

UNIVERSIDADE FEDERAL DA BAHIA INSTITUTO MULTIDISCIPLINAR EM SAÚDE



PROGRAMA DE PÓS-GRADUAÇÃO EM BIOCIÊNCIAS

LUCAS MAGALHÃES ALCANTARA

TRIAGEM DE GENES LIGNOCELULOLÍTICOS ATRAVÉS DE ANÁLISE METAGENÔMICA DE GENOMAS MICROBIANOS DO RÚMEN: UM ENFOQUE EM BIORREFINARIA

Vitória da Conquista, BA 2017

LUCAS MAGALHÃES ALCANTARA

TRIAGEM DE GENES LIGNOCELULOLÍTICOS ATRAVÉS DE ANÁLISE METAGENÔMICA DE GENOMAS MICROBIANOS DO RÚMEN: UM ENFOQUE EM BIORREFINARIA

Dissertação apresentada ao Programa de Pós-Graduação em Biociências, Universidade Federal da Bahia, como requisito para obtenção do título de Mestre em Biociências.

Orientadora: Prof.^a Dr.^a Andrea Lopes de Oliveira Ferreira Universidade Federal da Paraíba – UFPb

Co-orientador: Prof. Dr. Leandro Martins de Freitas Universidade Federal da Bahia – UFBA

Vitória da Conquista, BA 2017

LUCAS MAGALHÃES ALCANTARA

TRIAGEM DE GENES LIGNOCELULOLÍTICOS ATRAVÉS DE ANÁLISE METAGENÔMICA DE GENOMAS MICROBIANOS DO RÚMEN: UM ENFOQUE EM BIORREFINARIA

Esta dissertação foi julgada adequada à obtenção do grau de Mestre em Biociências e aprovada em sua forma final pelo Programa de Pós-graduação em Biociências, Universidade Federal da Bahia.

Vitória da Conquista - BA, 09/06/2017.

Prof^a. Dr^a. Andrea Lopes de Oliveira Ferreira (Orientadora) Universidade Federal da Paraíba

Prof. Dr. Leandro Martins de Freitas (Co-orientador) Universidade Federal da Bahia

Prof^a. Dr^a. Patrícia Lopes Leal (Examinadora) Universidade Federal da Bahia

Prof^a. Dr. Giovanilton Ferreira da Silva (Examinador) Instituto Federal da Bahia

AGRADECIMENTOS

Primeiramente agradeço a Deus, pois toda a boa dádiva e todo o dom vem d'Ele. Agradeço também à minha orientadora, Dr. Andrea Lopes de Oliveira Ferreira, pelo aprendizado e amizade que desenvolvemos ao decorrer desses anos juntos. Agradeço aos meus professores, colegas e alunos do tirocínio à docência que contribuíram ativamente no processo de ensino-aprendizagem. Agradeço à minha professora Dr. Patrícia Lopes Leal que nos deu a brilhante ideia para o desenvolvimento desta pesquisa. Agradeço ao meu professor e amigo Dr. Félix Gonçalves de Siqueira por ter despertado em mim a paixão por essa área da ciência e sempre ter me motivado a sonhar alto. Agradeço também a minha família pelo suporte e compreensão pelos longos momentos que precisei me fazer ausente. Por fim, e não menos importante, agradeço de coração ao meu co-orientador, Dr. Leandro Martins de Freitas, por ter aceitado o desafio de me co-orientar e ter feito isso da melhor forma possível, se colocando à total disposição para me corrigir e ensinar sem medir esforços. A todos vocês, meu muito obrigado!

"Se algum de vocês tem falta de sabedoria, peça-a a Deus, que a todos dá livremente, de boa vontade; e lhe será concedida". Tiago 1:5. Bíblia Sagrada.

ALCANTARA, Lucas Magalhães. Triagem de genes lignocelulolíticos através de análise metagenômica de genomas microbianos do rúmen: um enfoque em biorrefinaria. 72 f. il. 2017. Dissertação (Mestrado) – Instituto Multidisciplinar em Saúde, Universidade Federal da Bahia, Vitória da Conquista, 2017.

RESUMO

No rúmen, a população microbiana converte os complexos alimentos (plantas lignocelulósicas) em compostos mais simples, que podem ser facilmente capturados e metabolizados pelos ruminantes. Por conseguinte, a utilização direta da população microbiana do rúmen, ou enzimas obtidas a partir destes, pode proporcionar a solução para um problema industrial atual, isto é, a bioconversão de material lignocelulósico em combustíveis, produtos químicos e outros bioprodutos valiosos. Esta pesquisa realizou uma triagem de enzimas lignocelulolíticas de microrganismos do rúmen por meio de análises metagenômicas por homologia de sequência e estudou as vias de degradação da lignocelulose no rúmen. Um total de 410 proteomas preditos do microbioma do rúmen foram sistematicamente triados em diferentes famílias de Enzimas Carboidrato-Ativas (CAZymes) e Módulos de Ligação à Carboidratos (CBMs) em relação a diferentes microrganismos comparados com a base de dados CAZy. Observou-se que Firmicutes e Bacteroidetes são os filos responsáveis pela maioria das CAZymes e CBMs produzidos no rúmen. A ação corporativa entre os diferentes filos durante a degradação da biomassa foi observada, confirmando um sinergismo de sintonia refinada entre as diferentes espécies de rúmen. Grandes quantidades de CBM50 e quitinases chamam a atenção para a capacidade das bactérias do rúmen de usar polímeros quitinosos de paredes celulares fúngicas como uma fonte de carboidrato para fins de energia e crescimento. A ausência de enzimas de Atividade Auxiliar conhecidas por apresentarem atividades ligninolíticas reafirma a falta de habilidade da microbiota do rúmen na degradação da lignina. No entanto, a análise filogenética sugeriu que o único gene de Laccase encontrado foi adquirido por quatro Bacilli através de transferência horizontal de genes. Os resultados também mostraram a presença de enzimas que participam no afrouxamento da parede celular por clivagem de ligações entre a lignina e a hemicelulose, tornando assim a celulose mais acessível para hidrolases. Este estudo demonstra a importante relação entre microrganismos no rúmen, a qual contribui para as enzimas necessárias à degradação da parede celular vegetal em condições ambientais tão únicas.

Palavras-chave

Microbiota do Rúmen, Enzimas Carboidrato-Ativas, Módulos de Ligação à Carboidratos, Metagenômica

ALCANTARA, Lucas Magalhães. Screening of lignocullulolytic genes through metagenomic analysis of rumen microbial genomes: a biorefinery approach. 72 pp. il. 2017. Master Dissertation – Instituto Multidisciplinar em Saúde, Universidade Federal da Bahia, Vitória da Conquista, 2017.

ABSTRACT

In the rumen, microbial population converts the complex food stuffs (lingo-cellulosic plant materials) into simpler compounds, which can be easily taken up and metabolized by ruminants. Therefore, direct use of microbial population from rumen or enzymes obtained from them may provide the solution to a current industrial problem, i.e., the bioconversion of lignocellulosic material into fuels, chemicals, and other valuable bioproducts. This research performed a screening of lignocellulolytic enzymes of rumen microorganisms through metagenomic analyzes by sequence homology and studied the degradation pathways of lignocellulose in the rumen. A total of 410 predicted proteomes from the rumen microbiome were systematically screened for different families of Carbohydrate-Active Enzymes (CAZymes) and Carbohydrate-Binding Modules (CBMs) across different microorganisms compared with CAZy database. It was observed that Firmicutes and Bacteroidetes are the phylum responsible for the majority of CAZymes and CBMs produced in the rumen. Corporate action among different phyla during biomass degradation was observed, confirming the fine-tuning synergism among different species of rumen. Great amounts of CBM50 and chitinases call attention to the ability of rumen bacteria to use chitinous polymers from fungal cell walls as a source of carbohydrate for energy and growth purposes. The absence of auxiliary activity enzymes reported to present ligninolytic activities reaffirms the lack of ability of rumen microbiota to degrade lignin. Nonetheless, phylogenetic analysis suggested the only Laccase gene that was found was acquired by four Bacilli through horizontal gene transfer. Results also showed the presence of enzymes that participate in the loosening of cell wall by cleaving linkages between lignin and hemicellulose, thus making cellulose more accessible for hydrolases. This study demonstrates the important relation between microorganisms in the rumen that contributes to the enzymes required to degrade plant cell wall under such unique environmental condition.

Keywords

Rumen microbiota, Carbohydrate-Active Enzymes, Carbohydrate-binding modules, Metagenomics

1	INTRODUÇÃO10
2	REVISÃO DA LITERATURA12
2.1	LIGNOCELULOSE
2.1.1	LIGNINA
2.1.2	CELULOSE
2.1.3	HEMICELULOSE
2.1.4	RECALCITRÂNCIA DA BIOMASSA
2.1.5	PRÉ-TRATAMENTO DA BIOMASSA
2.2	RÚMEN BOVINO: UMA FONTE DE MICRORGANISMOS DE INTERESSE
INDU	STRIAL
2.3	ENZIMAS LIGNOCELULOLÍTICAS
2.3.1	CELULASES E HEMICELULASES
2.3.2	LIGNINASES
2.3.3	ENZIMAS DE ATIVIDADES AUXILIARES
2.4	METAGENÔMICA COMO FERRAMENTA DE DESCOBERTA DE NOVOS
GENE	27
3	JUSTIFICATIVA
4	OBJETIVO GERAL
4.1	OBJETIVOS ESPECÍFICOS
5	METODOLOGIA
5.1	RECUPERAÇÃO DE SEQUÊNCIAS
5.2	BUSCA POR SIMILARIDADE
5.3	CLASSIFICAÇÃO TAXONÔMICA
5.4	DETERMINAÇÃO DO POTENCIAL DE DEGRADAÇÃO DA BIOMASSA32
5.5	CLASSIFICAÇÃO DOS GENOMAS DO RÚMEN NOS GRUPOS DE GENES
ORTÓ	DLOGOS
6	REFERÊNCIAS
7	ARTIGO

SUMÁRIO

1 INTRODUÇÃO

O uso de material vegetal lignocelulósico como fonte de carbono para a produção de combustíveis e produtos químicos é atualmente um dos pontos focais da biotecnologia. O consumo total de petróleo no mundo, em 2014, foi de 92,42 milhões de barris/dia (BATISTA-GARCÍA *et al.*, 2016), e a projeção para 2030 é de crescimento de, aproximadamente, 25% pela demanda por petróleo bruto (BALAN, 2014). Em relação ao consumo de energia, cerca de 78% é derivado de combustíveis fósseis, enquanto apenas 19% é atribuído a fontes renováveis de energia, dos quais menos de 1% diz respeito a biocombustíveis já bem estabelecidos, como etanol e biodiesel (REN21, 2017).

Apesar de não haver dados exatos da quantidade de petróleo ainda disponível ao redor do mundo e por quanto tempo será possível depender dele, a sociedade já entende que é necessário adaptar-se e diversificar as fontes de energia para um futuro mais limpo e sustentável. Assim, é preciso que haja uma abordagem que considere a conversão de grandes partes da economia global em uma economia sustentável com bioenergia, biocombustíveis e bioprodutos como seus pilares principais (Figura 1).





Combustíveis e produtos químicos com base biológica, produzidos utilizando matéria prima residual, tal como a biomassa lignocelulósica, promove à sociedade diversos benefícios (GREENWELL; LLOYD-EVANS; WENNER, 2012), tais como:

- 1) Ser renovável e sustentável;
- Auxiliar indiretamente na fixação de dióxido de carbono atmosférico (o qual contribui negativamente para o aquecimento global);
- 3) Facilitar o estímulo e desenvolvimento da economia local;

- Reduzir a poluição atmosférica causada pela queima e degradação de biomassa nos campos;
- Trazer segurança energética para países dependentes da importação de combustíveis, e
- Criar empregos em áreas de alta tecnologia para engenheiros, biotecnologistas e cientistas.

A biomassa lignocelulósica pode ser considerada como única fonte para produção em larga escala de combustíveis renováveis, químicos e biomateriais (DEMARTINI *et al.*, 2013). No entanto, a parede celular da planta evoluiu para garantir sua estabilidade e robustez e não com a finalidade de armazenamento de carbono, apesar de possuir uma das mais abundantes fontes de carbono renovável da biosfera (SAUER; MARX; MATTANOVICH, 2012). Assim, apesar da lignocelulose ser uma matéria prima rica em macromoléculas essenciais para o desenvolvimento de uma biorrefinaria, ela possui uma estrutura complexa que precisa ser vencida para poder liberar o seu real potencial.

O maior obstáculo para a produção em escala industrial de biocombustíveis a partir de lignocelulose reside na desconstrução ineficiente do material vegetal devido à natureza recalcitrante do substrato frente a quebra enzimática, e à relativa baixa atividade das enzimas disponíveis atualmente (BLANCH *et al.*, 2008).

O grande número de estudos acerca das vias de degradação da lignocelulose não apenas permite entender os mecanismos e interações entre organismos na manutenção do balanço carbônico nos ciclos biogeoquímicos, mas permite também levar à descobertas potenciais de microrganismos não caracterizados e novas enzimas que, por sua vez, poderiam melhorar a conversão de biomassa vegetal subutilizada à biocombustíveis, químicos e outros materiais através de biorrefinarias (LEE, 1997; RUBIN, 2008).

Embora o sucesso na engenharia de proteínas para aumentar o desempenho de enzimas degradadoras de lignocelulose existentes tem sido limitado (WEN; NAIR; ZHAO, 2009), a recuperação de enzimas de comunidades microbianas, naturalmente envolvidas com a degradação de biomassa, oferece uma estratégia promissora à identificação de novas enzimas lignocelulolíticas com atividades potencialmente melhores (RUBIN, 2008).

A maioria das enzimas tem sido identificada utilizando laboriosos métodos de triagem e extensas caracterizações bioquímicas, mas evidências baseadas na bioinformática aumentaram a percepção que a abordagem *in silico* pode ser aplicada na identificação de novas enzimas a partir de sequências de genomas microbianos (WANG; FEWER; SIVONEN, 2011). A metagenômica consiste na análise direta do DNA de diferentes amostras de um determinado ambiente e ela representa uma estratégia para descobrir diversas enzimas codificadas na natureza (FERRER *et al.*, 2005; WANG *et al.*, 2009). A abordagem metagenômica tem sido utilizada com sucesso para identificar enzimas com atividades de interesse industrial (LI *et al.*, 2009).

Uma vez que o DNA metagenômico é obtido, a prospecção de novas enzimas ocorre através da triagem de bibliotecas de expressão, seja ela baseada em funções ou em similaridade das sequências (BATISTA-GARCÍA *et al.*, 2016). Em estudo, Li e colaboradores (2009) destacaram a relevância e vantagens do método de triagem metagenômico baseado em homologia de sequências. Segundo eles, ao contrário dos métodos baseados em funções, é possível detectar um número maior de genes alvo, uma vez que esse método independe da expressão gênica, do dobramento de proteínas no hospedeiro, das modificações póstraducionais, entre outros. Ou seja, esse método torna-se bastante atrativo, pois ele independe da integridade da sequência do gene alvo.

Nesse contexto, essa pesquisa estudou as vias de degradação da lignocelulose, as enzimas utilizadas nesse processo e realizou uma triagem de genes lignocelulolíticos de microrganismos do rúmen bovino através de análises metagenômicas por homologia de sequências.

2 REVISÃO DA LITERATURA

2.1 LIGNOCELULOSE

As plantas são compostas de mais de 35 diferentes tipos celulares que são distintos em composição, estrutura e ultraestrutura. Em células em crescimento, a parede é tipicamente uma camada fina e flexível (0,1 a 1 μ m) que consiste em, primordialmente, complexos polissacarídeos e pequenas quantidades de proteínas estruturais (DAVISON et al., 2013). Entretanto, todas as células têm uma grossa parede celular que promove rigidez e previne ataque por patógenos. As paredes celulares são tipicamente compostas por três camadas, a lamela média, parede celular primária e parede celular secundária. Esta última, a qual possui subcamadas (S1, externa; S2, média; e S3, externa), está presente apenas em certos tecidos e amadurece geralmente após final do crescimento, diferentemente das paredes celulares primárias, as quais estão presentes em todas as células (COSGROVE, 2005).

A composição química das paredes celulares difere significantemente entre monocotiledôneas e dicotiledôneas, o que influencia fortemente na sua susceptibilidade à desconstrução. Celulose (20–50%), hemicelulose (15–35%), e lignina (10–30%) são os principais constituintes da parede celular, enquanto que proteínas (3–10%), lipídios (1–5%), açúcares solúveis (1–10%), e minerais (5–10%) são componentes presentes em menores quantidades (todos com base em massa seca) (Figura 2) (PAULY; KEEGSTRA, 2008).



Figura 2. Estrutura da lignocelulose. O principal componente da lignocelulose é a celulose, uma cadeia de moléculas de glicose unidas por ligações $\beta(1-4)$. Ligações de hidrogênio entre diferentes camadas do polissacarídeo contribui para a resistência à degradação da celulose cristalina. Hemicelulose, o segundo componente mais abundante da lignocelulose, é composto de vários açúcares de 5 e 6 carbonos, como arabinose, galactose, glicose, manose e xilose. A lignina é composta de três componentes fenólicos principais, chamados de álcool *p*-cumarílico (H), álcool coniferílico (G) e álcool sinapílico (S). A lignina é sintetizada pela polimerização desses componentes e a sua proporção no polímero varia entre diferentes plantas, tecidos e camadas da parede celular. Celulose, hemicelulose e lignina formam estruturas chamadas de microfibras, as quais são organizadas em macrofibras que medeiam a estabilidade estrutural na parede celular vegetal (Adaptado de RUBIN, 2008).

No complexo lignocelulósico, a celulose retém a estrutura fibrosa cristalina, a qual é a parte central do complexo, enquanto que a hemicelulose posiciona-se entre as micro e macrofibras de celulose. A lignina proporciona um papel estrutural à matriz, na qual a celulose e a hemicelulose são incorporadas (POTUMARTHI; BAADHE; BHATTACHARYA, 2013).

Existem quatro tipos principais de ligações químicas descritas na literatura para o complexo lignocelulósico. São elas:

- 1) Éter;
- 2) Éster;

- 3) Carbono-carbono e
- 4) Pontes de hidrogênio.

Essas quatro ligações representam a maioria dos tipos de ligações que promovem as ligações intramoleculares, e entre diferentes componentes do material lignocelulósico para formar as ligações intermoleculares (Tabela 1).

Tabela 1. Visão geral das ligações entre as unidades de monômeros e polímeros para formar a lignocelulose (HARMSEN *et al.*, 2010)

Tipo de ligação	Ligações intramoleculares	Ligações intermoleculares
Éter (C-O-C)	Lignina Hemicelulose Celulose	Celulose-Lignina Hemicelulose-Lignina
Éster (R'-COOR")	Hemicelulose	Hemicelulose-Lignina
Carbono-carbono	Lignina	-
Pontes de Hidrogênio	Celulose	Celulose-Hemicelulose Celulose-Lignina Hemicelulose-Lignina

2.1.1 LIGNINA

A lignina consiste de três unidades principais de fenilpropano com um átomo de oxigênio na posição *p* (como OH ou O-C) e com um, dois ou nenhum grupo metoxílico nas posições orto desse átomo de oxigênio, são elas: 1) guaiacilpropano, 2) siringilpropano, e 3) *p*-hidroxifenilpropano (Figura 3) (BRUNOW, 2008).

