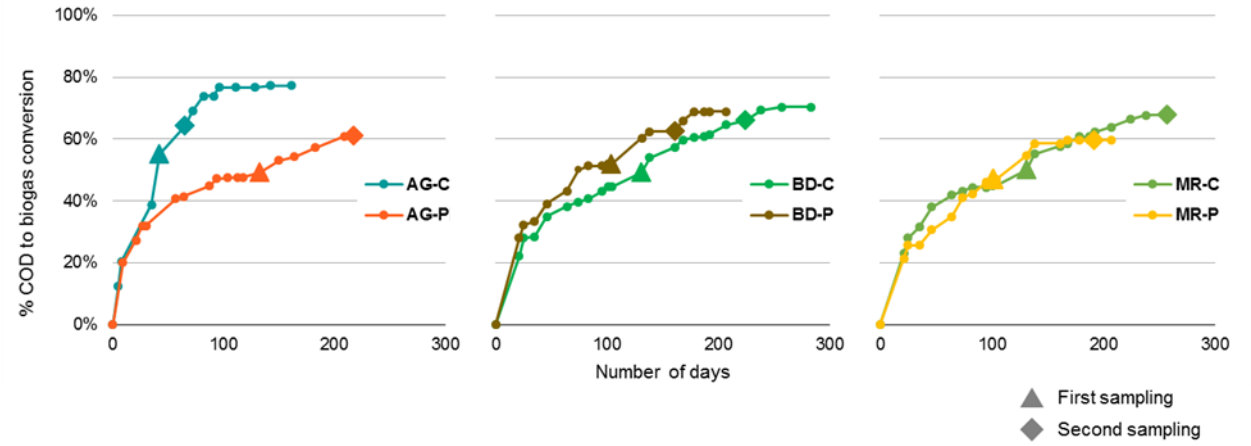


Figure 6.2 Biogas production profiles of phase 7, in which secretomes samples were harvested from the 1 L cellulose- and pretreated poplar-degrading microcosms. The

sampling was aimed at ~50% and ~80% (which is equivalent to ~64% as 80% of the culture remained after first sampling) COD conversion, indicated by the triangle and diamond markers, respectively. Pulp mill anaerobic granules, AG; beaver dropping, BD; cellulose, C; moose rumen, MR; pretreated poplar, P.

6.3.2 Estimation of protein content

Incongruence in estimated protein concentration occurred when the results generated from the Bradford and BCA assays were compared (Figure 6.3). The estimated protein concentration ranged between 0.02 and 1.13 µg/µL using the Bradford assay, while the range spanned from 0.07 to 52.1 µg/µL when the BCA assay was applied. The estimated concentrations not only varied in the numerical value, conflicting patterns also emerged when the protein concentrations of the secretomes collected from the microcosms at different COD conversion were compared. For instance, the protein concentrations of the secretomes from the cellulose- and pretreated poplar-fed microcosms from beaver droppings at 50% COD conversion were twice and 3.6 times of that at 80% respectively according to the Bradford assay. Contrarily, the results from BCA assay suggested that the protein contents of the secretomes from cellulose-fed cultures from beaver droppings were barely detected, and those collected from the pretreated poplar-fed microcosms at the two sampling points were relatively comparable with less than 25% difference. Furthermore, Bradford assay estimated a 40-times higher protein content in the secretome estimated from pretreated poplar-fed microcosm from moose rumen at 50% COD conversion than at 80%, while the BCA assay contrarily revealed an almost 6-times higher protein concentration at 80% COD conversion than at 50%. Indeed, such inconsistency was also highlighted in an earlier

attempt to quantify the protein content in complex cellulase-hemicellulase mixtures (Chundawat et al., 2011). Nonetheless, the two assays showed a similar trend in protein concentration for secretomes from other cultures. For instance, consistently low protein concentrations were estimated at both 50% and 80% COD conversions for cellulose-fed microcosms from anaerobic granules and moose rumen compared to other secretomes using either assay. Moreover, both assays agreed that the protein content in the secretome collected from the pretreated poplar-fed microcosm from anaerobic granules at 50% COD conversion was higher than that at 80%, and that it was overall the highest amongst all the prepared secretomes.

The observed incongruence in estimated protein concentration could be attributed to the fundamental differences of the underlying principles of the two assays. In Bradford method, Coomassie blue dye binds to the basic and aromatic amino acid residues (especially arginine) of proteins, and the formed dye-protein complex then absorbs light at 595 nm wavelength proportional to the amount formed (Bradford, 1976). In this context, secretomic proteins that do not share a similar fraction of basic and aromatic amino acids with the BSA standard would hence not be quantified accurately. Other potential sources of inaccuracy include the assay interference due to the presence of surfactants, polysaccharides and tannin (Compton and Jones, 1985; Marshall and Williams, 2004; Banik et al., 2009). Meanwhile, BCA assay is based on the reduction of Cu^{2+} ions by proteins to Cu^{+} ions, with which BCA forms purple-coloured chelates that absorbs light at 562 nm wavelength proportionate to the amount of protein present (Smith et al., 1985). Since the amino acid residues cysteine, cystine, tyrosine, tryptophan and the peptide backbone of proteins contribute to the reduction of Cu^{2+} ions, variability caused by the compositional differences of secretomic proteins and BSA standard was minimized. Nonetheless, reducing agents, copper chelators and high capacity buffer are known interfering agents for the

BCA assay (Brown et al., 1989). On the basis that the sample is suspended in HEPES buffer, while filtration and buffer exchange were performed to lower the content of degradation products, the results generated from the Bradford assay were deemed more reliable at this stage.

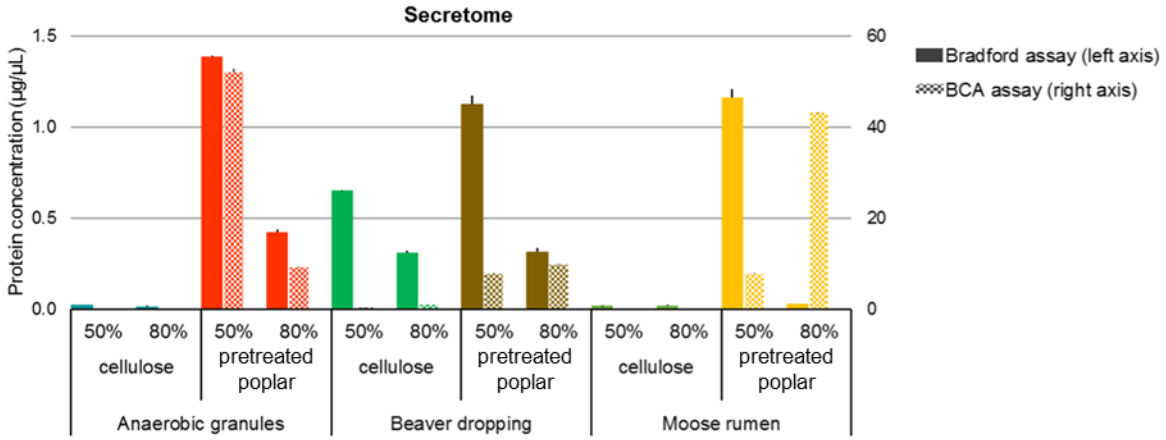


Figure 6.3 Incongruent protein content estimation using Bradford and BCA assays

To test the validity of the protein concentration estimated by Bradford assay, mixtures containing known volumes of BSA and secretomes were prepared. The presence of interfering substance in the secretomes from pretreated poplar-fed microcosms for Bradford assay was suggested by the consistently lower protein content of the mixtures containing the secretomes and BSA standard than the theoretical concentration (Figure 6.4). On average, the measured protein concentration of the secretomes-BSA mixtures from pretreated poplar-fed cultures was below 2/5 of the expected concentration. In the more extreme cases like pretreated poplar-fed microcosms from anaerobic granules (at 50% COD conversion) and moose rumen (at 80% COD conversion), the prepared secretome-BSA mixtures were estimated to have lower protein concentration than BSA alone. Additionally, the protein concentration from these pretreated poplar-enriched samples consistently increased after protein purification via ethanol precipitation except for the one from

beaver dropping microcosms at 50% COD conversion, which could be attributed to the protein loss due to the purification. Meanwhile, all secretome-BSA mixtures prepared from the cellulose enrichments attained ~20% higher than the theoretical concentrations, and only decreased slightly from that at varying degrees when the mixtures were prepared from ethanol purified secretomes. At this point, the required further investigation on the potential substrate-dependent presence of interfering substances in pretreated poplar-fed microcosms was halted due to the limited sample remaining. For future, it is worthwhile to perform protein precipitation of secretomes with BSA standards included as a control to establish a reference for the recovery rate of proteins. Afterwards, the protein concentration of the re-solubilized secretomes shall be tested at different dilutions using commercially available detergent compatible Bradford assay kit. Alternatively, other methods for protein purification can be explored. They include ammonium sulphate precipitation of proteins without denaturation (Feist and Hummon, 2015), or variations of trichloriacetic acid/acetone extraction optimized for removing interfering compounds in fungal secretome for a subset of the samples (Fragner et al., 2009; Adav et al., 2010). Moreover, Lowry protein assay represent a viable alternative quantification method that is substantially less influenced by the polyphenol content compared to Bradford assay (Redmile-Gordon et al., 2013).

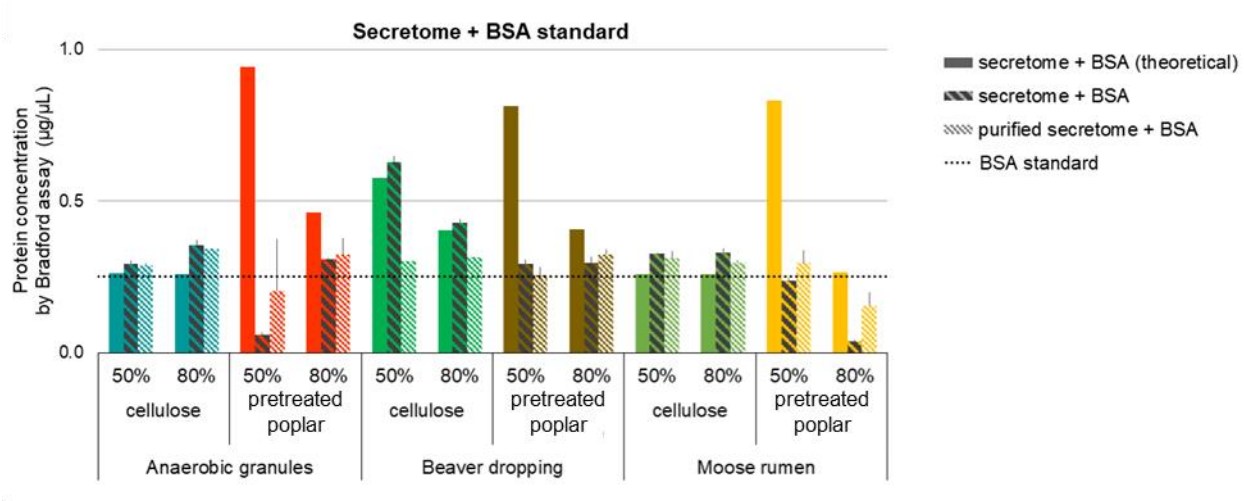


Figure 6.4 Protein quantification using Bradford assay. Presence of interfering agents in the secretomes from pretreated poplar-fed microcosms for Bradford assay was suggested by the lower protein content of the mixtures containing the secretomes and BSA standard than the theoretical protein concentration, and the increased values following the protein purification via ethanol precipitation for most samples.

6.3.3 Protein profiles of the secretomes

SDS-PAGE analysis of the concentrated culture supernatants revealed dissimilar protein profiles depending on the inocula, substrate, and to a lesser extent, the extent of COD conversion (Figure 6.5). To begin with, protein smears were consistently observed in secretomes from pretreated poplar-fed microcosms, and at a lower intensity, those collected from cellulose-fed microcosms at 80% COD conversion (Figure 6.5). The presence of smear could be reflective of the diversity of proteins secreted by the respective microflora. Note that these smears and the later described bands are considered secretomic proteins despite known interactions between Coomassie Blue and polysaccharides or tannin for multiple reasons: i) the reported complex formation has only been reported when the dye is directly mixed with the polysaccharides and tannins in protein assays (Marshall and Williams, 2004; Banik et al., 2009); ii) background interference has been consistently absent in polysaccharide-embedded (at 0.1 mg/mL) SDS-page gels that were stained with Coomassie blue (Vuong and Master, 2014; Foumani et al., 2015); iii) it remains unlikely that non-protein compounds that are between the size cutoff would be sufficiently abundant after both filtration and buffer exchange and negatively-charged for the gel electrophoresis to contribute to any bands. Moving on, the protein profiles of the secretomes from

microcosms from moose rumen appeared comparable across different substrates, with common protein bands at ~33 kDa, ~42 kDa and ~53 kDa. In contrast, the secretomes of the microcosms from anaerobic granules and beaver droppings displayed a greater variability between different substrate. Nonetheless, a protein with size ~40 kDa was present in all the secretome samples from the anaerobic granules microcosm. Next, the secretomes collected from the same microcosms at different extents of COD conversion displayed the greatest resemblance with common protein bands. Comparing the pairs collected at different extents of COD conversion, the profiles from the secretomes collected later typically featured bands with greater intensity along with some extra unique bands. Further insight could be derived from the protein profiles when compared across the microcosms from different inocula. Despite the dissimilar protein profiles across the secretomes, a common protein band at size ~37 kDa was observed. Another a protein with the size slightly smaller than 55 kDa was identified in all the secretomes from cellulose-fed microcosms at 80% COD conversion, albeit the band is strongest for the microcosm for beaver droppings, followed by moose rumen and anaerobic granules. As for the secretomes of pretreated poplar-fed microcosms at 80% COD conversion, the aforementioned protein sized ~40 kDa was actually present in all three microcosms from different inocula. Moreover, proteins with sizes > 170 kDa and ~33 kDa were present in the microcosms from anaerobic granules and beaver droppings.

The presence of smears along with the varying protein sizes forbade a reliable semi-quantitative estimation of the protein concentration via the comparison of band intensity with the BSA standards. Nonetheless, the overall intensity of the protein bands and smears partly correlated with the protein concentration estimated by the quantification assays conducted. Considering the secretomes from anaerobic granules cultures, the intensities of the protein bands

were highest for the secretomes collected from the pretreated poplar-fed culture at 50% COD conversion, followed by that at 80% COD conversion, and then the secretomes from cellulose-fed culture. Such trend was congruous with the results from the Bradford assay (Figure 6.3). Same is true for the secretomes from moose rumen, where both the protein profile shown by the SDS-Page gel and the estimation by BCA assay suggested the order of protein concentration (from high to low) to be pretreated poplar-fed microcosm at 80%, that at 50%, and the cellulose-fed microcosms.

While interesting insights regarding the diversity and quantity of proteins were enabled by the current results, further experiments are required for more substantial conclusions. To begin, the gel electrophoresis should be repeated with purified secretomes (via trichloriacetic acid/acetone extraction as mentioned in 6.3.2) to reflect, if any, the background interference due to non-protein components. Aside from that, the inclusion of a secondary ladder with a higher upper limit for molecular weight is also recommended, as some samples contained proteins with sizes beyond the 170 kDa cutoff of the current ladder. Moving on, the application of liquid chromatography-mass spectrometry (LC-MS/MS) on fractionated gels is a highly worthy avenue to pursue (Alessi et al., 2017). More specifically, the identities of the secreted proteins could be unveiled with the use of the in-house metagenomes of these microcosms, CAZy database and NCBI non-redundant protein database as references.

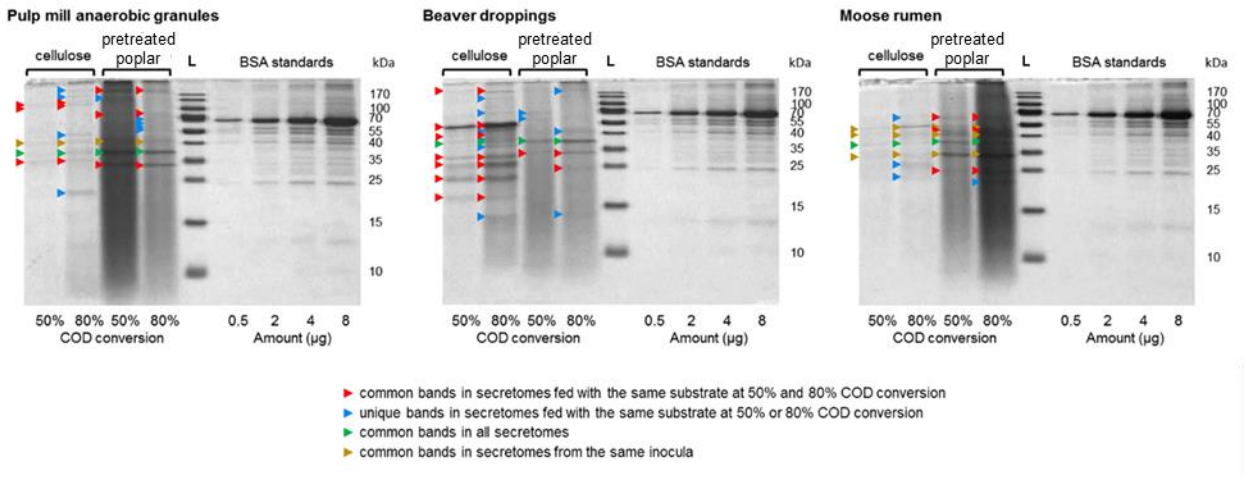


Figure 6.5 SDS-Page gel electrophoresis of secretomes (20 µL each) collected from the cellulose- and pretreated poplar-fed 1 L microcosms from pulp mill anaerobic granules, beaver droppings and moose rumen at ~50% and ~80% conversion of the fed COD in phase 7. The sizes (kDa) of the protein bands in the molecular weight ladder (L) are indicated on the right.

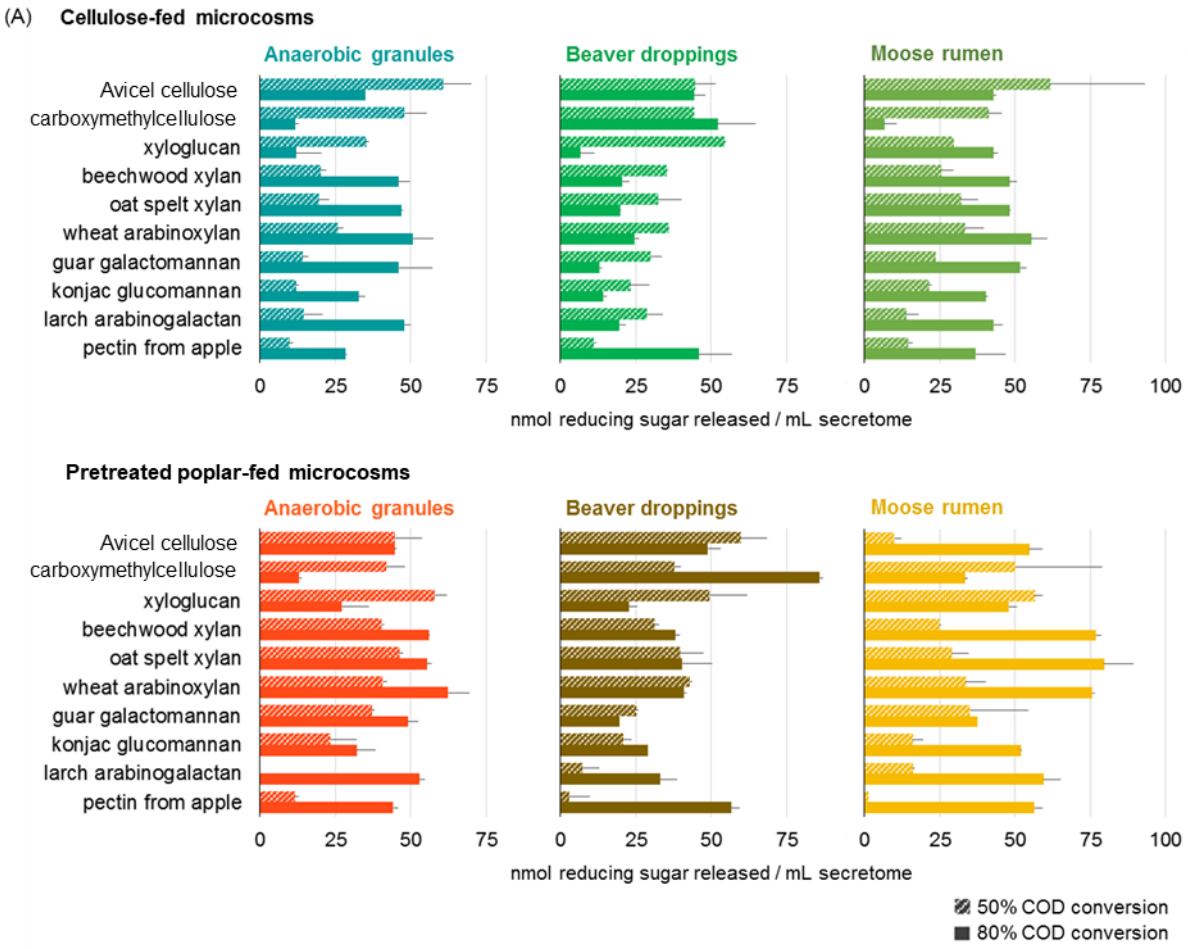
6.3.4 Saccharification capacities of the secretomes

Active across the tested polysaccharides regardless of the inocula or fed substrates (except for larch arabinogalactan for the secretomes collected from pretreated poplar-fed culture at 50% COD conversion) (Figure 6.6A), the secretomes displayed differential polysaccharide-degrading profiles primarily based on the extent of COD conversion when they were sampled (Figure 6.6B). The observed separation of secretomes collected at different % COD conversion was driven by the 3 to 37 times higher pectic degradation by the secretomes collected at 80% COD conversion than that at 50% (Figure 6.6A). Another driver of the separation was the greater activities on xylans and mannans by secretomes of cellulose and pretreated poplar-fed cultures from anaerobic

granules and moose rumen at 80% COD conversion, which are 1.2 to 3.3 times of that at 50% COD conversion (Figure 6.6A). Moreover, the secretomes collected from the cultures from anaerobic granules and moose rumen at the earlier COD conversion were marked by their greater ability to degrade carboxymethylcellulose by 1.5 to 6.2 times of those collected later (Figure 6.6A). The temporal alteration in the saccharification activities could be attributed by the succession of different microbial communities over the enrichment phase, as well as the possibility that the microflora expresses enzymes that preferentially hydrolyze glycosidic bonds between β-(1→4)-linked glucose residues before processing other polysaccharide structures. Indeed, the new protein bands emerged on the SDS-Page gel for secretomes collected at a later COD conversion might be the enzymes responsible for the added degradation activities (Figure 6.5).

A PCA analysis was also performed to further highlight distinctions between the activities of collected secretomes. These analyses again emphasized the extensive overlap between the secretomes collected at 80% COD regardless of the enrichment substrate (Figure 6.6B). The secretomes of pretreated poplar-fed cultures from anaerobic granules and moose rumen at 80% COD conversion clustered more closely, which was explained by their higher activities on xylans and mannans than those collected earlier, while those from beaver droppings clustered due to their consistently greater activity on carboxymethylcellulose at the later phase (Figure 6.6A). In comparison, less overlap occurred between the secretomes collected at 50% conversion (Figure 6.6B). Briefly, at 50% COD conversion, the secretomes of cellulose-fed cultures from anaerobic granules and moose rumen were most active on Avicel and carboxymethylcellulose; while those from cultures that were fed with pretreated poplar were both active on xyloglucans (Figure 6.6A). While the comparison of activities was made on the basis of volume of secretome added as the

further work is required for the quantification of protein content, the divergent saccharification profiles of secretomes from the scaled-up microcosms at different COD conversion showcased the success of the protocol developed in recovering functionally-active secretomes. The requirement for multiple activity assessments over the enrichment phase, as opposed to the single assessments reported in other metasecretomic analyses (Gladden et al., 2011; Jimenez et al., 2015), is also highlighted for a more complete understanding of the enzymatic successions expressed. In the context of enzyme discovery, such a temporal functional profile will guide the selection of secretomes from those collected at different time points for downstream LC-MS/MS analysis, and thereby increase the likelihood to unveil enzymes that correspond to the desired activities.



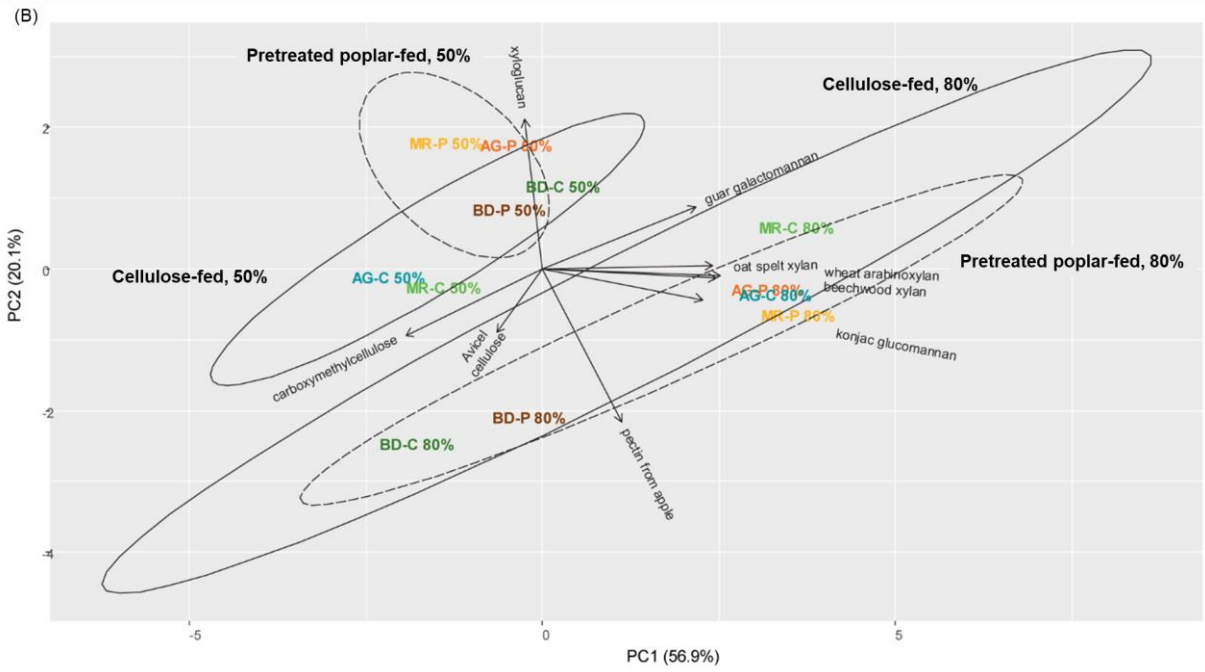


Figure 6.6 (A) Polysaccharide-degrading activities of the secretomes measured by nmol reducing sugar released per mL secretomes after 16-hour incubation at 37 °C, pH 7. (B) PCA biplot of the standardized saccharification activities of the secretomes. Ellipses show 95% confidence. C: cellulose, P: pretreated poplar.

6.4 Conclusions and future steps

To unveil the temporal dynamic in the polysaccharide-degrading capacity, the cellulose and pretreated poplar enrichments were successfully scaled up to ~1 L over two enrichment phases for the direct proteomic characterization of secretomes sampled at 50% and 80% COD conversion. While further improvement of protein purification is warranted due the presence of interfering substance to Bradford and BCA protein quantification assays, SDS-PAGE analysis of the supernatant revealed divergent protein profiles depending on the inocula, substrate, and to a lesser extent, the extent of COD conversion. Herein, the emergence of new bands in the secretomes collected at the later COD conversion coincided with their greater saccharification activities on pectin, xylans and mannans, suggesting a temporal dynamic in the enzyme expression over the enrichment phase.

Directions for future experiments could be devised based on the present results. To begin, further protein purification and quantification methods shall be explored to generate a reliable estimated protein content to provide a basis for the activity assays. Another obvious next step would be the identification of proteins present in the secretomes via LC-MS/MS using the various database as aforementioned; corresponding efforts have already been initiated. Concerning results from the saccharification assay, the exhibition of xylanase, mannanase, and pectinase activities from cultures that were exclusively enriched with cellulose was expected since catalytic domains with these activities were annotated via metagenomics analysis, though more frequent in cultures fed with pretreated poplar. In addition to that, functional promiscuity of enzymes within families or superfamilies is not uncommon in nature (Nobeli et al., 2009). As revealed by the active-site architecture analysis by Tian et al. (2016), the partial recognition between the active sites and substrates underlying functional promiscuity is common in families with fewer conserved amino

acid residues. Moreover, the ability for a carbohydrate-active enzyme to accommodate a diverse range of substrates could be facilitated by the presence of multiple domains (Nobeli et al., 2009), which were found to be plentiful from the metagenomics analysis. Nonetheless, the similar polysaccharide-degrading profiles of the secretomes collected from cultures fed with different substrates could potentially be a reflection of an insufficient dissociation of enzymes from the culture pellets (Feiz et al., 2006). In this study, a mild washing condition using high salt buffer was employed to liberate the lignocellulases and cell-bound cellulosomes from the culture pellets since more rigorous extraction methods increase the risk of sample contamination with intracellular proteins. To circumvent this problem, the supernatant-derived metasecretomes could be coupled with meta-surface-proteome (Alessi et al., 2017), wherein the surface-bound proteins are labelled with biotin tags, extracted with SDS, precipitated, purified using streptavidin sepharose media, and then analyzed by LC-MS/MS as described.

The execution of the enzymatic assays could also be readily improved in several ways to improve the resolution and accuracy of the lignocellulolytic activities. To show the reproducibility of the secretomes as well as a complete picture of their polysaccharide-degrading capacities, the secretomes harvested from replicate cultures should be assayed at more frequent time points across successive enrichment phases. Next, the use of vacuum filter unit with liquid handler could also be applied to prevent inaccuracy of absorbance reading due to the insoluble polysaccharides. Moreover, potential synergism with commercial enzyme cocktails could also be investigated by comparing the saccharification activities of the cocktails alone and cocktails with different dose of secretome proteins added. Last but not the least, non-reducing sugar releasing modification of polysaccharides after treatment with secretomes could be detected by high-performance anion exchange chromatography (HPAEC) and time-of-flight secondary ion mass spectrometry (ToF-

SIMS). The former is able to analyze the released mono- and oligosaccharides (Jurak et al., 2014), while the latter will allow an in-depth analysis of the surface modification (Goacher et al., 2013).

Chapter 7 Conclusions, engineering significance and future directions

7.1 Conclusions

To enable the diversification of value-added bioproducts from wood lignocelluloses, there is a need to explore microbiomes beyond those that degrade grasses and crops for new carbohydrate-active enzymes that could transform wood fibres that contain potent fermentation inhibitors from steam explosion pretreatment. There is also the need to assay microbes and enzymes on real, industrially-relevant substrates and substrate mixes.

Accordingly, multiple wood lignocellulose enrichments, including (i) cellulose, (ii) cellulose + lignosulphonate, (iii) cellulose + tannic acid, and (iv) pretreated poplar were cultured over 3 years. Subsequent pyrotag analysis of the 16S rRNA gene markers revealed a decrease in species richness after enrichment and a substrate-driven convergence of microbial composition which encompassed known biomass degraders along with several poorly defined lineages at high abundances. Focusing on the microcosms fed with more complex lignocelluloses (ii to iv), those uncharacterized microbes were grouped in *OPB54* order and *SJA-28* class in the gut microbiome microcosms amended with cellulose + lignosulphonate and cellulose + tannic acid, as well as *Cloacamonales* order, *TG3* and *BSV26* classes from the pulp mill anaerobic granules microcosms fed with cellulose + lignosulphonate and pretreated poplar. While functionally and taxonomically not well-defined, these lineages were found in habitats (such as termite gut, forest sediment, anaerobic bioreactor for energy crops and sludge) that involve anaerobic transformation of biomass. Overall, the observed substrate-driven convergence and proliferation of known and

potential biomass degraders confirms the simplification of community composition via enrichment, which in terms would facilitate the discovery of wood-degrading enzymes.

Downstream comparative metagenomic analysis targeting microcosms amended with cellulose and pretreated poplar enabled the identification of enzymes that are pertinent to wood fibre bioprocessing. Hierarchical clustering analysis of plant polysaccharide-active CAZyme families distinguished those found in the gut microbiomes from those reported in grass-feeding mammals, mixed plants foragers and the rumen from a wild moose (i.e. without enrichment), highlighting the impact of substrate enrichment. Meanwhile, the impact of the inocula was also signified by the greater similarity between the microcosms from the gut microbiomes when compared with those from pulp mill anaerobic granules. In addition to the broad substrate-active families GH2, GH3, GH5, and GH43 with cellulolytic and hemicellulolytic activities reported in earlier studies, the pretreated poplar-fed microcosms from gut microbiomes were comparatively enriched in sequences assigned to putative β-L-arabinofuranosidases from family GH127 and 4-O-methyl-glucuronoyl methylesterases from family CE15, whereas the pretreated poplar-fed microbiomes derived from pulp mill anaerobic granules were enriched in putative pectinolytic enzymes from families PL1, GH28, CE8, and GH105. First established in this study, the comprehensive screening further unveiled 365 catalytic cellulosomal sub-units, 861 proteins with unknown function positioned within the 416 predicted PULs and 72 proteins containing both a CBM and a DUF. While correlations between the amendment substrate and the catalytic cellulosomal subunits, as well as the proteins with unknown function and CAZy membership within the PUL are absent, two putative carbohydrate binding proteins with domain of unknown function (namely DUF3459-CBM48-GH13_10 and DUF5011-CBM4) are consistently present in microcosms fed with pretreated poplar only.

Metasecretomic analysis was also applied to the scaled up 1 L cellulose and pretreated poplar enrichments at different % COD conversion to decipher the functional fraction of polysaccharide-degrading CAZymes from the metagenomic prediction, and to capture the activities of catalytic domains that were not well-catalogued in the current database. While the anticipated substrate-based differences between the metasecretomes were only reflected on the protein profiles via the SDS-PAGE analysis, the emergence of new bands in the metasecretomes collected and greater saccharification activities on pectin, xylans and mannans at the later COD conversion implied a succession of microbial enzymes that preferentially hydrolyze glycosidic bonds between β-(1→4)-linked glucose residues before degrading other polysaccharide structures. Herein, the established protein concentration protocol that retains enzyme activity of secretomes from mixed cultures could be applied in other mixed cultures.

Collectively, the potential of lignocellulose-degrading microcosms serving as a repertoire of polysaccharide-degrading biocatalysts was confirmed. The comparative meta-analyses of the microcosms enriched with strategically chosen substrates revealed putative lignocellulolytic taxa and proteins specialized for processing forest fibre were revealed for downstream genomic and functional characterization. Furthermore, the functionally-active metasecretomes provide opportunities to determine the protein composition as reference for developing specialized enzymatic cocktails for pretreated wood fibres.

7.2 Engineering significance

1. New protein candidates with potential industrial relevance

A total of 12,962 CAZymes, 365 catalytic cellulosomal sub-units, 72 proteins containing both a CBM and a DUF, 861 proteins with unknown function positioned within the 416 predicted PULs were annotated from the lignocellulose enrichments. Of these, putative β-L-arabinofuranosidases (GH127), 4-O-methyl-glucuronoyl methylesterases (CE15), pectinases (PL1, GH28, CE8, and GH105), as well as two putative carbohydrate binding proteins with domains of unknown function (DUF3459-CBM48-GH13_10 and DUF5011-CBM4) were especially highlighted via the comparative analysis for their potential to improve enzymatic conversion of wood fibre.

