TAXONOMIC AND FUNCTIONAL CHARACTERIZATION OF BIOPOLYMER-DEGRADING MICROBIAL COMMUNITIES IN THE INTESTINAL TRACT OF BEAVERS

Dissertation

For the award of the degree

"Doctor rerum naturalium (Dr. rer. nat.)"

of the Georg-August-Universität Göttingen

within the doctoral program in Biology

of the Georg-August University School of Science (GAUSS)

submitted by

Rahadian Pratama

from Bogor, Indonesia

Göttingen, 2019

Thesis committee

Prof. Dr. Rolf Daniel, Department of Genomics and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August-Universität Göttingen

PD Dr. Michael Hoppert, Department of General Microbiology, Institute of Microbiology and Genetics, Georg-August-Universität Göttingen

Examination board

Reviewer:	Prof. Dr. Rolf Daniel, Department of Genomics and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August-Universität Göttingen
Second reviewer:	PD Dr. Michael Hoppert, Department of General Microbiology, Institute of Microbiology and Genetics, Georg-August-Universität Göttingen

Further member of the examination board

Prof. Dr. Stefanie Pöggeler, Department of Genetics of Eukaryotic Microorganisms, Institute of Microbiology and Genetics, Georg-August-Universität Göttingen

Prof. Dr. Gerhard Braus, Department of Molecular Microbiology and Genetics, Institute of Microbiology and Genetics, Georg-August-Universität Göttingen

Prof. Dr. Fabian Commichau, Department of General Microbiology, Institute of Microbiology and Genetics, Georg-August-Universität Göttingen

PD Dr. Christian Roos, Primate Genetics Laboratory, German Primate Center

Date of the oral examination: 2nd May 2019

Praise to **Allah SWT** and his messenger **Muhammad SAW**.

To my mother who always believe in me, and

my family for their support

Table of Contents

1. INTRODUCTION

1.1.World energy consumption trend and the use of bioethanol

The economic status of the country and the welfare of its citizen are strongly influenced by energy resources and their utilization. Barnes and Floor (1996) mentioned that the energy demand of a country is determined by its economic development and vice versa. The Statistical Review of World Energy showed that most of Europe's energy consumption comes from fossil fuel (oil, natural gas, and coal) (Figure 1. 1). The oil consumption was 49% in 2005 and 48% in 2017, followed by natural gas and coal, with hardly any change (BP, 2018).
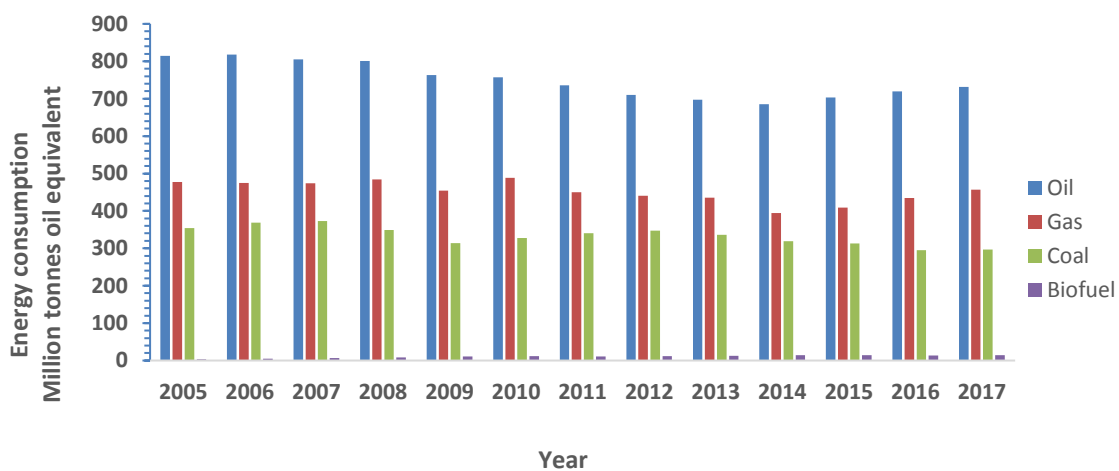
Figure 1. 1 Energy consumption in Europe by fuel type per year basis. Source: BP (2018)

Fossil fuels are mainly used for transportation (Wi et al., 2015). The fuel combustion from vehicles releases carbon dioxide (CO_2) as well as toxic particles, causing serious health problems, especially for children, which are most vulnerable (Perera, 2017). CO_2 and other greenhouse gases concentrated in the atmosphere are also major contributors to the climate change. In the ideal state, the concentrated CO_2 in the atmosphere is fixed by green plants, and in turn oxygen (O_2) is released, creating a state of equilibrium (Jain, 1993). The advance of human society and emerging technologies, however, also accelerated the consumption of fossil fuels, which led to higher CO_2 emissions into the atmosphere than nature can cope with (Jain, 1993). The economic development also promotes the transformation of forests into other systems (deforestation e.g. agriculture), which effectively reduces nature's capacity to fix CO_2. The deforestation contributes significantly to anthropogenic carbon emissions, soil degradation, biodiversity loss, and changes in the regional climate due to changes in the water balance (Robinet et al., 2018). Therefore, the use of fossil fuels should be reduced and a shift to cleaner energy sources for better living condition is needed.

5

Alternative energy sources or renewable energies offers cleaner energy than fossil fuels. The renewable energy sources must be able to reduce environmental impacts, produce little or no secondary waste, and be sustainable for the present and future needs of social and economic development (Owusu and Asumadu-Sarkodie, 2016). One form of renewable energy used as an alternative to fossil fuels is bioethanol. Currently, bioethanol is mainly produced from corn starch, in the USA and cane sugar, in Brazil. These two countries produced more than 90% of the world's bioethanol (Hood, 2016). However, the use of corn starch and cane sugar as bioethanol sources posed a further food security problem as these two materials were also used for animal feed and human food. Therefore, a non-food source for bioethanol production that does not compromise food security is required (Littlewood et al., 2014). A possible non-food source is plant-derived lignocellulosic biomass. Lignocellulosic biomass is abundant in nature, but of little value and comes as by-product from various industries. Lignocellulosic biomass stores a large energy potential that cannot be harnessed due to the recalcitrant properties of the lignocellulose polymer. If the lignocellulose polymer can be broken down into its pentose and hexose building blocks, these can be converted into bioethanol (Kang et al., 2014) (Figure 1. 2). In addition to the use of biofuels, lignocellulosic biomass was used as a material for the production of a variety of products in the pulp and paper, fiber and textile, nanocellulose, food, cosmetic, and medical industries (Zamani, 2015). These make lignocellulosic biomass a promising material from renewable resources.
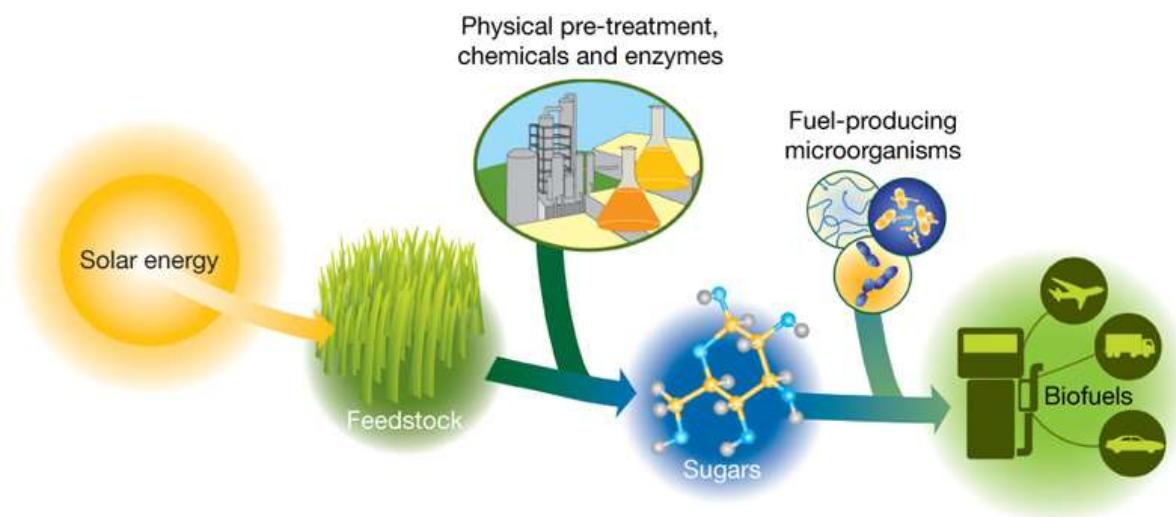
Figure 1. 2 Bioconversion of solar energy into biofuels. Decomposition of stored solar energy in the form of lignocellulosic material into simple pentose and hexose is achieved by pretreatment, followed by digestion by enzymes from biomass-degrading microorganism. The simple sugar can be subsequently converted into biofuels (Rubin, 2008). Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Nature, Genomics of cellulosic biofuels, Edward M. Rubin, ©2008.

1.1. Lignocellulose

Plant biomass is mainly composed of lignocellulose, which building blocks consist of cellulose, hemicellulose and lignin (Figure 1. 3). Cellulose is the most abundant component in plants biomass and the most abundant biopolymer on Earth (Behera et al., 2017). Cellulose is comprised of D-glucopyranose monomers linked via β-(1,4)-glycosidic bonds. Cellulose polymers stack together to form microfibrils, building the backbone of the cell wall (Chen, 2014). The gaps formed in the microfibrils are filled with hemicellulose and lignin. Hemicellulose, as the second most abundant biopolymer on earth, is comprised of heterogenous polysaccharides, which are characterized neither as cellulose nor pectin, and having equatorial β-(1,4)-linkages. Based on this configuration, hemicelluloses are grouped into xylans, xyloglucans, mannans, and glucomannans (Scheller and Ulvskov, 2010).
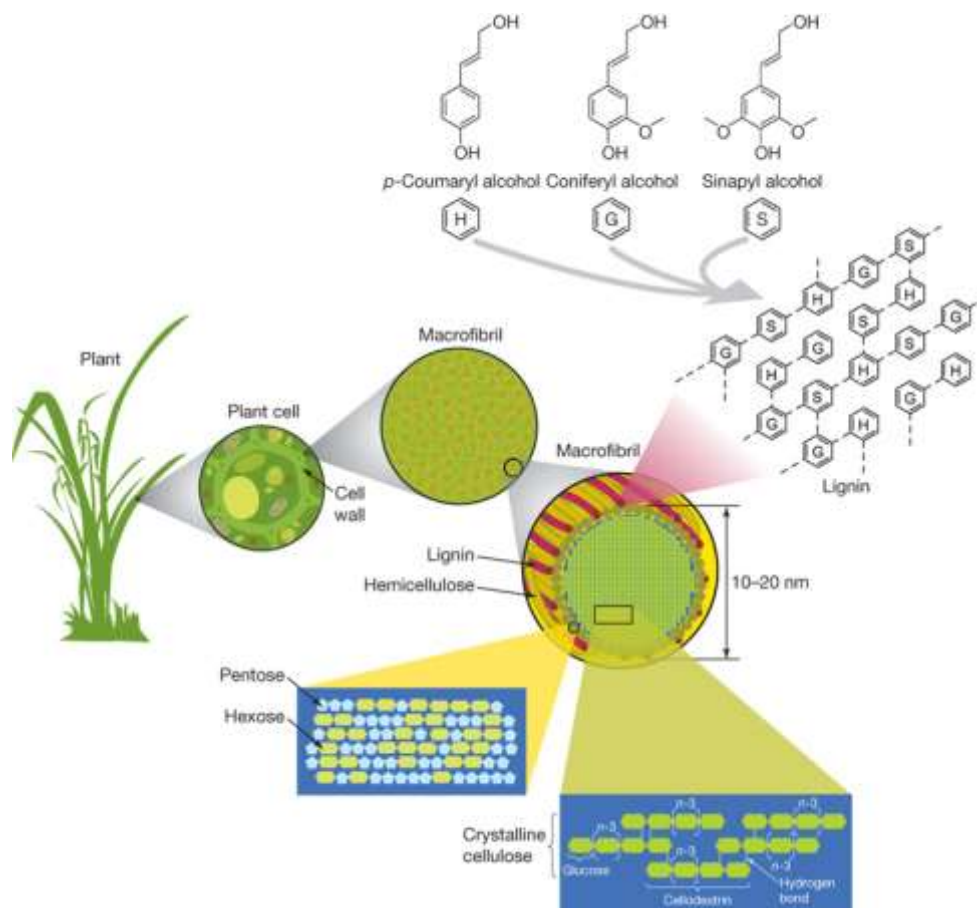
Figure 1. 3 The structure of lignocellulose. The plant cell wall constituent, cellulose, hemicellulose, and lignin are shown. The β-(1,4)-glycosidic linkage are the backbone of cellulose polymer. Source: Rubin (2008). Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Nature, Genomics of cellulosic biofuels, Edward M. Rubin, ©2008.

Lignin, which holds cellulose and hemicellulose together, is the third most abundant biopolymer. It further strengthens the cellulose-hemicellulose structure, makes it insoluble in water, and protects the cellulose against enzymatic attacks (Sanderson, 2011). Lignin is a highly

cross-linked polymer consisting of 4-hydroxy-phenylpropanoid monomers (monolignols) linked by various ether and carbon-carbon bonds (de Gonzalo et al., 2016). The phenylpropanoid monomers are coniferyl alcohol (guaiacyl propanol), coumaryl alcohol (p-hydroxyphenyl propanol), and sinapyl alcohol (syringyl alcohol), with the proportion of each monolignols depending on the plant species and tissue (Bajpai, 2016; de Gonzalo et al., 2016). During the conversion of cellulose and hemicellulose from biomass into an end product, e.g. ethanol, lignin restricts the enzymes access to these substrates, and thus effectively inhibit the conversion process. Therefore, lignin removal, also called delignification, is crucial in lignocellulose bioconversion. Delignification increases cell wall porosity and enables access for enzymatic hydrolysis (Wi et al., 2015). The proportion of cellulose, hemicellulose and lignin from biomass feedstock are given in Table 1. 1.

1.2. Lignocellulose biodegradation

To liberate simple sugars from the lignocellulose biomass, the physical barrier formed by lignin has to be removed. The removal of lignin polymer increases the access area for hydrolytic enzymes to attack. Subsequently, various enzymes work in concert to degrade cellulose and hemicellulose material and release simple sugars such as pentoses and hexoses (Pauly and Keegstra, 2008). In nature, fungi and bacteria have developed the necessary mechanisms to remove lignin from lignocellulosic biomass.

Table 1. 1 The composition of lignocellulosic biomass.

Lignocellulosic biomass	% of total dry weight		
	Cellulose	Hemicellulose	Lignin
Bamboo	49 – 50	18 – 20	23
Corn cobs	45	35	15
Corn stover	35 – 40	21 – 25	11 – 19
Grasses	25 – 40	35 – 50	10 – 30
Hardwood stems	40 – 50	24 – 40	18 – 25
Nut shells	25 – 30	25 – 30	30 – 40
Rice straw	29 – 35	23 – 26	17 – 19
Softwood stems	45 – 50	25 – 35	25 – 35
Sugar cane bagasse	25 – 45	28 – 32	15 – 25
Switch grass	30 – 50	10 – 40	5 – 20
Wheat straw	33 – 40	20 – 25	15 – 20

adapted from Mukhtam (2016)

The white-rot fungi are known to possess lignase, the key enzyme in the degradation of lignin. The peroxidase enzymes (lignin peroxidases (LiP), manganese peroxidases (MnP)) and phenol oxidase (laccases) from *Phanerochaete chrysosporium*, *Pleurotus ostreatus*, and *Trametes*

versicolor are the best studied phenol oxidases from white-rot fungi (Malherbe and Cloete, 2002). In addition, white-rot fungi have the ability to digest only lignin and hemicellulose and leave the cellulose intact (Amin et al., 2017). A protein homologous to LiP and MnP from white-rot fungi was not encountered in ligninolytic bacteria. In 1999, another type of peroxidases was isolated from *Bjerkandera adusta*, the dye-decolorizing peroxidases or DyPs (de Gonzalo et al., 2016). The DyPs were divided into four classes based on sequence characteristics, DyPs class A to D (de Gonzalo et al., 2016). DyPs of class A to C are bacteria-specific and DyPs of class D are found in fungi (Fawal et al., 2012). Although bacterial DyPs have a different protein structure compared to fungal peroxidases, they show similar catalytic properties. Moreover, both are secreted via the Tat secretion machinery (de Gonzalo et al., 2016).

1.2.1. Enzymes facilitate lignocellulosic biomass degradation

After lignin removal, the hemicellulose and cellulose matrix are accessible for enzymatic digestion. Hemicellulose and cellulose breakdown require a set of enzymes called hemicellulases and cellulases. Hemicellulose exists in different forms depending on the plant type and species, as mentioned above. Thus a wide range of hemicellulases such as endoxylanases (EC 3.2.1.8), endo-β-mannanases (EC 3.2.1.78), β-xylosidases (EC 3.2.1.37), β-glucosidases (EC 3.2.1.21), arabinosidases (EC 3.2.1.55), galactosidases (EC 3.2.1.23), mannosidases (EC 3.2.1.25), and glucuronidases (EC 3.2.1.31) are required to decompose the hemicellulose polymer, making the cellulose microfibril more accessible to cellulases (López-Mondéjar et al., 2016). The cellulose microfibrils in turn are digested by different types of cellulases due to their catalytic mode of action (Figure 1. 4). Endoglucanases (EC 3.2.1.4) randomly attack the cellulose chain in the amorphous region, exoglucanases or cellobiohydrolases (EC 3.2.1.91) attack at reducing or non-reducing ends of the cellulose chain, and β-glucosidases (EC 3.2.1.21) hydrolyse the product cellobiose of the aforementioned enzyme reactions (Liu et al., 2018; Nutt et al., 1998; Sadhu, 2013). Most aerobic cellulolytic bacteria, e.g. *Bacillus brevis* and *Pseudomonas fluorescens* secrete high amounts of extracellular cellulases (Singh and Kumar, 1998; Yamane and Suzuki, 1988), while anaerobic cellulolytic bacteria, e.g. *Clostridium thermocellum* produce a complex and efficient cellulolytic machinery called cellulosome (Behera et al., 2017; Himmel et al., 2010). The cellulosome was identified and characterized from *C. thermocellum* in the early 1980s by Bayer, Lamed and their colleagues (Bayer et al., 1985; Lamed et al., 1983).

Cellulosomes consist of scaffolding protein containing cohesin modules for the incorporation of different enzymes, e.g. endoglucanases, carbohydrate-binding modules and its complement module, dockerin (Figure 1. 5). The cohesin-dockerin interaction is important for cellulosome assembly as cellulosomes differ between bacterial species (Artzi et al., 2017). The cellulosomes usually bound to the cell surface via an anchoring scaffoldin or adaptor scaffoldin (in a more elaborate system), and also present as inherent free state outside the cell (Figure 1. 5). The hemicellulases and cellulases, and other type of enzymes that build, modify, and breakdown oligo- and polysaccharides are collectively referred to as Carbohydrate-Active enZymes (CAZymes) (Cantarel et al., 2009).
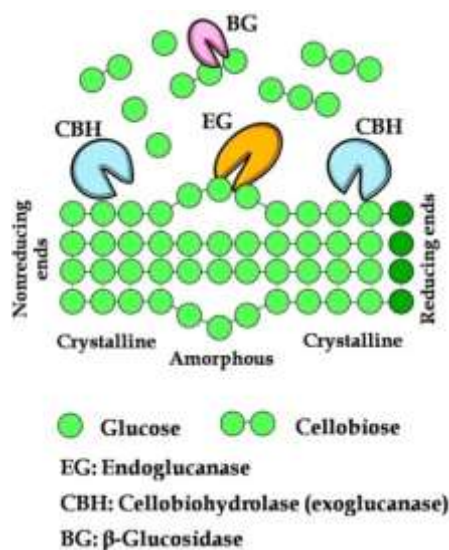
Figure 1. 4 Illustration of the mode of action from cellulases. The endoglucanase, exoglucanase, and β-glucosidase shown to act on the different part of cellulose microfibril. Source: adapted from Ratanakhanokchai et al. (2013).

1.2.2. CAZy database encompassing various known and novel CAZymes

The abundance of CAZymes in nature was classified based on protein sequence similarity and divided into different families. To date, several CAZymes families are registered within the CAZy database, an extensive database encompassing various known and novel CAZymes (http://www.cazy.org/) (Cantarel et al., 2009). The CAZy database includes families in the following classes of enzyme activities: 1) Glycoside hydrolases (GHs), responsible for the hydrolysis and/or transglycosylation of glycosidic bonds; 2) Glycosyltransferases (GTs), responsible for the biosynthesis of glycosidic bonds from phosphor-activated sugar donor; 3) Polysaccharide lyases (PLs), which cleave the glycosidic bonds of uronic acid-containing polysaccharides by a β-elimination mechanism; 4) Carbohydrate esterases (CEs), which remove ester-based modification present in mono-, oligo-, and polysaccharides, thereby facilitating the action of GHs on complex polysaccharides; 5) Carbohydrate-binding modules (CBMs), which

per se do not exhibit catalytic activity, but are known to support the activities of many aforementioned enzymes by targeting and promoting a prolonged interaction with the substrate (Cantarel et al., 2009). In 2013, the CAZy database was updated with an additional class, the Auxiliary Activities (AAs), which accommodate a broad range of enzyme mechanisms and substrates related to lignocellulose conversion, including lignin-degrading enzymes and polysaccharide lytic monooxygenates (Lombard et al., 2014). The CAZymes within each class were divided into several families (and some to subfamilies) to accommodate the differences in protein sequence and structure similarity (Cantarel et al., 2009).
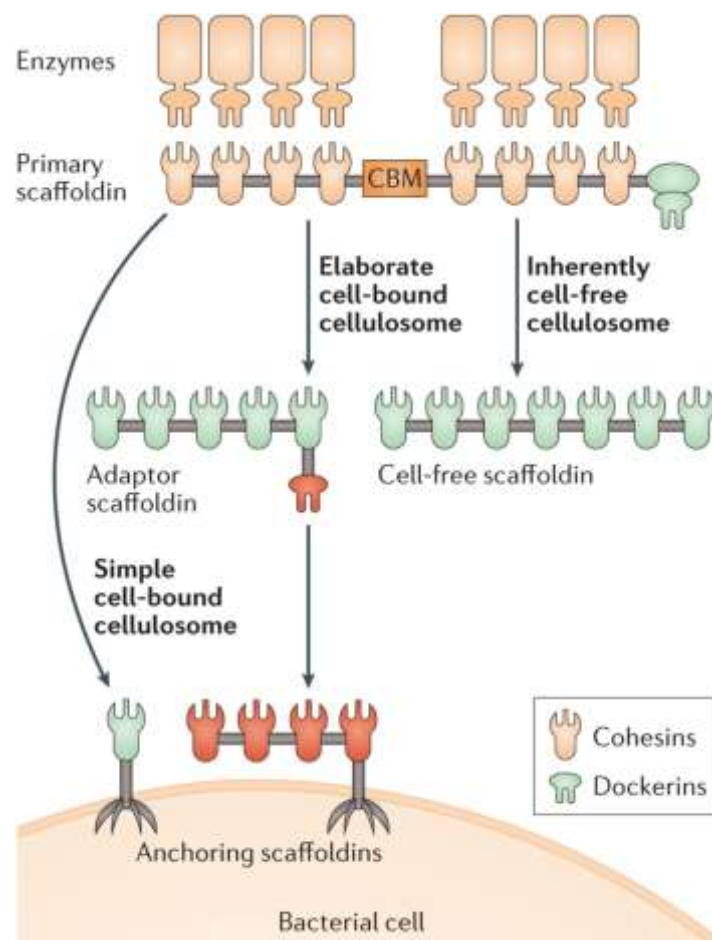
Nature Reviews | Microbiology

Figure 1. 5 The types of cellulosome system in various anaerobic cellulolytic bacteria. Source: Artzi *et al*. (2017). Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Nature Reviews Microbiology, Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides, Lior Artzi, Edward A. Bayer, Sarah Moraïs, ©2016.

Enzymes with cellulolytic activity belong to the families of GH 1, GH3, GH5, GH6, GH8, GH9, GH12, GH45, GH48, GH51 and GH74. In addition, enzymes associated with hemicellulolytic activity belong to GH2, GH10, GH11, GH16, GH26, GH30, GH31, GH39, GH42, GH43 and GH53 (López-Mondéjar et al., 2016). Both enzyme activities are present in various bacterial

communities especially those from soil, since lignocellulosic biomass is abundant in soil in the form of dead plant material or leaf litter (López-Mondéjar et al., 2016; Wang et al., 2009). Among GH families, GH5 is one of the largest families with a large variety of substrate specificities and is frequently found in the metagenomes of diverse microbial communities, as well as the genomes of individual organisms (Aspeborg et al., 2012). GH5, formerly known as cellulase family A, is the first cellulase family described. Many of the characterized or putative genes encoding cellulases from known cellulolytic bacteria, e.g. *C. cellulolyticum*, *C. clariflavum*, *C. ruminicola*, *C. thermocellum*, *Mucilaginibacter* L294, are GH5 cellulases (Artzi et al., 2015; Berger et al., 2007; Blouzard et al., 2007; Brumm et al., 2015; Cai et al., 2010; López-Mondéjar et al., 2016). The three cellulase enzyme types (endoglucanase, exoglucanase, beta-glucosidase) are included in this family. As of 19 February 2019, GH5 family contained a total of 13,292 protein sequences, 81.64% from bacteria, 16.68% from eukaryotse, 0.71% from archaea, 0.06% from viruses, and 0.91% were unclassified so far. The presence of genes encoding cellulases in bacterial genomes in 24% the sequenced genomes indicated that bacteria are potential cellulose degraders (Berlemont and Martiny, 2013).

1.3. Herbivorous gut as source of cellulolytic microorganisms and enzymes

The mammalian gut systems are home to complex microbial communities. The microbes and their genes in the gut form what we called nowadays the gut microbiome (Ursell et al., 2012). In humans alone, the gut microbiome outnumber the human cells by a factor of 10-100 (Bleich and Fox, 2015). The early thought of the microbiome residing in the mammalian gut was those of a commensal relationship, which neither benefits from the other or causing any harm. Continuous studies on the gut microbiome and its benefit to the host conclude that the relationship between host and gut microbiome tends to be mutualistic (Bäckhed et al., 2005; Chow et al., 2010; De Filippo et al., 2010; Macpherson et al., 2011). The gut microbiome provides benefits to its host by providing essential nutrients, defending against opportunistic pathogens, assisting in the development of intestinal architecture, as well as in the degradation of recalcitrant material such as lignocellulosic biomass (Hooper, 2001).

The herbivorous animal gut systems are considered rich in cellulolytic microorganisms, as their feed is high in lignocellulose content. The gut compartments such as the rumen in ruminants and the cecum in herbivorous non-ruminant animal harbor a plethora of microbial communities assisting them to degrade recalcitrant plant cell wall material (Montgomery and Macy, 1982; Wang and McAllister, 2002). To date, diverse herbivorous gut bacterial communities have been taxonomically and functionally characterized, and some of the bacterial key players for

degrading plant cell walls have been identified. The rumen is one of the most remarkable ecosystems for the degradation of lignocellulosic biomass, and many studies have been conducted to investigate the lignocellulolytic capacity of the microbial community therein (Brulc et al., 2009; Cai et al., 2010; Cai and Dong, 2010; Hess et al., 2011; Koike and Kobayashi, 2001; Tajima et al., 2000; Wang and McAllister, 2002). Another herbivorous gut microbiomes, which have been extensively explored, were that of the termites (Brune and Dietrich, 2015). Termites form symbiotic relationships with a dense assemblage of microorganism that contribute functions, which are lacking in the host, e.g. lignocellulosic biomass breakdown (Breznak and Brune, 1994). Due to their internal digestive organ structure and the gut microbiome, termites were divided into lower termites and higher termites. The lower termites host a dense and diverse population of bacteria and cellulose-digesting flagellate protozoa. The higher termites, which comprise three quarters of the termite species, also host a dense and diverse range of gut bacteria, but in general lack protozoa, and have a more elaborate external and internal anatomy and social organization (Breznak and Brune, 1994). In addition to ruminants and termites, the study of the gut bacterial communities from herbivorous animals, e.g. giant and red pandas (Kong et al., 2014; Li et al., 2015; Zhu et al., 2011), snails and slug (Cardoso et al., 2012; Joynson et al., 2017), beetles (Scully et al., 2013), rats (Montgomery and Macy, 1982), capybaras (García-Amado et al., 2012), and recently from North American beavers have been performed (Gruninger et al., 2016; Wong et al., 2016, 2017).

1.3.1. Eurasian beaver gut microbiome as a source of potentially new lignocellulases

The Eurasian beaver is a large semi-aquatic rodent that feeds on tree bark and some aquatic plants (Figure 1. 6). It is the second largest rodent after the capybara (*Hydrochoerus hydrochaeris*) (Nolet and Rosell, 1998). The Eurasian beaver (*Castor fiber*) is one of two remaining species of the genus *Castor*, the other being the North American beaver, *Castor canadensis* (Campbell-Palmer *et al*. 2015). Both species suffer from declining population at the end of 19th century, due to excessive hunting (Rosell et al., 2005). The beaver pelt was used in clothing as fur, felt or leather; their claws were used for decorations; the tail was used for meat and the skin was used for making pouches; and their castoreum gland was used for medical and fragrance purposes (Campbell-Palmer et al. 2015). The reintroduction program was initiated in the 1920s to recover both beaver species in their habitat (Rosell et al., 2005). By 2003, the estimated population of the Eurasian beaver was 639,000 (Halley and Rosell, 2003), while the North American beaver population in 1986 was estimated to be 6-12 million (Naiman et al., 1986).

Figure 1. 6 Eurasian beaver (*Castor fiber*). Source: Antje, Weber ©, *Büro Wildforschung & Artenschutz*, 39649 Gardelegen, Germany.

The Eurasian beaver's ability to digest hardwood is associated to its gut microbiome, which facilitates the degradation of recalcitrant lignocellulosic material. Studies on the degradation of lignocellulose by beavers have been of interest, since Currier (1958) conducted a cellulose degradation experiment with the fluid derived from the beaver cecum. From this time on, studies of beaver gut regarding its microbiome begun to emerge, especially regarding the cellulolytic capability (Armstrong et al., 2018; Gruninger et al., 2016; Hoover and Clarke, 1972; Wong et al., 2016, 2017). Recently, Gruninger *et al.* (2016) were able to classify bacterial and archaeal communities in the North American beaver gut system, showing the dominance of Firmicutes*,* Bacteroidetes and Methanosphaera. A better understanding of the beaver gut microbiome potential in degrading lignocellulose biomass, especially from the Eurasian beaver which is not yet explored, might reveal potential bacteria or enzymes to improve lignocellulose conversion to variety of product such as bioethanol, pulp and paper, fibre and textiles, food additive, and nanocelluloses among others (Zamani, 2015).

To identify bacteria, normally studies were performed through culturing using suitable media and subsequent identification by biochemical and phenotypic characteristics of the isolated bacteria. The gut bacteria in particular are difficult to culture (Wang et al., 2017). Over the last decade, the development of high-throughput sequencing has enabled the study of the whole microbial community from diverse environments. Metagenomic profiling using the 16S rRNA gene or whole metagenome shotgun sequencing have been used and successfully employed to analyze taxonomic and functional compositions of gut microbial communities. In addition,

metagenomics have been used to recover biotechnologically relevant genes and the corresponding gene products, including carbohydrate-active enzymes which are associated to lignocellulose breakdown (Cardoso et al., 2012; Joynson et al., 2017; Nacke et al., 2012).

1.4. Aim of the study

The culture-independent analysis of structure and function of the microbial communities in the gut system of the Eurasian beaver were in the focus of thesis. Prior to the study, only some information on the gut microbiome of beavers were derived from culture-independent approaches, but all of data were gathered from the North American beaver (Gruninger et al., 2016; Wong et al., 2016, 2017), and comparable information on the Eurasian beaver was lacking. In this study, the taxonomy and functional characterization of the bacterial communities were performed from the complete gut system of three Eurasian beavers. In this way, taxonomic and functional profiles of the bacterial communities along all parts of the gut systems were obtained and key organisms and genes involved in plant cell wall breakdown were identified. In addition, identification and characterization of (hemi)cellulases present in the gut microbiome of the Eurasian beaver were another focus of the thesis.

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Beavers gastrointestinal tract

The Eurasian beaver is a protected animal in Germany. No animal was harmed or killed in the course of this study. Samples were taken from three beaver carcasses in Lutherstadt-Wittenberg, Saxony-Anhalt Germany, which died from traffic accidents in the Biosphärenreservat Mittelelbe (Figure 2. 1A). The dead beavers, a juvenile (body weight 9.2 kg), sub-adult male (body weight 17.5 kg), and sub-adult female (body weight 14.6 kg) were collected by personnel of the Biosphärenreservat and stored at -20 °C until dissection. The samples covered the entire gut system, from stomach to colon (Figure 2. 1B). Gut contents from the stomach (Sto), duodenum (Duo), jejunum (Jej), ileum (Ile), front caecum (FC), back caecum (BC), upper colon (U.col), middle colon (M.col) and lower colon (L.col) were extracted from the sub-adult female beaver. The same gut contents were extracted from the male sub-adult and the juvenile beaver gut, except the duodenum part of sub-adult male and small intestine part (Duo, Jej, and Ile) of juvenile male beaver, as there was no gut content in these compartments. Samples were stored at -80 °C until further processing.


Figure 2. 1 Beaver samples and digestive system of the beaver. (**A**) Carcass of the Eurasian beaver (source: Antje, Weber ©, *Büro Wildforschung & Artenschutz*, 39649 Gardelegen, Germany). (**B**) The spread-out digestive system of the beaver showing the stomach (1), duodenum (2), jejunum (3), ileum (4), front cecum (5), back cecum (6), upper colon (7), middle colon (8) and lower colon (9).

2.1.2. Primers

Primer used in this study are listed in the Table 2. 1.

Table 2. 1 Sequences and references of primers used in this study. Sequences depicted in **bold** are the MiSeq overhang, normal sequence are the binding region.

Primer	Sequence (5' → 3')	Reference
S-D-Bact-0341-b-S-17	**TCGTCGGCAGCGTCAGATGTGTATAAGA GACAG**-CCTACGGGNGGCWGCAG	Klindworth et al., 2013

Table 2.1. Sequences and references of primers used in this study. continued		
S-D-Bact-0785-a-A-21	**GTCTCGTGGGCTCGGAGATGTGTATAA GAGACAG**-GACTACHVGGGTATCTAATCC	Klindworth et al., 2013
For_BC33a	ATGCGCATTGAGGCGATC	This study
Rev_BC33a	TCACGCCTTGGAGCGC	This study
For_BC33b	CACCATGCGCATTGAGGC	This study
Rev_BC33b	CGCCTTGGAGCGCAC	This study

2.2. Methods

2.2.1. Pretreatment of samples for direct metagenomic sequencing

The sample for direct metagenomic sequencing, back cecum and lower colon from both sub-adult beavers, were pretreated by filtration to remove plant debris and eukaryotic cells. Approximately one gram of each sample was mixed with 25 mL saline water (0.9% NaCl). This mixture was filtrated through 1.) coffee filter (size 4, Konos GmbH, Nossen, Germany), 2.) nylon filter (10 µm, Merck KGaA, Darmstadt, Germany), and 3.) glass microfiber filter (2.7 µm, Whatman GF/D, GE Healthcare, Freiburg, Germany). The filtrate was then centrifuged at 4,000 rpm (Centrifuge 5810R, Eppendorf Vertrieb Deutschland GmbH, Wesseling-Berzdorf, Germany) for 30 min to obtain a bacterial pellet. Subsequently, the metagenomic DNA was isolated directly

2.2.2. DNA extraction and nucleic acid purification

Metagenomic DNA was extracted from approximately 100 mg wet gut content (for 16S rRNA gene sequencing) or bacterial pellet (for direct sequencing) following the repeated bead-beating and column method with minor modifications (Yu and Morrison, 2004). In brief, cells were lysed by bead-beating (300 mg 0.1 mm glass bead, 100 mg 0.5 mm glass bead) in 500 mM NaCl, 50 mM Tris-HCl, 50 mM ethylenediaminetetraacetic acid (EDTA) and 4% sodium dodecyl sulfate (SDS). The bead beating process was done using a Micro Dismembrator for 30 s at 1,000 rpm (bbi-biotech GmbH, Berlin, Germany). After bead-beating, most of the impurities and SDS were removed by precipitation with 10 M ammonium acetate. To remove remaining contaminations such as protein, the resulting nucleic acid pellet was further purified using the DNeasy PowerClean Pro Cleanup Kit following the instructions of the manufacturer (Qiagen GmbH, Hilden, Germany).

2.2.3. Amplification and sequencing of bacterial 16S rRNA genes

Amplification of bacterial 16S rRNA genes was performed using the forward and reverse primers S-D-Bact-0341-b-S-17 and S-D-Bact-0785-a-A-21 targeting the V3 to V4 hypervariable region (Klindworth et al., 2013). Amplification was performed in a total volume of 50 µl containing 1 U Phusion high-fidelity DNA polymerase (Thermo Fisher Scientific, Massachusetts, USA), 10 µl of 5x Phusion GC Buffer, 0.2 mM of each primer, 10 mM dNTPs, 0.2 µl of 50 mM $MgCl_2$, 5% DMSO and 25 ng of metagenomic DNA. Thermal cycling was carried out as follows: initial denaturation for 1 min at 98 °C, followed by 25 cycles of 45 s at 98 °C, 45 s at 60 °C, 30 s at 72 °C and final elongation for 5 min at 72 °C.

The correct amplicon size (approximately 550 bp) was verified by agarose gel electrophoresis. Subsequently, the PCR products were purified using the magnetic bead kit NucleoMag 96 PCR as recommended by the manufacturer (Macherey-Nagel GmbH & Co. KG, Düren, Germany). Quantification of amplicons was conducted with the Qubit Fluorometer using the dsDNA HS assay kit (Invitrogen GmbH, Karlsruhe, Germany). Indices for Illumina sequencing were attached to the generated PCR products by using the Nextera XT index kit as recommended by the manufacturer (Illumina, San Diego, USA). Subsequently, the amplicons were sequenced by using the dual index paired-end approach for the MiSeq platform and v3 chemistry as recommended by the manufacturer (Illumina).