O que torna a estrutura da lignina mais complexa é que podem haver alterações significativas nas estruturas moleculares como, por exemplo, as posições orto podem, alternativamente, serem C-substituídas ou O-substituídas por outros grupos que não sejam metoxílicos (ZHAO, 2016).



Figura 3. Estrutura química das principais unidades de fenilpropano formadoras da lignina: 1) guaiacilpropano, 2) siringilpropano, e 3) *p*-hidroxifenilpropano.

Algumas poucas unidades aromáticas também podem ser substituídas, além de uma pequena porcentagem das moléculas estruturais também não serem unidades de fenilpropano. As unidades de fenilpropano são ligadas umas às outras por uma série de ligações características (β -O-4, β -5, β - β , α - α , 3-5, etc.), sem haver qualquer padrão de ligação em si, e também podem haver ligações entre a lignina e carboidratos (Figura 4) (BRUNOW; LUNDQUIST; GELLERSTEDT, 1999; ZHAO, 2016).



Figura 4. Exemplo de estrutura química da lignina com 20 unidades de fenilpropano ligadas entre si através de ligações do tipo β -O-4, β -5, β - β , etc.

O processo de lignificação tem origem no aparato de Golgi com a geração de três tipos de álcoois *p*-hidroxi-cinamílicos, 1) álcool *p*-cumarílico, 2) álcool coniferílico, 3) álcool sinapílico (Figura 5), a partir dos ácidos *p*-cumárico, ferúlico e sinápico, respectivamente.



Figura 5. Estruturas químicas dos álcoois precursores das unidades de fenilpropano constituintes da Lignina: 1) álcool *p*-cumarílico (H), 2) álcool coniferílico (G), 3) álcool sinapílico (S).

Os álcoois precursores são então transportados através da membrana até sítios de lignificação na parede celular e, para eles, liberados pela enzima β-glicosidase. Então, a lignina

é polimerizada e classificada de acordo com a polimerização dos álcoois *p*-hidroxi-cinamílicos (Tabela 1), promovendo rigidez, impermeabilidade e reforço estrutural para a planta (VOXEUR; WANG; SIBOUT, 2015).

Lignina	Álcool Base	Exemplo
Lignina Guaiacila (G)	Coniferílico	Coníferas
Lignina Guaiacila-Siringila (G-S)	Coniferílico Sinapílico	Folhosas
Lignina 4-Hidroxifenil-Guaiacila-Siringila (H-G-S)	Coniferílico Sinapílico <i>p</i> -Coumarílico	Gramíneas
Lignina 4-Hidroxifenil-Guaiacila (H-G)	Coniferílico <i>p</i> -Coumarílico	Madeira de compressão

Tabela 1. Classificação da lignina de acordo com a polimerização dos álcoois p-hidroxi-cinamílicos

2.1.2 CELULOSE

A celulose é uma complexa macromolécula composta de cadeias lineares de glicose com ligações do tipo β -1,4 que se agregam firmemente em microfibras (3 a 5 nm em diâmetro), as quais se mantêm seguras através de ligações de hidrogênio inter e intramoleculares e forças de Van der Waals (Figura 6).

O grau de polimerização da celulose varia, dependendo da sua fonte, de 2-6 mil e de 2-10 mil resíduos nas paredes primária e secundária, respectivamente, podendo ela ser cristalina e/ou amorfa, sendo a última mais fácil de ser degradada (CHUNDAWAT *et al.*, 2011). Devido à ausência de enrolamento ou ramificações, e pelo auxílio da conformação equatorial dos resíduos de glicose, a molécula de celulose adota uma conformação extensa em forma de bastão (KUBICEK, 2013).



Figura 6. Estrutura química da celulose. A cadeia da celulose é formada por unidades de glicose (delimitada por colchetes) ligadas através de ligações covalentes do tipo β -1,4. Cadeias de celulose são firmemente ligadas entre si por pontes de hidrogênio, tanto por ligações intramoleculares (linhas

tracejadas indicadas por setas), quanto por ligações intermoleculares (demais linhas tracejadas) (adaptado de KUBICEK, 2013).

A estrutura cristalina da celulose, a qual consiste em duas cadeias paralelas de glucano (monômeros de D-glicose unidos por ligações glicosídicas), são ligadas entre si por pontes de hidrogênio e forças de Van der Waals, as quais são então empilhadas em fibras que assumem formas de bastão. A hierarquia morfológica é definida por fibras elementares (1,5 a 3,5 nm), microfibras (10 a 30 nm) e macrofibras (~100 nm) (KUBICEK, 2013). Em sua forma cristalina, a celulose é insolúvel em água e praticamente inacessível à degradação enzimática devido à sua organização (BATISTA-GARCÍA et al., 2016).

A estereometria das cadeias de glicose empacotadas firmemente nesta morfologia sólida e cristalina é responsável pela baixa taxa de sacarificação (hidrólise) da celulose (IGARASHI; WADA; SAMEJIMA, 2007; WADA; IKE; TOKUYASU, 2010). Entretanto, uma explicação exata em nível molecular para as diferenças observadas nas taxas de digestão da celulose permanece uma questão em aberto (CHUNDAWAT *et al.*, 2011).

2.1.3 HEMICELULOSE

Hemicelulose é o segundo biopolímero mais abundante no reino vegetal depois da celulose. A sua diferença em relação à celulose está na sua constituição, a qual representa um tipo de hetero-polissacarídeo com estruturas complexas contendo, principalmente, glicose, xilose, manose, galactose e arabinose em várias quantidades dependendo da sua fonte (Figura 7) (REN; SUN, 2010).

Hemiceluloses formam ligações de hidrogênio com a celulose, ligações covalentes com a lignina, e ligações éster com unidades acetila e ácidos hidroxi-cinâmicos. Elas são polímeros ramificados de baixo peso molecular com grau de polimerização de 80 a 200 unidades. Suas fórmulas são, em geral, representadas por $(C_5H_8O_4)_n$ e $(C_6H_{10}O_5)_n$, as quais são chamadas de pentoses e hexoses, respectivamente (DAVISON *et al.*, 2013). Em sua grande maioria, um esqueleto de xiloses unidas por ligações glicosídicas β -1,4 é então ramificado com os outros açúcares, que foram listados acima, em posições β -1 ou β -6. Pectinas também são polímeros ramificados de ácido galacturônico ligados entre si por ligações glicosídicas β -1,4, nos quais arabinose e galactose são os açúcares mais abundantes nas ramificações (BATISTA-GARCÍA *et al.*, 2016).



Figura 7. Carboidratos presentes na lignocelulose, evidenciando as divergências entre os açúcares que compõem a celulose e a hemicelulose.

2.1.4 RECALCITRÂNCIA DA BIOMASSA

Para converter a biomassa lignocelulósica em qualquer bioproduto, os açúcares e fenóis que estão estocados como polímeros de alto peso molecular nas paredes celulares vegetais devem primeiro ser descontruídos e liberados em solução como monossacarídeos ou pequenos oligossacarídeos. A liberação eficiente dessas moléculas com grande rendimento é vital para a produção de biomoléculas com baixo custo. Para alcançar esse objetivo, uma das barreiraschave que devem ser ultrapassadas é a recalcitrância da biomassa, a qual se refere à resistência da parede celular à desconstrução por rotas químicas, enzimáticas e/ou microbianas (DEMARTINI *et al.*, 2013).

A complexidade de um determinado tipo de biomassa se reflete na relação entre os seus componentes estruturais e carboidratos. Os principais fatores que contribuem para a recalcitrância da biomassa, e consequentes diferenças na desconstrução da parede celular (HIMMEL *et al.*, 2007; MOSIER *et al.*, 2005), incluem:

- 1) Cristalinidade e grau de polimerização da celulose;
- 2) Porosidade (ou área superficial disponível);
- 3) Proteção da celulose pela lignina;
- 4) Revestimento da celulose pela hemicelulose;
- 5) Resistência da fibra e

 6) Inibidores enzimáticos que existem naturalmente na parede celular ou que são gerados durante o processo de desconstrução.

No contexto da biorrefinaria, estas características químicas e estruturais da biomassa afetam a penetração de líquidos e/ou a acessibilidade e atividade enzimática e, assim, os custos de conversão. Em princípio, um pré-tratamento eficaz provoca a ruptura destas barreiras de modo que as enzimas hidrolíticas possam penetrar e causar hidrólise na biomassa. Entretanto, estudos mostram que a dificuldade principal não é alcançar bons rendimentos na obtenção de bioprodutos, mas sim a obtenção de bons rendimentos com baixo consumo de energia (ZHU *et al.*, 2010; ZHU; PAN, 2010).

Analisando em nível molecular, a celulose cristalina é altamente resistente à hidrólise química e biológica devido à sua estrutura, na qual as cadeias de celobiose são dispostas com precisão. A conformação da cadeia dos resíduos de glicose em celulose força os grupos hidroxila em orientação radial (equatorial) e os átomos de hidrogênio alifáticos em posições axiais. Como resultado, existe uma forte ligação de hidrogênio intermolecular entre cadeias adjacentes numa folha de celulose e interações hidrofóbicas mais fracas entre folhas de celulose. A face hidrofóbica das folhas de celulose torna a celulose cristalina resistente à hidrólise ácida, porque contribui para a formação de uma camada densa de água próxima à superfície da celulose torna a folha de celulose cristalina resistente à hidrólise enzimática (NISHIYAMA; LANGAN; CHANZY, 2002), enquanto que a hemicelulose e a celulose amorfa são facilmente digeríveis. Numa escala microscópica e macroscópica, a complexa natureza heterogênea da biomassa cria evidentes limitações de transporte de massa para o fornecimento de catalisadores químicos ou bioquímicos.

Devido à sua forte associação com as microfibras de celulose, a lignina tem sido identificada como um importante impedimento à hidrólise enzimática e microbiana da biomassa lignocelulósica. De acordo com a literatura, a digestibilidade da biomassa aumenta proporcionalmente à remoção de lignina (LIMA *et al.*, 2013). Além de ser uma forte barreira física, os efeitos prejudiciais da lignina incluem: 1) adsorção não específica de enzimas hidrolíticas à lignina; 2) interferência e ligação não produtiva de enzimas celulolíticas à complexos de lignina-carboidratos; e 3) toxicidade aos microrganismos pelas moléculas de baixo peso molecular derivadas da quebra da lignina (AGBOR *et al.*, 2011; HIMMEL *et al.*, 2007).

Assim, a deslignificação (remoção da lignina) é uma das principais alternativas para diminuir a recalcitrância da biomassa. Esse processo causa a ruptura em diversos pontos da

estrutura da lignina, dilatação/inchaço na biomassa, aumento da área de superfície interna e aumento da acessibilidade das enzimas celulolíticas às fibras celulósicas. Embora nem todos os pré-tratamentos resultem em uma deslignificação substancial, em alguns casos alterações nas propriedades químicas da lignina já são suficientes para tornar a biomassa pré-tratada mais disponível para digestão do que a biomassa bruta (AGBOR *et al.*, 2011).

Apesar de grandes investimentos em pesquisa e avanços significativos em tecnologia para desconstrução da lignocelulose, esta continua a ser um dos pontos críticos para o estabelecimento de biorrefinarias economicamente viáveis. Geralmente, o material vegetal tem de ser submetido a um pré-tratamento físico e/ou químico seguido por hidrólise enzimática para se obter uma fonte de carbono acessível aos microrganismos (CLEMENTS; VAN DYNE, 2008).

2.1.5 PRÉ-TRATAMENTO DA BIOMASSA

As estratégias para o pré-tratamento da biomassa lignocelulósica foram desenvolvidas e refinadas há mais de 100 anos, originando-se de sistemas de pré-tratamento com ácido, seguidos de sistemas de vaporização e explosão à vapor para o fracionamento de biomassa desenvolvido na década de 1920. Mais recentemente, foram propostas estratégias envolvendo solventes e outros agentes químicos. Apesar dessa história relativamente longa, a pesquisa em tecnologias de pré-tratamento tem aumentado rapidamente nos últimos anos, em grande parte devido ao fato de que o pré-tratamento é a chave para desvendar o potencial de biocombustíveis celulósicos e produtos químicos orgânicos. Sem dúvida, os efeitos do pré-tratamento são extensos, impactando os rendimentos globais de bioprodutos, os custos de capital e de operação, a utilização de enzimas, a fermentação, a destilação e a eliminação de resíduos (SAVILLE, 2011).

Sabe-se que não existe um pré-tratamento que seja 100% eficaz devido às diferentes naturezas das matérias-primas, entretanto, é possível afirmar quais características são desejáveis em processo de desconstrução da biomassa (GALBE; ZACCHI, 2007). Ele deve:

- Resultar numa elevada recuperação de todos os hidratos de carbono, com um mínimo de degradação de produtos;
- 2) Ter um baixo custo operacional e de capital;
- Produzir uma fração de sólidos altamente digerível passível de subsequente hidrólise enzimática;

- Minimizar a necessidade de pré ou pós-processamento, quer sob a forma de redução mecânica do tamanho ou de desintoxicação à jusante;
- 5) Operar com uma carga de sólidos suficientemente elevada para evitar a diluição de açúcares e etanol que de outro modo afetariam negativamente os custos de processamento à jusante e
- 6) Ser capaz de processar uma grande variedade de matérias primas lignocelulósicas.

Uma vasta quantidade de diferentes métodos de pré-tratamento tem sido sugerida durante as últimas décadas, podendo eles ser divididos em categorias diferentes:

- 1) Físicos (moagem, trituração, irradiação, etc.);
- 2) Químicos (álcali, ácido diluído, agentes oxidantes, solventes orgânicos, etc.);
- 3) Físico-químicos (auto-hidrólise à vapor, hidrotermólise, oxidação húmida, etc.);
- 4) Biológicos ou
- 5) Combinações destes.

Embora seja difícil colocar os métodos em uma única categoria, em geral todos apresentam o mesmo objetivo final, que é o de vencer a recalcitrância da parede celular vegetal, afrouxando os componentes estruturais (Figura 8).





Apesar dos métodos de pré-tratamento físicos e químicos serem considerados como líderes nas tecnologias atuais de pré-tratamento, eles geralmente precisam de onerosos reatores resistentes à corrosão, processam grandes volumes de resíduos, lavam exaustivamente os sólidos tratados, e precisam desintoxicar compostos inibitórios aos microrganismos fermentadores (SAVILLE, 2011). Embora o pré-tratamento tenha um custo elevado, não pré-tratar é ainda mais oneroso (EGGEMAN; ELANDER, 2005). Assim, o pré-tratamento continua a ser um dos passos mais dispendiosos na produção de bioprodutos e é uma barreira significativa para sua comercialização (NAIK *et al.*, 2010).

Pré-tratamentos biológicos têm se destacado e sido principalmente associados à ação de microrganismos capazes de produzir enzimas que podem desconstruir a parede celular. Verificou-se que os fungos de podridão branca e macia (White-rot e Soft-rot, respectivamente) degradam o material de lignocelulose, sendo os de podridão branca os mais eficazes no prétratamento biológico da biomassa (Lee, 1997; Sun e Cheng, 2002). Os fungos de podridão marrom (Brown-rot) atacam principalmente a celulose, enquanto que os White e Soft-rot atacam tanto a lignina quanto a celulose através da produção de enzimas, tais como lignina peroxidases (LiP), polifenol oxidases, peroxidases manganês-dependente (MnP) e laccases (Lac) que degradam a lignina. Pesquisas descreveram que *Pleurotus ostreatus* e outros fungos de podridão branca possuem grande habilidade de degradar seletivamente a lignina, hemicelulose e celulose, respectivamente (FAIX et al., 1991; MARTÍNEZ et al., 2005; TANIGUCHI et al., 2005). Estudos relataram a deslignificação seletiva de madeira e palha de trigo por certos fungos de podridão-branca, tais como Phanerochaete chrysosporium, Phlebia radiata, Dichmitus squalens, Rigidosporus lignosus e Jungua separabilima. A despolimerização da lignina por estes fungos leva semanas para conseguir um resultado significativo, mas pode ser muito seletiva e eficiente (LUNDELL; MÄKELÄ; HILDÉN, 2010; RUIZ-DUENAS et al., 2013).

A baixa demanda de energia no pré-tratamento biológico tem gerado interesse na produção de biocombustíveis e bioprodutos a partir de materiais lignocelulósicos. O inconveniente mais significativo do pré-tratamento biológico é o longo tempo necessário para o processo. Entretanto, a combinação de pré-tratamento biológico com outros métodos de pré-tratamento pode reduzir o tempo necessário para que todo o processo se complete (SHIRKAVAND *et al.*, 2016).

A deslignificação biológica a partir de enzimas utiliza extratos enzimáticos e enzimas ligninolíticas purificadas ou parcialmente purificadas para realizar a degradação da lignina. A laccase é a ligninase mais utilizada, seguida de MnP e LiP, embora também tenham sido utilizadas misturas de duas ou três enzimas ligninolíticas. Entretanto, as enzimas ligninolíticas podem reduzir a conversão de celulose num processo de sacarificação. Para evitar a inibição das celulases, as enzimas ligninolíticas podem ser desativadas antes do processo de sacarificação (ESCAMILLA-ALVARADO et al., 2016).

As estratégias de pré-tratamento combinatórias são geralmente mais eficazes no aumento da digestibilidade da biomassa, do rendimento dos produtos (SHIRKAVAND et al., 2016), e muitas vezes são empregadas na concepção de tecnologias de pré-tratamento tidas como líderes. Combinações de pré-tratamentos incluem ao tratamento biológico, os tratamentos ácido (ISHOLA; ISROI; TAHERZADEH, 2014; KUHAR; NAIR; KUHAD, 2008; MA *et al.*,

2010), alcalino (YU *et al.*, 2010a, 2010b), explosão à vapor (TANIGUCHI *et al.*, 2010; VAIDYA; SINGH, 2012), solvente orgânico (FISSORE *et al.*, 2010; ITOH *et al.*, 2003; MONRROY *et al.*, 2010), extração com água quente (WAN; LI, 2011), ultrassônico com peróxido de hidrogênio (SEINO *et al.*, 2001; YU *et al.*, 2009), entre outros.

2.2 RÚMEN BOVINO: UMA FONTE DE MICRORGANISMOS DE INTERESSE INDUSTRIAL

Apesar dos grandes investimentos em pesquisa e avanços significativos na tecnologia, a desconstrução da parede celular continua a ser um dos pontos críticos para o estabelecimento de biorrefinarias econômicas. Geralmente, o material vegetal tem de ser submetido a um prétratamento seguido por hidrólise enzimática para se obter uma fonte de carbono acessível do ponto de vista microbiológico. Uma grande variedade de bactérias e fungos que servem como decompositores no ciclo natural do carbono secretam enzimas celulolíticas industrialmente úteis (WILSON, 2011).