2. Poorly characterized microbes highlighted for lignocellulose degradation

While the community analysis was conducted to validate the simplification of community composition via enrichment for the downstream discovery of wood-degrading enzymes within the scope of this thesis, the use of microbes for potential biological pretreatment represent an area of research focus for benefits such as low chemical and energy input (Zhuo et al., 2018). Therefore, the taxonomically and functionally less defined microbial lineages belonging to *BSV26*, *SJA-28*, *TG3* classes, *OPB54* and *Cloacamonales* orders that were dominant in the enrichments represent potential indicators for lignocellulose degradation or targets for further characterization. These microbes could potentially encode specialized catabolic activities relevant to pretreatment and detoxification of pretreated wood, or enhance the overall lignocellulolytic activities via non-catalytic secretary proteins.

3. **Blueprint for designing enzymatic cocktails for pretreated wood fibres**

   The potential of developing lignocellulose-degrading microcosms to serve as a repertoire of polysaccharide-degrading enzymes was validated. Moreover, the functionally-active metasecretomes provide opportunities to determine the identity and the relative fractions of the protein components for the development of enzymatic cocktails specialized for processing pretreated wood fibres.

4. **Metasecretomic analysis of microbial enrichments**

   The established metasecretome preparation method that retains the enzyme activity of secretomes from mixed cultures could be transferred to the investigations of other mixed cultures, such as those that degrade solid wood. Moreover, the unveiled differences in the capacity in polysaccharide degradation over the enrichment phase supported the need to perform temporal measurements of the enzyme activities for a complete functional profile. Together with the protein identification via LC-MS/MS, the functional screening will effectively narrow the number of candidates from the vast number of predicted polysaccharide-degrading enzymes for downstream characterization.

7.3 Future directions

To date, 24 protein candidates are selected for functional characterization as part of the Symbiomics project on the basis of the module organization, predicted substrate specificity inferred by the CBM family, putative activity inferred by the CAZyme family, existent descriptions about the DUFs, positions within the predicted PULs, as well as similarity to functionally-established enzymes based on the sequence identity and phylogeny (Table S7.1). In addition to that, proteomic data from the LC-MS/MS analysis of the secretome samples collected prior to the scale-up of culture were obtained for protein annotation based on the in-house metagenomes, CAZy database and NCBI non-redundant protein database; corresponding analyses are ongoing.

The following summarizes other worthy investigations to pursue:

1. **Characterization of genome bins from the metagenomes**

   27 quality-screened metagenome-assembled genomes (Bowers et al., 2017; Parks et al., 2017) were generated from the quality-trimmed metagenomic reads from Trimmomatic 0.32 (Bolger et al., 2014) and AfterQC 0.9.6 (Chen et al., 2017) using Metabat 2.12 (Kang et al., 2015) and Check M (Parks et al., 2015) (Table S7.2). Further characterization of the phylogeny, CAZyme profiles, and the metabolic pathways in carbohydrate utilization will provide insight into their potential application in bioprocessing (Sun et al., 2018). In addition to that, the mapping of PULs could be improved as some components might be missing due to assembly limitation.

2. **Identification of protein identities from the secretomes**

Complete secretome profiles will be achieved by applying LC-MS/MS on supernatant-derived secretomes and substrate-bound proteins, wherein the latter can be obtained via biotin tags and SDS extraction (Alessi et al., 2017). The use of the in-house metagenomes, CAZy database and NCBI non-redundant protein database as references would then unveil the identities of the secreted proteins. Metatranscriptomics analysis is also proposed herein as an alternative avenue to generate a database of transcribed genes from the scaled-up cultures; the fraction that encodes secretary proteins could then be determined by the presence of signal protein. In both cases, secretory proteins that are uniquely present or abundant in pretreated poplar-fed microcosms shall then become targets for recombinant expression and functional characterization.

3. **Further experiments with saccharification assays**
    a. The secretomes harvested from replicate cultures should be assayed at more frequent time points across successive enrichment phases to yield a reproducible and complete polysaccharide-degrading activity profile.
    b. Incorporation of vacuum filter unit with liquid handler could be applied to remove the insoluble substrates to improve precision and accuracy of absorbance reading.
    c. The saccharification activities of commercial cocktails on complex biomass substrates alone and with different dosage of secretome proteins. This comparative analysis could unveil the presence of novel proteins that boost the activity of known enzymes.
4. **Characterization of non-reducing sugar releasing modification using *p*NP substrates**

To profile the exo-activities of the secretomes, cleavage of the *p*NP-linked substrates by the catalytic enzymes present could be quantitatively measured by comparing the

absorbance at 410 nm with a *p*NP standard curve after an incubation at 36 °C for 2 to 4 hours.

5. **Characterization of polysaccharides after secretomic treatment**
    a. The mono- and oligosaccharides released from polysaccharides could be detected by HPAEC using appropriate standards (Jurak et al., 2014)
    b. Compositional variations in terms of the lignin, hemicellulose, and cellulose content before and after the secretomic treatment could be determined by Fourier transform infrared spectroscopy (Xu et al., 2013) or nuclear magnetic resonance (Tarmadi et al., 2018)
    c. In-depth spectral and spatial analysis of the surface modification could be unveiled using ToF-SIMS (Goacher et al., 2013; Tolbert and Ragauskas, 2017). Sample surfaces will be ionized with a primary ion beam, and the time of flight and mass-to-charge ratio of the secondary ion fragments will be analyzed to provide chemical information.

References

Abbott, D.W., and Boraston, A.B. (2007). The structural basis for exopolygalacturonase activity in a family 28 glycoside hydrolase. *J Mol Biol* 368(5)**,** 1215-1222. doi: 10.1016/j.jmb.2007.02.083.

Abbott, D.W., Hrynuik, S., and Boraston, A.B. (2007). Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterolitica*. *J Mol Biol* 367(4)**,** 1023-1033. doi: 10.1016/j.jmb.2007.01.030.

Abou Hachem, M., Karlsson, E.N., Bartonek-Roxa, E., Raghothama, S., Simpson, P.J., Gilbert, H.J., et al. (2000). Carbohydrate-binding modules from a thermostable *Rhodothermus marinus* xylanase: cloning, expression and binding studies. *Biochem J* 345**,** 53-60. doi: 10.1042/0264-6021:3450053.

Adams, A.S., Jordan, M.S., Adams, S.M., Suen, G., Goodwin, L.A., Davenport, K.W., et al. (2011). Cellulose-degrading bacteria associated with the invasive woodwasp *Sirex noctilio*. *ISME J* 5(8)**,** 1323-1331. doi: 10.1038/ismej.2011.14.

Adav, S.S., Li, A.A., Manavalan, A., Punt, P., and Sze, S.K. (2010). Quantitative iTRAQ secretome analysis of *Aspergillus niger* reveals novel hydrolytic enzymes. *J Proteome Res* 9(8)**,** 3932-3940. doi: 10.1021/pr100148j.

Agger, J.W., Busk, P.K., Pilgaard, B., Meyer, A.S., and Lange, L. (2017). A new functional classification of glucuronoyl esterases by peptide pattern recognition. *Front Microbiol* 8. doi: 10.3389/fmicb.2017.00309.

Agger, J.W., Isaksen, T., Várnai, A., Vidal-Melgosa, S., Willats, W.G., Ludwig, R., et al. (2014). Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proc Natl Acad Sci U S A* 111(17)**,** 6287-6292.

Ahmad, M., Roberts, J.N., Hardiman, E.M., Singh, R., Eltis, L.D., and Bugg, T.D.H. (2011). Identification of DypB from *Rhodococcus jostii* RHA1 as a lignin peroxidase. *Biochemistry* 50(23)**,** 5096-5107. doi: 10.1021/bi101892z.

Ahvazi, B., Radiotis, T., Bouchard, J., and Goel, K. (2007). Chemical pulping of steam-exploded mixed hardwood chips. *J Wood Chem Technol* 27(2)**,** 49-63. doi: 10.1080/02773810701486618.

Al-Dajani, W.W., Tschirner, U.W., and Jensen, T. (2009). Pre-extraction of hemicelluloses and subsequent kraft pulping Part II: Acid- and autohydrolysis. *Tappi J* 8(9)**,** 30-37.

Al-Masaudi, S., El Kaoutari, A., Drula, E., Al-Mehdar, H., Redwan, E.M., Lombard, V., et al. (2017). A metagenomics investigation of carbohydrate-active enzymes along the gastrointestinal tract of saudi sheep. *Front Microbiol* 8. doi: 10.3389/fmicb.2017.00666.

Alessi, A.M., Bird, S.M., Bennett, J.P., Oates, N.C., Li, Y., Dowle, A.A., et al. (2017). Revealing the insoluble metasecretome of lignocellulose-degrading microbial communities. *Sci Rep* 7(1)**,** 2356. doi: 10.1038/s41598-017-02506-5.

Allgaier, M., Reddy, A., Park, J.I., Ivanova, N., D'haeseleer, P., Lowry, S., et al. (2010). Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS One* 5(1). doi: 10.1371/Journal.Pone.0008812.

Amezcua-Allieri, M.A., Duran, T.S., and Aburto, J. (2017). Study of chemical and enzymatic hydrolysis of cellulosic material to obtain fermentable sugars. *J Chem*. doi: 10.1155/2017/5680105.

Anitori, R.P. (2012). *Extremophiles: Microbiology and Biotechnology*. Caister Academic Press.

Araki, T., Hashikawa, S., and Morishita, T. (2000). Cloning, sequencing, and expression in *Escherichia coli* of the new gene encoding beta-1,3-xylanase from a marine bacterium, *Vibrio* sp. strain XY-214. *Appl Environ Microbiol* 66(4)**,** 1741-1743.

Artzi, L., Bayer, E.A., and Morais, S. (2017). Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nat Rev Microbiol* 15(2)**,** 83-95. doi: 10.1038/nrmicro.2016.164.

Artzi, L., Morag, E., Barak, Y., Lamed, R., and Bayer, E.A. (2015). *Clostridium clariflavum*: key cellulosome players are revealed by proteomic analysis. *Mbio* 6(3). doi: 10.1128/mBio.00411-15.

Ashrafi, O., Yerushalmi, L., and Haghighat, F. (2015). Wastewater treatment in the pulp-and-paper industry: A review of treatment processes and the associated greenhouse gas emission. *J Environ Manage* 158**,** 146-157. doi: 10.1016/j.jenvman.2015.05.010.

Asiegbu, F.O., Paterson, A., Morrison, I.M., and Smith, J.E. (1995). Effects of cell wall phenolics and fungal metabolites on methane and acetate production under in vitro rumen conditions. *J Gen Appl Microbiol* 41(6)**,** 475-485. doi: 10.2323/Jgam.41.475.

Aspeborg, H., Coutinho, P.M., Wang, Y., Brumer, H., and Henrissat, B. (2012). Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol* 12. doi: 10.1186/1471-2148-12-186.

Avellar, B.K., and Glasser, W.G. (1998). Steam-assisted biomass fractionation. I. Process considerations and economic evaluation. *Biomass Bioenerg* 14(3)**,** 205-218. doi: 10.1016/S0961-9534(97)10043-5.

Bagenholm, V., Reddy, S.K., Bouraoui, H., Morrill, J., Kulcinskaja, E., Bahr, C.M., et al. (2017). Galactomannan catabolism conferred by a polysaccharide utilization locus of bacteroides ovatus- enzyme synergy and crystal structure of a beta-mannanase. *J Biol Chem* 292(1)**,** 229-243. doi: 10.1074/jbc.M116.746438.

Bajpai, P. (2004). Biological bleaching of chemical pulps. *Crit Rev Biotechnol* 24(1)**,** 1-58. doi: 10.1080/07388550490465817.

Bajwa, P.K., Shireen, T., D'Aoust, F., Pinel, D., Martin, V.J.J., Trevors, J.T., et al. (2009). Mutants of the pentose-fermenting yeast pichia stipitis with improved tolerance to inhibitors in hardwood spent sulfite liquor. *Biotechnol Bioeng* 104(5)**,** 892-900. doi: 10.1002/bit.22449.

Bakuzis, E.V., and Hansen, H.L. (1965). *Balsam fir: A monographic review*. University of Minnesota Press.

Banerjee, N., Bhatnagar, R., and Viswanathan, L. (1981). Inhibition of glycolysis by furfural in *Saccharomyces cerevisiae*. *Eur J Appl Microbiol* 11(4)**,** 226-228. doi: 10.1007/Bf00505872.

Banik, S.P., Pal, S., Ghorai, S., Chowdhury, S., and Khowala, S. (2009). Interference of sugars in the Coomassie Blue G dye binding assay of proteins. *Anal Biochem* 386(1)**,** 113-115. doi: 10.1016/j.ab.2008.12.006.

Barry, V.C., and Mitchell, P.W.D. (1955). Mechanism of osazone formation. *Nature* 175(4448)**,** 220-220. doi: 10.1038/175220a0.

Bateman, A., and Bycroft, M. (2000). The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *J Mol Biol* 299(4)**,** 1113-1119. doi: 10.1006/jmbi.2000.3778.

Batista-Garcia, R.A., Sanchez-Carbente, M.D., Talia, P., Jackson, S.A., O'Leary, N.D., Dobson, A.D.W., et al. (2016). From lignocellulosic metagenomes to lignocellulolytic genes:

trends, challenges and future prospects. *Biofuel Bioprod Bior* 10(6)**,** 864-882. doi: 10.1002/bbb.1709.

Baumann, M.J., Eklof, J.M., Michel, G., Kallas, A.M., Teeri, T.T., Czjzek, M., et al. (2007). Structural evidence for the evolution of xyloglucanase activity from xyloglucan endo-transglycosylases: Biological implications for cell wall metabolism. *Plant Cell* 19(6)**,** 1947-1963. doi: 10.1105/tpc.107.051391.

Bayer, E.A. (2017). Cellulosomes and designer cellulosomes: why toy with Nature? *Env Microbiol Rep* 9(1)**,** 14-15. doi: 10.1111/1758-2229.12489.

Bayer, E.A., Belaich, J.P., Shoham, Y., and Lamed, R. (2004). The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol* 58**,** 521-554. doi: 10.1146/annurev.micro.57.030502.091022.

Bayer, E.A., Lamed, R., White, B.A., and Flint, H.J. (2008). From cellulosomes to cellulosomics. *Chem Rec* 8(6)**,** 364-377. doi: 10.1002/Tcr.20160.

BCC Research (2017). "Biorefinery products: global markets". (MA, USA).

Bensoussan, L., Morais, S., Dassa, B., Friedman, N., Henrissat, B., Lombard, V., et al. (2017). Broad phylogeny and functionality of cellulosomal components in the bovine rumen microbiome. *Environ Microbiol* 19(1)**,** 185-197. doi: 10.1111/1462-2920.13561.

Bhatta, R., Uyeno, Y., Tajima, K., Takenaka, A., Yabumoto, Y., Nonaka, I., et al. (2009). Difference in the nature of tannins on in vitro ruminal methane and volatile fatty acid production and on methanogenic archaea and protozoal populations. *J Dairy Sci* 92(11)**,** 5512-5522. doi: 10.3168/jds.2008-1441.

Bianco, L., and Perrotta, G. (2015). Methodologies and Perspectives of Proteomics Applied to Filamentous Fungi: From Sample Preparation to Secretome Analysis. *Int J Mol Sci* 16(3)**,** 5803-5829. doi: 10.3390/ijms16035803.

Biely, P. (2016). Microbial glucuronoyl esterases: 10 years after discovery. *Appl Environ Microbiol* 82(24)**,** 7014-7018. doi: 10.1128/Aem.02396-16.

Biely, P., Malovikova, A., Uhliarikova, I., Li, X.L., and Wong, D.W.S. (2015). Glucuronoyl esterases are active on the polymeric substrate methyl esterified glucuronoxylan. *FEBS Lett* 589(18)**,** 2334-2339. doi: 10.1016/j.febslet.2015.07.019.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15)**,** 2114-2120. doi: 10.1093/bioinformatics/btu170.

Boraston, A.B., Bolam, D.N., Gilbert, H.J., and Davies, G.J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* 382**,** 769-781.

Boraston, A.B., Warren, R.A.J., and Kilburn, D.G. (2001). beta-1,3-glucan binding by a thermostable carbohydrate-binding module from *Thermotoga maritima*. *Biochemistry* 40(48)**,** 14679-14685. doi: 10.1021/bi015760g.

Borsenberger, V., Dornez, E., Desrousseaux, M.L., Massou, S., Tenkanen, M., Courtin, C.M., et al. (2014). A H-1 NMR study of the specificity of alpha-L-arabinofuranosidases on natural and unnatural substrates. *Bba-Gen Subjects* 1840(10)**,** 3106-3114. doi: 10.1016/j.bbagen.2014.07.001.

Boutros, S. (2012). Up-to-date insight on industrial enzymes applications and global market. *J Bioprocess Biotech*.

Bouws, H., Wattenberg, A., and Zorn, H. (2008). Fungal secretomes--nature's toolbox for white biotechnology. *Appl Microbiol Biotechnol* 80(3)**,** 381-388. doi: 10.1007/s00253-008-1572-5.

Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35(8)**,** 725-731. doi: 10.1038/nbt.3893.

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

Brown, R.E., Jarvis, K.L., and Hyland, K.J. (1989). Protein measurement using bicinchoninic acid - elimination of interfering substances. *Anal Biochem* 180(1)**,** 136-139. doi: 10.1016/0003-2697(89)90101-2.

Brulc, J.M., Antonopoulos, D.A., Miller, M.E.B., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A., et al. (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. doi: 10.1073/pnas.0806191105.

Brune, A. (2014). Symbiotic digestion of lignocellulose in termite guts. *Nat Rev Microbiol* 12(3)**,** 168-180. doi: 10.1038/nrmicro3182.

Bugg, T.D.H., and Rahmanpour, R. (2015). Enzymatic conversion of lignin into renewable chemicals. *Curr Opin Chem Biol* 29**,** 10-17. doi: 10.1016/j.cbpa.2015.06.009.

Buist, G., Steen, A., Kok, J., and Kuipers, O.R. (2008). LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol Microbiol* 68(4)**,** 838-847. doi: 10.1111/j.1365-2958.2008.06211.x.

Burrell, P.C., O'Sullivan, C., Song, H., Clarke, W.P., and Blackall, L.L. (2004). Identification, detection, and spatial resolution of *Clostridium* populations responsible for cellulose degradation in a methanogenic landfill leachate bioreactor. *Appl Environ Microbiol* 70(4)**,** 2414-2419. doi: 10.1128/Aem.70.4.2414-2419.2004.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5)**,** 335-336. doi: 10.1038/Nmeth.F.303.

Cara, C., Ruiz, E., Ballesteros, I., Negro, M.J., and Castro, E. (2006). Enhanced enzymatic hydrolysis of olive tree wood by steam explosion and alkaline peroxide delignification. *Process Biochem* 41(2)**,** 423-429. doi: 10.1016/j.procbio.2005.07.007.

Caramelo, L., Martinez, M.J., and Martinez, A.T. (1999). A search for ligninolytic peroxidases in the fungus *Pleurotus eryngii* involving alpha-keto-gamma-thiomethylbutyric acid and lignin model dimers. *Appl Environ Microbiol* 65(3)**,** 916-922.

Carroll, A., and Somerville, C. (2009). Cellulosic biofuels. *Annu Rev Plant Biol* 60**,** 165-182. doi: 10.1146/annurev.arplant.043008.092125.

Cartmell, A., Topakas, E., Ducros, V.M.A., Suits, M.D.L., Davies, G.J., and Gilbert, H.J. (2008). The *Cellvibrio japonicus* mannanase cjman26c displays a unique exo-mode of action that is conferred by subtle changes to the distal region of the active site. *J Biol Chem* 283(49)**,** 34403-34413. doi: 10.1074/jbc.M804053200.

Caufrier, F., Martinou, A., Dupont, C., and Bouriotis, V. (2003). Carbohydrate esterase family 4 enzymes: substrate specificity. *Carbohydr Res* 338(7)**,** 687-692.

Chakar, F.S., and Ragauskas, A.J. (2004). Review of current and future softwood kraft lignin process chemistry. *Ind Crop Prod* 20(2)**,** 131-141. doi: 10.1016/j.indcrop.2004.04.016.

Chandra, R., and Singh, R. (2012). Decolourisation and detoxification of rayon grade pulp paper mill effluent by mixed bacterial culture isolated from pulp paper mill effluent polluted site. *Biochem Eng J* 61**,** 49-58. doi: 10.1016/j.bej.2011.12.004.

Chaney, W.R. 2003. Why do animals eat the bark and wood of trees and shrubs? Available: https://www.extension.purdue.edu/extmedia/fnr/fnr_203.pdf [Accessed October 8, 2017].

Charnock, S.J., Brown, I.E., Turkenburg, J.P., Black, G.W., and Davies, G.J. (2002). Convergent evolution sheds light on the anti-beta-elimination mechanism common to family 1 and 10 polysaccharide lyases. *Proc Natl Acad Sci U S A* 99(19)**,** 12067-12072. doi: 10.1073/pnas.182431199.

Chen, S., Huang, T., Zhou, Y., Han, Y., Xu, M., and Gu, J. (2017). AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinformatics* 18(Suppl 3)**,** 80. doi: 10.1186/s12859-017-1469-3.

Chen, W., Zhang, C.K., Cheng, Y.M., Zhang, S.W., and Zhao, H.Y. (2013). A Comparison of methods for clustering 16S rRNA sequences into OTUs. *PLoS One* 8(8). doi: 10.1371/journal.pone.0070837.

Cherubini, F., Jungmeier, G., Wellisch, M., Willke, T., Skiadas, I., Van Ree, R., et al. (2009). Toward a common classification approach for biorefinery systems. *Biofuel Bioprod Bior* 3(5)**,** 534-546. doi: 10.1002/bbb.172.

Chouari, R., Le Paslier, D., Dauga, C., Daegelen, P., Weissenbach, J., and Sghir, A. (2005). Novel major bacterial candidate division within a municipal anaerobic sludge digester. *Appl Environ Microbiol* 71(4)**,** 2145-2153. doi: 10.1128/Aem.71.4.2145-2153.2005.

Chundawat, S.P.S., Lipton, M.S., Purvine, S.O., Uppugundla, N., Gao, D.H., Balan, V., et al. (2011). Proteomics-based compositional analysis of complex cellulase-hemicellulase mixtures. *J Proteome Res* 10(10)**,** 4365-4372. doi: 10.1021/pr101234z.

Clarke, J.H., Davidson, K., Rixon, J.E., Halstead, J.R., Fransen, M.P., Gilbert, H.J., et al. (2000). A comparison of enzyme-aided bleaching of softwood paper pulp using combinations of xylanase, mannanase and α-galactosidase. *Appl Microbiol Biotechnol* 53(6)**,** 661-667. doi: 10.1007/s002530000344.

Clesceri, L.S., Greenberg, A.E., and Eaton, A.D. (1998). *Standard Methods for the Examination of Water and Wastewater, 20th Edition.* Washington: American Public Health Association.

Colman, D.R., Toolson, E.C., and Takacs-Vesbach, C.D. (2012). Do diet and taxonomy influence insect gut bacterial communities? *Mol Ecol* 21(20)**,** 5124-5137. doi: 10.1111/j.1365-294X.2012.05752.x.

Compton, S.J., and Jones, C.G. (1985). Mechanism of dye response and interference in the Bradford protein assay. *Anal Biochem* 151(2)**,** 369-374. doi: 10.1016/0003-2697(85)90190-3.

Cortes-Tolalpa, L., Jimenez, D.J., Brossi, M.J.D., Salles, J.F., and van Elsas, J.D. (2016). Different inocula produce distinctive microbial consortia with similar lignocellulose degradation capacity. *Appl Microbiol Biotechnol* 100(17)**,** 7713-7725. doi: 10.1007/s00253-016-7516-6.

Cortes-Tolalpa, L., Salles, J.F., and van Elsas, J.D. (2017). Bacterial synergism in lignocellulose biomass degradation - complementary roles of degraders as influenced by complexity of the carbon source. *Front Microbiol* 8. doi: 10.3389/fmicb.2017.01628.

Coutinho, J.B., Gilkes, N.R., Warren, R.A.J., Kilburn, D.G., and Miller, R.C. (1992). The binding of cellulomonas-fimi endoglucanase-c (cenc) to cellulose and sephadex is mediated by the n-terminal repeats. *Mol Microbiol* 6(9)**,** 1243-1252. doi: 10.1111/j.1365-2958.1992.tb01563.x.

Cretoiu, M.S., Kielak, A.M., Abu Al-Soud, W., Sorensen, S.J., and van Elsas, J.D. (2012). Mining of unexplored habitats for novel chitinases-chiA as a helper gene proxy in metagenomics. *Appl Microbiol Biotechnol* 94(5)**,** 1347-1358. doi: 10.1007/s00253-012-4057-5.

Dai, X., Zhu, Y.X., Luo, Y.F., Song, L., Liu, D., Liu, L., et al. (2012). Metagenomic insights into the fibrolytic microbiome in yak rumen. *PLoS One* 7(7). doi: 10.1371/journal.pone.0040430.

Dale, B.E. (2003). 'Greening' the chemical industry: research and development priorities for biobased industrial products. *J Chem Technol Biotechnol* 78(10)**,** 1093-1103. doi: 10.1002/Jctb.850.

Damon, C., Lehembre, F., Oger-Desfeux, C., Luis, P., Ranger, J., Fraissinet-Tachet, L., et al. (2012). Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. *PLoS One* 7(1). doi: 10.1371/journal.pone.0028967.

Davidi, L., Morais, S., Artzi, L., Knop, D., Hadar, Y., Arfi, Y., et al. (2016). Toward combined delignification and saccharification of wheat straw by a laccase-containing designer cellulosome. *Proc Natl Acad Sci U S A* 113(39)**,** 10854-10859. doi: 10.1073/pnas.1608012113.

de Jong, E., Higson, A., Walsh, P., and Wellisch, M. (2012). "Bio-based chemicals: value added products from biorefineries". IEA Bioenergy Task42 Biorefinery).

De Santi, C., Willassen, N.P., and Williamson, A. (2016). Biochemical characterization of a family 15 carbohydrate esterase from a bacterial marine arctic metagenome. *PLoS One* 11(7). doi: 10.1371/journal.pone.0159345.

DeAngelis, K.M., Fortney, J.L., Borglin, S., Silver, W.L., Simmons, B.A., and Hazen, T.C. (2012). Anaerobic decomposition of switchgrass by tropical soil-derived feedstock-adapted consortia. *MBio* 3(1)**,** e00249. doi: 10.1128/mBio.00249-11.

DeAngelis, K.M., Silver, W.L., Thompson, A.W., and Firestone, M.K. (2010). Microbial communities acclimate to recurring changes in soil redox potential status. *Environ Microbiol* 12(12)**,** 3137-3149. doi: 10.1111/j.1462-2920.2010.02286.x.

den Camp, H.J.M.O., Verhagen, F.J.M., Kivaisi, A.K., Dewindt, F.E., Lubberding, H.J., Gijzen, H.J., et al. (1988). Effects of lignin on the anaerobic degradation of (ligno) cellulosic wastes by rumen microorganisms. *Appl Microbiol Biotechnol* 29(4)**,** 408-412. doi: 10.1007/BF00265827.

Dererie, D.Y., Trobro, S., Momeni, M.H., Hansson, H., Blomqvist, J., Passoth, V., et al. (2011). Improved bio-energy yields via sequential ethanol fermentation and biogas digestion of steam exploded oat straw. *Bioresour Technol* 102(6)**,** 4449-4455. doi: 10.1016/j.biortech.2010.12.096.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72(7)**,** 5069-5072. doi: 10.1128/AEM.03006-05.

Desjardins, E., and Beaulieu, C. (2003). Identification of bacteria contaminating pulp and a paper machine in a Canadian paper mill. *J Ind Microbiol Biotechnol* 30(3)**,** 141-145. doi: 10.1007/s10295-002-0017-x.

Despres, J., Forano, E., Lepercq, P., Comtet-Marre, S., Jubelin, G., Yeoman, C.J., et al. (2016). Unraveling the pectinolytic function of *Bacteroides xylanisolvens* using a RNA-seq approach and mutagenesis. *BMC Genomics* 17. doi: 10.1186/s12864-016-2758-3.

Di Risio, S., Hu, C.S., Saville, B.A., Liao, D., and Lortie, J. (2011). Large-scale, high-solids enzymatic hydrolysis of steam-exploded poplar. *Biofuel Bioprod Bior* 5(6)**,** 609-620. doi: 10.1002/bbb.323.

Ding, S.Y., Bayer, E.A., Steiner, D., Shoham, Y., and Lamed, R. (1999). A novel cellulosomal scaffoldin from *Acetivibrio cellulolyticus* that contains a family 9 glycosyl hydrolase. *J Bacteriol* 181(21)**,** 6720-6729.

Ding, S.Y., Bayer, E.A., Steiner, D., Shoham, Y., and Lamed, R. (2000). A scaffoldin of the *Bacteroides cellulosolvens* cellulosome that contains 11 type II cohesins. *J Bacteriol* 182(17)**,** 4915-4925. doi: 10.1128/Jb.182.17.4915-4925.2000.

Du, B.W., Sharma, L.N., Becker, C., Chen, S.F., Mowery, R.A., van Walsum, G.P., et al. (2010). Effect of varying feedstock-pretreatment chemistry combinations on the formation and accumulation of potentially inhibitory degradation products in biomass hydrolysates. *Biotechnol Bioeng* 107(3)**,** 430-440. doi: 10.1002/bit.22829.

Ed, J., Adrian, H., Patrick, W., and Maria, W. (2012). Product developments in the bio-based chemicals arena. *Biofuel Bioprod Bior* 6(6)**,** 606-624. doi: 10.1002/bbb.1360.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19)**,** 2460-2461. doi: 10.1093/bioinformatics/btq461.

Edwards, E.A., and Grbić-Galić, D. (1994). Anaerobic degradation of toluene and o-xylene by a methanogenic consortium. *Appl Environ Microbiol* 60(1)**,** 313-322.

Edwards, J.L., Smith, D.L., Connolly, J., McDonald, J.E., Cox, M.J., Joint, I., et al. (2010). Identification of carbohydrate metabolism genes in the metagenome of a marine biofilm community shown to be dominated by *Gammaproteobacteria* and *Bacteroidetes*. *Genes* 1(3)**,** 371-384. doi: 10.3390/genes1030371.

Excoffier, G., Toussaint, B., and Vignon, M.R. (1991). Saccharification of steam-exploded poplar wood. *Biotechnol Bioeng* 38(11)**,** 1308-1317. doi: 10.1002/bit.260381108.

Faure, D. (2002). The family-3 glycoside hydrolases: From housekeeping functions to host-microbe interactions. *Appl Environ Microbiol* 68(4)**,** 1485-1490. doi: 10.1128/Aem.68.4.1485-1490.2002.

Feist, P., and Hummon, A.B. (2015). Proteomic Challenges: Sample Preparation Techniques for Microgram-Quantity Protein Analysis from Biological Samples. *Int J Mol Sci* 16(2)**,** 3537-3563. doi: 10.3390/ijms16023537.

Feiz, L., Irshad, M., Pont-Lezica, R.F., Canut, H., and Jamet, E. (2006). Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from Arabidopsis hypocotyls. *Plant Methods* 2**,** 10. doi: 10.1186/1746-4811-2-10.

Fernandes, A.C., Fontes, C.M.G.A., Gilbert, H.J., Hazlewood, G.P., Fernandes, T.H., and Ferreira, L.M.A. (1999). Homologous xylanases from *Clostridium thermocellum*: evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes. *Biochem J* 342**,** 105-110. doi: 10.1042/0264-6021:3420105.

Field, J.A., and Lettinga, G. (1987). The methanogenic toxicity and anaerobic degradability of a hydrolyzable tannin. *Water Res* 21(3)**,** 367-374. doi: 10.1016/0043-1354(87)90217-X.

Field, J.A., Leyendeckers, M.J.H., Alvarez, R.S., Lettinga, G., and Habets, L.H.A. (1988). The methanogenic toxicity of bark tannins and the anaerobic biodegradability of water-soluble bark matter. *Water Sci Technol* 20(1)**,** 219-240.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44(D1)**,** D279-D285. doi: 10.1093/nar/gkv1344.

FitzGerald, J.A., Allen, E., Wall, D.M., Jackson, S.A., Murphy, J.D., and Dobson, A.D. (2015). Methanosarcina play an important role in anaerobic co-digestion of the seaweed ulva lactuca: Taxonomy and predicted metabolism of functional microbial communities. *PLoS One* 10(11)**,** e0142603. doi: 10.1371/journal.pone.0142603.

Flipphi, M.J., Visser, J., van der Veen, P., and de Graaff, L.H. (1993). Cloning of the *Aspergillus niger* gene encoding alpha-L-arabinofuranosidase A. *Appl Microbiol Biotechnol* 39(3)**,** 335-340.

Forest Products Association of Canada, and FPInnovations (2011). "The new face of the Canadian forest industry", in: *The Bio-pathways Project.* (Canada).

Foumani, M., Vuong, T.V., MacCormick, B., and Master, E.R. (2015). Enhanced polysaccharide binding and activity on linear beta-glucans through addition of carbohydrate-binding modules to either terminus of a glucooligosaccharide oxidase. *PLoS One* 10(5). doi: 10.1371/journal.pone.0125398.

Fragner, D., Zomorrodi, M., Kues, U., and Majcherczyk, A. (2009). Optimized protocol for the 2-DE of extracellular proteins from higher basidiomycetes inhabiting lignocellulose. *Electrophoresis* 30(14)**,** 2431-2441. doi: 10.1002/elps.200800770.

Freitas, A.C., Ferreira, F., Costa, A.M., Pereira, R., Antunes, S.C., Goncalves, F., et al. (2009). Biological treatment of the effluent from a bleached kraft pulp mill using basidiomycete

and zygomycete fungi. *Sci Total Environ* 407(10)**,** 3282-3289. doi: 10.1016/j.scitotenv.2009.01.054.

Fry, S.C. (1988). *The growing plant cell wall: chemical and metabolic analysis.* Longman Group Limited.

Fujita, K., Sakaguchi, T., Sakamoto, A., Shimokawa, M., and Kitahara, K. (2014). *Bifidobacterium longum* subsp longum exo-beta-1,3-galactanase, an enzyme for the degradation of type II arabinogalactan. *Appl Environ Microbiol* 80(15)**,** 4577-4584. doi: 10.1128/Aem.00802-14.

Fujita, K., Sakamoto, S., Ono, Y., Wakao, M., Suda, Y., Kitahara, K., et al. (2011). Molecular cloning and characterization of a beta-L-arabinobiosidase in *Bifidobacterium longum* that belongs to a novel glycoside hydrolase family. *J Biol Chem* 286(7)**,** 5143-5150. doi: 10.1074/jbc.M110.190512.

Fulling, E.H. (1956). Botanical Aspects of the paper-pulp and tanning industries in the United States - an economic and historical survey. *Am J Bot* 43(8)**,** 621-634. doi: 10.2307/2438878.

Gallardo, O., Fernandez-Fernandez, M., Valls, C., Valenzuela, S.V., Roncero, M.B., Vidal, T., et al. (2010). Characterization of a family GH5 xylanase with activity on neutral oligosaccharides and evaluation as a pulp bleaching aid. *Appl Environ Microbiol* 76(18)**,** 6290-6294. doi: 10.1128/Aem.00871-10.

Gao, D.H., Chundawat, S.P.S., Liu, T.J., Hermanson, S., Gowda, K., Brumm, P., et al. (2010). Strategy for identification of novel fungal and bacterial glycosyl hydrolase hybrid mixtures that can efficiently saccharify pretreated lignocellulosic biomass. *Bioenergy Res* 3(1)**,** 67-81. doi: 10.1007/s12155-009-9066-6.

Garg, G., Singh, A., Kaur, A., Singh, R., Kaur, J., and Mahajan, R. (2016). Microbial pectinases: an ecofriendly tool of nature for industries. *3 Biotech* 6. doi: 10.1007/s13205-016-0371-4.