2.2.4. Direct sequencing of beaver gut metagenomes

In eukaryotes, DNA methylation is performed to alter the gene expression (Chan et al., 2005; Law and Jacobsen, 2010). This characteristic was exploited to selectively remove the host and plant DNA contamination from the metagenomic DNA pool and enrich intact microbial DNA therein. The host DNA removal was done using the NEBNext Microbiome DNA Enrichment kit, targeting the CpG-methylated host and plant DNAs (New England Biolabs GmbH, Frankfurt, Germany). The removal of the eukaryotic DNA was conducted through the binding of methylated DNA to the MBD2-Fc protein, bound to the magnetic beads. Thus, the bead fraction containing host DNA is removed by using magnetic attraction.

The enriched beaver gut metagenome was sequenced using an Illumina HiSeq 2500 instrument (Illumina, San Diego, USA). Library preparations were done using the Nextera DNA kit and the Nextera Index kit following instructions of the manufacturer (Illumina). Index PCR was performed using 20 µl of fragmented metagenome DNA, 5 µl of both Index 1 (i7) and Index 2 (i5), 15 µl of Nextera PCR Mix, and 5 µl of PCR Primer Cocktail. The thermal cycling program

was carried out as follows: 72 °C for 3 min, 98 °C for 30 s, 5 cycles of 10 s at 98 °C, 30 s at 63 °C and 3 min at 72 °C. The metagenome DNA libraries were sequenced by using rapid run approach for the HiSeq 2500 platform and the v2 chemistry as recommended by the manufacturer (Illumina).

2.2.5. 16S rRNA bacterial community structure and diversity analysis

CASAVA data analysis software (Illumina) was used for demultiplexing and clipping of sequence adapters from raw sequences. Before removing sequences with an average quality score below 20 and unresolved bases with *split_libraries_fastq.py* from QIIME 1.9.1 (Caporaso et al., 2010), paired-end sequences were merged using PEAR v0.9.11 with default parameters (Zhang et al., 2014). Default settings of cutadapt 1.18 (Martin, 2011) were used for removal of non-clipped reverse and forward primer sequences. Generation of amplicon sequence variants (ASVs) (Callahan et al., 2017), chimera check, clustering, and creating abundance table were performed using VSEARCH v2.10.4 (Rognes et al., 2016). This included sorting by sequence length, size-filtering to ≥ 300 bp, and dereplication. Dereplicated ASVs were denoised using UNOISE3 with default settings, as well as chimera de novo removal with UCHIME. In addition, reference-based chimera removal was performed against the SILVA SSU v132 database (Quast et al., 2013). ASVs were clustered at 97% identity to generate operational taxonomic units (OTUs). Quality-filtered reads were mapped to OTUs to create OTU abundance tables. With *parallel_assign_taxonomy_blast.py* taxonomic classification of the OTU sequences against the SILVA database was done. *Filter_otu_table.py* was used for removal of chloroplasts, unclassified OTUs, and extrinsic domain OTUs. Finally, the lowest number of sequences by random subsampling (13,600 reads per sample) was used for sample comparison at the same surveying effort. Statistical test of alpha diversity (observed OTUs and phylogenetic diversity) from entire gut compartments and non-metric multidimensional scaling (NMDS) plots of the cecum and colon of the three beaver samples were calculated with the ampvis2 package in R (Andersen et al., 2018; R Core Team, 2018).

2.2.6. Comparison of herbivorous gut bacterial communities

The 16S rRNA gene datasets used for comparison to that of the Eurasian beaver included bovine, giant and red panda, termite, North American beaver, and human (Table 2. 2). Each 16S rRNA gene dataset was generated using different methods and approaches. In order to reduce bias when comparing these datasets, all datasets were preprocessed in a similar way to achieve comparable datasets and quality of 16S rRNA gene sequences. Datasets for which

sequence quality scores were available (pandas, termites, Eurasian beaver, North American beavers) were subjected to quality-filtering using *split_libraries.py* script from QIIME with default settings and minimal Q scores of 20. For bovine and human gut samples, according to the information of the authors, reads below 200 bp were excluded from subsequent analysis (Huttenhower et al., 2012; Jami et al., 2013). For the comparison with other 16S rRNA gene datasets obtained from cecum, rumen and fecal samples, we used only our beaver datasets from cecum and colon.

Table 2. 2 Dataset of 16S rRNA gene amplicon libraries used in this study.

Study	No. of samples	Sample type	16S Primers	Sequencing platform	Acc. number[g] & Reference
Bovine	21[a]	Rumen content	V2 – V3	Roche 454	4514864.3 - 868.3 (MG-RAST) (Jami et al., 2013)
Eurasian beaver	23[b]	Gut content	V3 – V4	Illumina MiSeq	PRJNA427255 (this study)
Human	10[c]	Feces	V3 – V5	Roche 454	SRS016152, 016437, 021664, 023914, 042290, 042703, 052196, 055137, 064276, 065665 (Huttenhower et al., 2012)
North American beaver (Grun)	4[d]	Gut content	V1 – V3	Illumina MiSeq	SRP069012, 069014 (Gruninger et al., 2016)
North American beaver (Wong)	3	Feces	V5 – V8	Roche 454	SRR2905007 (Wong et al., 2016)
Panda	11[e]	Feces	V1 – V3	Roche 454	SRR1766294 (Li et al., 2015)
Termite	19[f]	Gut content	V3 – V4	Roche 454	SAMN02228083 – 101 (Dietrich et al., 2014)

[a] Samples grouped to 5 age group: 1 day (n=3), 3 days (n=3), 2 months (n=5), 6 months (n=5), 2 years (n=5);
[b] gut compartment samples grouped to 3 individuals: male juvenile (n=6), male subadult (n=8), female subadult (n=9);
[c] samples grouped to 2 sex group: male (n=4), female (n=6);
[d] samples grouped to 2 sex group: male (n=2), female (n=2);
e samples grouped to 2 panda type: giant panda (n=5), red panda (n=6);
[f] samples grouped to 2 class: lower termite (n=8), higher termite (n=11);
[g] unless noted, all sample sequence was obtained from GenBank (Benson, 2003).

Open-reference OTU picking (*pick_open_reference_otus.py*) from QIIME was used to cluster the 16S rRNA genes of all studies. Open-reference OTU picking was performed with the non-redundant SILVA 132 SSU reference database at 97% sequence identity. The relative abundances at genus level calculated by QIIME *summarize_taxa.py* were used to perform

multivariate analysis using Bray-Curtis dissimilarities. The community structure and NMDS plot were calculated with ampvis2 package in R Studio (Andersen et al., 2018). ANOSIM from vegan package in R Studio (Dixon, 2003) was performed to measure the similarity of bacterial communities across all samples.

2.2.7. Metagenome reads quality filtering and assembly

CASAVA data analysis software (Illumina) was used for demultiplexing and clipping of sequence adapters from raw sequences. Raw metagenome reads were quality-filtered and screened for host DNA sequences using KneadData (v0.6.1). Reads with poor quality score (PHRED <20) as well as reads with short sequences (<50 bp) were removed. Reads without matching pair from either forward or reverse reads are separated and concatenated as single reads. Subsequently, potential contaminating host reads were removed by using the mouse genome (mouse C57BL) obtained from the KneadData websit, as reference. Read qualities were measured using FastQC. The high-quality reads (forwards, reverse and single reads) were then assembled using SPAdes with *--meta* as option for metagenome reads (Nurk et al., 2017). The assemblies of beaver gut metagenomes were evaluated using metaQUAST (Mikheenko et al., 2016).

2.2.8. Metagenome-derived microbial diversity of castor fiber gut

The contigs output from the previous assembly were screened against nr database of NCBI (downloaded 9 April 2018) using blastx within DIAMOND v0.9.21 (Buchfink et al., 2015; Wheeler et al., 2007). The resulted file (*.DAA) were transformed to MEGAN 6 (CE v6.11.1) format using *daa2rma* tools (Huson et al., 2016). To obtain comprehensive taxonomic information, the MEGAN mapping file of the NCBI taxonomy was used for protein-based taxonomic analysis (*prot_acc2tax-Mar2018X1*). Through the lowest common ancestor algorithm in Megan 6, taxonomic binning was performed by assigning reads to nodes in the NCBI taxonomy. The taxonomic rank was exported and heatmap was calculated at family level using Ampvis2 (Andersen et al., 2018).

2.2.9. Metagenome annotation and functional analysis

To assess metabolic activity from beaver gut, the assembled reads were annotated with prokka v1.13 (Seemann, 2014). The translated amino acid sequences were used as queries for protein family identification using Pfam database and the prokka software in order to assign a putative function (Finn et al., 2016). The functional assignment of the metagenome reads was done with

MEGAN 6 using mapping data of SEED (*acc2seed-May2015XX*) and eggNOG (*acc2eggnog-Oct2016X*). Further functional assignment was performed by annotating the predicted proteins against the carbohydrate active enzyme (CAZy) database and the dbCAN annotation server to identify the CAZymes family present in the beaver gut microbiome (Yin et al., 2012). The annotated ORFs were than mapped back to the contigs and taxonomic assignment was performed to determine which phyla are responsible for the lignocellulosic material breakdown. Resulted data were visualized with Alluvial package (R Core Team, 2018).

2.2.10. Cloning of a novel cellulase gene derived from the gut metagenome and its expression in *E.coli*

Based from the prokka annotation of genes associated to cellulase, a putative cellulase (endoglucanase) gene with the lowest sequence identity to the known cellulase in genbank was selected. The endoglucanase, designated as BC33, was amplified by semi-nested PCR from the metagenomic DNA of the lower colon of the sub-adult male beaver. The amplification was done using the specific primers for_BC33a and rev_BC33a. The PCR mix contained 31 µL ddH$_2$0, 10 µL 5x GC buffer, 1 µL dNTP, 1 µL for_BC33a, 1 µL rev_BC33a, 1 µL MgCl$_2$, 1,5 µL DMSO, and 0.5 µL Phusion DNA polymerase. Thermal cycling was carried out as follows: initial denaturation for 1 min at 98°C, followed by 25 cycles of 45 s at 98°C, 45 s at 68°C, 30 s at 72°C and final elongation for 5 min at 72°C. The resulting amplicon was analyzed on a 1 % agarose gel, and the band with correct size (993 bp) was purified using the PCR & Gel Purification kit (Macherey-Nagel, Düren, Germany). The amplicon sequence was verified by Sanger sequencing to confirm (Microsynth Seqlab, Göttingen, Germany). The amplicon was used as template for the second PCR using specific primers for_BC33b and rev_BC33b. This primer pair allowed directional cloning into pET101/D using the pET101/D directional TOPO expression kit (Invitrogen, Karlsruhe, Germany). The thermal cycling for the second amplification was carried out as follows: initial denaturation for 1 min at 98°C, followed by 25 cycles of 45 s at 98°C, 45 s at 67°C, 30 s at 72°C and final elongation for 5 min at 72°C. The resulting amplicon with a size of 994 bp was purified using NucleoSpin PCR Clean-up kit and cloned into pET 101/D, the resulting plasmid was designated as pBC33. The *E. coli* strains TOP10 and BL21 (DE3) were used as hosts for the cloning and for production of the putative cellulase BC33, respectively. The pBC33 which successfully cloned to TOP10 *E. coli* was used to transform BL21 *E. coli* to produce the endoglucanase BC33. The recombinant BL21 *E. coli* strains were grown in Lysogenic Broth (LB) medium supplemented with 100 µg/µl Ampicillin at 37°C under shaking at 180 rpm. To induce the cellulase BC33 expression, 0.35 mM (final

concentration) of isopropyl-β-ᴅ-thiogalactopyranoside (IPTG) was added to cultures at an OD600 of ~0.6. Subsequently, the cultures were incubated for 6 hours and then centrifuged at 8,000 and 4 °C for 30 min. The pellets were stored at -20°C until further analysis. The BL21 *E. coli* BL21 harboring cloning vector without insert was used as negative control for enzyme assays.

2.2.11. Preparation and purification of cellulase BC33

The cellulase BC33 expressed in the BL21 system was prepared for purification using French Press. The BL21 pellet was dissolved in 1x LEW buffer from Protino Ni-TED 2000 kit (Macherey-Nagel, Düren, Germany). Cells disruption through French Press was carried out twice at 6894757,23 pascals (Thermo Fisher Scientific, Massachusetts, USA). To remove cell debris, the lysate was filtered through two nylon membrane, pore size of 10 and 2.5 µm respectively. The resulting crude extract was purified with nickel column Protino Ni-TED 2000. The column purification was performed as recommended by the manufacturer (Macherey-Nagel), with an additional washing step using 1x LEW containing one mM imidazole. All purification steps were performed at 4°C. The cellulase concentration was measured from all fractions (crude extract, flow through, wash, and elution) using the Bradford assay (Bradford, 1976). The purity of the resulting protein BC33 was analyzed with SDS-PAGE.

2.2.12. BC33 activity assays

The activity of BC33 was determined by measuring the release of D-glucose equivalents from carboxymethyl cellulose (CMC low viscosity, Sigma-Aldrich) as substrate using the 3,5-dinitrosalicylic acid (DNS) method adapted from Lone *et al.* (2012). The standard reaction mixture contained 20 µL of BC33 enzyme and 2% (w/v) CMC in 80 µL of acetate buffer (0.2 M acetic acid, 0.2 M $C_2H_3O_2NA$, pH 5.0), was incubated at 50 °C for 1 h. The reaction was terminated by the addition of 120 µL of DNS and incubation at 98 °C for 5 min. The reaction mixture was cooled down on ice for 30 s. After addition of 800 µL dH_2O, the absorbance was measured at 540 nm.

The optimal temperature and pH for BC33 activity was measured using 2% CMC as substrate under the specified assay conditions. The optimum temperature was determined by incubation in 50 mM acetate buffer pH 5 at different temperatures ranging from 10 to 90 °C for 1 h. The commercial cellulase from *Trichoderma reesei* ATCC 26921 (dissolved in dH_2O, Sigma-Aldrich) was used as reference. The applied concentration for both cellulases was 0.004 µg/µL. The optimal pH for the activity of recombinant BC33 was determined by incubation at 60 °C

for 1 h in the following overlapping buffer systems of (each 50 mM): citrate-phosphate buffer (pH 2.6 – 6), phosphate buffer (pH 6 – 8), and Tris/HCl buffer (pH 8 – 9) (Gomori, 1955). One unit of cellulase activity was defined as the amount of enzyme required to release 1 nM of reduced sugar per min.

The substrate specificity of cellulase BC33 was analyzed with 1% CMC, 1 % Barley glucan (Megazyme, Bry, Ireland), and 1% lichenan (Sigma-Aldrich). To confirm whether cellulase BC33 is able to degrade xylan, 1% xylan from birch wood (Sigma-Aldrich) was used in the assay. The assay was performed in 50 mM citrate-phosphate buffer (pH 4) at 60 $^{\circ}$C for 1 h. All enzyme activity assays were performed in triplicate.

The kinetic parameters K_m and V_{max} were determined in 50 mM citrate-phosphate buffer (pH 4.0) containing 0.05 – 0.8 µg/µL Barley glucan at 60 $^{\circ}$C for 1 hour. The K_m and V_{max} values were calculated according to the Michaelis-Menten method.

2.2.13. BC33 phylogenetic analysis and structure prediction

The BC33 protein sequence was analyzed for its domains using the Conserved Domain Search Service (Marchler-Bauer et al., 2011). Based on the known domain (GH5), evolutionary analysis of the BC33 protein was performed against characterized bacterial GH5-containing proteins (282 protein sequences) downloaded from CAZy database (http://www.cazy.org/, February 2019). The alignments and construction of the phylogenetic tree were performed using MEGA v7 (Kumar et al., 2016). The alignment in MEGA was performed with Clustal using default parameters. The alignment result was then used to construct the phylogenetic tree using the Neighbor-Joining method, employing 500 bootstrap replicates and using the number of differences approach to compute the evolutionary distances. The resulted phylogenetic tree was analyzed with Booster web server (http://booster.c3bi.pasteur.fr) to calculate the branch bootstrap support (Lemoine et al., 2018). The tree visualization was calculated using iTOL server (Letunic and Bork, 2016). The BC33 protein structure was predicted using I-TASSER server based on the available PDB database (Roy et al., 2012; Yang and Zhang, 2015; Zhang, 2009). The I-TASSER is a tool for protein structure and function prediction, which will report the biological function of the protein, e.g. the ligand binding sites, the associated ligand, the enzyme commission number, as well as the homologous Gene Ontology

3. RESULTS AND DISCUSSION

3.1. Diversity and composition of the Eurasian beaver gut bacterial community

3.1.1. Diversity of intestinal bacterial communities

The Eurasian beaver gut bacterial communities in different parts of the gastrointestinal tract were characterized by analysis of the 16S rRNA gene amplicons. A total of 2,599,870 high-quality paired-end reads with an average read length of 450 bp were obtained. We identified a total of 277 unique OTUs at 97% genetic identity (species level) across the entire dataset, which comprised 23 samples. In general, the main fraction of the beaver gut bacterial community was covered by the surveying effort indicated by the saturation of rarefaction curves (Figure 3. 1). The bacterial community of the subadult beaver stomach compartments was more diverse than that in cecum and colon (Figure 3. 2B). This result is explained considering that the stomach is the entry point of plant material and the associated diverse microbes into the digestive system. In contrast to the subadult beaver, the stomach of the male juvenile beaver has the lowest diversity compared to its cecum and colon. Considering the age of the male juvenile, the bacterial community continues to develop in its gut system and stabilizes when the beaver reaches adulthood (Jami et al., 2013; Rodríguez et al., 2015). The diversity of the bacterial community in the cecum and colon of the three beavers varied. The communities in the cecum and colon of the male subadult beaver were more diverse than the communities in the subadult female and male juvenile beaver (Figure 3. 2B) .
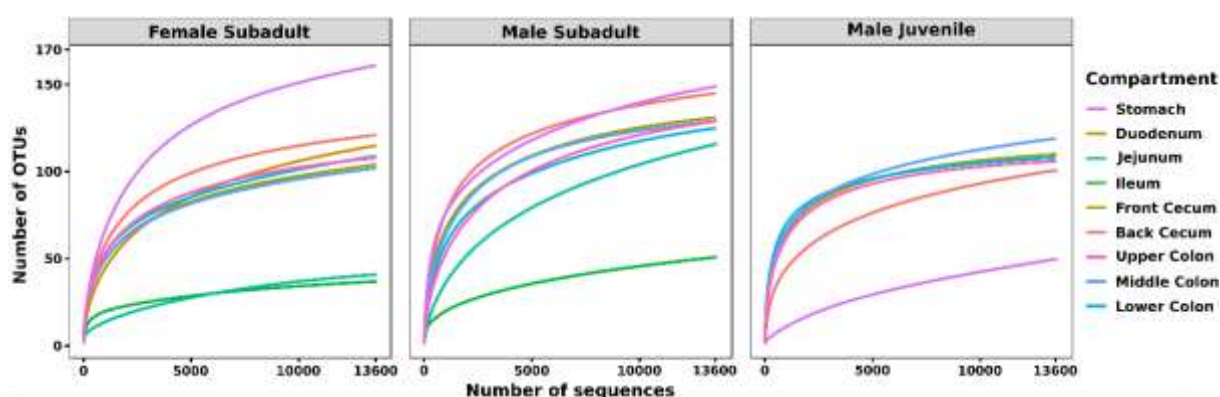

Figure 3. 1 Rarefaction curves from the gut systems of the three analyzed beavers. All samples were randomly subsampled to the least abundant reads (13600 reads).

The similarity in bacterial community composition of the cecum and colon was analyzed via non-metric multidimensional scaling (NMDS, Figure 3. 3). Based on the NMDS, differences in the bacterial communities were associated with the different compartments and individual

beavers. The comparison of the cecum and colon compartment based on the Bray-Curtis distance showed that the bacterial community of cecum and colon between the three beavers, as well as the bacterial community between cecum and colon of male beavers differed in both structure and abundance of OTUs. Similar results were obtained for the bacterial communities in the cecum and colon of North American beavers (Gruninger et al., 2016).
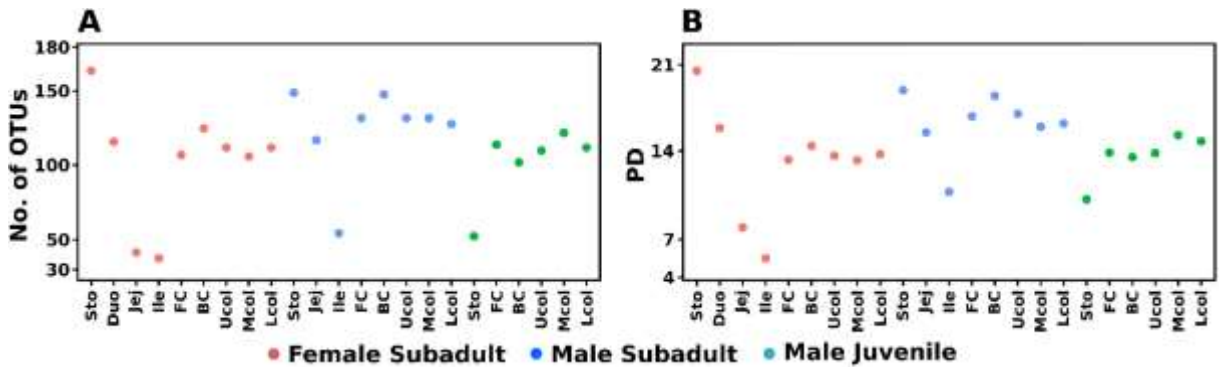

Figure 3. 2 Diversity estimates of the bacterial community composition along the beaver gut samples, observed OTUs (A) and phylogenetic diversity (B). The phylogenetic tree was midpoint-rooted using phangorn R package (Schliep, 2011) before alpha diversity calculation in ampvis2.
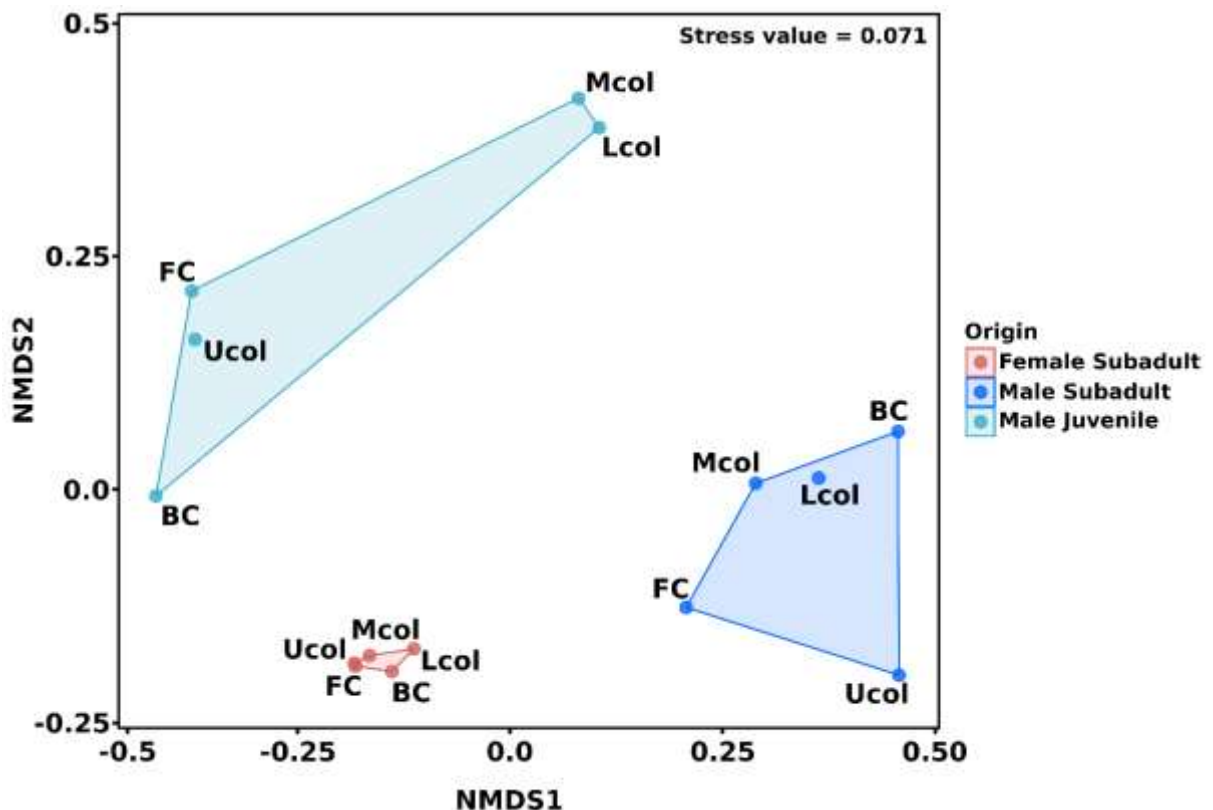
Figure 3. 3 Non-metric multidimensional scaling (NMDS) of cecum and colon bacterial communities in the male and female Eurasian beaver. The ordination was calculatd based on Bray-Curtis distance measure. The 15 gut compartment samples of cecum and colon were grouped according to beaver individuals before distance measure calculation.

3.1.2. The Eurasian beaver gut bacterial community is dominated by Firmicutes and Actinobacteria

In the entire dataset, members of 8 bacterial phyla were detected. Most of the classified sequences belonged to Firmicutes (41.2%), Actinobacteria (23.6%), and Proteobacteria (Alpha- and Gamma-, total of 12.6%). The other phyla were the Verrucomicrobia (9.5%), Fusobacteria (7.1%), Bacteroidetes (5.5%), and Tenericutes (0.1%) (Figure 3. 4A). Compared to the typical mammalian gut bacterial communities that primarily comprise Firmicutes and Bacteroidetes (Ley et al., 2008), the Eurasian beaver gut system is dominated by Firmicutes and Actinobacteria.

Although the Firmicutes was the dominant phylum of the Eurasian beaver gut system, its relative abundance varied along the different gut compartments. In the gut of female subadult beaver, the relative abundance of Firmicutes in the stomach and small intestine (Duo, Jej, Ile) was 75.1% and more than 90%, respectively, whereas in the cecum and colon it was less than 30%. Within the cecum and colon system, we identified members of different Firmicutes families, which were known to possess lignocellulolytic activity, e.g. Clostridiaceae 1, Ruminococcaceae, and Lachnospiraceae (Figure 3. 4B) (Flint et al., 2012; Lee R. Lynd, Paul J. Weimer, Willem H. van Zyl, 2002). Some members of these families such as *Clostridium* and *Ruminococcus* form cellulosomes. Members of *Clostridium* including *Cl. cellulolyticum*, *Cl. cellulovorans*, *Cl. josui*, *Cl. papyrosolvens*, and *Cl. thermocellum* are probably the best studied with regard to cellulose breakdown by cellulosome complexes (Bayer et al., 1985; Blouzard et al., 2007; Doi et al., 1994; Kakiuchi et al., 1998; Nölling et al., 2001; Pohlschröder et al., 1995). Within the genus *Ruminococcus*, *R. Albus,* and *R. flavefaciens* are known as cellulosome producers in the bovine rumen (Ding et al., 2001; Lamed et al., 1987). Their cellulolytic activity in the rumen comprises also degradation of recalcitrant lignocellulose (Flint et al., 2008). Clostridiaceae were abundant in the stomach (25.9%) and cecum (15%) of male subadult beaver (Figure 3. 4B).

In contrast to the male subadult beaver, high abundance of Clostridiaceae was observed in the female subadult beaver in the stomach (66.9%) and small intestine (5.9 − 77.8%), and in the male juvenile beaver in lower relative abundance throughout the colon compartment (<5%). Compared to Clostridiaceae, Ruminococcaceae were abundant in the cecum and colon compartment of the three beavers, with the female subadult showing the highest relative abundance (13.5 − 18.8%), followed by male subadult beaver (6 − 16.9%), and male juvenile beaver (2.9 − 14.6%). The family Lachnospiraceae, which was present throughout the cecum

27

and colon of the three beavers, showed the highest relative abundance in the colon of the male juvenile beaver (1.6 – 20.7%). Among the characterized members of Lachnospiraceae, cellulolytic activity is known for *Cellulosilyticum ruminicola*, isolated from the rumen content of a yak and *C. lentocellum* (formerly *Clostridium lentocellum*) isolated from river sediment (Cai and Dong, 2010; Miller et al., 2011). *C. ruminicola* was present in the small intestine of female subadult beaver. The presence of Ruminococcaceae and Lachnospiraceae in the cecum and colon of the male juvenile beaver (Figure 3. 4B) indicates that members of these families colonize the beaver gut from early age on.
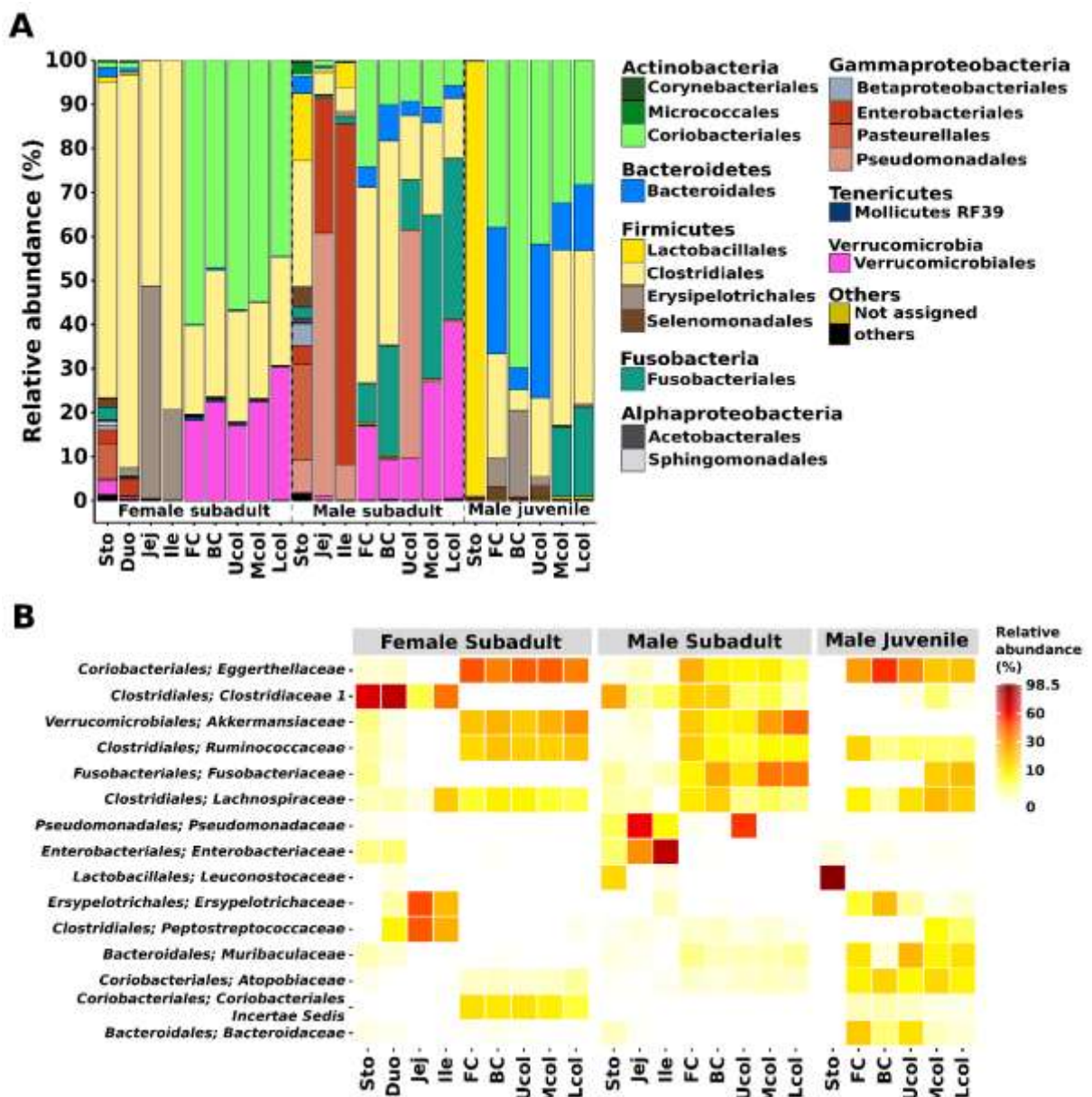
Figure 3. 4 Bacterial community composition in the gut of Eurasian beaver. (**A**) Relative abundance of bacterial communities from male subadult, female subadult, and male juvenile beaver at order level. The figure represents the relative abundance of OTUs at 97% identity. (**B**) The top 15 most abundant bacterial from (**A**) were aggregated at family level.

28

The second most abundant phylum, Actinobacteria, was found mainly in the cecum and colon compartments in all three beavers (Figure 3. 4A). In these compartments, the *Eggerthellaceae* was the most abundant family, accounting for over 60% of Actinobacteria. Dominant genera among the Eggerthellaceae, were *Enterorhabdus* (76.5 – 96.8%), *Adlercreutzia* (<3%), *Gordonibacter* (<2.5%), and uncultured bacteria (0.5 – 20.7%). To date, *Enterorhabdus* consists of *E. mucosicola*, *E. caecimuris*, and *E. muris* isolated from mice intestine (Clavel et al., 2009, 2010; Lagkouvardos et al., 2016). These species have so far only been found enriched in the mice and hamster gut (Clavel et al., 2014), and has not been reported from other gut systems, suggesting that the members of this genus are host-specific. *Enterorhabdus* comprises aerotolerant bacteria that grow under anoxic conditions and utilize a variety of amino acid derivatives as energy source (Clavel et al., 2009). Although, *Enterorhabdus* present in high relative abundance in the Eurasian beaver gut, this did not apply for the North American beaver in which *Enterorhabdus* was not detected during our analysis. These results suggest that *Enterorhabdus* species are specifically associated with the Eurasian beaver.

The phylum Proteobacteria was detected in both subadult beaver gut systems, especially in the male subadult beaver small intestine (jejunum and ileum, Figure 3. 4A), Based on the 16S rRNA gene analysis of jejunum and ileum of the male subadult beaver, high relative abundances of *Pseudomonas* sp. (59.6 and 8%, respectively) and *Escherichia-Shigella* sp. (30.7 and 77.5%, respectively) were recorded. The genus *Pseudomonas* is ubiquitous in soil and aquatic environments and can be isolated worldwide in all types of environments (Peix et al., 2009). It includes species capable of using various organic and inorganic compounds, including cellulose. Of these species, *P. fluorescens var. cellulosa*, *P. nitroreducens*, and the newly isolated *P. coleopterorum* sp. *nov* were reported to exhibit cellulolytic activity (Hazlewood et al., 1992; Huang et al., 2012; Menéndez et al., 2015; Yamane et al., 1971; Yamane and Suzuki, 1988). In general, *Pseudomonas* tend to live in aerobic, mesophilic and neutral pH environments (Moore et al., 2006). The presence of *Pseudomonas* in the subadult males small intestine may indicate the presence of *P. aeruginosa* as this species is able to grow anaerobically in the presence of nitrate, nitrite, and nitrous oxide (Wu et al., 2005). Thus, *Pseudomonas* could also play a vital role for nitrogen metabolism in the gut. Like the termites, the Eurasian beavers are dependent on microbial nitrogen fixation to fulfill their N demand, as their diet is low in nitrogen. The presence of nitrogen-fixing bacteria in the small intestine of subadult beavers such as *Pseudomonas* could be the result of beavers coprophagy to utilize these biological nitrogen sources (Vecherskii et al., 2009). In addition. members of *Escherichia and Shigella,* which are closely related and share many common characteristic (Devanga Ragupathi et al., 2018), were

present. They are known to contaminate water bodies (Jun et al., 2016; Probert et al., 2017). The presence of *Escherichia-Shigella* in the beaver gut might originated from a contaminated water source in the beaver habitat.

The Verrucomicrobia phylum was detected in the cecum and colon of both subadult beavers (9 to 17.1 % and 30.2 to 40.3%, respectively), but *was* not detected in male juvenile beaver samples (Figure 3. 4A). This suggests that the colonization of Verrucomicrobia has not started in the juvenile beaver. Based on the study of *Akkermansia* sp., the colonization of Verrucomicrobia will increase and reach its maximum abundance when the host becomes adult (Derrien et al., 2008). Verrucomicrobia were also reported to inhabit the human gut (Flint et al., 2012), the bovine rumen (Li et al., 2012b), and the North American beaver gut (Gruninger et al., 2016). The Verrucomicrobia in the Eurasian beaver gut consisted solely of the genus *Akkermansia*. To date, *Akkermansia* comprises two species, *A. muciniphila* and *A. glycaniphila* isolated from human feces sample (Derrien et al., 2004) and *a* python feces sample (Ouwerkerk et al., 2016), respectively. The *Akkermansia muciniphila* plays an essential role in the human gut by supporting glucose homeostasis, blood lipid formation and body composition after calorie restrictions to maintain a healthy metabolic status (Dao et al., 2016).

The presence of Fusobacteria in the gut of both male subadult and male juvenile beavers was recorded (Figure 3. 4A). The Fusobacteria are abundant in both cecum (9 to 25%) and colon (11.3 to 37.2%) of the male subadult beaver, and the colon (15.8 – 20.1%) of the male juvenile beaver (Figure 3. 4A). The Fusobacteria were represented by members of a single genus, *Fusobacterium*. The presence of *Fusobacterium* in the gut system is often linked to pathogenicity, e.g. *F. necrophorum*, causes Lemierre's disease (Riordan, 2007) and *F. nucleatum* is enriched in patients with chronic gut inflammation (Allen-Vercoe et al., 2011). whereas *F. varium* provides butyrate and acetate that are important to maintain a healthy colon (Potrykus et al., 2007).

Members of Bacteroidetes were present in low abundance in subadult beavers (<10%) but in high relative abundance in cecum and colon of the juvenile male beaver (5% - 34.9%) (Figure 3. 4A). Certain species of gut-associated Bacteroidetes are known to possess a large number of genes that encode carbohydrate active enzymes (Flint et al., 2012). Among Bacteroidetes, the Muribaculaceae (part of Bacteroidales *S24-7*) and Bacteroidaceae are the most abundant families in the male juvenile cecum (0.6 – 11.3% and 3.8 – 16.3%, respectively) and colon (9 – 21.2 and 0.7 – 11.9%, respectively). Earlier studies of 57 unique animal species with respect to Bacteroidales *S24-7* showed that 96% of these animals harboring Bacteroidales *S24-7* were

herbivores, or omnivores mainly consuming herbivorous food (Ormerod et al., 2016). As with the Bacteroidales S24-7 family, Bacteroidaceae was also enriched in the guts of Korean adolescents, whose food consisted mainly of plant-based and fermented foods (Jang et al., 2017). Members of the dominant genus *Bacteroides* are among the most common in the human gut bacterial community (Ramakrishna, 2013). The presence of cellulolytic *Bacteroides* e.g., *B. cellulosilyticus,* was reported previously from human gut (Robert et al., 2007). The recently isolated *B. luti* from methanogenic sludge also possess cellulolytic activity (Hatamoto et al., 2014). Further analysis of the bacterial community in the Eurasian beaver might explain the shift of Bacteroidetes abundance from juvenile to subadult beavers.