Ao se analisar os ruminantes, a situação é bastante semelhante. A ruminação nada mais é do que o pré-tratamento da biomassa para torná-la enzimaticamente acessível. Durante a ruminação, a biomassa é degradada no rúmen pela comunidade microbiana complexa e convertida em compostos, que são úteis para o ruminante. Esse processo é bastante interessante, pois inclui, entre outros, um pré-tratamento físico elegante e eficaz; a operação em alta carga de sólidos em condições não-assépticas; exigências mínimas de nutrientes, além da própria biomassa vegetal; fermentação eficiente de quase todos os componentes da planta; recuperação eficiente de produtos finais de fermentação primária; e produção de coprodutos úteis (WEIMER; RUSSELL; MUCK, 2009). Isso é o que toda biorrefinaria moderna almeja.

O rúmen é um dos mais complexos e fascinantes ecossistemas microbianos na natureza. Os ruminantes desenvolveram uma relação simbiótica com um microbioma complexo constituído por bactérias, arqueias, fungos, protozoários e vírus localizados no rúmen, os quais são responsáveis pela degradação microbiana da lignocelulose. A fermentação dos açúcares solúveis liberados produz ácidos graxos de cadeia curta que são absorvidos pelo epitélio do rúmen e utilizados pelo ruminante para crescimento (CREEVEY *et al.*, 2014; SAUER; MARX; MATTANOVICH, 2012).

Essa simbiose é estrategicamente importante, pois permite que estes animais utilizem o componente lignocelulósico do material vegetal como principal fonte de energia. Assim, o rúmen se assemelha perfeitamente a uma solução para um problema industrial atual: a biorrefinaria, que visa a bioconversão de material lignocelulósico em combustíveis e produtos químicos.

Dado o potencial desse microbioma para aplicações industriais, foi iniciado um importante projeto chamado Hungate1000 que tem como objetivo produzir um conjunto de sequências de referência de mil genomas microbianos do rúmen através do sequenciamento dos genomas de bactérias do rúmen e arqueias metanogênicas, juntamente com culturas representativas de fungos anaeróbios do rúmen e protozoários ciliados (CREEVEY *et al.*, 2014). O projeto conta com um banco de dados que já contém cerca de 450 genomas sequenciados, catalogados e depositados no Joint Genome Institute (JGI), do Departamento de Energia dos EUA (http://genome.jgi.doe.gov/TheHunmicrobiome), os quais serão utilizados nesse trabalho para entender o potencial degradador da biomassa lignocelulósica pelas enzimas produzidas por estes microrganismos.

2.3 ENZIMAS LIGNOCELULOLÍTICAS

Uma grande variedade de bactérias e fungos que servem como decompositores no ciclo natural do carbono secretam enzimas industrialmente úteis, as quais atuam corporativamente e atacam a estrutura complexa da biomassa lignocelulósica (LYND *et al.*, 2002; VAN DYK; PLETSCHKE, 2012). Enzimas com diferentes especificidades são necessárias para degradar todos os componentes da lignocelulose (Tabela 2), sendo possível encontrar muitos trabalhos sobre este tema na literatura (BANERJEE; SCOTT-CRAIG; WALTON, 2010; GILBERT, 2010; GILBERT; STÅLBRAND; BRUMER, 2008; LOMBARD et al., 2014; LYND et al., 2002; SAHA, 2003; ZHANG; LYND, 2004).

Macromolécula		Principais enzimas
Lignina	Lignina peroxidase Laccase	Manganês peroxidase
Destine	Pectina metilesterase	Poligalacturonase
recuna	Pectato liase	Rhamnogalacturonana liase
	Endo-xilanase	α -L-arabinofuranosidase
	Acetilxilana esterase	α- glucuronidase
Hemicelulose	α-xilosidase	Ácido ferúlico esterase
	Endo-mannanase	α -galactosidase
	α-mannosidase	Ácido Q-coumárico esterase
Coluloso	Celobiohidrolase	a alucosidase
Celulose	Endoglucanase	u-giucosidase

Tabela 2. Principais enzimas relacionadas à desconstrução da parede celular vegetal.

Existem principalmente quatro grupos de enzimas identificadas como sendo relevantes para as biorrefinarias: amilases, celulases, xilanases e lignina-peroxidases. Estas têm aplicações em indústrias como têxtil, alimentos, celulose e papel, e biocombustíveis, entre outras. Outras enzimas derivadas da biomassa vegetal com propriedades notáveis e potencial promissor em biorrefinarias são lipases, proteases e pectinases (UÇKUN KIRAN *et al.*, 2014).

Segundo relatório publicado pela empresa *Markets-and-Markets* (a maior empresa de pesquisa de mercado mundial em termos de relatórios de pesquisa de mercado), o mercado de enzimas industriais (por exemplo, celulases, proteases, lipases, catalases e tanases) foi avaliado em 420 milhões de dólares em 2014 e projetado para crescer anualmente a 7,0% a partir de 2015 para 2020 (ESCAMILLA-ALVARADO *et al.*, 2016). As aplicações destas enzimas, além da produção de biocombustíveis, incluem alimentos e bebidas, agentes de limpeza, ração animal, processamento de amido, têxteis, couro, entre outros.

É possível encontrar na literatura diversas revisões exaustivas que descrevem profundamente cada tipo de enzima citada anteriormente, sua cinética, mecanismo de ação, substratos, produtos e seus microrganismos produtores, (CHUNDAWAT et al., 2011; HORN et al., 2012; HOWARD et al., 2003; KASHYAP et al., 2001; LYND et al., 2002; POLLEGIONI; TONIN; ROSINI, 2015; TETER et al., 2008), portanto, serão descritas apenas as informações mais relevantes sobre elas.

2.3.1 CELULASES E HEMICELULASES

As celulases e hemicelulases são as principais enzimas para valorização da biomassa residual por sacarificação. De acordo com o banco de dados ExPASy (http://www.expasy.org), as celulases são uma família constituída principalmente por 1) endoglucanases (EC 3.2.1.4), que atacam as ligações β -1,4 nas regiões amorfas dentro da cadeia de glucano, 2) exoglucanases (EC 3.2.1.74 e EC 3.2.1.91), que separam a celobiose das extremidades redutoras e não redutoras, e 3) β -glucosidases (EC 3.2.1.21), que convertem a celobiose em monômeros de glicose. Estas enzimas atuam sinergicamente para despolimerizar a celulose efetivamente (MENON; RAO, 2012).

Considerando que, como descrito anteriormente, a xilana é o polissacarídeo mais abundante na hemicelulose, as hemicelulases mais importantes são a β -1,4-endoxilanase (EC 3.2.1.8), que cliva as ligações β -1,4-D-xilopiranosil das xilanas, e β -1,4-xilosidases (EC 3.2.1.37), que hidrolisam xiloligossacarídeos curtos em xilose. As hemicelulases envolvidas na clivagem dos demais resíduos da hemicelulose são as α -L-arabinofuranosidases (EC 3.2.1.55),

 α -glucuronidases e acetilxilana esterases (EC 3.1.1.72). Essas enzimas são conhecidas como enzimas acessórias e, do ponto de vista da sacarificação, elas são importantes não por causa de sua abundância, mas pela abertura de poros na estrutura da xilana que permite o ataque de xilanases (GÍRIO *et al.*, 2010).

2.3.2 LIGNINASES

Lignina-Peroxidase (LiP) (EC1.11.1.14) e Manganês-Peroxidase (MnP) (EC1.11.1.13) possuem um alto potencial redox e atuam na oxidação de compostos fenólicos e não-fenólicos da lignina gerando radicais aromáticos. A LiP degrada unidades de lignina não-fenólicas (até 90% do polímero), enquanto que a MnP gera Mn³⁺, que atua como um oxidante difuso em unidades de lignina fenólicas ou não-fenólicas através de reações de peroxidação lipídica (MARTÍNEZ *et al.*, 2005). Laccases (EC 1.10.3.2) são enzimas oxidases que contém cobre e agem sobre fenóis e moléculas semelhantes, realizando oxidações de um elétron (POLLEGIONI; TONIN; ROSINI, 2015). Devido à natureza complicada e heterogênea da biomassa lignocelulósica, o papel e o mecanismo de ação exato das enzimas ligninolíticas na degradação das matérias primas de biomassa ainda não foi definido. Contudo, as enzimas oxidativas podem ser ativas na superfície da parede celular e induzir a formação de moléculas de baixa peso molecular, tais como radicais e ácidos oxálicos (WAN; LI, 2012).

2.3.3 ENZIMAS DE ATIVIDADES AUXILIARES

Recentes evidências destacam o papel fundamental de enzimas alternativas que estão envolvidas na oxidação de outros componentes da parede celular vegetal. Originalmente classificadas como celulases, as enzimas da família 61 das Glicosídeo Hidrolases (GH61) foram descritas recentemente como sendo Mono-oxigenases Líticas de Polissacarídeo Dependentes de Cobre (LPMO), as quais aumentam a degradação da celulose em consonância com celulases clássicas. Tais catalisadores receberam o nome de enzimas de Atividades Auxiliares (AA), existindo atualmente mais de treze mil módulos, classificados em treze famílias, depositados no banco de dados CAZy (www.cazy.org) (CANTAREL *et al.*, 2009).

Além das Expansinas (COSGROVE, 2015) e Swolleninas (GOURLAY et al., 2013), outra classe de proteínas que está diretamente ligada à desconstrução da parede celular é o Módulo de Ligação aos Carboidratos (CBM), o qual é definido como uma sequência de aminoácido contíguo contida em uma Enzima Carboidrato-Ativa (CAZyme) com uma discreta região que possui atividade de ligação a carboidratos, a qual refina o reconhecimento e adesão aos carboidratos da parede celular vegetal pelas CAZymes (BORASTON *et al.*, 2004).

2.4 METAGENÔMICA COMO FERRAMENTA DE DESCOBERTA DE NOVOS GENES

O maior obstáculo para a produção em escala industrial de biocombustíveis bioprodutos a partir de lignocelulose reside na desconstrução ineficiente do material vegetal devido à natureza recalcitrante do substrato frente a quebra enzimática, e à relativa baixa atividade das enzimas disponíveis atualmente (BLANCH *et al.*, 2008).

O grande número de estudos acerca das vias de degradação da lignocelulose não apenas permite entender os mecanismos e interações entre organismos na manutenção do balanço carbônico nos ciclos biogeoquímicos, mas permite também levar à descobertas potenciais de microrganismos não caracterizados e novas enzimas que, por sua vez, poderiam melhorar a conversão de biomassa vegetal subutilizada à biocombustíveis, químicos e outros materiais através de biorefinarias (LEE, 1997; RUBIN, 2008).

Embora o sucesso na engenharia de proteínas para aumentar o desempenho de enzimas degradadoras de lignocelulose existentes tem sido limitado (WEN; NAIR; ZHAO, 2009), a recuperação de enzimas de comunidades microbianas naturalmente envolvidas com a degradação de biomassa oferece uma estratégia promissora à identificação de novas enzimas lignocelulolíticas com atividades potencialmente melhores (RUBIN, 2008).

A maioria das enzimas tem sido identificadas utilizando laboriosos métodos de triagem e extensas caracterizações bioquímicas, mas evidências baseadas na bioinformática aumentaram a percepção que a abordagem *in silico* pode ser aplicada na identificação de novas enzimas a partir de sequências de genomas microbianos (WANG; FEWER; SIVONEN, 2011).

A metagenômica consiste na análise direta do DNA de diferentes amostras de um determinado ambiente e ela representa uma estratégia para descobrir diversas enzimas codificadas na natureza (FERRER *et al.*, 2005; WANG *et al.*, 2009). A abordagem metagenômica tem sido utilizada com sucesso para identificar enzimas com atividades de interesse industrial (LI *et al.*, 2009).

Uma vez que o DNA metagenômico é obtido, a prospecção de novas enzimas ocorre através da triagem de bibliotecas de expressão, seja ela baseada em funções ou em similaridade das sequências. Em estudo, Li e colaboradores (2009) destacam a relevância e vantagens do método de triagem metagenômico baseado em homologia de sequências. Segundo eles, ao contrário dos métodos baseados em funções, é possível detectar um número maior de genes

alvo, uma vez que esse método independe da expressão gênica, do dobramento de proteínas no hospedeiro, das modificações pós-traducionais, entre outros. Ou seja, esse método torna-se bastante atrativo pois ele independe da integridade da sequência do gene alvo.

Nesse contexto, essa pesquisa estudou as vias de degradação da lignocelulose, as enzimas utilizadas nesse processo e realizou uma triagem de genes lignocelulolíticos de microrganismos do rúmen através de análises metagenômicas por homologia de sequências.

3 JUSTIFICATIVA

Lignocelulose é um constituinte básico da biomassa vegetal e representa uma das mais abundantes fontes de carbono recicláveis da biosfera. Sua complexa estrutura consiste principalmente de polímeros de carboidrato: celulose, hemicelulose e lignina. Na natureza, a degradação da lignocelulose requer múltiplas enzimas produzidas por diversos microrganismos, os quais atuam corporativamente e atacam a estrutura complexa da biomassa lignocelulósica. O grande número de estudos acerca das vias de degradação da lignocelulose não apenas permite entender os mecanismos e interações entre microrganismos na manutenção do balanço carbônico nos ciclos biogeoquímicos, mas permite também levar a descobertas potenciais de microrganismos não caracterizados e novas enzimas que, por sua vez, poderiam melhorar a conversão de biomassa vegetal subutilizada à biocombustíveis, químicos e outros materiais através de biorrefinarias. A maioria das enzimas tem sido identificadas utilizando laboriosos métodos de triagem e extensas caracterizações bioquímicas, mas evidências baseadas na bioinformática aumentaram a noção que a abordagem in silico pode ser aplicada na identificação de novas enzimas a partir de sequências de genomas microbianos. A metagenômica, análise direta do DNA de diferentes amostras de um determinado ambiente, representa uma estratégia para descobrir diversas enzimas codificadas na natureza. O rúmen é um dos mais complicados e mais fascinantes ecossistemas microbianos na natureza. Os ruminantes desenvolveram uma relação simbiótica com um complexo microbioma (bactérias, arqueias, fungos, protozoários e vírus) localizado no pré-estômago que permite estes animais utilizarem o componente lignocelulósico de materiais vegetais como sua principal fonte de energia. Dessa forma, o rúmen pode contribuir ativamente com as biorrefinarias, levando à identificação de enzimas-chave para o desenvolvimento de bioprocessos de bioconversão do material lignocelulósico em combustíveis e químicos.

4 OBJETIVO GERAL

Identificar enzimas com potencial atividade lignocelulósica presentes em organismos oriundos do rúmen utilizando técnicas de bioinformática para metagenomas.

4.1 OBJETIVOS ESPECÍFICOS

- Classificar os genomas do rúmen quanto à taxonomia;
- Identificar enzimas responsáveis pela degradação da lignocelulose presentes nos genomas dos microrganismos do Projeto Hungate1000;
- Classificar as enzimas de acordo com sua atividade lignocelulósica em consonância com a classificação do CAZyDB (GH, CBM, PL, CE ou AA);
- Determinar o potencial dos microrganismos quanto à degradação da biomassa lignocelulósica;
- Classificar os genomas do rúmen nos grupos de genes ortólogos (COGs) e observar a porcentagem de dedicação dos mesmos.

5 METODOLOGIA

5.1 RECUPERAÇÃO DE SEQUÊNCIAS

O Projeto Hungate1000 (CREEVEY *et al.*, 2014) é composto de 436 genomas de microrganismos presentes no rúmen de bovinos, os quais estão depositados na base de dados do DOE Joint Genome Institute, sendo considerados para esse estudo apenas os genomas completos e devidamente anotados. O download dos mesmos foi feito conforme instruído pelo JGI (http://genome.jgi.doe.gov/help/download.jsf) via API (*Application Programming Interface*), plataforma que promove uma interface de programação que, nesse caso, possibilitou o acesso ao banco de dados do JGI através de linhas de comando por ele definidas. O script utilizado foi escrito em Perl e baseado em quatro etapas principais:

 Cada microrganismo possui um portal único, no qual todas as informações pertinentes a ele estão depositadas. Por exemplo, os dados do microrganismo *Treponema sp.* C6A8 estão depositados no endereço <http://genome.jgi.doe.gov/TreponemaspC6A8_FD>, logo o nome do portal desse microrganismo é *TreponemaspC6A8_FD*. Assim, a primeira etapa foi identificar o nome do portal de cada microrganismo em estudo, o qual foi gerado pelo próprio JGI e exportado em formato CSV para que então pudessem ser feitas iterações de todos os microrganismos de forma automatizada.

 Para baixar os arquivos é necessário estar conectado ao JGI, o que foi feito através do comando

> curl 'https://signon.jgi.doe.gov/signon/create' --data-urlencode 'login=USER_NAME' --data-urlencode 'password=USER_ PASSWORD' -c cookies > /dev/null

substituindo USER_NAME e USER_PASSWORD com os valores apropriados.

 Uma vez conectado, foram baixadas as listas de arquivos disponíveis para os portais dos microrganismos em estudo e exportando-os para XML através do comando

> curl 'http://genome.jgi.doe.gov/ext-api/downloads/get-directory? organism=PORTAL_NAME' -b cookies > files_list.xml

substituindo PORTAL_NAME pelo nome do portal do microrganismo desejado.

 O arquivo XML exportado foi estudado e, ao se identificar o URL do arquivo que contém o proteoma do microrganismo em FASTA, este foi baixado através do comando

> curl 'http://genome.jgi.doe.gov/ext-api/downloads/get_tape_file? blocking=true&url=/FILE_URL' -b cookies > proteome_file.tar .gz

substituindo FILE_URL pelo endereço apropriado.

O banco de dados CAZy contém uma coletânea de inúmeros módulos catalíticos (enzimas) subdivididos em várias classes de enzimas e suas famílias que catalisam a quebra, biossíntese e/ou modificação de glicoconjugados, oligo e polissacarídeos, além de famílias de módulos encontrados em conjunto com os módulos catalíticos (CBM) (CANTAREL *et al.*, 2009). O CAZyDB organiza, em forma de tabela em HTML, algumas informações sobre a proteína, como seu nome, organismo produtor, GenBank ID, entre outros.

Assim, a primeira etapa baseou-se no desenvolvimento de um script em Perl para recuperar todos os GenBank IDs depositados, conforme descrito a seguir:

 As páginas do CAZyDB seguem o seguinte padrão: http://www.cazy.org/'CLASSE''FAMILIA'_'DOMINIO'.html, onde, por exemplo, todas as proteínas da classe Glicosil Hidrolase (GH), família 1, produzidas por bactérias estão depositadas no endereço http://www.cazy.org/GH1_bacteria.html. Dessa forma, foi possível criar um script que baixasse todas as páginas do banco de dados em formato HTML utilizando um *loop* para combinar as variáveis "Classe" (GH, PL, CE, AA, CBM), "Família" (144, 26, 16, 13, 81), e "Domínio" (archaea, bacteria, eukaryota, viruses), respectivamente.

2. Foi utilizado o módulo HTML::TableExtract para Perl afim de extrair da tabela em HTML os IDs de todas as proteínas presentes na coluna "GenBank" e salvá-los em um arquivo TXT. Para auxiliar na análise dos dados posteriormente, o script foi desenhado para organizar as proteínas em diferentes arquivos classificados por classe e domínio (ex.: AA_archaea.txt). Dentro de cada um deles os IDs estavam seguidos de suas respectivas classes e famílias (ex.: AFS80362.1;AA5).