Garron, M.L., and Cygler, M. (2010). Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology* 20(12)**,** 1547-1573. doi: 10.1093/glycob/cwq122.

Ghribi, M., Meddeb-Mouelhi, F., and Beauregard, M. (2016). Microbial diversity in various types of paper mill sludge: identification of enzyme activities with potential industrial applications. *Springerplus* 5. doi: 10.1186/s40064-016-3147-8.

Gladden, J.M., Allgaier, M., Miller, C.S., Hazen, T.C., VanderGheynst, J.S., Hugenholtz, P., et al. (2011). Glycoside hydrolase activities of thermophilic bacterial consortia adapted to switchgrass. *Appl Environ Microbiol* 77(16)**,** 5804-5812. doi: 10.1128/AEM.00032-11.

Glasser, W.G., and Wright, R.S. (1998). Steam-assisted biomass fractionation. II. Fractionation behavior of various biomass resources. *Biomass Bioenerg* 14(3)**,** 219-235. doi: 10.1016/S0961-9534(97)10037-X.

Gloster, T.M., Ibatullin, F.M., Macauley, K., Eklof, J.M., Roberts, S., Turkenburg, J.P., et al. (2007). Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12. *J Biol Chem* 282(26)**,** 19177-19189. doi: 10.1074/jbc.M700224200.

Goacher, R.E., Tsai, A.Y., and Master, E.R. (2013). Towards practical time-of-flight secondary ion mass spectrometry lignocellulolytic enzyme assays. *Biotechnol Biofuels* 6(1)**,** 132. doi: 10.1186/1754-6834-6-132.

Goel, G., Puniya, A., Aguilar, C., and Singh, K. (2005). Interaction of gut microflora with tannins in feeds. *Naturwissenschaften* 92(11)**,** 497-503.

Gogola, W., Giżejewski, Z., Demiaszkiewicz, A., Lachowicz, J., Miltko, R., Kowalik, B., et al. (2011). "Preliminary studies on the activity of microorganisms inhabiting the lower gut of the beaver (*Castor fiber*)," in *From Research To Practice Application,* eds. P. Peter, P. Javorský, J. Kopečný & G. Avguštin. (Slovakia: Institute of Animal Physiology, Slovak Academy of Science), 81.

Gold, M.H., and Alic, M. (1993). Molecular biology of the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. *Microbiol Rev* 57(3)**,** 605-622.

Goldstein, J.L., and Swain, T. (1965). The inhibition of enzymes by tannins. *Phytochemistry* 4(1)**,** 185-192. doi: 10.1016/S0031-9422(00)86162-2.

Gourlay, K., Arantes, V., and Saddler, J.N. (2012). Use of substructure-specific carbohydrate binding modules to track changes in cellulose accessibility and surface morphology during the amorphogenesis step of enzymatic hydrolysis. *Biotechnol Biofuels* 5. doi: 10.1186/1754-6834-5-51.

Grbić-Galić, D., and Vogel, T.M. (1987). Transformation of toluene and benzene by mixed methanogenic cultures. *Appl Environ Microbiol* 53(2)**,** 254-260.

Greetham, D. (2015). Presence of low concentrations of acetic acid improves fermentations using *Saccharomyces cerevisiae*. *J Bioprocess Biotech* 5(1). doi: 10.4172/2155-9821.1000192.

Grondin, J.M., Tamura, K., Dejean, G., Abbott, D.W., and Brumer, H. (2017). Polysaccharide utilization loci: Fueling microbial communities. *J Bacteriol* 199(15). doi: 10.1128/JB.00860-16.

Haddaway, A. (2014). "Pulp & paper: A look at wastewater treatment trends and technologies", in: *Industrial WaterWorld.* (Oklahoma, USA: PennWell Corporation).

Hagelqvist, A. (2013). Batchwise mesophilic anaerobic co-digestion of secondary sludge from pulp and paper industry and municipal sewage sludge. *Waste Manage* 33(4)**,** 820-824. doi: 10.1016/j.wasman.2012.11.002.

Hahn-Hagerdal, B., Karhumaa, K., Fonseca, C., Spencer-Martins, I., and Gorwa-Grauslund, M.F. (2007). Towards industrial pentose-fermenting yeast strains. *Appl Microbiol Biotechnol* 74(5)**,** 937-953. doi: 10.1007/s00253-006-0827-2.

Hammel, K.E., Jensen, K.A., Mozuch, M.D., Landucci, L.L., Tien, M., and Pease, E.A. (1993). Ligninolysis by a purified lignin peroxidase. *J Biol Chem* 268(17)**,** 12274-12281.

Harris, P.J., and Stone, B.A. (2009). "Chemistry and molecular organization of plant cell walls," in *Biomass Recalcitrance,* ed. M. Himmel. (New Jersey: Blackwell Publishing Ltd.), 61-93.

He, S.M., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R.H., et al. (2013). Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS One* 8(4). doi: 10.1371/journal.pone.0061126.

Healy, F.G., Ray, R.M., Aldrich, H.C., Wilkie, A.C., Ingram, L.O., and Shanmugam, K.T. (1995). Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl Microbiol Biotechnol* 43(4)**,** 667-674.

Heiss-Blanquet, S., Fayolle-Guichard, F., Lombard, V., Hebert, A., Coutinho, P.M., Groppi, A., et al. (2016). Composting-like conditions are more efficient for enrichment and diversity of organisms containing cellulase-encoding genes than submerged cultures. *PLoS One* 11(12). doi: 10.1371/journal.pone.0167216.

Hess, M., Sczyrba, A., Egan, R., Kim, T.W., Chokhawala, H., Schroth, G., et al. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 331(6016)**,** 463-467. doi: 10.1126/science.1200387.

Holtzapple, M.T. (2003). "Lignin," in *Encyclopedia of Food Science and Nutrition,* eds. B. Caballero, L. Trugo & P.M. Finglas. (MA, USA: Academic Press).

Hongoh, Y., Deevong, P., Hattori, S., Inoue, T., Noda, S., Noparatnaraporn, N., et al. (2006). Phylogenetic diversity, localization, and cell morphologies of members of the candidate phylum *TG3* and a subphylum in the phylum *Fibrobacteres*, recently discovered bacterial groups dominant in termite guts. *Appl Environ Microbiol* 72(10)**,** 6780-6788. doi: 10.1128/Aem.00891-06.

Hood, G.A., and Bayley, S.E. (2009). A comparison of riparian plant community response to herbivory by beavers (*Castor canadensis*) and ungulates in Canada's boreal mixed-wood forest. *For Ecol Manage* 258(9)**,** 1979-1989. doi: 10.1016/j.foreco.2009.07.052.

Horn, S., Durka, W., Wolf, R., Ermala, A., Stubbe, A., Stubbe, M., et al. (2011). Mitochondrial genomes reveal slow rates of molecular evolution and the timing of speciation in beavers (*Castor*), one of the largest rodent species. *PLoS One* 6(1)**,** e14622. doi: 10.1371/journal.pone.0014622.

Howard, R., Abotsi, E., Van Rensburg, E.J., and Howard, S. (2004). Lignocellulose biotechnology: issues of bioconversion and enzyme production. *Afr J Biotechnol* 2(12)**,** 602-619.

Huang, C.I. (2012). *Molecular Ecology of Free-Living Chemoautotrophic Microbial Communities at a Shallow-sea Hydrothermal Vent.* Universität Bremen.

Huang, X.F., Bakker, M.G., Judd, T.M., Reardon, K.F., and Vivanco, J.M. (2013). Variations in diversity and richness of gut bacterial communities of termites (*Reticulitermes flavipes*) fed with grassy and woody plant substrates. *Microb Ecol* 65(3)**,** 531-536. doi: 10.1007/s00248-013-0219-y.

Huang, Y., Niu, B.F., Gao, Y., Fu, L.M., and Li, W.Z. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26(5)**,** 680-682. doi: 10.1093/bioinformatics/btq003.

Hunter, C. (2018). *Biorefineries: Helping Transform Canada's Forest Industry* [Online]. Canada: Natural Resources Canada. Available: http://www.nrcan.gc.ca/simply-science/20336 [Accessed 25-5-2018 2018].

Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11**,** 119. doi: 10.1186/1471-2105-11-119.

Iqbal, H.M.N., Kyazze, G., and Keshavarz, T. (2013). Advances in the valorization of lignocellulosic materials by biotechnology: An overview. *Bioresources* 8(2)**,** 3157-3176.

Ishaq, S.L., and Wright, A.D. (2014). High-throughput DNA sequencing of the ruminal bacteria from moose (*Alces alces*) in Vermont, Alaska, and Norway. *Microb Ecol* 68(2)**,** 185-195. doi: 10.1007/s00248-014-0399-0.

Ishaq, S.L., and Wright, A.D.G. (2012). Insight into the bacterial gut microbiome of the North American moose (*Alces alces*). *BMC Microbiol* 12. doi: 10.1186/1471-2180-12-212.

Ishii, S., Yamamoto, M., Kikuchi, M., Oshima, K., Hattori, M., Otsuka, S., et al. (2009). Microbial populations responsive to denitrification-inducing conditions in rice paddy soil, as revealed by comparative 16S rRNA gene analysis. *Appl Environ Microbiol* 75(22)**,** 7070-7078. doi: 10.1128/Aem.01481-09.

Itoh, T., Ochiai, A., Mikami, B., Hashimoto, W., and Murata, K. (2006). A novel glycoside hydrolase family 105: The structure of family 105 unsaturated rhamnogalacturonyl hydrolase complexed with a disaccharide in comparison with family 88 enzyme complexed with the disaccharide. *J Mol Biol* 360(3)**,** 573-585. doi: 10.1016/j.jmb.2006.04.047.

Jiang, C., Hao, Z.Y., Jin, K., Li, S.X., Che, Z.Q., Ma, Q.C., et al. (2010). Identification of a metagenome-derived β-glucosidase from bioreactor contents. *J Mol Catal B Enzym* 63(1-2)**,** 11-16.

Jimenéz, D.J., Brossi, M.J.D., Schuckel, J., Kracun, S.K., Willats, W.G.T., and van Elsas, J.D. (2016). Characterization of three plant biomass-degrading microbial consortia by metagenomics- and metasecretomics-based approaches. *Appl Microbiol Biotechnol* 100(24)**,** 10463-10477. doi: 10.1007/s00253-016-7713-3.

Jimenez, D.J., Dini-Andreote, F., and van Elsas, J.D. (2014a). Metataxonomic profiling and prediction of functional behaviour of wheat straw degrading microbial consortia. *Biotechnol Biofuels* 7**,** 92. doi: 10.1186/1754-6834-7-92.

Jimenez, D.J., Korenblum, E., and van Elsas, J.D. (2014b). Novel multispecies microbial consortia involved in lignocellulose and 5-hydroxymethylfurfural bioconversion. *Appl Microbiol Biotechnol* 98(6)**,** 2789-2803. doi: 10.1007/s00253-013-5253-7.

Jimenez, D.J., Maruthamuthu, M., and van Elsas, J.D. (2015). Metasecretome analysis of a lignocellulolytic microbial consortium grown on wheat straw, xylan and xylose. *Biotechnol Biofuels* 8**,** 199. doi: 10.1186/s13068-015-0387-8.

Johnston, J.B. (2015). Bridging the gap: Fostering competitiveness in the Canadian biobased industry. *Ind Biotechnol* 11(5)**,** 253-258. doi: 10.1089/ind.2015.29012.jbj.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W.Z., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9)**,** 1236-1240. doi: 10.1093/bioinformatics/btu031.

Jonsson, L.J., and Martin, C. (2016). Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresour Technol* 199**,** 103-112. doi: 10.1016/j.biortech.2015.10.009.

Jorgensen, H., and Pinelo, M. (2017). Enzyme recycling in lignocellulosic biorefineries. *Biofuel Bioprod Bior* 11(1)**,** 150-167. doi: 10.1002/bbb.1724.

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., et al. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front Microbiol* 7. doi: 10.3389/fmicb.2016.00459.

Ju, F., and Zhang, T. (2015). Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. *ISME J* 9(3)**,** 683-695. doi: 10.1038/ismej.2014.162.

Jurak, E., Kabel, M.A., and Gruppen, H. (2014). Carbohydrate composition of compost during composting and mycelium growth of Agaricus bisporus. *Carbohydr Polym* 101**,** 281-288. doi: 10.1016/j.carbpol.2013.09.050.

Kamali, M., Gameiro, T., Costa, M.E.V., and Capela, I. (2016). Anaerobic digestion of pulp and paper mill wastes - An overview of the developments and improvement opportunities. *Chem Eng J* 298**,** 162-182. doi: 10.1016/j.cej.2016.03.119.

Kamra, D.N., Agarwal, N., and Chaudhary, L.C. (2006). Inhibition of ruminal methanogenesis by tropical plants containing secondary compounds. *Int Congr Ser* 1293**,** 156-163. doi: 10.1016/j.ics.2006.02.002.

Kang, D.D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3**,** e1165. doi: 10.7717/peerj.1165.

Kanokratana, P., Uengwetwanit, T., Rattanachomsri, U., Bunterngsook, B., Nimchua, T., Tangphatsornruang, S., et al. (2011). Insights into the phylogeny and metabolic potential of a primary tropical peat swamp forest microbial community by metagenomic analysis. *Microb Ecol* 61(3)**,** 518-528. doi: 10.1007/s00248-010-9766-7.

Karkas, M.D., Matsuura, B.S., Monos, T.M., Magallanes, G., and Stephenson, C.R.J. (2016). Transition-metal catalyzed valorization of lignin: the key to a sustainable carbon-neutral future. *Organic & Biomolecular Chemistry* 14(6)**,** 1853-1914. doi: 10.1039/c5ob02212f.

Kato, S., Haruta, S., Cui, Z.J., Ishii, M., and Igarashi, Y. (2005). Stable coexistence of five bacterial strains as a cellulose-degrading community. *Appl Environ Microbiol* 71(11)**,** 7099-7106. doi: 10.1128/Aem.71.11.7099-7106.2005.

Kenealy, W.R., and Jeffries, T.W. (2003). "Enzyme processes for pulp and paper: A review of recent developments," in *Wood Deterioration and Preservation*. (Washington: ACS Publications), 210-239.

Kim, D. (2018). Physico-Chemical Conversion of Lignocellulose: Inhibitor Effects and Detoxification Strategies: A Mini Review. *Molecules* 23(2). doi: 10.3390/molecules23020309.

Kim, Y., Kreke, T., Hendrickson, R., Parenti, J., and Ladisch, M.R. (2013). Fractionation of cellulase and fermentation inhibitors from steam pretreated mixed hardwood. *Bioresour Technol* 135**,** 30-38. doi: 10.1016/j.biortech.2012.10.130.

King, B.C., Donnelly, M.K., Bergstrom, G.C., Walker, L.P., and Gibson, D.M. (2009). An optimized microplate assay system for quantitative evaluation of plant cell wall-degrading enzyme activity of fungal culture extracts. *Biotechnol Bioeng* 102(4)**,** 1033-1044. doi: 10.1002/bit.22151.

Kirk, T.K., and Jeffries, T.W. (1996). "Roles for microbial enzymes in pulp and paper processing," in *Enzymes for Pulp and Paper Processing*. American Chemical Society), 2-14.

Kirker, G.T., Blodgett, A.B., Arango, R.A., Lebow, P.K., and Clausen, C.A. (2013). The role of extractives in naturally durable wood species. *Int Biodeterior Biodegrad* 82**,** 53-58. doi: 10.1016/j.ibiod.2013.03.007.

Kobayashi, H., and Fukuoka, A. (2013). Synthesis and utilisation of sugar compounds derived from lignocellulosic biomass. *Green Chem* 15(7)**,** 1740-1763. doi: 10.1039/c3gc00060e.

Kohl, K.D., Skopec, M.M., and Dearing, M.D. (2014). Captivity results in disparate loss of gut microbial diversity in closely related hosts. *Conserv Physiol* 2(1). doi: 10.1093/conphys/cou009.

Kong, W., Fu, X., Wang, L., Alhujaily, A., Zhang, J., Ma, F., et al. (2017). A novel and efficient fungal delignification strategy based on versatile peroxidase for lignocellulose bioconversion. *Biotechnol Biofuels* 10**,** 218. doi: 10.1186/s13068-017-0906-x.

Konn, J., Holmbom, B., and Nickull, O. (2002). Chemical reactions in chemimechanical pulping: Material balances of wood components in a CTMP process. *J Pulp Pap Sci* 28(12)**,** 395-399.

Koontz, L. (2014). "TCA precipitation," in *Methods Enzymol,* ed. J. Lorsch. Academic Press), 3-10.

Koropatkin, N.M., Cameron, E.A., and Martens, E.C. (2012). How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* 10(5)**,** 323-335. doi: 10.1038/nrmicro2746.

Krober, M., Bekel, T., Diaz, N.N., Goesmann, A., Jaenicke, S., Krause, L., et al. (2009). Phylogenetic characterization of a biogas plant microbial community integrating clone library 16S-rDNA sequences and metagenome sequence data obtained by 454-pyrosequencing. *J Biotechnol* 142(1)**,** 38-49. doi: 10.1016/j.jbiotec.2009.02.010.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res* 19(9)**,** 1639-1645. doi: 10.1101/gr.092759.109.

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. doi: 10.1139/m87-041.

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. doi: 10.1007/s10295-008-0327-8.

Lacourt, M.W. (2011). *Enrichment of methanogenic microcosms on recalcitrant lignocellulosic biomass.* Master, University of Toronto.

Lamed, R., Naimark, J., Morgenstern, E., and Bayer, E.A. (1987). Specialized cell-surface structures in cellulolytic bacteria. *J Bacteriol* 169(8)**,** 3792-3800.

Lamed, R., Setter, E., and Bayer, E.A. (1983). Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *J Bacteriol* 156(2)**,** 828-836.

Langston, J.A., Shaghasi, T., Abbate, E., Xu, F., Vlasenko, E., and Sweeney, M.D. (2011). Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Appl Environ Microbiol* 77(19)**,** 7007-7015. doi: 10.1128/AEM.05815-11.

Larsbrink, J., Izumi, A., Ibatullin, F.M., Nakhai, A., Gilbert, H.J., Davies, G.J., et al. (2011). Structural and enzymatic characterization of a glycoside hydrolase family 31 alpha-xylosidase from Cellvibrio japonicus involved in xyloglucan saccharification. *Biochem J* 436(3)**,** 567-580. doi: 10.1042/BJ20110299.

Larsbrink, J., Zhu, Y., Kharade, S.S., Kwiatkowski, K.J., Eijsink, V.G.H., Koropatkin, N.M., et al. (2016). A polysaccharide utilization locus from Flavobacterium johnsoniae enables conversion of recalcitrant chitin. *Biotechnol Biofuels* 9. doi: 10.1186/s13068-016-0674-z.

Larsson, S., Palmqvist, E., Hahn-Hägerdal, B., Tengborg, C., Stenberg, K., Zacchi, G., et al. (1999). The generation of fermentation inhibitors during dilute acid hydrolysis of softwood. *Enzyme Microb Technol* 24(3)**,** 151-159. doi: 10.1016/S0141-0229(98)00101-X.

Li, C.M., Wang, Y., and Yu, W.X. (2011). Dynamic changes of phenolic compound contents in leaf and bark of poplar during autumn temperature drop. *J Forest Res* 22(3)**,** 481. doi: 10.1007/s11676-011-0191-7.

Li, L.L., McCorkle, S.R., Monchy, S., Taghavi, S., and van der Lelie, D. (2009). Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. *Biotechnol Biofuels* 2**,** 10. doi: 10.1186/1754-6834-2-10.

Lilburn, T.C., Kim, K.S., Ostrom, N.E., Byzek, K.R., Leadbetter, J.R., and Breznak, J.A. (2001). Nitrogen fixation by symbiotic and free-living spirochetes. *Science* 292(5526)**,** 2495-2498. doi: 10.1126/science.1060281.

Limam, R.D., Chouari, R., Mazeas, L., Wu, T.D., Li, T.L., Grossin-Debattista, J., et al. (2014). Members of the uncultured bacterial candidate division WWE1 are implicated in anaerobic digestion of cellulose. *MicrobiologyOpen* 3(2)**,** 157-167. doi: 10.1002/mbo3.144.

Lin, S.H., and Guidotti, G. (2009). Purification of membrane proteins. *Methods Enzymol* 463**,** 619-629. doi: 10.1016/S0076-6879(09)63035-4.

Liu, D.Y., Li, J., Zhao, S., Zhang, R.F., Wang, M.M., Miao, Y.Z., et al. (2013). Secretome diversity and quantitative analysis of cellulolytic Aspergillus fumigatus Z5 in the presence of different carbon sources. *Biotechnol Biofuels* 6. doi: 10.1186/1754-6834-6-149.

Liu, Y., Qiao, J.T., Yuan, X.Z., Guo, R.B., and Qiu, Y.L. (2014). Hydrogenispora ethanolica gen. nov., sp. nov., an anaerobic carbohydrate-fermenting bacterium from anaerobic sludge. *Int J Syst Evol Microbiol* 64(Pt 5)**,** 1756-1762. doi: 10.1099/ijs.0.060186-0.

Liu, Z.Z., DeSantis, T.Z., Andersen, G.L., and Knight, R. (2008). Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res* 36(18). doi: 10.1093/nar/gkn491.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42(D1)**,** D490-D495. doi: 10.1093/nar/gkt1178.

Lopez-Mondejar, R., Zuhlke, D., Becher, D., Riedel, K., and Baldrian, P. (2016). Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by the action of structurally variable enzymatic systems. *Sci Rep* 6**,** 25279. doi: 10.1038/srep25279.

Lopez, M.J., Nichols, N.N., Dien, B.S., Moreno, J., and Bothast, R.J. (2004). Isolation of microorganisms for biological detoxification of lignocellulosic hydrolysates. *Appl Microbiol Biotechnol* 64(1)**,** 125-131. doi: 10.1007/s00253-003-1401-9.

Lovering, A.L., Lee, S.S., Kim, Y.W., Withers, S.G., and Strynadka, N.C.J. (2005). Mechanistic and structural analysis of a family 31 alpha-glycosidase and its glycosyl-enzyme intermediate. *J Biol Chem* 280(3)**,** 2105-2115. doi: 10.1074/jbc.M410468200.

Lucas, R., Kuchenbuch, A., Fetzer, I., Harms, H., and Kleinsteuber, S. (2015). Long-term monitoring reveals stable and remarkably similar microbial communities in parallel full-scale biogas reactors digesting energy crops. *FEMS Microbiol Ecol* 91(3). doi: 10.1093/femsec/fiv004.

Luo, C.D., Brink, D.L., and Blanch, H.W. (2002). Identification of potential fermentation inhibitors in conversion of hybrid poplar hydrolyzate to ethanol. *Biomass Bioenerg* 22(2)**,** 125-138. doi: 10.1016/S0961-9534(01)00061-7.

MacCormick, B., Vuong, T.V., and Master, E.R. (2018). Chemo-enzymatic Synthesis of Clickable Xylo-oligosaccharide Monomers from Hardwood 4-O-Methylglucuronoxylan. *Biomacromolecules* 19(2)**,** 521-530. doi: 10.1021/acs.biomac.7b01642.

Machovic, M., and Janecek, S. (2008). Domain evolution in the GH13 pullulanase subfamily with focus on the carbohydrate-binding module family 48. *Biologia* 63(6)**,** 1057-1068. doi: 10.2478/s11756-008-0162-4.

Mai-Gisondi, G., Maaheimo, H., Chong, S.L., Hinz, S., Tenkanen, M., and Master, E. (2017). Functional comparison of versatile carbohydrate esterases from families CE1, CE6 and CE16 on acetyl-4-O-methylglucuronoxylan and acetyl-galactoglucomannan. *Bba-Gen Subjects* 1861(9)**,** 2398-2405. doi: 10.1016/j.bbagen.2017.06.002.

Maijala, P., Kleen, M., Westin, C., Poppius-Levlin, K., Herranen, K., Lehto, J.H., et al. (2008). Biomechanical pulping of softwood with enzymes and white-rot fungus *Physisporinus rivulosus*. *Enzyme Microb Technol* 43(2)**,** 169-177. doi: 10.1016/j.enzmictec.2007.11.017.

Maki, M.L., Broere, M., Leung, K.T., and Qin, W. (2011). Characterization of some efficient cellulase producing bacteria isolated from paper mill sludges and organic fertilizers. *Int J Biochem Mol Biol* 2(2)**,** 146-154.

Makkar, H.P.S. (2003). Effects and fate of tannins in ruminant animals, adaptation to tannins, and strategies to overcome detrimental effects of feeding tannin-rich feeds. *Small Ruminant Res* 49(3)**,** 241-256. doi: 10.1016/S0921-4488(03)00142-1.

Malherbe, S., and Cloete, T.E. (2002). Lignocellulose biodegradation: Fundamentals and applications. *Rev Environ Sci Biotechnol* 1(2)**,** 105-114.

Mardis, E.R. (2008). Next-generation DNA sequencing methods. *Annu Rev Genom Hum Genet* 9**,** 387-402. doi: 10.1146/annurev.genom.9.081307.164359.

Marques, A.P., Evtuguin, D.V., Magina, S., Amado, F.M.L., and Prates, A. (2009). Chemical composition of spent liquors from acidic magnesium-based sulphite pulping of *Eucalyptus globulus*. *J Wood Chem Technol* 29(4)**,** 322-336. doi: 10.1080/02773810903207754.

Marshall, T., and Williams, K.M. (2004). Interference in the Coomassie Brilliant Blue and Pyrogallol Red protein dye-binding assays is increased by the addition of sodium dodecyl sulfate to the dye reagents. *Anal Biochem* 331(2)**,** 255-259. doi: 10.1016/j.ab.2004.04.029.

Martens, E.C., Koropatkin, N.M., Smith, T.J., and Gordon, J.I. (2009). Complex glycan catabolism by the human gut microbiota: The *Bacteroidetes* sus-like paradigm. *J Biol Chem* 284(37)**,** 24673-24677. doi: 10.1074/jbc.R109.022848.

Martin-Sampedro, R., Eugenio, M.E., Revilla, E., Martin, J.A., and Villar, J.C. (2011). Integration of Kraft Pulping on a Forest Biorefinery by the Addition of a Steam Explosion Pretreatment. *Bioresources* 6(1)**,** 513-528.

McGregor, N., Yin, V., Tung, C.C., Van Petegem, F., and Brumer, H. (2017). Crystallographic insight into the evolutionary origins of xyloglucan endotransglycosylases and endohydrolases. *Plant J* 89(4)**,** 651-670. doi: 10.1111/tpj.13421.

McKendry, P. (2002). Energy production from biomass (part 1): overview of biomass. *Bioresour Technol* 83(1)**,** 37-46.

McKie, V.A., Black, G.W., MillwardSadler, S.J., Hazlewood, G.P., Laurie, J.I., and Gilbert, H.J. (1997). Arabinanase A from *Pseudomonas fluorescens* subsp cellulosa exhibits both an endo- and an exo- mode of action. *Biochem J* 323**,** 547-555. doi: 10.1042/bj3230547.

Meehan, C.J., and Beiko, R.G. (2014). A phylogenomic view of ecological specialization in the *Lachnospiraceae*, a family of digestive tract-associated bacteria. *Genome Biol Evol* 6(3)**,** 703-713. doi: 10.1093/gbe/evu050.

Menzel, P., Ng, K.L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* 7**,** 11257. doi: 10.1038/ncomms11257.

Mesnage, S., Fontaine, T., Mignot, T., Delepierre, M., Mock, M., and Fouet, A. (2000). Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. *EMBO J* 19(17)**,** 4473-4484. doi: 10.1093/emboj/19.17.4473.

Metsalu, T., and Vilo, J. (2015). ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res* 43(W1)**,** W566-W570. doi: 10.1093/nar/gkv468.

Mewis, K., Lenfant, N., Lombard, V., and Henrissat, B. (2016). Dividing the large glycoside hydrolase family 43 into subfamilies: A motivation for detailed enzyme characterization. *Appl Environ Microbiol* 82(6)**,** 1686-1692. doi: 10.1128/Aem.03453-15.

Meyer, T., and Edwards, E.A. (2015). Corrigendum to "Anaerobic digestion of pulp and paper mill wastewater and sludge" [Water Res. 65 (2014) 321-349]. *Water Res* 68**,** 849. doi: 10.1016/j.watres.2014.10.059.

Michel, G., Barbeyron, T., Kloareg, B., and Czjzek, M. (2009). The family 6 carbohydrate-binding modules have coevolved with their appended catalytic modules toward similar substrate specificity. *Glycobiology* 19(6)**,** 615-623. doi: 10.1093/glycob/cwp028.

Mills, T.Y., Sandoval, N.R., and Gill, R.T. (2009). Cellulosic hydrolysate toxicity and tolerance mechanisms in *Escherichia coli*. *Biotechnol Biofuels* 2**,** 26. doi: 10.1186/1754-6834-2-26.

Mirescu, A., and Prusse, U. (2007). A new environmental friendly method for the preparation of sugar acids via catalytic oxidation on gold catalysts. *Appl Catal B-Environ* 70(1-4)**,** 644-652. doi: 10.1016/j.apcatb.2006.01.017.

Mitchell, V.D., Taylor, C.M., and Bauer, S. (2014). Comprehensive analysis of monomeric phenolics in dilute acid plant hydrolysates. *Bioenergy Res* 7(2)**,** 654-669. doi: 10.1007/s12155-013-9392-6.

Mizutani, K., Fernandes, V.O., Karita, S., Luis, A.S., Sakka, M., Kimura, T., et al. (2012). Influence of a mannan binding family 32 carbohydrate binding module on the activity of the appended mannanase. *Appl Environ Microbiol* 78(14)**,** 4781-4787. doi: 10.1128/Aem.07457-11.

Mohnen, D. (2008). Pectin structure and biosynthesis. *Curr Opin Plant Biol* 11(3)**,** 266-277. doi: 10.1016/j.pbi.2008.03.006.

Montella, S., Amore, A., and Faraco, V. (2016). Metagenomics for the development of new biocatalysts to advance lignocellulose saccharification for bioeconomic development. *Crit Rev Biotechnol* 36(6)**,** 998-1009. doi: 10.3109/07388551.2015.1083939.

Morag, E., Halevy, I., Bayer, E.A., and Lamed, R. (1991). Isolation and properties of a major cellobiohydrolase from the cellulosome of *Clostridium-Thermocellum*. *J Bacteriol* 173(13)**,** 4155-4162.

Morais, S., Morag, E., Barak, Y., Goldman, D., Hadar, Y., Lamed, R., et al. (2012). Deconstruction of lignocellulose into soluble sugars by native and designer cellulosomes. *Mbio* 3(6). doi: 10.1128/mBio.00508-12.

Moretti, R., and Thorson, J.S. (2008). A comparison of sugar indicators enables a universal high-throughput sugar-1-phosphate nucleotidyltransferase assay. *Anal Biochem* 377(2)**,** 251-258. doi: 10.1016/j.ab.2008.03.018.

Mottiar, Y., Vanholme, R., Boerjan, W., Ralph, J., and Mansfield, S.D. (2016). Designer lignins: harnessing the plasticity of lignification. *Curr Opin Biotechnol* 37**,** 190-200. doi: 10.1016/j.copbio.2015.10.009.

Mudgal, R., Sandhya, S., Chandra, N., and Srinivasan, N. (2015). De-DUFing the DUFs: Deciphering distant evolutionary relationships of domains of unknown function using sensitive homology detection methods. *Biol Direct* 10. doi: 10.1186/s13062-015-0069-2.

Muller, B., Sun, L., Westerholm, M., and Schnurer, A. (2016). Bacterial community composition and fhs profiles of low- and high-ammonia biogas digesters reveal novel syntrophic acetate-oxidising bacteria. *Biotechnol Biofuels* 9**,** 48. doi: 10.1186/s13068-016-0454-9.

Nakai, R., Abe, T., Baba, T., Imura, S., Kagoshima, H., Kanda, H., et al. (2012). Microflorae of aquatic moss pillars in a freshwater lake, East Antarctica, based on fatty acid and 16S rRNA gene analyses. *Polar Biol* 35(3)**,** 425-433. doi: 10.1007/s00300-011-1090-2.

Nascimento, M.S., Santana, A.L.B.D., Maranhão, C.A., Oliveira, L.S., and Bieber, L. (2013). "Phenolic Extractives and Natural Resistance of Wood," in *Biodegradation - Life of Science*, eds. R. Chamy & F. Rosenkranz. (Croatia: InTech).

National Center for Biotechnology Information (2016a). *PubChem Compound Database; CID=16129778* [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/compound/tannic_acid [Accessed 24 April 2016].

National Center for Biotechnology Information (2016b). *PubChem Compound Database; CID=25113562* [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/compound/25113562 [Accessed 24 April 2016].

Natural Resources Canada (2016). *Forest bioeconomy, bioenergy and bioproducts* [Online]. Canada: Natural Resources Canada. Available: http://www.nrcan.gc.ca/forests/industry/bioproducts/13315 [Accessed 25-5-2018 2019].

Natural Resources Canada (2017). "The State of Canada's Forests". (Ottawa: Canadian Forest Service).

Natural Resources Canada (2018). *Statistical data* [Online]. Government of Canada. Available: http://cfs.nrcan.gc.ca/statsprofile [Accessed 6/3/2018].

Ndeh, D., Rogowski, A., Cartmell, A., Luis, A.S., Basle, A., Gray, J., et al. (2017). Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* 544(7648)**,** 65-70. doi: 10.1038/nature21725.

Neumuller, K.G., de Souza, A.C., van Rijn, J.H.J., Streekstra, H., Gruppen, H., and Schols, H.A. (2015). Positional preferences of acetyl esterases from different CE families towards acetylated 4-O-methyl glucuronic acid-substituted xylo-oligosaccharides. *Biotechnol Biofuels* 8. doi: 10.1186/s13068-014-0187-6.

Nobeli, I., Favia, A.D., and Thornton, J.M. (2009). Protein promiscuity and its implications for biotechnology. *Nat Biotechnol* 27(2)**,** 157-167. doi: 10.1038/nbt1519.

Nogue, V.S.I., Narayanan, V., and Gorwa-Grauslund, M.F. (2013). Short-term adaptation improves the fermentation performance of *Saccharomyces cerevisiae* in the presence of acetic acid at low pH. *Appl Microbiol Biotechnol* 97(16)**,** 7517-7525. doi: 10.1007/s00253-013-5093-5.

O'Connor, R.M., Fung, J.M., Sharp, K.H., Benner, J.S., McClung, C., Cushing, S., et al. (2014). Gill bacteria enable a novel digestive strategy in a wood-feeding mollusk. *Proc Natl Acad Sci U S A* 111(47)**,** E5096-5104. doi: 10.1073/pnas.1413110111.

O'Neill, M.A., Ishii, T., Albersheim, P., and Darvill, A.G. (2004). Rhamnogalacturonan II: Structure and function of a borate cross-linked cell wall pectic polysaccharide. *Annu Rev Plant Biol* 55**,** 109-139. doi: 10.1146/annurev.arplant.55.031903.141750.

Olsen, I. (2014). "The family *Fusobacteriaceae*," in *The Prokaryotes,* eds. E. Rosenberg, E. DeLong, S. Lory, E. Stackebrandt & F. Thompson. (Berlin Heidelberg: Springer), 109-132.

Olsson, L., and Hahn-Hagerdal, B. (1996). Fermentation of lignocellulosic hydrolysates for ethanol production. *Enzyme Microb Technol* 18(5)**,** 312-331. doi: 10.1016/0141-0229(95)00157-3.

Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G.A., Papanikolaou, N., Kotoulas, G., et al. (2015). Metagenomics: Tools and Insights for Analyzing Next-Generation Sequencing Data Derived from Biodiversity Studies. *Bioinform Biol Insig* 9**,** 75-88. doi: 10.4137/Bbi.S12462.