Based on the bacterial community composition, it is concluded that the degradation of lignocellulosic plant material occurred within the cecum and colon compartments, as putative cellulolytic members of Ruminococcaceae*,* and Lachnospiraceae, are present almost exclusively in cecum and colon of the beavers. This also suggests that they colonized the beaver cecum and colon from an early age on in order to prepare juvenile beavers to adapt to lignocellulosic diet after weaning.

3.1.3. Potential functional capabilities of the beaver gut microbiome

We used Tax4Fun to predict the functional profile from our 16S rRNA gene datasets (Aßhauer et al., 2015). The Tax4Fun prediction has been shown to provide a good correlation of functional profile with the metagenome profile derived from direct sequencing (Aßhauer et al., 2015). In addition, Tax4Fun prediction returned a high coverage of mammalian gut bacterial community, a functional profile could be predicted for about 95% of the OTUs (Aßhauer et al., 2015). So far, functional profiles have been predicted from various 16S rRNA gene datasets derived from different environments using Tax4Fun (Berkelmann et al., 2018; Hasegawa et al., 2017; Kaiser et al., 2016; Wemheuer et al., 2017; Xu et al., 2018).

We were able to assign functional profiles for 74.85% of our OTUs (Appendix 1). Based on the ability of the beaver microbiome to degrade hardwood, we focused our analysis on metabolic functions associated with cellulolytic activity derived from the KEGG pathway database. We recorded enriched abundance of endoglucanase and beta-glucosidase genes in the cecum and colon of the three beavers (Figure 3. 5). Since plant material also consists of storage polysaccharides, e.g. starch, the predicted genes associated with starch modification (starch phosphorylase and 4-alpha-glucanotransferase) were also abundant in the cecum and colon (Figure 3. 5). All genes encoding the above-mentioned enzymes were also present in the

31

stomach and small intestine but in lower relative abundance than in the cecum and colon. This could be the result of coprophagy in the beavers, since beavers require nutrients from microbial metabolism. The presence of endoglucanase, beta-glucosidase, starch phosphorylase and 4-alpha-glucanotransferase, together with the abundance of cellulolytic bacteria from Firmicutes phylum in the cecum and colon compartment, indicates that the breakdown of lignocellulosic plant material takes place in these gut compartments.
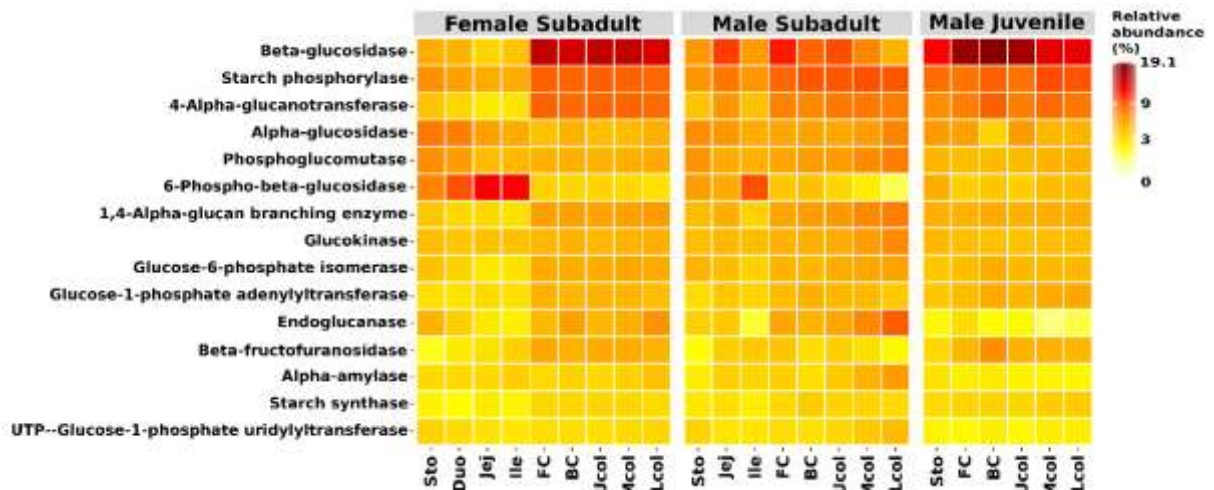

Figure 3. 5 Top 15 genes function prediction under starch and sucrose metabolism (KEGG) of 16S beaver gut datasets with Tax4Fun. The relative abundance of predicted genes was shown, including those encoding cellulases (endoglucanase and beta-glucosidase).

3.1.4. Eurasian beaver gut microbiome in comparison with North American beavers, herbivorous animals, and humans

The comparative analyses of the gut bacterial communities, between Eurasian (Eu) beaver and two North American (NA) beavers derived from different studies, NA_Grun and NA_Wong (Gruninger et al., 2016; Wong et al., 2016), were conducted. Additionally, the gut bacterial communities of other herbivorous animals, such as bovine (Jami et al., 2013), giant and red panda (Li et al., 2015), and termite (Dietrich et al., 2014) were included (Table 2. 2). The dataset of human gut bacterial communities (Huttenhower et al., 2012) were also included to compare the beaver gut bacterial communities with those of omnivores. To create comparable bacterial community 16S rRNA gene datasets of the different studies, the datasets were divided into two groups: group A comprised the datasets from Eu beaver cecum, NA_Grun beaver cecum, bovine rumen, and termite hindgut as this compartment has been known to harbor abundant lignocellulolytic bacterial community; and group B comprised the datasets from fecal samples of giant and red pandas, NA_Wong beaver, and human, in addition to the colon of Eu and NA_Grun beaver, which is most equal to fecal samples. The different grouping of rumen/cecum and fecal

samples was intended, as the bacterial community derived from fecal material cannot be directly regarded as representative for the gut bacterial community (Ingala et al., 2018; Kohl et al., 2014).

The comparison of group A (Figure 3. 6A) shows that the cecum bacterial community of the Eu beavers differed from that of the North American beavers (NA$_{Grun}$) (R=0.881, P<0.05), the bovine (R=0.908, P<0.05), the higher termites (R=0.976, P<0.05) and the lower termites (R=0.998, P<0.05). This is mainly due to differences in the composition of the dominant phyla, in which Firmicutes and Bacteroidetes were generally dominant. The NA$_{Grun}$ cecum and bovine rumen, and partly also the gut of higher termites such as *A. trestus*, *Macrotermes* sp, and *Odontotermes* sp, were dominated by Firmicutes and Bacteroidetes as previously reported (Flint et al., 2012) (Appendix 2). This was distinct from the Eu beaver cecum in which members of the Actinobacteria are more abundant than that of Bacteroidetes. The Firmicutes families, Ruminococcaceae and Lachnospiraceae, were present in both Eu and NA beaver cecums as well as in bovine rumen, and the majority of termite guts (Appendix 3), suggesting that it is common family associated with herbivorous gut microbiomes (Ferrario et al., 2017; Jami and Mizrahi, 2012; Singh et al., 2017). The Streptococcaceae was the major family of the rumen bacterial community of one-day old bovine (Appendix 3). In accordance with previous studies on mammals this family is among the earliest colonizers of the infant gut, transmitted from the mother by breastfeeding (Rodríguez et al., 2015). For bovine aged two months to two years, members of Bacteroidetes, the Prevotellaceae, dominate in the rumen (59%) and play an important role in the intake of carbohydrates (Kim et al., 2011; Ramakrishna, 2013). The presence of Prevotellaceae has also been detected in the cecum of Eu and NA beaver, but in relative abundances of less than 10%. The abundance of the dominant Actinobacteria family, Eggerthellaceae, was observed only within the Eu beaver cecum (Appendix 3), which supports our earlier hypothesis that the member of this family, *Enterorhabdus* might be specific for the Eu beaver gut. The phylum Spirochaetae was also abundant in several termite guts. Spirochaetae account for up to 50% of all prokaryotes present in some termites, and are involved in the degradation of cellulose and hemicellulose (Droge et al., 2006; Dubinina et al., 2015; Lilburn et al., 1999; Sravanthi et al., 2015).
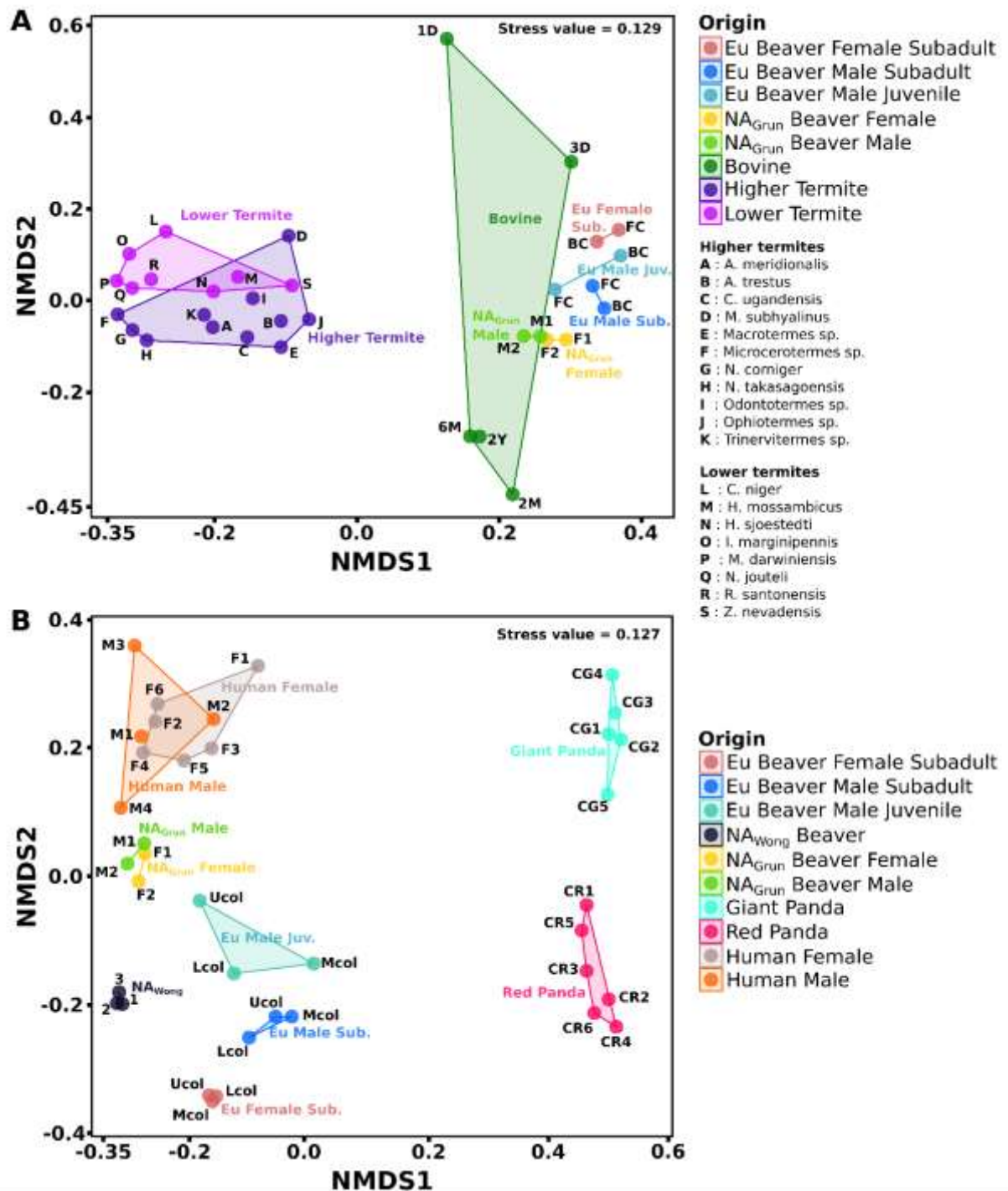
Figure 3. 6 NMDS of gut bacterial communities of herbivore animals. The combined datasets separated based on sample source: group A consist of cecum, rumen, and termite gut samples (**A**); group B consist of fecal samples (**B**). The addition of human data was to compare with non-herbivore gut bacterial community. The distance between the ordination of one sample to another indicates the dissimilarity of the gut bacterial community between those samples.

The feces-derived bacterial communities (group B, Figure 3. 6B) showed that the bacterial community of the Eu beaver differed from that of the giant panda (R=1, P<0.05), red panda (R=0.986, P<0.05), NA$_{Grun}$ beaver (R=0.889, P<0.05), NA$_{Wong}$ beaver (R=0.801, P<0.05), and human (R=0.974, P<0.05). The Firmicutes were generally present in high abundance in all

samples, particularly in the giant and red pandas, in which Firmicutes account in average more than 90% of the total bacterial community (Appendix 4). The exceptions were NA$_{Wong}$ (less than 10%) and human male 3 (2.1%). Within the two colon samples of Eu and NA$_{Grun}$ beavers, Ruminococcaceae and Lachnospiraceae were the common Firmicutes families (Appendix 5). These families are also generally present in human fecal samples. As already mentioned, the members of Ruminococcaceae and Lachnospiraceae facilitates breakdown of plant cell walls in the host gut and are common in herbivore guts. Since humans are omnivores, a dietary change to a fiber-rich diet increase the abundance of Ruminococcaceae and Lachnospiraceae (Muegge et al., 2011; Singh et al., 2017).

The giant and red panda fecal samples were dominated by Streptococcaceae and Clostridiaceae 1 family, respectively. Members of the Streptococcaceae were not present in the Eu beaver colon samples at all, but members of Clostridiaceae 1 were present in low abundance (below 7%). The giant pandas are known to possess digestive tracts like carnivores, with a simple stomach, degenerated cecum, rapid transit time and gut microbiome similar to those of bears (Li et al., 2015; Xue et al., 2015). On the other hand, the bacterial communities of the red panda gut did not differ in diversity compared to the giant panda (Appendix 5).

Verrucomicrobia was discovered in human and all beavers (except juvenile Eu beaver, Appendix 4). The low abundance of Verrucomicrobia in both North American beaver and human gut might be related to a poor health status of the host, as the abundance of *Akkermansia* sp. has been inversely correlated to several disease states (Geerlings et al., 2018). Members of the Fusobacteria were abundant in the colon samples Eu and NA$_{Wong}$ beaver, and also in human male 2 (Appendix 4). Since the Fusobacteria have been reported to be abundant in the gut of marine carnivorous mammals, their presence in the Eu beaver gut systems needs to be further investigated in order to ascertain their role in the gut system. Interestingly, the fecal samples of NA$_{Wong}$ beaver and the upper colon of the male subadult Eu beaver showed a high abundance of *Pseudomonadaceae* (Appendix 5). Although some species of *Pseudomonas* are able to degrade cellulose (Hazlewood et al., 1992; Huang et al., 2012; Lejeune et al., 1986), their presence have not been observed in other beaver gut systems.

Based on the comparative analysis, the Eurasian beaver gut indicated the bacterial community structure that is differed to the herbivorous animals and humans gut bacterial communities included in this study. However, the Eurasian beaver gut bacterial community was closer to the North American beaver. This result suggests that both the Eurasian and North American beaver

shares common bacterial species inhabiting their gut, which make them distinct to the other herbivorous gut bacterial communities.

3.2. Metagenomic analysis of the Eurasian beaver gut bacterial community reveals novel cellulase

3.2.1. Bacterial community structure derived from metagenome sequences

The direct metagenome sequencing of the female subadult back cecum (FSBC) and lower colon (FSLC) beaver, and male subadult back cecum (MSBC) and lower colon (MSLC) beaver resulted in a total of 6,200,436 high quality reads. The *de novo* assembly of metagenomic reads using metaSpades resulted in a total of 101,060 contigs with the largest contig of 198,219 bp. The assembly statistics are shown in Table 3. 1.

Table 3. 1 Assembly statistic of beaver gut microbiome.

	FSBC	FSLC	MSBC	MSLC
Total reads after filtering*	1323903	2338236	1621314	916983
Total contigs	9545	30197	28885	32433
Total contigs ($\geq$ 1 kb)	4385	10433	8574	10968
Largest contig (bp)	153114	166334	198219	115974
N50	3284	1818	2201	1744
Predicted ORFs	13461	26578	27527	28770

*: total of both paired-end reads.

The metagenomic community analysis showed that over 99% of reads from each metagenome sample were assigned to *Bacteria*, while *Archaea*, *Eukaryota*, and *Viruses* accounted for less than 1%. This indicated that pre-filtering and enrichment treatment to limit the number of host and plant DNA contaminants was successful. Taxonomic analysis with MEGAN 6 showed that the majority of the bacterial community was assigned to Proteobacteria, Firmicutes, and Actinobacteria. Most of the reads were assigned to the Gammaproteobacteria (58.7% on FSBC, 36.5% on FSLC, 23.6% on MSBC, and 66.8% on MSLC), and the Pseudomonadaceae family therein. Members of this family are highly adapted to various environmental conditions. Other families with high representation in our metagenome reads were the Clostridiaceae (38.7% FSLC, 14.4% MSBC, 9.2% MSLC), the Eggerthellaceae (19.4% FSBC, 10.2% FSLC, 17.5% MSBC, 7.7% MSLC), the Lachnospiraceae (19.4% MSBC), and the Ruminococcaceae (7.4% MSBC, 9.5% MSLC) (Figure 3. 7).
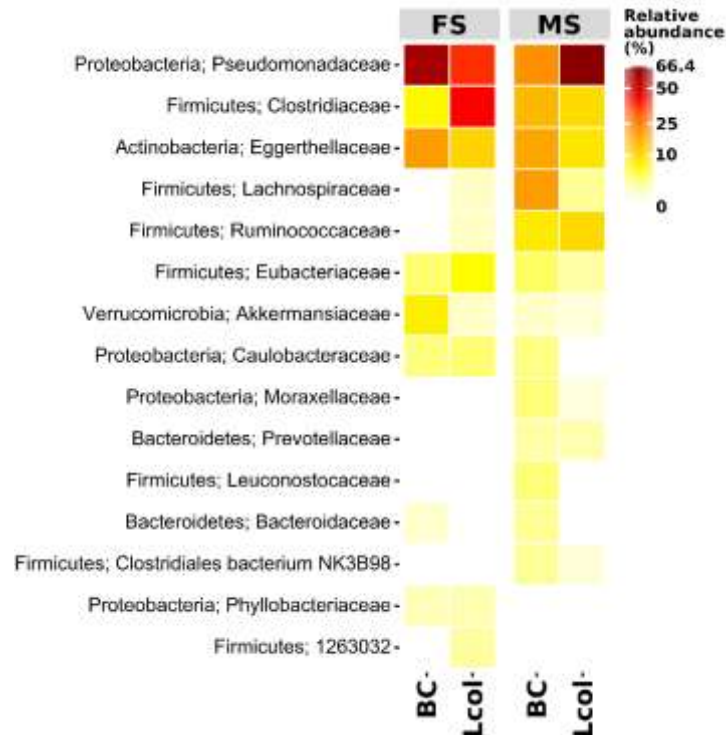
Figure 3. 7 The relative abundance of the top 15 families bacterial community from the back cecum (BC) and lower colon (Lcol) of female subadult (FS) and male subadult (MS) beaver.

Over 99% of the Pseudomonadaceae belonged to the genus *Pseudomonas*. Members of the genus *Pseudomonas* are widely distributed in nature, and less than 4% are culturable so far (Li et al., 2013). The genus *Pseudomonas* is known for its pathogenicity, for example lung infection/respiratory disease caused by *P. aeruginosa,* necrotizing hepatitis and haemolytic activity caused by *P. fluorescens,* urinary tract infections caused by *P. putida*, and several other *Pseudomonas sp.* (Peix et al., 2009). On the other hand, *Pseudomonas* species with nitrogenase activity have been described (Haahtela et al., 1983; Yan et al., 2008), which makes *Pseudomonas* an ideal candidate for nitrogen fixation in the beaver gut. A recent study on the distribution of nitrogenase-like sequences amongst microbial genomes reveal that Proteobacteria were one of the phyla harboring many nitrogenase-like genes (80 proposed species) (Dos Santos et al., 2012). In addition to nitrogenase activity, some *Pseudomonas* species also exhibit cellulolytic properties, e.g. *P. fluorescens* var. *cellulosa* and novel *P. coleopterorum* (Bakare et al., 2005; Menéndez et al., 2015). The high abundance of Proteobacteria has not only been observed in the Eurasian beaver gut, but also in the North American beaver gut (Appendix 4) (Wong et al., 2016). Other families that are particularly helpful in supporting the beaver's digestion include Clostridiaceae and Ruminococcaceae. The genus *Clostridium* of the family Clostridiaceae was present in all samples (5.3% FSBC, 38.7% FSLC, 14.2% MSBC, 8.8% MSLC), while *Ruminococcus* was abundant in the male subadult beaver (4.9% MSBC, 9.5% MSLC). The genera *Clostridium* and *Ruminococcus* are known for

37

their cellulolytic properties, for example *C. thermocellum*, *R. albus* and *R. flavefaciens a*s described in section 3.1.2.

3.2.2. Metabolic potential derived from the metagenome of subadult beavers

To assess the metabolic potential within the beaver's gut microbiome, functional annotations of the assembled metagenome sequences were performed using MEGAN 6, mapped against the SEED and eggNOG database. The male and female beaver generally showed similar metabolic profiles (Figure 3. 8). The genes function annotation of the SEED subsystem showed that metabolic pathway associated with amino acids, carbohydrates, and cofactors/vitamins metabolism were the most abundant, 9.2 – 10.8%, 9.5 – 10.4%, and 9.3 – 9.8% respectively (Figure 3. 8). Similar to the SEED annotation, the eggNOG annotation indicated that genes associated with amino acids metabolism are well represented (10.3 – 12.5%), followed by genes involved in DNA replication, recombination and repair (10.1 – 12.3%), and genes associated with energy production and conversion (8.2 – 10.6%). The genes annotated to the carbohydrate metabolic pathway based on eggNOG annotation were 5.6 to 9.4%. The abundance of carbohydrate-associated genes from both database annotations are shown on Figure 3. 9.
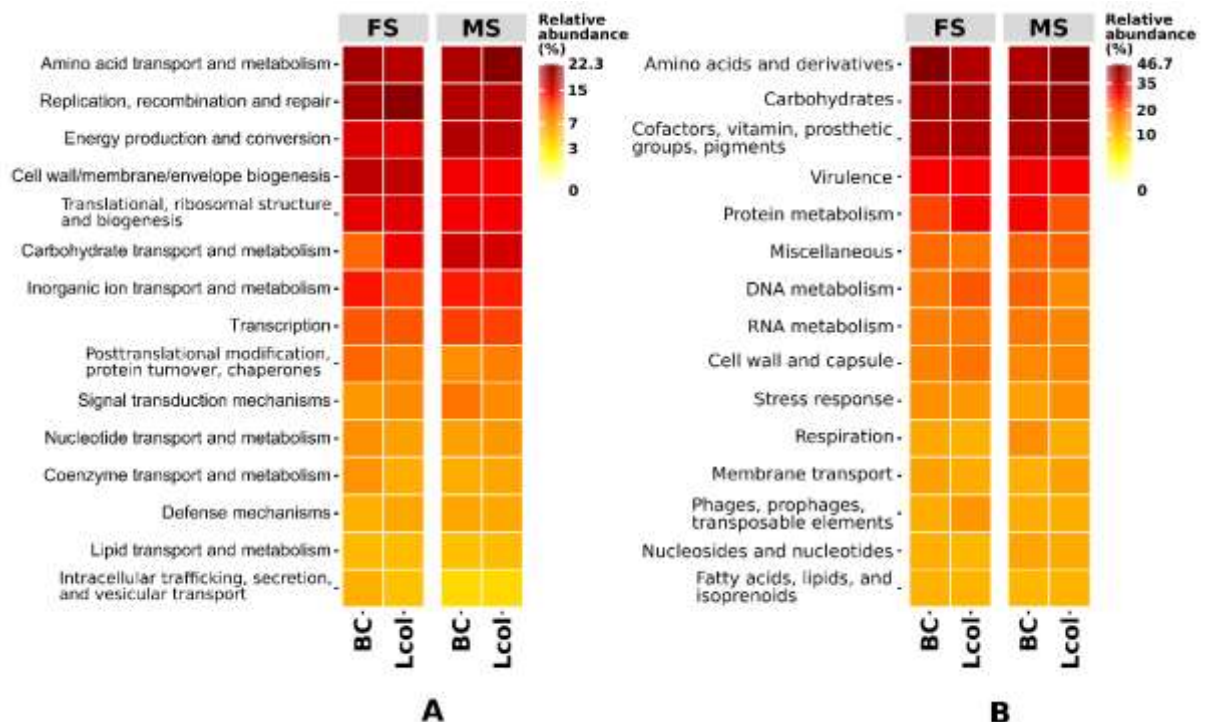
Figure 3. 8 The relative abundance of the top 15 metabolic pathway derived from (A) eggNOG and (B) SEED analysis of beaver gut metagenomes. FS = female subadult beaver, MS = male subadult beaver, BC = back cecum, Lcol = lower colon.
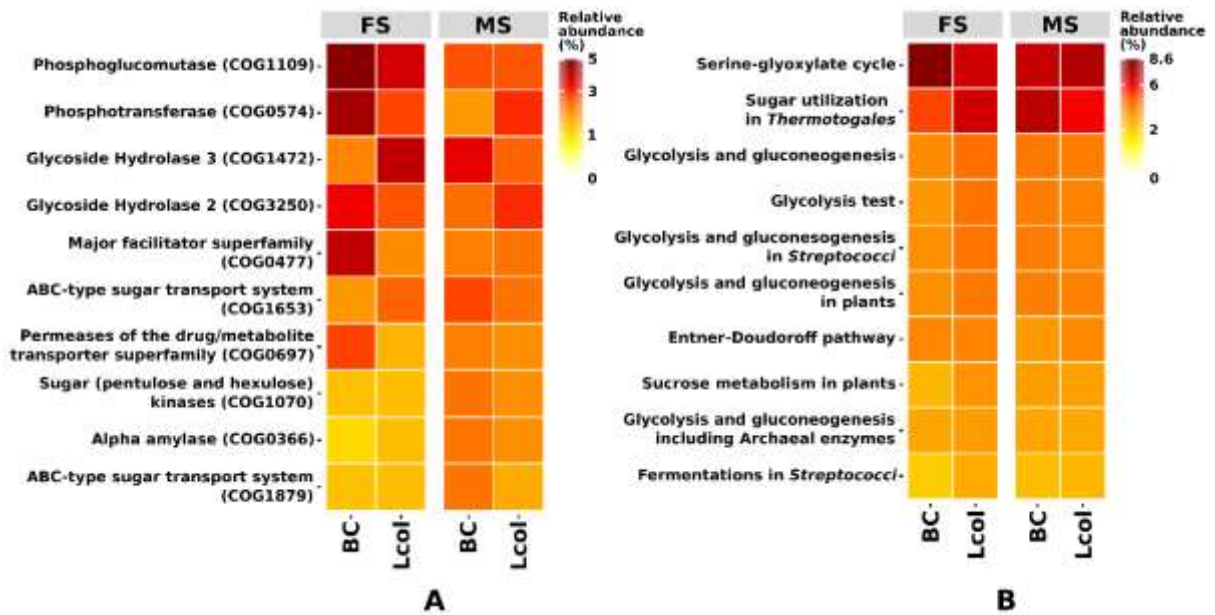
Figure 3. 9 The relative abundance of top 10 genes function associated with carbohydrate metabolic pathway from (A) eggNOG annotation and (B) SEED annotation. FS = female subadult beaver, MS = male subadult beaver, BC = back cecum, Lcol = lower colon.

The gene repertoire of the Eurasian beaver gut microbiome includes many sequences associated with defense mechanism (Figure 3. 10A), with genes associated to the ATP binding cassette (ABC) transporter being the most represented (over 20%). The ABC transporter genes are present in all the domains of life, which is essential for the uptake of various substances such as amino acids and sugars, as well as for the export of waste products, toxins, or cellular component that function outside the plasma membrane (Jiao and Zheng, 2011). In addition to the ABC transporter, genes associated to multidrug-resistance efflux pumps were present (11.6 – 21.5%) (Figure 3. 10A). The multidrug-resistance efflux pumps are essential for the cell survival by actively exporting toxic compounds that are harmful from the cell (Piddock, 2006). Since the beaver diet consists of plants, genes associated with resistance to plant toxins (e.g. secondary metabolite of phenolic compounds) and oxidative stress, were present under stress response. Phenolic compounds in the plants could bind covalently to digestive enzymes and thus inactivate the enzymes (Sánchez-Sánchez and Morquecho-Contreras, 2017). These indicates that the Eurasian beaver gut microbiome could play an important role in resistance to toxic compound, e.g. heavy metal and antibiotics, as well as an agent to detoxify plant secondary metabolite.

Interestingly, the SEED annotation of the Eurasian beaver gut microbiome represented genes associated with virulence, disease, and defense which are relatively abundant (Figure 3. 10B). The genes that are over represented were related to the resistome of *Legionella pneumophila* (over 30%). *L. pneumophila* is a ubiquitous bacterium that colonizes environmental water and

is pathogenic to humans, causing nosocomial and community pneumonia (De Giglio et al., 2015). The anthropogenic impact on environmental water system, e.g. antibiotic contamination from medical or veterinary practices, could promote the development of antibiotic resistance genes of *L. pneumonia* (De Giglio et al., 2015). In addition to the *L. pneumophila* resistome, genes associated with methicillin-resistant *Staphylococcus aureus* (MRSA, 7.4 – 11.8%) were observed. MRSA is a commensal bacteria which is also the main cause of endocarditis, bacteremia, osteomyelitis as well as skin and soft tissue infections (Turner et al., 2019).
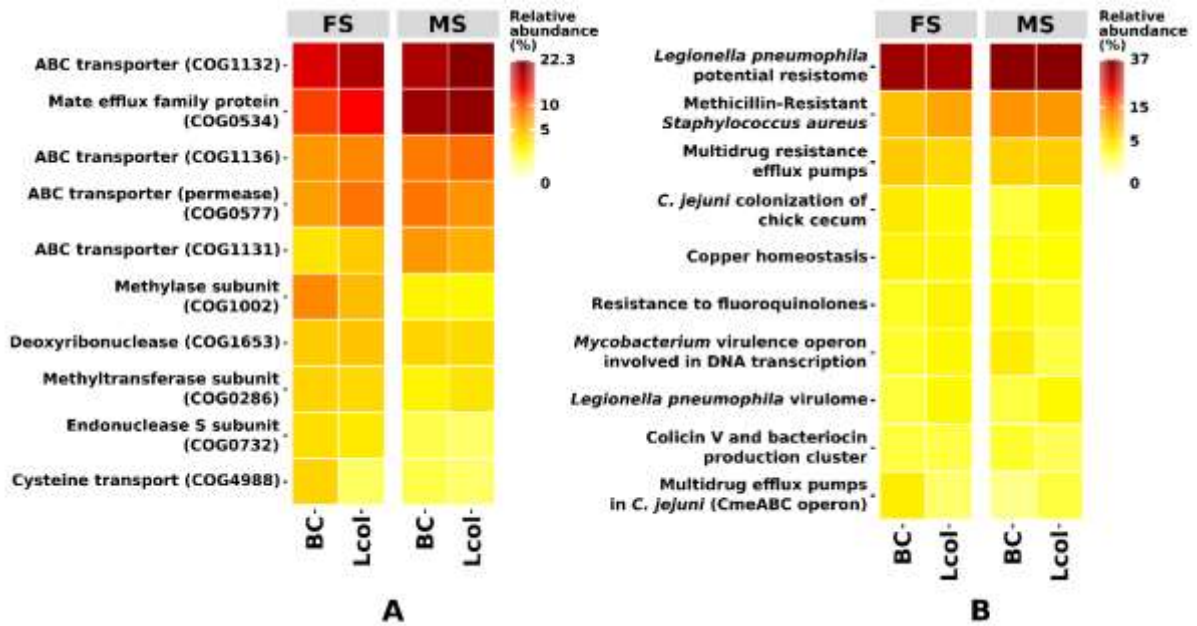
Figure 3. 10 The relative abundance of top 10 genes associated with (A) defense mechanism and (B) virulence, disease, and defense. FS = female subadult beaver, MS = male subadult beaver, BC = back cecum, Lcol = lower colon.

Genes associated with carbohydrate metabolism, in particular numerous glycoside hydrolase families (GHs), have been observed from the eggNOG genes annotation (Figure 3. 11A). These includes GH3 (20 – 31.1%, broad range of enzymes includes celulases and hemicellulases), GH5 (2.5 – 7.5%, cellulases and hemicellulases), and GH31 (11.3 – 15%, mainly α-glucosidases). Several hits to the phosphotransferase system (PTS) were observed within the eggNOG functional assignment. The PTS was originally proposed to catalyze sugar transport and sugar phosphorylation (Kundig et al., 1964). PTS has been extensively studied and is now known as a complex system with various functions within cellular physiology (Saier, 2015). In addition to carbohydrate metabolic functions, genes associated with nitrogen fixation, NifL, were recorded with relative abundance of 11.5 – 19.6% (Figure 3. 11B). The nitrogenase regulator, NifL, in the presence of ammonium and oxygen directly regulates the NifA activity, controlling nitrogenase formation (Dixon et al., 1995; Xie et al., 2006). Microbial nitrogen

fixation is important for beaver survival as described in section 3.1.2. Additionally, the synthesis of various vitamins and cofactors, including the molybdenum cofactor biosynthesis, which is needed for nitrogen fixation were observed.
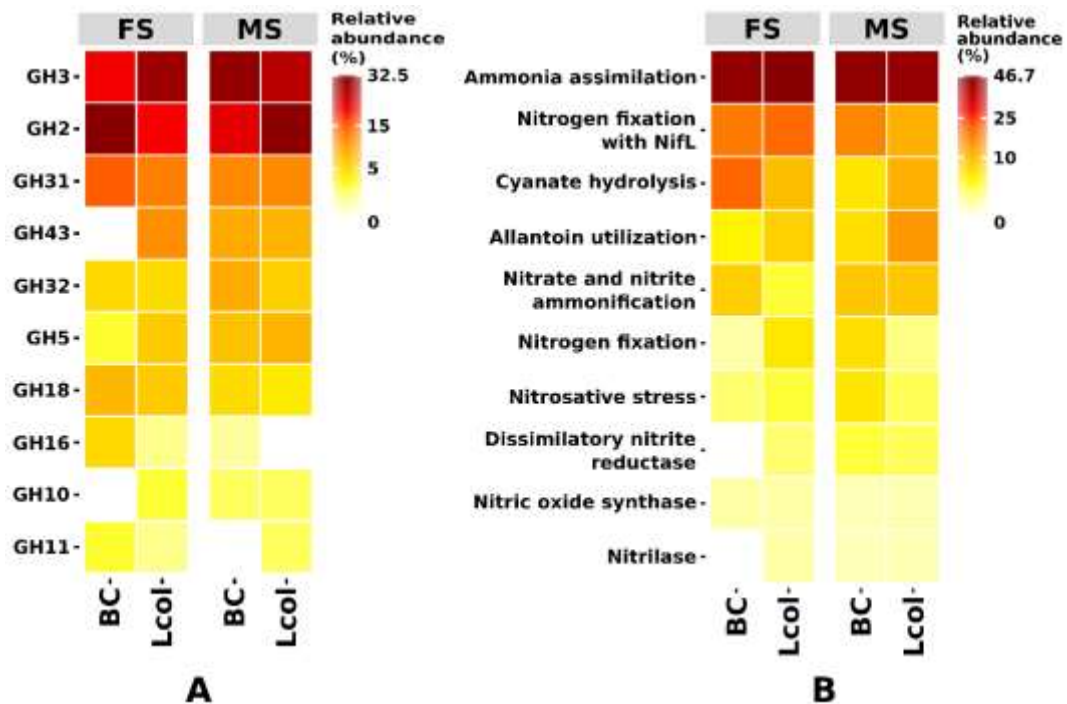
Figure 3. 11 The relative abundance of top 10 genes associated with glycoside hydrolases (GHs, A) and nitrogen metabolism (B). FS = female subadult beaver, MS = male subadult beaver, BC = back cecum, Lcol = lower colon.