Na segunda etapa foram recuperadas do site do *National Center for Biotechnology Information* (NCBI) as sequências de aminoácidos em formato FASTA de cada uma das proteínas depositadas no CAZyDB, de acordo com os respectivos GenBank IDs obtidos pelo script acima.

1. Esse processo foi feito via API (E-utilities) de acordo com instruções fornecidas pelo próprio NCBI (https://www.ncbi.nlm.nih.gov/home/tools), utilizando o comando:

curl 'https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db= protein&id="GENBANK_ID"&retmode=text&rettype=fasta' -b cookies > file.fasta

substituindo GENBANK_ID pelo ID apropriado.

2. As sequências recuperadas foram salvas em arquivos FASTA organizados da mesma forma citada acima, classificados por classe e domínio (ex.: AA_archaea.fasta)

As sequências de proteínas com atividade de ligninase depositadas no banco de dados do UniProt correspondentes ao termo de busca 'laccase' foram baixadas em formato FASTA diretamente do seu website, entretanto, apenas as 135 proteínas cujas sequências apresentaram classificação revisada pelo Swiss-Prot foram utilizadas. Todos os dados mencionados acima foram coletados em janeiro de 2017.

5.2 BUSCA POR SIMILARIDADE

A busca por sequências similares às enzimas com potencial de degradação da lignocelulose foi realizada usando as sequências de aminoácidos dos proteomas recuperados do Projeto Hungate1000 comparando-as com as sequências das enzimas dos bancos de dados CAZy e UniProt. Para tanto, a mesma foi realizada através do programa BLASTp 2.6.1+ via script em Perl segundo especificações indicadas pelo manual do usuário

(https://www.ncbi.nlm.nih.gov/books/NBK279675), utilizando um *e-value* menor que 10⁻⁶ (ALTSCHUL et al., 1990, 1997; CAMACHO et al., 2009).

5.3 CLASSIFICAÇÃO TAXONÔMICA

Os genomas que apresentaram similaridade em conformidade com os padrões supracitados foram classificados de acordo com seus níveis taxonômicos (domínio, filo, classe, ordem, família, gênero e espécie), utilizando informações taxonômicas dos genomas depositados no NCBI-Taxonomy.

5.4 DETERMINAÇÃO DO POTENCIAL DE DEGRADAÇÃO DA BIOMASSA

Os resultados do BLASTp foram analisados via script desenhado em linguagem R no programa RStudio (RSTUDIO TEAM, 2015) utilizando um filtro para selecionar e quantificar apenas resultados (*hits*) de proteínas com *e-value* menor que 10⁻⁵⁰ para uma análise mais rigorosa. A tabela criada foi exportada para o Microsoft® Excel 2016 para análise e criação de gráficos e demais tabelas. Assim, foi possível identificar e quantificar as proteínas presentes nos proteomas dos microrganismos do rúmen bovino correspondentes às classes e famílias das enzimas depositadas no CAZyDB e no UniProt. Dessa forma, determinou-se o potencial de degradação dos microrganismos, bem como a diversidade microbiana presente no processo de degradação da lignocelulose e seu caráter coorporativo de acordo com informações disponíveis no próprio CAZyDB e na literatura.

5.5 CLASSIFICAÇÃO DOS GENOMAS DO RÚMEN NOS GRUPOS DE GENES ORTÓLOGOS

A classificação dos genomas foi realizada através da comparação por similaridade entre as proteínas dos genomas do JGI e os Grupos de Genes Ortólogos (COG). A busca por similaridade e classificação funcional COG foi realizada pelo programa COGnitor (KRISTENSEN *et al.*, 2010). Os resultados foram analisados e organizados via script em R no programa RStudio (RSTUDIO TEAM, 2015) utilizando um filtro para selecionar e quantificar apenas dados com *e-value* menor que 10⁻¹⁰. A tabela criada foi exportada para o Microsoft® Excel 2016 para análise e criação de gráficos e demais tabelas. Os grupos de sequências foram separados por filo para observar a porcentagem de dedicação dos genomas nos COGs.

6 REFERÊNCIAS

AGBOR, V. B. et al. Biomass pretreatment: Fundamentals toward application. **Biotechnology Advances**, v. 29, n. 6, p. 675–685, nov. 2011.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, out. 1990.

ALTSCHUL, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic acids research**, v. 25, n. 17, p. 3389–402, set. 1997.

BALAN, V. Current challenges in commercially producing biofuels from lignocellulosic biomass. **ISRN biotechnology**, v. 2014, p. 31, 2014.

BANERJEE, G.; SCOTT-CRAIG, J. S.; WALTON, J. D. Improving Enzymes for Biomass Conversion: A Basic Research Perspective. **BioEnergy Research**, v. 3, n. 1, p. 82–92, 9 mar. 2010.

BATISTA-GARCÍA, R. A. et al. From lignocellulosic metagenomes to lignocellulolytic genes: trends, challenges and future prospects. **Biofuels, Bioproducts and Biorefining**, v. 10, n. 6, p. 864–882, nov. 2016.

BLANCH, H. W. et al. Addressing the need for alternative transportation fuels: the Joint BioEnergy Institute. **ACS chemical biology**, v. 3, n. 1, p. 17–20, 2008.

BORASTON, A. B. et al. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. **The Biochemical journal**, v. 382, n. Pt 3, p. 769–81, 15 set. 2004.

BRUNOW, G. Lignin Chemistry and its Role in Biomass Conversion. In: **Biorefineries -Industrial Processes and Products**. Weinheim: Wiley-VCH Verlag GmbH, 2008. p. 151–163.

BRUNOW, G.; LUNDQUIST, K.; GELLERSTEDT, G. Lignin. In: SJÖSTRÖM, E.; ALÉN, R. (Eds.). . Analytical Methods in Wood Chemistry, Pulping, and Papermaking. 1. ed. Heidelberg: Springer Berlin Heidelberg, 1999. p. 77–124.

CAMACHO, C. et al. BLAST+: architecture and applications. **BMC Bioinformatics**, v. 10, n. 1, p. 421, 2009.

CANTAREL, B. I. et al. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. **Nucleic Acids Research**, v. 37, n. SUPPL. 1, 2009.

CHUNDAWAT, S. P. S. et al. Deconstruction of Lignocellulosic Biomass to Fuels and Chemicals. **Annual Review of Chemical and Biomolecular Engineering**, v. 2, n. 1, p. 121–145, 15 jul. 2011.

CLEMENTS, L. D.; VAN DYNE, D. L. The Lignocellulosic Biorefinery - A Strategy for Returning to a Sustainable Source of Fuels and Industrial Organic Chemicals. In: **Biorefineries** - **Industrial Processes and Products**. Weinheim, Germany: Wiley-VCH Verlag GmbH, 2008. v. 1–2p. 115–128.

COSGROVE, D. J. Growth of the plant cell wall. **Nature reviews. Molecular cell biology**, v. 6, n. 11, p. 850–61, nov. 2005.

COSGROVE, D. J. Plant expansins: diversity and interactions with plant cell walls. **Current Opinion in Plant Biology**, v. 25, p. 162–172, 2015.

CREEVEY, C. J. et al. Determining the culturability of the rumen bacterial microbiome. **Microbial Biotechnology**, v. 7, n. 5, p. 467–479, set. 2014.

DAVISON, B. H. et al. Plant Cell Walls: Basics of Structure, Chemistry, Accessibility and the Influence on Conversion. In: WYMAN, C. E. (Ed.). Aqueous Pretreatment of Plant Biomass for Biological and Chemical Conversion to Fuels and Chemicals. 1. ed. Oxford: John Wiley & Sons, Inc., 2013. p. 23–38.

DEMARTINI, J. D. et al. Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. **Energy & Environmental Science**, v. 6, n. 3, p. 898, 2013.

EGGEMAN, T.; ELANDER, R. T. Process and economic analysis of pretreatment technologies. **Bioresource Technology**, v. 96, n. 18, p. 2019–2025, dez. 2005.

ESCAMILLA-ALVARADO, C. et al. An overview of the enzyme potential in bioenergyproducing biorefineries. **Journal of Chemical Technology & Biotechnology**, out. 2016.

FAIX, O. et al. Monitoring of chemical changes in white-rot degraded beech wood by pyrolysis—gas chromatography and Fourier-transform infrared spectroscopy. **Journal of Analytical and Applied Pyrolysis**, v. 21, n. 1–2, p. 147–162, set. 1991.

FERRER, M. et al. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. **Environmental Microbiology**, v. 7, n. 12, p. 1996–2010, 16 set. 2005.

FISSORE, A. et al. Evaluation of a combined brown rot decay--chemical delignification process as a pretreatment for bioethanol production from Pinus radiata wood chips. **Journal of Industrial Microbiology & Biotechnology**, v. 37, n. 9, p. 893–900, 2010.

GALBE, M.; ZACCHI, G. Pretreatment of Lignocellulosic Materials for Efficient Bioethanol Production. In: OLSSON, L. (Ed.). . **Biofuels**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 41–65.

GILBERT, H. J. The biochemistry and structural biology of plant cell wall deconstruction. **Plant physiology**, v. 153, n. 2, p. 444–455, 2010.

GILBERT, H. J.; STÅLBRAND, H.; BRUMER, H. How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. **Current opinion in plant biology**, v. 11, n. 3, p. 338–48, jun. 2008.

GÍRIO, F. M. et al. Hemicelluloses for fuel ethanol: A review. **Bioresource Technology**, v. 101, n. 13, p. 4775–4800, 2010.

GOURLAY, K. et al. Swollenin aids in the amorphogenesis step during the enzymatic hydrolysis of pretreated biomass. **Bioresource Technology**, v. 142, p. 498–503, ago. 2013.

GREENWELL, H. C.; LLOYD-EVANS, M.; WENNER, C. Biofuels, science and society. **Interface Focus**, v. 3, n. 1, p. 20120093–20120093, 21 dez. 2012.

HARMSEN, P. F. H. et al. Literature Review of Physical and Chemical Pretreatment Processes for Lignocellulosic Biomass. Disponível em: http://edepot.wur.nl/150289>.

Acesso em: 10 fev. 2017.

HIMMEL, M. E. et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. **Science (New York, N.Y.)**, v. 315, n. 5813, p. 804–7, 2007.

HORN, S. J. et al. Novel enzymes for the degradation of cellulose. **Biotechnology for Biofuels**, v. 5, n. 1, p. 45, 2012.

HOWARD, R. L. et al. Lignocellulose biotechnology: issues of bioconversion and enzyme production. African Journal of Biotechnology, v. 2, n. 12, p. 602–619, 31 dez. 2003.

IGARASHI, K.; WADA, M.; SAMEJIMA, M. Activation of crystalline cellulose to cellulose III(I) results in efficient hydrolysis by cellobiohydrolase. **The FEBS journal**, v. 274, n. 7, p. 1785–92, abr. 2007.

ISHOLA, M. M.; ISROI; TAHERZADEH, M. J. Effect of fungal and phosphoric acid pretreatment on ethanol production from oil palm empty fruit bunches (OPEFB). **Bioresource Technology**, v. 165, p. 9–12, ago. 2014.

ITOH, H. et al. Bioorganosolve pretreatments for simultaneous saccharification and fermentation of beech wood by ethanolysis and white rot fungi. **Journal of Biotechnology**, v. 103, n. 3, p. 273–280, 2003.

KAMM, B. et al. Biorefinery Systems – An Overview. In: **Biorefineries-Industrial Processes** and **Products**. Weinheim, Germany: Wiley-VCH Verlag GmbH, 2008. v. 1p. 1–40.

KASHYAP, D. . et al. Applications of pectinases in the commercial sector: a review. **Bioresource Technology**, v. 77, n. 3, p. 215–227, maio 2001.

KRISTENSEN, D. M. et al. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. **Bioinformatics (Oxford, England)**, v. 26, n. 12, p. 1481–7, 15 jun. 2010.

KUBICEK, C. P. The Plant Biomass. In: KUBICEK, C. P. (Ed.). . Fungi and Lignocellulosic Biomass. 1. ed. Oxford: John Wiley & Sons, Inc., 2013. p. 1–28.

KUHAR, S.; NAIR, L. M. N. M.; KUHAD, R. C. C. Pretreatment of lignocellulosic material with fungi capable of higher lignin degradation and lower carbohydrate degradation improves substrate acid hydrolysis and the eventual conversion to ethanol. **Canadian Journal of Microbiology**, v. 54, n. 4, p. 305–313, 2008.

LEE, J. Biological conversion of lignocellulosic biomass to ethanol. Journal of Biotechnology, v. 56, n. 1, p. 1–24, jul. 1997.

LI, L.-L. et al. Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. **Biotechnology for Biofuels**, v. 2, n. 1, p. 10, 2009.

LIMA, M. A. et al. Effects of pretreatment on morphology, chemical composition and enzymatic digestibility of eucalyptus bark: a potentially valuable source of fermentable sugars for biofuel production – part 1. **Biotechnology for Biofuels**, v. 6, n. 1, p. 75, 2013.

LOMBARD, V. et al. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research, v. 42, n. D1, 2014.

LUNDELL, T. K.; MÄKELÄ, M. R.; HILDÉN, K. Lignin-modifying enzymes in filamentous basidiomycetes: ecological, functional and phylogenetic review. **Journal of Basic Microbiology**, v. 50, n. 1, p. 5–20, fev. 2010.

LYND, L. R. et al. Microbial Cellulose Utilization: Fundamentals and Biotechnology. **Microbiology and Molecular Biology Reviews**, v. 66, n. 3, p. 506–577, 1 set. 2002.

MA, F. et al. Combination of biological pretreatment with mild acid pretreatment for enzymatic hydrolysis and ethanol production from water hyacinth. **Bioresource Technology**, v. 101, n. 24, p. 9600–9604, dez. 2010.

MARTÍNEZ, A. T. et al. Biodegradation of lignocellulosics: microbial, chemical, and enzymatic aspects of the fungal attack of lignin. **International microbiology : the official journal of the Spanish Society for Microbiology**, v. 8, n. 3, p. 195–204, set. 2005.

MENON, V.; RAO, M. Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & amp; biorefinery concept. **Progress in Energy and Combustion Science**, v. 38, n. 4, p. 522–550, ago. 2012.

MONRROY, M. et al. Bioorganosolv pretreatments of P. radiata by a brown rot fungus (Gloephyllum trabeum) and ethanolysis. **Enzyme and Microbial Technology**, v. 47, n. 1–2, p. 11–16, jul. 2010.

MOSIER, N. et al. Features of promising technologies for pretreatment of lignocellulosic biomass. **Bioresource technology**, v. 96, n. 6, p. 673–86, abr. 2005.

NAIK, S. N. et al. Production of first and second generation biofuels: A comprehensive review. **Renewable and Sustainable Energy Reviews**, v. 14, n. 2, p. 578–597, 2010.

NISHIYAMA, Y.; LANGAN, P.; CHANZY, H. Crystal structure and hydrogen-bonding system in cellulose I-beta from synchrotron X-ray and neutron fiber diffraction. Journal of the American Chemical Society, v. 124, n. 31, p. 9074–82, 7 ago. 2002.

PAULY, M.; KEEGSTRA, K. Cell-wall carbohydrates and their modification as a resource for biofuels. **The Plant journal: for cell and molecular biology**, v. 54, n. 4, p. 559–68, maio 2008.

POLLEGIONI, L.; TONIN, F.; ROSINI, E. Lignin-degrading enzymes. **FEBS Journal**, v. 282, n. 7, p. 1190–1213, abr. 2015.

POTUMARTHI, R.; BAADHE, R. R.; BHATTACHARYA, S. Fermentable Sugars from Lignocellulosic Biomass: Technical Challenges. In: GUPTA, V. K.; TUOHY, M. G. (Eds.). . **Biofuel Technologies: Recent Developments**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 3–27.

REN, J.; SUN, R. Hemicelluloses. In: SUN, R. (Ed.). . Cereal Straw as a Resource for Sustainable Biomaterials and Biofuels. 1. ed. Oxford: Elsevier B.V., 2010. p. 73–130.

REN21. Renewable Energy Policy Network for the 21st century - 2016 Global Status Report. Disponível em: http://www.ren21.net/REN21Activities/GlobalStatusReport.aspx>.

RSTUDIO TEAM. **RStudio: Integrated Development for R. RStudio, Inc**. Boston, MA, 2015. Disponível em: http://www.rstudio.com>

RUBIN, E. M. Genomics of cellulosic biofuels. Nature, v. 454, n. 7206, p. 841-845, 14 ago.

2008.

RUIZ-DUENAS, F. J. et al. Lignin-degrading peroxidases in Polyporales: an evolutionary survey based on 10 sequenced genomes. **Mycologia**, v. 105, n. 6, p. 1428–1444, 1 nov. 2013.

SAHA, B. C. Hemicellulose bioconversion. Journal of Industrial Microbiology and Biotechnology, v. 30, n. 5, p. 279–291, 1 maio 2003.

SAUER, M.; MARX, H.; MATTANOVICH, D. From rumen to industry. Microbial Cell Factories, v. 11, n. 1, p. 121, 2012.

SAVILLE, B. A. Pretreatment Options. In: HOOD, E. E.; NELSON, P.; POWELL, R. (Eds.). . **Plant Biomass Conversion**. 1. ed. Oxford: John Wiley & Sons, Inc., 2011. p. 199–226.

SEINO, T. et al. ESR studies of radicals generated by ultrasonic irradiation of lignin solution. An application of the spin trapping method. **Wood Science and Technology**, v. 35, n. 1, p. 97–106, 2001.

SHIRKAVAND, E. et al. Combination of fungal and physicochemical processes for lignocellulosic biomass pretreatment - A review. **Renewable and Sustainable Energy Reviews**, v. 54, p. 217–234, 2016.

TANIGUCHI, M. et al. Evaluation of pretreatment with Pleurotus ostreatus for enzymatic hydrolysis of rice straw. **Journal of Bioscience and Bioengineering**, v. 100, n. 6, p. 637–643, dez. 2005.

TANIGUCHI, M. et al. Effect of steam explosion pretreatment on treatment with Pleurotus ostreatus for the enzymatic hydrolysis of rice straw. **Journal of Bioscience and Bioengineering**, v. 110, n. 4, p. 449–452, out. 2010.

TETER, S. A. et al. Enzymes for Biorefineries. In: **Biorefineries-Industrial Processes and Products**. Weinheim, Germany: Wiley-VCH Verlag GmbH, 2008. v. 1p. 356–383.

UÇKUN KIRAN, E. et al. Enzyme Production from Food Wastes Using a Biorefinery Concept. **Waste and Biomass Valorization**, v. 5, n. 6, p. 903–917, 2014.

VAIDYA, A.; SINGH, T. Pre-treatment of Pinus radiata substrates by basidiomycetes fungi to enhance enzymatic hydrolysis. **Biotechnology Letters**, v. 34, n. 7, p. 1263–1267, 2012.

VAN DYK, J. S.; PLETSCHKE, B. I. A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes—Factors affecting enzymes, conversion and synergy. **Biotechnology Advances**, v. 30, n. 6, p. 1458–1480, nov. 2012.

VOXEUR, A.; WANG, Y.; SIBOUT, R. Lignification: different mechanisms for a versatile polymer. **Current opinion in plant biology**, v. 23, p. 83–90, fev. 2015.