Oyserman, B.O., Martirano, J.M., Wipperfurth, S., Owen, B.R., Noguera, D.R., and McMahon, K.D. (2017). Community assembly and ecology of activated sludge under photosynthetic feast famine conditions. *Environ Sci Technol* 51(6)**,** 3165-3175. doi: 10.1021/acs.est.6b03976.

Palmqvist, E., and Hahn-Hagerdal, B. (2000). Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition. *Bioresour Technol* 74(1)**,** 25-33. doi: 10.1016/S0960-8524(99)00161-3.

Pampulha, M.E., and Loureiro-Dias, M.C. (2000). Energetics of the effect of acetic acid on growth of *Saccharomyces cerevisiae*. *FEMS Microbiol Lett* 184(1)**,** 69-72.

Pandey, S., Gulati, S., Goyal, E., Singh, S., Kumar, K., Nain, L., et al. (2016). Construction and screening of metagenomic library derived from soil for β-1, 4-endoglucanase gene. *Biocatal Agric Biotechnol* 5**,** 186-192. doi: 10.1016/j.bcab.2016.01.008.

Pandey, S., Kushwah, J., Tiwari, R., Kumar, R., Somvanshi, V.S., Nain, L., et al. (2014). Cloning and expression of beta-1, 4-endoglucanase gene from *Bacillus subtilis* isolated from soil long term irrigated with effluents of paper and pulp mill. *Microbiol Res* 169(9-10)**,** 693-698. doi: 10.1016/j.micres.2014.02.006.

Park, J.I., Steen, E.J., Burd, H., Evans, S.S., Redding-Johnson, A.M., Batth, T., et al. (2012). A thermophilic ionic liquid-tolerant cellulase cocktail for the production of cellulosic biofuels. *PLoS One* 7(5)**,** e37010. doi: 10.1371/journal.pone.0037010.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25(7)**,** 1043-1055. doi: 10.1101/gr.186072.114.

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.A., Woodcroft, B.J., Evans, P.N., et al. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology* 2(11)**,** 1533-1542. doi: 10.1038/s41564-017-0012-7.

Patrascu, O., Beguet-Crespel, F., Marinelli, L., Le Chatelier, E., Abraham, A.L., Leclerc, M., et al. (2017). A fibrolytic potential in the human ileum mucosal microbiota revealed by functional metagenomic. *Sci Rep* 7. doi: 10.1038/srep40248.

Peng, F., Peng, P., Xu, F., and Sun, R.C. (2012). Fractional purification and bioconversion of hemicelluloses. *Biotechnol Adv* 30(4)**,** 879-903. doi: 10.1016/j.biotechadv.2012.01.018.

Picart, P., Domínguez de María, P., and Schallmey, A. (2015). From gene to biorefinery: microbial β-etherases as promising biocatalysts for lignin valorization. *Front Microbiol* 6(916). doi: 10.3389/fmicb.2015.00916.

Pold, G., Billings, A.F., Blanchard, J.L., Burkhardt, D.B., Frey, S.D., Melillo, J.M., et al. (2016). Long-term warming alters carbohydrate degradation potential in temperate forest soils. *Appl Environ Microbiol* 82(22)**,** 6518-6530. doi: 10.1128/Aem.02012-16.

Pollegioni, L., Tonin, F., and Rosini, E. (2015). Lignin-degrading enzymes. *FEBS J* 282(7)**,** 1190-1213. doi: 10.1111/febs.13224.

Pope, P.B., Denman, S.E., Jones, M., Tringe, S.G., Barry, K., Malfatti, S.A., et al. (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. doi: 10.1073/pnas.1005297107.

Pope, P.B., Mackenzie, A.K., Gregor, I., Smith, W., Sundset, M.A., McHardy, A.C., et al. (2012). Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. *PLoS One* 7(6)**,** e38571. doi: 10.1371/journal.pone.0038571.

Potters, G., Van Goethem, D., and Schutte, F. (2010). Promising biofuel resources: lignocellulose and algae. *Nature Education* 3(9)**,** 14.

Prasad, S., Singh, A., and Joshi, H.C. (2007). Ethanol as an alternative fuel from agricultural, industrial and urban residues. *Resour Conserv Recy* 50(1)**,** 1-39. doi: 10.1016/j.resconrec.2006.05.007.

Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 35(9)**,** 833-844. doi: 10.1038/nbt.3935.

R Development Core Team (2010). "R: A language and environment for statistical computing". (Vienna, Austria: R Foundation for Statistical Computer).

Rahikainen, J., Mikander, S., Marjamaa, K., Tamminen, T., Lappas, A., Viikari, L., et al. (2011). Inhibition of enzymatic hydrolysis by residual lignins from softwood--study of enzyme binding and inactivation on lignin-rich surface. *Biotechnol Bioeng* 108(12)**,** 2823-2834. doi: 10.1002/bit.23242.

Rahman, N.A., Parks, D.H., Vanwonterghem, I., Morrison, M., Tyson, G.W., and Hugenholtz, P. (2016). A Phylogenomic Analysis of the Bacterial Phylum *Fibrobacteres*. *Front Microbiol* 6. doi: 10.3389/fmicb.2015.01469.

Rahmanpour, R., and Bugg, T.D.H. (2015). Characterisation of Dyp-type peroxidases from *Pseudomonas fluorescens* Pf-5: Oxidation of Mn(II) and polymeric lignin by Dyp1B. *Arch Biochem Biophys* 574**,** 93-98. doi: 10.1016/j.abb.2014.12.022.

Rajulapati, V., and Goyal, A. (2017). Molecular cloning, expression and characterization of pectin methylesterase (CtPME) from *Clostridium thermocellum*. *Mol Biotechnol* 59(4-5)**,** 128-140. doi: 10.1007/s12033-017-9997-7.

Ransom-Jones, E., McCarthy, A.J., Haldenby, S., Doonan, J., and McDonald, J.E. (2017). Lignocellulose-degrading microbial communities in landfill sites represent a repository of unexplored biomass-degrading diversity. *mSphere* 2(4). doi: 10.1128/mSphere.00300-17.

Ravachol, J., Borne, R., Tardif, C., de Philip, P., and Fierobe, H.P. (2014). Characterization of all family-9 glycoside hydrolases synthesized by the cellulosome-producing bacterium *Clostridium cellulolyticum*. *J Biol Chem* 289(11)**,** 7335-7348. doi: 10.1074/jbc.M113.545046.

Razeq, F.M. (2017). *Discovery and biochemical characterization of new CAZYmes from PULs and metagenome sequences.* Master, University of Toronto.

Razeq, F.M., Jurak, E., Stogios, P.J., Yan, R.Y., Tenkanen, M., Kabel, M.A., et al. (2018). A novel acetyl xylan esterase enabling complete deacetylation of substituted xylans. *Biotechnol Biofuels* 11. doi: 10.1186/s13068-018-1074-3.

Redmile-Gordon, M.A., Armenise, E., White, R.P., Hirsch, P.R., and Goulding, K.W.T. (2013). A comparison of two colorimetric assays, based upon Lowry and Bradford techniques, to estimate total protein in soil extracts. *Soil Biol Biochem* 67**,** 166-173. doi: 10.1016/j.soilbio.2013.08.017.

Rennie, E.A., and Scheller, H.V. (2014). Xylan biosynthesis. *Curr Opin Biotechnol* 26**,** 100-107. doi: 10.1016/j.copbio.2013.11.013.

Resolute Forest Products (2017). "Resolute forest products-2016 annual report". (Montreal, Canada).

Rizzi, A., Crotti, E., Borruso, L., Jucker, C., Lupi, D., Colombo, M., et al. (2013). Characterization of the bacterial community associated with larvae and adults of anoplophora chinensis collected in italy by culture and culture-independent methods. *Biomed Res Int*. doi: 10.1155/2013/420287.

Robert, D., Bardet, M., Lapierre, C., and Gellerstedt, G. (1988). Structural-changes in aspen lignin during steam explosion treatment. *Cell Chem Technol* 22(2)**,** 221-230.

Rocha-Martin, J., Martinez-Bernal, C., Perez-Cobas, Y., Reyes-Sosa, F.M., and Garcia, B.D. (2017). Additives enhancing enzymatic hydrolysis of lignocellulosic biomass. *Bioresour Technol* 244(Pt 1)**,** 48-56. doi: 10.1016/j.biortech.2017.06.132.

Roehr, M. (2001). *The biotechnology of ethanol: classical and future applications.* Wiley-VCH.

Rosenthal, A.Z., Matson, E.G., Eldar, A., and Leadbetter, J.R. (2011). RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture. *ISME J* 5(7)**,** 1133-1142. doi: 10.1038/Ismej.2011.3.

Roske, I., Sabra, W., Nacke, H., Daniel, R., Zeng, A.P., Antranikian, G., et al. (2014). Microbial community composition and dynamics in high-temperature biogas reactors using industrial bioethanol waste as substrate. *Appl Microbiol Biotechnol* 98(21)**,** 9095-9106. doi: 10.1007/s00253-014-5906-1.

Rosnow, J.J., Anderson, L.N., Nair, R.N., Baker, E.S., and Wright, A.T. (2016). Profiling microbial lignocellulose degradation and utilization by emergent omics technologies. *Crit Rev Biotechnol***,** 1-15. doi: 10.1080/07388551.2016.1209158.

Ross, K.A., Feazel, L.M., Robertson, C.E., Fathepure, B.Z., Wright, K.E., Turk-MacLeod, R.M., et al. (2012). Phototrophic phylotypes dominate mesothermal microbial mats associated

with hot springs in yellowstone national park. *Microb Ecol* 64(1)**,** 162-170. doi: 10.1007/s00248-012-0012-3.

Rossmassler, K., Dietrich, C., Thompson, C., Mikaelyan, A., Nonoh, J.O., Scheffrahn, R.H., et al. (2015). Metagenomic analysis of the microbiota in the highly compartmented hindguts of six wood- or soil-feeding higher termites. *Microbiome* 3. doi: 10.1186/s40168-015-0118-1.

Rothberg, J.M., and Leamon, J.H. (2008). The development and impact of 454 sequencing. *Nat Biotechnol* 26(10)**,** 1117-1124. doi: 10.1038/nbt1485.

Roumpeka, D.D., Wallace, R.J., Escalettes, F., Fotheringham, I., and Watson, M. (2017). A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data. *Front Genet* 8. doi: 10.3389/fgene.2017.00023.

Royer, J.C., and Nakas, J.P. (1987). Production of mycelial protein and hydrolytic enzymes from paper-mill sludges by cellulolytic fungi. *J Ind Microbiol* 2(1)**,** 9-13. doi: 10.1007/Bf01569400.

Ruiz, H.A., Rodríguez-Jasso, R.M., Hernandez-Almanza, A., Contreras-Esquivel, J.C., and Aguilar, C.N. (2017). "3 - Pectinolytic Enzymes," in *Current Developments in Biotechnology and Bioengineering*, eds. S. Negi & C.R. Soccol. Elsevier), 47-71.

Sakamoto, T., Nishimura, Y., Makino, Y., Sunagawa, Y., and Harada, N. (2013). Biochemical characterization of a GH53 endo-beta-1,4-galactanase and a GH35 exo-beta-1,4-galactanase from *Penicillium chrysogenum*. *Appl Microbiol Biotechnol* 97(7)**,** 2895-2906. doi: 10.1007/s00253-012-4154-5.

Samsudin, A.A., Evans, P.N., Wright, A.D.G., and Al Jassim, R. (2011). Molecular diversity of the foregut bacteria community in the dromedary camel (*Camelus dromedarius*). *Environ Microbiol* 13(11)**,** 3024-3035. doi: 10.1111/j.1462-2920.2011.02579.x.

San Martin-Davison, J., Ballesteros, M., Manzanares, P., Sepulveda, X.P.B., and Vergara-Fernandez, A. (2015). Effects of temperature on steam explosion pretreatment of poplar hybrids with different lignin contents in bioethanol production. *Int J Green Energy* 12(8)**,** 832-842. doi: 10.1080/15435075.2014.887569.

Scharf, M.E., and Tartar, A. (2008). Termite digestomes as sources for novel lignocellulases. *Biofuel Bioprod Bior* 2(6)**,** 540-552. doi: 10.1002/Bbb.107.

Scheller, H.V., and Ulvskov, P. (2010). Hemicelluloses. *Annu Rev Plant Biol* 61**,** 263-289. doi: 10.1146/annurev-arplant-042809-112315.

Schmid, A., Dordick, J.S., Hauer, B., Kiener, A., Wubbolts, M., and Witholt, B. (2001). Industrial biocatalysis today and tomorrow. *Nature* 409(6817)**,** 258-268. doi: 10.1038/35051736.

Schols, H.A., Visser, R.G.F., and Voragen, A.G.J. (2009). *Pectins and pectinases*. Wageningen Academic Publishers.

Scully, E.D., Geib, S.M., Hoover, K., Tien, M., Tringe, S.G., Barry, K.W., et al. (2013). Metagenomic profiling reveals lignocellulose degrading system in a microbial community associated with a wood-feeding beetle. *PLoS One* 8(9)**,** e73827. doi: 10.1371/journal.pone.0073827.

Shah, H.N., and Gharbia, S.E. (2010). "6.4.2 Secreted proteins," in *Mass Spectrometry for Microbial Proteomics*. (UK: Wiley).

Shallom, D., Leon, M., Bravman, T., Ben-David, A., Zaide, G., Belakhov, V., et al. (2005). Biochemical characterization and identification of the catalytic residues of a family 43 beta-D-xylosidase from *Geobacillus stearothermophilus T-6*. *Biochemistry* 44(1)**,** 387-397. doi: 10.1021/bi048059w.

Shanks, O.C., Kelty, C.A., Archibeque, S., Jenkins, M., Newton, R.J., McLellan, S.L., et al. (2011). Community structures of fecal bacteria in cattle from different animal feeding operations. *Appl Environ Microbiol* 77(9)**,** 2992-3001. doi: 10.1128/Aem.02988-10.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11)**,** 2498-2504. doi: 10.1101/gr.1239303.

Sheldon, R.A. (2014). Green and sustainable manufacture of chemicals from biomass: state of the art. *Green Chem* 16(3)**,** 950-963. doi: 10.1039/c3gc41935e.

Shelton, J.L., Akob, D.M., McIntosh, J.C., Fierer, N., Spear, J.R., Warwick, P.D., et al. (2016). Environmental drivers of differences in microbial community structure in crude oil reservoirs across a methanogenic gradient. *Front Microbiol* 7. doi: 10.3389/fmicb.2016.01535.

Shiro, S. (2001). "Chemical composition and distribution," in *Wood and Cellulosic Chemistry*, ed. D.N.H.a.N. Shiraishi. (New York: Marcel Dekker, Inc), 51-82.

Sierra-Alvarez, R., and Lettinga, G. (1991). The methanogenic toxicity of wastewater lignins and lignin related compounds. *J Chem Technol Biotechnol* 50(4)**,** 443-455. doi: 10.1002/jctb.280500403.

Sierraalvarez, R., Field, J.A., Kortekaas, S., and Lettinga, G. (1994). Overview of the anaerobic toxicity caused by organic forest industry waste-water pollutants. *Water Sci Technol* 29(5-6)**,** 353-363.

Simmons, C.W., Reddy, A.P., Simmons, B.A., Singer, S.W., and VanderGheynst, J.S. (2014). Effect of inoculum source on the enrichment of microbial communities on two lignocellulosic bioenergy crops under thermophilic and high-solids conditions. *J Appl Microbiol* 117(4)**,** 1025-1034. doi: 10.1111/jam.12609.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Res* 19(6)**,** 1117-1123. doi: 10.1101/gr.089532.108.

Singh, A., Yadav, R.D., Kaur, A., and Mahajan, R. (2012). An ecofriendly cost effective enzymatic methodology for deinking of school waste paper. *Bioresour Technol* 120**,** 322-327. doi: 10.1016/j.biortech.2012.06.050.

Singh, B., Bhat, T.K., and Sharma, O.P. (2001). Biodegradation of tannic acid in an in vitro ruminal system. *Livest Prod Sci* 68(2-3)**,** 259-262. doi: 10.1016/S0301-6226(00)00227-X.

Singh, R., Kumar, M., Mittal, A., and Mehta, P.K. (2016). Microbial enzymes: industrial progress in 21st century. *3 Biotech* 6. doi: 10.1007/s13205-016-0485-8.

Singhania, R.R., Patel, A.K., Sukumaran, R.K., Larroche, C., and Pandey, A. (2013). Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. *Bioresour Technol* 127**,** 500-507. doi: 10.1016/j.biortech.2012.09.012.

Sixta, H. (2006). *Handbook of pulp.* Wiley.

Sjöström, E. (1993). *Wood chemistry: Fundamentals and applications.* Academic Press.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., et al. (1985). Measurement of protein using bicinchoninic acid. *Anal Biochem* 150(1)**,** 76-85. doi: 10.1016/0003-2697(85)90442-7.

Smith, S.P., Bayer, E.A., and Czjzek, M. (2017). Continually emerging mechanistic complexity of the multi-enzyme cellulosome complex. *Curr Opin Struct Biol* 44**,** 151-160. doi: 10.1016/j.sbi.2017.03.009.

Soergel, D.A.W., Dey, N., Knight, R., and Brenner, S.E. (2012). Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J* 6(7)**,** 1440-1444. doi: 10.1038/ismej.2011.208.

Song, Y.H., Lee, K.T., Baek, J.Y., Kim, M.J., Kwon, M.R., Kim, Y.J., et al. (2017). Isolation and characterization of a novel glycosyl hydrolase family 74 (GH74) cellulase from the black goat rumen metagenomic library. *Folia Microbiol* 62(3)**,** 175-181. doi: 10.1007/s12223-016-0486-3.

Sorokin, D.Y., Tourova, T.P., Sukhacheva, M.V., Mardanov, A.V., and Ravin, N.V. (2012). Bacterial chitin utilisation at extremely haloalkaline conditions. *Extremophiles* 16(6)**,** 883-894. doi: 10.1007/s00792-012-0484-6.

Stam, M.R., Danchin, E.G.J., Rancurel, C., Coutinho, P.M., and Henrissat, B. (2006). Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. *Protein Eng Des Sel* 19(12)**,** 555-562. doi: 10.1093/protein/gzl044.

Stenius, P., Gullichsen, J., and Paulapuro, H. (2000). *Papermaking Science and Technology: Forest products chemistry. Book 3.* Jyväskylä: Fapet Oy.

Stone, B. (2001). "Cellulose: Structure and distribution," in *eLS*. John Wiley & Sons, Ltd).

Subramaniyan, S., and Prema, P. (2002). Biotechnology of microbial xylanases: Enzymology, molecular biology, and application. *Crit Rev Biotechnol* 22(1)**,** 33-64. doi: 10.1080/07388550290789450.

Sun, C., Zhang, S., Xin, F., Shanmugam, S., and Wu, Y.R. (2018). Genomic comparison of *Clostridium* species with the potential of utilizing red algal biomass for biobutanol production. *Biotechnol Biofuels* 11**,** 42. doi: 10.1186/s13068-018-1044-9.

Svartström, O., Alneberg, J., Terrapon, N., Lombard, V., de Bruijn, I., Malmsten, J., et al. (2017). Ninety-nine de novo assembled genomes from the moose (*Alces alce*s) rumen microbiome provide new insights into microbial plant biomass degradation. *ISME J*. doi: 10.1038/ismej.2017.108.

Symons, G.E., and Buswell, A.M. (1933). The methane fermentation of carbohydrates. *J Am Chem Soc* 55**,** 2028-2036. doi: 10.1021/ja01332a039.

Tailford, L.E., Crost, E.H., Kavanaugh, D., and Juge, N. (2015). Mucin glycan foraging in the human gut microbiome. *Front Genet* 6**,** 81. doi: 10.3389/fgene.2015.00081.

Takasaki, K., Miura, T., Kanno, M., Tamaki, H., Hanada, S., Kamagata, Y., et al. (2013). Discovery of glycoside hydrolase enzymes in an avicel-adapted forest soil fungal community by a metatranscriptomic approach. *PLoS One* 8(2)**,** e55485. doi: 10.1371/journal.pone.0055485.

Tang, Y.Q., Shigematsu, T., Ikbal, Morimura, S., and Kida, K. (2004). The effects of micro-aeration on the phylogenetic diversity of microorganisms in a thermophilic anaerobic municipal solid-waste digester. *Water Res* 38(10)**,** 2537-2550. doi: 10.1016/j.watres.2004.03.012.

Tarmadi, D., Tobimatsu, Y., Yamamura, M., Miyamoto, T., Miyagawa, Y., Umezawa, T., et al. (2018). NMR studies on lignocellulose deconstructions in the digestive system of the lower termite *Coptotermes formosanus* Shiraki. *Sci Rep* 8. doi: 10.1038/s41598-018-19562-0.

Tauseef, S.M., Abbasi, T., and Abbasi, S.A. (2013). Energy recovery from wastewaters with high-rate anaerobic digesters. *Renew Sust Energ Rev* 19**,** 704-741. doi: 10.1016/j.rser.2012.11.056.

Tejirian, A., and Xu, F. (2011). Inhibition of enzymatic cellulolysis by phenolic compounds. *Enzyme Microb Technol* 48(3)**,** 239-247. doi: 10.1016/j.enzmictec.2010.11.004.

Temple, M.J., Cuskin, F., Baslé, A., Hickey, N., Speciale, G., Williams, S.J., et al. (2017). A *Bacteroidetes* locus dedicated to fungal 1,6-β-glucan degradation: unique substrate conformation drives specificity of the key endo-1,6-β-glucanase. *J Biol Chem*. doi: 10.1074/jbc.M117.787606.

Terrapon, N., Lombard, V., Drula, E., Lapebie, P., Al-Masaudi, S., Gilbert, H.J., et al. (2017). PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res*. doi: 10.1093/nar/gkx1022.

Terrapon, N., Lombard, V., Gilbert, H.J., and Henrissat, B. (2015). Automatic prediction of polysaccharide utilization loci in *Bacteroidetes* species. *Bioinformatics* 31(5)**,** 647-655. doi: 10.1093/bioinformatics/btu716.

Thompson, G., Swain, J., Kay, M., and Forster, C.F. (2001). The treatment of pulp and paper mill effluent: a review. *Bioresour Technol* 77(3)**,** 275-286. doi: 10.1016/S0960-8524(00)00060-2.

Tian, L., Liu, S.J., Wang, S., and Wang, L.S. (2016). Ligand-binding specificity and promiscuity of the main lignocellulolytic enzyme families as revealed by active-site architecture analysis. *Sci Rep* 6. doi: 10.1038/srep23605.

Tien, M., and Kirk, T.K. (1983). Lignin-degrading enzyme from the Hymenomycete *Phanerochaete chrysosporium* Burds. *Science* 221(4611)**,** 661-662. doi: 10.1126/science.221.4611.661.

Tiwari, R., Nain, L., Labrou, N.E., and Shukla, P. (2018). Bioprospecting of functional cellulases from metagenome for second generation biofuel production: a review. *Crit Rev Microbiol* 44(2)**,** 244-257. doi: 10.1080/1040841x.2017.1337713.

Toczylowska-Maminska, R. (2017). Limits and perspectives of pulp and paper industry wastewater treatment - A review. *Renew Sust Energ Rev* 78**,** 764-772. doi: 10.1016/j.rser.2017.05.021.

Tolbert, A., and Ragauskas, A.J. (2017). Advances in understanding the surface chemistry of lignocellulosic biomass via time-of-flight secondary ion mass spectrometry. *Energy Sci Eng* 5(1)**,** 5-20. doi: 10.1002/ese3.144.

Tomme, P., Creagh, A.L., Kilburn, D.G., and Haynes, C.A. (1996). Interaction of polysaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi CenC* .1. Binding specificity and calorimetric analysis. *Biochemistry* 35(44)**,** 13885-13894. doi: 10.1021/bi961185i.

Tong, H., Chen, M.J., Li, F.B., Liu, C.S., and Liao, C.Z. (2017). Changes in the microbial community during repeated anaerobic microbial dechlorination of pentachlorophenol. *Biodegradation* 28(2-3)**,** 219-230. doi: 10.1007/s10532-017-9791-z.

Tracy, B.P., Jones, S.W., Fast, A.G., Indurthi, D.C., and Papoutsakis, E.T. (2012). *Clostridia*: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. *Curr Opin Biotechnol* 23(3)**,** 364-381. doi: 10.1016/j.copbio.2011.10.008.

Tringe, S.G., and Hugenholtz, P. (2008). A renaissance for the pioneering 16S rRNA gene. *Curr Opin Microbiol* 11(5)**,** 442-446. doi: 10.1016/j.mib.2008.09.011.

Tu, M.B., Pan, X.J., and Saddler, J.N. (2009). Adsorption of cellulase on cellulolytic enzyme lignin from lodgepole pine. *J Agric Food Chem* 57(17)**,** 7771-7778. doi: 10.1021/Jf901031m.

Turriff, S.L. (2015). "Ten years on, forest biorefinery a “modest success”", in: *Canadian Biomass.* (Simcoe, Canada: Canadian Biomass).

Vaaje-Kolstad, G., Westereng, B., Horn, S.J., Liu, Z., Zhai, H., Sorlie, M., et al. (2010). An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* 330(6001)**,** 219-222. doi: 10.1126/science.1192231.

van der Gast, C.J., Walker, A.W., Stressmann, F.A., Rogers, G.B., Scott, P., Daniels, T.W., et al. (2011). Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J* 5(5)**,** 780-791. doi: 10.1038/ismej.2010.175.

van der Lelie, D., Taghavi, S., McCorkle, S.M., Li, L.L., Malfatti, S.A., Monteleone, D., et al. (2012). The metagenome of an anaerobic microbial community decomposing poplar wood chips. *PLoS One* 7(5)**,** e36740. doi: 10.1371/journal.pone.0036740.

Van Heiningen, A. (2006). Converting a kraft pulp mill into an integrated forest biorefinery. *Pulp Pap-Canada* 107(6)**,** 38-43.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. (2010). Lignin Biosynthesis and Structure. *Plant Physiol* 153(3)**,** 895-905. doi: 10.1104/pp.110.155119.

Vanwonterghem, I., Jensen, P.D., Dennis, P.G., Hugenholtz, P., Rabaey, K., and Tyson, G.W. (2014). Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. *ISME J* 8(10)**,** 2015-2028. doi: 10.1038/ismej.2014.50.

Venditto, I., Luis, A.S., Rydahl, M., Schuckel, J., Fernandes, V.O., Vidal-Melgosa, S., et al. (2016). Complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition. *Proc Natl Acad Sci U S A* 113(26)**,** 7136-7141. doi: 10.1073/pnas.1601558113.

Vidal, G., and Diez, M.C. (2003). Influence of feedstock and bleaching technologies on methanogenic toxicity of kraft mill wastewater. *Water Sci Technol* 48(6)**,** 149-155.

Voragen, A.G.J., Coenen, G.J., Verhoef, R.P., and Schols, H.A. (2009). Pectin, a versatile polysaccharide present in plant cell walls. *Struct Chem* 20(2)**,** 263-275. doi: 10.1007/s11224-009-9442-z.

Vuong, T.V., and Master, E.R. (2014). Fusion of a xylan-binding module to gluco-oligosaccharide oxidase increases activity and promotes stable immobilization. *PLoS One* 9(4). doi: 10.1371/journal.pone.0095170.

Walter, A., Silberberger, S., Juarez, M.F.D., Insam, H., and Franke-Whittle, I.H. (2016). Biomethane potential of industrial paper wastes and investigation of the methanogenic communities involved. *Biotechnol Biofuels* 9. doi: 10.1186/s13068-016-0435-z.

Wang, C., Dong, D., Wang, H.S., Muller, K., Qin, Y., Wang, H.L., et al. (2016a). Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of *Actinobacteria* in lignocellulose decomposition. *Biotechnol Biofuels* 9. doi: 10.1186/s13068-016-0440-2.

Wang, C.C., Lee, J.C., Luo, S.Y., Kulkarni, S.S., Huang, Y.W., Lee, C.C., et al. (2007). Regioselective one-pot protection of carbohydrates. *Nature* 446(7138)**,** 896-899. doi: 10.1038/nature05730.

Wang, K., Pereira, G.V., Cavalcante, J.J.V., Zhang, M.L., Mackie, R., and Cann, I. (2016b). *Bacteroides intestinalis DSM 17393*, a member of the human colonic microbiome, upregulates multiple endoxylanases during growth on xylan. *Sci Rep* 6. doi: 10.1038/srep34360.

Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., et al. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450(7169)**,** 560-U517. doi: 10.1038/nature06269.

White, B.A., Lamed, R., Bayer, E.A., and Flint, H.J. (2014). Biomass utilization by gut microbiomes. *Annu Rev Microbiol* 68**,** 279-296. doi: 10.1146/annurev-micro-092412-155618.

Wilkens, C., Busk, P.K., Pilgaard, B., Zhang, W.J., Nielsen, K.L., Nielsen, P.H., et al. (2017). Diversity of microbial carbohydrate-active enzymes in Danish anaerobic digesters fed with wastewater treatment sludge. *Biotechnol Biofuels* 10. doi: 10.1186/s13068-017-0840-y.

Wong, M.T., Wang, W., Couturier, M., Razeq, F.M., Lombard, V., Lapebie, P., et al. (2017). Comparative metagenomics of cellulose- and poplar hydrolysate-degrading microcosms from gut microflora of the Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*) after long-term enrichment. *Front Microbiol* 8(2504). doi: 10.3389/fmicb.2017.02504.

Wong, M.T., Wang, W.J., Lacourt, M., Couturier, M., Edwards, E.A., and Master, E.R. (2016). Substrate-driven convergence of the microbial community in lignocellulose-amended

enrichments of gut microflora from the Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*). *Front Microbiol* 7. doi: 10.3389/Fmicb.2016.00961.

Wongwilaiwalin, S., Laothanachareon, T., Mhuantong, W., Tangphatsornruang, S., Eurwilaichitr, L., Igarashi, Y., et al. (2013). Comparative metagenomic analysis of microcosm structures and lignocellulolytic enzyme systems of symbiotic biomass-degrading consortia. *Appl Microbiol Biotechnol* 97(20)**,** 8941-8954. doi: 10.1007/s00253-013-4699-y.

Xiao, C.W., and Anderson, C.T. (2013). Roles of pectin in biomass yield and processing for biofuels. *Front Plant Sci* 4. doi: 10.3389/fpls.2013.00067.

Ximenes, E., Kim, Y., Mosier, N., Dien, B., and Ladisch, M. (2010). Inhibition of cellulases by phenols. *Enzyme Microb Technol* 46(3-4)**,** 170-176. doi: 10.1016/j.enzmictec.2009.11.001.

Ximenes, E., Kim, Y., Mosier, N., Dien, B., and Ladisch, M. (2011). Deactivation of cellulases by phenols. *Enzyme Microb Technol* 48(1)**,** 54-60. doi: 10.1016/j.enzmictec.2010.09.006.

Xu, F., Yu, J.M., Tesso, T., Dowell, F., and Wang, D.H. (2013). Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review. *Appl Energ* 104**,** 801-809. doi: 10.1016/j.apenergy.2012.12.019.

Yang, G., and Jaakkola, P. (2011). "Wood chemistry and isolation of extractives from wood". (South Karelia, Finland: Saimaa University of Applied Sciences).

Yu, Y., and Christopher, L.P. (2017). Detoxification of hemicellulose-rich pretreated poplar by polymeric resins for improved ethanol fermentability. *Fuel* 203**,** 187-196. doi: 10.1016/j.fuel.2017.04.118.

Zhang, G., Liu, P., Zhang, L., Wei, W., Wang, X., Wei, D., et al. (2016). Bioprospecting metagenomics of a microbial community on cotton degradation: Mining for new glycoside hydrolases. *J Biotechnol* 234**,** 35-42. doi: 10.1016/j.jbiotec.2016.07.017.

Zhang, H., Fangel, J.U., Willats, W.G.T., Selig, M.J., Lindedam, J., Jorgensen, H., et al. (2014a). Assessment of leaf/stem ratio in wheat straw feedstock and impact on enzymatic conversion. *Gcb Bioenergy* 6(1)**,** 90-96. doi: 10.1111/gcbb.12060.

Zhang, K., Chen, X., Schwarz, W.H., and Li, F. (2014b). Synergism of glycoside hydrolase secretomes from two thermophilic bacteria cocultivated on lignocellulose. *Appl Environ Microbiol* 80(8)**,** 2592-2601. doi: 10.1128/AEM.00295-14.

Zhang, M.L., Chekan, J.R., Dodd, D., Hong, P.Y., Radlinski, L., Revindran, V., et al. (2014c). Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. *Proc Natl Acad Sci U S A* 111(35)**,** E3708-E3717. doi: 10.1073/pnas.1406156111.

Zhu, L.F., Wu, Q., Dai, J.Y., Zhang, S.N., and Wei, F.W. (2011). Evidence of cellulose metabolism by the giant panda gut microbiome. *Proc Natl Acad Sci U S A* 108(43)**,** 17714-17719. doi: 10.1073/pnas.1017956108.

Zhu, Y., Lee, Y.Y., and Elander, R.T. (2005). Optimization of dilute-acid pretreatment of corn stover using a high-solids percolation reactor. *Appl Biochem Biotechnol* 121-124**,** 1045-1054.

Zhuo, S., Yan, X., Liu, D., Si, M., Zhang, K., Liu, M., et al. (2018). Use of bacteria for improving the lignocellulose biorefinery process: importance of pre-erosion. *Biotechnol Biofuels* 11(1)**,** 146. doi: 10.1186/s13068-018-1146-4.

Zverlov, V.V., Volkov, I.Y., Velikodvorskaya, G.A., and Schwarz, W.H. (2001). The binding pattern of two carbohydrate-binding modules of laminarinase Lam16A from *Thermotoga neapolitana*: differences in beta-glucan binding within family CBM4. *Microbiology-Sgm* 147**,** 621-629. doi: 10.1099/00221287-147-3-621.

Appendices

Supplementary Method

Supplementary method for the calculation of stoichiometric maximum biogas volume per mg COD carbohydrates for chapters 3 and 5

According to Buswell's equation (Symons and Buswell, 1933),

$C_nH_aO_bN_d + (n-a/4-b/c+3d/4)\ H_2O \rightarrow (n/2+a/8-b/4-3d/8)\ CH_4 + (n/2-a/8+b/4+3d/8)\ CO_2 + dNH_3$

Using glucose as an example for carbohydrate, we have $C_6H_{12}O_6 \rightarrow 3\ CH_4 + 3\ CO_2$. Assuming ideal conditions and no dissolution of CO_2 in the medium, 1 mole of glucose yields 6 mole of gases. Using the ideal gas law (PV = nRT or V = nRT÷P), 1 mole of glucose yields 6 mol × 298 K × 0.082057 L atm K^{-1} mol^{-1} ÷ 1 atm = 146.7 L biogas, or 1 mg of glucose yields 0.8142 ml biogas. Consider the oxidation of glucose: $C_6H_{12}O_6 + 6\ O_2 \rightarrow 6\ CO_2 + 6\ H_2O$, the COD of glucose = 6 × 32 g/mol O_2 ÷ 180.1559 g/mol glucose = 1.066 g O_2/g glucose. Therefore, the stoichiometric maximum biogas yield is 0.8142 ml biogas/mg glucose ÷ 1.066 g O_2/g glucose = 0.764 ml biogas/mg COD.