3.2.3. The gut microbiome of the beaver as reservoir for genes encoding of carbohydrate-active enzymes (CAZymes)

To identify the proteins associated with plant cell wall breakdown in gut system, functional annotation of the metagenome sequences was performed using the automatic annotation dbCAN web server, which is an HMM-based database for CAZymes. A total of 2,352 CAZymes were identified from the four beaver gut metagenomes (Figure 3. 12). The Kaiju analysis of the putative CAZymes revealed the affiliation to 456 bacterial taxa. Most of the identified CAZymes were associated with Proteobacteria, Firmicutes, and Actinobacteria. The cellulose hydrolysis family, GHs, associated with Proteobacteria (63.8%), Firmicutes (16.4%), and Actinobacteria (15.5%) (Figure 3. 13). Among the CAZymes families from beaver metagenome sequences, the building blocks of cellulosome complexes were observed. Cellulosomes are multi-enzyme complexes which comprise dockerin-harboring enzymes and cohesion-containing structural proteins called scaffoldins (Artzi et al., 2017) (Figure 3. 14A). The scaffoldin subunit is connected to another type of functional domain, the Carbohydrate-Binding Module (CBM). The catalytic domain (GHs, polysaccharide lyases (PLs) and CEs) is

41

bound to the dockerin protein, which will interact with cohesin in the scaffoldin subunit (Fontes and Gilbert, 2010). Finally, the anchoring subunit contains an S-layer homology (SLH) domain which anchors the cellulosome complex at the bacterial surface (Figure 3. 14A) (Doi and Kosugi, 2004).
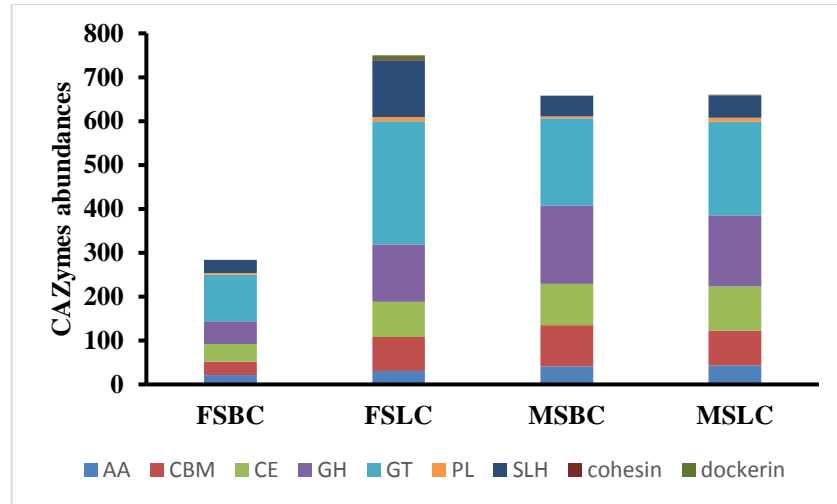

Figure 3. 12 Identified CAZymes from Eurasian beaver metagenomes. FSBC = female subadult beaver cecum, FSLC = female subadult lower colon, MSBC = male subadult beaver cecum, MSLC = male subadult lower colon.
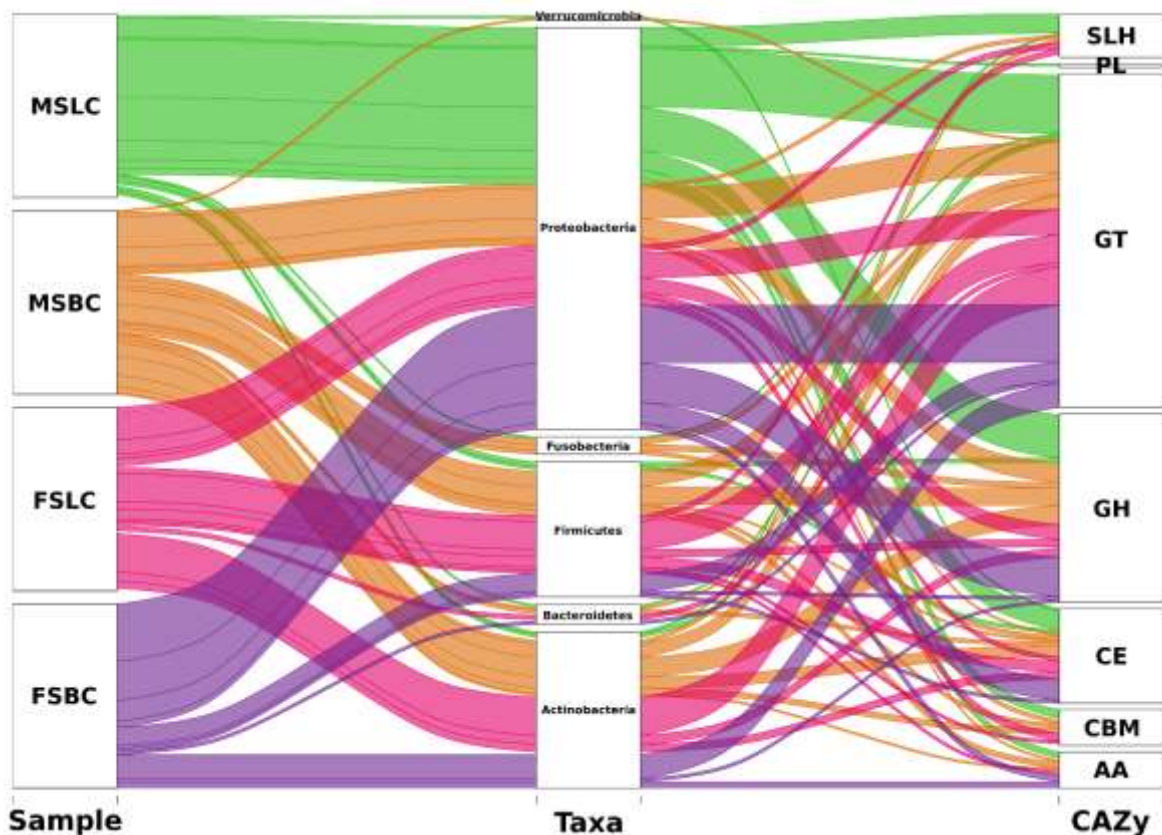
Figure 3. 13 Parallel diagram of CAZymes-associated taxa inferred from metagenome sequences. The rectangle size of Taxa and CAZymes correspond to their abundances. FSBC = female subadult beaver cecum; FSLC = female subadult lower colon; MSBC = male subadult beaver cecum; MSLC = male subadult lower colon; SLH = s-layer homology; PL = polysaccharide lyase; GT = glycosyl

42

transferase; GH = glycoside hydrolase; CE = carbohydrate esterase; CBM = carbohydrate-binding module; AA = auxiliary activity.

The proteins related to lignocellulose breakdown, in particular those associated with cellulosome complexes were observed. As many as 3 putative cohesins and 9 putative dockerins, as well as a type-I-cohesin like and type-I-dockerin like protein were identified from the lower colon compartment of female subadult beaver (Figure 3. 12). The type-I-cohesin contains CBM and binds to the type I dockerin which contains the catalytic enzyme, to form the primary scaffoldin backbone of the cellulosome complex (Figure 3. 14A) (Carvalho et al., 2003). Based on the search result of the Conserved Domain Database, the gene coding for dockerin was confirmed with the presence of Ca binding site, which is essential for the cohesin-dockerin interaction (Figure 3. 14B) (Fontes and Gilbert, 2010). To degrade lignocellulose, the incorporation of catalytic GHs into dockerin within the cellulosome complex is important. Members of the GH families are responsible to hydrolyze the glycosidic linkage in polysaccharides (Garvey et al., 2013). From the identified GH families in the metagenomes, GHs which act on peptidoglycan (GH23 and 73) were the most abundant (Figure 3. 15). The enzymes in these families are lytic transglycosylases and lysozymes that cleave the β-1,4-linkage between *N*-acetylmuramyl and *N*-acetylglucosaminyl residues (Figure 3. 16A & B) (Blackburn and Clarke, 2000; Lipski et al., 2015).
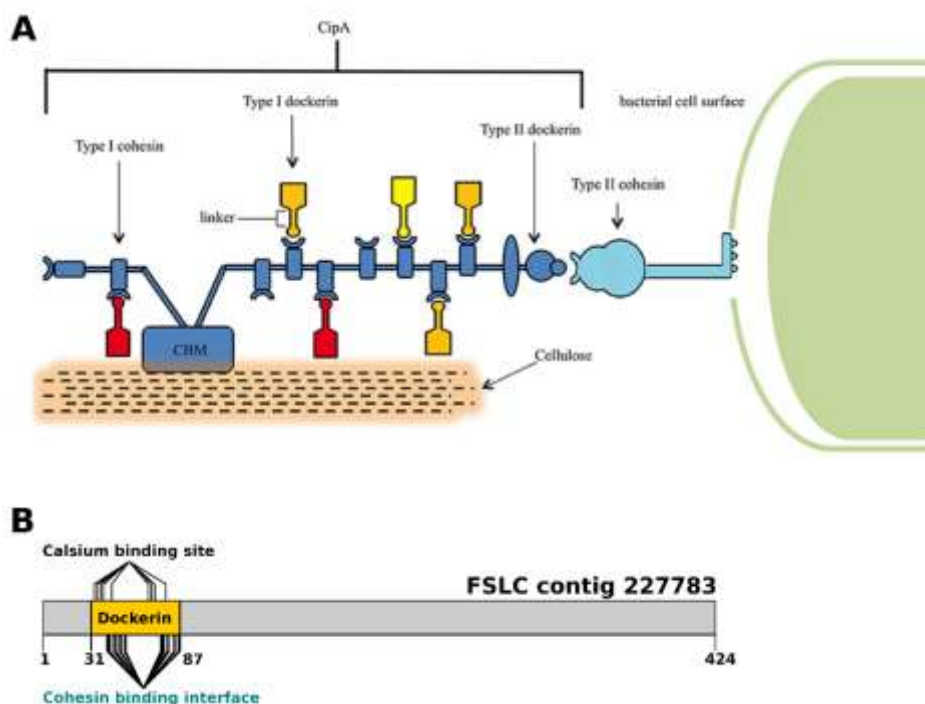
Figure 3. 14 (A) Structure of cellulosome complex from *C. thermocellum*, and (B) the domain of Dockerin with calcium binding site from Eurasian beaver gut metagenome. The central component of cellulosome (CipA, the scaffoldin subunit) is bound to the cell surface via binding of type II cohesin and type II dockerin domains. also contain CBM, to bind cellulose, and type I cohesin that bind type I dockerin containing catalytic unit. Source: Akinosho et al. (2014).
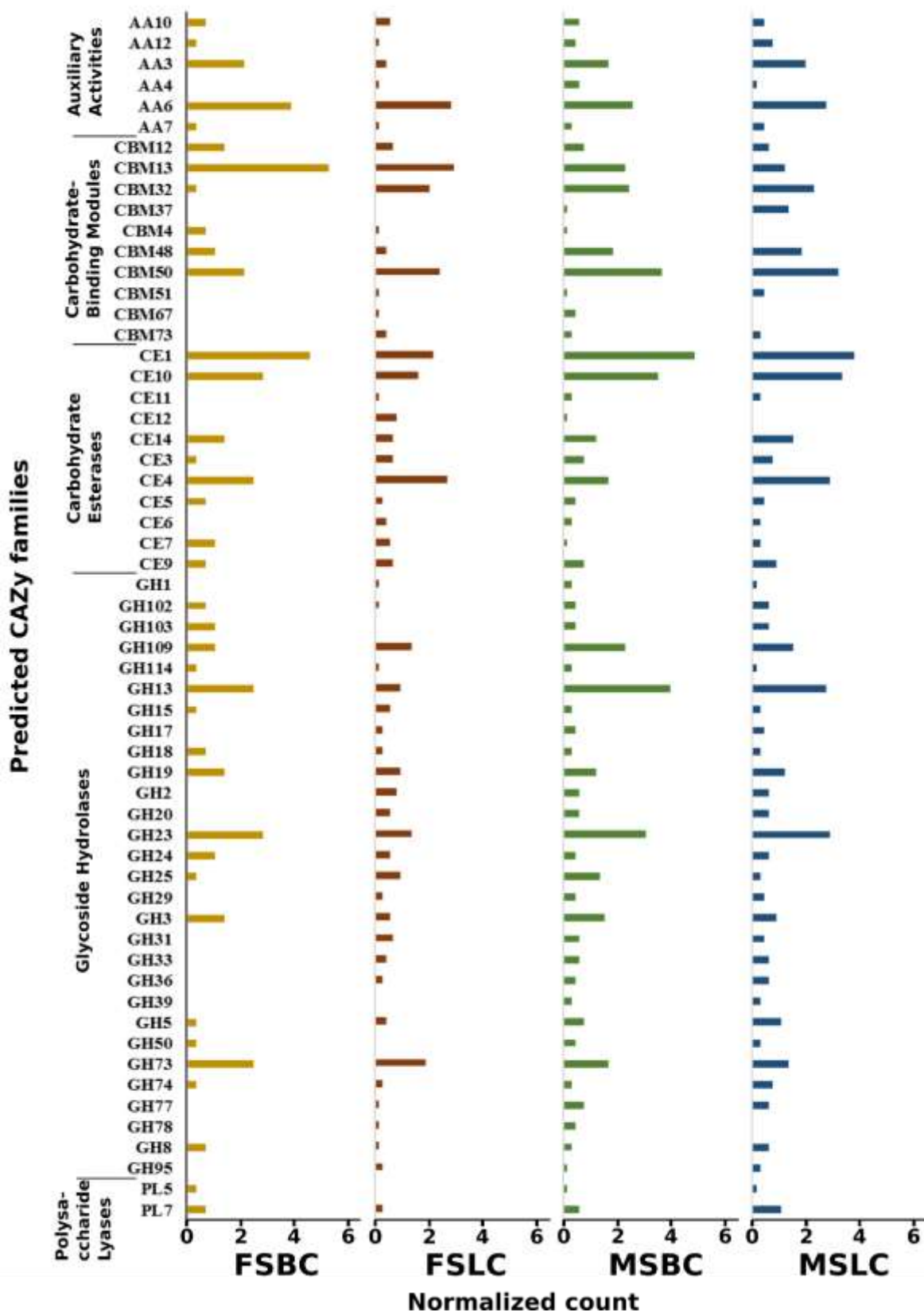
Figure 3. 15 The CAZymes families present in the Eurasian beaver metagenomes. FSBC = female subadult back cecum, FSLC = female subadult lower colon, MSBC = male subadult back cecum, MSLC = male subadult lower colon. The families of glycosyl transferase (GTs), dockerin, cohesin, and SLH were omitted. Count normalization was done through Ampvis2 package in R. For the count values and complete CAZymes families, see Appendix 6.

44
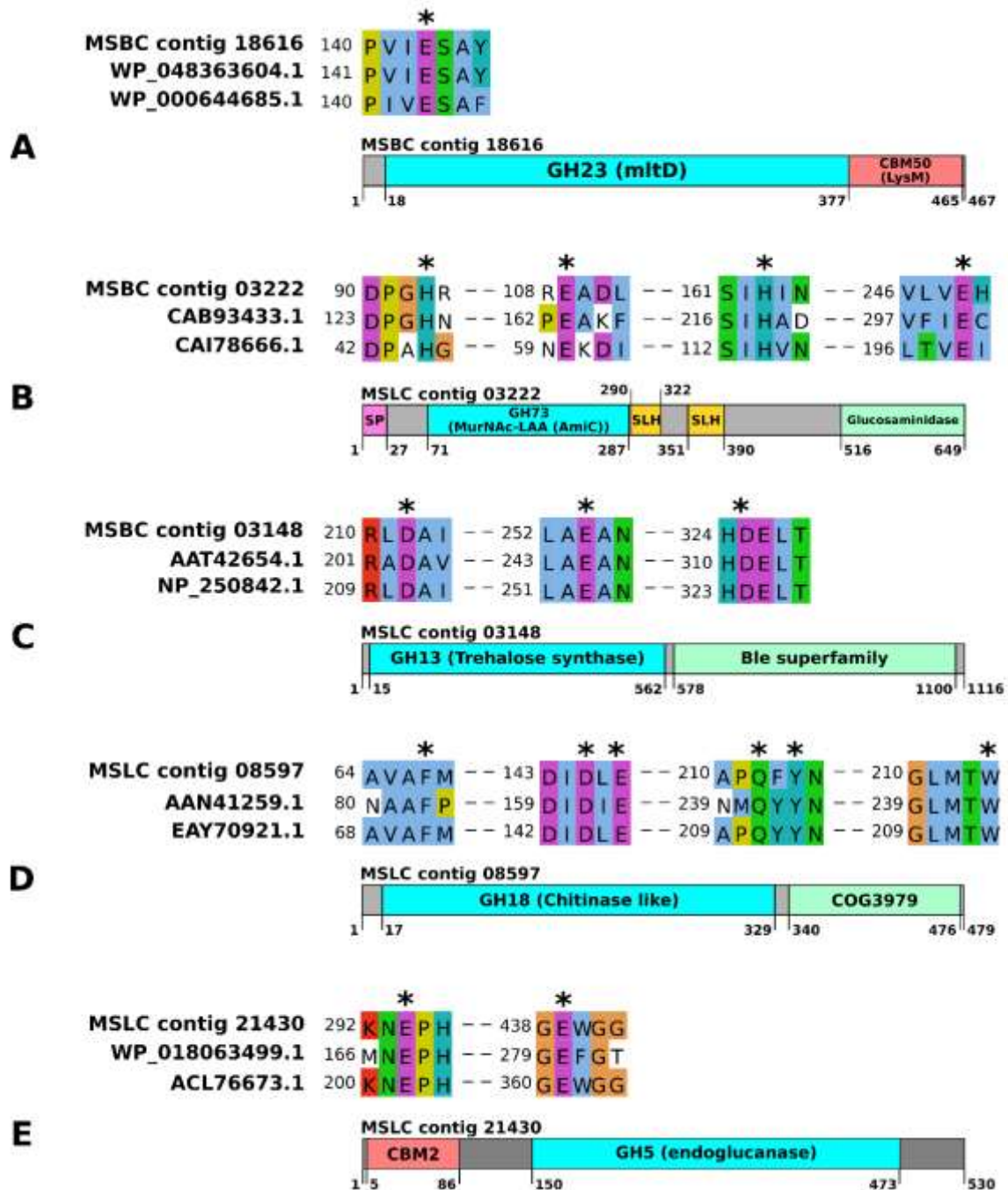
Figure 3. 16 Representative of several identified glycoside hydrolases (GHs) from Eurasian beaver gut metagenomes and the sequences alignment against reference proteins. The catalytic residues were marked with asterisk (*). (**A**) GH23 (membrane-bound lytic transglycosylase D) and CBM50 (lysin motif) (MSBC contig 18616), LysM peptidoglycan binding domain from *Pseudomonas weihenstephanensis* (WP_048363604.1), murein transglycosylase D from Proteobacteria (WP_000644685.1). (**B**) GH73 (N-acetylmuramoyl-L-alanine amidase) (MSBC contig 03222), putative membrane protein from *Streptomyces coelicolor* A3 (CAB93433.1), N-acetylmuramoyl-L-alanine amidase from uncultured delta proteobacterium (CAI78666.1); SP = signal peptide; SLH = S-layer homology (**C**) GH13 (MSBC contig 03148), trehalose synthase from *Picrophilus torridus* DSM 9790 (AAT42654.1), trehalose synthase from *P. aeruginosa* PAO1 (NP_250842.1); Ble superfamily = predicted trehalose synthase. (**D**) GH18 (MSLC contig 08597), chitinase from *Beauveria bassiana* (AAN41259.1), Chitinase from *Burkholderia dolosa* AU0158 (EAY70921.1); COG3979 (predicted Chitodextrinase). (**E**) GH5 and CBM2 (MSLC contig 21430), GH5 protein from *Martelella mediterranea* (WP_018063499.1), GH5 protein from *Ruminiclostridium cellulolyticum* H10 (ACL76673.1).

45

The lytic transglycosylases, also known as peptidoglycan lyases, are essential to modify bacterial cell wall during growth (Vollmer et al., 2008). Although these GH families did not contain cellulases, it was reported that the active site of some cellulases are structural similar to the GH23 enzymes (Figure 3. 16A) (Brás et al., 2011). The GH families acting on the alpha-glycosydic linkage, GH13 and GH109, were also found. The GH13 is the largest GH family acting on alpha-glycosidic linkages in polysaccharides like starch, glycogen and related oligosaccharides (Figure 3. 16C) (Stam et al., 2006; van der Maarel et al., 2002), while GH109 cleaves α-N-acetylgalactosaminyl residues of various glycolipids (Ohta et al., 2000). The GH family members acting on cellulose were found in GH5 and GH3. The GH5 family (Figure 3. 16E), also known as "cellulase family A", has been reported to act on various substrates including cellulose, chitin, mannan, xylan, glucan and lichenin (Aspeborg et al., 2012; Elifantz et al., 2008). The putative genes associated with GH5 family were relatively abundant in the lower colon metagenome of the male subadult beaver (Figure 3. 15, Appendix 6). The GH3 family includes wide range of enzymes that act on cellulosic substances, the modification of glycoconjugates and cell wall polymers (Cournoyer and Faure, 2003).

Cardoso *et al.* (2012) conducted a metagenomic comparative study of the GH families in snail, termite and wallaby showing that GH5 is the most abundant family with respect to cellulose (and hemicellulose) breakdown, although hemicellulose breakdown requires a concerted action of several GH families, e.g. GH2, GH10, GH11, GH16, GH26, GH30, GH31, GH39, GH42, GH43 and GH53. Hemicellulose, is composed of many heterogenous oligo- and polysaccharides, e.g. xylan, mannan, β-(1,3:1,4)-glucan, and xyloglucan (Broeker et al., 2018). Hardwood, which is one of the beavers main food source in winter, is mainly composed of xylan (Pauly and Keegstra, 2008). Metagenome sequences associated with GH families able to breakdown xylan were detected, and included GH3, GH5, GH8, GH28, GH30, and GH53 (Figure 3. 15). In addition to GHs, other CAZymes families such as CEs and auxiliary activities (AAs) also support the degradation of plant polysaccharides. The CEs catalyze de-O- or de-N-acetylation to remove ester substitutions from carbohydrates, which are common in hemicelluloses. The xylan backbone, normally carries substitutions such as acetyl groups linked by ester bonds at the 2-*O* and/or 3-*O* positions, thereby preventing the glycosyl hydrolase from degrading xylans. The CE1 family remove these substitutions and enables xylan degradation (Nakamura et al., 2017). Another CE family, CE4, catalyze the O- or N-deacetylation of xylan and was also reported to act on chitin and peptidoglycans (Nakamura et al., 2017). Both the putatives CE1 and CE4 families were present in relatively high abundance (Figure 3. 15, Appendix 6). Lastly the AA10 family was observed, a family of lytic polysaccharide mono-

oxygenases, capable of oxidizing polysaccharide chains at the surface of the crystalline polymer, and disrupt the structure to enable further enzymatic action and degradation processes (Hemsworth et al., 2014).

In order to degrade insoluble and complex plant polysaccharides, the catalytic modules (e.g. GHs) are often bound to the non-catalytic module of CBM (Fujimoto, 2013). The CBM binds to the insoluble polysaccharides, such as cellulose, and increases catalytic efficiency by delivering the substrate to the adjacent catalytic module (Santos et al., 2012). Sequences belonging to CBM13, 32, 48, and 50 were the most abundant in the metagenomes of Eurasian beavers (Figure 3. 15, Appendix 6). CBM13 and CBM32 were often found bound to hemicellulases. CBM13 shows specificity for xylan (endo-1,4-β-xylanases), while CBM32 is bound to GH5 mannanase (Mizutani et al., 2014). CBM13 is bound to several other GHs, e.g. endo-1,3(4)-β-glucanase, chitinase, α-galactosidase, α-N-acetylgalactosaminidase, α-L-arabinofuranosidase, β-agarase, and β-L-arabinosidase (Fujimoto, 2013). It has been reported that CBM48 is bound to GH13 with specificity for starch (Machovič and Janeček, 2008). The abundance of putative CBM48 was positively correlated to the abundance of putative GH13 in the beaver metagenomes (Figure 3. 15). CBM50, also known as LysM domain, is involved in the binding of N-acetylglucoseamine residues in bacterial peptidoglycans and chitin. CBM50-bound GH23 and GH73 play an important role in cell division by localizing these enzymes to the divisional site (Figure 3. 16A) (Visweswaran et al., 2013). Another study reported that a GH18 chitinase (Figure 3. 16D) containing CBM50 exhibited antifungal activity (Onaga and Taira, 2008). The abundance of CBM50 in the beaver metagenomes was in accordance with the abundance of GH23 and GH73 (Figure 3. 15).

To enable cellulose breakdown by cellulases, the lignin polymers must first be degraded. Lignin is a highly cross-linked polymer of 4-hydroxyphenyl propanoid monomers (monolignols). The aromatic nature and highly branched structure make lignin inert to degradation, thereby limiting the access to cellulose (de Gonzalo et al., 2016). In the past, studies of lignin-degrading organism focused on fungi, especially white-rot fungi, which developed a rich collection of extracellular oxidative enzymes to degrade lignin (Malherbe and Cloete, 2002). Equivalent to fungi, the lignin-degrading peroxidases were also produced from bacteria, called dye-decolorizing peroxidases (DyPs) (van Bloois et al., 2010). So far, a large number of bacterial DyPs have been described (Lambertz et al., 2016). Several *Pseudomonas* sp. are known to produce DyP-type peroxidases, each with a different activity in lignin decomposition (Li et al., 2012a; Rahmanpour and Bugg, 2015; Santos et al., 2014). Enzymes that are also known to be

involved in bacterial lignin degradation are the glutathione S-transferases, laccases, ring cleaving dioxygenases (Allocati et al., 2009; Masai et al., 2003), monooxygenases and phenol oxidases (Perestelo et al., 1989). A partial sequence of DyP-like protein in the lower colon of the male subadult metagenome was recorded. The similarity analysis of this DyP-like protein showed a 100% identity with a DyP-type peroxidase secreted from *Pseudomonas fragi*. Within the CAZy database, the domain of lignin-degrading proteins is preserved under auxiliary activity (AA) family (Lombard et al., 2014). The AA6 family was relatively abundant from the back cecum and lower colon metagenomes of the Eu beaver (Figure 3. 15, Appendix 6). The AA6 is a family of 1,4-benzoquinone reductase, which catalyze the reduction of benzoquinone derived from lignin breakdown (Akileswaran et al., 1999).

3.3. Analysis of beaver cellulase candidate 33 (BC33) – a new GH5 family cellulase

A total of 33 putative endo-β-glucanases (FSBC: 3, FSLC: 10, MSBC: 7, MSLC: 13) were predicted with Prokka from the gut metagenomes, designated as beaver cellulase (BC) 01 to BC33 (Appendix 7). From 33 putative cellulases, only 16 proteins were confirmed to have domain associated to cellulase (endoglucanase) based on the conserved domain analysis and were assigned to the GH families (Appendix 7). The rest of the putative proteins were assigned to the CBM family (one protein), cohesin and SLH (15 proteins), and a non-CAZyme related protein (one protein) (Appendix 7). According to the sequence similarity to the CAZy database, the identified GHs were from the families GH5, GH8, and GH16 (Figure 3. 17). The putative endoglucanases associated with the GH5 family (four proteins) were from the lower colon of the male subadult beaver. The putative GH5 endoglucanases were further subjected to similarity analysis to find homologous sequences in the NCBI database. Of the four putative GH5 endoglucanases, the endoglucanase from MSLC contigs 23174 (BC33, 310 amino acids) has the lowest identity score. BC33 shares 58.98% sequence identity with the GH5 endoglucanases from *Ruminococcus* sp. 1xD21-23 (Accession number RKJ34599.1) (Figure 3. 17). Further BLAST against the UniprotKB database resulted in 60.1% sequence identity with a GH5 cellulase from Ruminococcaceae bacterium (Figure 3. 17). This suggest that the endoglucanase domain in BC33 could be a new representative of GH5 family. Conserved domain analysis on the BC33 protein sequence confirmed the presence of the GH5 protein domain with a specific hit toward pfam00150 (E-value of 2.17e-73).
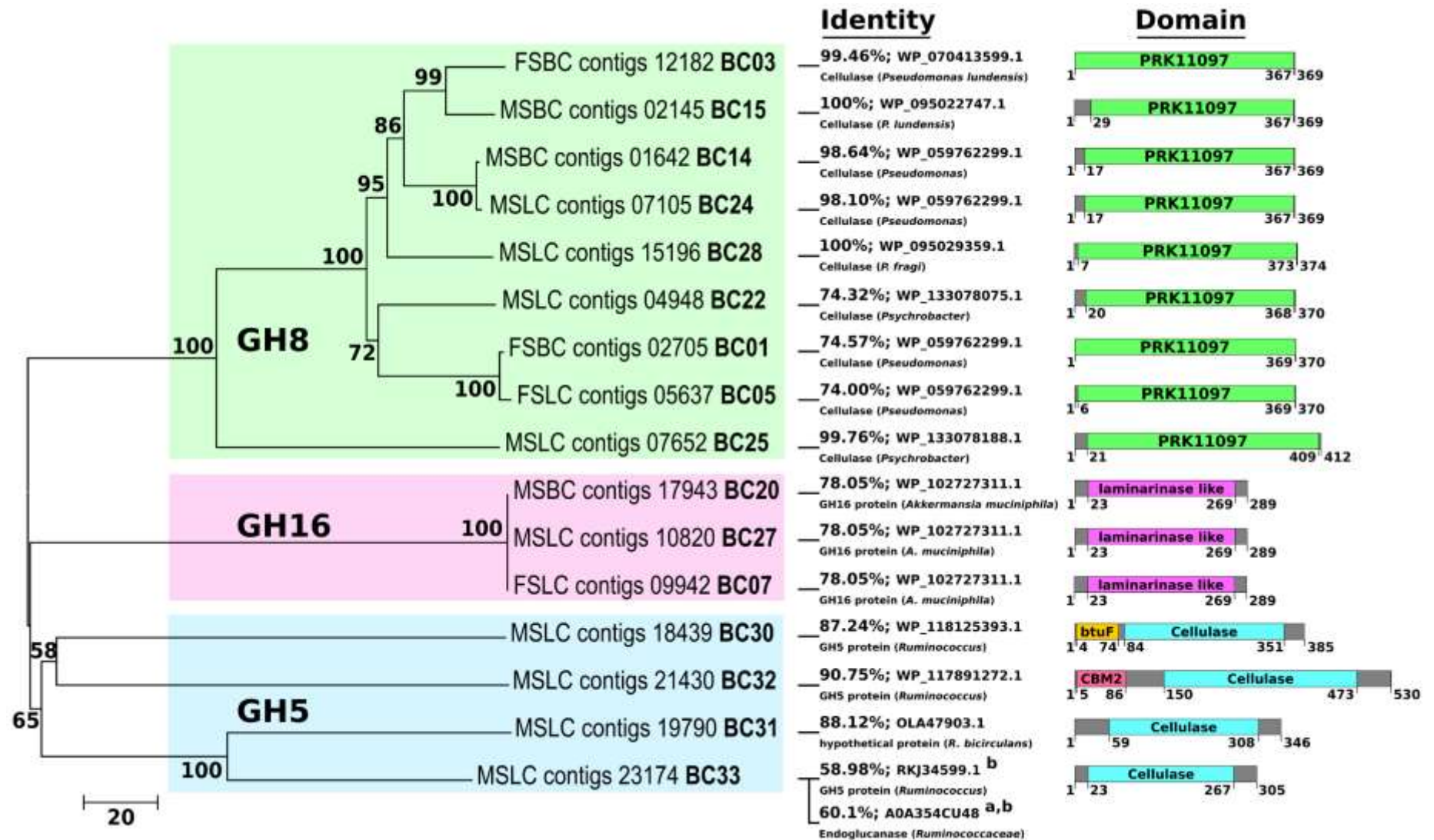
Figure 3. 17 Unrooted phylogenetic tree of putative beaver cellulase based on sequence similarity. Homologous genes search of putative beaver cellulase (BC) was performed against GenBank using blastp with default settings. The domain of putative cellulases were identified by using Conserved Domain (CD) - search. [a] Homologous search was performed against UniProt using blastp with default settings. [b] For sequence alignment, see Appendix 9.

Analysis with SignalP 4.0 server (Petersen et al., 2011) revealed no putative signal peptide in the sequence of BC33 endoglucanase, with predicted molecular mass of 36 kDa. Based on the sequence similarity (Figure 3. 17), BC33 endoglucanase was similar to endoglucanase from *Ruminococcus* sp., a Gram-positive bacterium. The secretory pathway, whose mechanisms are clearly defined, refers to Gram-negative bacteria (Yan and Wu, 2013). The secretion of BC33 endoglucanase could have been performed through another secretory pathway. The other putative GH5 endoglucanases shared identity scores of 87.24% (BC30), 88.12% (BC31), and 90.75% (BC32) to the closest similar endoglucanase sequences in the GenBank (Figure 3. 17). The BC33 putative GH5 endoglucanase was thus selected to be functionally expressed in *Eschericia coli* (Appendix 8). The GH5 family in general has two catalytic residues, catalytic nucleophile (Glu) and general acid/base (Glu) (Henrissat et al., 1995). Multiple sequence alignment of the BC33 protein sequence revealed the two putative catalytic residues as Glu167 and Glu254 (Figure 3. 18).
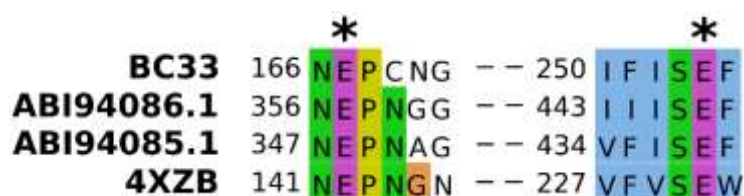
Figure 3. 18 Multiple alignment of BC33 with reference GH5 proteins. The catalytic residues were marked with asterisk (*). ABI94085.1 & ABI94086.1 = endo-β-1,4-glucanase from uncultured bacterium; 4XZB = endoglucanase GsCelA P1 from *Geobacillus* sp. 70PC53.

3.3.1. The phylogenetic tree and structure of BC33

From the 282 bacterial GH5 reference proteins used in phylogenetic analysis, Firmicutes and Proteobacteria were dominant, 45.7% and 17.8% respectively. The remaining reference proteins are derived from Actinobacteria (9.7%), Bacteroidetes (5.4%), Fibrobacter (3.6%), Thermotogae (2.2%), and uncultured bacteria (15.6%). The phylogenetic relationship of BC33 endoglucanase to these bacterial GH5 reference proteins is presented in Figure 3. 19. Based on the sequence similarity to the reference GH5, the BC33 is part of GH family 5 and subfamily 2 (GH5_2, Figure 3. 19). Subfamily 2, which is the largest in GH5 family (Aspeborg et al., 2012), is closely related to the GH5 subfamily 26 and 48. Both GH5_2 and GH5_26 displayed endo-β-1,4-glucanase activity toward cellulose, however, one activity toward both CMC and oat spelt xylan was reported from GH5_2 family (Aspeborg et al., 2012). Interestingly, a representative of GH5_2 has been reported to be a chitosanase with transglycosylation activity (Tanabe et al., 2003), which might explain the relation to GH5_48 that have specificity for chitin and chitosan derivatives (Aspeborg et al., 2012).
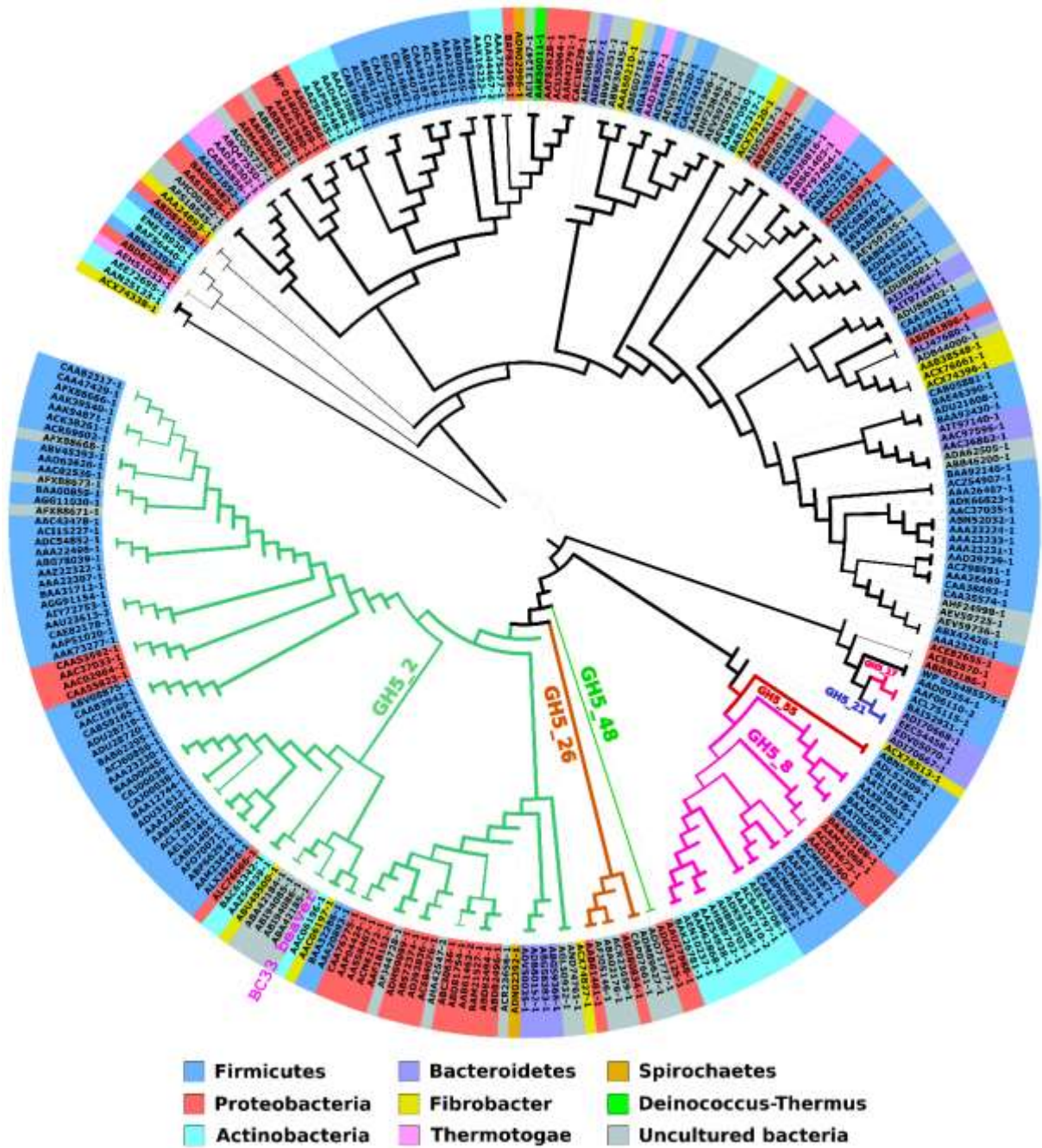
Figure 3. 19 The phylogenetic tree of recombinant BC33 to the bacterial GH5 reference proteins (282 proteins) based on the sequences similarity. The thickness of the branches corresponds to the bootstrap value. For the complete names of the 282 protein references, see Appendix 10.

BC33 is closely related to four representatives of the GH5 endoglucanases from uncultured bacteria (accession numbers of ABA42185.1, ABI94086.1, ABI94085.1, and ABA42184.1; Figure 3. 19). These four GH5 endoglucanases were isolated from rabbit cecum and were able to degrade various β-glucans such as lichenan, barley glucan, carboxymethylcellulose and xylan (Feng et al., 2007). Similar to BC33, the four endoglucanases contain one catalytic cellulase domain and no carbohydrate-binding module (Figure 3. 20).