WADA, M.; IKE, M.; TOKUYASU, K. Enzymatic hydrolysis of cellulose I is greatly accelerated via its conversion to the cellulose II hydrate form. **Polymer Degradation and Stability**, v. 95, n. 4, p. 543–548, abr. 2010.

WAN, C.; LI, Y. Effect of hot water extraction and liquid hot water pretreatment on the fungal degradation of biomass feedstocks. **Bioresource Technology**, v. 102, n. 20, p. 9788–9793, out. 2011.

WAN, C.; LI, Y. Fungal pretreatment of lignocellulosic biomass. **Biotechnology Advances**, v. 30, n. 6, p. 1447–1457, nov. 2012.

WANG, F. et al. Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. **Microbiological research**, v. 164, n. 6, p. 650–7, nov. 2009.

WANG, H.; FEWER, D. P.; SIVONEN, K. Genome Mining Demonstrates the Widespread Occurrence of Gene Clusters Encoding Bacteriocins in Cyanobacteria. **PLoS ONE**, v. 6, n. 7, p. e22384, 20 jul. 2011.

WEIMER, P. J.; RUSSELL, J. B.; MUCK, R. E. Lessons from the cow: What the ruminant animal can teach us about consolidated bioprocessing of cellulosic biomass. **Bioresource Technology**, v. 100, n. 21, p. 5323–5331, nov. 2009.

WEN, F.; NAIR, N. U.; ZHAO, H. **Protein engineering in designing tailored enzymes and microorganisms for biofuels productionCurrent Opinion in Biotechnology**, 2009.

WILSON, D. B. Microbial diversity of cellulose hydrolysis. Current Opinion in Microbiology, v. 14, n. 3, p. 259–263, 2011.

YU, H. et al. Fungal treatment of cornstalks enhances the delignification and xylan loss during mild alkaline pretreatment and enzymatic digestibility of glucan. **Bioresource Technology**, v. 101, n. 17, p. 6728–6734, set. 2010a.

YU, H. et al. Evaluation of white-rot fungi-assisted alkaline/oxidative pretreatment of corn straw undergoing enzymatic hydrolysis by cellulase. **Journal of Bioscience and Bioengineering**, v. 110, n. 6, p. 660–664, dez. 2010b.

YU, J. et al. Combinations of mild physical or chemical pretreatment with biological pretreatment for enzymatic hydrolysis of rice hull. **Bioresource Technology**, v. 100, n. 2, p. 903–908, jan. 2009.

ZHANG, Y.-H. P.; LYND, L. R. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. **Biotechnology and Bioengineering**, v. 88, n. 7, p. 797–824, 30 dez. 2004.

ZHAO, Q. Lignification: Flexibility, Biosynthesis and Regulation. **Trends in plant science**, v. 21, n. 8, p. 713–21, ago. 2016.

ZHU, J. Y. et al. Pretreatment of woody biomass for biofuel production: energy efficiency, technologies, and recalcitrance. **Applied Microbiology and Biotechnology**, v. 87, n. 3, p. 847–857, 15 jul. 2010.

ZHU, J. Y.; PAN, X. J. Woody biomass pretreatment for cellulosic ethanol production: Technology and energy consumption evaluation. **Bioresource Technology**, v. 101, n. 13, p. 4992–5002, 2010.

7 ARTIGO

Screening of lignocellulolytic enzymes through analysis of microbial rumen proteomes indicates *Firmicutes* and *Bacteroidetes* as main CAZymes and CBM producers

Lucas Magalhães Alcantara^{1*}, Leandro Martins de Freitas¹, Andrea Lopes de Oliveira Ferreira²

 Núcleo de Biointegração, Universidade Federal da Bahia, Vitória da Conquista, BA, Brazil
 Departamento de Engenharia Química, Centro de Tecnologia da Universidade Federal da Paraíba (CT/UFPB)
 *Correspondence: lucas.ma@outlook.com leandromartins@ufba.br andreaferreira456@gmail.com

Abstract

Background: In the rumen, microbial population converts the complex food stuffs (lingo-cellulosic plant materials) into simpler compounds, which can be easily taken up and metabolized by ruminants. Therefore, direct use of microbial population from rumen or enzymes obtained from them may provide the solution to a current industrial problem, i.e., the bioconversion of lignocellulosic material into fuels, chemicals, and other valuable bioproducts. This research performed a screening of lignocellulolytic enzymes of rumen microorganisms through metagenomic analyzes by sequence homology and studied the degradation pathways of lignocellulose in the rumen.

Results: A total of 410 predicted proteomes from the rumen microbiome were systematically screened for different families of Carbohydrate-Active Enzymes Carbohydrate-Binding Modules (CAZymes) and (CBMs) across different microorganisms compared with CAZy database. It was observed that Firmicutes and Bacteroidetes are the phylum responsible for the majority of CAZymes and CBMs produced in the rumen. Corporate action among different phyla during biomass degradation was observed, confirming the fine-tuning synergism among different species of rumen. Great amounts of CBM50 and chitinases call attention to the ability of rumen bacteria to use chitinous polymers from fungal cell walls as a source of carbohydrate for energy and growth purposes. The absence of auxiliary activity enzymes reported to present ligninolytic activities reaffirms the lack of ability of rumen microbiota to degrade lignin. Nonetheless, phylogenetic analysis suggested the only Laccase gene that was found was acquired by four Bacilli through horizontal gene transfer. Results also showed the presence of enzymes that participate in the loosening of cell wall by cleaving linkages between lignin and hemicellulose, thus making cellulose more accessible for hydrolases. Conclusions: This study demonstrates the important relation between microorganisms in the rumen that contributes to the enzymes required to degrade plant cell wall under such unique environmental condition. It also highlights the uniqueness of in silico techniques for big data analysis.

Keywords

Rumen microbiota, Carbohydrate-Active Enzymes, Carbohydrate-binding modules, Metagenomics

1 Background

The use of lignocellulosic plant material as a source of carbon for the production of fuels and chemicals is currently one of the focal points of biotechnology. The total oil consumption in the world in 2014 was 92.42 million barrels per day [1], and the projection for 2030 is an approximately 25% growth in demand for crude oil [2]. Regarding energy consumption, about 78% is derived from fossil fuels, while only 19% is attributed to renewable energy sources, of which less than 1% refers to well-established biofuels, such as ethanol and biodiesel [3].

Although there is no accurate data on the amount of oil still available around the world, and how long it will be possible to depend on it, society already understands that it is necessary to adapt and diversify energy sources for a cleaner and more sustainable future. Thus, it is needed to develop an approach that considers converting large parts of the global economy into a sustainable economy with bioenergy, biofuels and bioproducts as its main pillars.

Biological-based fuels and chemicals produced using residual feedstock, such as lignocellulosic biomass, provide society with a number of benefits, such as: 1) being renewable and sustainable; 2) assisting indirectly in the fixation of atmospheric carbon dioxide (which contributes negatively to global warming); 3) facilitating the stimulation and development of the local economy; 4) reducing air pollution caused by burning and degradation of biomass in fields; 5) bringing energy security to countries dependent on imported fuel, and 6) creating jobs in high-tech areas for engineers, biotechnologists and scientists [4].

Lignocellulosic biomass can be considered as the only source for large-scale production of renewable fuels, chemicals and biomaterials [5]. However, the plant cell wall has evolved to ensure its stability and robustness and not for the purpose of carbon storage, despite having one of the most abundant sources of renewable carbon in the biosphere [6]. Thus, although lignocellulose is a rich raw material in macromolecules essential for the development of a biorefinery, it has a complex structure that needs to be overcome in order to release its real potential.

The major obstacle for the industrial scale production of biofuels from lignocellulose lies in the inefficient deconstruction of plant material due to the recalcitrant nature of the substrate against enzymatic breakdown and the relatively low activity of currently available enzymes [7].

The large number of studies on lignocellulose degradation pathways not only allows understanding the mechanisms and interactions between organisms in the maintenance of the carbon balance in the biogeochemical cycles, but also leads to potential discoveries of uncharacterized microorganisms and new enzymes which, in turns, could improve the conversion of underutilized biomass to biofuels, chemicals and other materials through biorefineries [8,9].

Although the success in engineering proteins to increase the performance of existing lignocellulosic degrading enzymes has been limited [10], the recovery of enzymes from microbial communities naturally involved with biomass degradation offers a promising strategy for identification of new lignocellulolytic enzymes with potentially better activities [9].

Most enzymes have been identified using laborious screening methods and extensive biochemical characterizations, but evidences based on bioinformatics have raised the perception that *in silico* metagenomics approach can be applied to identify new enzymes from sequences of microbial genomes [11].

The metagenomics consists on the direct analysis of the DNA from different samples of a given environment and it represents a strategy to discover several enzymes encoded in nature [12,13]. The metagenomics approach has been used successfully to identify enzymes with activities of industrial interests [14].

Once metagenomic DNA is obtained, the prospection of new enzymes occurs through the screening of expression libraries, whether based on functions or on sequence similarity. Zhao *et al.* (2013) used a sequence homology screening method to comparatively analyze fungal genomes and their ability to degrade plant polysaccharide materials using CAZymes. Their study provided insights into the variety and expansion of fungal CAZyme classes and revealed the relationship of CAZyme size and diversity with their nutritional strategy and host specificity. Sequence homology analysis was also used [15] to perform genome mining of the complete and partial genome sequences of 229 ruminal microorganisms, and determine the distribution and diversity of bacteriocin gene clusters. The *in silico* analysis provided evidence of novel biosynthetic gene clusters in bacterial species not previously related to bacteriocin production.

In a study, Li *et al.* (2009) highlighted the relevance and advantages of the metagenomics screening method based on sequence homology. According to them, unlike function-based methods, it is possible to detect a greater number of target genes, since this method does not rely on gene expression, protein folding in the host, post-translational modifications, among others. Therefore, this method becomes quite attractive because it does not dependent on the integrity of the target gene sequence.

In this context, a screening of more than five hundred thousand lignocellulolytic enzymes and CBMs was performed on 410 publicly available bovine rumen metagenomes by *in silico* sequence homology analysis and the degradation pathways of lignocellulose in this environment were studied.

2 Methodology

2.1 Data collection

Genomes of microorganisms present in the rumen from the Hungate1000 Project [16] were retrieved as amino acid sequences from the DOE Joint Genome Institute (JGI) database and classified according to their taxonomy based on the NCBI-Taxonomy database. Revised Swiss-Prot protein sequences with ligninase activity were recovered from UniProt using the search term "laccase". Protein sequences of Glycoside Hydrolase (GH), Carbohydrate Esterase (CE), Polysaccharide Lyase (PL), Auxiliary Activity (AA) and Carbohydrate-Binding Module (CBM) families were retrieved from the Carbohydrate-Active Enzymes database (CAZyDB) (www.cazy.org) [17]. All data was collected in January 2017.

2.2 Functional gene annotation and metabolic pathway analysis

Sequence similarity searching to identify homologous enzymes with the potential for lignocellulose degradation in the proteome sequences recovered from Hungate1000 were performed against to the enzyme sequences from the CAZy and UniProt databases. Search was performed using the BLAST program, using an e-value cutoff of 1e-50 [18–20].

Proteins resulting from the similarity analysis were classified among the enzymes responsible for the degradation of lignocellulose components in their main pathways, according to information deposited on the CAZyDB. Therefore, the degradation potential of the microorganisms was determined, as well as the microbial diversity present in the lignocellulose degradation process and its corporate character.

Classification of the genomes was performed by comparing the similarity between the proteins of the JGI genomes and the Clusters of Orthologous Genes (COG) using the COGnitor program using an e-value cutoff of 1e-10 [21].

Phylogenetic analysis was carried out with Multiple Sequence Comparison by Log-Expectation (MUSCLE) Web Service under default settings [22,23]

3 Results

3.1 Taxonomic classification and microbial diversity of the rumen metagenome

Knowing the microbial diversity of an environment is one of the first steps to understand what are the players in the cell wall degradation, so that one can determine their roles in the process. From the 437 microorganisms publicly available on the JGI database, 410 had a completed annotation status (Additional file: Table 1).

The taxonomic distribution of microorganisms in the 410 studied bovine rumen metagenomes (Figure 1) is majorly composed of *Bacteria* (99%), and only a small percentage of *Archaea* is present (1%), being *Euryarchaeota* its only Phylum. *Firmicutes* is the most prevalent phylum among *Bacteria* (74.5%), followed by *Bacteroidetes* (12.13%), *Actinobacteria* (7.43%), *Proteobacteria* (4.46%), *Spirochaetes* (0.99%), *Fibrobacteres* (0.25%) and *Fusobacteria* (0.25%).

3.2 Distribution of CAZyme families

A total of 410 predicted proteomes from the rumen microbiome were systematically screened for different families of Carbohydrate-Active Enzymes (CAZymes) and CBMs across different microorganisms compared with CAZyDB. From a total of 280 CAZyme families (excluding un-classified families, e.g. GH0), 180 families from the following five functional classes were identified: 95 GHs, 16 PLs, 13 CEs, 4 AA and 52 CBMs. Although 13 microorganisms present only one CAZyme family, the majority of microorganisms (207) contain more than 70 CAZymes, being *Methanomicrobium mobile* DSM 1539 (NCBI Taxon ID 694440) the one with less CAZymes (3 GHs) and *Bacteroides ovatus* NLAE-zl-C57 (NCBI

Taxon ID 28116) the most abundant in CAZymes, with 438 in total. GH13 and GH19 were present among all proteomes. We also observed some CAZyme families appeared to be phylum-specific. For example, GH102, CBM69, PL8 and PL5 were only present in *Proteobacteria*, whereas 20 other families appeared to be *Firmicutes*-specific. Nonetheless, *Euryarchaeota*, *Fusobacteria* and *Spirochaetes* are the only phyla that no specificity was found (Table 1).

Family	Actinobacteria	Bacteroidetes	Euryarchaeota	Fibrobacteres	Firmicutes	Fusobacteria	Proteobacteria	Spirochaetes	Family	Actinobacteria	Bacteroidetes	Euryarchaeota	Fibrobacteres	Firmicutes	Fusobacteria	Proteobacteria	Spirochaetes
AA7	0	0	0	0	10	0	0	0	GH97	0	3	0	0	0	0	0	0
CBM8	0	0	0	0	10	0	0	0	GH102	0	22	0	0	0	0	0	0
CBM36	0	0	0	0	16	0	0	0	GH113	0	0	0	1	0	0	0	0
CBM46	0	0	0	0	42	0	0	0	GH114	0	0	0	2	0	0	0	0
CBM54	0	0	0	0	21	0	0	0	GH116	0	0	0	0	1	0	0	0
CBM58	0	5	0	0	0	0	0	0	GH117	0	0	0	0	2	0	0	0
CBM63	0	0	0	0	13	0	0	0	GH120	0	0	0	0	3	0	0	0
CBM69	0	0	0	0	0	0	7	0	GH123	0	0	0	0	49	0	0	0
CBM79	0	0	0	0	5	0	0	0	GH126	1	0	0	0	0	0	0	0
CE3	0	0	0	0	2	0	0	0	GH128	0	0	0	0	1	0	0	0
CE5	3	0	0	0	0	0	0	0	GH137	6	0	0	0	0	0	0	0
GH12	0	0	0	0	1	0	0	0	GH141	0	5	0	0	0	0	0	0
GH44	0	0	0	0	0	0	11	0	GH142	0	0	0	0	20	0	0	0
GH45	0	0	0	0	9	0	0	0	GH144	0	12	0	0	0	0	0	0
GH52	1	0	0	0	0	0	0	0	PL4	0	4	0	0	0	0	0	0
GH66	0	0	0	1	0	0	0	0	PL5	0	0	0	0	0	0	1	0
GH68	0	4	0	0	0	0	0	0	PL6	0	4	0	0	0	0	0	0
GH70	0	0	0	0	6	0	0	0	PL8	0	0	0	0	0	0	1	0
GH71	0	0	0	0	2	0	0	0	PL13	0	7	0	0	0	0	0	0
GH79	0	0	0	0	10	0	0	0	PL14	0	0	0	1	0	0	0	0
GH85	0	1	0	0	0	0	0	0	PL15	0	15	0	0	0	0	0	0
GH89	0	0	0	0	1	0	0	0	PL21	0	4	0	0	0	0	0	0
GH91	0	9	0	0	0	0	0	0									

Table 1. Phylum specific CAZyme and CBM families

3.2.1 Glycoside hydrolases

GHs hydrolyze the glycosidic bond between two or more carbohydrates, or between a carbohydrate and a non-carbohydrate moiety, such as protein, or a lipid. There are 144 GH families grouped according to their amino acid sequence, from which 95 were identified in the microorganisms from rumen, with the most predominant families being GH1 and GH13 (Figure 2). Moreover, our results show the families that are present in the majority of the microorganisms are the ones that have more enzymes identified as GH. It is important to highlight that, although *Bacteroidetes* represent only 12% of total rumen microbiota, they are accountable for 20% of produced GHs. Nonetheless, Firmicutes are the responsible for the vast majority (70%) of GHs in the rumen (Additional file: Table 4).

3.2.2 Polysaccharide Lyases

PLs mainly degrade glycosaminoglycans and pectin, and are classified into 26 families in the CAZyDB. According to our study, the microbiota from the rumen is able to produce 16 from the 26 PL families, with the most abundant family being PL12 (Figure 2). Among the eight phylum studied, *Euryarchaeota*, *Fusobacteria* and *Spirochaetes* lack any PL, whereas nine PL families (4, 5, 6, 8, 13, 14, 15 and 21) appeared to be phylum-specific, mostly within *Bacteroidetes* (Table 1).

3.2.3 Carbohydrate esterases

CEs catalyze the de-O or de-N-acylation of esters or amides and other substituted saccharides in which sugars play the role of alcohol and amine. There are 16 CE families deposited on CAZyDB, and our study revealed that rumen microbiota has 13 of them, being CE4 the one with more number of CEs (Figure 2). As in the previously described families, *Firmicutes* is the phylum which we detected more CEs, having about 60% more enzymes than the second most abundant, *Bacteroidetes* (Additional file: Table 4). Interestingly, only CE3 and

CE5 seems to be phylum-specific, being them present only in *Firmicutes* and *Actinobacteria*, respectively (Table 1). Among the studied Archaea, 33% of them lack any CEs, against less than 1% of all Bacteria (Additional file: Table 3). Our study showed that CE4 is the most prevalent CE family, being present in 390 proteomes, and that CE5 and CE3 are the least, which were detected in only three and two proteomes, respectively (Figure 2).

3.2.4 Carbohydrate-binding modules

CBMs are appended to carbohydrate active enzymes that degrade insoluble polysaccharides. Our study identified 64% of all CBM families classified according to the CAZyDB. The highest counts were assigned to CBM50 and CBM48, which were also the only ones present among all phyla (Figure 2). We detected the highest abundance of CBMs within *Firmicutes* (Additional file: Table 4), and, interestingly, all detected CBM families were distributed among all phyla, but only the archaea *Methanomicrobium mobile* DSM 1539 (NCBI Taxon ID 694440) lack any CBM (Additional file: Table 3).