Consider sodium lignosulphonate ($C_{20}H_{24}Na_2O_{10}S_2$) (National Center for Biotechnology Information, 2016b), 1 mole of this substrate yields 489.1 L biogas containing 10.5 mole of CH_4 and 9.5 mole of CO_2. Based on the measured COD of sodium lignosulphonate (Table S3.1), the stoichiometric maximum biogas yield is 0.58 ml biogas/mg COD. Similarly, 1 mole of tannic acid ($C_{76}H_{52}O_{46}$) (National Center for Biotechnology Information, 2016a) generates 1858.4 L biogas containing 33 moles of CH_4 and 43 mole of CO_2 based on stoichiometry, and the calculated maximum biogas yield is 0.90 ml biogas/mg COD using the measured COD of tannic acid (Table S3.2). In proportion of the substrates added at each enrichment phase, the stoichiometric

maximum biogas yields (ranged from 0.69 to 0.82 ml biogas/mg COD substrate) were derived (Table S3.2). Actual yield may differ owing to factors such as limited biodegradability of compounds and solubility of CO_2 in liquid.

Supplementary Tables

Table S3.1 Measured COD of enrichment substrates (± standard deviation)

Enrichment substrate	*COD content (g COD/g substrate)*
Cellulose	1.22 ± 0.12
Sodium lignosulphonate	1.54 ± 0.11
Tannic acid	1.22 ± 0.00
Pretreated poplar	0.93 ± 0.30

Table S3.2 Lignocellulosic amendments for enrichment microcosms and stoichiometric maximum biogas yields

Enrichment	Cellulose (C)		Lignosulphonate (L), Tannic acid (T), or Poplar hydrolysate (PH)		Total COD (mgCOD)	Estimated stoichiometric maximum biogas yield (ml biogas/mg COD substrate)
	mg per bottle	mgCOD per bottle	mg per bottle	mgCOD per bottle		
Phase 1						
C	28.4	33.7	-	-	33.7	0.76
CL	22.8	27	5.3	8.1	35.1	0.72
CT	22.8	27	17.7	22	49	0.82
P	-	-	28.4	26.6	26.6	-
Phase 2-3						
C	142.2	168.5	-	-	168.5	0.76
CL	113.8	134.9	26.3	40.5	175.4	0.72
CT	136.5	161.8	17.7	22	183.8	0.78
P	-	-	142.2	132.8	132.8	-
Phase 4						
C	113.8	134.9	-	-	134.9	0.76
CL	113.8	134.9	26.3	40.5	175.4	0.72
CT	113.8	134.9	88.5	109.9	244.8	0.82
P	-	-	136.5	127.5	127.5	-
Phase 5-10						
C	125	152.5	-	-	134.9	0.76
CL	125	152.5	90	109.8	262.3	0.69
CT	125	152.5	26.3	40.5	193	0.79
P	-	-	136	126.5	126.5	-

Table S3.3 Dates and duration of microcosm enrichment phases 1 to 4, 7, and 9

	Inocula	***Beaver dropping***				***Moose rumen***			
	Substrates	**C**	**CL**	**CT**	**P**	**C**	**CL**	**CT**	**P**
Phase 1	Start date[1]	10-02-12	10-02-12	10-02-12	10-02-12	09-10-13	09-10-13	09-10-13	09-10-13
	Duration (d)	124	124	124	124	201	201	201	201
Phase 2	Start date	10-07-23	10-07-23	10-07-23	10-07-23	10-08-12	10-08-12	10-08-12	10-08-12
	Duration (d)	55	55	55	55	62	62	62	62
Phase 3	Start date	10-09-30	10-09-30	10-09-30	10-09-30	10-11-17	10-11-17	10-11-17	10-11-17
	Duration (d)	90	90	90	90	58	58	58	58
Phase 4	Start date	11-01-06	11-01-06	11-01-06	11-01-06	11-01-15	11-01-15	11-01-15	11-01-15
	Duration (d)	30	30	30	30	30	30	30	30
Phase 7	Start date	11-12-05	11-12-05	11-12-05	11-12-05	11-12-05	11-12-05	11-12-05	11-12-05
	Duration (d)	259	176	259	259	259	176	259	259
Phase 9	Start date	12-11-06	12-11-12	12-11-01	12-10-20	12-11-06	12-11-08	12-11-01	12-10-20
	Duration (d)	78	71	82	94	78	75	82	94

[1] Dates are displayed as YY-MM-DD.

Table S3.4 Multiplex barcodes, DNA concentration, and numbers of reads for amplicon samples prepared from beaver dropping, moose rumen and their enrichment microcosms

Enrichment condition	*Multiplex barcode*	*Amplicon DNA concentration (µg/mL)*	*Number of reads*
Beaver dropping			
Inoculum	TAGTGTAGAT	31.7	4504
	TCGCACTAGT	38.2	5116
	TCTATACTAT	43.4	5213
Cellulose	ACATACGCGT	138	7471
	ACGCGAGTAT	116.3	7828
	ACTACTATGT	168.9	8312
Cellulose + Lignosulphonate	TACGAGTATG	158.9	9475
	TAGAGACGAG	133.1	8191
	TCGTCGCTCG	90.3	5921
Cellulose + Tannic acid	ATAGAGTACT	29.1	-
	CACGCTACGT	30.5	-
Pretreated poplar	AGTACGCTAT	99.1	8699
	ATAGAGTACT	116.4	9232
	CACGCTACGT	51.2	4441
Moose rumen			
Inoculum	CGACGTGACT	34.2	3198
	TACACGTGAT	25.8	3444
	TACGCTGTCT	20.1	5339
Cellulose	TGTACTACTC	154.9	8051
	ACGACTACAG	133.8	6972
	CGTAGACTAG	163.6	8523
Cellulose + Lignosulphonate	TGATACGTCT	154.3	8694
	CGAGAGATAC	116.9	6378
	TCTACGTAGC	129.4	7100
Cellulose + Tannic acid	AGCGTCGTCT	23.6	9144
	AGTACGCTAT	65.2	9128
Pretreated poplar	CAGTAGACGT	85.2	8522
	TACACGTGAT	68.2	5739
	TACGCTGTCT	100.5	8673

Table S3.5 Statistics of 16S rRNA pyrotag sequencing runs

Sample	*Total number of reads*	*Number of OTUs*	*Number of OTUs with abundance ≥ 0.5% (sum of represented abundances)*
Beaver droppings			
Inoculum	14564	415	22 (72.9%)
Cellulose	23268	685	21 (77.6%)
Cellulose + Lignosulphonate	23024	840	23 (70.2%)
Cellulose + Tannic acid[1]	12	-	-
Pretreated poplar	22077	581	21 (77.9%)
Moose rumen			
Inoculum	11291	967	29 (42.2%)
Cellulose	23091	734	22 (73.3%)
Cellulose + Lignosulphonate	21869	644	18 (76.4%)
Cellulose + Tannic acid	18003	346	17 (80.5%)
Pretreated poplar	22614	588	18 (76.0%)

[1] Dataset for beaver droppings enriched on cellulose plus tannic acid was removed from downstream analysis due to low number of reads (less than 15).

Table S4.1 Metagenomic DNA extraction for enrichment cultures from beaver dropping and moose rumen

Sample	Concentration (ng/μL)	A_{260}/A_{280}
beaver dropping-cellulose	73.7	2.00
beaver dropping-pretreated poplar	47.7	2.00
moose rumen-cellulose	43.8	1.95
moose rumen-pretreated poplar	64.9	1.86

Table S4.2 Annotation of CAZymes and CBMs in the metagenomes

	BD-C	BD-P	MR-C	MR-P
CBM2	1	6	2	6
CBM3	33	32	31	141
CBM4	17	12	10	6
CBM6	15	24	21	22
CBM8	0	0	1	2
CBM9	24	8	8	18
CBM13	4	6	8	5
CBM14	1	0	0	0
CBM16	12	4	2	6
CBM17	0	0	2	0
CBM20	3	18	5	6
CBM22	7	6	6	7
CBM23	4	0	1	0
CBM25	7	4	6	9
CBM26	0	3	0	3
CBM27	1	0	1	0
CBM28	0	0	1	0
CBM30	2	2	2	4
CBM32	23	33	15	26
CBM34	5	6	0	4
CBM35	5	10	5	12
CBM38	2	5	1	2
CBM40	0	0	0	1
CBM41	3	1	0	2
CBM44	1	1	1	1
CBM45	1	0	0	0

CBM47	1	0	0	3
CBM48	32	42	18	28
CBM50	162	179	114	173
CBM51	14	0	0	1
CBM54	0	0	4	0
CBM57	13	0	0	1
CBM58	0	3	0	4
CBM61	0	4	0	4
CBM62	0	4	2	2
CBM63	1	1	1	1
CBM66	32	11	6	5
CBM67	13	3	3	3
CBM70	1	0	0	1
CBM77	0	1	1	0
CE1	28	63	13	51
CE2	5	4	1	5
CE3	8	6	4	9
CE4	31	41	20	44
CE6	7	6	2	6
CE7	16	10	6	9
CE8	7	7	5	1
CE9	23	19	12	11
CE11	9	14	5	11
CE12	3	5	4	2
CE14	46	25	24	30
CE15	8	15	1	9
CE16	0	4	0	0
GT1	21	18	15	14
GT2	432	336	308	326
GT3	1	8	2	2
GT4	344	307	298	263
GT5	26	19	15	16
GT6	1	0	0	1
GT7	0	0	0	1
GT8	3	2	5	3
GT9	84	24	41	22
GT10	0	0	1	3
GT11	7	3	2	4
GT17	1	0	0	0
GT19	12	19	6	13
GT20	3	10	9	9
GT21	4	0	1	0

GT25	2	2	5	0
GT26	20	8	10	8
GT28	39	29	26	36
GT30	13	14	7	13
GT32	4	1	4	1
GT35	25	36	17	19
GT39	3	7	3	4
GT41	59	1	0	0
GT48	2	0	0	0
GT50	0	1	0	1
GT51	33	54	26	41
GT55	0	1	0	0
GT56	0	0	0	1
GT66	1	2	12	3
GT73	1	0	1	0
GT74	2	0	2	0
GT76	8	2	4	5
GT81	6	6	6	7
GT83	28	22	14	22
GT84	2	7	3	4
GT87	3	0	0	0
GT89	8	0	5	1
PL1	19	4	5	2
PL2	2	0	0	0
PL3	1	0	0	0
PL6	1	0	2	0
PL7	2	0	0	0
PL8	0	1	0	0
PL9	26	13	4	4
PL10	2	1	0	0
PL11	9	2	1	3
PL12	15	2	4	7
PL14	0	0	0	1
PL15	3	0	0	1
PL17	4	1	1	2
PL21	2	0	0	0
PL22	2	0	0	3
GH1	28	7	20	21
GH2	50	88	35	55
GH3	63	102	36	90
GH4	24	11	25	15
GH5	14	5	2	11

GH5_1	4	5	4	13
GH5_2	2	5	2	4
GH5_4	0	10	1	12
GH5_5	0	7	0	0
GH5_7	6	6	1	7
GH5_8	0	0	0	1
GH5_12	2	3	4	1
GH5_13	1	5	1	3
GH5_17	0	0	0	1
GH5_19	1	0	1	0
GH5_22	1	3	6	3
GH5_25	3	2	3	1
GH5_36	1	3	1	1
GH5_37	0	2	1	3
GH5_40	0	1	0	1
GH5_44	0	1	1	0
GH5_45	0	1	0	0
GH5_46	2	4	2	6
GH5_52	3	0	2	0
GH8	5	5	2	13
GH9	28	36	25	64
GH10	19	20	15	13
GH11	2	4	4	4
GH12	0	2	0	3
GH13	36	39	20	36
GH13_2	1	0	0	0
GH13_3	0	4	0	0
GH13_4	4	1	2	0
GH13_6	0	0	0	1
GH13_7	0	0	0	1
GH13_8	1	4	2	2
GH13_9	13	7	6	5
GH13_10	1	5	0	0
GH13_11	6	8	5	2
GH13_13	0	0	0	1
GH13_14	4	5	1	4
GH13_16	6	8	3	2
GH13_18	6	6	4	2
GH13_19	0	2	0	1
GH13_20	12	8	7	11
GH13_21	1	1	0	1
GH13_23	6	11	5	9

GH13_26	1	5	0	0
GH13_29	1	0	0	0
GH13_31	5	3	1	2
GH13_33	0	4	0	0
GH13_36	2	3	1	4
GH13_38	1	5	2	1
GH13_39	0	1	0	1
GH13_41	1	0	0	0
GH14	1	0	0	0
GH15	8	7	9	6
GH16	11	22	6	11
GH17	2	2	1	2
GH18	21	27	19	38
GH19	0	2	0	1
GH20	35	18	13	17
GH23	57	57	31	42
GH24	2	11	2	3
GH25	19	9	5	5
GH26	8	11	8	9
GH27	9	24	3	18
GH28	13	20	9	4
GH29	25	27	5	5
GH30	3	2	1	9
GH30_1	3	2	4	3
GH30_2	1	1	1	0
GH30_3	4	9	2	1
GH30_8	1	1	2	2
GH31	17	31	19	34
GH32	10	10	4	6
GH33	23	14	1	2
GH35	12	17	3	12
GH36	21	18	8	15
GH37	1	1	0	1
GH38	23	17	5	20
GH39	23	8	7	9
GH42	12	5	4	8
GH43	0	68	1	1
GH43_1	1	0	2	2
GH43_2	0	0	0	4
GH43_3	1	0	3	0
GH43_4	1	0	0	1
GH43_5	1	0	0	1

GH43_8	1	0	0	0
GH43_10	2	0	3	2
GH43_11	3	0	0	4
GH43_12	7	0	3	3
GH43_16	1	0	1	1
GH43_17	1	0	6	5
GH43_18	1	0	0	0
GH43_20	1	0	1	1
GH43_22	2	0	0	0
GH43_24	1	0	1	3
GH43_26	2	0	1	4
GH43_28	1	0	0	10
GH43_29	1	0	1	1
GH43_30	1	0	0	0
GH43_31	2	0	2	1
GH43_33	1	0	0	0
GH43_34	2	0	0	0
GH43_35	2	0	1	4
GH44	2	2	2	3
GH46	0	0	0	1
GH48	2	2	6	18
GH50	4	4	3	6
GH51	20	28	12	20
GH52	2	2	5	2
GH53	3	20	3	21
GH55	2	4	0	1
GH57	41	19	19	18
GH62	1	1	1	1
GH63	14	7	8	4
GH65	5	12	9	17
GH66	2	1	3	1
GH67	6	10	3	3
GH73	18	12	2	10
GH74	33	3	20	7
GH76	3	5	4	2
GH77	15	25	14	15
GH78	21	32	3	4
GH81	3	1	1	1
GH84	5	1	0	0
GH87	0	2	1	1
GH88	19	6	1	2
GH89	1	1	1	0

GH92	18	43	9	9
GH93	0	4	1	0
GH94	24	22	18	26
GH95	8	17	3	4
GH96	0	0	0	1
GH97	10	14	2	4
GH98	0	1	0	0
GH99	2	1	0	1
GH100	1	0	0	0
GH102	0	1	0	3
GH103	1	9	0	1
GH105	9	18	12	8
GH106	4	23	0	0
GH107	1	0	0	0
GH108	6	1	1	0
GH109	7	8	1	5
GH110	3	1	0	0
GH112	1	0	0	0
GH113	5	1	4	3
GH114	1	0	0	0
GH115	5	9	3	4
GH116	7	7	4	4
GH117	2	0	0	0
GH119	1	0	0	0
GH120	2	0	1	1
GH121	0	0	1	0
GH123	11	4	0	3
GH125	4	5	5	4
GH126	0	0	1	0
GH127	14	27	2	21
GH128	0	2	2	0
GH129	6	0	0	1
GH130	26	27	19	21
GH133	4	12	5	3
GH139	1	0	0	0
GH140	0	2	0	0
GH141	1	2	1	1
GH142	1	1	0	0

Table S4.3 Relative abundances of polysaccharide-active CAZymes that contributed to > 0.2 or < -0.2 component loading in PC1, PC2, and PC3 in the PCA plot

	Component loading[1]			Relative abundances of plant polysaccharide-active CAZys annotated in metagenomes (%)										
	PC1	PC2	PC3	BD-C	MR-C	BD-P	MR-P	Moose	Cow	Reindeer	Wallaby	Sheep	Panda	Termite
GH3	0.36	-0.49	-0.25	2.4	2.2	4.0	4.3	3.5	4.7	5.5	4.9	3.5	1.5	5.1
GH5	0.31			1.5	2.0	2.5	3.2	2.2	2.7	2.2	1.7	1.6	0.3	6.2
GH2	0.29	0.31		1.9	2.1	3.4	2.6	3.9	4.3	4.2	4.0	5.2	1.1	2.8
GH43	0.29	0.36	-0.47	1.4	1.6	2.7	2.3	5.2	4.5	6.1	4.0	2.9	2.4	3.3
CE1			0.22	1.1	0.8	2.5	2.4	1.0	1.9	1.6	1.3	1.5	0.1	0.3
GH9		-0.28		1.1	1.5	1.4	3.0	0.5	1.6	0.9	0.4	0.6	0.0	2.1
GH94		-0.38	-0.27	0.9	1.1	0.9	1.2	0.6	1.1	1.2	1.3	0.8	0.9	4.3
GH10		-0.33	-0.23	0.7	0.9	0.8	0.6	0.8	1.6	1.0	0.8	0.8	0.1	4.1
GH78		0.26		0.8	0.2	1.3	0.2	1.3	1.6	1.2	2.4	2.1	0.3	0.2
CE4	-0.31			1.2	1.2	1.6	2.1	0.6	0.9	0.4	1.0	1.6	4.4	0.9
GH1	-0.58		-0.61	1.1	1.2	0.3	1.0	0.8	0.4	0.6	3.4	0.7	7.4	1.2

[1]Only values > 0.2 or < -0.2 are shown.

Table S4.4 Domain architecture, number of gene count and relative abundances of top 20 most abundant multi-modular CAZymes in the metagenomes

	BD-C		BD-P		MR-C		MR-P	
Multimodular CAZymes	**Count**	**%**	**Count**	**%**	**Count**	**%**	**Count**	**%**
CBM50-CBM50-GH18	2	1.2%	12	6.2%	5	5.8%	11	5.0%
CBM48-GH13_9	13	7.6%	7	3.6%	6	7.0%	4	1.8%
CBM48-GH13_11	6	3.5%	7	3.6%	5	5.8%	2	0.9%
CBM48-CE1	1	0.6%	8	4.1%	0	0.0%	10	4.6%
CBM67-GH78	9	5.3%	2	1.0%	3	3.5%	3	1.4%
GT4-GT4	4	2.3%	4	2.1%	1	1.2%	7	3.2%
GT2-GT2	7	4.1%	0	0.0%	5	5.8%	3	1.4%
GH9-CBM3-CBM3	0	0.0%	0	0.0%	0	0.0%	14	6.4%
GH9-CBM3	2	1.2%	2	1.0%	2	2.3%	7	3.2%
GT84-GH94	1	0.6%	6	3.1%	2	2.3%	3	1.4%
GH23-CBM50-CBM50	2	1.2%	4	2.1%	3	3.5%	3	1.4%
CBM48-GH13	3	1.8%	2	1.0%	4	4.7%	2	0.9%
GH43_28-CBM32	0	0.0%	5	2.6%	0	0.0%	6	2.8%
CBM34-GH13_20	3	1.8%	4	2.1%	0	0.0%	3	1.4%
CBM66-PL9	10	5.8%	0	0.0%	0	0.0%	0	0.0%
CBM66-PL1	10	5.8%	0	0.0%	0	0.0%	0	0.0%
CBM20-CBM20-GH77	1	0.6%	7	3.6%	2	2.3%	0	0.0%
GT2-GT4	5	2.9%	1	0.5%	3	3.5%	0	0.0%
GH9-CE4	1	0.6%	2	1.0%	1	1.2%	5	2.3%
CBM50-GH18	3	1.8%	1	0.5%	1	1.2%	3	1.4%

Table S4.5 Domain architecture for all CAZy-dockerins identified in the metagenomes

Domain architecture	BD-C	BD-P	MR-C	MR-P
CBM13-DOC1	1	1	1	1
CBM30-GH9-DOC1	2	2	2	2
CBM35-DOC1	1	1	1	1
CBM35-GH26-DOC1	1	1	1	1
CBM4-DOC1	0	0	2	0
CBM4-GH9-DOC1	1	1	1	1
CBM6-DOC1	0	0	1	0
CBM9-DOC1	1	1	1	4
CE12-DOC1	2	2	2	2
CE15-DOC1	1	1	1	1
CE1-DOC1	2	2	2	2
CE3-DOC1	3	3	2	3
CE8-DOC1	1	1	1	1
DOC1-CBM4-GH16	1	1	1	1
DOC1-CBM63	1	1	1	1
DOC1-CE2	1	1	1	1
DOC1-CE6	1	1	1	1
DOC1-GH3	0	0	0	1
DOC1-GH3-CBM4-CBM4-CBM4-GH9	1	1	1	1
GH10-DOC1-CE3	1	1	1	1
GH11-CBM6-DOC1-CE6	1	1	1	1
GH11-DOC1-GH10	1	1	1	1
GH16-DOC1	1	1	1	1
GH18-DOC1	1	1	1	1
GH28-DOC1	1	1	1	1
GH30_8-DOC1	1	1	1	1
GH30-DOC1	1	1	1	1
GH31-CBM32-DOC1	1	0	0	0
GH43_10-CBM6-DOC1	1	1	1	1
GH43_20-DOC1	1	1	1	1
GH43_24-CBM13-DOC1	0	0	1	0
GH43_24-DOC1	1	1	0	1
GH43_26-DOC1	1	1	1	1
GH44-DOC1	1	1	1	1
GH48-DOC1	2	2	2	2
GH5_1-DOC1	2	2	2	2
GH53-DOC1	1	1	1	1

GH5-CBM32-DOC1	1	1	1	1
GH5-DOC1	0	0	0	1
GH62-CBM6-DOC1-CE3	1	1	1	1
GH8-DOC1	1	1	1	1
GH8-DOC1-CE4	1	1	1	1
GH9-CBM3-CBM3-DOC1	3	3	3	3
GH9-CBM3-DOC1	7	7	7	7
GH9-DOC1	3	3	3	3
PL11-DOC1	2	2	1	2
PL1-DOC1-PL9	1	1	1	1

Table S4.6 Complete list of components in the predicted PULs in the metagenomes

	BD-C	BD-P	MR-C	MR-P
Anti-σ	13	29	9	10
AraC	0	1	0	0
ECF-σ	14	25	10	10
GntR	0	1	1	0
GT2	1	1	0	0
HTCS	2	28	2	7
MFS	2	18	5	10
Pept_CA	0	2	0	0
Pept_MC	0	1	2	2
Pept_MH	0	2	0	0
Pept_na	0	0	2	2
Pept_PB	0	1	1	1
Pept_PC	0	0	1	1
Pept_SC	1	7	0	8
Pept_SE	0	2	0	0
Sulf_1	0	5	1	0
Sulf_4	0	1	0	4
SusC	35	196	37	67
SusD	35	187	35	65
SusR	0	12	3	4
CBM4	0	3	1	0
CBM6	0	9	5	3
CBM13	0	1	1	0
CBM20	0	0	1	1
CBM32	0	9	2	3
CBM35	0	3	1	0
CBM38	0	4	1	2
CBM48	1	9	0	7
CBM58	0	3	0	4
CBM62	0	1	1	0
CBM66	0	3	3	0
CBM67	0	1	2	0
CBM77	0	1	1	0
CE1	2	28	0	24
CE4	0	2	0	1
CE6	0	3	0	3
CE7	1	1	0	0

CE8	0	5	4	0
CE12	0	2	2	0
CE15	0	1	0	3
GH2	2	34	5	3
GH3	5	23	6	8
GH5	0	8	5	0
GH9	0	6	0	3
GH10	0	8	2	0
GH11	0	1	1	0
GH13	1	13	0	14
GH16	1	10	3	4
GH20	0	3	2	0
GH26	1	4	2	0
GH27	0	5	2	1
GH28	0	4	2	0
GH29	0	5	1	0
GH30_3	2	3	2	0
GH31	0	9	6	4
GH32	1	2	1	1
GH33	0	4	0	0
GH35	0	7	1	1
GH36	0	6	0	1
GH38	0	1	2	0
GH42	0	1	0	0
GH43	1	40	7	13
GH50	0	1	0	1
GH51	0	13	1	4
GH53	0	9	0	2
GH55	0	3	0	0
GH63	1	0	0	0
GH66	0	1	1	0
GH67	0	2	1	1
GH76	2	2	2	0
GH77	0	0	1	1
GH78	1	10	2	0
GH81	0	1	1	0
GH87	0	1	1	0
GH88	0	2	1	0
GH92	4	10	6	0
GH93	0	3	1	0
GH95	0	5	1	1

GH97	0	8	1	1
GH98	0	1	0	0
GH99	0	1	0	0
GH105	0	8	1	0
GH106	0	3	0	0
GH108	0	1	0	0
GH109	1	1	0	0
GH115	0	3	2	0
GH116	0	1	0	0
GH123	0	2	0	0
GH125	1	0	1	0
GH127	0	9	0	3
GH128	0	2	2	0
GH130	2	8	2	1
PL1	0	3	4	0
PL6	0	0	2	0
PL8	0	1	0	0
PL9	0	2	0	0
PL10	0	1	0	0
PL12	0	0	1	0
PL17	0	1	1	0

Table S5.1 Start dates, duration, and transfer ratios for the enrichment phases. C: cellulose, CL: cellulose + lignosulphonate, CT: cellulose + tannic acid, P: pretreated poplar.

	Inoculum	Pulp mill anaerobic granules				
	Substrates	C	CL	CT	P	Transfer ratio[1]
Phase 1	Start date[2]	2010-08-24	2010-08-24	2010-08-24	2010-08-24	1:2
	Duration (d)	60	60	60	60	
Phase 2	Start date	2010-10-23	2010-10-23	2010-10-23	2010-10-23	1:2
	Duration (d)	75	75	75	75	
Phase 3	Start date	2011-01-06	2011-01-06	2011-01-06	2011-01-06	1:2
	Duration (d)	34	34	34	34	
Phase 4	Start date	2011-02-09	2011-02-09	2011-02-09	2011-02-09	1:1
	Duration (d)	246	246	246	246	
Phase 5	Start date	2011-10-13	2011-10-13	2011-10-13	2011-10-13	1:1
	Duration (d)	53	53	53	53	
Phase 6	Start date	2011-12-05	2011-12-05	2011-12-05	2011-12-05	1:1
	Duration (d)	241	70	240	320	
Phase 7	Start date	2012-08-02	2012-02-13	2012-08-01	2012-10-20	1:1
	Duration (d)	96	114	253	122	
Phase 8	Start date	2012-11-06	2012-06-06	2013-04-11	2013-02-19	1:1
	Duration (d)	78	67	82	94	

[1] Transfer by dilution was conducted through phases 1-7 for cultures fed with cellulose, cellulose + lignosulphonate, cellulose + tannic acid, and phases 1-6 for cultures fed with pretreated poplar. Following transfers were conducted by centrifugation and supernatant displacement with fresh medium and substrates.

[2] Dates are displayed as YYYY-MM-DD.

Table S5.2 Multiplex barcodes, amplicon DNA concentration, numbers of reads and OTUs for amplicon samples prepared from the pulp mill anaerobic granules and the enrichment microcosms

Sample (enrichment length)	Multiplex barcode	Amplicon DNA concentration (μg/ml)	Number of reads	Number of OTUs
Inoculum	ATAGAGTACT	106.8	3921	963
	CACGCTACGT	71.4	4546	1030
	CAGTAGACGT	107.8	4594	1020
Cellulose (2 yrs)	ATATCGCGAG	67.0	5112	467
	CTCGCGTGTC	77.1	4594	524
	TCTCTATGCG	65.1	4085	482
Cellulose (3 yrs)	ATATCGCGAG	146.7	7379	533
	CTCGCGTGTC	154.8	7038	460
	TCTCTATGCG	205.4	9463	561
Cellulose + Lignosulphonate (2 yrs)	ACGAGTGCGT	100.4	4621	557
	AGCACTGTAG	105.9	4448	524
	ATCAGACACG	83.6	4379	495
Cellulose + Lignosulphonate (3 yrs)	ACGAGTGCGT	137.1	7140	568
	AGCACTGTAG	139.3	6478	479
	ATCAGACACG	118.3	5808	412
Cellulose+Tannic acid (2 yrs)	ACTGTACAGT	70.0	4543	429
	AGACTATACT	38.2	4693	477
Cellulose+Tannic acid (3 yrs)	TAGTGTAGAT	68.8	7742	852
	TCTATACTAT	60.5	7347	602
	TGACGTATGT	76.4	7534	631
Pretreated poplar (2 yrs)	AGCGTCGTCT	48.1	4348	698
	AGTACGCTAT	68.9	4306	708
Pretreated poplar (3 yrs)	ACTGTACAGT	34.6	5320	838
	AGACTATACT	60.1	8562	1130
	AGCGTCGTCT	58.0	6346	940

Table S5.3 Statistics of the whole genome shotgun sequencing

	AG-C	AG-P
Number of quality trimmed reads (% passed)	70,817,105 (99.8%)	78,824,461 (99.7%)
Total Mbp assembled (Mbp)	40.1	76.3
Number of contigs	5,447	16,540
Longest contig (bp)	1,292,613	1,128,798
N50 (bp)	89,739	45,242
Number of open reading frames (ORFs)	40,477	81,343

Table S5.4 Annotation of (A) CBMs, (B) CEs, (C) GHs, (D) GTs, and (E) PLs in the metagenomes

(A)

	AG-C	AG-P
CBM2	1	6
CBM3	31	34
CBM4	9	18
CBM6	14	33
CBM9	10	16
CBM11	0	3
CBM13	4	9
CBM16	2	4
CBM20	3	13
CBM21	0	1
CBM22	4	4
CBM25	6	4
CBM26	0	5
CBM28	0	1
CBM30	2	5
CBM32	10	16
CBM34	0	1
CBM35	10	13
CBM38	1	1
CBM41	0	2
CBM44	1	1
CBM47	3	0
CBM48	13	27
CBM50	111	154
CBM51	2	0
CBM57	0	1
CBM61	0	1
CBM62	0	3
CBM63	1	1
CBM64	0	3
CBM66	2	8
CBM67	2	5
CBM77	0	1

(B)

	AG-C	AG-P
CE1	21	35
CE2	4	5
CE3	4	8
CE4	27	43
CE6	2	7
CE7	2	4
CE8	2	12
CE9	9	8
CE11	6	23
CE12	3	9
CE14	13	10
CE15	2	7

(C)

	AG-C	AG-P
GH1	6	5
GH2	20	50
GH3	25	64
GH4	9	14
GH5	1	8
GH5_1	4	4
GH5_10	1	2
GH5_12	1	3
GH5_13	0	2
GH5_19	0	1
GH5_2	0	19
GH5_22	2	0
GH5_25	1	4
GH5_26	0	1
GH5_36	1	1
GH5_37	1	1
GH5_38	0	2
GH5_4	2	3
GH5_40	0	1
GH5_41	1	0
GH5_45	0	2
GH5_46	1	2
GH5_53	0	1
GH5_7	1	1
GH5_8	1	2
GH8	3	11
GH9	23	39
GH10	9	19
GH11	2	5
GH12	0	1
GH13	8	25
GH13_10	0	1
GH13_11	2	3
GH13_14	1	6
GH13_18	4	1
GH13_20	3	10
GH13_23	2	1
GH13_31	5	2
GH13_36	2	0
GH13_38	3	3

	AG-C	AG-P
GH13_39	0	2
GH13_4	0	1
GH13_6	0	1
GH13_8	2	3
GH13_9	3	6
GH15	1	2
GH16	6	24
GH17	0	1
GH18	20	11
GH19	1	1
GH20	11	16
GH23	27	50
GH24	1	10
GH25	1	9
GH26	6	13
GH27	3	7
GH28	4	19
GH29	7	8
GH30	1	2
GH30_1	2	0
GH30_2	1	0
GH30_3	2	9
GH30_5	0	1
GH30_8	1	3
GH31	11	13
GH32	5	8
GH33	3	2
GH35	2	6
GH36	5	7
GH37	0	1
GH38	6	14
GH39	4	5
GH42	1	2
GH43	0	2
GH43_1	2	2
GH43_10	2	7
GH43_11	1	0
GH43_12	4	1
GH43_15	0	2
GH43_16	1	2

	AG-C	AG-P
GH43_17	1	2
GH43_18	0	1
GH43_19	0	1
GH43_2	0	1
GH43_20	1	1
GH43_24	1	3
GH43_26	3	1
GH43_27	1	0
GH43_28	1	1
GH43_29	2	3
GH43_3	1	2
GH43_31	1	2
GH43_32	0	1
GH43_33	1	1
GH43_34	1	0
GH43_35	3	2
GH43_4	2	0
GH43_8	1	0
GH44	2	3
GH45	0	1
GH46	0	1
GH48	2	2
GH50	0	1
GH51	10	17
GH52	1	2
GH53	2	6
GH55	0	1
GH57	13	31
GH62	1	2
GH63	3	5
GH65	3	3
GH66	0	1
GH67	3	5
GH73	19	9
GH74	15	15
GH76	2	4
GH77	8	26
GH78	6	13
GH79	1	0
GH81	0	1

	AG-C	AG-P
GH84	2	0
GH87	0	1
GH88	3	6
GH89	0	2
GH92	13	15
GH93	0	2
GH94	8	14
GH95	4	5
GH97	2	15
GH99	2	0
GH102	0	2
GH103	1	1
GH105	3	13
GH106	0	2
GH108	1	10
GH109	2	4
GH110	1	7
GH113	1	1
GH115	1	4
GH116	0	3
GH119	0	3
GH120	2	1
GH123	1	5
GH125	1	1
GH126	1	0
GH127	7	2
GH128	0	2
GH130	12	18
GH133	2	14
GH139	0	0
GH140	0	3
GH141	1	2
GH142	0	0

(D)

	AG-C	AG-P
GT1	10	12
GT2	193	401
GT3	2	8
GT4	147	416
GT5	12	24
GT8	3	0
GT9	24	68
GT10	0	1
GT11	1	2
GT19	5	24
GT20	4	5
GT21	0	1
GT22	0	1
GT25	2	1
GT26	6	4
GT28	15	40
GT30	8	20
GT32	5	3
GT35	11	41
GT39	3	4
GT41	0	1
GT50	0	1
GT51	26	55
GT55	0	1
GT56	0	1
GT62	2	1
GT66	5	1
GT74	1	1
GT76	1	2
GT81	1	4
GT83	14	17
GT84	2	3
GT89	1	1

(E)

	AG-C	AG-P
PL1	1	19
PL6	0	8
PL8	0	1
PL9	5	17
PL10	0	1
PL11	1	4
PL12	2	6
PL14	0	1
PL17	0	2
PL22	1	0

Table S5.5 Domain architecture, number of gene count and relative abundances of top 20 most abundant multi-modular CAZymes in the metagenomes

	AG-C		AG-P	
Multi-modular CAZymes	**count**	**%**	**count**	**%**
GH9-CBM3-DOC1	7	6.4%	7	3.7%
CBM48-GH13_9	3	2.8%	6	3.2%
GH23-CBM50-CBM50-CBM50	0	0.0%	7	3.7%
CBM20-CBM20-GH77	1	0.9%	5	2.6%
CBM48-GH13	2	1.8%	4	2.1%
CBM67-GH78	2	1.8%	4	2.1%
CBM50-CBM50-GH18	3	2.8%	3	1.6%
GH23-CBM50-CBM50	3	2.8%	3	1.6%
GH9-CBM3-CBM3-DOC1	3	2.8%	3	1.6%
GH9-DOC1	3	2.8%	3	1.6%
GH23-CBM50	0	0.0%	5	2.6%
CE3-DOC1	2	1.8%	3	1.6%
GH73-CBM50-CBM50	0	0.0%	4	2.1%
GT3-GT35	1	0.9%	3	1.6%
GT84-GH94	1	0.9%	3	1.6%
CBM30-GH9-DOC1	2	1.8%	2	1.1%
CE12-DOC1	2	1.8%	2	1.1%
CE1-DOC1	2	1.8%	2	1.1%
GH48-DOC1	2	1.8%	2	1.1%
GH5_1-DOC1	2	1.8%	2	1.1%
GH9-CBM3	2	1.8%	2	1.1%
GT2-GT4	3	2.8%	1	0.5%

Table S6.1 Lignocellulosic enrichment during the scale up process

Phase	Prior to scale-up	1	2	3	4	5	6	7
Culture volume (mL) per bottle	60	120	1000	1000	1000	1000	1000	1000
Scale up across transfer	-	1:2	1:2.8	1:1	1:1	1:1	1:1	1:1
mg cellulose added	125	250	1125	1500	1500	1200	1500	1500
mg COD cellulose added	152.5	305.0	1372.5	1830.0	1830.0	1464.0	1464.0	1464.0
mg pretreated poplar added	136	204	918	1500	1224	979.2		
mg COD pretreated poplar added	126.5	189.7	853.7	1395.0	1138.3	910.7		

Table S7.1 Description of the protein candidates selected from the metagenomes for biochemical characterization.