51
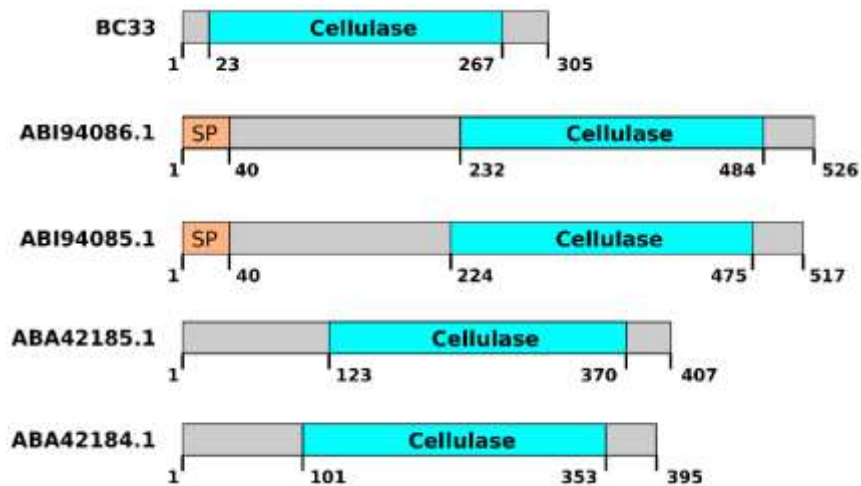
Figure 3. 20 The domain structure of BC33 and four closely related GH5_2 endoglucanases derived from rabbit cecum. ABI94085.1 & ABI94086.1 = endo-β-1,4-glucanase from uncultured bacterium; ABA42184.1 & ABA42185.1 = β-glucanase from uncultured bacterium; SP = signal peptide.

The BC33 3D structure was generated using I-TASSER server. The modelling resulted in five predicted models, with the highest confidence level (C-score) of -0.70 (2 is the highest). Based on the predicted protein structure, BC33 endoglucanase is similar to the endoglucanase GsCelA P1 from *Geobacillus* sp. 70PC53 (PDB: 4xzb), with the protein structure common to cellulase, the $(\alpha/\beta)_8$ TIM barrel fold structure (Figure 3. 21). The structural similarity score between these two proteins (TM-score) was 0.911 (1 is the highest), suggesting that the native protein structure of BC33 endoglucanase could be similar to that of GSCelA P1 endoglucanase.
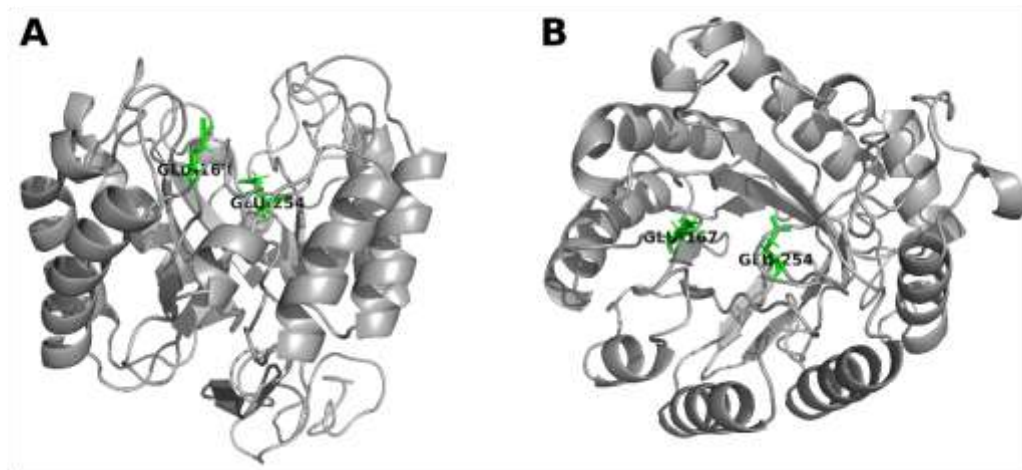
Figure 3. 21 The predicted 3D structure of recombinant BC33 from side (A) and top (B). The common TIM barrel fold with β-barrel in the inside surrounded by α-helices were shown. The conserved putative catalytic glutamate residues are colored in green.

3.3.2. Enzymatic assay of overexpressed BC33 protein

The recombinant BC33 expressed in *E. coli* BL21 (DE3) resulted in a protein with a molecular mass of 36.8 kDa (Figure 3. 22A & B, line E1). Recombinant endoglucanase BC33 displayed

an optimal temperature of 60 °C (Figure 3. 23A). Enzyme activity gradually increased from 10 to 60 °C. Enzyme activity decreased rapidly at higher temperatures, and lost its activity at 90 °C. At temperatures between 40 to 70 °C, the enzyme retained more than 50% of its activity. When compared to the emdoglucanase from *Trichoderma reesei* ATCC 26921 (Sigma), BC33 showed higher relative activity at all temperatures. The optimum pH for BC33 was evaluated at 60 °C under standard conditions. BC33 was active within a pH range of pH 3 to 8 (Figure 3. 23B), showing the highest activity at pH 4. The BC33 retained over 70% of its activity at pH 5 and pH 6. Thus, the optimal conditions for activity of the recombinant cellulase BC33 are pH 4 and 60 °C.
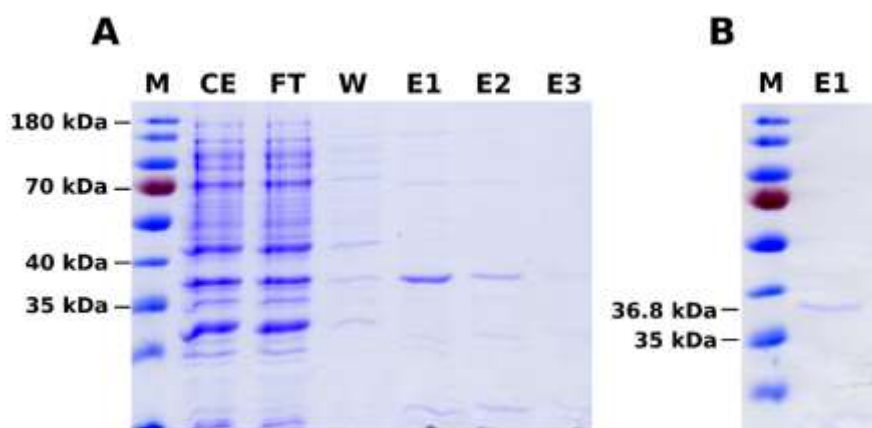
Figure 3. 22 (A) SDS-PAGE analysis for the purified recombinant cellulase BC33 , and (B) after size exclusion. M = protein ladder, CE = crude extract, FT = flow through, W = wash, E1-3 = elution 1-3.

3.3.3. Substrate specificity and kinetics of recombinant BC33

The specificity of BC33 with different substrates harboring β-glycosidic bonds was assayed at optimal conditions, citrate-phosphate buffer pH 4 and temperature of 60 °C for one hour. The enzyme displayed the highest activity toward barley β-glucan (3.157 ± 0.28 U/mg), followed with lichenan (3.024 ± 0.31 U/mg), CMC (1.585 ± 0.09 U/mg) and lastly xylan (1.578 ± 0.09 U/mg) (Figure 3. 24). The optimal condition for BC33 activity is in general similar to most cellulases with optimal temperatures between 45 and 55 °C and pH in the range of 4-5 (Pino et al., 2018). The other cellulases isolated from the rabbit's cecum for example, showed optimal activity toward neutral pH in the range of 6.0 - 7 and at optimal temperature of 50 °C (Feng et al., 2007). Kinetic parameters were determined using barley β-glucan as substrate. The Km and Vmax of BC33 were 0.205 ± 0.022 µM/min and 0.732 ± 0.026 µM/min, respectively.
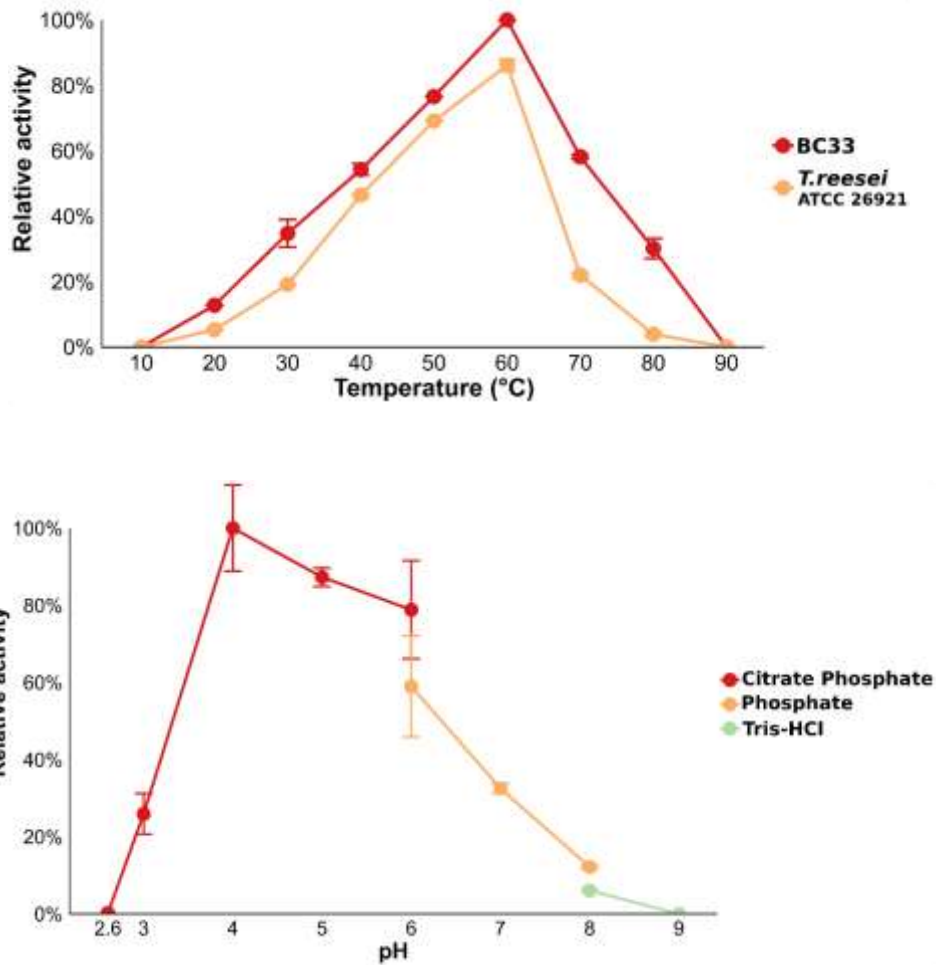
Figure 3. 23 The effect of temperature (A) on activity of recombinant BC33 and cellulase from *Trichoderma reesei* ATCC 26921 (Sigma), and the effect of pH (B) on activity of recombinant BC33. The relative activity of 100% for the temperature profile was 0.091 ± 0.002 SD, and for pH profile was 0.126 ± 0.014 SD.
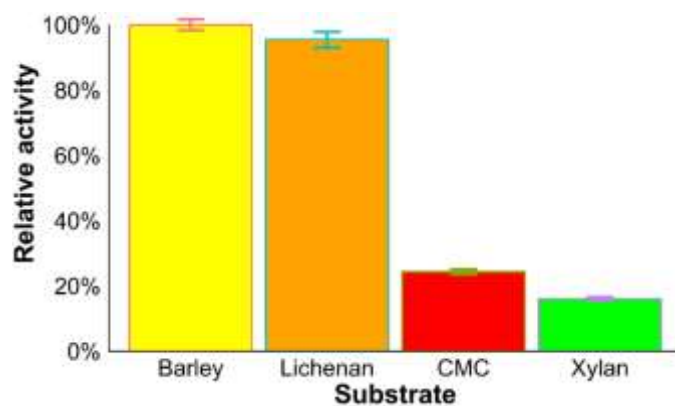
Figure 3. 24 The substrate specificity of recombinant BC33. Tests were performed at optimum temperature of 60 °C and pH 4 (citrate-phosphate) for one hour. The relative activity for 100% was 0.098 ± 0.015 SD.

4. GENERAL DISCUSSION

The bacterial community in the gut system provides additional metabolic capabilities to the respective host. For example, commensal bacteria help the breakdown of recalcitrant material, which the host cannot digest on its own (Hooper, 2001; Savage, 1977). This applies in particular to the gut system of herbivores, where by harboring richer bacterial communities than carnivores is a defining characteristic of herbivores, to compensate for the difficulty in digesting structural carbohydrates from plant materials (Choat and Clements, 1998). This symbiosis partly contributed to the adaptability of the class *Mammalia*, allowing them to occupy diverse environmental niches (Ley et al., 2008). The differences in habitat and diet have been investigated, and found to affect the bacterial community composition in mammalian gut (Nelson et al., 2013). A study with 60 mammalian species reveal that their gut bacterial communities are mainly dominated by Firmicutes (65.7%) and Bacteroidetes (16.3%), accompanied with less dominant phyla, e.g. Proteobacteria (8.8%), Actinobacteria (4.7%), Verrucomicrobia (2.2%), Fusobacteria (0.67%), Spirochaetes (0.46%), DSSl (0.35%), Fibrobacteres (0.13%), TM7 (0.13%), Cyanobacteria (0.10%), Planctomycetes (0.08%), Defferibacteres (0.05%) and Lentisphaerae (0.04%) (Ley et al., 2008). The herbivorous gut in general also shows high relative abundance of Firmicutes and Bacteroidetes, depending on the host species, as shown in several gut bacterial community studies, e.g. bovine (Jami et al., 2013), capybara (García-Amado et al., 2012), humus- and soil-feeding higher termites (Rossmassler et al., 2015), and North American beavers (Gruninger et al., 2016). The beaver hosts specialized gut microbial communities to degrade recalcitrant hardwood. Studies with gene marker and direct metagenome sequencing from North American beaver gut microbial community unravel a plethora of CAZymes and its distribution into several bacterial phyla, e.g. Firmicutes, Bacteroidetes, Chloroflexi, and Proteobacteria. (Gruninger et al., 2016; Wong et al., 2016, 2017).

In this study, the bacterial community of different compartments of the Eurasian beavers has been identified (Chapter 3.1). Functional prediction based on 16S rRNA gene-derived OTUs related to the cellulose breakdown, and comparison of the Eurasian beaver's gut bacterial community structure with other herbivores were also performed. Metagenome analysis of potential CAZymes including the characterization of a novel endoglucanase derived from the gut system of the Eurasian beaver was discussed (Chapter 3.2). CAZymes annotation from the

beaver gut metagenome could provide information on how the bacterial community associated with Eurasian beaver mediates plant fiber degradation.

4.1. Metagenome-guided discovery of new cellulase from Eurasian beaver gut microbiome

Direct sequencing of the gut metagenome provides in-depth analysis of gene functions, and enables screening for potential cellulases, especially in situations where function-based screening approaches are not feasible (e.g. due to plant material contamination). Additionally, by using marker gene (16S rRNA) studies, a detailed insight into the gut bacterial community structure can be obtained. Data generated from marker gene sequencing (amplicon sequencing), can be used to select which gut compartments to screen for potential cellulase-encoding genes by metagenome sequencing. Furthermore, with the assistance of functional profile predictions based on marker gene data, such as Tax4Fun, a more precise selection of the gut compartment is feasible.

The gut bacterial community of Eurasian beavers which was dominated by Firmicutes and Actinobacteria, differed from those of general terrestrial herbivores (Figure 4. 1). A study of terrestrial and marine mammalian gut discovered that, in general, Firmicutes and Bacteroidetes predominated the gut bacterial communities (Nelson et al., 2013). The guts of terrestrial carnivores were an exception, whereby Proteobacteria was more dominant than Bacteroidetes. In the herbivorous gut, Firmicutes and Bacteroidetes play major role to assists their hosts in digesting recalcitrant lignocellulose polymer (Flint et al., 2008). Most known cellulolytic bacterial species are from class Clostridia, e.g. *Clostridium thermocellum*, *C. cellulolyticum*, *C. cellulovorans*, *Ruminococcus albus*, and *R. flavefaciens* among other cellulolytic Clostridia (Flint et al., 2008). These cellulolytic bacterial species are able to digest cellulose and hemicellulose polymers, and subsequently release simple pentose and hexose sugars. The members of Bacteroidetes, e.g. *Prevotella* sp., interact synergistically with cellulolytic bacteria to improve polysaccharides degradation in plants, by degrading the oligosaccharides released from cellulose and hemicellulose degradation by cellulolytic Clostridia (Flint and Bayer, 2008; Matsui et al., 2000). While Bacteroidetes was abundant in the juvenile Eurasian beaver gut, its abundance in the subadult Eurasian beaver gut was relatively small, less than 8% in cecum and less than 4% in colon (Appendix 2 & 4).

Interestingly, the study of herbivores guts bacterial communities from terrestrial and marine mammals showed relatively low abundance for Fusobacteria (Figure 4. 1), which was relatively abundant in the Eurasian beaver gut, especially in the male beavers (Figure 3. 4; Appendix 2 &

4). The member of Fusobacteria were ranging from facultative anaerobes to obligate anaerobes, and are able to ferment carbohydrates to produce various organic acids, including butyrate and acetate (Potrykus et al., 2007). Member of Fusobacteria are also often linked to pathogenicity in the human gut (Allen-Vercoe et al., 2011; Riordan, 2007). Further studies investigating the abundant presence of Fusobacteria in the Eurasian beaver gut, which have not been recorded in other herbivores, are needed.
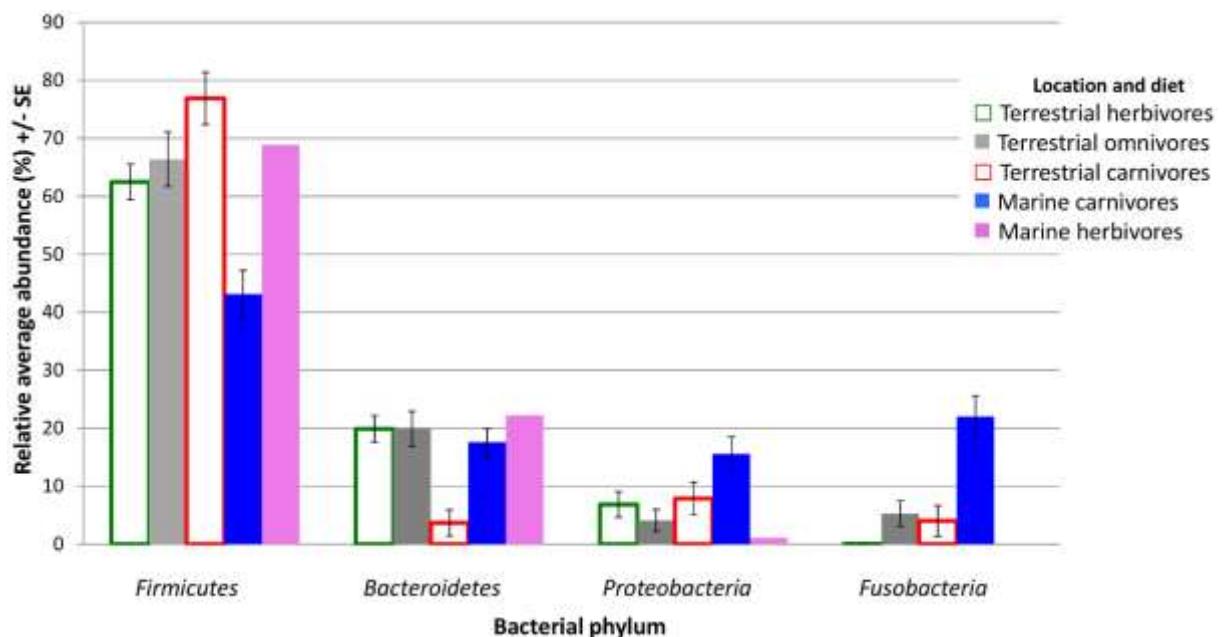

Figure 4. 1 The abundance of dominant phyla in the gut bacterial community of mammals based on groupings of habitat and diet. Error bars represent standard errors (SE). Source: Nelson et al. (2013).

The bacterial community structure and predicted functional profile analysis of the entire beaver gut compartments showed that the cecum and colon were selected for further metagenome sequencing. This was based on the high abundance of Clostridia, which are known for their cellulolytic capabilities, as well as the abundance of potential genes associated with cellulose breakdown, such as β-glucosidase and endoglucanase. The analysis of metagenomes generated from the back cecum and lower colon showed a different bacterial community profile, with Proteobacteria being the most abundant phylum, compared to the marker gene analysis where Firmicutes was the most abundant. Differences in the community structure are most likely due to the differences in the DNA extraction procedure. Since metagenome sequencing does not target marker genes like amplicon sequencing, a high amount of eukaryotic DNA will reduce the amount of bacterial sequences in a sample (Marotz et al., 2018). In this study, eukaryotic DNA most likely derives from the host (beaver) and digested plant material. Thus, prefiltering and bacterial DNA enrichment treatments at the DNA extraction step, might have altered the

microbial community structure, as mentioned in other studies (Henderson et al., 2013; Zielińska et al., 2017). The relative abundance of Gammaproteobacteria was higher in direct sequenced metagenomes compared to PCR-based marker gene analysis. Similar results were reported in a study performed by Tessler *et al.*(2017), in which the differences between 16S rRNA gene amplicon and shotgun sequencing were compared. However, Gammaproteobateria*,* Firmicutes and Actinobacteria, remained the dominant phyla independently of the approach applied.

The metagenomes of the back cecum and lower colon microbiome from male, and female beavers provide in-depth analysis of functional profiles at gene level. Functional profile analyses on carbohydrate metabolism using eggNOG and SEED databases in Megan (Figure 3. 9) showed similar results to the predicted functional profiles from marker genes (Figure 3. 5). This suggests that the functional profile analysis from gut microbiome with Tax4Fun, based on 16S rRNA genes data, shows good correlation with the functional profile from metagenome analyses. Both functional profiles suggest that putative genes associated with the membrane transport system were the most abundant, accompanied by genes involved in general metabolism (e.g. carbohydrate, protein, DNA). However, metagenome functional analysis provided detailed gene function annotations, which were not assigned by Tax4Fun, such as the stress response to environmental factors, biosynthesis of cofactors and coenzymes, and putative genes associated to energy metabolism. The stress response from SEED analysis of the Eurasian beaver gut metagenome, revealed the bacterial responses to the toxic compounds such as chemical defenses in plants. Since herbivorous diets mainly consist of plants, their guts are prone to secondary metabolites produced by plants. Plant secondary metabolites, which includes terpenoids, alkaloids, flavonoids, and saponins, are the most important chemical defense against herbivores, and pathogens (Mithöfer and Boland, 2012). The stress response was also displayed in the functional analysis of the snail (*Acatina fulica*) gut metagenome (Figure 4. 2A), in which many genes were associated with oxidative and osmotic stress, heat shock, and detoxification (Cardoso et al., 2012). The Eurasian beaver which lives near water environments, is also susceptible to contamination by anthropogenic factors such as antibiotics and heavy metals (Khatri and Tyagi, 2015; Na et al., 2018). Genes associated with resistance to these compounds were represented in the Eurasian beaver gut metagenome, similar to findings from *A. fulica* gut metagenome analysis (Figure 4. 2B). This indicates that the gut bacterial community in herbivores might play an important role in protecting the hosts from chemical compounds in plant secondary metabolites, as well as from other toxic compounds which contaminate water environments.
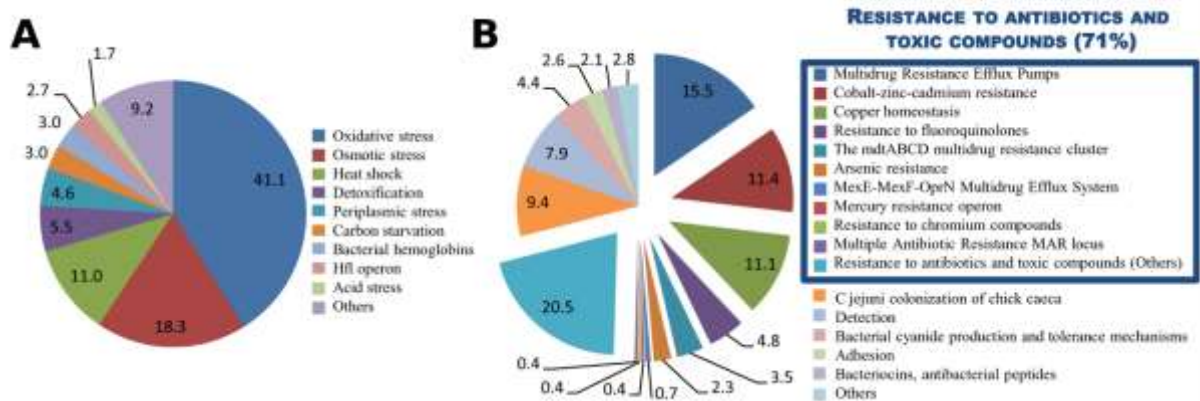
Figure 4. 2 (A) SEED analysis of the snail crop microbiome, indicating percent of genes in the "Stress Response" subsystem, and (B) percent of genes assigned to the "Virulence, Disease and Defense". Source: adapted from Cardoso et al. (2012).

Metagenome-derived analysis of carbohydrate-active enzymes (CAZymes) families showed the presence of various glycoside hydrolase (GHs) families. Information on GH families associated with the beaver gut metagenome could shed light on how plant fiber was digested. GH3 and GH5 were the most abundant families acting on cellulosic materials (Appendix 6). This is in accordance with the result based of the CAZy-database annotation (Figure 3. 15). The higher relative abundance of GH3 over GH5 was also presented by other herbivores such as snail, termite, wallaby and panda, as well as the human gut metagenome, which showed a similar pattern (Table 4. 1) (Cardoso et al., 2012). This is not unexpected, as GH3, which includes β-glucosidases, plays an important role in converting products of cellulose degradation by endoglucanases and cellobiohydrolases, that can otherwise inhibit cellulose degradation (Xiao et al., 2004). The GH3 family, primarily acting on oligosaccharides, was overrepresented in both the back cecum and lower colon metagenomes of both subadult beavers (Figure 3. 15; Appendix 6). The GH3 family was frequently observed in bacteria, plants and fungi, but rarely or absent in archaea and animals (Cournoyer and Faure, 2003). Tax4Fun predictions also fit to the abundance of putative genes associated with the GH3 family, due to the overrepresented β-glucosidases in the cecum and colon of all Eurasian beavers in this study (Figure 3. 5). On the other hand, the GH5 family covers a wide range of enzymes with different substrate specificities, such as cellulose, mannan, xylan, xyloglucan, and arabinoxylan, and cellobiose (Aspeborg et al., 2012). Table 4. 1 show that GH5 CAZymes were relatively abundant in the herbivores gut metagenome, except in panda, in which only one GH5-associated gene was found. A study on the giant panda gut reveal the carnivore-like gut microbiota, which was predominated by *Streptococcus* (Xue et al., 2015). Our finding also supports this statement that Streptococcaceae relative abundance was in general over 90% (Appendix 5), with a few members of Clostridia were present, except relatively low abundance

of Clostridiaceae 1 (less than 6%, one sample with 15.3% relative abundance), and Lachnospiraceae (less than 0.5%). Termites which have the highest number of identified GH5-associated genes (Table 4. 1), probably benefit from a gut microbiota rich in cellulolytic bacteria such as members of Firmicutes and Spirochaetes (Appendix 2). Many GH families which act on cellulose (e.g. GH5, 9, 30) were founds in Spirochaetes, such as *Spirochaeta thermophila* (Angelov et al., 2011; Ransom-Jones et al., 2017). In addition, metatranscriptomic studies of higher termite gut (*Nasutitermes* sp.), identified GH5 as one of the most highly expressed glycoside hydrolases (Marynowska et al., 2017).

Table 4. 1 Comparison of the glycoside hydrolase (GHs) profiles targeting plant structural polysaccharides in the snail, termite, giant panda, wallaby, and human metagenomes.

CAZy Family	Known activity	Snail	Termite	Wallaby	Panda	Human
Cellulases						
GH5	cellulase	36	125	27	1	7
GH6	endoglucanase	4	0	0	0	0
GH7	endoglucanase	0	0	0	0	0
GH9	endoglucanase	15	43	5	0	0
GH44	endoglucanase	0	0	0	0	0
GH45	endoglucanase	0	6	0	0	0
GH48	endo-processive cellulase	2	0	0	0	0
Total		57 (2)	174 (16)	32 (4)	1 (0.5)	0
Oligosaccharide-degrading enzymes						
GH1	β-glucosidases & other β-linked dimers	294	27	94	41	54
GH2	β-galactosidases & other β-linked dimers	66	32	39	4	29
GH3	mainly β-glucosidases	219	109	101	11	55
GH29	α-L-fucosidase	70	12	5	0	7
GH35	β-galactosidases	32	7	8	1	4
GH38	α-mannosidase	18	18	3	8	6
GH39	β-xylosidases	6	13	3	8	2
GH42	β-galactosidases	54	33	17	7	15
GH43	arabinases & xylosidases	185	63	72	13	34
GH52	β-xylosidases	0	3	0	0	0
Total		944 (36)	317 (28)	342 (39)	93 (41)	206 (36)

The numbers in parentheses represent the percentage of these groups relative to the total number of GHs identified in the metagenomic dataset (2590 for snail, 872 for wallaby, 1117 for termite, 227 for panda (from 3 samples), and 580 for human (from 2 samples)). Source: adapted from Cardoso et al. (2012).

Recent metagenomic study of cellulose- and poplar hydrolysate- degrading microcosm from the gut microbial community of the North American beaver reveals that, GH5 and GH9, which act on cellulose, are present in relatively high abundance from cultures enriched on poplar hydrolysate (Wong et al., 2017). This suggest that GH5 and GH9 CAZymes in the North American beaver gut could work in synergy to digest plant materials. We can conclude that

GH5 cellulases could be the key players in degrading cellulose in the Eurasian beaver gut. Further studies on the expression level of the CAZymes from Eurasian beaver gut metatranscriptomes might uncover the synergistic relationship between CAZymes families in assisting plants material degradation.

By combining taxonomy classification and the CAZymes annotation, CAZymes found in the cecum and colon can be associated to the bacterial taxa, thus, the bacterial taxa responsible for cellulose degradation can be narrowed down. The high relative abundance of Clostridia in the beaver cecum and colon (Clostridiaceae I, Ruminococcaceae, and Lachnospiraceae families) support early findings that Clostridia were involved in the breakdown of plant cell wall material. The genus *Clostridium* has been extensively studied and known to harbor various GH families associated with cellulose breakdown. Various GH families acting on cellulose (GH5, 8, 9, and 48), and hemicellulose (GH10, 11, 16, 18, 26, 30, 53, 81, and 74) have been identified from the *Clostridium thermocellum* draft genomes (Demain et al., 2005). Additionally, genes which act on cellulose, and hemicellulose belonging to GH families (GH9, 11, and 27) have been characterized from *C. cellulolyticum* (Blouzard et al., 2007). Similar to *C. thermocellum*, the draft genome of *Ruminococcus flavefaciens* FD-1 contains a diverse set of genes encoding proteins that belong to GH families acting on cellulose (GH3, 5, 9, 48, 44, 48, and 74), and hemicellulose (GH2, 10, 11, 16, 26, 42, 43, 44, and 53), with GH5 and GH9 as the most abundant families (Berg Miller et al., 2009). Given this finding, it is apparent that class Clostridia plays an important role in assisting beavers to digest plant material in their cecum and colon. The high abundances of GH3 and GH5 over other GH families, suggests that the bacterial communities in beaver cecum and colon utilize cellulases from these families to degrade plant material. In addition to the predicted GH3 and GH5 families, the GH5 subfamily 2 endoglucanase from the lower colon of male beaver was characterized, supporting the previous statement. The GH5 endoglucanase of the Eurasian beaver showed optimal activity at a higher temperature (60 °C) and in an acidic environment (pH 4) than other GH5 endoglucanases from different gut system, such as swamp buffalo rumen (45 °C, pH 5.5). The high activity of BC33 endoglucanase to β-glucan was in accordance with characteristic of the GH5 subfamily 2. In addition, activity on CMC and xylan, which BC33 demonstrated, was also demonstrated by endoglucanase from GH5_2 *Fibrobacter succinogenes* S85, but no other representative of GH5_2 exhibit activity against xylan was observed.

5. SUMMARY

The mammalian gut harbors a complex and specialized microbiota, which includes all tree domains of life (Archaea, Bacteria, and Eukarya). Bacteria are the most abundant group within the gut system, and they take part in metabolic processes that expand the hosts metabolic potential. One of the most beneficial traits of this relationship is that bacteria help to degrade recalcitrant plant materials. In this study, the microbiome of three Eurasian beaver gut systems (male and female subadult, and male juvenile) were analysed by applying metagenomics using 16S rRNA gene marker-based and direct sequencing-based approaches, to gain insights into the diversity, structure and function of the gut-inhabiting bacterial communities and genes involved in cellulose breakdown.

Metagenomic DNA was isolated from the entire gut system of three Eurasian beavers, covering different compartments (stomach, small intestine, cecum, and colon). The taxonomic compositions of the bacterial communities within these compartments were assessed using the hypervariable regions V3-V4 of the 16S rRNA gene, amplified from the isolated DNA by PCR. Subsequently, amplicon-based analysis of 2,599,870 high-quality paired end reads revealed 277 unique OTUs in the entire dataset. The bacterial diversity in the cecum and colon was higher in the male subadult beaver than in the female subadult and male juvenile beaver. The gut bacterial community was dominated by Firmicutes and Actinobacteria, followed by Proteobacteria, Verrucomicrobia, Fusobacteria, Bacteroidetes, and Tenericutes. A bacterial community shift from the juvenile to subadult beaver was indicated by the decrease of members of Bacteroidetes and an increase in Verrucomicrobia members. In addition, the presence and abundance of some phyla could be associated with sex, e.g. Fusobacteria, were detected in both male beavers but not in the female. However, further studies of Eurasian beaver gut microbial communities are necessary to confirm these trends. The presence of Clostridiaceae 1, Lachnospiraceae*, and* Ruminococcaceae in the cecum and colon of the beaver indicated that plant cell wall breakdown is mainly performed in these compartments. The predicted functional profiles showed an increased relative abundance in genes necessary for cellulose breakdown and uptake of degradation products, which is in accordance to the presence of potential cellulolytic bacterial species. The abundance of unclassified Clostridiaceae 1, Lachnospiraceae*, and* Ruminococcaceae in the cecum and colon of the beaver as well as unclassified Bacteroidaceae in juvenile beaver suggest the presence of novel species exhibiting cellulolytic activity. In comparison to its North American relative, the Eurasian beaver has a higher relative abundance

of Actinobacteria and Verrucomicrobia, and a lower abundance of Bacteroidetes in the gut system. In addition, high relative abundances of Ruminococcaceae and Lachnospiraceae were detected in the cecum and colon compartments of the Eurasian North American beaver (NA$_{Grun}$). In general, the bacterial community structure in the cecum was similar to that of the colon in both the Eurasian and North American beavers.

In-depth analysis of the microbiome of the back cecum and lower colon of both subadult beavers was performed through direct metagenome sequencing. A total of 6,200,436 high-quality paired-end reads were obtained. Metagenome assembly resulted in a total of 101,060 contigs with the largest contig of 198,219 bp. The bacterial community structure derived from the metagenome sequencing differed to that based on amplicon sequencing. The relative abundance of Proteobacteria was higher in the metagenome sequences. However, Proteobacteria, Firmicutes, and Actinobacteria remained the dominant phyla. The functional analysis of the metagenomes indicates the metabolic ability of the Eurasian beaver gut microbiome to neutralize toxic compounds from plants and heavy metals, as well as resistance to pathogens and antibiotics. Functional analysis revealed a high diversity of CAZymes, especially glycoside hydrolases (GHs) for the degradation of polysaccharides. A high relative abundance of genes associated to the cellulolytic process, especially GH3 and GH5 family members, was detected in the metagenomes. In addition, the CAZyme family involved in lignin breakdown, AA2, and an annotated DyP-like protein responsible for lignin degradation were also present in the gut microbiome of the beaver. Through combination of marker gene community analysis and metagenome analysis, a novel cellulase from GH family 5 was identified and isolated. The beaver cellulase candidate 33 (BC33) was characterized and exhibit optimal catalytic activity at pH 4 and 60 °C, with high affinity towards barley glucan (K_m 0.205 ± 0.022 µM/min, V_{max} 0.732 ± 0.026 µM/min).

Since Eurasian beaver gut system is mainly an anoxic environment, the anaerobic cellulolytic bacteria, e.g. Clostridia could also produce cellulosomes. Parts of the building block for cellulosomes such as dockerin and cohesin were present in the Eurasian beaver gut metagenomes. Further studies are needed to confirm the existence of cellulosome system and to identify which CAZymes families are bound to the cellulosome, thus, enabling a better understanding of plant material digestion in the Eurasian beaver gut.

6. REFERENCES

Akileswaran, L., Brock, B. J., Cereghino, J. L., and Gold, M. H. (1999). 1,4-benzoquinone reductase from *Phanerochaete chrysosporium*: cDNA cloning and regulation of expression. *Appl. Environ. Microbiol.* 65, 415–21. doi: 10.1007/s00253-006-0820-9

Akinosho, H., Yee, K., Close, D., and Ragauskas, A. (2014). The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications. *Front. Chem.* 2, 1–18. doi:10.3389/fchem.2014.00066.

Allen-Vercoe, E., Strauss, J., and Chadee, K. (2011). *Fusobacterium nucleatum*. *Gut Microbes* 2, 294–298. doi:10.4161/gmic.2.5.18603.

Allocati, N., Federici, L., Masulli, M., and Di Ilio, C. (2009). Glutathione transferases in bacteria. *FEBS J.* 276, 58–75. doi:10.1111/j.1742-4658.2008.06743.x.

Amin, F. R., Khalid, H., Zhang, H., Rahman, S. u, Zhang, R., Liu, G., et al. (2017). Pretreatment methods of lignocellulosic biomass for anaerobic digestion. *AMB Express* 7, 72. doi:10.1186/s13568-017-0375-4.

Andersen, S. K., Kirkegaard, R. H., Karst, S. M., and Albertsen, M. (2018). ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. *bioRxiv* 22, 1–3. doi:10.1101/299537.

Angelov, A., Loderer, C., Pompei, S., and Liebl, W. (2011). Novel Family of Carbohydrate-Binding Modules Revealed by the Genome Sequence of *Spirochaeta thermophila* DSM 6192. *Appl. Environ. Microbiol.* 77, 5483–5489. doi:10.1128/AEM.00523-11.