3.2.5 Auxiliary activities

AAs class groups both ligninolytic enzymes and lytic polysaccharide mono-oxygenases (LPMOs). We detected only five, out of 13 AA families from CAZyDB, and more than 90% of all proteomes have no AA. Acting synergistically with cellulases [24], AA10 (previously known as CBM33) is the most abundant family, with 55% of detected AAs (Figure 2). AA10 is a LPMO found in bacteria and it is known to carry out oxidative cleavage of glycoside bonds in cellulose, thus boosting the activity of well-known hydrolytic depolymerizing enzymes [25]. Overall, almost 64% of AAs were found in *Firmicutes*, and no AA was identified among *Bacteroidetes*, *Fibrobacteres*, *Fusobacteria* and *Spirochaetes* (Additional file: Table 4).

4 Discussion

4.1 Biomass-degrading metabolic potential in rumen microbiome

The microbial community found in the rumen is capable of producing various types of enzymes required to convert cellulose, hemicellulose, and lignin into different types of monosaccharides and aromatic compounds that are essential energy sources for ruminants. Most lignocellulose-degrading enzymes found in our study belongs to *Firmicutes* and *Bacteroidetes* in several different genera. Enzyme mapping analysis (Figure 3) showed the class *Clostridiales* is the most dominant producer of the main lignocellulolytic enzymes, being the genus *Butyrivibrio* alone responsible for nearly 18% of them (Additional file: Table 5).

4.1.1 Cellulose

Cellulose degradation in the rumen microbiome is performed by the attack of different enzymes, mainly from GH5, 9, 44, and 45 families, which act as both exo- and endo-glucanases on the highly ordered crystalline fraction and also on the amorphous regions [26]. The yielded cello-oligosaccharides, or cellodextrins, and cellobioses are further processed by enzymes involved in the hydrolysis of beta-linked dimers of oligosaccharides, such as β -glucosidases, glucan- β -1,4-glucanase, and cellobiose phosphorylase from the GH1, 3, 30, 94 and 116 families. Our study found 15,154 GHs in the rumen microbiome, and 41% of them are related to 28 GH families that have lignocellulose-degrading enzymatic activities (Table 2).

Breakdown of cellulose in this environment seems to happen under the action of enzymes from four main GH families, being the majority of them assigned to the core family GH5 (80%), which contains numerous enzymes known to participate in the degradation of cellulose to cellodextrin (endo- β -1,4-glucanase, EC 3.2.1.4), from cellodextrin to cellobiose (β -glucosidase, EC 3.2.1.21; exo- β -1,4-glucanase, EC 3.2.1.74), and from cellulose to cellobiose (cellulose β -1,4-cellobiosidase, EC 3.2.1.91). We found *Firmicutes* as being the phylum with

more enzymes from GH5 family, and *Lachnospiraceae bacterium* NLAE-zl-G231 (NCBI Taxon ID 186803) is the bacteria with more enzymes from this GH family (Additional file: Table 3). GH9, which also participates in the cellulose degradation with β -glucosidases and exo- β -1,4-glucanases, appeared to be more present in *Bacteroidetes*. Interestingly, cellulose degradation by GH44 and GH45 seems to be specific to *Fibrobacteres*. Despite the oligosaccharide-degrading enzymes are mainly found in *Firmicutes*, the majority of GH30 enzymes were identified within Bacteroidetes, and GH116 appears to be specific to *Fibrobacteres*.

4.1.2 Hemicellulose

Hemicellulose contains a greater variety of carbohydrate polymers, such as xylan, mannan, and pectin, and thus requires a broader range of exo- and endo-acting enzymes to degrade them [1].

4.1.2.1 Xylan

Xylan is one of the main components of hemicellulose and it is formed by a backbone of β-1,4-D-xylopyranose units with carbohydrate chains on the C2- or C3-position, such as Dglucuronic acid or L-arabinose [27]. A number of enzymes are responsible for the breakdown of xylan, and those include glycoside hydrolases from several families (Table 2). The glycosidic bonds in the xylan backbone are cleaved by endo-1,4-β-xylanases (1,4-β-D-xylan xylanohydrolase; EC 3.2.1.8), and our results showed this enzyme distributed among the families GH5, 8, 10, and 51. There were 65 enzymes detected in family GH10, the most important for bacteria [28], and, despite *Actinobacteria, Spirochaetes* and *Bacteroidetes* presented some of them, *Firmicutes* is the leading phylum with 85% of the endo-1,4-βxylanases. The oligosaccharides yielded are than cleaved by a pool of enzymes, mainly from β-xylosidase (EC 3.2.1.37) belonging to the families GH3, 30, 39, 43, 51 and 52 [27]. From these, GH43 and GH3 are the most abundant families. Interestingly, whereas GH3 has 84% of its enzymes assigned to *Firmicutes*, GH43 has them well distributed among *Bacteroidetes* (41%) and *Firmicutes* (50%).

Many other accessory enzymes participate in the xylan degradation, such as α -Larabinofuranosidase (EC 3.2.1.55, GH51) [29], and α -glucuronidase (EC 3.2.1.139, GH67) [30]. Carbohydrate esterases also take part in this support by the action of enzymes from families CE1 and CE4, such as Acetyl xylan esterase (EC 3.1.1.6) and Ferulic acid esterase (EC 3.1.1.73) [1]. Special attention can be given to CE4, which has 27% of all CEs found in the rumen microbiota, with 80% of them belonging to *Firmicutes* (Figure 4, Additional file: Table 8). One type of covalent linkages connecting lignin and hemicellulose in plant cell walls is the ester linkage between D-glucuronic or 4-O-methyl-D-glucuronic acid of glucuronoxylans and lignin alcohols [31]. An enzyme that could hydrolyze such linkages is named glucuronoyl esterase and it is classified into family CE 15, which was detected in 31% of the studied proteomes, representing 7% of all identified CEs (Additional file: Table 1).

4.1.2.2 Mannan

Together with Xylan, Mannan is also a major hemicellulosic polysaccharide found in nature. The backbone of this heterogeneous β -1,4-linked polysaccharide can consist exclusively of mannose (mannans) or a random sequence of glucose and mannose residues (glucomannan) [32]. Both mannans and glucomannans are often acetylated and can also be decorated with α -1,6-galactose residues (galactomannan and galactoglucomannan, respectively) [33]. Heteromannans are first degraded by β -mannanases (EC 3.2.1.78) that are allocated in the GH families 5 and 26, yielding dimers and oligomers of mannose [34]. Interestingly, our results showed no β -mannanases (GH26) from *Firmicutes*, but only from *Actinobacteria* and *Bacteroidetes*, being the former with more enzymes detected (Figure 4, Additional file: Table 8). However, corporate action among microorganisms is observed in the further processing of mannan, since the majority (72%) of α -mannosidases (EC 3.2.1.24), enzymes known to degrade mannose oligosaccharides (GH38), were identified within *Firmicutes*, and evenly distributed amounts among *Proteobacteria* and *Bacteroidetes*.

4.1.2.3 Pectin

Pectin generally consists of a backbone of α -1,4-linked D-galacturonic acid residues, of which about 20% can also be replaced by other residues [35]. According to the nature of the monomer composition of this backbone and also of the side chains, pectins are distinguished into homogalacturonans, rhamnogalacturonans I and II. xylogalacturonan, and apiogalacturonan [36]. Pectic substances can be hydrolyzed by polygalacturonases (EC 3.2.1.15 and 3.2.1.67), or cleaved via β -elimination by pectin lyases (EC 4.2.2.10) and pectate lyases (EC 4.2.2.2) [37]. Polygalacturonases comprise endo- and exo-acting enzymes, both of which are accommodated in glycosyl hydrolase family GH28 and GH4, respectively. Interestingly, our study shows endo-acting polygalacturonidases to be more present among Bacteroidetes than in *Firmicutes*, however, the opposite is true for exo-acting polygalacturonases from family GH4.

Pectin lyases (PL1) participate in the degradation of natural pectin and highly (98%) methylesterified pectin into pectate via β -elimination, and also cleave pectate into digalacturonate, with the aid of pectate lyases (PL10) and exopolygalacturonases (EC 4.2.2.9, PL9) [35]. This nonhydrolytic cleavage of pectin in the rumen seems to be accomplished mainly by *Bacteroidetes*, whereas pectate elimination is performed by *Firmicutes*. Another enzyme in this process of pectin degradation is the pectin esterase (EC 3.1.1.11) from CE8 family, which catalyses the specific demethylesterification of homogalacturonans for further hydrolysis by polygalacturonases [38] and, according to our results, CE8 enzymes are more abundant in *Firmicutes*. These results clearly show the corporate action of enzymes in the pectin degradation performed by *Bacteroidetes* and *Firmicutes* in the rumen, either by hydrolysis or β -elimination.

4.1.2.4 Lignin

In contrast to the polysaccharides present in the plant cell wall, lignin is a polyphenylpropan polymer in which the linkages between its building blocks are the result of radical reactions and not due to removal of water [39]. Consequently, lignin degradation does not require hydrolytic enzymes. It is known that enzymes participating in the lignin degradation process are grouped into the AA class, specially laccases (EC 1.10.3.2, AA1), manganese peroxidases (EC 1.11.1.13, AA2), versatile peroxidases (EC 1.11.1.16, AA2), lignin peroxidases (EC 1.11.1.14, AA2), and vanillyl-alcohol oxidases (EC 1.1.3.38, AA4) [40]. Fungi are thought to play a primary role in the biodegradation of lignocellulose in natural environments, and also in the rumen because of their ability to penetrate plant tissues [41,42]. However, other studies have shown that the rumen bacteria also participate in the lignin degradation [43].

Laccase is a polyphenol oxidase that possesses moderately low redox potentials and does not require hydrogen peroxide, representing the largest subgroup of blue multicopper oxidases [44]. According to POLLEGIONI and colleagues (2015), only fungal laccases are used in biotechnological processes and most of the characterized enzymes have been isolated from fungi. To date, few bacterial laccases have been studied, although rapid progress in genome analysis suggests that these enzymes are widespread in bacteria [45–49].

Our results showed no enzymes from families AA1, 2 and 4 using the CAZyDB, so we run BLAST against 135 reviewed Laccases from the UniProt/SwissProt database (e-value < 1e-50). We than found one copy of a Laccase (UniProtKB Entry - D4GPK6 | LACC_HALVD) in each of the following four *Firmicutes*: *Bacillus licheniformis* VTM3R78 (NCBI Taxon ID 1402), *Enterococcus casseliflavus* NLAE-zl-C414 (NCBI Taxon ID 37734), *Enterococcus gallinarum* SKF1 (NCBI Taxon ID 1353), and *Lactobacillus plantarum* AG30 (NCBI Taxon ID 1590) (Additional file: Table 6). To investigate the presence of Laccase in these four bacteria, we performed phylogenetic analysis and it seems they were inherited by a *Bacilli* common ancestor (Figure 5). According to UniProt/SwissProt database, such Laccase was isolated from the Archaea *Haloferax volcanii* (NCBI Taxon ID 309800), which lead us to hypothesize this enzyme was acquired through horizontal gene transfer to an ancestor more related to *Bacillus licheniformis* VTM3R78, and then it was retained through evolution among the other three bacteria. It is noteworthy that these bacteria represent only 6.5% of total *Bacilli* (61) from rumen, which gives more support for horizontal gene transfer hypothesis, as it is less likely all other 57 *Bacilli* would lose the Laccase gene over time. However, further studies need to be accomplished to better understand this process.

Other enzymes from the AA families 3 and 5 play an important role in lignin degradation, although indirectly, by generating hydrogen peroxide required for enzymes from AA2 family. Such cases are the glyoxal oxidase (EC 1.2.3.15, AA5), cellobiose dehydrogenase (EC 1.1.99.18, AA3), aryl alcohol oxidase (EC 1.1.3.7, AA3), and pyranose oxidase (EC 1.1.3.10, AA3). We found copies of these enzymes in the rumen microbiota, distributed among *Actinobacteria*, *Euryarchaeota*, *Firmicutes*, and *Proteobacteria*, accounting for 17% of all detected AAs. We also identified copies of the enzyme 1,4-benzoquinone reductase (EC 1.6.5.6, AA6), which is involved in the biodegradation of aromatic compounds and in the protection of fungal cells from reactive quinone compounds.

It in long known that ruminants shows incomplete recovery of compounds traditionally referred to as lignin, leading to the belief that a limited digestion of lignin occur under anaerobic conditions, with fungi playing a predominant role in the process [50,51]. Researchers used radio labeled [¹⁴C] lignocellulosic biomass preparations to investigate the mineralization and degradation processes of the polysaccharides and lignin components of plant tissues [43]. Their results demonstrated that little lignin could be completely degraded and converted to CO₂ and

CH₄. In another study, two typical products of lignin, 3,4-dimethoxyphenol and 3,4dimethoxybenzoic acid, were identified in the rumen liquor. However, no substances with the structural features of lignin decomposition products were detected in the degraded samples [51]. These studies suggest that lignin digestion in the rumen is more likely to be loosen, rather than fully degraded, which supports our results for ligninolytic enzymes.

4.2 Microbial interactions in the rumen

Metabolic interactions among different types of microorganisms are essential for sustaining the microbial community in the rumen. There are many types of interactions, and attention can be given to amensalism, which is based on the production of a compound by one organism that is inhibitory to the growth of another. This effect has been found in cocultures of anaerobic fungi with rumen cellulolytic bacteria [52].

Although the cell wall composition varies among fungal species, chemogenomic comparative analysis have led to a better understanding of the genes and mechanisms involved in the construction of the fungal cell wall, which has a common central core composed of branched β -1,3 glucan-chitin [53].

We found an abundancy of chitinases produced by rumen bacteria belonging to GH families 18, 19 and 23 (Additional file: Table 1). Interestingly, chitinases from family GH19 were found among all proteomes, including Archaeas. Altogether, chitinases from these families represented almost 10% of all GHs detected in the rumen microbiota. The carbohydrate-binding modules from families 12, 14, 18, 50 and 55 are found among chitinases, where they contribute to the antifungal activity of chitinases through their binding ability to chitinous component of the fungal cell wall [54]. Our results showed CBM50 as the most abundant (11%) and prevalent (97%) module, being *Bacillus* the genus with more copy numbers. The diversity of enzymes in the chitinolytic complex of rumen bacteria shows the

importance of chitin as a growth substrate for these bacteria, which, in the rumen ecosystem, the permanent source of chitin is represented by the cell walls of anaerobic fungi.

4.3 Meta-analysis of functional gene contents

We further studied the genes possessed by the rumen microbiota through classification of them into different Clusters of Orthologous Groups (COGs) [55] to gain an overall picture of their functions according to the different phylum (Additional file: Table 9, Figure 6). We observed the majority of DNA sequences are assigned to genes involved in metabolic processes among all phyla (approx. 42%), mainly in carbohydrate transport and metabolism (COG class G), amino acid transport and metabolism (E), and Energy production and conversion (C). *Euryarchaeota* presents a very low percentage (3%) of its genes dedicated to carbohydrate transport and metabolism (G). However, a significant portion of its genes are involved in energy production and conversion (10.16%), and coenzyme transport and metabolism (10%), which does not occur in any other phylum.

Interestingly, genes involved in cellular processing and signalling (22.85%), and in information storage (23.19%) are similarly distributed in the rumen metagenome. As expected, cell wall, membrane, envelope biogenesis (M), and translation, ribosomal structure and biogenesis (J) genes are more abundant within their respective functional categories. Overall, translation, ribosomal structure and biogenesis (J) genes are still poorly characterized with their functions unknown (S).

4.4 Biorefinery perspectives for rumen microbial enzymes

In the current scenario, enzymes become important tools for various industries as their use in various industrial processes not only improve quality and quantity of the product, but also fasten the processes. In light of the identification of new sources for the enzyme production, researchers found rumen as an interesting source of enzymes due to the presence of a wide variety of microbial population of obligate anaerobic microorganisms, including fungi, protozoa, and bacteria [56]. These microbial populations have evolved their capacity for the efficient utilization of complex and recalcitrant plant polymers such as cellulose and hemicellulose. They make the ruminants able to degrade the various carbohydrates usually polysaccharides and structural polysaccharides constituting the major part of the ruminants' diets and the primary source of the energy in forage-based diets.

Because of the presence of diversified microbial population in rumen, activities of enzymes found in rumen are diverse, but most commonly fibrolytic enzymes, which include cellulases, xylanases, β -glucanases, and pectinases. Although not covered in this work, other enzyme activities are also identified in rumen, which include amylases, proteases, phytases, and specific plant toxin-degrading enzymes (tannases) [57]. All these enzymes have different valuable commercial applications. However, till date, no report of commercial application of rumen enzyme is available [56].

The use of various ligninolytic enzymes from rumen microbiota in the feed of nonruminants increases the food utilization by enhancing the degradation of various plant cell wall polymers into simple sugars [58]. Such enzymes utilization allows the availability of carbon to the non-rumen livestock as their intestinal enzymes are not able to make these sources available to the livestock [57]. The presence of non-starch polysaccharides (NSP) (i.e., arabinoxylan, β glucan, etc.) in the foodstuff interferes with the digestion by increasing the viscosity of the digest in the animals' small intestine. Therefore, utilization of a wide range of enzymes such as cellulases, glucanases, and pectinases decreases the anti-nutritional compounds from the foodstuff (i.e., high fiber content). The addition of other enzymes such a xylanase and β glucanase hydrolyze the NSP to reduce the digest's viscosity and also release the nutrients (proteins and starch) [59]. In most of the lignocellulose-based industries, plant materials need to be subjected to a pretreatment (mechanical and/or chemical) followed by enzymatic hydrolysis to obtain a microbiologically accessible carbon source. This phenomenon in biorefineries is quite similar to a phenomenon that happen in the ruminants [60]. Thus, the enzymes obtained from rumen, i.e., cellulases, hemicellulases, lyases, esterases, etc., have a tremendous potential for being used in various lignocellulose-based industries, such as biofuel and paper industries. However, further researches are needed for the development of catalytically efficient enzymes and the bioprocess for making the biofuel production more efficient and cost-effective so that these biofuels can complement or even replace the utilization of fossil fuels.

5 Conclusions

In the rumen, microbial population converts the complex food stuffs (lingo-cellulosic plant materials) into simpler compounds, which can be easily taken up and metabolized by ruminants. Therefore, direct use of microbial population from rumen or enzymes obtained from them may provide the solution to a current industrial problem, i.e., the bioconversion of lignocellulosic material into fuels, chemicals, and other valuable bioproducts. It was identified glycoside hydrolases, polysaccharide lyases, carbohydrate esterases, and auxiliary activities enzymes, as well as carbohydrate-binding modules from predicted proteomes of 410 representative microorganisms belonging to *Euryarchaeota*, *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Proteobacteria*, *Spirochaetes*, *Fibrobacteres*, and *Fusobacteria*. Comparative analysis revealed that they exhibit tremendous diversity in number and variety of CAZymes and CBMs. It was observed that *Firmicutes* and *Bacteroidetes* are the phylum responsible for the majority of CAZymes and CBMs produced in the rumen. Besides, corporate action among different phyla during biomass degradation was observed, confirming the fine-tuning synergism among different species of rumen. It is noteworthy that the great amount of CBM50, a module found attached to various enzymes from GH families related to chitin degradation,

highlighting the ability of rumen bacteria to use chitinous polymers from fungal cell walls as a source of carbohydrate for energy and growth purposes. The absence of auxiliary activity enzymes reported to present ligninolytic activities reaffirms the lack of ability of rumen microbiota to degrade lignin. Nonetheless, our study showed the presence of the same Laccase within four Bacilli, and phylogenetic analysis suggests they were acquired through horizontal gene transfer. Results also showed the presence of enzymes that participate in the loosening of cell wall by cleaving linkages between lignin and hemicellulose, thus making cellulose more accessible for hydrolases. COG analysis revealed that rumen microbiota has their genomes similarity dedicated across all functional categories. This demonstrates an important relation between microorganisms in the rumen that contribute to the enzymes required to degrade plant cell wall under such unique environmental condition.