Description of protein candidates	Protein ID
Multi-modular xylanases containing GH43 with additional modules related to xylan modification - contains a GH43 module - contains additional modules related to xylan modification	BDPH__42662_79__GH43-CBM32 BDPH__45339_3__GH43-CBM32 BDPH__45575_19__GH43-CBM32-CBM48-CE1 MR-C_16832_17_m_GH43-GH43
GH43 xylanases from the predicted non-xylan related PULs - contains GH43 module - positioned within PULs that do not contain 10-12 components, of which at least two are xylan-active CAZymes (e.g. GH8, GH10, GH11, GH43, GH54, GH115, CE2, CE3, CE6)	AG-PH_36_48426_10_GH43 BD-PH_36_42180_24_GH43 BD-PH_154_46169_20_GH43 MR_C_26_18908_57_CE12-GH43
Multi-modular GH43 xylanases with non-xylan related modules - contains a GH43 module - contains additional module unrelated to xylan modification - contains a signal sequence - does not contain a transmembrane domain	BDPH_42662_81_GH43-GH95 BD-PH_45181_268_GH43-GH16 MRPH__21800_1__GH43_17-GH50
GH115 xylanases - contains a GH115 module - low similarity to characterized GH115 members in CAZyDB	BDC_16082_12__GH115 BDPH__45034_39__GH115 BDPH__46267_3__GH115
Carbohydrate-binding proteins with domain of unknown function - contains xylan-related CBM - contains domains of unknown function with functional implication - preference for candidates exclusively present in pretreated poplar enrichments	MRPH__19969_56__CBM9__DUF362 MRPH__20505_57__CBM3__DUF362 MRPH__21177_69__CBM9__DUF362

Proteins with unknown function within the predicted xylan-related PULs - positioned within PULs that contain 10-12 components, of which at least two are xylan-active CAZymes (e.g. GH8, GH10, GH11, GH43, GH54, GH115, CE2, CE3, CE6) - positioned near other CAZymes - contains signal sequence - representative of other proteins of unknown function using similarity-based CD-hit clustering	BD-PH_PUL21_41672_129 BD-PH_PUL30_41953_3 BD-PH_PUL127_45181_263
Non-CAZyme, carbohydrate-binding proteins with unknown function positioned within the predicted PULs - contains signal sequence - representative of its clade in a phylogenetic tree - contains CBM that targets hemicellulose or potentially new GH or CBM	AG-PH_CBM-UNK B-50065_20 MR-C_CBM-UNK G- 18049_203 MR-C_CBM-UNK H- 18697_2 MR-C_CBM-UNK-I-18697_6

Table S7.2 Taxonomic assignment and quality of the metagenome-assembled genomes.

Sample	bin ID	Marker lineage	Genome size (Mbp)	%Completeness (Parks et al, 2015)	%Contamination (Parks et al, 2015)	%Completeness - %Contamination x5 > 50% (Parks et al, 2017)	Quality (Bowers et al, 2017)
AG-C	15	k__Bacteria	2.53	96.9	0.0	Y (96.9%)	High
AG-P	25	k__Bacteria	2.97	98.9	1.3	Y (92.2%)	High
	15	k__Bacteria	1.71	87.3	1.1	Y (81.9%)	Medium
	42	k__Archaea	3.22	99.1	6.5	Y (66.4%)	Medium
	49	c__Clostridia	0.97	58.1	0.8	Y (54.1%)	Medium
	52	o__Clostridiales	1.28	54.9	0.3	Y (53.2%)	Medium
	47	k__Bacteria	1.98	50.0	0.5	N (47.3%)	Medium
BD-C	44	k__Bacteria	5.23	86.2	1.2	Y (80.5%)	Medium
	28	k__Bacteria	1.23	78.3	0.5	Y (75.9%)	Medium
	31	k__Bacteria	1.84	73.6	3.6	Y (55.4%)	Medium
	45	k__Bacteria	1.31	60.0	1.8	Y (50.9%)	Medium
	39	f__Lachnospiraceae	2.81	60.6	2.0	Y (50.6%)	Medium
BD-P	42	k__Bacteria	2.38	97.3	0.8	Y (93.3%)	High
	47	k__Archaea	2.96	99.1	4.7	Y (75.7%)	High
	19	o__Clostridiales	3.05	79.5	1.2	Y (73.4%)	Medium
	26	o__Clostridiales	4.04	67.1	0.0	Y (67.1%)	Medium
	28	p__Bacteroidetes	2.99	71.0	1.4	Y (63.8%)	Medium
	43	k__Bacteria	2.89	67.2	1.6	Y (59.0%)	Medium
	50	k__Bacteria	0.96	57.7	0.0	Y (57.7%)	Medium
	45	c__Clostridia	2.68	56.8	0.8	Y (52.8%)	Medium
MR-C	44	k__Bacteria	3.34	96.4	6.1	Y (66.1%)	Medium
	40	k__Bacteria	1.31	64.3	2.3	Y (52.8%)	Medium
MR-P	23	o__Clostridiales	4.23	69.1	0.7	Y (65.8%)	Medium
	37	k__Bacteria	1.99	75.5	2.7	Y (61.8%)	Medium
	14	p__Firmicutes	4.02	70.6	2.4	Y (58.6%)	Medium
	29	p__Firmicutes	2.97	64.6	1.5	Y (57.2%)	Medium
	42	k__Bacteria	2.25	61.1	0.9	Y (56.5%)	Medium

Supplementary Figures

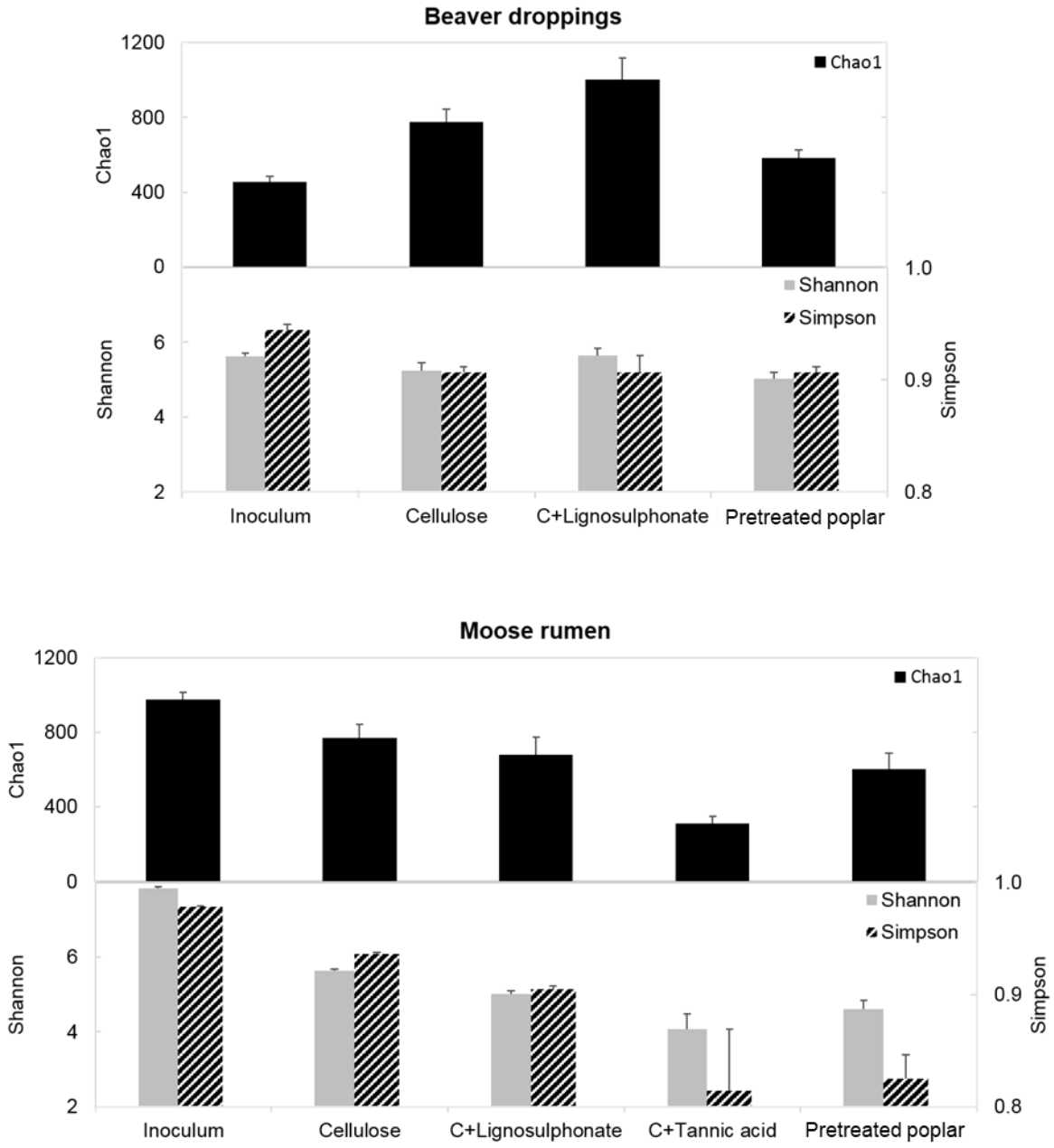


Figure S3.1 Diversity indices of inocula and corresponding enrichments amended with various lignocellulosic substrates. Error bars indicate standard deviation; n=3 except for the moose rumen cultures enriched with cellulose plus tannic acid, where n=2.

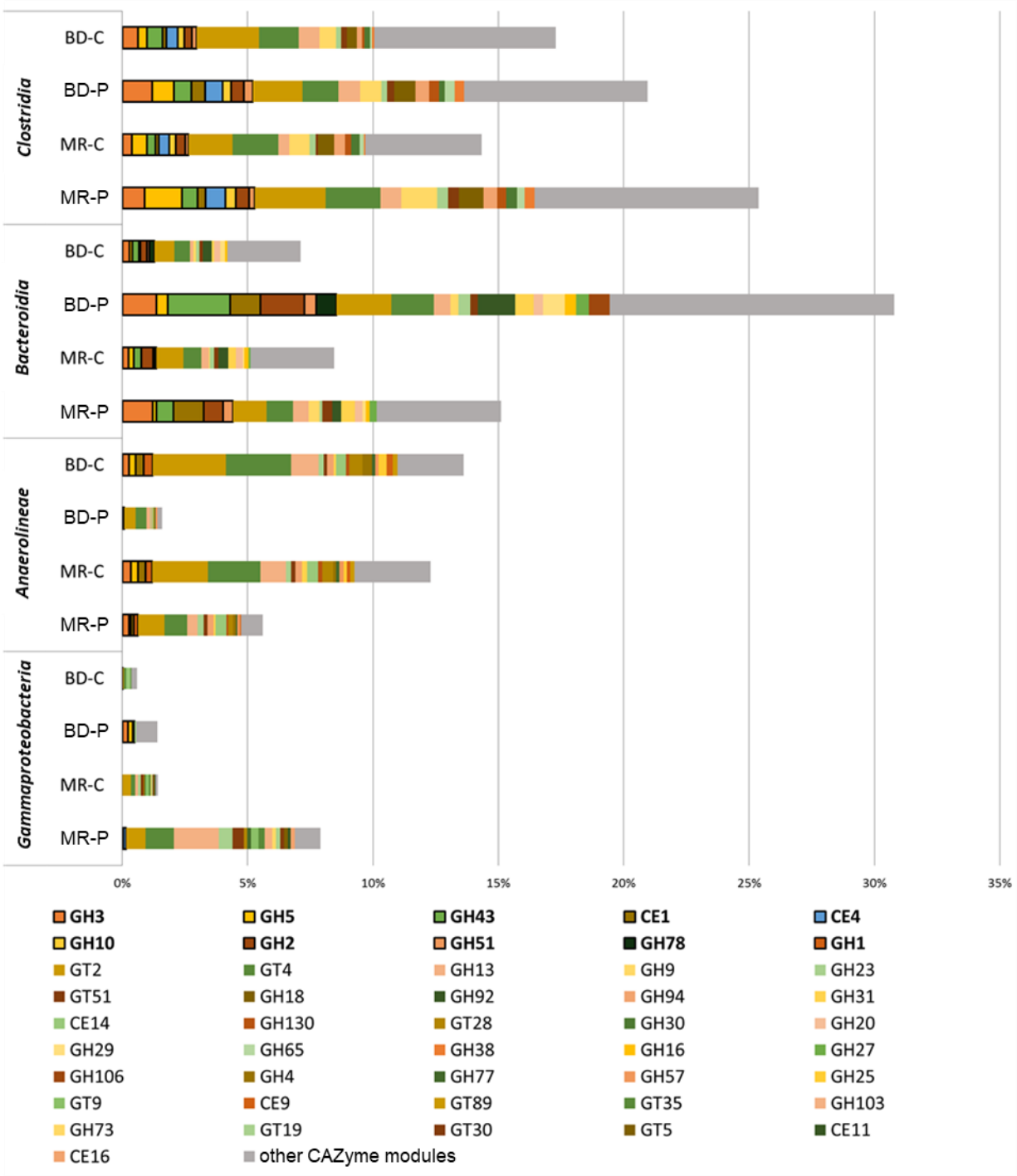


Figure S4.1 Top 20 CAZyme families assigned to most abundant identified classes: *Clostridia*, *Bacteroidia*, *Anaerolineae*, and *Gammaproteobacteria*. CAZyme families predicted to act on plant polysaccharides are outlined in black.

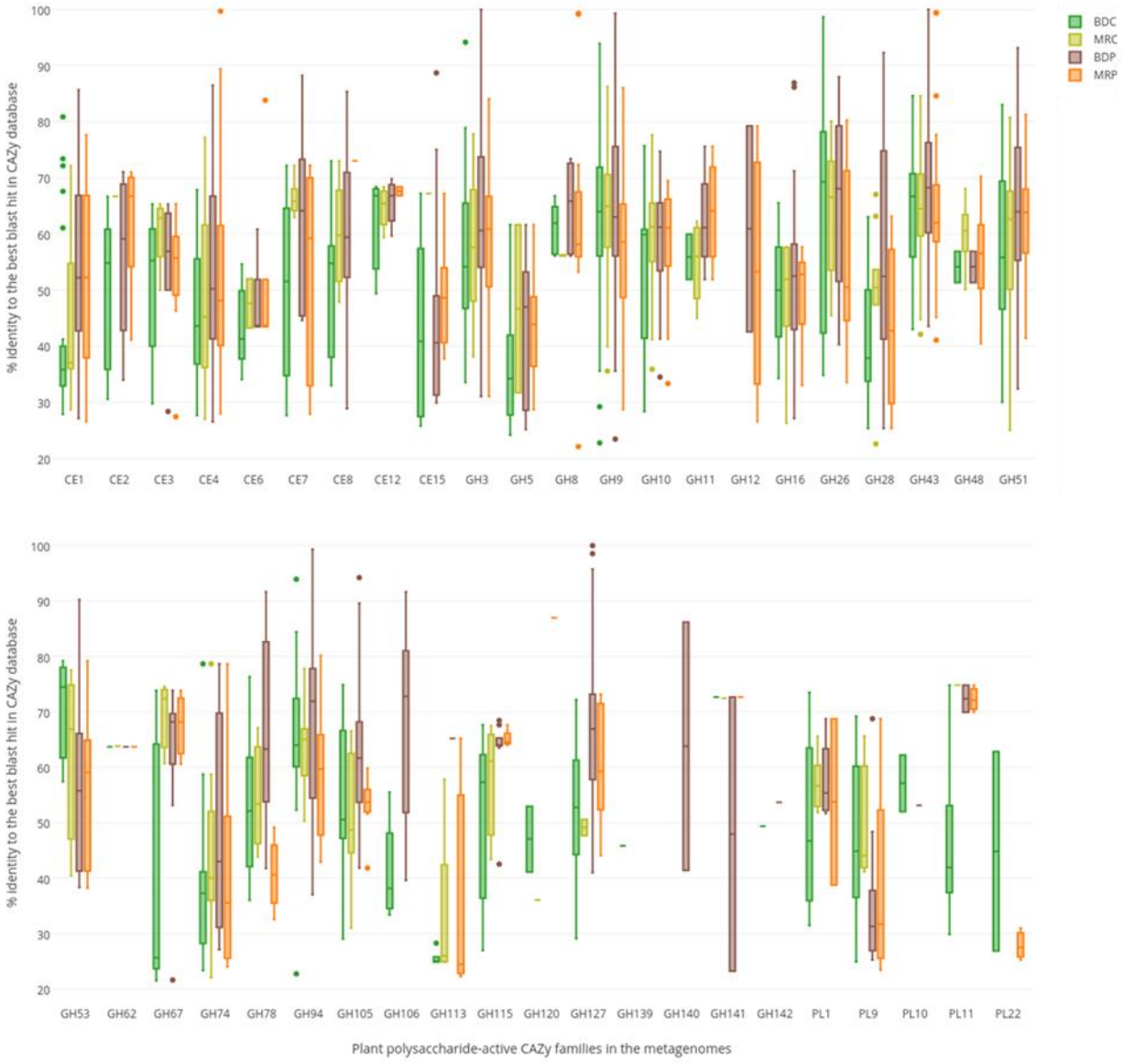


Figure S4.2 Distribution of percentage identities of the plant polysaccharide-active CAZyme families in beaver dropping (BD) and moose rumen (MR) microcosms enriched with cellulose (C) and pretreated poplar (P) with respective best blast hits in the CAZy database

Sample	PUL Nº	Modularity
BD-C	Predicted 1	ECF-σ ▶ SusC ▶ SusD ▶
BD-C	Predicted 2	HTCS ▶ SusC ▶ SusD ▶ GH32 ▶ unk ▶ MFS ▶ unk ▶ CE7 ▶
BD-C	Predicted 3	GH2 ▶ unk ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ ECF-σ ▶
BD-C	Predicted 4	unk ▶ GH130 ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH26 ▶
BD-C	Predicted 5	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ GH76 ▶ GH76 ▶ GH125 ▶
BD-C	Predicted 6	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
BD-C	Predicted 7	SusC ▶ SusD ▶
BD-C	Predicted 8	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶
BD-C	Predicted 9	SusC ▶ SusD ▶
BD-C	Predicted 10	SusC ▶ SusD ▶
BD-C	Predicted 11	SusC ▶ SusD ▶ SusC ▶ SusD ▶ SusC ▶ SusD ▶
BD-C	Predicted 12	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ GH63 ▶ unk ▶
BD-C	Predicted 13	SusC ▶ SusD ▶
BD-C	Predicted 14	SusC ▶ SusD ▶ Pept_SC ▶
BD-C	Predicted 15	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶
BD-C	Predicted 16	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ GH2|GH130 ▶
BD-C	Predicted 17	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶
BD-C	Predicted 18	◀ ECF-σ Anti-σ ▶ SusC ▶ SusD ▶
BD-C	Predicted 19	◀ unk ◀ GH92 SusC ▶ SusD ▶
BD-C	Predicted 20	SusC ▶ SusD ▶
BD-C	Predicted 21	◀ GH78 ◀ HTCS unk ▶ SusC ▶ SusD ▶ GH3 ▶ CE1 ▶ CE1 ▶
BD-C	Predicted 22	SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶
BD-C	Predicted 23	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶
BD-C	Predicted 24	SusC ▶ SusD ▶ GH30_3 ▶ unk ▶ GH30_3 ▶ GH3 ▶
BD-C	Predicted 25	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ GH92 ▶ GH92 ▶ GH92 ▶
BD-C	Predicted 26	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ GH109 ▶
BD-C	Predicted 27	SusC ▶ SusD ▶ unk ▶ CBM48|GH13 ▶

BD-C	Predicted 28	SusC ▶ SusD ▶
BD-C	Predicted 29	SusC ▶ SusD ▶ unk ▶ GH3 ▶ unk ▶ GH43_28 ▶ unk ▶
BD-C	Predicted 30	SusC ▶ SusD ▶
BD-C	Predicted 31	◀ ECF-σ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ GT2 ▶ unk ▶ MFS ▶ unk ▶ unk ▶ unk ▶ unk ▶
BD-C	Predicted 32	GH3 ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ GH16 ▶ GH3 ▶
BD-C	Predicted 33	SusC ▶ SusD ▶
BD-PH	Predicted 1	SusC ▶ SusD ▶
BD-PH	Predicted 2	SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 3	SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 4	SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 5	◀ HTCS SusC ▶ SusD ▶ unk ▶ unk ▶ GH43 ▶ GH127 ▶ unk ▶ unk ▶
BD-PH	Predicted 6	◀ GH93 ◀ GH43 SusC ▶ SusD ▶ GH43 ▶ HTCS ▶ GH2 ▶ GH93 ▶ GH43 ▶
BD-PH	Predicted 7	ECF-σ ▶ CE1 ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH5_36 ▶ GH28 ▶ unk ▶ unk ▶ unk ▶
BD-PH	Predicted 8	GH78 ▶ SusC ▶ SusD ▶ unk ▶ GH2 ▶ unk ▶ unk ▶ unk ▶ MFS ▶ GH36 ▶ GH36 ▶ unk ▶
BD-PH	Predicted 9	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ GH2 ▶ GH105 ▶
BD-PH	Predicted 10	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶
BD-PH	Predicted 11	Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ GH29 ▶ unk ▶ unk ▶ unk ▶ GH3 ▶ unk ▶ unk ▶ unk ▶ unk ▶ GH95 ▶ unk ▶ unk ▶
BD-PH	Predicted 12	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶
BD-PH	Predicted 13	SusC ▶ SusD ▶ SusC ▶ SusD ▶ Pept_SC ▶ Pept_CA ▶ Pept_MH ▶ Pept_MH ▶ unk ▶ CE1 ▶ unk ▶
BD-PH	Predicted 14	SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 15	SusC ▶ SusD ▶ unk ▶ unk ▶
BD-PH	Predicted 16	unk ▶ unk ▶ SusC ▶ SusD ▶
BD-PH	Predicted 17	unk ▶ SusC ▶ SusD ▶ unk ▶

BD-PH	Predicted 18	MFS ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ CBM58|GH13 ▶ unk ▶ GH13 ▶ GH13 ▶ unk ▶ unk ▶ unk ▶ unk ▶
BD-PH	Predicted 19	SusC ▶ SusD ▶ unk ▶ unk ▶
BD-PH	Predicted 20	SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ unk ▶
BD-PH	Predicted 21	SusC ▶ SusD ▶ unk ▶ GH28 ▶ CE12 ▶ GH105 ▶ CE8 ▶ PL1 ▶ PL1 ▶ CE8 ▶ unk ▶ PL1|CBM77 ▶ CE8 ▶ unk ▶
BD-PH	Predicted 22	SusC ▶ SusD ▶ unk ▶ unk ▶ GH26 ▶ GH26 ▶ GH130 ▶ GH27 ▶ GH5_25 ▶ MFS ▶ unk ▶ GH5_2 ▶ unk ▶ GH3 ▶ ◀ unk GH20 ▶ unk ▶ GH20 ▶
BD-PH	Predicted 23	unk ▶ GH76 ▶ SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 24	unk ▶ SusC ▶ SusD ▶
BD-PH	Predicted 25	SusC ▶ SusD ▶
BD-PH	Predicted 26	◀ GH5_36 ◀ GH106 ◀ unk ◀ Pept_CA unk ▶ ◀ GH105 ◀ unk unk ▶ SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 27	SusC ▶ SusD ▶
BD-PH	Predicted 28	SusR ▶ unk ▶ unk ▶ SusC ▶ SusD ▶ GH99 ▶ GH97 ▶
BD-PH	Predicted 29	SusC ▶ SusD ▶
BD-PH	Predicted 30	SusC ▶ SusD ▶ GH43 ▶ CE1|CE6 ▶ GH9 ▶ CE1|CBM48|CBM32|GH43 ▶ CE1|CBM48 ▶ GH51 ▶ CE1 ▶ CE1 ▶ unk ▶ CBM6|GH43 ▶ unk ▶
BD-PH	Predicted 31	SusC ▶ SusD ▶ unk ▶ ◀ GH2 ◀ GH53|GH53 ◀ unk
BD-PH	Predicted 32	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ MFS ▶ unk ▶ GH9|CE4 ▶ unk ▶ GT2 ▶ unk ▶ unk ▶ MFS ▶
BD-PH	Predicted 33	SusC ▶ SusD ▶ SusD ▶ Sulf_1 ▶ Sulf_1 ▶ GH33 ▶
BD-PH	Predicted 34	SusC ▶ SusD ▶ CBM67 ▶
BD-PH	Predicted 35	◀ HTCS SusC ▶ SusD ▶
BD-PH	Predicted 36	SusC ▶ SusD ▶ GH2 ▶ GH43 ▶ unk ▶ GH35 ▶ HTCS ▶
BD-PH	Predicted 37	HTCS ▶ GH2 ▶ SusC ▶ SusD ▶ unk ▶ GH88 ▶ ◀ GH92
BD-PH	Predicted 38	SusC ▶ SusD ▶
BD-PH	Predicted 39	unk ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ GH43 ▶ GH97 ▶ GH92 ▶ unk ▶ ◀ GH43 ◀ GH3
BD-PH	Predicted 40	SusC ▶ SusD ▶

BD-PH	Predicted 41	◀ HTCS GH130 ▶ MFS ▶ unk ▶ SusC ▶ SusD ▶ SusC ▶ SusD ▶
BD-PH	Predicted 42	unk ▶ SusC ▶ SusD ▶ unk ▶ GH53 ▶ GH2 ▶ unk ▶
BD-PH	Predicted 43	SusC ▶ SusD ▶ unk ▶ unk ▶ ◀ unk ◀ GH95 GH43 ▶ unk ▶ GH3 ▶ GH92 ▶ GH95|GH43 ▶ GH3 ▶ CBM32|GH43 ▶ GH92 ▶ GH2 ▶ GH3 ▶ ◀ GH2
BD-PH	Predicted 44	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ GH55 ▶ GH9 ▶ GH55 ▶ GH55 ▶ GH50 ▶ unk ▶ unk ▶ GH32 ▶
BD-PH	Predicted 45	SusC ▶ SusD ▶ unk ▶ GH16 ▶ GH3 ▶
BD-PH	Predicted 46	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ GH105 ▶ unk ▶
BD-PH	Predicted 47	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶
BD-PH	Predicted 48	unk ▶ SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 49	SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ GH130 ▶ Pept_SE ▶ unk ▶ unk ▶
BD-PH	Predicted 50	unk ▶ SusC ▶ SusD ▶
BD-PH	Predicted 51	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH88 ▶ unk ▶ PL17 ▶ PL8 ▶
BD-PH	Predicted 52	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 53	unk ▶ unk ▶ unk ▶ unk ▶ GH36 ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
BD-PH	Predicted 54	SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 55	unk ▶ SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 56	SusC ▶ SusD ▶ unk ▶ unk ▶ GH26|CBM35 ▶ GH26 ▶ GH130 ▶ MFS ▶ unk ▶ unk ▶
BD-PH	Predicted 57	unk ▶ SusC ▶ SusC ▶ SusD ▶ SusD ▶
BD-PH	Predicted 58	SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 59	SusC ▶ SusD ▶
BD-PH	Predicted 60	◀ ECF-σ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 61	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ MFS ▶ GH51 ▶ GH51 ▶ Sulf_1 ▶ Pept_SC ▶
BD-PH	Predicted 62	SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 63	SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶
BD-PH	Predicted 64	SusC ▶ SusD ▶

BD-PH	Predicted 65	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ CBM32|GH2 ▶ ◀ unk ◀ GH3 |Pept_SE
BD-PH	Predicted 66	SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 67	SusC ▶ SusD ▶
BD-PH	Predicted 68	SusC ▶ SusD ▶
BD-PH	Predicted 69	SusC ▶ SusD ▶ GH29 ▶ SusC ▶ SusD ▶
BD-PH	Predicted 70	SusC ▶ SusD ▶ unk ▶ GH123 ▶ GH109 ▶ GH36 ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ Sulf_1 ▶ unk ▶ ◀ unk GH36 ▶ unk ▶ unk ▶ unk ▶
BD-PH	Predicted 71	Anti-σ ▶ unk ▶ SusC ▶ SusC ▶ SusD ▶
BD-PH	Predicted 72	HTCS ▶ GH31 ▶ GH2 ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH5_4 ▶
BD-PH	Predicted 73	unk ▶ SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 74	SusC ▶ SusD ▶ unk ▶ unk ▶ GH116|CBM38 ▶ unk ▶ unk ▶
BD-PH	Predicted 75	◀ unk ◀ unk ◀ GH105 unk ▶ unk ▶ SusC ▶ SusD ▶ GH95 ▶ SusC ▶ SusC ▶ SusD ▶ unk ▶ CE8|PL10 ▶ GH2 ▶ CE7 ▶ GH28 ▶ CE12 ▶ GH43 ▶ CE8 ▶ HTCS ▶
BD-PH	Predicted 76	GH29 ▶ ◀ GH2 ◀ unk SusC ▶ SusD ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH16 ▶ HTCS ▶ GH16 ▶ GH123 ▶ Pept_SC ▶ GH2 ▶ GH42 ▶
BD-PH	Predicted 77	unk ▶ GH92 ▶ SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 78	SusC ▶ SusD ▶ unk ▶ unk ▶
BD-PH	Predicted 79	SusC ▶ SusD ▶
BD-PH	Predicted 80	SusC ▶ SusD ▶ GH2 ▶ GH43 ▶ unk ▶ GH35 ▶ HTCS ▶
BD-PH	Predicted 81	◀ HTCS SusC ▶ SusD ▶
BD-PH	Predicted 82	◀ ECF-σ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
BD-PH	Predicted 83	SusR ▶ SusC ▶ SusD ▶
BD-PH	Predicted 84	SusC ▶ SusD ▶
BD-PH	Predicted 85	◀ GH2 ◀ GH28 ◀ unk ◀ SusD ◀ SusC HTCS ▶ ◀ unk SusC ▶ SusD ▶ unk ▶ unk ▶
BD-PH	Predicted 86	SusC ▶ SusD ▶ unk ▶ unk ▶
BD-PH	Predicted 87	unk ▶ unk ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ GH2 ▶ unk ▶
BD-PH	Predicted 88	ECF-σ ▶ ◀ unk Anti-σ ▶ SusC ▶ SusD ▶ GH92 ▶ CE15 ▶ GH92 ▶

BD-PH	Predicted 89	SusC ▶ SusD ▶ unk ▶ GH13 ▶ GH13 ▶
BD-PH	Predicted 90	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ GH30_3 ▶ unk ▶
BD-PH	Predicted 91	ECF-σ ▶ ◀ unk Anti-σ ▶ SusC ▶ SusD ▶ GH127 ▶ GH127 ▶ unk ▶ unk ▶
BD-PH	Predicted 92	CBM38 ▶ MFS ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ CE1 ▶ GH13 ▶ SusC|CE1| CBM48 ▶
BD-PH	Predicted 93	SusC ▶ SusD ▶
BD-PH	Predicted 94	◀ unk ◀ GH130 SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶
BD-PH	Predicted 95	SusC ▶ SusD ▶ Pept_MC ▶
BD-PH	Predicted 96	CBM6|GH128 ▶ unk ▶ CBM6|GH81|CBM38 ▶ GH3|CBM32 ▶ CBM6| GH5_46 ▶ unk ▶ SusC ▶ SusD ▶ GH16 ▶ CBM6|GH128 ▶ GH2 ▶ GH16 ▶
BD-PH	Predicted 97	HTCS ▶ ◀ unk GH95 ▶ GH98|CBM35 ▶ SusC ▶ SusD ▶ SusC ▶ SusD ▶ unk ▶ GH10|CBM4|GH10 ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ GH115| GH10 ▶ GH43 ▶ unk ▶ GH29 ▶ GH43 ▶ unk ▶ unk ▶ GH108 ▶
BD-PH	Predicted 98	SusC ▶ SusD ▶ CBM48 ▶
BD-PH	Predicted 99	unk ▶ SusC ▶ SusC ▶ SusD ▶ SusD ▶
BD-PH	Predicted 100	SusC ▶ SusD ▶
BD-PH	Predicted 101	SusC ▶ SusD ▶ unk ▶ unk ▶
BD-PH	Predicted 102	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ GH5_4 ▶ GH31 ▶ GH31 ▶ unk ▶ GH29 ▶ ◀ GH2 SusC ▶ SusD ▶ unk ▶ GH10|CBM4 ▶ GH35 ▶ GH115 ▶ GH67 ▶ unk ▶ MFS ▶ GH10 ▶ GH11 ▶ GH43 ▶ GH3 ▶ unk ▶
BD-PH	Predicted 103	GH3 ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
BD-PH	Predicted 104	SusR ▶ ◀ GH87 unk ▶ SusC ▶ SusD ▶ GH97 ▶
BD-PH	Predicted 105	SusC ▶ SusD ▶ unk ▶ GH16 ▶ GH3 ▶
BD-PH	Predicted 106	SusC ▶ SusD ▶
BD-PH	Predicted 107	SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 108	GH3 ▶ SusC ▶ SusD ▶ unk ▶ GH30_3 ▶ GH30_3 ▶
BD-PH	Predicted 109	SusC ▶ SusD ▶ unk ▶ GH10|CBM4 ▶ GH35 ▶ unk ▶ MFS ▶ GH10 ▶ GH43 ▶ GH67 ▶ GH43 ▶
BD-PH	Predicted 110	SusC ▶ SusD ▶ unk ▶