Armstrong, Z., Mewis, K., Liu, F., Morgan-Lang, C., Scofield, M., Durno, E., et al. (2018). Metagenomics reveals functional synergy and novel polysaccharide utilization loci in the *Castor canadensis* fecal microbiome. *ISME J.* 12, 2757–2769. doi:10.1038/s41396-018-0215-9.

Artzi, L., Bayer, E. A., and Moraïs, S. (2017). Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nat. Rev. Microbiol.* 15, 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, e00411-15. doi:10.1128/mBio.00411-15.

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, 186. doi:10.1186/1471-2148-12-186.

Aßhauer, K. P., Wemheuer, B., Daniel, R., and Meinicke, P. (2015). Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 31, 2882–4. doi:10.1093/bioinformatics/btv287.

Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., and Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* 307, 1915–20. doi:10.1126/science.1104816.

Bajpai, P. (2016). *Pretreatment of Lignocellulosic Biomass for Biofuel Production*. Singapore: Springer Singapore doi:10.1007/978-981-10-0687-6.

Bakare, M. K., Adewale, I. O., and Shonukan, O. O. (2005). Purification and characterization of cellulase from the wild-type and two improved mutants of *Pseudomonas fluorescens*. *African J. Biotechnol.* 4, 898–904. doi: 10.1007/s00253-006-0820-9.

Barnes, D. F., and Floor, W. M. (1996). Rural energy in developing countries: A challenge for economic development. *Annu. Rev. Energy Environ.* 21, 497–530. doi:10.1146/annurev.energy.21.1.497.

Bayer, E. A., Setter, E., and Lamed, R. (1985). Organization and distribution of the cellulosome in *Clostridium thermocellum*. *J. Bacteriol.* 163, 552–9. doi: 10.1007/s00253-006-0820-9.

Behera, B. C., Sethi, B. K., Mishra, R. R., Dutta, S. K., and Thatoi, H. N. (2017). Microbial cellulases – Diversity & biotechnology with reference to mangrove environment: A review. *J. Genet. Eng. Biotechnol.* 15, 197–210. doi:10.1016/j.jgeb.2016.12.001.

Benson, D. A. (2003). GenBank. *Nucleic Acids Res.* 31, 23–27. doi:10.1093/nar/gkg057.

Berg Miller, M. E., Antonopoulos, D. A., Rincon, M. T., Band, M., Bari, A., Akraiko, T., et al. (2009). Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. *PLoS One* 4, e6650. doi:10.1371/journal.pone.0006650.

Berger, E., Zhang, D., Zverlov, V. V., and Schwarz, W. H. (2007). Two noncellulosomal cellulases of *Clostridium thermocellum*, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. *FEMS Microbiol. Lett.* 268, 194–201. doi:10.1111/j.1574-6968.2006.00583.x.

Berkelmann, D., Schneider, D., Engelhaupt, M., Heinemann, M., Christel, S., Wijayanti, M., et al. (2018). How rainforest conversion to agricultural systems in Sumatra (Indonesia) affects active soil bacterial communities. *Front. Microbiol.* 9, 1–13. doi:10.3389/fmicb.2018.02381.

Berlemont, R., and Martiny, A. C. (2013). Phylogenetic distribution of potential cellulases in bacteria. *Appl. Environ. Microbiol.* 79, 1545–1554. doi:10.1128/AEM.03305-12.

Blackburn, N. T., and Clarke, A. J. (2000). Assay for lytic transglycosylases: a family of peptidoglycan lyases. *Anal. Biochem.* 284, 388–393. doi:10.1006/abio.2000.4707.

Bleich, A., and Fox, J. G. (2015). The mammalian microbiome and its importance in laboratory animal research. *ILAR J.* 56, 153–158. doi:10.1093/ilar/ilv031.

Blouzard, J. C., Bourgeois, C., De Philip, P., Valette, O., Bélaïch, A., Tardif, C., et al. (2007). Enzyme diversity of the cellulolytic system produced by *Clostridium cellulolyticum* explored by two-dimensional analysis: Identification of seven genes encoding new dockerin-containing proteins. *J. Bacteriol.* 189, 2300–2309. doi:10.1128/JB.00917-06.

BP (2018). *BP Statistical Review of World Energy*, *67th edition*. Available at: https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html.

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.

doi:10.1016/0003-2697(76)90527-3.

Brás, J. L. A., Cartmell, A., Carvalho, A. L. M., Verzé, G., Bayer, E. A., Vazana, Y., et al. (2011). Structural insights into a unique cellulase fold and mechanism of cellulose hydrolysis. *Proc. Natl. Acad. Sci. U. S. A.* 108, 5237–42. doi:10.1073/pnas.1015006108.

Breznak, J. a, and Brune, A. (1994). Role of microorganisms in the digestion of lignocellulose by termites. *Annu. Rev. Entomol.* 39, 453–487. doi:10.1146/annurev.en.39.010194.002321.

Broeker, J., Mechelke, M., Baudrexl, M., Mennerich, D., Hornburg, D., Mann, M., et al. (2018). The hemicellulose-degrading enzyme system of the thermophilic bacterium *Clostridium stercorarium*: comparative characterisation and addition of new hemicellulolytic glycoside hydrolases. *Biotechnol. Biofuels* 11, 229. doi:10.1186/s13068-018-1228-3.

Brulc, J. M., Antonopoulos, D. A., Berg Miller, M. E., 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.* 106, 1948–1953. doi:10.1073/pnas.0806191105.

Brumm, P., Hermanson, S., Gowda, K., Xie, D., and Mead, D. (2015). *Clostridium thermocellum* Cel5L – cloning and characterization of a new, thermostable GH5 cellulase. *Int. J. Biochem. Res. Rev.* 6, 62–74. doi:10.9734/IJBCRR/2015/15199.

Brune, A., and Dietrich, C. (2015). The gut microbiota of termites: digesting the diversity in the light of ecology and evolution. *Annu. Rev. Microbiol.* 69, 145–166. doi:10.1146/annurev-micro-092412-155715.

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi:10.1038/nmeth.3176.

Cai, S., and Dong, X. (2010). *Cellulosilyticum ruminicola* gen. nov., sp. nov., isolated from the rumen of yak, and reclassification of *Clostridium lentocellum* as *Cellulosilyticum lentocellum* comb. nov. *Int. J. Syst. Evol. Microbiol.* 60, 845–849. doi:10.1099/ijs.0.014712-0.

Cai, S., Li, J., Ze Hu, F., Zhang, K., Luo, Y., Janto, B., et al. (2010). *Cellulosilyticum ruminicola*, a newly described rumen bacterium that possesses redundant Fibrolytic-protein-encoding genes and degrades lignocellulose with multiple carbohydrate borne fibrolytic enzymes. *Appl. Environ. Microbiol.* 76, 3818–3824. doi:10.1128/AEM.03124-09.

Callahan, B. J., McMurdie, P. J., and Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* 11, 2639–2643. doi:10.1038/ismej.2017.119.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi:10.1093/nar/gkn663.

Cardoso, A. M., Cavalcante, J. J. V, Cantao, M. E., Thompson, C. E., Flatschart, R. B., Glogauer, A., et al. (2012). metagenomic analysis of the microbiota from the crop of an invasive snail reveals a rich

reservoir of novel genes. *PLoS One* 7. doi:10.1371/journal.pone.0048505.

Carvalho, A. L., Dias, F. M. V., Prates, J. A. M., Nagy, T., Gilbert, H. J., Davies, G. J., et al. (2003). Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proc. Natl. Acad. Sci.* 100, 13809–13814. doi:10.1073/pnas.1936124100.

Chan, S. W. L., Henderson, I. R., and Jacobsen, S. E. (2005). Erratum: Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat. Rev. Genet.* 6, 351–360. doi:10.1038/nrg1601.

Chen, H. (2014). *Biotechnology of Lignocellulose*. Dordrecht: Springer Netherlands doi:10.1007/978-94-007-6898-7.

Choat, J. H., and Clements, K. D. (1998). Vertebrate herbivores in marine and terrestrial environments: a nutritional ecology perspective. *Annu. Rev. Ecol. Syst.* 29, 375 – 403. doi: 10.1146/annurev.ecolsys.29.1.375.

Chow, J., Lee, S. M., Shen, Y., Khosravi, A., and Mazmanian, S. K. (2010). Host-bacterial symbiosis in health and disease. *Adv. Immunol.* 107, 243–74. doi:10.1016/B978-0-12-381300-8.00008-3.

Clavel, T., Charrier, C., Braune, A., Wenning, M., Blaut, M., and Haller, D. (2009). Isolation of bacteria from the ileal mucosa of TNFdeltaARE mice and description of *Enterorhabdus mucosicola* gen. nov., sp. nov. *Int. J. Syst. Evol. Microbiol.* 59, 1805–1812. doi:10.1099/ijs.0.003087-0.

Clavel, T., Desmarchelier, C., Haller, D., Gérard, P., Rohn, S., Lepage, P., et al. (2014). Intestinal microbiota in metabolic diseases. *Gut Microbes* 5, 544–551. doi:10.4161/gmic.29331.

Clavel, T., Duck, W., Charrier, C., Wenning, M., Elson, C., and Haller, D. (2010). *Enterorhabdus caecimuris* sp. nov., a member of the family Coriobacteriaceae isolated from a mouse model of spontaneous colitis, and emended description of the genus *Enterorhabdus Clavel* et al. *Int. J. Syst. Evol. Microbiol.* 60, 1527–31. doi:10.1099/ijs.0.015016-0.

Cournoyer, B., and Faure, D. (2003). Radiation and functional specialization of the family-3 glycoside hydrolases. *J. Mol. Microbiol. Biotechnol.* 5, 190–198. doi:10.1159/000070269.

Currier, A. A. (1958). A preliminary study of cellulose digestion in the beaver.

Dao, M. C., Everard, A., Aron-Wisnewsky, J., Sokolovska, N., Prifti, E., Verger, E. O., et al. (2016). *Akkermansia muciniphila* and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. *Gut* 65, 426–436. doi:10.1136/gutjnl-2014-308778.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., et al. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci.* 107, 14691–14696. doi:10.1073/pnas.1005963107.

De Giglio, O., Napoli, C., Lovero, G., Diella, G., Rutigliano, S., Caggiano, G., et al. (2015). Antibiotic susceptibility of *Legionella pneumophila* strains isolated from hospital water systems in Southern Italy. *Environ. Res.* 142, 586–590. doi:10.1016/j.envres.2015.08.013.

de Gonzalo, G., Colpa, D. I., Habib, M. H. M., and Fraaije, M. W. (2016). Bacterial enzymes involved in lignin degradation. *J. Biotechnol.* 236, 110–119. doi:10.1016/j.jbiotec.2016.08.011.

Demain, A. L., Newcomb, M., and Wu, J. H. D. (2005). Cellulase, clostridia, and ethanol. *Microbiol. Mol. Biol. Rev.* 69, 124–54. doi:10.1128/MMBR.69.1.124-154.2005.

Derrien, M., Collado, M. C., Ben-Amor, K., Salminen, S., and de Vos, W. M. (2008). The mucin degrader *Akkermansia muciniphila* is an abundant resident of the human intestinal tract. *Appl. Environ. Microbiol.* 74, 1646–1648. doi:10.1128/AEM.01226-07.

Derrien, M., Vaughan, E. E., Plugge, C. M., and de Vos, W. M. (2004). *Akkermansia municiphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int. J. Syst. Evol. Microbiol.* 54, 1469–1476. doi:10.1099/ijs.0.02873-0.

Devanga Ragupathi, N. K., Muthuirulandi Sethuvel, D. P., Inbanathan, F. Y., and Veeraraghavan, B. (2018). Accurate differentiation of *Escherichia coli* and *Shigella serogroups*: challenges and strategies. *New Microbes New Infect.* 21, 58–62. doi:10.1016/j.nmni.2017.09.003.

Dietrich, C., Kohler, T., and Brune, A. (2014). The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl. Environ. Microbiol.* 80, 2261–2269. doi:10.1128/AEM.04206-13.

Ding, S.-Y., Rincon, M. T., Lamed, R., Martin, J. C., McCrae, S. I., Aurilia, V., et al. (2001). Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. *J. Bacteriol.* 183, 1945–1953. doi:10.1128/JB.183.6.1945-1953.2001.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 14, 927–930. doi:10.1111/j.1654-1103.2003.tb02228.x.

Dixon, R., Austin, S., Eydmann, T., Jones, T., Söderbäck, E., and Hill, S. (1995). "Regulation of nitrogen fixation by the NIFL and NIFA proteins from *Azotobacter vinelandii*," in *10th International Congress on Nitrogen Fixation*, 171–176.

Doi, R. H., Goldstein, M., Hashida, S., Park, J.-S., and Takagi, M. (1994). The *Clostridium cellulovorans* cellulosome. *Crit. Rev. Microbiol.* 20, 87–93. doi:10.3109/10408419409113548.

Doi, R. H., and Kosugi, A. (2004). Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat. Rev. Microbiol.* 2, 541–551. doi:10.1038/nrmicro925.

Dos Santos, P. C., Fang, Z., Mason, S. W., Setubal, J. C., and Dixon, R. (2012). Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics* 13, 162. doi:10.1186/1471-2164-13-162.

Droge, S., Frohlich, J., Radek, R., and Konig, H. (2006). spirochaeta coccoides sp. nov., a novel coccoid spirochete from the hindgut of the Termite neotermes castaneus. *Appl. Environ. Microbiol.* 72, 392–397. doi:10.1128/AEM.72.1.392-397.2006.

Dubinina, G., Grabovich, M., Leshcheva, N., Gronow, S., Gavrish, E., and Akimov, V. (2015). *Spirochaeta sinaica* sp. nov., a halophilic spirochaete isolated from a cyanobacterial mat. *Int. J. Syst. Evol. Microbiol.* 65, 3872–3877. doi:10.1099/ijsem.0.000506.

Elifantz, H., Waidner, L. A., Michelou, V. K., Cottrell, M. T., and Kirchman, D. L. (2008). Diversity and abundance of glycosyl hydrolase family 5 in the North Atlantic Ocean. *FEMS Microbiol. Ecol.*

63, 316–327. doi:10.1111/j.1574-6941.2007.00429.x.

Fawal, N., Li, Q., Savelli, B., Brette, M., Passaia, G., Fabre, M., et al. (2012). PeroxiBase: a database for large-scale evolutionary analysis of peroxidases. *Nucleic Acids Res.* 41, D441–D444. doi:10.1093/nar/gks1083.

Feng, Y., Duan, C.-J., Pang, H., Mo, X.-C., Wu, C.-F., Yu, Y., et al. (2007). Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. *Appl. Microbiol. Biotechnol.* 75, 319–328. doi:10.1007/s00253-006-0820-9.

Ferrario, C., Statello, R., Carnevali, L., Mancabelli, L., Milani, C., Mangifesta, M., et al. (2017). How to feed the mammalian gut microbiota: bacterial and metabolic modulation by dietary fibers. *Front. Microbiol.* 8, 1–11. doi:10.3389/fmicb.2017.01749.

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, D279–D285. doi:10.1093/nar/gkv1344.

Flint, H. J., and Bayer, E. A. (2008). Plant cell wall breakdown by anaerobic microorganisms from the mammalian digestive tract. *Ann. N. Y. Acad. Sci.* 1125, 280–288. doi:10.1196/annals.1419.022.

Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R., and White, B. A. (2008). Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat. Rev. Microbiol.* 6, 121–131. doi:10.1038/nrmicro1817.

Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P., and Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* 3, 289–306. doi:10.4161/gmic.19897.

Fontes, C. M. G. A., and Gilbert, H. J. (2010). Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.* 79, 655–681. doi:10.1146/annurev-biochem-091208-085603.

Fujimoto, Z. (2013). Structure and function of carbohydrate-binding module families 13 and 42 of glycoside hydrolases, comprising a β-trefoil fold. *Biosci. Biotechnol. Biochem.* 77, 1363–1371. doi:10.1271/bbb.130183.

García-Amado, M. A., Godoy-Vitorino, F., Piceno, Y. M., Tom, L. M., Andersen, G. L., Herrera, E. a., et al. (2012). Bacterial diversity in the cecum of the world's largest living rodent (*Hydrochoerus hydrochaeris*). *Microb. Ecol.* 63, 719–725. doi:10.1007/s00248-011-9963-z.

Garvey, M., Klose, H., Fischer, R., Lambertz, C., and Commandeur, U. (2013). Cellulases for biomass degradation: Comparing recombinant cellulase expression platforms. *Trends Biotechnol.* 31, 581–593. doi:10.1016/j.tibtech.2013.06.006.

Geerlings, S., Kostopoulos, I., de Vos, W., and Belzer, C. (2018). *Akkermansia muciniphila* in the human gastrointestinal tract: when, where, and how? *Microorganisms* 6, 75. doi:10.3390/microorganisms6030075.

Gomori, G. (1955). "Preparation of buffers for use in enzyme studies," in *Methods Enzymol*, 138–146.

doi:10.1016/0076-6879(55)01020-3.

Gruninger, R. J., McAllister, T. A., and Forster, R. J. (2016). Bacterial and archaeal diversity in the gastrointestinal tract of the north american beaver (*Castor canadensis*). *PLoS One* 11, e0156457. doi:10.1371/journal.pone.0156457.

Haahtela, K., Kari, K., and Sundman, V. (1983). Nitrogenase activity (acetylene reduction) of root-associated, cold-climate A*zospirillum*, *Enterobacter*, *Klebsiella*, and *Pseudomonas* species during growth on various carbon sources and at various partial pressures of oxygen. *Appl. Environ. Microbiol.* 45, 563–70. doi:10.1177/1359104503008001012.

Hasegawa, K., Stewart, C. J., Mansbach, J. M., Linnemann, R. W., Ajami, N. J., Petrosino, J. F., et al. (2017). Sphingolipid metabolism potential in fecal microbiome and bronchiolitis in infants: a case-control study. *BMC Res. Notes* 10, 325. doi:10.1186/s13104-017-2659-9.

Hatamoto, M., Kaneshige, M., Nakamura, A., and Yamaguchi, T. (2014). *Bacteroides luti* sp. nov., an anaerobic, cellulolytic and xylanolytic bacterium isolated from methanogenic sludge. *Int. J. Syst. Evol. Microbiol.* 64, 1770–4. doi:10.1099/ijs.0.056630-0.

Hazlewood, G. P., Laurie, J. I., Ferreira, L. M. A., and Gilbert, H. J. (1992). *Pseudomonas fluorescens* subsp. cellulosa : an alternative model for bacterial cellulase. *J. Appl. Bacteriol.* 72, 244–251. doi:10.1111/j.1365-2672.1992.tb01830.x.

Hemsworth, G. R., Henrissat, B., Davies, G. J., and Walton, P. H. (2014). Discovery and characterization of a new family of lytic polysaccharide monooxygenases. *Nat. Chem. Biol.* 10, 122–126. doi:10.1038/nchembio.1417.

Henderson, G., Cox, F., Kittelmann, S., Miri, V. H., Zethof, M., Noel, S. J., et al. (2013). Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS One* 8, e74787. doi:10.1371/journal.pone.0074787.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., and Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci.* 92, 7090–7094. doi:10.1073/pnas.92.15.7090.

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 (80-.).* 331, 463–467. doi:10.1126/science.1200387.

Himmel, M. E., Xu, Q., Luo, Y., Ding, S.-Y., Lamed, R., and Bayer, E. A. (2010). Microbial enzyme systems for biomass conversion: emerging paradigms. *Biofuels* 1, 323–341. doi:10.4155/bfs.09.25.

Hood, E. E. (2016). Plant-based biofuels. *F1000Research* 5, 1–9. doi:10.12688/f1000research.7418.1.

Hooper, L. V (2001). Commensal host-bacterial relationships in the gut. *Science (80-.).* 292, 1115–1118. doi:10.1126/science.1058709.

Hoover, W. H., and Clarke, S. D. (1972). Fiber digestion in the beaver. *J. Nutr.* 102, 9–15.

Huang, S., Sheng, P., and Zhang, H. (2012). Isolation and identification of cellulolytic bacteria from the gut of *Holotrichia parallela* Larvae (Coleoptera: Scarabaeidae). *Int. J. Mol. Sci.* 13, 2563–2577.

doi:10.3390/ijms13032563.

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., et al. (2016). MEGAN Community Edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLOS Comput. Biol.* 12, e1004957. doi:10.1371/journal.pcbi.1004957.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., et al. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. doi:10.1038/nature11234.

Ingala, M. R., Simmons, N. B., Wultsch, C., Krampis, K., Speer, K. A., and Perkins, S. L. (2018). Comparing microbiome sampling methods in a wild mammal: fecal and intestinal samples record different signals of host ecology, evolution. *Front. Microbiol.* 9, 1–13. doi:10.3389/fmicb.2018.00803.

Jain, P. C. (1993). Greenhouse effect and climate change: scientific basis and overview. *Renew. Energy* 3, 403–420. doi:10.1016/0960-1481(93)90108-S.

Jami, E., Israel, A., Kotser, A., and Mizrahi, I. (2013). Exploring the bovine rumen bacterial community from birth to adulthood. *ISME J.* 7, 1069–79. doi:10.1038/ismej.2013.2.

Jami, E., and Mizrahi, I. (2012). Composition and similarity of bovine rumen microbiota across individual animals. *PLoS One* 7, e33306. doi:10.1371/journal.pone.0033306.

Jang, H. B., Choi, M., Kang, J. H., Park, S. I., and Lee, H. (2017). Association of dietary patterns with the fecal microbiota in Korean adolescents. 1–11. doi:10.1186/s40795-016-0125-z.

Jiao, N., and Zheng, Q. (2011). The microbial carbon pump: from genes to ecosystems. *Appl. Environ. Microbiol.* 77, 7439–7444. doi:10.1128/AEM.05640-11.

Joynson, R., Pritchard, L., Osemwekha, E., and Ferry, N. (2017). Metagenomic analysis of the gut microbiome of the common black slug *Arion ater* in search of novel lignocellulose degrading enzymes. *Front. Microbiol.* 8, 2181. doi:10.3389/fmicb.2017.02181.

Jun, J. W., Giri, S. S., Kim, H. J., Yun, S. K., Chi, C., Chai, J. Y., et al. (2016). Bacteriophage application to control the contaminated water with Shigella. *Sci. Rep.* 6, 22636. doi:10.1038/srep22636.

Kaiser, K., Wemheuer, B., Korolkow, V., Wemheuer, F., Nacke, H., Schöning, I., et al. (2016). Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. *Sci. Rep.* 6, 33696. doi:10.1038/srep33696.

Kakiuchi, M., Isui, A., Suzuki, K., Fujino, T., Fujino, E., Kimura, T., et al. (1998). Cloning and DNA sequencing of the genes encoding *Clostridium josui* scaffolding protein CipA and cellulase CelD and identification of their gene products as major components of the cellulosome. *J. Bacteriol.* 180, 4303–8. doi:10.1038/emm.2003.11.

Kang, Q., Appels, L., Tan, T., and Dewil, R. (2014). Bioethanol from lignocellulosic biomass: current findings determine research priorities. *Sci. World J.* 2014, 1–13. doi:10.1155/2014/298153.

Khatri, N., and Tyagi, S. (2015). Influences of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas. *Front. Life Sci.* 8, 23–39.

doi:10.1080/21553769.2014.933716.

Kim, M., Morrison, M., and Yu, Z. (2011). Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiol. Ecol.* 76, 49–63. doi:10.1111/j.1574-6941.2010.01029.x.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41, 1–11. doi:10.1093/nar/gks808.

Kohl, K. D., Miller, A. W., Marvin, J. E., Mackie, R., and Dearing, M. D. (2014). Herbivorous rodents (*Neotoma* spp.) harbour abundant and active foregut microbiota. *Environ. Microbiol.* 16, 2869–2878. doi:10.1111/1462-2920.12376.

Koike, S., and Kobayashi, Y. (2001). Development and use of competitive PCR assays for the rumen cellulolytic bacteria: *Fibrobacter succinogenes*, *Ruminococcus albus* and *Ruminococcus flavefaciens*. *FEMS Microbiol. Lett.* 204, 361–366. doi:10.1111/j.1574-6968.2001.tb10911.x.

Kong, F., Zhao, J., Han, S., Zeng, B., Yang, J., Si, X., et al. (2014). Characterization of the gut microbiota in the red panda (*Ailurus fulgens*). *PLoS One* 9, e87885. doi:10.1371/journal.pone.0087885.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054.

Kundig, W., Ghosh, S., and Roseman, S. (1964). Phosphate bound to histidine in a protein as an intermediate in a novel phospho-transferase system. *Proc. Natl. Acad. Sci. U. S. A.* 52, 1067–74. doi: 10.1073/pnas.52.4.1067.

Lagkouvardos, I., Pukall, R., Abt, B., Foesel, B. U., Meier-Kolthoff, J. P., Kumar, N., et al. (2016). The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nat. Microbiol.* 1, 16131. doi:10.1038/nmicrobiol.2016.131.

Lambertz, C., Ece, S., Fischer, R., and Commandeur, U. (2016). Progress and obstacles in the production and application of recombinant lignin-degrading peroxidases. *Bioengineered* 7, 145–154. doi:10.1080/21655979.2016.1191705.

Lamed, R., Naimark, J., Morgenstern, E., and Bayer, E. A. (1987). Specialized cell surface structures in cellulolytic bacteria. *J. Bacteriol.* 169, 3792–3800. doi:10.1128/jb.169.8.3792-3800.1987.

Lamed, R., Setter, E., Kenig, R., and Bayer, E. A. (1983). The cellulosome: a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. *Biotechnol. Prog.* 13, 163–181.

Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–20. doi:10.1038/nrg2719.

Lee R. Lynd, Paul J. Weimer, Willem H. van Zyl, I. S. P. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* 66, 506–577. doi:10.1128/MMBR.66.3.506.

Lejeune, A., Colson, C., and Eveleigh, D. E. (1986). Cloning of an endoglucanase gene from

Pseudomonas fluorescens var.cellulosa into *Escherichia coli* and *Pseudomonas fluorescens*. *J. Ind. Microbiol.* 1, 79–86. doi:10.1007/BF01569315.

Lemoine, F., Domelevo Entfellner, J.-B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., et al. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 556, 452–456. doi:10.1038/s41586-018-0043-0.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi:10.1093/nar/gkw290.

Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., et al. (2008). Evolution of mammals and their gut microbes. *Science (80-.).* 320, 1647–1651. doi:10.1126/science.1155725.

Li, J., Liu, C., Li, B., Yuan, H., Yang, J., and Zheng, B. (2012a). Identification and molecular characterization of a novel DyP-type peroxidase from *Pseudomonas aeruginosa* PKE117. *Appl. Biochem. Biotechnol.* 166, 774–785. doi:10.1007/s12010-011-9466-x.

Li, L., Abu Al-Soud, W., Bergmark, L., Riber, L., Hansen, L. H., Magid, J., et al. (2013). Investigating the diversity of *Pseudomonas* spp. in soil using culture dependent and independent techniques. *Curr. Microbiol.* 67, 423–430. doi:10.1007/s00284-013-0382-x.

Li, R. W., Connor, E. E., Li, C., Baldwin Vi, R. L., and Sparks, M. E. (2012b). Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools. *Environ. Microbiol.* 14, 129–139. doi:10.1111/j.1462-2920.2011.02543.x.

Li, Y., Guo, W., Han, S., Kong, F., Wang, C., Li, D., et al. (2015). The evolution of the gut microbiota in the giant and the red pandas. *Sci. Rep.* 5, 10185. doi:10.1038/srep10185.

Lilburn, Schmidt, and Breznak (1999). Phylogenetic diversity of termite gut spirochaetes. *Environ. Microbiol.* 1, 331–345. doi:10.1046/j.1462-2920.1999.00043.x.

Lipski, A., Herve, M., Lombard, V., Nurizzo, D., Mengin-Lecreulx, D., Bourne, Y., et al. (2015). Structural and biochemical characterization of the -N-acetylglucosaminidase from *Thermotoga maritima*: Toward rationalization of mechanistic knowledge in the GH73 family. *Glycobiology* 25, 319–330. doi:10.1093/glycob/cwu113.

Littlewood, J., Guo, M., Boerjan, W., and Murphy, R. J. (2014). Bioethanol from poplar: a commercially viable alternative to fossil fuel in the European Union. *Biotechnol. Biofuels* 7, 113. doi:10.1186/1754-6834-7-113.

Liu, Y.-J., Liu, S., Dong, S., Li, R., Feng, Y., and Cui, Q. (2018). Determination of the native features of the exoglucanase Cel48S from *Clostridium thermocellum*. *Biotechnol. Biofuels* 11, 6. doi:10.1186/s13068-017-1009-4.

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

lone M.A , wani M.R , Bhat N.A , Sheikh S.A, R. M. . (2012). Evaluation of cellulase enzyme secreted by some common and stirring rhizosphere fungi of *Juglans regia* L. by DNS method. *J. Enzym. reserach* 3, 18–22.

López-Mondéjar, R., Zühlke, 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.

Machovič, M., and Janeček, Š. (2008). Domain evolution in the GH13 pullulanase subfamily with focus on the carbohydrate-binding module family 48. *Biologia (Bratisl).* 63. doi:10.2478/s11756-008-0162-4.

Macpherson, A. J., Geuking, M. B., and McCoy, K. D. (2011). Immunoglobulin A: a bridge between innate and adaptive immunity. *Curr. Opin. Gastroenterol.* 27, 529–33. doi:10.1097/MOG.0b013e32834bb805.

Malherbe, S., and Cloete, T. E. (2002). Lignocellulose biodegradation: Fundamentals and applications. *Rev. Environ. Sci. Bio/Technology* 1, 105–114. doi:10.1023/A:1020858910646.

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., et al. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 39, D225–D229. doi:10.1093/nar/gkq1189.

Marotz, C. A., Sanders, J. G., Zuniga, C., Zaramela, L. S., Knight, R., and Zengler, K. (2018). Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome* 6, 42. doi:10.1186/s40168-018-0426-3.

Marynowska, M., Goux, X., Sillam-Dussès, D., Rouland-Lefèvre, C., Roisin, Y., Delfosse, P., et al. (2017). Optimization of a metatranscriptomic approach to study the lignocellulolytic potential of the higher termite gut microbiome. *BMC Genomics* 18, 681. doi:10.1186/s12864-017-4076-9.

Masai, E., Ichimura, A., Sato, Y., Miyauchi, K., Katayama, Y., and Fukuda, M. (2003). Roles of the enantioselective glutathione s-transferases in cleavage of -aryl ether. *J. Bacteriol.* 185, 1768–1775. doi:10.1128/JB.185.6.1768-1775.2003.

Matsui, H., Ogata, K., Tajima, K., Nakamura, M., Nagamine, T., Aminov, R. I., et al. (2000). Phenotypic characterization of polysaccharidases produced by four *Prevotella* type strains. *Curr. Microbiol.* 41, 45–49. doi:10.1007/s002840010089.

Menéndez, E., Ramírez-Bahena, M. H., Fabryová, A., Igual, J. M., Benada, O., Mateos, P. F., et al. (2015). *Pseudomonas coleopterorum* sp. nov., a cellulase-producing bacterium isolated from the bark beetle *Hylesinus fraxini*. *Int. J. Syst. Evol. Microbiol.* 65, 2852–8. doi:10.1099/ijs.0.000344.

Mikheenko, A., Saveliev, V., and Gurevich, A. (2016). MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics* 32, 1088–1090. doi:10.1093/bioinformatics/btv697.

Miller, D. A., Suen, G., Bruce, D., Copeland, A., Cheng, J. F., Detter, C., et al. (2011). Complete genome sequence of the cellulose-degrading bacterium *Cellulosilyticum lentocellum*. *J. Bacteriol.* 193, 2357–2358. doi:10.1128/JB.00239-11.

Mithöfer, A., and Boland, W. (2012). Plant defense against herbivores: chemical aspects. *Annu. Rev. Plant Biol.* 63, 431–450. doi:10.1146/annurev-arplant-042110-103854.

Mizutani, K., Sakka, M., Kimura, T., and Sakka, K. (2014). Essential role of a family-32 carbohydrate-binding module in substrate recognition by *Clostridium thermocellum* mannanase Ct Man5A. *FEBS Lett.* 588, 1726–1730. doi:10.1016/j.febslet.2014.03.022.

Montgomery, L., and Macy, J. M. (1982). Characterization of rat cecum cellulolytic bacteria. *Appl. Environ. Microbiol.* 44, 1435–43. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC242207/.

Moore, E. R. B., Tindall, B. J., Martins Dos Santos, V. A. P., Pieper, D. H., Ramos, J.-L., and Palleroni, N. J. (2006). "Nonmedical: Pseudomonas," in *The Prokaryotes* (New York, NY: Springer New York), 646–703. doi:10.1007/0-387-30746-X_21.

Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., Gonzalez, A., Fontana, L., et al. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science (80-.).* 332, 970–974. doi:10.1126/science.1198719.

Muktham, R., K. Bhargava, S., Bankupalli, S., and S. Ball, A. (2016). A review on 1st and 2nd generation bioethanol production-recent progress. *J. Sustain. Bioenergy Syst.* 06, 72–92. doi:10.4236/jsbs.2016.63008.

Na, G., Lu, Z., Gao, H., Zhang, L., Li, Q., Li, R., et al. (2018). The effect of environmental factors and migration dynamics on the prevalence of antibiotic-resistant *Escherichia coli* in estuary environments. *Sci. Rep.* 8, 1663. doi:10.1038/s41598-018-20077-x.

Nacke, H., Engelhaupt, M., Brady, S., Fischer, C., Tautzt, J., and Daniel, R. (2012). Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. *Biotechnol. Lett.* 34, 663–675. doi:10.1007/s10529-011-0830-2.

Nakamura, A. M., Nascimento, A. S., and Polikarpov, I. (2017). Structural diversity of carbohydrate esterases. *Biotechnol. Res. Innov.* 1, 35–51. doi:10.1016/j.biori.2017.02.001.

Nelson, T. M., Rogers, T. L., and Brown, M. V (2013). The gut bacterial community of mammals from marine and terrestrial habitats. *PLoS One* 8, e83655. doi:10.1371/journal.pone.0083655.

Nolet, B. A., and Rosell, F. (1998). Comeback of the beaver Castor fiber: An overview of old and new conservation problems. *Biol. Conserv.* 83, 165–173. doi:10.1016/S0006-3207(97)00066-9.

Nölling, J., Breton, G., Omelchenko, M. V, Makarova, K. S., Zeng, Q., Gibson, R., et al. (2001). Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum.* *J. Bacteriol.* 183, 4823–38. doi:10.1128/JB.183.16.4823-4838.2001.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi:10.1101/gr.213959.116.

Nutt, A., Sild, V., Pettersson, G., and Johansson, G. (1998). Progress curves. A mean for functional classification of cellulases. *Eur. J. Biochem.* 258, 200–206. doi:10.1046/j.1432-

1327.1998.2580200.x.

Ohta, M., Ohnishi, T., Ioannou, Y. A., Hodgson, M. E., Matsuura, F., and Desnick, R. J. (2000). Human -N-acetylgalactosaminidase: site occupancy and structure of N-linked oligosaccharides. *Glycobiology* 10, 251–261. doi:10.1093/glycob/10.3.251.

Onaga, S., and Taira, T. (2008). A new type of plant chitinase containing LysM domains from a fern (Pteris ryukyuensis): Roles of LysM domains in chitin binding and antifungal activity. *Glycobiology* 18, 414–423. doi:10.1093/glycob/cwn018.

Ormerod, K. L., Wood, D. L. A., Lachner, N., Gellatly, S. L., Daly, J. N., Parsons, J. D., et al. (2016). Genomic characterization of the uncultured Bacteroidales family S24-7 inhabiting the guts of homeothermic animals. *Microbiome* 4, 36. doi:10.1186/s40168-016-0181-2.

Ouwerkerk, J. P., Aalvink, S., Belzer, C., and de Vos, W. M. (2016). *Akkermansia glycaniphila* sp. nov., an anaerobic mucin-degrading bacterium isolated from reticulated python faeces. *Int. J. Syst. Evol. Microbiol.* 66, 4614–4620. doi:10.1099/ijsem.0.001399.

Owusu, P. A., and Asumadu-Sarkodie, S. (2016). A review of renewable energy sources, sustainability issues and climate change mitigation. *Cogent Eng.* 3, 1–14. doi:10.1080/23311916.2016.1167990.

Pauly, M., and Keegstra, K. (2008). Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J.* 54, 559–568. doi:10.1111/j.1365-313X.2008.03463.x.

Peix, A., Ramírez-Bahena, M.-H., and Velázquez, E. (2009). Historical evolution and current status of the taxonomy of genus *Pseudomonas*. *Infect. Genet. Evol.* 9, 1132–1147. doi:10.1016/j.meegid.2009.08.001.

Perera, F. P. (2017). Multiple threats to child health from fossil fuel combustion: impacts of air pollution and climate change. *Environ. Health Perspect.* 125, 141–148. doi:10.1289/EHP299.

Perestelo, F., Falcón, M. A., Pérez, M. L., Corominas Roig, E., and de la Fuente Martin, G. (1989). Bioalteration of kraft pine lignin by *Bacillus megaterium* isolated from compost piles. *J. Ferment. Bioeng.* 68, 151–153. doi:10.1016/0922-338X(89)90066-4.

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–6. doi:10.1038/nmeth.1701.

Piddock, L. J. V. (2006). Multidrug-resistance efflux pumps - not just for resistance. *Nat. Rev. Microbiol.* 4, 629–36. doi:10.1038/nrmicro1464.

Pino, M. S., Rodríguez-Jasso, R. M., Michelin, M., Flores-Gallegos, A. C., Morales-Rodriguez, R., Teixeira, J. A., et al. (2018). Bioreactor design for enzymatic hydrolysis of biomass under the biorefinery concept. *Chem. Eng. J.* 347, 119–136. doi:10.1016/j.cej.2018.04.057.

Pohlschröder, M., Canale-Parola, E., and Leschine, S. B. (1995). Ultrastructural diversity of the cellulase complexes of *Clostridium papyrosolvens* C7. *J. Bacteriol.* 177, 6625–9. doi:10.1128/jb.177.22.6625-6629.1995.

Potrykus, J., Mahaney, B., White, R. L., and Bearne, S. L. (2007). Proteomic investigation of glucose metabolism in the butyrate-producing gut anaerobe *Fusobacterium varium*. *Proteomics* 7, 1839–

53. doi:10.1002/pmic.200600464.