6 List of abbreviations

AA - Auxiliary Activity

CAZyDB - Carbohydrate-Active Enzymes Database

CAZymes - Carbohydrate-Active Enzymes

CBM - Carbohydrate-Binding Module

CE - Carbohydrate Esterase

COG - Clusters of Orthologous Genes

GH - Glycoside Hydrolase

JGI - Joint Genome Institute

LPMO - Lytic Polysaccharide Mono-Oxygenases

MUSCLE - Multiple Sequence Comparison by Log-Expectation

NSP - Non-Starch Polysaccharides

PL - Polysaccharide Lyase

7 Declarations

7.1 Ethics approval and consent to participate

Not applicable

7.2 Consent for publication

Not Applicable

7.3 Availability of data and material

The datasets analysed during the current study are available in the Carbohydrate-Active Enzymes Database, [http://www.cazy.org], in the Joint Genome Institute (Genome Portal), [http://genome.jgi.doe.gov/TheHunmicrobiome], and in the Universal Protein Resource Repository, [http://www.uniprot.org]. All data generated during this study are included in this published article and its supplementary information file.

7.4 Competing interests

The authors declare that they have no competing interests.

7.5 Funding

Authors would like to acknowledge the Bahia State Agency for Research Funding (FAPESB – Brazil) for the financial support for L.M.A.

7.6 Authors' contributions

LMA and LMF conceived the idea. LMA and LMF wrote the scripts. LMA carried out the bioinformatics analyses. All authors are involved in interpreting the results, developing the discussion section and writing the manuscript.

7.7 Acknowledgements

Authors would like to thank Dra. Patricia Lopes Leal for her valuable thoughts on the initial design of this research.

8 References

1. Batista-García RA, del Rayo Sánchez-Carbente M, Talia P, Jackson SA, O'Leary ND, Dobson ADW, et al. From lignocellulosic metagenomes to lignocellulolytic genes: trends, challenges and future prospects. Biofuels, Bioprod. Biorefining [Internet]. 2016;10:864–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2571

2. Balan V. Current challenges in commercially producing biofuels from lignocellulosic biomass. ISRN Biotechnol. [Internet]. 2014;2014:31. Available from: http://www.hindawi.com/journals/isrn/2014/463074/

3. REN21. Renewable Energy Policy Network for the 21st century - 2016 Global Status Report[Internet].2017.Availablefrom:http://www.ren21.net/REN21Activities/GlobalStatusReport.aspx

4. Greenwell HC, Lloyd-Evans M, Wenner C. Biofuels, science and society. Interface Focus [Internet]. 2012;3:20120093–20120093. Available from: http://rsfs.royalsocietypublishing.org/cgi/doi/10.1098/rsfs.2012.0093

5. DeMartini JD, Pattathil S, Miller JS, Li H, Hahn MG, Wyman CE. Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. Energy Environ. Sci. [Internet]. 2013;6:898. Available from: http://xlink.rsc.org/?DOI=c3ee23801f

6. Sauer M, Mattanovich D. Construction of microbial cell factories for industrial bioprocesses. J. Chem. Technol. Biotechnol. [Internet]. 2012;87:445–50. Available from: http://doi.wiley.com/10.1002/jctb.3711

7. Blanch HW, Adams PD, Andrews-Cramer KM, Frommer WB, Simmons BA, Keasling JD.Addressing the need for alternative transportation fuels: the Joint BioEnergy Institute. ACSChem.Biol.[Internet].2008;3:17–20.Availablefrom:http://pubs.acs.org/doi/abs/10.1021/cb700267s

8. Lee J. Biological conversion of lignocellulosic biomass to ethanol. J. Biotechnol. [Internet]. 1997;56:1–24. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0168165697000734

9. Rubin EM. Genomics of cellulosic biofuels. Nature [Internet]. 2008;454:841–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18704079

10. Wen F, Nair NU, Zhao H. Protein engineering in designing tailored enzymes and microorganisms for biofuels production. Curr. Opin. Biotechnol. 2009. p. 412–9.

11. Wang H, Fewer DP, Sivonen K. Genome Mining Demonstrates the Widespread Occurrence of Gene Clusters Encoding Bacteriocins in Cyanobacteria. Blazquez MA, editor. PLoS One [Internet]. 2011;6:e22384. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21799841

12. Ferrer M, Golyshina O V., Chernikova TN, Khachane AN, Reyes-Duarte D, Santos VAPM Dos, et al. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. Environ. Microbiol. [Internet]. 2005;7:1996–2010. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16309396

13. Wang F, Li F, Chen G, Liu W. Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. Microbiol. Res. [Internet]. 2009;164:650–7. Available from:

http://linkinghub.elsevier.com/retrieve/pii/S0944501308000724

14. Li L-L, McCorkle SR, Monchy S, Taghavi S, van der Lelie D. Bioprospecting

metagenomes: glycosyl hydrolases for converting biomass. Biotechnol. Biofuels [Internet]. 2009;2:10. Available from:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2694162&tool=pmcentrez&rende rtype=abstract

15. Azevedo AC, Bento CBP, Ruiz JC, Queiroz M V., Mantovani HC. Distribution and Genetic Diversity of Bacteriocin Gene Clusters in Rumen Microbial Genomes. Nojiri H, editor. Appl. Environ. Microbiol. [Internet]. 2015;81:7290–304. Available from: http://aem.asm.org/lookup/doi/10.1128/AEM.01223-15

16. Creevey CJ, Kelly WJ, Henderson G, Leahy SC. Determining the culturability of the rumen bacterial microbiome. Microb. Biotechnol. [Internet]. 2014 [cited 2017 May 23];7:467–79. Available from: http://doi.wiley.com/10.1111/1751-7915.12141

17. Cantarel BI, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res. 2009;37.

18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool.J.Mol.Biol.[Internet].1990;215:403–10.Availablehttp://www.sciencedirect.com/science/article/pii/S0022283605803602

19. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics [Internet]. 2009;10:421. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20003500%5Cnhttp://download.springer.com/static/pdf /101/art%253A10.1186%252F1471-2105-10-

421.pdf?originUrl=http%253A%252F%252Fbmcbioinformatics.biomedcentral.com%252Fart icle%252F10.1186%252F1471-2105-10-421&token2=exp=14

20. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.

21. Kristensen DM, Kannan L, Coleman MK, Wolf YI, Sorokin A, Koonin E V., et al. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. Bioinformatics [Internet]. 2010;26:1481–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20439257

22. Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res. 2015;43:W580-4.

23. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113.

24. Forsberg Z, Vaaje-Kolstad G, Westereng B, Bunæs AC, Stenstrøm Y, MacKenzie A, et al. Cleavage of cellulose by a CBM33 protein. Prot Sci. [Internet]. 2011;20. Available from: http://dx.doi.org/10.1002/pro.689

25. Agger JW, Isaksen T, Várnai A, Vidal-Melgosa S, Willats WGT, Ludwig R, et al. Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. Proc. Natl. Acad. Sci. U. S. A. [Internet]. 2014;111:6287–92. Available from: http://www.pnas.org/content/111/17/6287.short

26. Kim IJ, Lee HJ, Choi IG, Kim KH. Synergistic proteins for the enhanced enzymatic hydrolysis of cellulose by cellulase. Appl. Microbiol. Biotechnol. 2014. p. 8469–80.

27. Pollet A, Delcour JA, Courtin CM. Structural determinants of the substrate specificities of

xylanases from different glycoside hydrolase families. Crit. Rev. Biotechnol. [Internet]. 2010;30:176–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20225927

28. Chakdar H, Kumar M, Pandiyan K, Singh A, Nanjappan K, Kashyap PL, et al. Bacterial xylanases: biology to biotechnology. 3 Biotech [Internet]. 2016;6:150. Available from: http://dx.doi.org/10.1007/s13205-016-0457-z

29. Shoham Y. Glycoside Hydrolase Family 51 [Internet]. 2017 [cited 2017 May 27]. Available from:

http://www.cazypedia.org/index.php?title=Glycoside_Hydrolase_Family_51&oldid=8782

30. Gilbert H. Glycoside Hydrolase Family 67 [Internet]. 2017 [cited 2017 May 27]. Available from:

http://www.cazypedia.org/index.php?title=Glycoside_Hydrolase_Family_67&oldid=7523

31. Das NN, Das SC, Mukherjee AK. On the ester linkage between lignin and 4-O-methyl-d-
glucurono-d-xylan in jute fiber (Corchorus capsularis). Carbohydr. Res. [Internet].1984;127:345–8.Availablefrom:from:

http://www.sciencedirect.com/science/article/pii/0008621584853690

32. Howard RL, Abotsi E, Jansen van REL, Howard S. Lignocellulose biotechnology: issues of bioconversion and enzyme production. African J. Biotechnol. [Internet]. 2003;2:602–19. Available from: http://www.academicjournals.org/AJB

33. Malgas S, van Dyk JS, Pletschke BI. A review of the enzymatic hydrolysis of mannans and synergistic interactions between β -mannanase, β -mannosidase and α -galactosidase. World J. Microbiol. Biotechnol. [Internet]. 2015;31:1167–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26026279

34. Centeno MSJ, Guerreiro CIPD, Dias FM V, Morland C, Tailford LE, Goyal A, et al.Galactomannan hydrolysis and mannose metabolism in Cellvibrio mixtus. FEMS Microbiol.Lett.[Internet].2006;261:123–32.Availablehttp://www.ncbi.nlm.nih.gov/pubmed/16842369

35. Bonnin E, Garnier C, Ralet M-C. Pectin-modifying enzymes and pectin-derived materials: applications and impacts. Appl. Microbiol. Biotechnol. [Internet]. 2014;98:519–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24270894

36. Vincken J-P, Schols HA, Oomen RJFJ, McCann MC, Ulvskov P, Voragen AGJ, et al. If homogalacturonan were a side chain of rhamnogalacturonan I. Implications for cell wall architecture. Plant Physiol. [Internet]. 2003;132:1781–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12913136

37. Lagaert S, Beliën T, Volckaert G. Plant cell walls: Protecting the barrier from degradation by microbial enzymes. Semin. Cell Dev. Biol. 2009;20:1064–73.

38. Pelloux J, Rustérucci C, Mellerowicz EJ. New insights into pectin methylesterase structure and function. Trends Plant Sci. 2007;12:267–77.

39. Dashtban M, Schraft H, Syed TA, Qin W. Fungal biodegradation and enzymatic modification of lignin. Int. J. Biochem. Mol. Biol. [Internet]. 2010;1:36–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21968746

40. Pollegioni L, Tonin F, Rosini E. Lignin-degrading enzymes. FEBS J. [Internet]. 2015;282:1190–213. Available from: http://doi.wiley.com/10.1111/febs.13224

41. Kajikawa H, Kudo H, Kondo T, Jodai K, Honda Y, Kuwahara M, et al. Degradation of benzyl ether bonds of lignin by ruminal microbes. FEMS Microbiol. Lett. 2000;187:15–20.

42. Bauchop T. Rumen anaerobic fungi of cattle and sheep. Appl. Environ. Microbiol. 1979;38:148–58.

43. Akin DE, Benner R. Degradation of Polysaccarides and Lignin by Ruminal Bacteria and Fungi. Appl. Environ. Microbiol. 1988;54:1117–25.

44. Giardina P, Faraco V, Pezzella C, Piscitelli A, Vanhulle S, Sannia G. Laccases: a neverending story. Cell. Mol. Life Sci. 2010;67:369–85.

45. Hullo MF, Moszer I, Danchin A, Martin-Verstraete I. CotA of Bacillus subtilis is a copperdependent laccase. J. Bacteriol. 2001;183:5426–30.

46. Martins LO, Soares CM, Pereira MM, Teixeira M, Costa T, Jones GH, et al. Molecular and biochemical characterization of a highly stable bacterial laccase that occurs as a structural component of the Bacillus subtilis endospore coat. J. Biol. Chem. 2002;277:18849–59.

47. Enguita FJ, Martins LO, Henriques AO, Carrondo MA. Crystal structure of a bacterial endospore coat component. A laccase with enhanced thermostability properties. J. Biol. Chem. 2003;278:19416–25.

48. Koschorreck K, Schmid RD, Urlacher VB. Improving the functional expression of a Bacillus licheniformis laccase by random and site-directed mutagenesis. BMC Biotechnol. 2009;9:12.

49. Beloqui A, Pita M, Polaina J, Martinez-Arias A, Golyshina O V., Zumarraga M, et al. Novel Polyphenol Oxidase Mined from a Metagenome Expression Library of Bovine Rumen: BIOCHEMICAL PROPERTIES, STRUCTURAL ANALYSIS, AND PHYLOGENETIC RELATIONSHIPS. J. Biol. Chem. 2006;281:22933–42.

50. Chen F, Zhu Y, Dong X, Liu L, Huang L, Dai X. Lignocellulose degrading bacteria and their genes encoding cellulase/hemicellulase in rumen: a review. Wei Sheng Wu Xue Bao [Internet]. 2010;50:981–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20931863

51. Susmel P, Stefanon B. Aspects of lignin degradation by rumen microorganisms. J. Biotechnol. [Internet]. 1993;30:141–8. Available from: http://www.sciencedirect.com/science/article/pii/016816569390035L

52. Bernalier A, Fonty G, Bonnemoy F, Gouet P. Inhibition of the cellulolytic activity of Neocallimastix frontalis by Ruminococcus flavefaciens. J. Gen. Microbiol. England; 1993;139:873–80.

53. Latge J-P. The cell wall: a carbohydrate armour for the fungal cell. Mol. Microbiol. England; 2007;66:279–90.

54. Ohnuma T, Taira T. Carbohydrate Binding Module Family 50 [Internet]. 2017 [cited 2017May7].Availablefrom:http://www.cazypedia.org/index.php/Carbohydrate_Binding_Module_Family_50

55. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin E V, et al. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. England; 2003;4:41.

56. Puniya AK, Singh R, Kamra DN, editors. Rumen Microbiology: From Evolution to Revolution [Internet]. New Delhi: Springer India; 2015. Available from: http://link.springer.com/10.1007/978-81-322-2401-3

57. Cheng KJ, Lee SS, Bae HD, Ha JK. Industrial Application of Rumen Microbes. Asian-Australian J. Anim. Sci. 1999. p. 84–92.

58. Badhan A, Wang Y, Gruninger R, Patton D, Powlowski J, Tsang A, et al. Formulation of enzyme blends to maximize the hydrolysis of alkaline peroxide pretreated alfalfa hay and barley straw by rumen enzymes and commercial cellulases. BMC Biotechnol. [Internet]. 2014;14:31. Available from: http://bmcbiotechnol.biomedcentral.com/articles/10.1186/1472-6750-14-31

59. Ribeiro GO, Gruninger RJ, Badhan A, McAllister TA. Mining the rumen for fibrolytic feed enzymes. Anim. Front. [Internet]. 2016;6:20. Available from: https://dl.sciencesocieties.org/publications/af/abstracts/6/2/20

60. Sauer M, Marx H, Mattanovich D. From rumen to industry. Microb. Cell Fact. [Internet]. 2012;11:121. Available from: http://www.microbialcellfactories.com/content/11/1/121



Figure 1. Taxonomic distribution of rumen microbiota.