BD-PH	Predicted 111	SusC ▶ SusD ▶ GH106 ▶
BD-PH	Predicted 112	◀ SusC SusC ▶ SusD ▶
BD-PH	Predicted 113	SusC ▶ SusD ▶
BD-PH	Predicted 114	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶
BD-PH	Predicted 115	SusC ▶ SusD ▶ GH105 ▶
BD-PH	Predicted 116	SusR ▶ SusC ▶ SusD ▶ unk ▶ GH16 ▶ GH3 ▶ unk ▶
BD-PH	Predicted 117	unk ▶ SusC ▶ SusD ▶ unk ▶ GH2 ▶ GH2 ▶ GH43 ▶ unk ▶ unk ▶ GH115 ▶ CBM66|GH32 ▶ unk ▶ CBM66|GH51 ▶ GH31 ▶ GH43 ▶ CBM62 ▶ GH78 ▶ CBM32 ▶ GH92 ▶ unk ▶ CBM13|GH76|CBM6 ▶ CBM66 ▶
BD-PH	Predicted 118	GH127 ▶ SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 119	SusC ▶ SusD ▶ Pept_SC ▶
BD-PH	Predicted 120	SusC ▶ SusD ▶ ◀ unk ◀ unk GH106 ▶
BD-PH	Predicted 121	unk ▶ SusC ▶ SusD ▶
BD-PH	Predicted 122	SusC ▶ SusD ▶ unk ▶ unk ▶ GH2 ▶ unk ▶ unk ▶ unk ▶ GH92 ▶ unk ▶ unk ▶ unk ▶ MFS ▶ GH92 ▶
BD-PH	Predicted 123	AraC ▶ GH78 ▶ GH78 ▶ Sulf_4|GH20 ▶ unk ▶ SusC ▶ SusD ▶ SusC ▶ SusD ▶
BD-PH	Predicted 124	SusC ▶ SusD ▶ unk ▶ CBM6 ▶ unk ▶ unk ▶ GH127 ▶ GH78 ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
BD-PH	Predicted 125	SusC ▶ SusD ▶ GH33 ▶ unk ▶ GH33 ▶ GH33 ▶ MFS ▶
BD-PH	Predicted 126	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ GH105 ▶ ◀ GH78 ◀ GH78 ◀ GH78 ◀ GH78 GntR ▶
BD-PH	Predicted 127	GH105 ▶ unk ▶ ◀ GH2 ◀ HTCS SusC ▶ SusD ▶ GH43|GH16 ▶ ◀ unk ◀ unk GH2 ▶ HTCS ▶ ◀ GH43 ◀ SusD ◀ SusC ◀ unk ◀ unk GH43 ▶ ◀ unk GH97 ▶ GH51 ▶ unk ▶ unk ▶ unk ▶ unk ▶ GH51 ▶ GH43 ▶ GH43 ▶ GH35|CBM32 ▶ GH51 ▶
BD-PH	Predicted 128	ECF-σ ▶ Anti-σ ▶ SusC ▶ unk ▶ unk ▶ SusC ▶ unk ▶ SusC ▶ SusD ▶
BD-PH	Predicted 129	SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 130	SusC ▶ unk ▶ SusD ▶ unk ▶
BD-PH	Predicted 131	SusR ▶ SusC ▶ SusD ▶ unk ▶ GH27 ▶ GH97 ▶

BD-PH	Predicted 132	◀ GH3 ◀ unk CE1 ▶ CE1 ▶ ◀ HTCS SusC ▶ SusD ▶ GH43 ▶ CE6|CE1 ▶ GH9 ▶ GH43|CBM32|CBM48|CE1 ▶ CBM48|CE1 ▶ GH51 ▶ CE1 ▶ CE1 ▶ unk ▶ GH43|CBM6 ▶ unk ▶
BD-PH	Predicted 133	SusC ▶ SusD ▶ unk ▶ unk ▶ GH9 ▶ MFS ▶ GH10 ▶ GH3 ▶ ◀ CE1 ◀ CE1 ◀ GH3 ◀ GH51
BD-PH	Predicted 134	SusC ▶ SusD ▶
BD-PH	Predicted 135	MFS ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH13|CBM58 ▶ unk ▶ GH13 ▶ GH13 ▶ unk ▶ unk ▶ unk ▶ unk ▶
BD-PH	Predicted 136	◀ HTCS GH2 ▶ unk ▶ SusC ▶ SusD ▶ PL9 ▶
BD-PH	Predicted 137	unk ▶ SusC ▶ SusD ▶ CE1 ▶ GH3 ▶ GH51 ▶ MFS ▶ unk ▶ unk ▶
BD-PH	Predicted 138	SusC ▶ SusD ▶ unk ▶ ◀ GH2 ◀ GH53|GH53 ◀ unk
BD-PH	Predicted 139	SusC ▶ SusD ▶ unk ▶ ◀ GH2 ◀ GH53|GH53 ◀ GH2 ◀ HTCS GH3 ▶ unk ▶ ◀ unk GH51 ▶
BD-PH	Predicted 140	◀ GH3 ◀ unk CE1 ▶ CE1 ▶ ◀ HTCS SusC ▶ SusD ▶ GH43 ▶ CE1|CE6 ▶ GH9 ▶ CE1|CBM48|CBM32|GH43 ▶ CE1|CBM48 ▶ GH51 ▶ CE1 ▶ CE1 ▶ unk ▶ CBM6|GH43 ▶ unk ▶
BD-PH	Predicted 141	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ Pept_SC ▶ GH31 ▶ unk ▶
BD-PH	Predicted 142	SusC ▶ SusD ▶ CBM32|GH43 ▶ Anti-σ ▶ SusC ▶
BD-PH	Predicted 143	SusC ▶ SusD ▶ unk ▶ ◀ GH2 ◀ GH53|GH53 ◀ GH2 ◀ HTCS GH3 ▶ unk ▶ ◀ unk GH51 ▶
BD-PH	Predicted 144	HTCS ▶ unk ▶ unk ▶ GH127 ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
BD-PH	Predicted 145	SusR ▶ SusC ▶ SusD ▶ unk ▶ GH16 ▶ GH3 ▶ unk ▶
BD-PH	Predicted 146	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ Pept_SC ▶ GH31 ▶ unk ▶
BD-PH	Predicted 147	SusR ▶ SusC ▶ SusD ▶ unk ▶ GH66 ▶ GH31 ▶ CBM35 ▶
BD-PH	Predicted 148	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ Pept_SC ▶ GH31 ▶ unk ▶
BD-PH	Predicted 149	GH127 ▶ SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 150	◀ HTCS GH2 ▶ unk ▶ SusC ▶ SusD ▶ PL9 ▶
BD-PH	Predicted 151	GH127 ▶ SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 152	GH127 ▶ SusC ▶ SusD ▶ unk ▶
BD-PH	Predicted 153	GH43 ▶ ◀ unk ◀ GH130 SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶

BD-PH	Predicted 154	GH43 ▶ ◀ unk ◀ GH130 SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶
BD-PH	Predicted 155	MFS ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ CBM58|GH13 ▶ unk ▶ GH13 ▶ GH13 ▶ unk ▶ unk ▶ unk ▶ unk ▶
BD-PH	Predicted 156	ECF-σ ▶ Anti-σ ▶ unk ▶ SusC ▶ SusC ▶ unk ▶ SusD ▶
BD-PH	Predicted 157	◀ unk ◀ GH130 SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶
BD-PH	Predicted 158	unk ▶ unk ▶ unk ▶ GH36 ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ ◀ CE4
BD-PH	Predicted 159	SusC ▶ SusD ▶ GH2 ▶ GH43 ▶ unk ▶ GH35 ▶ HTCS ▶
BD-PH	Predicted 160	◀ HTCS SusC ▶ SusD ▶ Pept_PB ▶ ◀ HTCS CBM38 ▶ MFS ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ CE1 ▶ GH13 ▶ SusC|CBM48|CE1 ▶
BD-PH	Predicted 161	◀ HTCS SusC ▶ SusD ▶
BD-PH	Predicted 162	SusC ▶ SusD ▶ GH2 ▶ GH43 ▶ unk ▶ GH35 ▶ HTCS ▶
BD-PH	Predicted 163	SusR ▶ SusC ▶ SusD ▶ unk ▶ GH27 ▶ GH97 ▶
BD-PH	Predicted 164	SusR ▶ SusC ▶ SusD ▶
BD-PH	Predicted 165	SusR ▶ SusC ▶ SusD ▶ unk ▶ GH16 ▶ GH3 ▶ unk ▶
BD-PH	Predicted 166	SusC ▶ SusD ▶ GH43 ▶ Sulf_1 ▶ GH31 ▶ HTCS ▶ GH2 ▶ GH93 ▶ GH43 ▶ GH5_13 ▶ GH38 ▶
BD-PH	Predicted 167	SusR ▶ SusC ▶ SusD ▶ unk ▶ GH27 ▶ GH97 ▶
BD-PH	Predicted 168	SusR ▶ SusC ▶ SusD ▶ unk ▶ GH27 ▶ GH97 ▶
BD-PH	Predicted 169	SusC ▶ SusD ▶ GH78 ▶
BD-PH	Predicted 170	SusC ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ ◀ unk SusC ▶ SusD ▶
BD-PH	Predicted 171	SusC ▶ SusD ▶
MR-C	Predicted 1	unk ▶ MFS ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ Pept_PB ▶ ◀ CBM20 |GH77
MR-C	Predicted 2	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
MR-C	Predicted 3	SusC ▶ SusD ▶ unk ▶
MR-C	Predicted 4	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ Pept_MC ▶
MR-C	Predicted 5	SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ unk ▶

MR-C	Predicted 6	SusC ▶ SusD ▶ unk ▶ GH28 ▶ CE12 ▶ GH105 ▶ CE8 ▶ PL1 ▶ PL1 ▶ CE8 ▶ unk ▶ PL1|CBM77 ▶ CE8 ▶ unk ▶
MR-C	Predicted 7	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ Pept_PC ▶ unk ▶ Pept_MC ▶ unk ▶ unk ▶ Pept_na ▶ Pept_na ▶
MR-C	Predicted 8	unk ▶ SusC ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
MR-C	Predicted 9	SusC ▶ SusD ▶ GH43_3 ▶ GH31 ▶ unk ▶ unk ▶ unk ▶ GH43_31 ▶ Sulf_1 ▶ GH31 ▶ HTCS ▶ GH2 ▶ GH93 ▶ GH43_3 ▶ GH5_13 ▶ GH38 ▶
MR-C	Predicted 10	SusR ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH27 ▶ GH66 ▶ GH31 ▶ CBM35 ▶
MR-C	Predicted 11	GH3 ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
MR-C	Predicted 12	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ GH92 ▶ MFS ▶ unk ▶ unk ▶
MR-C	Predicted 13	SusR ▶ ECF-σ ▶ unk ▶ GH38 ▶ GH92 ▶ GH78|CBM67 ▶ GH92 ▶ GH92 ▶ GH130 ▶ SusD ▶ unk ▶ unk ▶ SusC ▶ SusD ▶ SusC ▶
MR-C	Predicted 14	SusC ▶ SusD ▶
MR-C	Predicted 15	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶
MR-C	Predicted 16	SusC ▶ SusD ▶
MR-C	Predicted 17	SusC ▶ SusD ▶
MR-C	Predicted 18	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusC ▶ SusD ▶ unk ▶
MR-C	Predicted 19	SusC ▶ SusD ▶ unk ▶ unk ▶ GH26 ▶ GH26 ▶ GH130 ▶ GH27 ▶ GH5_25 ▶ MFS ▶ unk ▶ GH5_2 ▶ unk ▶ GH3 ▶
MR-C	Predicted 20	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ GH20 ▶ unk ▶ GH20 ▶
MR-C	Predicted 21	ECF-σ ▶ SusC ▶ SusD ▶ unk ▶ Anti-σ ▶
MR-C	Predicted 22	unk ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ GH2 ▶ GH2 ▶ GH43_3 ▶ unk ▶ unk ▶ GH115 ▶ GH32|CBM66 ▶ unk ▶ GH51|CBM66 ▶ GH31 ▶ GH43_31 ▶ CBM62 ▶ GH78|CBM67 ▶ CBM32 ▶ GH92 ▶ unk ▶ CBM6|GH76|CBM13 ▶ CBM66 ▶
MR-C	Predicted 23	SusC ▶ SusD ▶ unk ▶
MR-C	Predicted 24	GH128|CBM6 ▶ unk ▶ CBM38|GH81|CBM6 ▶ CBM32|GH3 ▶ GH5_46|CBM6 ▶ unk ▶ SusC ▶ SusD ▶ GH16 ▶ GH128|CBM6 ▶ unk ▶ GH2 ▶ GH16 ▶
MR-C	Predicted 25	SusC ▶ SusD ▶

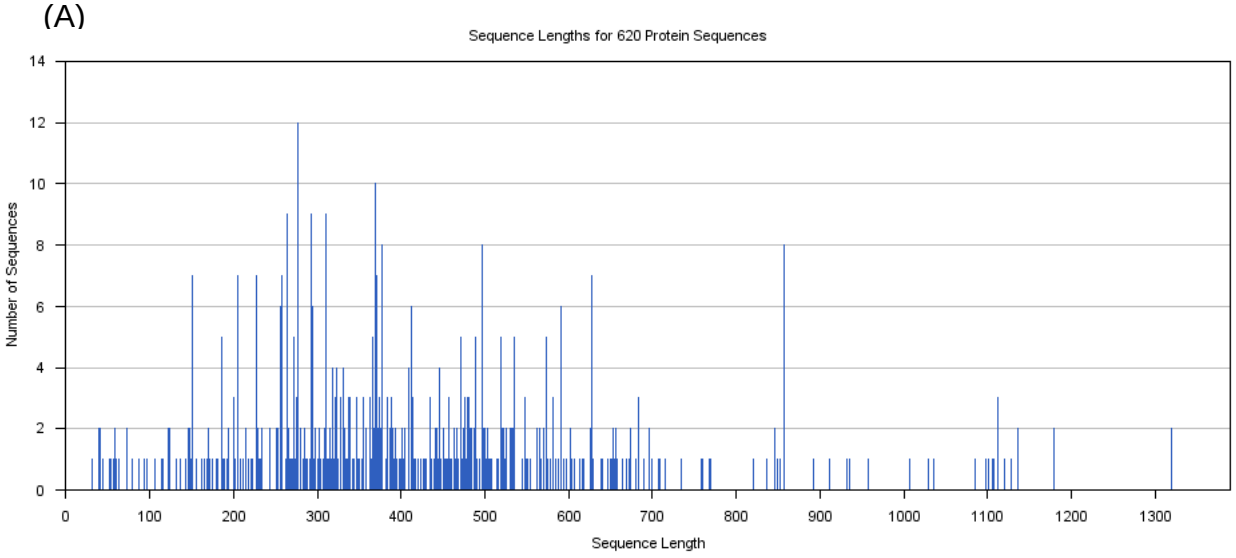
MR-C	Predicted 26	◀ HTCS PL1 ▶ SusC ▶ SusD ▶ unk ▶ GH28 ▶ unk ▶ CE8 ▶ GH95 ▶ CE12|GH43_10 ▶
MR-C	Predicted 27	ECF-σ ▶ Anti-σ ▶ SusC ▶ unk ▶ SusD ▶ unk ▶ GH5_4 ▶ GH31 ▶ GH31 ▶ GH29 ▶ ◀ GH2 SusC ▶ SusD ▶ unk ▶ GH10|CBM4 ▶ SusC ▶ SusD ▶ GH35 ▶ GH115 ▶ GH67 ▶ unk ▶ MFS ▶ GH10 ▶ GH11 ▶ GH43_1 ▶ GH3 ▶ unk ▶
MR-C	Predicted 28	GH3 ▶ SusC ▶ SusD ▶ unk ▶ GH30_3 ▶ GH30_3 ▶
MR-C	Predicted 29	GntR ▶ SusC ▶ SusD ▶ unk ▶ PL6 ▶ GH88 ▶ unk ▶ unk ▶ PL12 ▶ unk ▶ unk ▶ unk ▶ PL6 ▶ PL17 ▶ unk ▶ MFS ▶ unk ▶
MR-C	Predicted 30	SusR ▶ ◀ GH87 unk ▶ SusC ▶ SusD ▶ GH97 ▶
MR-C	Predicted 31	SusC ▶ SusD ▶ unk ▶ GH16 ▶ GH3 ▶
MR-C	Predicted 32	unk ▶ GH76 ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ GH92 ▶ GH125 ▶
MR-PH	Predicted 1	SusC ▶ SusD ▶ unk ▶
MR-PH	Predicted 2	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ Pept_MC ▶
MR-PH	Predicted 3	SusC ▶ SusD ▶ unk ▶
MR-PH	Predicted 4	◀ GH3 ◀ unk CE1 ▶ CE1 ▶ ◀ HTCS SusC ▶ SusD ▶ GH43_17 ▶ CE1|CE6 ▶ GH9 ▶ CE1|CBM48|CBM32|GH43_28 ▶ CE1|CBM48 ▶ GH51 ▶ CE1 ▶ CE1 ▶ unk ▶ CBM6|GH43_2 ▶ unk ▶ CE15 ▶
MR-PH	Predicted 5	SusR ▶ SusC ▶ SusD ▶ unk ▶ GH27 ▶ GH97 ▶
MR-PH	Predicted 6	unk ▶ unk ▶ unk ▶ GH36 ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ ◀ CE4
MR-PH	Predicted 7	◀ GH95 ◀ unk ◀ GH67|GH43_10 ◀ HTCS MFS ▶ GH43_1 ▶ ◀ unk ◀ GH130 SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶ unk ▶
MR-PH	Predicted 8	◀ GH3 ◀ unk CE1 ▶ CE1 ▶ ◀ HTCS SusC ▶ SusD ▶ GH43_17 ▶ CE6|CE1 ▶ GH9 ▶ GH43_28|CBM32|CBM48|CE1 ▶ CBM48|CE1 ▶ GH51 ▶ CE1 ▶ CE1 ▶ unk ▶ GH43_2|CBM6 ▶ unk ▶ CE15 ▶
MR-PH	Predicted 9	SusC ▶ SusD ▶ GH2 ▶ GH43_24 ▶ unk ▶ GH35 ▶ HTCS ▶
MR-PH	Predicted 10	◀ HTCS GH32|CBM38 ▶ MFS ▶ unk ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ GH13_18 ▶ MFS ▶ unk ▶ unk ▶
MR-PH	Predicted 11	ECF-σ ▶ SusC ▶ SusD ▶ unk ▶ Anti-σ ▶
MR-PH	Predicted 12	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ Pept_PC ▶ unk ▶ Pept_MC ▶ unk ▶ unk ▶ Pept_na ▶ Pept_na ▶

MR-PH	Predicted 13	SusC ▶ SusD ▶ unk ▶ ◀ GH2 ◀ GH53|GH53 ◀ GH2
MR-PH	Predicted 14	SusC ▶ SusD ▶
MR-PH	Predicted 15	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶
MR-PH	Predicted 16	unk ▶ SusC ▶ SusD ▶
MR-PH	Predicted 17	unk ▶ MFS ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ Pept_PB ▶ ◀ GH77 |CBM20
MR-PH	Predicted 18	SusC ▶ SusD ▶
MR-PH	Predicted 19	SusC ▶ SusD ▶
MR-PH	Predicted 20	SusC ▶ SusD ▶
MR-PH	Predicted 21	SusC ▶ SusD ▶
MR-PH	Predicted 22	SusC ▶ SusD ▶ unk ▶
MR-PH	Predicted 23	unk ▶ SusC ▶ SusD ▶ CE1 ▶ GH3 ▶ GH51 ▶ MFS ▶ unk ▶ unk ▶
MR-PH	Predicted 24	◀ GH3 ◀ unk CE1 ▶ CE1 ▶ ◀ HTCS SusC ▶ SusD ▶ GH43_17 ▶ CE1| CE6 ▶ GH9 ▶ CE1|CBM48|CBM32|GH43_28 ▶ CE1|CBM48 ▶ GH51 ▶ CE1 ▶ CE1 ▶ unk ▶ CBM6|GH43_2 ▶ unk ▶ CE15 ▶
MR-PH	Predicted 25	SusC ▶ SusD ▶ unk ▶
MR-PH	Predicted 26	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶
MR-PH	Predicted 27	SusC ▶ SusD ▶
MR-PH	Predicted 28	SusC ▶ SusD ▶ unk ▶
MR-PH	Predicted 29	SusC ▶ SusD ▶ unk ▶ unk ▶ GH50|GH43_17 ▶
MR-PH	Predicted 30	GH127 ▶ SusC ▶ SusD ▶ unk ▶
MR-PH	Predicted 31	SusC ▶ SusD ▶
MR-PH	Predicted 32	SusC ▶ SusD ▶
MR-PH	Predicted 33	SusC ▶ SusD ▶
MR-PH	Predicted 34	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ Pept_SC ▶ GH31 ▶ unk ▶
MR-PH	Predicted 35	SusC ▶ SusD ▶ unk ▶
MR-PH	Predicted 36	SusC ▶ SusD ▶ unk ▶
MR-PH	Predicted 37	SusC ▶ SusD ▶ unk ▶
MR-PH	Predicted 38	SusC ▶ SusD ▶ unk ▶

MR-PH	Predicted 39	MFS ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ CBM58|GH13 ▶ unk ▶ GH13 ▶ GH13 ▶ unk ▶ unk ▶ unk ▶ unk ▶
MR-PH	Predicted 40	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ Pept_SC ▶ GH31 ▶ unk ▶
MR-PH	Predicted 41	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ Pept_SC ▶ GH31 ▶ unk ▶
MR-PH	Predicted 42	Pept_SC ▶ unk ▶ SusC ▶ SusD ▶
MR-PH	Predicted 43	SusC ▶ SusD ▶ unk ▶ GH16 ▶ GH3 ▶ unk ▶ unk ▶ ◀ Sulf_4
MR-PH	Predicted 44	SusC ▶ SusD ▶ unk ▶ GH16 ▶ GH3 ▶ unk ▶ unk ▶ ◀ Sulf_4
MR-PH	Predicted 45	Pept_SC ▶ unk ▶ SusC ▶ SusD ▶
MR-PH	Predicted 46	SusR ▶ SusC ▶ SusD ▶ unk ▶ GH16 ▶ GH3 ▶ unk ▶ unk ▶ ◀ Sulf_4
MR-PH	Predicted 47	Pept_SC ▶ unk ▶ SusC ▶ SusD ▶
MR-PH	Predicted 48	unk ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
MR-PH	Predicted 49	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ Pept_SC ▶ GH31 ▶ unk ▶
MR-PH	Predicted 50	◀ HTCS CBM38 ▶ MFS ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ CE1 ▶ GH13 ▶ SusC|CBM48|CE1 ▶
MR-PH	Predicted 51	◀ SusC SusC ▶ SusD ▶ unk ▶ unk ▶
MR-PH	Predicted 52	SusR ▶ SusC ▶ SusD ▶ unk ▶
MR-PH	Predicted 53	SusR ▶ SusC ▶ SusD ▶ unk ▶
MR-PH	Predicted 54	unk ▶ GH127 ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
MR-PH	Predicted 55	unk ▶ unk ▶ GH127 ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
MR-PH	Predicted 56	MFS ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ CBM58|GH13 ▶ unk ▶ GH13 ▶ GH13 ▶ unk ▶ unk ▶ unk ▶ unk ▶
MR-PH	Predicted 57	MFS ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ CBM58|GH13 ▶ unk ▶ GH13 ▶ GH13 ▶ unk ▶ unk ▶ unk ▶ unk ▶
MR-PH	Predicted 58	MFS ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH13|CBM58 ▶ unk ▶ GH13 ▶ GH13 ▶ unk ▶ unk ▶ unk ▶ unk ▶
MR-PH	Predicted 59	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
MR-PH	Predicted 60	SusC ▶ SusD ▶ unk ▶ unk ▶ SusC ▶ SusD ▶
MR-PH	Predicted 61	SusC ▶ SusD ▶
MR-PH	Predicted 62	SusC ▶ SusD ▶
MR-PH	Predicted 63	Pept_SC ▶ unk ▶ SusC ▶ SusD ▶

Figure S4.3 Predicted PULs from the metagenomes from the cellulose (C)- and pretreated poplar (PH)-fed microcosms from beaver droppings (BD) and moose rumen (MR)

(A)


Sequence Lengths for 620 Protein Sequences
Number of Sequences
Sequence Length
0
2
4
6
8
10
12
14
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300

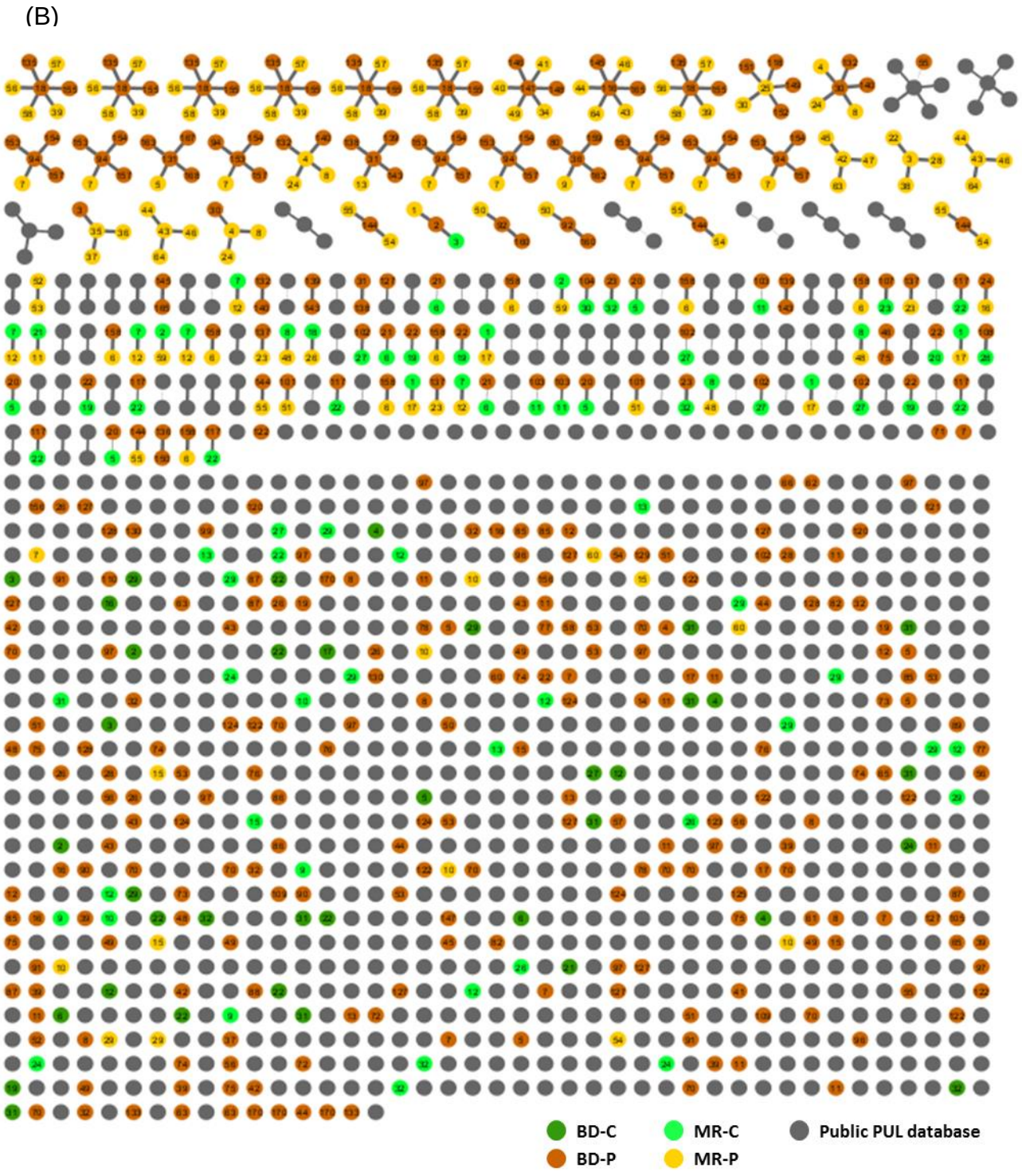
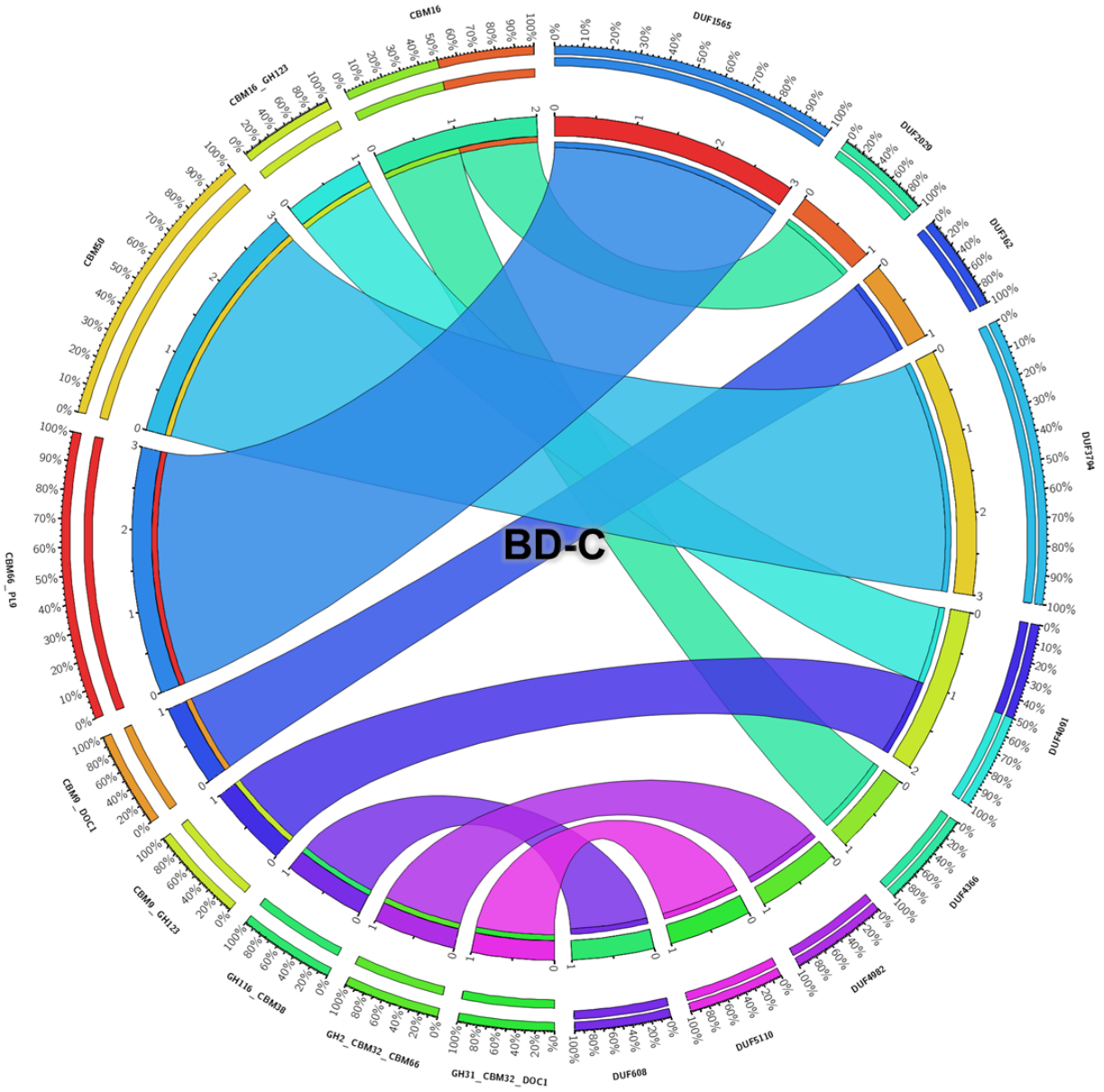
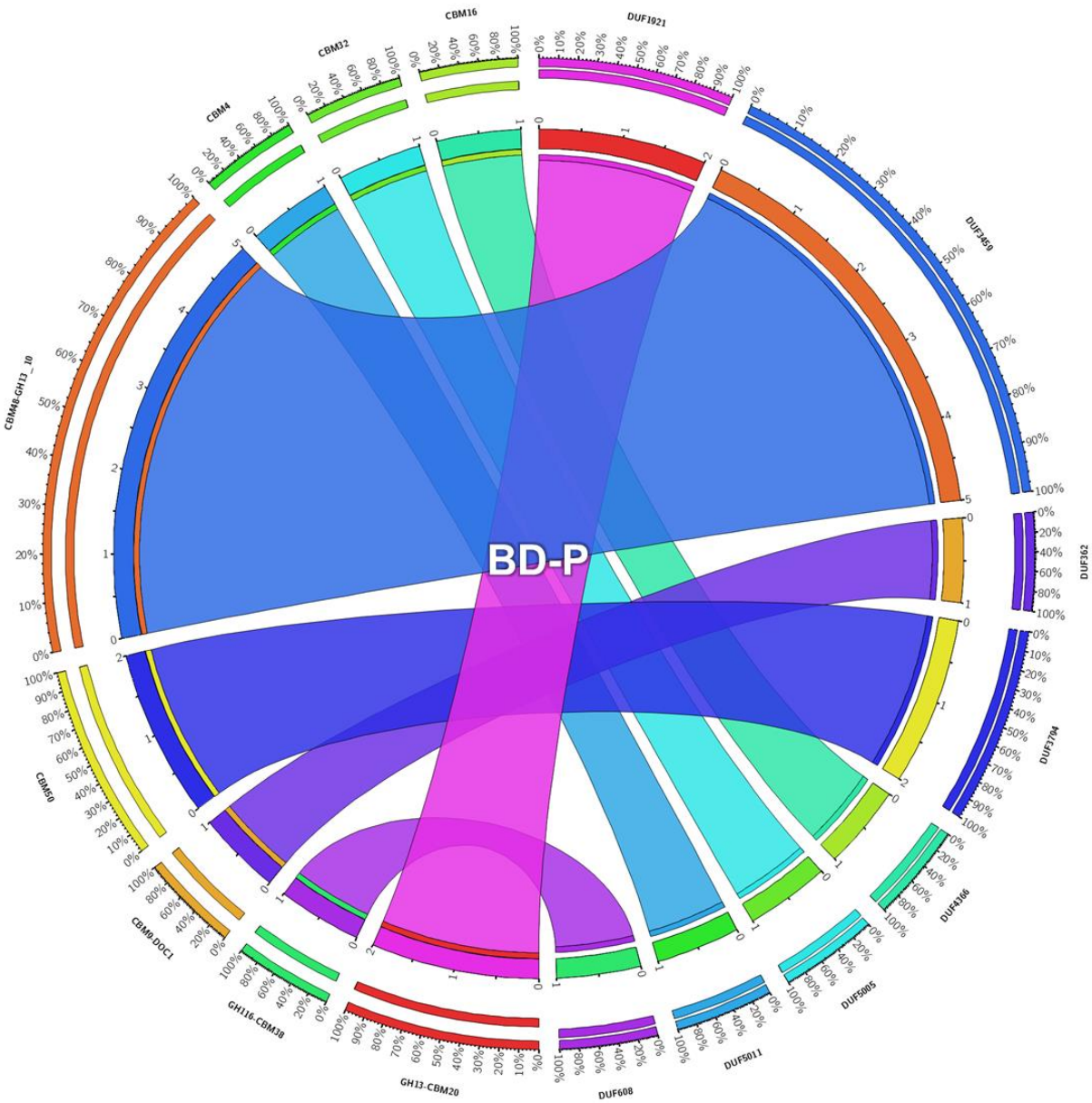
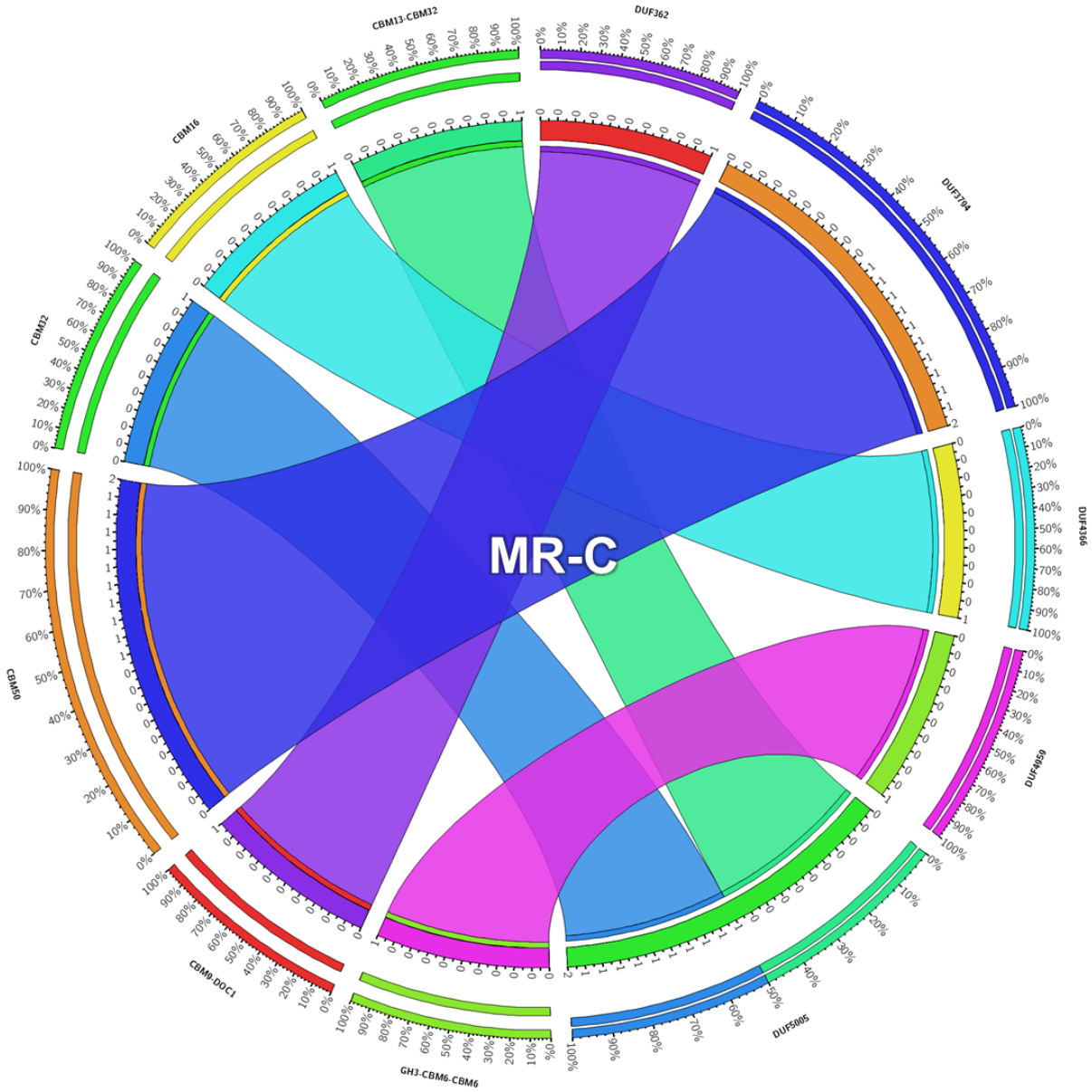


Figure S4.4 (A) Distribution of sequence length for proteins with unknown function positioned within predicted PULs. (B) Similarity-based clustering (≥ 70%) of proteins with

unknown function positioned within PULs predicted herein and the public PUL database (http://www.cazy.org/PULDB/). Each dot represents a protein with unknown function with its PUL identifier shown on top; the thickness of the edges correlates with percent identity.