Probert, W. S., Miller, G. M., and Ledin, K. E. (2017). Contaminated stream water as source for *Escherichia coli* O157 illness in children. *Emerg. Infect. Dis.* 23, 1216–1218. doi:10.3201/eid2307.170226.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41, 590–596. doi:10.1093/nar/gks1219.

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.

Ramakrishna, B. S. (2013). Role of the gut microbiota in human nutrition and metabolism. *J. Gastroenterol. Hepatol.* 28, 9–17. doi:10.1111/jgh.12294.

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, 1–13. doi:10.1128/mSphere.00300-17.

Ratanakhanokchai, K., Waeonukul, R., Pason, P., Tachaapaikoon, C., Lay, K., Sakka, K., et al. (2013). *Paenibacillus curdlanolyticus* strain B-6 multienzyme complex: a novel system for biomass utilization. *Biomass Now - Cultivation and Utilization* (InTech), 64. doi:10.5772/51820.

Riordan, T. (2007). Human infection with *Fusobacterium necrophorum* (Necrobacillosis), with a focus on Lemierre's syndrome. *Clin. Microbiol. Rev.* 20, 622–659. doi:10.1128/CMR.00011-07.

Robert, C., Chassard, C., Lawson, P. A., and Bernalier-Donadille, A. (2007). *Bacteroides cellulosilyticus* sp. nov., a cellulolytic bacterium from the human gut microbial community. *Int. J. Syst. Evol. Microbiol.* 57, 1516–1520. doi:10.1099/ijs.0.64998-0.

Robinet, J., Minella, J. P. G., de Barros, C. A. P., Schlesner, A., Lücke, A., Ameijeiras-Mariño, Y., et al. (2018). Impacts of forest conversion and agriculture practices on water pathways in Southern Brazil. *Hydrol. Process.* 32, 2304–2317. doi:10.1002/hyp.13155.

Rodríguez, J. M., Murphy, K., Stanton, C., Ross, R. P., Kober, O. I., Juge, N., et al. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb. Ecol. Health Dis.* 26, 26050. doi:10.3402/mehd.v26.26050.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584. doi:10.7717/peerj.2584.

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, 56. doi:10.1186/s40168-015-0118-1.

Roy, A., Yang, J., and Zhang, Y. (2012). COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res.* 40, W471–W477. doi:10.1093/nar/gks372.

Rubin, E. M. (2008). Genomics of cellulosic biofuels. *Nature* 454, 841–845. doi:10.1038/nature07190.

Sadhu, S. (2013). Cellulase production by bacteria: a review. *Br. Microbiol. Res. J.* 3, 235–258. doi:10.9734/BMRJ/2013/2367.

Saier, M. H. (2015). The bacterial phosphotransferase system: new frontiers 50 years after its discovery. *J. Mol. Microbiol. Biotechnol.* 25, 73–8. doi:10.1159/000381215.

Sánchez-Sánchez, H., and Morquecho-Contreras, A. (2017). "Chemical plant defense against herbivores," in *Herbivores* (InTech), 64. doi:10.5772/67346.

Sanderson, K. (2011). Lignocellulose: A chewy problem. *Nature* 474, S12–S14. doi:10.1038/474S012a.

Santos, A., Mendes, S., Brissos, V., and Martins, L. O. (2014). New dye-decolorizing peroxidases from *Bacillus subtilis* and *Pseudomonas putida* MET94: towards biotechnological applications. *Appl. Microbiol. Biotechnol.* 98, 2053–2065. doi:10.1007/s00253-013-5041-4.

Santos, C. R., Paiva, J. H., Sforça, M. L., Neves, J. L., Navarro, R. Z., Cota, J., et al. (2012). Dissecting structure–function–stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168. *Biochem. J.* 441, 95–104. doi:10.1042/BJ20110869.

Savage, D. C. (1977). Microbial ecology of the gastrointestinal tract. *Annu. Rev. Microbiol.* 31, 107–133. doi:10.1146/annurev.mi.31.100177.000543.

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

Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593. doi:10.1093/bioinformatics/btq706.

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, 1–22. doi:10.1371/journal.pone.0073827.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi:10.1093/bioinformatics/btu153.

Singh, R. K., Chang, H.-W., Yan, D., Lee, K. M., Ucmak, D., Wong, K., et al. (2017). Influence of diet on the gut microbiome and implications for human health. *J. Transl. Med.* 15, 73. doi:10.1186/s12967-017-1175-y.

Singh, V., and Kumar, A. (1998). Production and purification of an extracellular cellulase from *Bacillus brevis* vs-1. *IUBMB Life* 45, 443–452. doi:10.1080/15216549800202832.

Sravanthi, T., Ramana, C. V., Tushar, L., and Sasikala, C. (2015). *Spirochaeta odontotermitis* sp. nov., an obligately anaerobic, cellulolytic, halotolerant, alkaliphilic spirochaete isolated from the termite *Odontotermes obesus* (Rambur) gut. *Int. J. Syst. Evol. Microbiol.* 65, 4589–4594. doi:10.1099/ijsem.0.000616.

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 -amylase-related proteins. *Protein Eng. Des. Sel.* 19, 555–562. doi:10.1093/protein/gzl044.

Tajima, K., Arai, S., Ogata, K., Nagamine, T., Matsui, H., Nakamura, M., et al. (2000). Rumen bacterial

community transition during adaptation to high-grain diet. *Anaerobe* 6, 273–284. doi:10.1006/anae.2000.0353.

Tanabe, T., Morinaga, K., Fukamizo, T., and Mitsutomi, M. (2003). Novel chitosanase from *Streptomyces griseus* HUT 6037 with transglycosylation activity. *Biosci. Biotechnol. Biochem.* 67, 354–364. doi:10.1271/bbb.67.354.

Tessler, M., Neumann, J. S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L. F. M., et al. (2017). Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Sci. Rep.* 7, 6589. doi:10.1038/s41598-017-06665-3.

Turner, N. A., Sharma-Kuinkel, B. K., Maskarinec, S. A., Eichenberger, E. M., Shah, P. P., Carugati, M., et al. (2019). Methicillin-resistant *Staphylococcus aureus*: an overview of basic and clinical research. *Nat. Rev. Microbiol.* 17, 203–218. doi:10.1038/s41579-018-0147-4.

Ursell, L. K., Metcalf, J. L., Parfrey, L. W., and Knight, R. (2012). Defining the human microbiome. *Nutr. Rev.* 70 Suppl 1, S38-44. doi:10.1111/j.1753-4887.2012.00493.x.

van Bloois, E., Torres Pazmiño, D. E., Winter, R. T., and Fraaije, M. W. (2010). A robust and extracellular heme-containing peroxidase from *Thermobifida fusca* as prototype of a bacterial peroxidase superfamily. *Appl. Microbiol. Biotechnol.* 86, 1419–1430. doi:10.1007/s00253-009-2369-x.

van der Maarel, M. J. E. ., van der Veen, B., Uitdehaag, J. C. ., Leemhuis, H., and Dijkhuizen, L. (2002). Properties and applications of starch-converting enzymes of the α-amylase family. *J. Biotechnol.* 94, 137–155. doi:10.1016/S0168-1656(01)00407-2.

Vecherskii, M. V., Naumova, E. I., Kostina, N. V., and Umarov, M. M. (2009). Assimilation of biological nitrogen by European beaver. *Biol. Bull.* 36, 92–95. doi:10.1134/S1062359009010142.

Visweswaran, G. R. R., Steen, A., Leenhouts, K., Szeliga, M., Ruban, B., Hesseling-Meinders, A., et al. (2013). AcmD, a homolog of the major autolysin AcmA of *Lactococcus lactis*, binds to the cell wall and contributes to cell separation and autolysis. *PLoS One* 8, e72167. doi:10.1371/journal.pone.0072167.

Vollmer, W., Joris, B., Charlier, P., and Foster, S. (2008). Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol. Rev.* 32, 259–286. doi:10.1111/j.1574-6976.2007.00099.x.

Wang, F., Li, F., Chen, G., and Liu, W. (2009). Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. *Microbiol. Res.* 164, 650–657. doi:10.1016/j.micres.2008.12.002.

Wang, J., Fan, H., Han, Y., Zhao, J., and Zhou, Z. (2017). Characterization of the microbial communities along the gastrointestinal tract of sheep by 454 pyrosequencing analysis. *Asian-Australasian J. Anim. Sci.* 30, 100–110. doi:10.5713/ajas.16.0166.

Wang, Y., and McAllister, T. A. (2002). Rumen microbes, enzymes and feed digestion-a review. *Asian-Australasian J. Anim. Sci.* 15, 1659–1676. doi:10.5713/ajas.2002.1659.

Wemheuer, B., Wemheuer, F., Meier, D., Billerbeck, S., Giebel, H.-A., Simon, M., et al. (2017). Linking

compositional and functional predictions to decipher the biogeochemical significance in DFAA turnover of abundant bacterioplankton lineages in the north sea. *microorganisms* 5, 68. doi:10.3390/microorganisms5040068.

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., et al. (2007). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 36, D13–D21. doi:10.1093/nar/gkm1000.

Wi, S. G., Cho, E. J., Lee, D.-S., Lee, S. J., Lee, Y. J., and Bae, H.-J. (2015). Lignocellulose conversion for biofuel: a new pretreatment greatly improves downstream biocatalytic hydrolysis of various lignocellulosic materials. *Biotechnol. Biofuels* 8, 228. doi:10.1186/s13068-015-0419-4.

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, 1–14. doi:10.3389/fmicb.2017.02504.

Wong, M. T., Wang, W., 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, 1–13. doi:10.3389/fmicb.2016.00961.

Wu, M., Guina, T., Brittnacher, M., Nguyen, H., Eng, J., and Miller, S. I. (2005). The *Pseudomonas aeruginosa* proteome during anaerobic growth. *J. Bacteriol.* 187, 8185–8190. doi:10.1128/JB.187.23.8185-8190.2005.

Xiao, Z., Zhang, X., Gregg, D. J., and Saddler, J. N. (2004). Effects of sugar inhibition on cellulases and β-glucosidase during enzymatic hydrolysis of softwood substrates. *Appl. Biochem. Biotechnol.* 115, 1115–1126. doi:10.1385/ABAB:115:1-3:1115.

Xie, Z., Dou, Y., Ping, S., Chen, M., Wang, G., Elmerich, C., et al. (2006). Interaction between NifL and NifA in the nitrogen-fixing *Pseudomonas stutzeri* A1501. *Microbiology* 152, 3535–3542. doi:10.1099/mic.0.29171-0.

Xu, G.-D., Cai, L., Ni, Y.-S., Tian, S.-Y., Lu, Y.-Q., Wang, L.-N., et al. (2018). Comparisons of effects on intestinal short-chain fatty acid concentration after exposure of two glycosidase inhibitors in mice. *Biol. Pharm. Bull.* 41, 1024–1033. doi:10.1248/bpb.b17-00978.

Xue, Z., Zhang, W., Wang, L., Hou, R., Zhang, M., Fei, L., et al. (2015). The bamboo-eating giant panda harbors a carnivore-like gut microbiota, with excessive seasonal variations. *MBio* 6, 1–12. doi:10.1128/mBio.00022-15.

Yamane, K., and Suzuki, H. (1988). Cellulases of *Pseudomonas fluorescens* var. cellulosa. in, 200–210. doi:10.1016/0076-6879(88)60121-2.

Yamane, K., Yoshikawa, T., Suzuki, H., and Nisizawa, K. (1971). Localization of cellulase components in *Pseudomonas fluorescens* var. cellulosa. *J. Biochem.* 69, 771–80. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/4995445.

Yan, S., and Wu, G. (2013). Secretory pathway of cellulase: a mini-review. *Biotechnol. Biofuels* 6, 177. doi:10.1186/1754-6834-6-177.

Yan, Y., Yang, J., Dou, Y., Chen, M., Ping, S., Peng, J., et al. (2008). Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated *Pseudomonas stutzeri* A1501. *Proc. Natl. Acad. Sci.* 105, 7564–7569. doi:10.1073/pnas.0801093105.

Yang, J., and Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* 43, W174–W181. doi:10.1093/nar/gkv342.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40, W445–W451. doi:10.1093/nar/gks479.

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

Zamani, A. (2015). "Introduction to Lignocellulose-based Products," in, 1–36. doi:10.1007/978-3-319-14033-9_1.

Zhang, Y. (2009). I-TASSER: Fully automated protein structure prediction in CASP8. *Proteins Struct. Funct. Bioinforma.* 77, 100–113. doi:10.1002/prot.22588.

Zhu, L., Wu, Q., Dai, J., Zhang, S., and Wei, F. (2011). Evidence of cellulose metabolism by the giant panda gut microbiome. *Proc. Natl. Acad. Sci. U. S. A.* 108, 17714–9. doi:10.1073/pnas.1017956108.

Zielińska, S., Radkowski, P., Blendowska, A., Ludwig-Gałęzowska, A., Łoś, J. M., and Łoś, M. (2017). The choice of the DNA extraction method may influence the outcome of the soil microbial community structure analysis. *Microbiologyopen* 6, e00453. doi:10.1002/mbo3.453.

7. APPENDICES
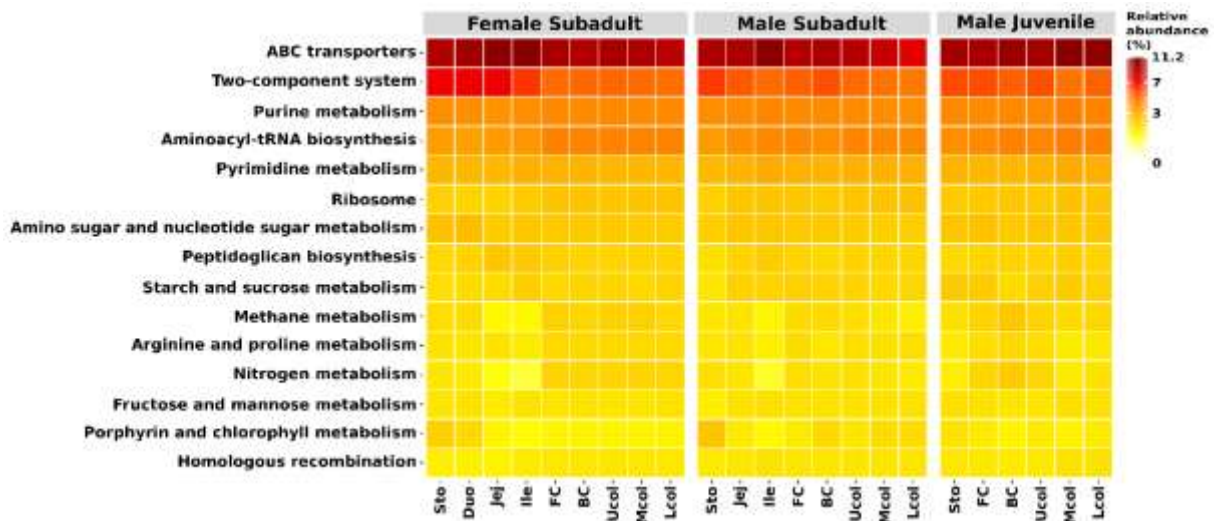
Appendix 1. The relative abundance of top 15 predicted metabolic functions derived from OTUs using Tax4Fun

	Eurasian Beaver						NAgrun Beaver				Bovine					Higher Termite											Lower Termite							
	FS FC	FS BC	MS FC	MS BC	MJ FC	MJ BC	Female 1	Female 2	Male 1	Male 2	Bovine 1D	Bovine 3D	Bovine 2M	Bovine 6M	Bovine 2Y	A. meridionalis	A. trestus	C. ugandensis	M. subhyalinus	Macrotermes sp.	Microcerotermes sp.	N. corniger	N. takasagoensis	Odontotermes sp.	Ophiotermes sp.	Trinervitermes sp.	C. niger	H. mossambicus	H. sjoestedti	I. marginipennis	M. darwiniensis	N. jouteli	R. santonensis	Z. nevadensis
Firmicutes	21.8	29.9	47.5	48.5	37.2	26.9	48.1	53.2	43.6	48.9	51.6	11.7	16	28.7	42.1	56.4	41.3	50.2	19.3	27.5	8.3	5.1	9.2	31.4	69.3	23.6	8.6	23.9	21.5	2.2	8	11.1	6.8	16.3
Bacteroidetes	0.3	0.7	4.4	8	28.9	5.1	49.9	44.1	53.7	49.6	11.6	19.9	47.4	59.5	55.3	0.7	24.7	10.4	24.7	37.8	3.9	4.4	5	24.9	6.2	15	66.5	18.2	8.3	20.1	9.1	5.8	1.4	25.3
Spirochaetes	0	0	0	0	0	0	0.2	0.5	0.6	0.1	0	0	0.2	0.2	0.1	13.4	8.6	8.7	0.5	1.8	64.8	69.9	73.5	9.3	4	49.8	11.9	16.8	52.9	49.7	56.9	45.5	64.5	20.9
Actinobacteria	58.9	45.9	23.5	10.6	33.7	67.6	0.4	0.7	0.9	0.6	1.3	1.7	0.3	0.4	0.2	12.1	4.5	3.7	6.7	1.5	2.5	0.9	1.4	3.6	7.2	2.3	0.3	3.6	0.9	3.5	0.4	1.7	1.3	1.5
Proteobacteria	0.1	0.1	0.3	0.2	0.1	0.3	0.8	0.7	0.5	0.5	35.3	59.3	35.5	10.1	1.9	5.4	7.2	19.5	18	5	4	2.6	2.2	9	6.8	3.7	1.9	6.1	7.3	18.7	5	11.2	7.1	10.6
Verrucomicrobia	17.7	22.5	15.3	8.6	0	0	0	0.4	0.1	0.1	0	6.8	0	0	0	0.1	0.7	0.8	0.2	0.3	0	0.1	0	1.8	0.4	0	0.1	0.1	0.3	0.1	5.3	0.8	1	6
Elusimicrobia	0	0	0.1	0.1	0	0	0	0	0	0	0	0	0	0	0	0.3	0.7	0.4	0.6	0.6	0.2	0	0	0.4	0.1	0	0.7	24.7	3.2	0.4	7.2	12.1	13.9	6
Patescibacteria	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0.2	0.1	1.4	1.1	0.7	12.4	4.9	0.5	1.2	0.6	5.7	1.9	0.6	3	3.8	1.3	0.6	2.7	2.2	1.8	6.4
Fibrobacteres	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0	2.7	0	0.1	0	1.9	13.2	13.1	7.3	1.3	0.1	2.2	0	0.4	0	0	0	0.8	0	0
Fusobacteria	0.1	0	8.8	23.7	0	0	0	0	0	0	0	0.7	0	0.1	0	0	0	0.3	0	0.1	0	0.1	0	2.2	0	0	5	0	0	0.2	1.1	0	0	0

Relative abundance (%)
73.5
50
30
10
0

Appendix 2. The bacterial community of herbivore group A based on cecum/rumen and gut samples of termites at the phylum level. The top 10 phyla are shown. The color corresponds to the relative abundance.

Appendix 3. The bacterial community of herbivore group A based on cecum/rumen and gut samples of termites at the family level. The top 15 families are shown. The color corresponds to the relative abundance.

Family	FS FC	FS BC	MS FC	MS BC	MJ FC	MJ BC	Female 1	Female 2	Male 1	Male 2	Bovine 1D	Bovine 3D	Bovine 2M	Bovine 6M	Bovine 2Y	A. meridionalis	A. trestus	C. ugandensis	M. subhyalinus	Macrotermes sp	Microcerotermes sp	N. corniger	N. takasagoensis	Odontotermes sp	Ophiotermes sp	Trinervitermes sp	C. niger	H. mossambicus	H. sjoestedti	I. marginipennis	M. darwiniensis	N. jouteli	R. santonensis	Z. nevadensis
Spirochaetaceae	0	0	0	0	0	0	0.2	0.5	0.6	0.1	0	0	0.2	0.2	0.1	13.4	8.5	8.2	0.5	1.8	64.8	69.8	73.4	9.3	3.6	49.8	11.9	16.2	52.7	49.7	56.9	45.5	64.4	20.6
Ruminococcaceae	14.5	19.6	17.4	8.7	16.6	4.3	10.1	10.2	9.3	10.4	0.1	1.1	4.7	9	9.3	36.2	14.6	25.4	5.6	11.4	1.1	2.1	5.1	8.7	31.1	8.8	1.8	7.9	7	0.3	4.3	5	3.2	7.2
Lachnospiraceae	6.9	9.5	12	16.1	11.8	2.5	37.7	42.7	33.9	38.1	0.1	0.7	9.2	10.4	22.1	4.8	6.7	13.1	3.6	4.2	1	0.4	0.7	10.5	14.2	3.8	4.6	0.9	3.4	0.2	0.2	0.3	1.8	4
Prevotellaceae	0.1	0.2	1.3	6.4	6.1	3.3	4.8	4.4	5.6	5.1	0	2.9	47	59	54.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Eggerthellaceae	44.9	33.9	22.5	9.8	25.7	53.7	0.4	0.6	0.9	0.5	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
p-251-o5	0	0	0	0	0	0	27.5	23.1	28	28.1	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bacteroidaceae	0.1	0.2	0.1	0.2	10.6	1.3	13	11.4	13.8	12.1	1.8	13.5	0	0	0	0	0	0.1	0.1	0.8	0	0	0	0.2	0	0	0	0.4	0.3	1	0	0.1	0	22.7
Dysgonomonadaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2	8.3	2.8	0.2	5.3	0.1	0.5	0.5	5.6	3.4	3.2	65.2	0	1.9	0.6	3.2	0.2	0.2	0.1
Rikenellaceae	0	0	0	0	0	0	0.1	0.2	0.1	0.1	0	0	0.1	0.2	0.1	0.2	2.1	2.9	21.9	26.8	2.2	1.2	1.4	12	0.3	3	0.0	4.1	1.7	9.9	0.2	2.5	0.3	0.4
Akkermansiaceae	17.7	22.5	15.3	8.6	0	0	0	0.4	0.1	0.1	0	6.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Endomicrobiaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0.3	0.1	0.2	0.1	0.2	0	0	0.2	0.1	0	0.6	24.7	3.1	0.4	7.2	11.6	13.9	5.8
Streptococcaceae	0	0	0	0	0	0	0	0	0	0	42.8	2.2	0.1	0	0	0.1	1.9	0.2	0.5	0	0.3	0	0	1.7	0	3	0.9	1.9	0.2	0	0	0.1	0	0
Burkholderiaceae	0	0	0	0	0	0	0.1	0.2	0.2	0.1	2.5	45.8	0	0	0	0.3	1.6	0.3	0.3	0.1	0.1	0	0	1	0.3	0.2	0	0.3	0.2	0.1	0	0.1	0	0
Succinivibrionaceae	0	0	0	0	0	0	0	0	0	0	0	0	35.3	9.8	1.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Coriobacteriales Incertae Sedis	13.2	10.8	0	0.1	1.5	2.2	0	0	0	0	0	0	0	0	0	2.2	0.4	1.3	3.2	1.4	0.1	0.2	0.1	0.4	0.4	0.2	0.2	1.9	0.8	0.5	0.2	0.7	1	0.9

Column groups: **Eurasian Beaver** (FS FC, FS BC, MS FC, MS BC, MJ FC, MJ BC); **NA_grus Beaver** (Female 1, Female 2, Male 1, Male 2); **Bovine** (Bovine 1D, Bovine 3D, Bovine 2M, Bovine 6M, Bovine 2Y); **Higher Termite** (A. meridionalis, A. trestus, C. ugandensis, M. subhyalinus, Macrotermes sp, Microcerotermes sp, N. corniger, N. takasagoensis, Odontotermes sp, Ophiotermes sp, Trinervitermes sp); **Lower Termite** (C. niger, H. mossambicus, H. sjoestedti, I. marginipennis, M. darwiniensis, N. jouteli, R. santonensis, Z. nevadensis).

Relative abundance (%): 73.4 / 50 / 30 / 10 / 0

Phylum	Eurasian Beaver									NA$_{grun}$ Beaver				NA$_{wong}$ Bea.			Human										Giant Panda					Red Panda					
	FS Ucol	FS Mcol	FS Lcol	MS Ucol	MS Mcol	MS Lcol	MJ Ucol	MJ Mcol	MJ Lcol	Female 1	Female 2	Male 1	Male 2	1	2	3	Female 1	Female 2	Female 3	Female 4	Female 5	Female 6	Male 1	Male 2	Male 3	Male 4	CG 1	CG 2	CG 3	CG 4	CG 5	CR 1	CR 2	CR 3	CR 4	CR 5	CR 6
Firmicutes	27.1	23.7	26.2	17.3	24.4	16.1	25.8	43.2	39.8	56.6	65.4	53.5	71.9	8.4	7.9	7.3	19.2	24.5	62.9	21.1	50.4	26.9	24.2	11.1	2.1	53.5	97.2	99.9	99.8	99.9	97.8	99.2	99.4	82.2	99	93.5	93.6
Bacteroidetes	0.4	0.2	0.3	3.6	3.5	3.2	36.1	10.4	14.3	39.9	30.9	43.3	25.2	10.6	10.9	12.2	41.6	70.7	33.9	74.9	47.2	68.4	73.7	75	97.7	46.3	0	0	0	0	0	0	0	0.1	0	0	0
Proteobacteria	0.1	0.1	0.1	48.2	0	0	0.1	0.1	0.1	1.1	0.4	1.2	0.9	68	67.7	67.2	24.8	1.1	1.1	2.2	1.5	4.6	2	9.2	0.2	0	2.8	0.1	0.1	0	2.2	0.7	0.2	16.8	0.2	2.1	6.2
Actinobacteria	54.9	53.6	44.2	10.1	11.8	6.3	37.7	30.4	26	0.6	0.5	0.9	1.4	0.5	0.7	0.8	14.4	0.3	2.2	0	0.1	0.1	0	0.5	0	0.1	0	0	0	0	0	0	0	0.4	0	0	0
Verrucomicrobia	16.5	21.7	28.9	9.3	24.8	38.4	0	0	0	0.9	2.1	0.2	0.2	2.2	2	2.4	0	3.5	0	0.2	0.8	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0
Fusobacteria	0.1	0.1	0.1	11.3	35.3	35.5	0.2	15.7	19.7	0	0	0	0	9.3	9.9	9.4	0	0	0	0	0	0	0	4.2	0	0	0	0	0	0	0	0	0	0	0	0	0
Epsilonbacteraeota	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0.3	0.3	0.7	4.4	6.2
Tenericutes	0.9	0.5	0.4	0	0.1	0.1	0.1	0.1	0	0.1	0.1	0.1	0.1	0.5	0.5	0.4	0	0	0	1.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cyanobacteria	0	0	0	0	0.1	0.1	0	0	0	0.5	0.4	0.5	0.2	0.3	0.2	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Spirochaetes	0	0	0	0	0	0	0	0	0	0.3	0.2	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Relative abundance (%): 99.9 / 50 / 30 / 10 / 0

Appendix 4. The bacterial community of herbivore group B based on feces sample at the phylum level. The top 10 phyla are shown. The color corresponds to the relative abundance.

Family	FS Ucol	FS Mcol	FS Lcol	MS Ucol	MS Mcol	MS Lcol	MJ Ucol	MJ Mcol	MJ Lcol	Female 1	Female 2	Male 1	Male 2	Beaver 1	Beaver 2	Beaver 3	Female 1	Female 2	Female 3	Female 4	Female 5	Female 6	Male 1	Male 2	Male 3	Male 4	CG 1	CG 2	CG 3	CG 4	CG 5	CR 1	CR 2	CR 3	CR 4	CR 5	CR 6
	Eurasian Beaver									**NA grun Beaver**				**NA wong Bea.**			**Human**										**Giant Panda**					**Red Panda**					
Bacteroidaceae	0.1	0.1	0.1	0	0.1	0.1	11.3	0.7	0.7	12.3	8.7	15.4	7.3	1.5	1.8	2	37.5	62.3	14.8	44.5	27.8	85.1	57.6	63.7	88.8	7.3	0	0	0	0	0	0	0	0	0	0	0
Clostridiaceae 1	0.1	0.1	0	4.7	6.3	1.7	0.5	4.2	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.8	5.2	2.8	0.7	15.3	57	80.8	73.1	80.1	52.1	76.7
Streptococcaceae	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0.1	0	0.1	0	0	0	0	0	0	91.3	91.4	93.8	94.4	75.9	7.2	0.9	0.1	0	1.2	0
Lachnospiraceae	9.2	6.9	5.5	4.2	6.2	3.2	14.4	22.2	17.5	38.4	40.8	34.8	53.8	0.8	1	1	0	2.2	7.1	6.4	16.3	19.2	8.7	3.5	1.4	13.2	0	0	0	0	0.5	6.9	11.5	5.9	18.8	11.6	16.5
Ruminococcaceae	17.4	16.4	19.9	6.9	9.1	9.1	5.6	4.8	5.5	17.8	24.3	18	17.5	2.1	1.9	1.9	5.1	21.3	39	13.8	32.5	5.5	12.4	0.6	0.5	11.5	0	0	0	0	0	0	0	0	0	0	0.1
Eggerthellaceae	41.8	41.9	34.7	9.1	10.3	5.2	30.5	16.9	17.4	0.5	0.4	0.8	1.2	0.1	0.2	0.2	0	0	0.1	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0
Pseudomonadaceae	0	0	0	48.1	0	0	0	0	0	0	0	0	0	40.8	39.6	41.5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Akkermansiaceae	16.5	21.7	28.9	9.3	24.8	38.4	0	0	0	0.9	2.1	0.2	0.2	2.2	2	2.4	0	3.5	0	0.2	0.9	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0
Fusobacteriaceae	0.1	0.1	0.1	11.3	35.3	35.5	0.2	15.7	19.7	0	0	0	0	9.3	9.9	9.4	0	0	0	0	0	0	0	4.2	0	0	0	0	0	0	0	0	0	0	0	0	0
Muribaculaceae	0.1	0.1	0.1	1.4	1.4	2.6	21.4	8.6	12.3	4	5.7	7.7	3.1	6.8	6.5	7.8	0	0	0	2.1	0	0	0	0	0	10.5	0	0	0	0	0	0	0	0	0	0	0
Lactobacillaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.3	0	0	0	0	0	0	0	0	0	0	1.7	2.3	2.8	0	26.1	0.1	0	0	27.8	0
Burkholderiaceae	0	0	0	0	0	0	0	0	0	0.4	0	0.8	0.6	13.6	13.4	12.3	8.5	1	0	2.1	0.9	4.5	2	0.8	0.2	0	0	0	0	0	0	0	0	0.3	0	0	0.1
p-251-o5	0	0	0	0	0	0	0	0	0	18.8	13.3	15.5	12.1	0.1	0.1	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rikenellaceae	0	0	0	0	0	0	0	0	0	0.1	0.1	0.2	0	0.1	0	0.1	0.5	0	14.2	17.1	11.9	2.7	8.2	1.7	1	0	0	0	0	0	0	0	0	0	0	0	0
Prevotellaceae	0.1	0.1	0.1	2.1	2	0.4	3.4	1	0.4	4.3	2.8	4.2	2.5	0.5	0.7	0.6	0.7	0	3.7	2.4	0	0	0	0	0	20.9	0	0	0	0	0	0	0	0	0	0	0

Relative abundance (%): 94.4 / 50 / 30 / 10 / 0

Appendix 5. The bacterial community of herbivore group B based on feces samples at the family level. The top 15 families are shown. The color corresponds to the relative abundance.

Appendix 6. Identified CAZymes and its relative abundance from Eurasian beaver gut metagenomes. Count was normalized in Ampvis2 package in R

CAZy	FSBC	FSLC	MSBC	MSLC
AA10	0,70	0,53	0,61	0,45
AA12	0,35	0,13	0,46	0,76
AA3	2,11	0,40	1,67	1,97
AA4	0,00	0,13	0,61	0,15
AA6	3,87	2,80	2,58	2,73
AA7	0,35	0,13	0,30	0,45
CBM12	1,41	0,67	0,76	0,61
CBM13	5,28	2,93	2,28	1,21
CBM16	0,00	0,13	0,15	0,00
CBM2	0,00	0,00	0,15	0,15
CBM20	0,00	0,00	0,46	0,00
CBM22	0,00	0,27	0,00	0,00
CBM31	0,00	0,27	0,00	0,00
CBM32	0,35	2,00	2,43	2,27
CBM35	0,00	0,13	0,15	0,30
CBM37	0,00	0,00	0,15	1,36
CBM4	0,70	0,13	0,15	0,00
CBM48	1,06	0,40	1,82	1,82
CBM5	0,00	0,00	0,00	0,15
CBM50	2,11	2,40	3,65	3,18
CBM51	0,00	0,13	0,15	0,45
CBM54	0,00	0,00	0,15	0,00
CBM56	0,00	0,00	0,00	0,15
CBM6	0,00	0,00	0,15	0,00
CBM61	0,00	0,13	0,30	0,00
CBM62	0,00	0,00	0,30	0,00
CBM67	0,00	0,13	0,46	0,00
CBM73	0,00	0,40	0,30	0,30
CBM80	0,00	0,00	0,15	0,00
CBM9	0,00	0,13	0,15	0,15
CE1	4,58	2,13	4,86	3,79
CE10	2,82	1,60	3,50	3,33
CE11	0,00	0,13	0,30	0,30
CE12	0,00	0,80	0,15	0,00
CE14	1,41	0,67	1,22	1,52
CE2	0,00	0,13	0,00	0,45
CE3	0,35	0,67	0,76	0,76
CE4	2,46	2,67	1,67	2,88
CE5	0,70	0,27	0,46	0,45
CE6	0,00	0,40	0,30	0,30
CE7	1,06	0,53	0,15	0,30
CE8	0,00	0,00	0,15	0,15
CE9	0,70	0,67	0,76	0,91
cohesin	0,00	0,40	0,15	0,00
dockerin	0,00	1,20	0,00	0,15
GH1	0,00	0,13	0,30	0,15
GH102	0,70	0,13	0,46	0,61
GH103	1,06	0,00	0,46	0,61

Appendix 6. continued				
GH105	0,00	0,13	0,00	0,00
GH109	1,06	1,33	2,28	1,52
GH110	0,00	0,13	0,15	0,15
GH112	0,00	0,13	0,00	0,00
GH113	0,00	0,00	0,00	0,15
GH114	0,35	0,13	0,30	0,15
GH120	0,00	0,00	0,15	0,00
GH123	0,00	0,13	0,15	0,15
GH125	0,00	0,13	0,00	0,00
GH127	0,00	0,13	0,00	0,15
GH13	2,46	0,93	3,95	2,73
GH130	0,00	0,00	0,00	0,15
GH133	0,00	0,00	0,15	0,00
GH15	0,35	0,53	0,30	0,30
GH16	0,00	0,13	0,15	0,15
GH17	0,00	0,27	0,46	0,45
GH18	0,70	0,27	0,30	0,30
GH19	1,41	0,93	1,22	1,21
GH2	0,00	0,80	0,61	0,61
GH20	0,00	0,53	0,61	0,61
GH23	2,82	1,33	3,04	2,88
GH24	1,06	0,53	0,46	0,61
GH25	0,35	0,93	1,37	0,30
GH27	0,00	0,00	0,15	0,00
GH28	0,00	0,27	0,00	0,30
GH29	0,00	0,27	0,46	0,45
GH3	1,41	0,53	1,52	0,91
GH30	0,00	0,00	0,15	0,15
GH31	0,00	0,67	0,61	0,45
GH32	0,00	0,13	0,00	0,30
GH33	0,00	0,40	0,61	0,61
GH35	0,00	0,27	0,15	0,15
GH36	0,00	0,27	0,46	0,61
GH38	0,00	0,00	0,15	0,00
GH39	0,00	0,00	0,30	0,30
GH4	0,00	0,13	0,00	0,30
GH42	0,00	0,27	0,00	0,00
GH43	0,00	0,27	0,15	0,15
GH5	0,35	0,40	0,76	1,06
GH50	0,35	0,00	0,46	0,30
GH51	0,00	0,13	0,00	0,00
GH53	0,00	0,13	0,00	0,00
GH57	0,00	0,27	0,15	0,15
GH65	0,35	0,00	0,00	0,00
GH73	2,46	1,87	1,67	1,36
GH74	0,35	0,27	0,30	0,76
GH75	0,00	0,13	0,15	0,15
GH77	0,00	0,13	0,76	0,61
GH78	0,00	0,13	0,46	0,00
GH8	0,70	0,13	0,30	0,61
GH84	0,00	0,13	0,15	0,15

Appendix 6. continued				
GH89	0,00	0,13	0,15	0,15
GH94	0,00	0,00	0,00	0,30
GH95	0,00	0,27	0,15	0,30
GH97	0,00	0,00	0,15	0,00
GH99	0,00	0,13	0,00	0,00
GT1	0,00	0,00	0,15	0,15
GT10	0,00	0,40	0,30	0,30
GT101	0,00	0,13	0,15	0,15
GT103	0,00	0,13	0,15	0,15
GT104	0,35	0,00	0,46	0,61
GT11	0,00	0,00	0,15	0,00
GT12	0,00	0,00	0,15	0,00
GT14	0,00	0,13	0,00	0,00
GT19	0,70	0,80	0,76	0,76
GT2	16,55	15,60	10,49	10,76
GT21	0,00	0,13	0,00	0,00
GT25	0,00	0,13	0,15	0,30
GT26	0,35	0,40	0,15	0,15
GT27	0,00	0,27	0,00	0,45
GT28	2,11	2,93	1,67	1,36
GT30	0,70	0,40	0,61	0,76
GT32	0,70	0,53	0,61	0,45
GT35	0,35	0,67	0,76	0,61
GT39	0,00	0,27	0,00	0,00
GT4	9,51	10,00	5,17	6,97
GT41	0,35	0,13	0,00	0,00
GT5	0,35	0,67	0,76	1,06
GT51	1,76	1,73	2,74	2,12
GT56	0,00	0,13	0,00	0,00
GT7	0,00	0,00	0,15	0,15
GT76	0,00	0,13	0,00	0,00
GT8	0,00	0,53	0,46	0,30
GT81	1,06	0,53	0,30	0,15
GT83	1,76	0,53	1,98	3,33
GT9	0,70	0,13	1,67	1,36
GT92	0,00	0,00	0,15	0,00
PL11	0,00	0,53	0,00	0,00
PL12	0,00	0,13	0,00	0,00
PL2	0,00	0,00	0,00	0,15
PL22	0,35	0,13	0,00	0,00
PL26	0,00	0,13	0,00	0,00
PL5	0,35	0,00	0,15	0,15
PL7	0,70	0,27	0,61	1,06
SLH	10,56	17,20	6,99	7,73

Appendix 7. List of beaver cellulase (BC) annotated with prokka. The annotation was performed by using prokka, and the domain was confirmed with conserved domain (CD) search.