0 0		0	200	400	600	800	1000	1200	1400		0	200	400	600	800	1000	1200	1400
	CH12	_								GH9								
	CH10									GH24	Ε.							
0H73 0H33 0H73 0H34 0H34 0H34 0H35 0H34 0H36 0H34 0H37 0H34 0H36 0H34 0H37 0H34 0H38 0H39 0H39 0H34 0H34 0H34 0H35 0H34 0H36 0H34 0H37 0H34 0H38 0H34 0H39 0H34 0H36 0H34 0H37 0H34 0H38 0H34 0H39 0H34 0H39 0H34 0H36 0H34 0H37 0H34 0H38 0H34 0H39	CH36									GH30								
0H13 0H29 0H29 0H3 0H39 0H39 0H39 0H39 0H39 0H39 0H39 0H39 0H30 0H39 0H39 0H31 0H39 0H39 0H30 0H39 0H39 0H31 0H39 0H39 0H31 0H39 0H39 0H31 0H39 0H39 0H31 0H39 0H39 0H32 0H39 0H39 0H33 0H39 0H39 0H34 0H39 0H39	CH22									GH63	•							
0H12 0H109	GH/7				-					GH29	•							
0H22 0H2 0H2 0H2 0H1 0H13 0H2 0H13 0H2 0H13 0H2 0H13 0H2 0H13 0H2 0H14 0H13 0H15 0H14 0H14 0H15 0H14 0H12 0H16 0H13 0H16 0H14 0H12 0H14 0H14 0H12 0H14 0H12 0H14 0H12 0H14 0H13 0H13 0H14 0H14 0H12 0H14 0H13 0H14 0H14 0H14 0H15 0H14 0H14 0H14 0H14 <t< td=""><td>GH43</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>GH109</td><td>-</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>	GH43									GH109	-							
0H139 0H139 0H139 0H130 0H139 0H139 0H131 0H144 0H144 0H134 0H144 0H144 0H136 0H144 0H144 0H137 0H126 0H126 0H126 0H126 0H126 0H126 0H126 0H126 0H127 0H126 0H126 0H128 0H126 0H126 0H128 0H126 0H126 0H128 0H126 0H126 0H128 0H141 0H126 0H129 0H141 0H141	GH3Z									GH92	÷							
GH1 GH133 GH3 GH134 GH3 GH134 GH3 GH144 GH3 GH144 GH3 GH144 GH3 GH144 GH3 GH135 GH3 GH134 GH12 GH123 GH2 GH124 GH13 GH133 GH3 GH134 GH13 GH134 GH13 GH134 GH13 GH134 GH13 GH134 GH13 GH134 GH134 GH134 GH135 GH141 GH136 GH136 GH137 GH136 GH138 GH136 GH134 GH136 GH135 GH136 GH136 GH136 GH137 GH136 GH138 GH141 GH135 GH137 GH33 GH136 GH34 GH136 GH33 GH137 GH34 GH136 GH135 GH137	GHZ5									GH138	÷							
GH12 GH14 GH14 GH13 GH14 GH14 GH14 GH14 GH14 GH13 GH14 GH14 GH13 GH14 GH14 GH13 GH14 GH14 GH13 GH14 GH13 GH2 GH14 GH13 GH2 GH13 GH13 GH2 GH13 GH13 GH2 GH13 GH13 GH2 GH13 GH13 GH2 GH14 GH14 GH13 GH14 GH14 GH22 GH14 GH14 </td <td>GH3</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>GH139</td> <td>÷ .</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	GH3									GH139	÷ .							
GH13 GH13 GH14 GH3 GH14 GH14 GH3 GH14 GH14 GH3 GH14 GH12 GH3 GH14 GH12 GH3 GH14 GH13 GH3 GH14 GH12 GH3 GH13 GH13 GH3 GH14 GH13 GH3 GH14 GH14 GH2 GH14 GH13 GH3 GH14 GH14 GH13 GH13 GH14 GH13 GH14 GH14 GH13 GH15 GH14 GH14 GH14 GH14 GH15 GH14 GH14 GH15 GH14 GH14 GH15 GH14 GH15 GH15 GH14 GH15 GH12 GH14 GH15 GH13 GH14 GH15 GH12 GH14 GH15 GH15 GH14 GH16 GH12 GH15 GH14 GH12 GH13 GH14	GHI								•	GHS	£							
GH73 GH144 GH144 GH73 GH144 GH144 GH73 GH144 GH144 GH73 GH144 GH144 GH3 GH144 GH144 GH3 GH144 GH144 GH3 GH129 GH120 GH120 GH120 GH120 GH121 GH120 GH124 GH123 GH120 GH120 GH124 GH120 GH120 GH125 GH120 GH120 GH126 GH120 GH120 GH128 GH120 GH120 GH138 GH120 GH120 GH138 GH120 GH120 GH138 GH120 GH120 GH138 GH120 GH120 GH139 GH144 GH120 GH139 GH144 GH120 GH139 GH144 GH120 GH139 GH120 GH120 GH120 GH120 GH120 GH120 GH120 GH120 GH120 GH120 GH120	GH23									GH15	£							
GH31 GH31 GH31 GH31 GH32 GH32 GH33 GH33 GH34 GH37 GH33 GH34 GH35 GH34 GH36 GH36 GH37 GH37 GH33 GH34 GH35 GH35 GH35 GH36 GH36 GH36 GH36 GH36 GH36 GH36 GH37 GH37 GH33 GH36 GH37	GH94									GH144	.							
GH35 GH35 GH35 GH35 GH34 GH37 GH37 GH44 GH129 GH121 GH35 GH129 GH121 GH36 GH120 GH121 GH37 GH129 GH121 GH37 GH129 GH121 GH36 GH120 GH121 GH37 GH121 GH129 GH44 GH129 GH120 GH38 GH120 GH120 GH39 GH141 GH130 GH130 GH131 GH131 GH31 GH120 GH120 GH32 GH131 GH131 GH33 GH131 GH131 GH34 GH133 GH131 GH35 GH141 GH120 GH36 GH131 GH131 GH137 GH133 GH133 GH135 GH137 GH133	GH/3				•					GH91	-							
GH33 GH4 GH4 GH102 GH30 GH102 GH127 GH123 GH42 GH113 GH127 GH123 GH42 GH123 GH43 GH124 GH44 GH125 GH42 GH127 GH42 GH128 GH43 GH129 GH124 GH120 GH28 GH126 GH33 GH126 GH33 GH136 GH138 GH136 GH139 GH141 GH130 GH120 GH131 GH126 GH28 GH126 GH139 GH141 GH120 GH120 GH121 GH120 GH122 GH121 GH123 GH121 GH139 GH131 GH149 GH121 GH120 GH121 GH121 GH121 GH122 GH121 GH139 GH137 GH149 GH137 GH130 <t< td=""><td>GH51</td><td></td><td></td><td>-</td><td></td><td></td><td></td><td></td><td></td><td>GH95</td><td>1 - C</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>	GH51			-						GH95	1 - C							
GH3 GH3 GH126 GH127 GH128 GH129 GH42 GH129 GH124 GH42 GH129 GH124 GH43 GH129 GH124 GH44 GH120 GH126 GH43 GH124 GH124 GH44 GH124 GH124 GH44 GH126 GH124 GH45 GH124 GH126 GH46 GH126 GH124 GH45 GH124 GH126 GH138 GH136 GH137 GH138 GH127 GH124 GH139 GH144 GH144 GH139 GH144 GH147 GH139 GH144 GH144 GH139 GH144 GH144 GH139 GH144 GH144 GH120 GH121 GH124 GH120 GH124 GH124 GH139 GH124 GH124 GH139 GH124 GH124 GH139 GH124 GH124 GH139 GH137 GH144 <td>GH35</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>GH26</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	GH35									GH26								
GH14 GH102 GH130 GH122 GH130 GH137 GH127 GH129 GH128 GH129 GH129 GH121 GH120 GH129 GH121 GH129 GH122 GH121 GH123 GH129 GH124 GH129 GH125 GH120 GH126 GH120 GH136 GH120 GH137 GH136 GH138 GH137 GH14 GH141 GH120 GH121 GH131 GH131 GH132 GH136 GH133 GH131 GH14 GH121 GH131 GH131 GH132 GH131 GH133 GH131 GH134 GH131 GH135 GH131 GH135	GH5									GH97								
GH2 GH120 GH127 GH123 GH2 GH123 GH2 GH123 GH2 GH123 GH2 GH123 GH2 GH124 GH125 GH141 GH120 GH141 GH120 GH141 GH133 GH141 GH134 GH120 GH35 GH136 GH138 GH136 GH138 GH136 GH138 GH136 GH138 GH137 GH28 GH141 GH129 GH142 GH136 GH142 GH136 GH142 GH127 GH85 GH14 GH142 GH120 GH141 GH121 GH15 GH122 GH123 GH123 GH144 GH124 GH128 GH137 GH137 GH138 GH137 GH140 GH128 GH133 GH137 GH134 GH137 GH135 <t< td=""><td>GH4</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>GH102</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>	GH4									GH102								
GH137 GH37 GH24 GH17 GH27 GH121 GH37 GH129 GH42 GH141 GH104 GH129 GH37 GH129 GH37 GH129 GH39 GH141 GH104 GH120 GH36 GH37 GH37 GH36 GH38 GH37 GH38 GH17 GH38 GH17 GH39 GH17 GH39 GH17 GH39 GH36 GH17 GH68 GH17 GH68 GH123 GH16 GH124 GH16 GH125 GH124 GH39 GH26 GH31 GH124 GH124 GH46 GH27 GH128 GH38 GH39 GH39 GH27 GH12 GH26 GH12 GH44 GH12 GH12 GH13 GH12 GH14 GH14	GH2									GH120 GH113	÷ .							
GH127 GH129 GH28 GH111 GH39 GH129 GH39 GH141 GH39 GH129 GH39 GH134 GH39 GH134 GH39 GH134 GH133 GH134 GH134 GH135 GH135 GH136 GH136 GH137 GH137 GH137 GH138 GH177 GH129 GH142 GH42 GH142 GH142 GH142 GH142 GH144 GH123 GH123 GH124 GH28 GH125 GH124 GH136 GH137 GH128 GH144 GH129 GH144 GH120 GH124 GH126 GH126 GH127 GH128 GH138 GH137 GH140 GH128 GH137 GH137 GH138 GH137 GH140 GH128 GH139 GH144 GH15	GH130			-						GH37	1.							
GH42 GH41 GH105 GH11 GH105 GH141 GH105 GH141 GH105 GH141 GH13 GH141 GH13 GH141 GH18 GH136 GH18 GH137 GH18 GH117 GH2 GH117 GH2 GH117 GH2 GH117 GH2 GH141 GH2 GH142 GH3 GH123 GH3 GH141 GH12 GH123 GH3 GH12 GH3 GH12 GH3 GH12 GH3 GH142 GH3 GH144 GH12 GH124 GH13 GH123 <	GH127									GH129								
GH105 GH120 GH34 GH30 GH35 GH30 GH38 GH30 GH38 GH30 GH38 GH30 GH38 GH33 GH38 GH30 GH38 GH31 GH38 GH33 GH38 GH31 GH38 GH31 GH38 GH117 GH38 GH117 GH38 GH117 GH39 GH120 GH39 GH121 GH39 GH46 GH27 GH31 GH38 GH46 GH27 GH31 GH31 GH46 GH32 GH46 GH33 GH31 GH33 GH31 GH34 GH45 GH38 GH31 GH38 GH32 GH38 GH33 GH38 GH33 GH38 GH33 GH38 GH33 GH38 GH33 GH38 GH33 GH	GH42		-							GH141	1							
GH67 GH87 GH84 GH84 GH87 GH84 GH50 GH50 GH84 GH104 GH104 GH133 GH14 GH104 GH134 GH136 GH136 GH135 GH16 GH17 GH28 GH17 GH17 GH49 GH17 GH16 GH29 GH17 GH17 GH29 GH17 GH17 GH29 GH17 GH12 GH30 GH16 GH17 GH31 GH16 GH17 GH20 GH16 GH17 GH20 GH12 GH16 GH21 GH12 GH16 GH22 GH12 GH14 GH13 GH14 GH12 GH33 GH14 GH12 GH34 GH12 GH12 GH35 GH13 GH13 GH36 GH37 GH36 GH37 GH38 GH14 GH155 GH17 GH37 GH36 GH37 GH36	GH105									GH120								
GH39 GH39 GH39 GH39 GH39 GH30 GH317 GH30 GH317 GH317 GH316 GH317 GH316 GH317 GH316 GH317 GH316 GH317 GH316 GH316 GH316 GH317 GH316 GH316 GH316 GH316 GH316 GH316 GH42 GH42 GH435 GH316 GH317 GH316 GH317 GH316 GH316 GH316 GH435 GH316 GH435 GH316 GH317 GH316 GH317 GH316 GH316 GH317 GH316 GH317 GH316 GH317 GH316 GH44 GH326 GH317 GH316 GH317 GH316 GH317 GH316 GH317 GH316 GH317 GH316 GH317 GH317 GH44 GH326 GH317 GH317 GH316 GH317 GH316 GH317 GH316 GH317 GH317 GH317 <td>GH67</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>GH84</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	GH67									GH84								
GH65 GH104 GH133 GH136 GH134 GH136 GH135 GH39 GH28 GH35 GH16 GH106 GH33 GH17 GH49 GH68 GH49 GH46 GH49 GH47 GH49 GH48 GH33 GH49 GH34 GH45 GH35 GH45 GH34 GH45 GH35 GH45 GH34 GH45 GH35 GH45 GH36 GH45 GH37 GH45 GH34 GH45 GH35 GH45 GH36 GH45 GH37 GH45 GH38 GH45 GH39 GH45 GH30 GH45 GH31 GH45 GH33 GH44 GH34 GH45 GH35 GH45 GH36 GH44 GH37 GH45 GH38 GH44 GH	GH39		•							GH50								
GH133 GH133 GH133 GH133 GH28 GH17 GH35 GH16 GH36 GH17 GH48 GH142 GH36 GH143 GH37 GH185 GH48 GH17 GH48 GH142 GH20 GH17 GH21 GH485 GH20 GH12 GH21 GH485 GH22 GH12 GH12 GH45 GH12 GH16 GH12 GH137 GH38 GH137 GH38 GH137 GH38 GH137 GH38 GH137 GH38 GH137 GH38 GH137 GH39 GH137 GH135 GH14	GH65		-							GH104 GH108								
GH13 GH38 GH38 GH38 GH37 GH15 GH49 GH57 GH49 GH57 GH49 GH49 GH40 GH40 GH40 GH40 GH40 GH44 GH45 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12	GH133									GH136								
GH88 GH188 GH16 GH16 GH17 GH85 GH17 GH49 GH49 GH49 GH49 GH49 GH40 GH20 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12	GH18		_							GH89								
GH28 GH85 GH16 GH16 GH33 GH17 GH49 GH7 GH49 GH7 GH49 GH46 GH20 GH12 GH21 GH45 GH22 GH12 GH12 GH45 GH12 GH46 GH12 GH44 GH12 GH16 GH12 GH14 GH12 GH14 GH12 GH14 GH12 GH14 GH12 GH14 GH12 GH14 GH13 GH14 GH14 GH14 GH13 GH14 GH14 GH14 GH13 GH14 GH14 GH13 GH13 GH13 GH14 GH52 GH70 GH71 GH15 GH71 GH15 GH71 GH15 GH71 GH15 GH71	GH88		•							GH117	1							
GH16 GH33 GH37 GH49 GH20 GH20 GH27 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12	GH28		•							GH85								
GH53 GH57 GH49 GH49 GH27 GH27 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12	GH16		-							GH106								
GH57 GH17 GH49 GH68 GH20 GH66 GH21 GH66 GH33 GH45 GH34 GH66 GH35 GH124 GH12 GH66 GH14 GH16 GH12 GH12 GH12 GH12 GH12 GH12 GH14 GH12 GH17 GH16 GH12 GH12 GH14 GH12 GH14 GH12 GH14 GH12 GH14 GH12 GH14 GH12 GH15 GH137 GH15 GH71 GH70 GH71 GH15 GH71 Distribution Abundance	GH53									GH142								
GH49 GH49 GH40 GH20 GH27 GH27 GH27 GH33 GH33 GH33 GH34 GH34 GH34 GH34 GH34 GH34 GH34 GH34 GH34 GH14 GH14 GH14 GH128 GH123 GH123 GH123 GH124 GH14 GH14 GH124 GH124 GH123 GH123 GH123 GH123 GH124 GH14 GH125 GH124 GH24 G	GH57	-								GH17								
GH20 GH27 GH27 GH12 GH12 GH12 GH33 GH33 GH34 GH14 GH14 GH14 GH16 GH12 GH16 GH12 GH12 GH12 GH16 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH14 GH14 GH14 GH14 GH14 GH14 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH14 GH14 GH14 GH14 GH14 GH14 GH14 GH14 GH14 GH12 GH17 GH38 GH37 GH38 GH37 GH22	GH49	-								GH/6								
GH27 GH112 GH112 GH12 GH12 GH13 GH31 GH14 GH14 GH16 GH10 GH10 GH10 GH12 GH12 GH12 GH12 GH12 GH12 GH12 GH12	GH20									GH55								
GH112 GH33 GH34 GH34 GH14 GH125 GH12 GH126 GH128 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH120 GH22 GH220 GH20 GH	GH27									GH123								
GH33 GH31 GH125 GH10 GH126 GH10 GH10 GH128 GH128 GH128 GH128 GH128 GH128 GH128 GH128 GH128 GH128 GH128 GH128 GH128 GH128 GH128 GH128 GH128 GH128 GH128 GH129 GH128 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH129 GH29 GH29 GH29 GH29 GH29 GH29 GH29 GH	GH112									GH66	1							
GH31 GH125 GH10 GH126 GH78 GH38 GH38 GH128 GH128 GH128 GH128 GH128 GH128 GH127 GH144 GH140 GH15 GH70 GH71 GH71 GH71 GH79 Distribution Abundance	GH33									GH45	1							
GH125 GH10 GH78 GH78 GH78 GH38 GH44 GH140 GH140 GH70 GH70 GH70 GH71 GH71 GH71 GH71 GH71 GH71 GH79 Distribution Abundance	GH31	_								GH114								
GH10 GH78 GH78 GH38 GH38 GH127 GH44 GH44 GH44 GH52 GH70 GH71 GH71 GH71 GH71 GH71 GH79 Distribution Abundance	GH125	_								GH116								
GH78 GH38 GH38 GH140 GH140 GH70 GH70 GH71 GH71 GH71 GH71 GH79 Distribution Abundance	GH10	-								GH12								
GH38 GH140 GH70 GH70 GH115 CH115 CH115 CH71 GH71 GH79 CH79 CH79 CH79 CH79 CH79 CH79 CH79 C	GH78	-								GH128 GH127								
GH140 GH52 GH52 GH71 GH71 GH79 Distribution Abundance Distribution Abundance	GH38	-								GH44								
GH70 GH71 GH71 GH79 Distribution Abundance Distribution Abundance	GH140									GH52								
GH115 GH79 GH79 Distribution Abundance Distribution Abundance	GH70	-								GH71								
Distribution Abundance Distribution Abundance	GH115									GH79								
Distribution Abundance Distribution Abundance																		
				Distri	oution	Abund	ance						Distrik	oution	Abund	ance		

Figure 2. Distribution and abundance of CAZyme family in the rumen microbiome.









Figure 2. Continue.



C)% 2	0%	40%	60%	80%	100%	C)%	20%	40%	60%	80%	100%
GH001							GH067						
GH002							GH068						
GH003							GH070						
GH004							GH071						
GH005							GH073						
GH008							GH076						
GH009							GH077						
GH010							GH078						
CH012							GH079						
GH012							GH084						
GH015							GH085						
CHOIC							GH088						
GHU10							GH089						
GHU17							GH091						
GHUI8							GH092						
GH019							GH094						
GHUZU							GH095						
GH023							GH097						
GH024							GH102						
GH025							GH103						
GH026							GH104						
GH027							GH105						
GH028							GH106						
GH029							GH108						
GH030							GH109						
GH031							GH112						
GH032							GH113						
GH033							GH114						
GH035							GH115						
GH036							GH116						
GH037							GH117						
GH038							GH120						
GH039							GH123						
GH042							GH125						
GH043							GH126						
GH044							GH127						
GH045							GH128						
GH046							GH129						
GH049							GH130						
GH050							GH133						
GH051							GH136						
GH052							GH137						
GH053							GH138						
GH055							GH139						
GH057							GH140						
GH063							GH141						
GH065							GH147						
GHUEE							GH142						
311000			1	1	1		011144						

Figure 4. Comparative relative abundance of enzymes from the rumen microbiome involved in lignocellulose degradation.

Actinobacteria Bacteroidetes Euryarchaeota Fibrobacteres Firmicutes Fusobacteria

Proteobacteria Spirochaetes

0%	20%	40%	60%	80%	100%		0%	20%	40%	60%	80%	100%
AA03						CBM01						
AA05						CBM02						
AA06						CBM03						
AA07						CBM04						
AA10						CBM05						
						CBM06						
CE01						CBM08						
CE02						CBM09						
CE03						CBM10						
CE04						CBM11						
CE05						CBM12						
CE06						CBM13						
						CBM16						
						CBM20						
						CBM22						
CE11						CBM23						
CE12						CBM25						
CE14						CBM26						
					-	CBM27						
PI 01						CBM30						
PL04						CBM32						
PL05						CBM34						
PL06						CBM35						
PL08						CBM36						
PL09						CBM37						
PL10						CBM38						
PL11						CBM40						
PL12 💻						CBM41						
PL13						CBIVI42						
PL14					_	CBIVI46						
PL15 🛑						CBIVI47						
PL17						CDIVI40						
PL21						CBIVI50						
PL22						CDIVIDI						
PL26						CDIVISS						
						CDIVID4						
						CDIVISO						
						CBM59						
						CBM50						
						CBM61						
						CBM62						
						CDIVIOZ						
						CBM65						
						CBM66						
						CBM67						
						CBM69						
						CBM70						
						CBM71						
						CBM72						
						CBM77						
						CBM79						
						00111/0						

Figure 4. Continue.

Actinobacteria Bacteroidetes Euryarchaeota Fibrobacteres Firmicutes Fusobacteria

Proteobacteria Spirochaetes

Figure 5. Hierarchical clustering of Laccase enzyme found in the rumen microbiota. Protein IDs are shown in parenthesis, and branch lengths in brackets.



Figure 6. Comparison of Clusters of Orthologous Groups (COGs) in the rumen microbiome.