BD-C
CBM16
DUF1565
DUF2029
DUF362
DUF3794
DUF4001
DUF4366
DUF4982
DUF5110
DUF608
GH31_CBM32_DOC1
GH2_CBM32_CBM66
GH116_CBM38
CBM9_GH123
CBM9_DOC1
CBM66_PL9
CBM60
CBM16_GH123


BD-P
CBM16
DUF1921
CBM32
CBM4
DUF3459
CBM48-GH13_10
DUF362
DUF5704
CBM50
DUF4366
CBM9-DOC1
DUF5005
GH116-CBM38
DUF5011
GH13-CBM20
DUF608


MR-C
DUF362
DUF3784
DUF4366
DUF4959
DUF5005
GH3-CBM6-CBM6
CBMp-DOC1
CBM50
CBM32
CBM16
CBM13-CBM32

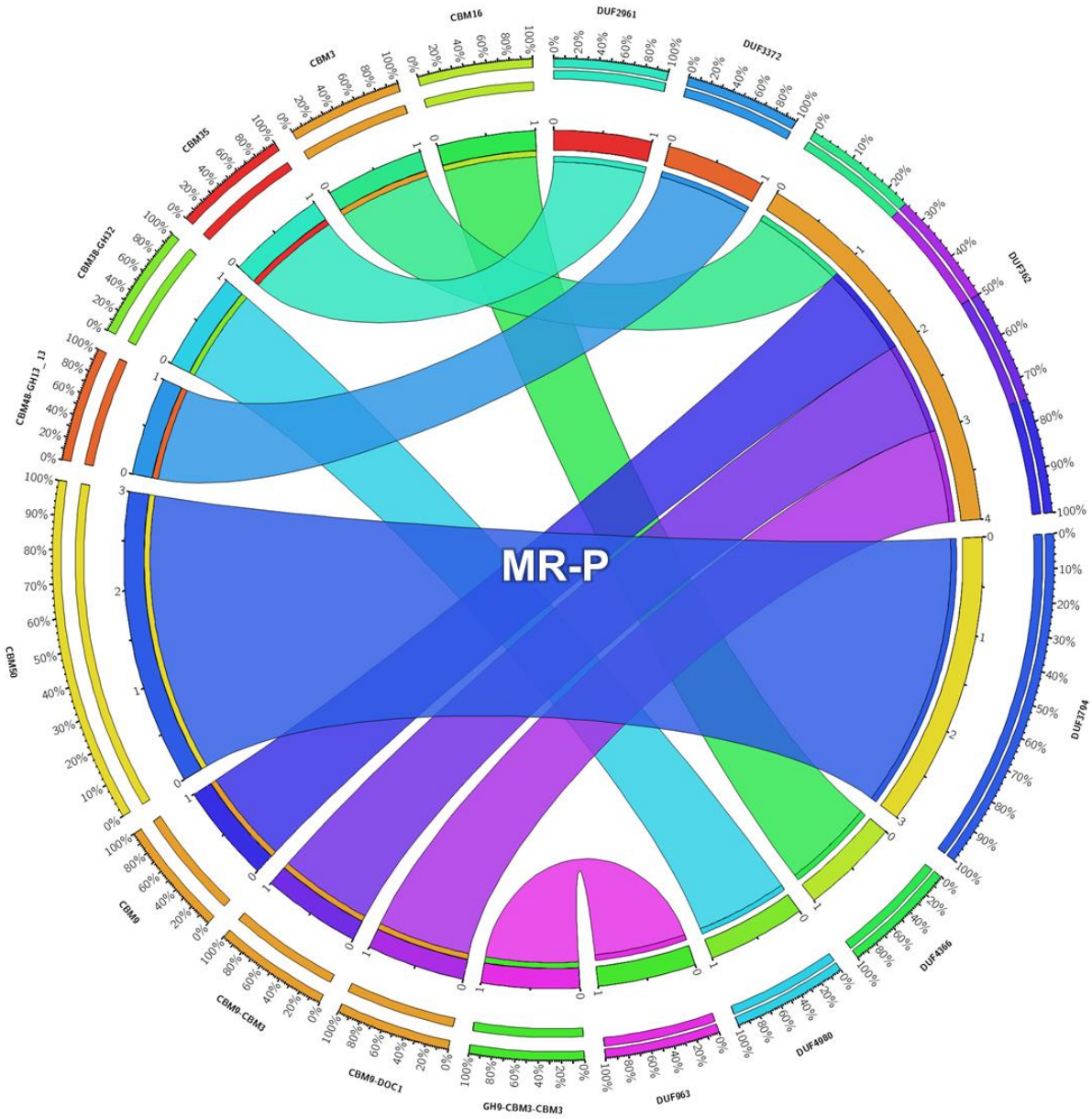


Figure S4.5 Carbohydrate-active proteins with domains of unknown functions identified in each metagenome

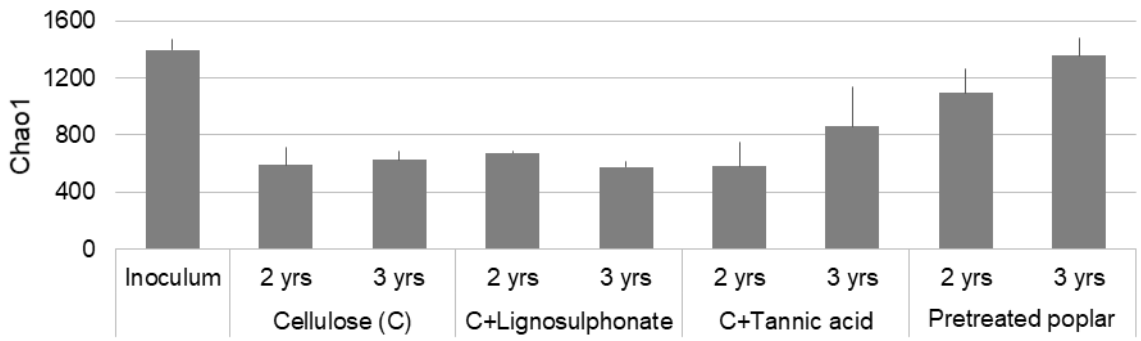


Figure S5.1 Chao1 index of the pulp mill anaerobic granule inocula and the corresponding enrichments amended with various lignocellulosic substrates. Error bars indicate standard deviation; n=3.

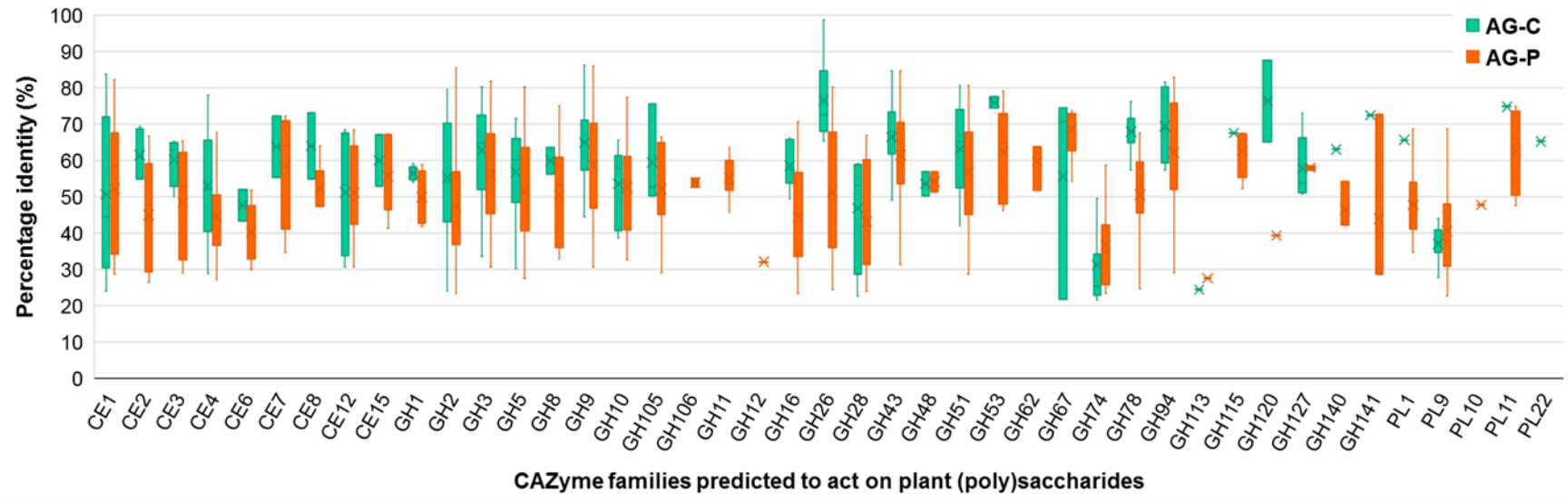


Figure S5.2 Distribution of percentage identities of the plant polysaccharide-active CAZyme families in cellulose (C)- and pretreated poplar (P)-enriched microcosms from pulp mill anaerobic granule (AG) with respective best blast hits in the CAZy database

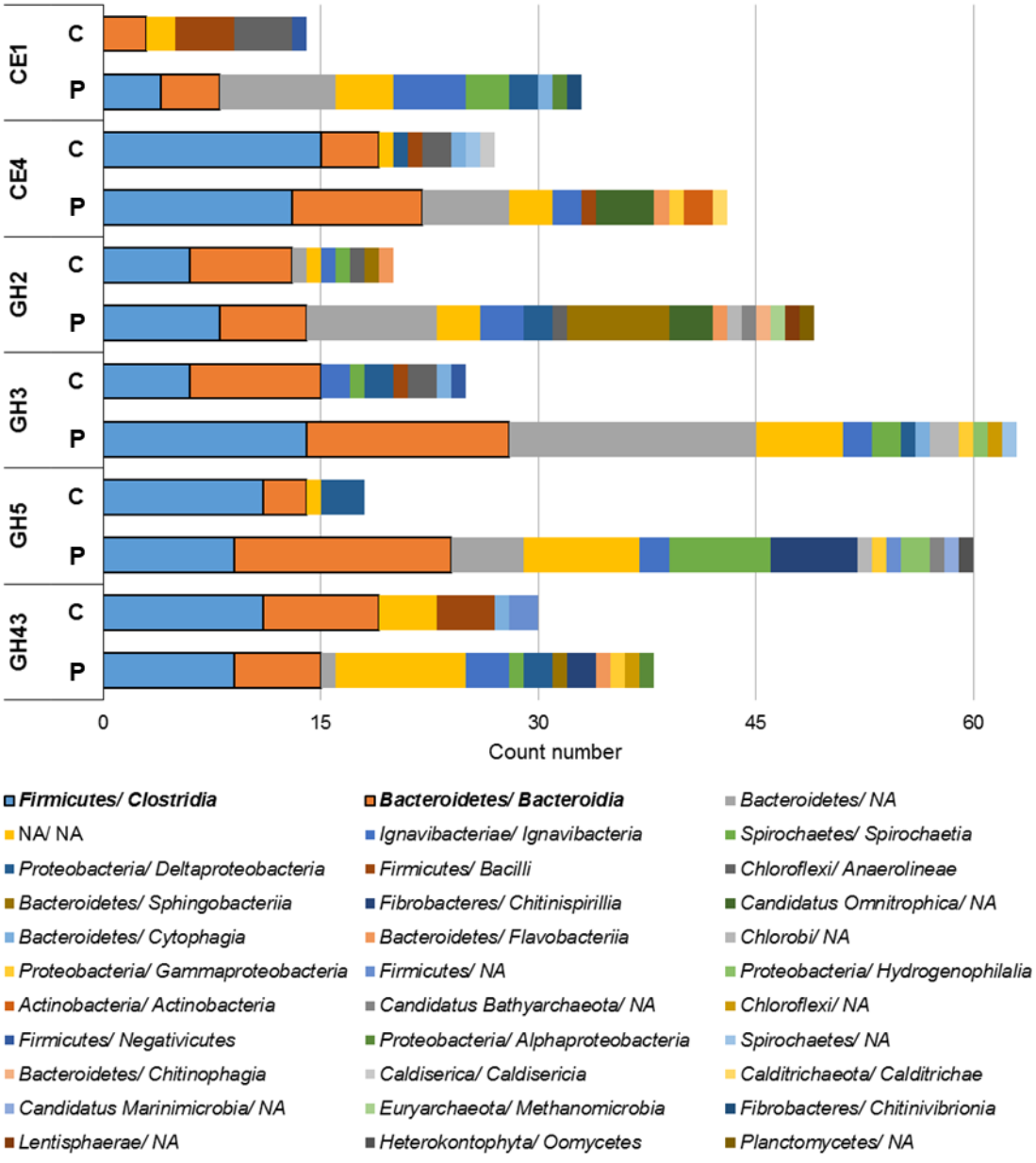


Figure S5.3 Phylogenetic distribution (phyla/class) of selected CAZymes predicted to act on plant (poly)saccharides. Outlined in black are the fractions assigned to *Clostridia* and *Bacteroidia* classes. C: cellulose, P: pretreated poplar, NA: assigned to unknown lineages.

Sample	PUL Nº	Modularity
AG-C	Predicted PUL 1	◀ SusD ◀ SusC
AG-C	Predicted PUL 2	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶
AG-C	Predicted PUL 3	◀ ECF-σ Anti-σ ▶ unk ▶ SusC ▶ SusD ▶
AG-C	Predicted PUL 4	◀ SusD ◀ SusC GH92 ▶ unk ▶
AG-C	Predicted PUL 5	◀ GH109 ◀ SusD ◀ SusC ◀ Anti-σ ◀ ECF-σ
AG-C	Predicted PUL 6	SusC ▶ SusD ▶ unk ▶ CBM48 GH13 ▶
AG-C	Predicted PUL 7	unk ▶ SusC ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
AG-C	Predicted PUL 8	◀ unk ◀ unk ◀ unk ◀ unk ◀ MFS ◀ unk ◀ GT2 ◀ unk ◀ unk ◀ unk ◀ SusD ◀ SusC ◀ Anti-σ ECF-σ ▶
AG-C	Predicted PUL 9	◀ SusD ◀ SusC
AG-C	Predicted PUL 10	◀ SusD ◀ SusC ◀ Pept_MG ◀ SusD ◀ SusC
AG-C	Predicted PUL 11	◀ SusD ◀ SusC
AG-C	Predicted PUL 12	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶
AG-C	Predicted PUL 13	◀ SusD ◀ SusC
AG-C	Predicted PUL 14	◀ SusD ◀ SusC ◀ Pept_MG ◀ SusD ◀ SusC
AG-C	Predicted PUL 15	◀ CBM32 Pept_MA ◀ CBM32 ◀ SusD ◀ SusC ◀ Anti-σ ◀ ECF-σ
AG-C	Predicted PUL 16	ECF-σ ▶ unk ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH43_31 ▶ GH2 ▶
AG-C	Predicted PUL 17	unk ▶ SusC ▶ SusD ▶
AG-C	Predicted PUL 18	◀ SusD ◀ SusC
AG-C	Predicted PUL 19	SusC ▶ SusD ▶ unk ▶ GH3 ▶ unk ▶ GH43_28 ▶ CE1 ▶
AG-C	Predicted PUL 20	SusC ▶ SusD ▶
AG-C	Predicted PUL 21	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶
AG-C	Predicted PUL 22	◀ SusD ◀ SusC
AG-C	Predicted PUL 23	CBM20 GH77 ▶ ◀ Pept_PB ◀ unk ◀ unk ◀ SusD ◀ SusC ◀ SusC ◀ unk ◀ MFS ◀ unk
AG-C	Predicted PUL 24	ECF-σ ▶ SusC ▶ SusD ▶ unk ▶ Anti-σ ▶
AG-C	Predicted PUL 25	SusC ▶ SusD ▶
AG-C	Predicted PUL 26	unk ▶ GH130 ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH26 ▶
AG-C	Predicted PUL 27	GH2 ▶ unk ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ ECF-σ ▶
AG-C	Predicted PUL 28	HTCS ▶ SusC ▶ SusD ▶ GH32 ▶ unk ▶ MFS ▶ unk ▶ CE7 ▶
AG-C	Predicted PUL 29	◀ unk ◀ unk ◀ unk ◀ unk ◀ unk ◀ unk ◀ SusD ◀ SusC
AG-C	Predicted PUL 30	◀ SusD ◀ SusC
AG-C	Predicted PUL 31	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶
AG-C	Predicted PUL 32	SusC ▶ SusD ▶ Pept_MG ▶ SusC ▶ SusD ▶
AG-C	Predicted PUL 33	SusC ▶ SusD ▶
AG-C	Predicted PUL 34	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
AG-C	Predicted PUL 35	◀ GH125 ◀ GH76 ◀ GH76 ◀ unk ◀ SusD ◀ SusC ◀ Anti-σ ◀ ECF-σ
AG-C	Predicted PUL 36	◀ SusD ◀ SusC ◀ Anti-σ ◀ ECF-σ
AG-C	Predicted PUL 37	SusC ▶ SusD ▶ unk ▶ Pept_MH ▶ unk ▶
AG-C	Predicted PUL 38	◀ SusD ◀ SusC

AG-C	Predicted PUL 39	SusC ▶ SusD ▶ SusC ▶ SusD ▶ SusC ▶ SusD ▶
AG-C	Predicted PUL 40	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ GH63 ▶ unk ▶
AG-C	Predicted PUL 41	◀ SusD ◀ SusC
AG-C	Predicted PUL 42	◀ Pept_na ◀ Pept_na ◀ unk ◀ unk ◀ Pept_MC ◀ unk ◀ Pept_PC ◀ unk ◀ unk ◀ SusD ◀ SusC ◀ Anti-σ ◀ ECF-σ
AG-C	Predicted PUL 43	SusC ▶ SusD ▶ Pept_MG ▶ SusC ▶ SusD ▶
AG-C	Predicted PUL 44	SusC ▶ SusD ▶
AG-C	Predicted PUL 45	GH3 ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ GH16 ▶ GH3 ▶
AG-C	Predicted PUL 46	◀ GH3 ◀ GH30_3 ◀ unk ◀ GH30_3 ◀ SusD ◀ SusC
AG-C	Predicted PUL 47	◀ unk ◀ SusD ◀ SusC
AG-C	Predicted PUL 48	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ Pept_MC ▶
AG-C	Predicted PUL 49	◀ unk ◀ unk ◀ SusD ◀ SusC ◀ Anti-σ ◀ ECF-σ
AG-C	Predicted PUL 50	◀ SusD ◀ SusC ◀ SusC
AG-C	Predicted PUL 51	◀ GH78 ◀ HTCS unk ▶ SusC ▶ SusD ▶ GH3 ▶ CE1 ▶ CE1 ▶
AG-C	Predicted PUL 52	SusC ▶ SusD ▶ Pept_SC ▶
AG-C	Predicted PUL 53	SusC ▶ SusD ▶
AG-C	Predicted PUL 54	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ GH130|GH2 ▶
AG-C	Predicted PUL 55	ECF-σ ▶ SusC ▶ SusD ▶
AG-C	Predicted PUL 56	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ GH92 ▶ GH92 ▶ GH92 ▶
AG-PH	Predicted PUL 1	◀ GH16 ◀ unk ◀ SusD ◀ SusC
AG-PH	Predicted PUL 2	◀ unk ◀ SusD ◀ SusC
AG-PH	Predicted PUL 3	SusC ▶ SusD ▶
AG-PH	Predicted PUL 4	◀ unk ◀ SusD ◀ SusC
AG-PH	Predicted PUL 5	◀ GH30_3 ◀ GH30_3 ◀ unk ◀ SusD ◀ SusC ◀ GH3
AG-PH	Predicted PUL 6	SusC ▶ SusC ▶ SusD ▶ unk ▶
AG-PH	Predicted PUL 7	◀ unk ◀ SusD ◀ SusC ◀ GH36 ◀ unk ◀ unk ◀ unk ◀ unk ◀ unk
AG-PH	Predicted PUL 8	◀ SusD ◀ SusC
AG-PH	Predicted PUL 9	SusC ▶ SusD ▶ unk ▶ GH3 ▶ GH16 ▶
AG-PH	Predicted PUL 10	GntR ▶ SusC ▶ SusD ▶ unk ▶ PL6 ▶ GH88 ▶ unk ▶ unk ▶ unk ▶ PL12 ▶ unk ▶ PL6 ▶ unk ▶ unk ▶ unk ▶ unk ▶ PL6 ▶ PL17 ▶ unk ▶ MFS ▶ unk ▶
AG-PH	Predicted PUL 11	◀ unk ◀ PL12 ◀ unk ◀ unk ◀ unk ◀ GH88 ◀ PL6 ◀ unk ◀ SusD ◀ SusC ◀ GntR
AG-PH	Predicted PUL 12	◀ unk ◀ PL12 ◀ unk ◀ unk ◀ unk ◀ GH88 ◀ PL6 ◀ unk ◀ SusD ◀ SusC ◀ GntR
AG-PH	Predicted PUL 13	◀ SusD ◀ SusC
AG-PH	Predicted PUL 14	◀ unk ◀ SusD ◀ unk ◀ SusC ◀ unk unk ▶ GH105 ▶ Pept_CA ▶ unk ▶ GH106 ▶ GH5_36 ▶
AG-PH	Predicted PUL 15	SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ unk ▶
AG-PH	Predicted PUL 16	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH76 ▶ GH92 ▶
AG-PH	Predicted PUL 17	◀ unk ◀ unk ◀ SusD ◀ SusC ◀ unk ◀ GH3
AG-PH	Predicted PUL 18	SusR ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH66 ▶ GH31 ▶ unk ▶
AG-PH	Predicted PUL 19	◀ unk ◀ SusD ◀ SusC ◀ GH3 ◀ GH3
AG-PH	Predicted PUL 20	◀ unk ◀ unk ◀ SusD ◀ SusC ◀ GH31

AG-PH	Predicted PUL 21	◀ HTCS ◀ unk ◀ unk ◀ GH32 ◀ unk ◀ GH32 ◀ unk ◀ SusD ◀ SusC
AG-PH	Predicted PUL 22	◀ unk ◀ GH9 ◀ GH31 ◀ GH31 ◀ GH5_4 ◀ unk ◀ unk ◀ SusD ◀ SusC ◀ Anti-σ ◀ ECF-σ
AG-PH	Predicted PUL 23	◀ GH27 ◀ unk ◀ unk ◀ unk ◀ unk ◀ unk ◀ SusD ◀ SusC ◀ GH92 ◀ unk ◀ unk
AG-PH	Predicted PUL 24	GH97 ▶ GH116 ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH130 ▶ unk ▶ MFS ▶ GH94 ▶
AG-PH	Predicted PUL 25	◀ Sulf_1 ◀ GH27|Sulf_1 ◀ Sulf_1 ◀ unk ◀ Sulf_1 ◀ Sulf_1 ◀ Sulf_1 ◀ GH110 ◀ SusD ◀ SusC
AG-PH	Predicted PUL 26	◀ GH30_3 ◀ GH30_3 ◀ GH30_3 ◀ GH30_3 ◀ unk ◀ SusD ◀ SusC ◀ GH3
AG-PH	Predicted PUL 27	◀ GH97 ◀ SusD ◀ SusC ◀ unk GH87 ▶ unk ▶ ◀ SusR
AG-PH	Predicted PUL 28	SusC ▶ SusD ▶ unk ▶ unk ▶ MFS ▶ unk ▶
AG-PH	Predicted PUL 29	◀ unk ◀ unk ◀ MFS ◀ GH92 ◀ GH38 ◀ unk ◀ unk ◀ unk ◀ unk ◀ SusD ◀ unk ◀ SusC
AG-PH	Predicted PUL 30	SusC ▶ SusD ▶ unk ▶ unk ▶
AG-PH	Predicted PUL 31	SusC ▶ SusD ▶ CBM48 ▶ ◀ unk ◀ unk ◀ CE4 ◀ unk
AG-PH	Predicted PUL 32	unk ▶ SusC ▶ SusD ▶ SusC ▶
AG-PH	Predicted PUL 33	SusC ▶ SusD ▶ unk ▶ unk ▶
AG-PH	Predicted PUL 34	SusC ▶ SusD ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
AG-PH	Predicted PUL 35	unk ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ GH3 ▶ GH3 ▶
AG-PH	Predicted PUL 36	◀ unk ◀ GH97|GH28 ◀ GH43_18 ◀ GH28 ◀ unk ◀ unk ◀ SusD ◀ SusC ◀ unk
AG-PH	Predicted PUL 37	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ CBM32|GH2 ▶ ◀ unk ◀ GH3|Pept_SE
AG-PH	Predicted PUL 38	SusC ▶ SusD ▶ unk ▶ GH28 ▶ CE12 ▶ GH105 ▶ CE8 ▶ PL1 ▶ PL1 ▶ CE8 ▶ unk ▶ PL1|CBM77 ▶ CE8 ▶ unk ▶
AG-PH	Predicted PUL 39	◀ SusD ◀ SusC
AG-PH	Predicted PUL 40	SusC ▶ SusD ▶
AG-PH	Predicted PUL 41	SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ unk ▶
AG-PH	Predicted PUL 42	GH128|CBM6 ▶ unk ▶ CBM38|GH81|CBM6 ▶ CBM32|GH3 ▶ GH5_46|CBM6 ▶ unk ▶ unk ▶ ◀ unk SusC ▶ SusD ▶ GH16 ▶ GH128|CBM6 ▶ unk ▶ unk ▶ GH2 ▶ unk ▶ GH16 ▶
AG-PH	Predicted PUL 43	ECF-σ ▶ Anti-σ ▶ SusC ▶ SusD ▶ unk ▶ GH20 ▶ GH20 ▶
AG-PH	Predicted PUL 44	GH29 ▶ ◀ GH2 ◀ GH2 SusC ▶ SusD ▶ unk ▶ GH10|CBM4 ▶ GH35 ▶
AG-PH	Predicted PUL 45	◀ HTCS ◀ GH31 ◀ Sulf_1 ◀ Sulf_1 ◀ GH43_31 ◀ SusD ◀ SusC
AG-PH	Predicted PUL 46	SusC ▶ SusD ▶ unk ▶ unk ▶ GH26 ▶ GH26 ▶ GH26 ▶
AG-PH	Predicted PUL 47	◀ unk ◀ GH20 ◀ SusD ◀ SusC
AG-PH	Predicted PUL 48	SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ GH5_2 ▶
AG-PH	Predicted PUL 49	GntR ▶ SusC ▶ SusD ▶ unk ▶ PL6 ▶ GH88 ▶ unk ▶ unk ▶ PL12 ▶ unk ▶
AG-PH	Predicted PUL 50	◀ unk ◀ unk ◀ unk ◀ SusD ◀ SusC
AG-PH	Predicted PUL 51	SusC ▶ SusD ▶
AG-PH	Predicted PUL 52	◀ unk ◀ GH13 ◀ unk ◀ unk ◀ SusD ◀ SusC unk ▶ CBM9 ▶ unk ▶
AG-PH	Predicted PUL 53	◀ SusD ◀ SusC
AG-PH	Predicted PUL 54	◀ GntR ◀ GH63 SusC ▶ SusD ▶
AG-PH	Predicted PUL 55	CE6 ▶ HTCS ▶ GH88 ▶ SusC ▶ SusD ▶ unk ▶ unk ▶
AG-PH	Predicted PUL 56	SusC ▶ SusD ▶ GH43_31 ▶ HTCS ▶ GH2 ▶ GH93 ▶ GH43_3 ▶ GH5_13 ▶
AG-PH	Predicted PUL 57	◀ unk ◀ unk ◀ unk ◀ unk ◀ unk ◀ SusD ◀ SusC
AG-PH	Predicted PUL 58	◀ SusD ◀ SusC

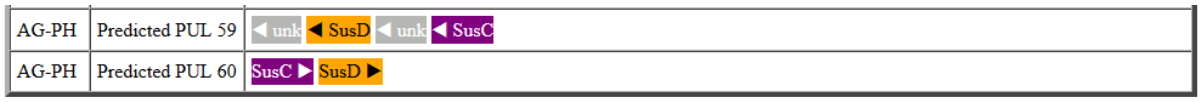


Figure S5.4 Predicted PULs from the metagenomes from cellulose (C)- and pretreated poplar (PH)-fed microcosms from pulp mill anaerobic granules

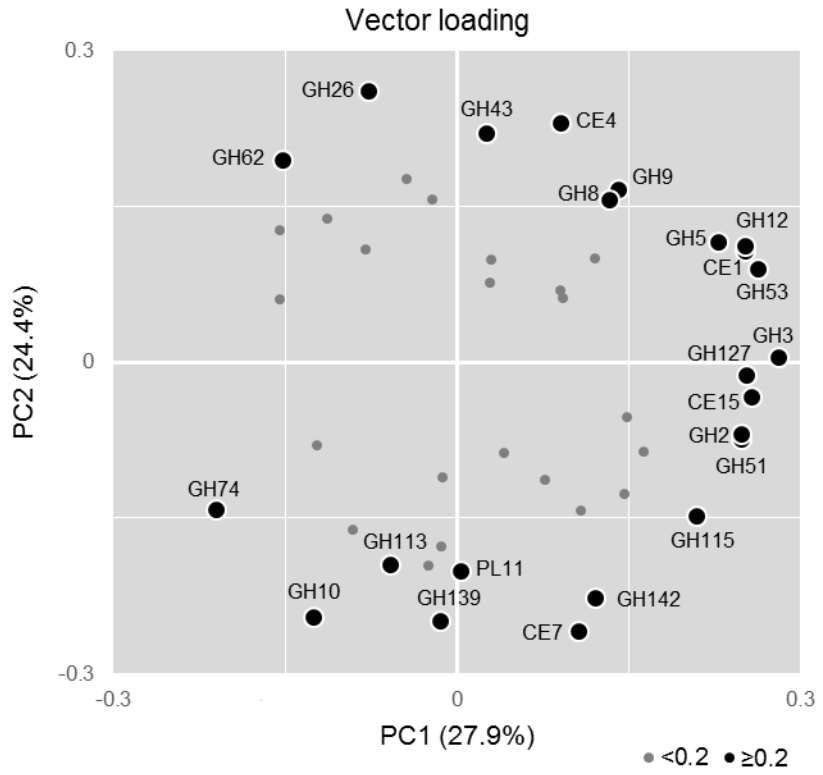
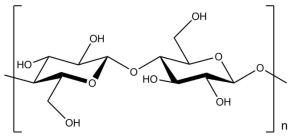
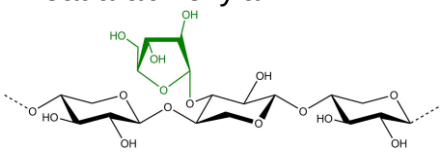


Figure S5.5 PCA loadings based on the relative abundances of CAZyme families predicted to degrade plant polysaccharides

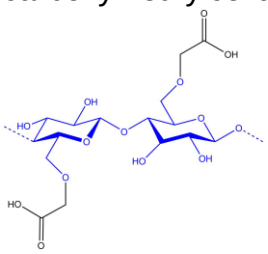
cellulose



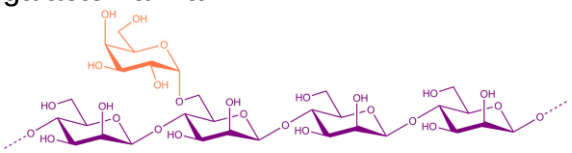
wheat arabinoxylan



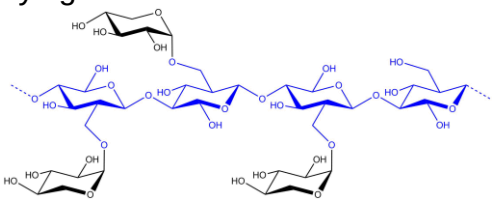
carboxymethylcellulose



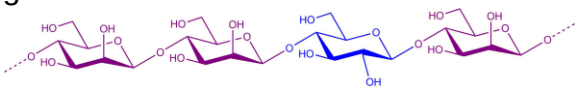
galactomannan



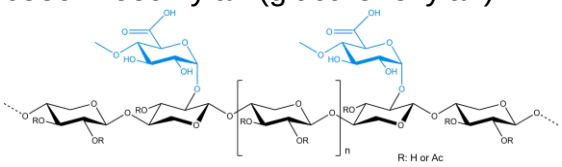
xyloglucan



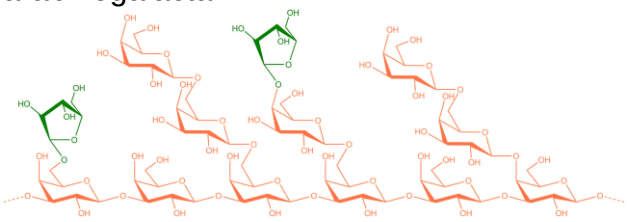
glucomannan



beechwood xylan (glucuronoxylan)



arabinogalactan



oat spelt xylan (arabinoglucuronoxylan)

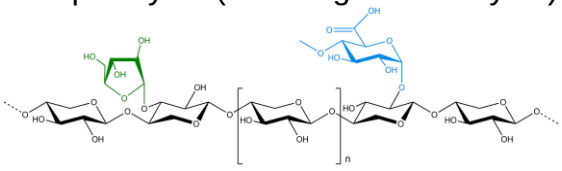


Figure S6.1 Schematic diagram of the cellulose and hemicellulose structures tested