BC	Contig	Seq. length (aa)	Prokka annotation	Domain
1	FSBC contig 02705	370	Endoglucanase	GH8 (PRK11097)
2	FSBC contig 05976	1241	Endoglucanase	SLH
3	FSBC contig 12182	369	Endoglucanase	GH8 (PRK11097)
4	FSLC contig 16199	381	Endoglucanase	SLH & Type_I_cohesin_like
5	FSLC contig 05637	370	Endoglucanase	GH8 (PRK11097)
6	FSLC contig 09012	1071	Endoglucanase	SLH
7	FSLC contig 09942	289	Beta-glucanase	GH16 (laminarinase_like)
8	FSLC contig 10126	789	Endoglucanase	Flg_new
9	FSLC contig 11709	213	Endoglucanase	SLH
10	FSLC contig 14433	599	Endoglucanase	bacterial CAP & SLH
11	FSLC contig 17119	503	Endoglucanase	SLH
12	FSLC contig 20138	424	Endoglucanase	SLH
13	FSLC contig 25969	290	Endoglucanase	Big_2 & SLH
14	MSBC contig 01642	369	Endoglucanase	GH8 (PRK11097)
15	MSBC contig 02145	369	Endoglucanase	GH8 (PRK11097)
16	MSBC contig 07099	187	Endoglucanase	SLH
17	MSBC contig 09418	503	Endoglucanase	SLH
18	MSBC contig 09425	1242	Endoglucanase	SLH
19	MSBC contig 14583	599	Endoglucanase	bacterial CAP & SLH
20	MSBC contig 17943	289	Beta-glucanase	GH16 (laminarinase_like)
21	MSLC contig 03384	187	Endoglucanase	SLH
22	MSLC contig 04948	370	Endoglucanase	GH8 (PRK11097)
23	MSLC contig 06147	503	Endoglucanase	SLH
24	MSLC contig 07105	369	Endoglucanase	GH8 (PRK11097)
25	MSLC contig 07652	412	Endoglucanase	GH8 (PRK11097)
26	MSLC contig 10152	599	Endoglucanase	bacterial CAP, Big_2, & SLH
27	MSLC contig 10820	289	Beta-glucanase	GH16 (laminarinase_like)
28	MSLC contig 15196	374	Endoglucanase	GH8 (PRK11097)
29	MSLC contig 16254	774	Endoglucanase	SLH
30	MSLC contig 18439	385	Endoglucanase 1	GH5_4 Cellulase
31	MSLC contig 19790	346	Endoglucanase 5	GH5_2 Cellulase
32	MSLC contig 21430	530	Endoglucanase E1	CBM2 & GH5_1 Cellulase
33	MSLC contig 23174	305	Endoglucanase 4	GH5_2 Cellulase

GH = glycoside hydrolases; SLH = S-layer homology; Fig_new = lysteria-*Bacteroides* repeat domain; CAP = cysteine-rich secretory protein; Big_2 = Bacterial Ig-like domain.

```
MRIEAIEVRVVGKNAAAPGSTANHYVTASGASSGALSVKGTRLVDKSGRTVQLRG
1|        10|        20|        30|        40|        50|

VSTHGLAWFPQYVNEAAFKTLRDDWKVNTIRLALYTAEYGGYCTGGDQANLLSIV
   60|        70|        80|        90|        100|        110|

DRGVQAATKLGMYVIIDWHILSEGSPLVYQAQALNFFRTVSAKYKAYDNVLYEIC
     120|        130|        140|        150|        160|

NEPCNGATWSQVKQYANQVIPAIRANDPDAVVIVGTPTWSQDVDQAAASPLSYSN
 170|        180|        190|        200|        210|        220|

VMYSLHFYANTHTEWLRNRAASALVKGLPIFISEFGACDASGSGGNNFTSASAWL
     230|        240|        250|        260|        270|

SWAESNNIPYVAWNLSNKWETSALISGNTSATSGWSASQLTQWGQWYRNWVRSKA
 280|        290|        300|        310|        320|        330|
```

Appendix 8. The protein sequence of BC33 endoglucanase validated with sanger sequencing.

A

```
RKJ34599.1      1    MGEKQFFLFCLEIFINYSLGGIFMNHQKHRSLSALFLTVLLLAAHVFSGAKLRVLADT--   58
BC33            1    ------------------------------MR---------IEAIEVRVVGHNAA        16
                                                  :         :.. ::**:...

RKJ34599.1      59   ----------QNTPFSIHGALKVDGTNLVDSQGEKFQLRGMSTHGIAWFPQYVNREAFQTL  109
BC33            17   APGSTANHYVTASGASSGALSVKGTRLVDKSGRTVQLRGVSTHGLAWFPQYVNEAAFKTL   76
                              .:  :  ***.*.**.***..*..:****;****;********. **;**

RKJ34599.1      110  RDDWQVNCIRLALYTEEYNGYCTGGHQDELHTLVKKGVDYATELGMYVILDWHVLNDRDP  169
BC33            77   RDDWHVNTIRLALYTAEYGGYCTGGDQANLLSIVDRGVQAATKLGMYVIIDWHILSEGSP  136
                     ****;** ******* **.******:* :* ::*.:**: **;*****:***;*.: .*

RKJ34599.1      170  NVHKEEALKFFQEMSSLYKDQSNVLYEICNEPNSYATWDSIRAYGKEVIPVIRANDPDAV  229
BC33            137  LVVYQAQALNFFRTVSAKYKAYDNVLYEICNEPCNGATWSQVKQYANQVIPAIRANDPDAV 196
                     *:: :**:**: :*: ** .********** . ***.::: *.::***.*********

RKJ34599.1      230  IIVGTPTWSQDIHQAAANPLEYDNLLYALHFYAATHTDWLRDRLKTCVAQGLPVFVSEFG  289
BC33            197  VIVGTPTWSQDVDQAAASPLSYSNVMYSLHFYANTHTEWLRNRAASALVKGLPIFISEFG  256
                     ;**********;.****.**.*.*:;*;***** ***;***:*  :.:.:***;*;****

RKJ34599.1      290  TCDASGNGGNNFDQTRQWLQLLDDYSIGYCCWNLANKAESSSVIVPECSKTYGWSREDLS  349
BC33            257  ACDASGSGGNNFTSASAWLSWAESNNIPYVAWNLSNKWETSALISGNTSATSGWSASQLT  316
                     ;*****.***** .:  **. :. .* * .***;** *;*:;:*  :  * * *** .:*:

RKJ34599.1      350  ESGHWIFEYFRSHN                                               363
BC33            317  QWGQWYRNWVRSKA                                               330
                     : *:*  ::.**:
```

B

```
A0A354CU48|A0A3..  1    MRIKIKSALSLIAAAALALTGCSSSEPDVMLFDEPSVSTPFSETDITEITDIITTVVTAA  60
BC33               1    ------------------------------------------------------------  0

A0A354CU48|A0A3..  61   VSDTEQTAASESKEAETSAAEGAQTSDKQPGGNSSGGSGNNSGNNSGNNSGNNSGSNSGN  120
BC33               1    ------------------------------------------------------------  0

A0A354CU48|A0A3..  121  NSGNNPGNNSGGNSGGSGGGSGSGNNSGSGSFEAGWLNGTTWESGGNYCGTIELKITNKT  180
BC33               1    ------------------------------------------------------------  0

A0A354CU48|A0A3..  181  GSALSSWTVSFTVPSGFEIVNGWNGKYSVSGTTVTVKNESYNGDLANGASIQIGFQYSSS  240
BC33               1    ------------------------------------------------------------  0

A0A354CU48|A0A3..  241  SKFTPPSNITINGNPSGSGSGGGNSGGGNSGGGNGNSPSVTEKPAEPLPAFALNGLVKEH  300
BC33               1    MRIE-AIEVRVVG--------------------KNAAAPGS--------TANHYVTASGASS  33
                        ::     :: : *                      *. :*.           :.. .

A0A354CU48|A0A3..  301  GHLSVKGAQLVDKNGKNFQLKGMSTHGIAWFPDFINEGAFKTLRDDWNTNVVRLAMYTAE  360
BC33               34   GALSVKGTRLVDKSGRTVQLRGVSTHGLAWFPQYVNEAAFKTLRDDWKVNTIRLALYTAE  93
                        * *****;;****.*;.;.**;*****;**** ****;:**.***********:.*;***;

A0A354CU48|A0A3..  361  YGGYCSGGDKSYLKGLIDKGVQAATDLGMYVIIDWHILSDNNPQTNKAEAKSFFMEVSAK  420
BC33               94   YGGYCTGGDQANLLSIVDRGVQAATKLGMYVIIDWHILSEGSPLVYQAQALNFFRTVSAK  153
                        *****;***:: *  .:;*:******. ************:..* .  ;*:* .**. ****

A0A354CU48|A0A3..  421  YKNYDNVLYEICNEPNGGTSWDTIKSYAEEIIPVIRHNSPDSVIIVGTPTWSQDIDQAAA  480
BC33               154  YKAYDNVLYEICNEPCNGATWSQVKQYANQVIPAIRANDPDAVVIVGTPTWSQDVDQAAA  213
                        ** ***********  *::*. :*.**:;***  *.**;*  *.**;********;*****

A0A354CU48|A0A3..  481  KPLKEKNVMYALHFYAATHIDWLRQRLIDCYKNGLPIFVTEFGCCDASGNGANDFGQTRQ  540
BC33               214  SPLSYSNVMYSLHFYANTHTEWLRNRAASALVKGLPIFISEFGACDASGSGGNNFTSASA  273
                        .**. .****;***** ***;***;* :..  ******;;***.*****.*.*;* .:

A0A354CU48|A0A3..  541  WLNLLDSYGISYCNWNLANHKDESSSAFKSNASPNGNWSESVMSEGGHWMRKWFRGEQ    597
BC33               274  WLSWAESNNIPYVAWNLSNKWETSALISGNTSATSGWSASQLTQWGQWYRNWVRSKA    330
                        **. :* .* *  ***;** *;*; :..*;* ...** * ::: *;* *;*.*,:
```

Appendix 9. (A) Sequence alignment of BC33 endoglucanase with GH5 protein from *Ruminococcus* sp. 1xD21-23 (GenBank), and (B) with endoglucanase from Ruminococcaceae **bacterium (UniProt).**

Appendix 10. List of reference GH5 family members used in the study.

Accession no.	Protein and bacteria source
Firmicutes	
AAA20893.1	endoglucanase A partial *Butyrivibrio fibrisolvens*
AAA22304.1	endo-beta-14-glucanase *Bacillus* sp-
AAA22307.1	cellulase *Bacillus subtilis*
AAA22408.1	endo-beta-14-glucanase *Paenibacillus lautus*
AAA22496.1	endo-beta-14-glucanase *Bacillus subtilis*
AAA22631.1	endo-beta-14-glucanase *Paenibacillus polymyxa*
AAA23220.1	endo-14-beta-glucanase *Hungateiclostridium thermocellum*
AAA23221.1	endo-beta-14-glucanase -celCCA- *Ruminiclostridium cellulolyticum*
AAA23224.1	endo-beta-14-glucanase *Hungateiclostridium thermocellum*
AAA23225.1	cellulase precursor -EC 3-2-1-4- *Hungateiclostridium thermocellum*
AAA23230.1	endo-beta-14-glucanase precursor *Clostridium saccharobutylicum*
AAA23231.1	endoglucanase *Clostridium cellulovorans* 743B
AAA23233.1	endoglucanase *Clostridium cellulovorans*
AAA26467.1	beta-14-D-glucanase *Ruminococcus albus*
AAA26469.1	endo-14-beta-glucanase *Ruminococcus albus*
AAA71887.1	beta-mannanase *Caldicellulosiruptor saccharolyticus*
AAB40891.1	endo-14-beta glucanase EngF *Clostridium cellulovorans*
AAC02536.1	endo-b-14-glucanase *Bacillus* sp- 79-23
AAC19169.1	alkaline cellulase Cel5A *Bacillus agaradhaerens*
AAC37035.1	endo-14-beta-D-glucanase *Clostridium longisporum*
AAC43478.1	bifunctional cellulase precursor *Bacillus* sp-
AAC71692.1	beta-14-mannanase *Geobacillus stearothermophilus*
AAD09354.1	s-layer associated multidomain endoglucanase *Caldanaerobius polysaccharolyticus* DSM 13641
AAD39739.1	endoglucanase EngE *Clostridium cellulovorans*
AAF06110.2	mannanase A *Clostridium cellulovorans* 743B
AAF22274.1	multidomain beta-14-mannanase precursor *Caldibacillus cellulovorans*
AAK39540.1	alkali tolerable cellulase *Bacillus subtilis*
AAK73277.1	cellulase *Bacillus* sp- NBL420
AAK94871.1	endo-14-beta-glucanase *Bacillus subtilis*
AAL83749.1	endo-beta-14-glucanase *Paenibacillus* sp- KCTC8848P
AAM23649.1	Cellulase *Caldanaerobacter subterraneus* subsp- *tengcongensis* MB4
AAO63626.1	cellulase C *Bacillus subtilis*
AAP51020.1	Cel5A *Bacillus licheniformis*
AAT06599.1	beta-mannanase precursor *Bacillus* sp- N16-5
AAT39478.1	mannanase *Bacillus circulans*
AAU23613.2	endo-14-beta-glucanase Glycoside hydrolase Family 5 *Bacillus licheniformis* DSM 13 ATCC 14580
AAU40777.1	endoglucanase CelB *Bacillus licheniformis* DSM 13 ATCC 14580
AAX87002.1	mannanase *Bacillus circulans*
AAX87003.1	mannanase *Bacillus circulans*
AAZ22322.1	endo-beta-14-glucanase *Bacillus subtilis*
ABG78039.1	endo-14-beta-glucanase *Bacillus subtilis*
ABN51772.1	glycoside hydrolase family 5 *Hungateiclostridium thermocellum* ATCC 27405
ABN52032.1	glycoside hydrolase family 5 *Hungateiclostridium thermocellum* ATCC 27405
ABN52056.1	coagulation factor 5/8 type domain protein *Hungateiclostridium thermocellum* ATCC 27405
ABN52701.1	glycoside hydrolase family 5 *Hungateiclostridium thermocellum* ATCC 27405
ABN53395.1	Carbohydrate binding family 6 *Hungateiclostridium thermocellum* ATCC 27405
ABN54070.1	glycoside hydrolase family 5 *Hungateiclostridium thermocellum* ATCC 27405
ABP66297.1	Cellulase *Caldicellulosiruptor saccharolyticus* DSM 8903
ABP66692.1	Cellulose 14-beta-cellobiosidase *Caldicellulosiruptor saccharolyticus* DSM 8903
ABV08875.1	Cel5A *Paenibacillus polymyxa*
ABV08876.1	Cel5B *Paenibacillus polymyxa*

93

Appendix 10. continued	
ABV45393.1	cellulase *Bacillus subtilis*
ABX41541.1	Cellulase *Lachnoclostridium phytofermentans* ISDg
ABX42426.1	Cellulase *Lachnoclostridium phytofermentans* ISDg
ACI15227.1	beta-14-endo-glucanase precursor *Bacillus subtilis*
ACI18520.1	endoglucanase H *Dictyoglomus* thermophilum H-6-12
ACJ60856.1	CelA precursor partial *Geobacillus* sp- 70PC53
ACK38261.1	endo-14-beta-glucanase *Bacillus subtilis*
ACK41955.1	glycoside hydrolase family 5 *Dictyoglomus turgidum* DSM 6724
ACK41956.1	glycoside hydrolase family 5 *Dictyoglomus turgidum* DSM 6724
ACL74811.1	glycoside hydrolase family 5 *Ruminiclostridium cellulolyticum* H10
ACL75115.1	cellulosome protein dockerin type I *Ruminiclostridium cellulolyticum* H10
ACL75118.1	glycoside hydrolase family 5 *Ruminiclostridium cellulolyticum* H10
ACL75216.1	glycoside hydrolase family 5 *Ruminiclostridium cellulolyticum* H10
ACL76673.1	glycoside hydrolase family 5 *Ruminiclostridium cellulolyticum* H10
ACM60947.1	glycoside hydrolase family 5 *Caldicellulosiruptor bescii* DSM 6725
ACM60953.1	glycoside hydrolase family 9 *Caldicellulosiruptor bescii* DSM 6725
ACM60954.1	Mannan endo-14-beta-mannosidase- Cellulase *Caldicellulosiruptor bescii* DSM 6725
ACR59602.1	cellulase *Bacillus subtilis*
ACZ54907.1	xyloglucan-specific endo-beta-14-glucanase uncultured bacterium
ACZ98591.1	endoglucanase partial *Cellulosilyticum ruminicola* JCM 14822
ADC54852.1	endoglucanase *Bacillus amyloliquefaciens*
ADD62401.1	cellulase *Bacillus* sp- BG-CS10
ADK66823.1	endo-/exo-type cellulase Clostridiaceae bacterium AN-C16-KBRB
ADL52309.1	protein of unknown function DUF291 *Clostridium cellulovorans* 743B
ADL52789.1	protein of unknown function DUF291 *Clostridium cellulovorans* 743B
ADU21608.1	Cellulase *Ruminococcus albus* 7 DSM 20455
ADU28719.1	Cellulase *Bacillus cellulosilyticus* DSM 2522
ADU28720.1	Cellulase *Bacillus cellulosilyticus* DSM 2522
ADU31612.1	LPXTG-motif cell wall anchor domain protein *Bacillus cellulosilyticus* DSM 2522
AEB00655.1	b-14-endoglucanase *Paenibacillus* sp- ICGEB2008
AEL31246.1	endoglucanase Cel5A bacterium enrichment culture clone FV-Cel5A
AFC68970.1	cellulase *Paenibacillus xylanilyticus*
AFO70071.1	CelB *Caldicellulosiruptor* sp- F32
AGA35556.1	Man5B *Caldanaerobius polysaccharolyticus*
AGG11030.1	endo-beta-14-glucanase *Bacillus subtilis* subsp- subtilis
AGG91154.1	glucanase partial *Bacillus megaterium*
AIY72753.1	endo-beta-14-glucanase partial *Bacillus licheniformis*
ALX38276.1	CelDZ1a *Thermoanaerobacterium* sp- A57Txylan
BAA00045.1	alkaline cellulase *Bacillus* sp-
BAA00859.1	CMCase *Bacillus* subtilis
BAA12744.1	endo-14-beta-glucanase *Bacillus* sp-
BAA25878.1	mannanase *Bacillus circulans*
BAA31712.1	cellulase *Bacillus* sp- 5H
BAA32286.1	endo-14-beta-glucanase *Ruminococcus albus*
BAA92146.1	beta-14-endoglucanase V *Ruminococcus albus*
BAA92430.1	beta-14-endoglucanase VII *Ruminococcus albus*
BAB04322.1	endo-beta-14-glucanase -celulase B- *Bacillus halodurans* C-125
BAB62295.1	endoglucanase N252 *Bacillus* sp- KSM-N252
BAD99527.1	mannanase *Bacillus* sp- JAMB-602
BAE44526.1	xyloglucanase *Paenibacillus* sp- KM21
BAE46390.1	endoglucanase cel5A *Eubacterium cellulosolvens*
BAI52931.1	mannanase *Ruminiclostridium josui*
CAA27266.1	unnamed protein product *Hungateiclostridium thermocellum*
CAA31936.1	celB polypeptide precursor *Caldicellulosiruptor saccharolyticus*
CAA35574.1	endoglucanase *Butyrivibrio fibrisolvens*
CAA38693.1	endoglucanse B *Ruminococcus albus*
CAA47429.1	cellulase *Bacillus subtilis*

Appendix 10. continued	
CAA49187.1	Endo-1,4-beta-glucanase *Hungateiclostridium thermocellum*
CAA73113.1	cellulase *Paenibacillus barcinonensis*
CAA82317.1	endo-14-beta-glucanase *Bacillus subtilis*
CAA83942.1	endo-beta-14-glucanase *Bacillus* sp-
CAB01405.2	endoglucanase *Caldicellulosiruptor bescii* DSM 6725
CAB05881.1	endoglucanase A *Ruminococcus flavefaciens* 17
CAB59165.1	unnamed protein product *Bacillus* sp-
CAB76938.1	14-beta-cellobiohydrolase *Hungateiclostridium thermocellum*
CAC27410.1	endo-14-glucanase *Hungateiclostridium thermocellum*
CAD61244.1	putative cellulase *Bacillus agaradhaerens*
CAE82178.1	endoglucanase precursor *Bacillus licheniformis*
CAJ00038.1	unnamed protein product *Bacillus* sp- KSM-S237
CAJ00039.1	unnamed protein product *Bacillus* sp- KSM-64
CBL16523.1	Endoglucanase *Ruminococcus champanellensis* 18P13 JCM 17042
CBL16847.1	Endoglucanase *Ruminococcus champanellensis* 18P13 JCM 17042
CBL18180.1	Cellulase -glycosyl hydrolase family 5- *Ruminococcus champanellensis* 18P13 JCM 17042
EGC04285.1	cellulase -glycosyl hydrolase family 5- *Ruminococcus albus* 8
WP 026485575.1	hypothetical protein *Caldanaerobius polysaccharolyticus*
Proteobacteria	
AAA61980.1	beta-14-endoglucanase *Ralstonia solanacearum*
AAB61461.1	cellulase A *Cellvibrio mixtus*
AAB61462.2	cellulase B *Cellvibrio mixtus*
AAC02964.2	beta-14--glucan glucanohydrolase precursor *Pectobacterium carotovorum* subsp- carotovorum
AAC37033.1	cellulase *Pectobacterium carotovorum*
AAF18152.1	endo-14-beta-glucanase precursor *Dickeya chrysanthemi*
AAF83628.1	endo-14-beta-glucanase *Xylella fastidiosa* 9a5c
AAF85505.1	endo-14-beta-glucanase *Xylella fastidiosa* 9a5c
AAG44364.1	endoglucanase precursor *Sinorhizobium meliloti*
AAM41068.1	mannan endo-14-beta-mannosidase *Xanthomonas campestris* pv- campestris str- ATCC 33913
AAM42791.1	cellulase *Xanthomonas campestris* pv- campestris str- ATCC 33913
AAO31760.1	endo-b14-mannanase 5B *Cellvibrio japonicus*
AAP04424.1	CelX psychrophilic marine bacterium DY3
AAS19695.1	Man5A *Cellvibrio mixtus*
AAS58467.1	cellulase *Pseudoalteromonas* sp- MB-1
AAU27988.1	hypothetical protein lpg1918 *Legionella pneumophila* subsp- pneumophila str- Philadelphia 1
ABB92850.1	endoglucanase *Azoarcus* sp- BH72
ABC30636.1	Endoglucanase *Hahella chejuensis* KCTC 2396
ABD80834.1	putative retaining b-glycosidase *Saccharophagus degradans* 2-40
ABD81750.1	endoglucanase-like protein *Saccharophagus degradans* 2-40
ABD81754.1	putative endoglucanase *Saccharophagus degradans* 2-40
ABD81896.1	endoglucanase-like protein *Saccharophagus degradans* 2-40
ABD82186.1	putative retaining b-glycosidase *Saccharophagus degradans* 2-40
ABD82280.1	putative retaining b-glycosidase *Saccharophagus degradans* 2-40
ABD82494.1	putative endoglucanase *Saccharophagus degradans* 2-40
ABD82496.1	putative endoglucanase *Saccharophagus degradans* 2-40
ABS72374.1	bifunctional beta 14-endoglucanase/cellobiohydrolase *Teredinibacter turnerae*
ABZ70413.1	Caul_1283 (Cel5A) *Caulobacter* sp. K31
ACE82655.1	Endo-1,4-beta mannanase man5C *Cellvibrio japonicus* Ueda107
ACE82870.1	endo-14-beta glucanase cel5A *Cellvibrio japonicus* Ueda107
ACE84076.1	endo-14-beta glucanase cel5B *Cellvibrio japonicus* Ueda107
ACE84673.1	endo- 14-beta-mannanase man5A *Cellvibrio japonicus* Ueda107
ACJ71329.1	EngV partial *Salinivibrio* sp- NTU-05
ACN62172.1	cellulase *Pseudoalteromonas* sp- NO3

Appendix 10. continued	
ACU30064.1	endo-glucanase *Xanthomonas* sp- EC102
ADG43125.1	glycoside hydrolyase family 5 protein partial *Pseudomonas putida*
ADJ93836.1	endo-14-beta glucanase cel5A *Vibrio xiamenensis*
ADM99099.1	endo-14-beta-glucanase precursor *Dickeya dadantii* 3937
AEM45646.1	cellulase *Pseudomonas* sp- MM15
ALC76666.1	cellulase *Photobacterium panuliri*
BAA25188.1	beta-14-mannanase *Vibrio* sp- MA-138
BAF87299.1	endoglucanase *Azorhizobium caulinodans* ORS 571
BAG69482.2	beta-14-mannanase *Vibrio* sp- MA-138
BAM21527.1	cellulase *Saccharophagus* sp- Myt-1
CAA53592.1	endoglucanase *Pectobacterium carotovorum*
CAA55823.1	cellulase *Pectobacterium carotovorum*
CAA76775.1	cellulase *Pseudoalteromonas haloplanktis*
CAC18529.1	extracellular endoglucanase -ENGXCA protein- *Xanthomonas campestris* pv- campestris
WP 018063499.1	glycoside hydrolase family 5 protein *Martelella mediterranea*
Fibrobacter	
AAA24893.1	endoglucanase 3 precursor *Fibrobacter succinogenes*
AAA50210.1	cellodextrinase *Fibrobacter succinogenes* subsp- succinogenes S85
AAB38548.1	endoglucanase CelG *Fibrobacter succinogenes* subsp- succinogenes S85
AAC06197.1	CMC-xylanase partial *Fibrobacter succinogenes* subsp- succinogenes S85
ABU45500.1	cellulase *Fibrobacter succinogenes* subsp- succinogenes S85
ACX74396.1	Cellulase *Fibrobacter succinogenes* subsp- succinogenes S85
ACX74827.1	Cellulase *Fibrobacter succinogenes* subsp- succinogenes S85
ACX75120.1	glycoside hydrolase family 5 *Fibrobacter succinogenes* subsp- succinogenes S85
ACX76513.1	Carbohydrate-binding CenC domain protein *Fibrobacter succinogenes* subsp- succinogenes S85
ACX76661.1	Cellulase *Fibrobacter succinogenes* subsp- succinogenes S85
Actinobacteria	
AAA23089.1	endo-14-beta-D-glucanase *Cellulomonas fimi*
AAA26710.2	mannanase *Streptomyces lividans* 1326
AAA75477.1	E I beta-14-endoglucanase precursor *Acidothermus cellulolyticus*
AAB67050.1	endoglycoceramidase II *Rhodococcus* sp-
AAC06196.1	endoglucanase *Actinomyces* sp- 40
AAD48494.3	endo-14-beta-glucanase *Cellulomonas flavigena*
AAK16222.1	cellulase CelA *Clavibacter sepedonicus*
AAP56348.1	endoglucanase *Thermobifida fusca* TM51
AAZ54938.1	Mannan endo-14-beta-mannosidase- Glycosyl Hydrolase family 5 *Thermobifida fusca* YX
AAZ54939.1	Cellulase- Glycosyl Hydrolase family 5 *Thermobifida fusca* YX
AAZ56745.1	Cellulase- Glycosyl Hydrolase family 5 *Thermobifida fusca* YX
ACS46797.1	secreted beta-mannosidase *Bifidobacterium animalis* subsp- lactis Bl-04
ADK91085.1	beta-14-mannanase *Streptomyces* sp- S27
AEE43708.1	beta-14-mannanase *Cellulosimicrobium* sp- HY-13
AEN10237.1	cellulose-binding family II *Streptomyces* sp- SirexAA-E
AHB89702.1	endomannanse *Thermobifida fusca* TM51
AHB89703.1	endomannanse *Thermobifida cellulosilytica* TB100
BAB17317.1	C9 endoglycoceramidase *Rhodococcus* sp- C9
BAC65342.1	chitosanase II *Streptomyces griseus*
BAF56440.1	endogalactosylceramidase partial *Rhodococcus hoagii*
BAK26781.1	beta-mannanase *Streptomyces thermolilacinus*
BAM62868.1	mannanase *Streptomyces thermoluteus*
CAA44467.2	Cellulase partial -plasmid- *Clavibacter michiganensis* subsp- michiganensis NCPPB 382
EME18930.1	endoglycosylceramidase *Rhodococcus triatomae* BKS 15-14
Bacteroidetes	
AAC36862.1	xylanase *Prevotella ruminicola*
AAC97596.1	B-14-endoglucanase *Prevotella bryantii* B14

Appendix 10. continued	
ABG58383.1	CHU large protein endoglucanase glycoside hydrolase family 5 protein *Cytophaga hutchinsonii* ATCC 33406
ABG59366.1	endoglucanase glycoside hydrolase family 5 protein *Cytophaga hutchinsonii* ATCC 33406
ADB80152.1	endoglucanase *Cellulophaga* sp- QY3
ADE83057.1	glycosyl hydrolase family 5 *Prevotella ruminicola* 23
ADI70667.1	Xyn5B *Prevotella bryantii* B14
ADI70668.1	Xyn5A *Prevotella bryantii* B14
ADV50035.1	Cellulase *Cellulophaga algicola* DSM 14237
AIJ19564.1	GH5 Bacteroidetes bacterium AC2a
AIT97140.1	gycosyl hydrolase 5 uncultured Bacteroidetes bacterium
AIT97141.1	gycosyl hydrolase 5 uncultured Bacteroidetes bacterium
ALJ47680.1	Endoglucanase A precursor *Bacteroides ovatus*
EDV05070.1	bacterial group 2 Ig-like protein *Bacteroides intestinalis* DSM 17393
EEC54456.1	cellulase -glycosyl hydrolase family 5- *Bacteroides eggerthii* DSM 20697
Spirochaeta	
ADN02392.1	endo-1-4-beta glucanase glycosyl hydrolase family 5 *Spirochaeta thermophila* DSM 6192
ADN02996.1	glycosyl hydrolase family 5 cellulase CBM64 *Spirochaeta thermophila* DSM 6192
Thermotogae	
AAD36302.1	endo-14-beta-mannosidase *Thermotoga maritima* MSB8
AAD36816.1	endoglucanase *Thermotoga maritima* MSB8
ABQ47550.1	Mannan endo-14-beta-mannosidase- Glycosyl Hydrolase family 5 *Thermotoga petrophila* RKU-1
ABS61403.1	glycoside hydrolase family 5 *Fervidobacterium nodosum* Rt17-B1
AFY97404.1	Cel5A *Fervidobacterium gondwanense*
CAB56856.1	beta-mannosidase *Thermotoga neapolitana*
Deinococcus.Thermus	
AAK60011.1	endocellulase (CelA) (Cel5A) *Thermus caldophilus* GK24
Uncultured bacterium	
AAA91966.1	cellulase bacterium
ABA02176.1	cellulase uncultured bacterium
ABA42184.1	beta-glucanase uncultured bacterium
ABA42185.1	beta-glucanase uncultured bacterium
ABB46200.1	endoglycosidase precursor protein uncultured bacterium
ABB51612.1	endo-14-beta-D-glucanase uncultured bacterium
ABE60666.1	endo-14-beta-D-glucanase uncultured bacterium
ABE60714.1	endo-14-beta-D-glucanase uncultured bacterium
ABE60715.1	cellodextrinase uncultured bacterium
ABI94085.1	endo-beta-14-glucanase uncultured bacterium
ABI94086.1	endo-beta-14-glucanase uncultured bacterium
ABW39345.1	glycoside hydrolase family 5 uncultured bacterium
ABW39351.1	glycoside hydrolase family 5 uncultured bacterium
ACO55737.1	endoglucanase uncultured bacterium
ACR23656.1	cellulase bacterium enrichment culture clone CelA10
ACR23659.1	cellulase bacterium enrichment culture clone CelA24
ADA62505.1	bifunctional mannanase-xyloglucanase uncultured bacterium
ADB44000.1	exo-xyloglucanase uncultured bacterium
ADD71777.1	endo-beta-14-glucanase uncultured bacterium
ADM89627.1	endo-glucanase uncultured rumen bacterium
ADU86901.1	cellulase uncultured bacterium
ADU86902.1	cellulase uncultured bacterium
AEL31247.1	endoglucanase Cel5B bacterium enrichment culture clone FXVII-Cel5B
AEV59725.1	putative cellulase uncultured bacterium
AEV59731.1	putative cellulase uncultured bacterium
AEV59734.1	putative cellulase uncultured bacterium
AEV59735.1	putative cellulase uncultured bacterium

Appendix 10. continued	
AEV59736.1	putative cellulase uncultured bacterium
AEV59739.1	putative cellulase uncultured bacterium
AFJ05146.1	cellulase uncultured bacterium
AFJ44728.1	beta-14--glucan glucanohydrolase uncultured bacterium
AFS18545.1	endoglucanase uncultured bacterium
AFX88668.1	endo-14-beta-glucanase uncultured bacterium
AFX88671.1	endo-14-beta-glucanase uncultured bacterium
AFX88673.1	endo-14-beta-glucanase uncultured bacterium
AGL50932.1	cellulase precursor uncultured bacterium
AHA42547.2	CelE1 uncultured bacterium
AHC00282.1	beta-endo-14-mannosidase precursor uncultured bacterium
AHF23845.1	cellulase -glycosyl hydrolase family 5- uncultured bacterium Contig15
AHF24998.1	cellulase uncultured bacterium Contig33
AID57617.1	endoglucanase GH5 25 uncultured bacterium
AND74761.1	cellulase GH5 uncultured bacterium
CAP07661.1	beta-glucanase uncultured rumen bacterium

Appendix 11.

ACKNOWLEDGEMENT

This thesis was finally possible thanks to the god, "**Allah SWT**" who gave me the opportunity to finish it.

First of all, I would like to thank Prof. Dr. Rolf Daniel. To accept me under his supervision, enabled me to venture into this interesting and amazing world of research, and for the valuable advice and suggestion during the course of this study. The experiences and knowledges I gained for these past years means a lot for me and my carrier in the future, thank you very much.

I am also grateful to PD Dr. Michael Hoppert that he is my second supervisor. Thank you very much for your constructive advice in the course of my PhD thesis. My thanks also go to the members of the examination committee, Prof. Dr. Stefanie Pöggeler, Prof. Dr. Gerhard Braus, apl. Prof. Dr. Fabian Commichau, and Dr. Christian Roos for evaluating my thesis

Many thanks to all the postdoc in our department, to my bioinformatic mentor and my supervisor, Dr. Dominik Schneider, for his great teachings, scripts, advices, and various help regarding data processing that I used in this study. Learning bioinformatics was originally my main goal to study in Germany, and I learned a lot from you. Whenever I met a dead-end with my data, you keep on helping me with patience even though you were very busy to begin with. I can proudly say that it is a pleasure to be in "The Dominikans"!

I would like to thank to the two person who help me maintain my well-being in the lab, Dr. Jacqueline Hollensteiner and Dr. Birgit Pfeiffer. Many thanks to both of you, I feel like being with my own family. Thank you Jacky for always taken care of me, help me go through different phases at the end of my thesis, and thank you Birgit for helping me to correct my thesis from the messy English grammar and the general support that I receive in my thesis, you were a great help.

I would like to thank Dr. Silja Brady who welcomed me to the department in 2014. Even though our time was limited, I appreciate all your support and kindness. I would also like to thank Dr. Andrea Thürmer and Dr. Anja Poehlein who helped with the sequencing part. Also many thanks to Dr. Robert Hertel for his general support and Dr. Heiko Nacke who gave me advices on cellulases.

I would like to thank Dipl.-Biol. Antje Weber from *Büro Wildforschung & Artenschutz*, Jeggau 44a, 39649 Gardelegen-Germany, who provided the beaver gut samples used in this study, and kindly provided some of the beaver pictures used in this thesis. I am very grateful for the opportunity to receive Erasmus Mundus EXPERTS4Asia scholarships award which supporting me to pursue PhD in Göttingen. Thank you for the EXPERTS coordinator and team from both Uni-Göttingen and IPB. Also I am thankful to GAUSS for the family bridging program that gave me additional support to continue my PhD.

I am very grateful to my friends, both of whom already graduated and are still struggling, for the past years of my PhD. Thanks to Kristin, Amelie, Richard, Genis, Mingji, Sarah, Simone, Dirk, Avril, Ines, Tatiana, and Stefanie. All of you are amazing! I will miss the irreplaceable moments we share together. I have to thank both Amelie and Genis for their great help during my PhD. Both of you help me countless time to guide me, to teach, to share your knowledge, and to help me finish although yourself was finishing your PhD. Now my PhD time is over, I can only say thank you for all your support and the friendship we have built during that time. I hope we have the opportunity to meet again in the future! I also would like to thanks Inka, Cynthia, Alejandra, Katia, Sylvia (to name just a few) for being an awesome friend! Thanks also goes to Melanie and Mechtild who keep the labs in order and provide help whenever needed.

I would like to thank my former student Tim, who wrote his bachelor thesis under my supervision and contributed to the results of this thesis. I would like to thank the students who become part of our big family, to Tobi, Karo, Katrina, Katha, Jorn, Till, Mazidah (to name just a few), thanks to make the department such a nice place! I would like to thank the other people in the AG Daniel, who present as well as gone. Thank you for your support and warm friendship, to Petra, Daniela, and Nicole, who is a nice partner for learning German.

Who am I without my family? I can't really put it into word, how grateful I am that my family supports me all the years of studying in Germany. I would like to present this PhD for my mother, who never tired keep encouraging me to pursue my study, keep pushing me to not give up, to the point visiting me in Germany to give her support. You are the biggest reason why I am able to finish this PhD. For my father, sisters and brother, thank you for your family love and support. For my wife, and our daughter born here in Göttingen, you give me a strong strength to always keep me up, no matter how many failures I have experienced, all for our little family. Thank you so much for being here with me, I love you all.

Appendix 13.

<div align="center">

THESIS DECLARATION

</div>

Declaration of plagiarism

I hereby declare that the PhD thesis entitled "Taxonomic and Functional Characterization of Biopolymer-degrading Microbial Communities in the Intestinal Tract of Beavers" has been written independently and with no other sources and aids than quoted. Moreover, I have not submitted this thesis previously in any form for another degree at any institution or any university.

Rahadian Pratama

Göttingen, 11 April 2019