UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA CELULAR
PÓS-GRADUAÇÃO EM BIOLOGIA MOLECULAR

GENÔMICA FUNCIONAL DE *Clostridium thermocellum* EM DIFERENTES CONDIÇÕES DE CULTIVO

Brenda Rabello de Camargo

Brasília

2017

UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA CELULAR
PÓS-GRADUAÇÃO EM BIOLOGIA MOLECULAR

GENÔMICA FUNCIONAL DE *Clostridium thermocellum* EM DIFERENTES CONDIÇÕES DE CULTIVO

Brenda Rabello de Camargo

Orientador: Dra. Eliane Ferreira Noronha

Tese de Doutorado apresentada ao programa de Pós-graduação em Biologia Molecular da Universidade de Brasília, como parte dos requisitos para obtenção do título de Doutor em Biologia Molecular.

Brasília

2017

BRENDA RABELLO DE CAMARGO

GENÔMICA FUNCIONAL DE *Clostridium thermocellum* EM DIFERENTES CONDIÇÕES DE CULTIVO

Tese de Doutorado apresentada ao programa de Pós-graduação em Biologia Molecular da Universidade de Brasília, como parte dos requisitos para obtenção do título de Doutor em Biologia Molecular.

Banca examinadora:

_____

Profa. Dra. Eliane Ferreira Noronha (UnB)

_____

Profa. Dra. Pérola Magalhães (UnB)

_____

Profa. Dra. Nádia Skorupa Parachin (UnB)

_____

Prof. Dr. Roberto Nascimento Silva (USP-Ribeirão Preto)

_____

Prof. Dr. Marcelo Ramada (UCB-Brasília)

Agradecimento

Agradeço primeiramente à Deus e Nossa Senhora por acalmar meu espírito, e me dar a certeza de que não estou sozinha durante a caminhada. Agradeço aos meus pais, Mara Rabello e Lamartine Rabello que são meu alicerce e fonte de energia. À minha amada irmã por estar sempre ao meu lado, meu cunhado Alessandro Ribeiro, e à nossa pequena Valentia que dentro do ventre, nos dá esperança. Ao meu amor Leonardo Assis da Silva, que sem ele, não acreditaria na verdade e não teria tido forças nas horas difíceis. Obrigada! Á minha Orientadora Eliane Noronha, pela oportunidade de aprendizado, companheirismo e amizade durante longos anos. Ao Servé Kengen, pela oportunidade de trabalhar em um grupo que estuda no limiar da ciência. Aos meus colegas de trabalho e amigos, em Wageningen e Brasília, pela amizade a paciência.

Muito obrigada!

Livres-pensadores são aqueles que estão dispostos a usar suas mentes sem prejuízo e sem receio de entender as coisas que se chocam com seus próprios costumes, privilégios ou crenças. Este estado de espírito não é comum, mas é essencial para pensar direito.

Leon Tolstói

Resumo

O Brasil possui biomassa de origem lignocelulósica, de abundância, como os resíduos da indústria da cana de açúcar que, podem ser convertidos em produtos de alto valor agregado. *Clostridium thermocellum* é um microrganismo com grande potencial para degradação de celulose e biomassa lignocelulósica, devido à capacidade dessa bactéria em secretar um complexo multi-enzimático (celulossoma), formado por um arsenal de enzimas entre glicosil hidrolases, carboidrato esterases, e polissacarídeo liases, considerado carboidrato-ativas. Além disso, essa bactéria fermenta hexoses (C6) e não pentoses (C5), produzindo principalmente etanol, acetato e lactato. A regulação da expressão de genes, regulados pela fonte de carbono, é uma temática de contínuo estudo. Assim, foi realizado o proteoma descritivo de *Clostridium thermocellum* quando crescido por 96 horas em diferentes fontes de carbono: celulose, bagaço e palha de cana. Foram analisadas diferentes frações de conteúdo de proteína; sobrenadante (FS), ligado ao substrato (FES), e parcialmente purificado em coluna de exclusão molecular (CPP). Em todas as frações foram encontradas proteínas relacionadas à degradação da biomassa, em maior número em amostras de CPP de celulose, seguido pela amostra FES de palha. As exoglucanases CelS, CelK, CbhA, e proteínas estruturais CipA e OlpB foram encontradas em todos os substratos. Na amostra de bagaço foram identificadas as endoglucanases CelN, CelA, CelB, CelR, CelJ, CelQ, CelW, CelG e hemicelulases, Xgh74A, XynC, XynZ, Cthe_0798, sendo a última, presente também em palha. O transcritoma (RNAseq) e análise diferencial dos genes nas mesmas condições, porém em 37 horas, foi realizado. Das proteínas encontradas no proteoma, os genes *cthe_0798* e *xgh74A* foram encontrados diferencialmente expressos em bagaço quando comparado com celulose, seguido pelos genes codificadores de CelP, CtMan5A, CelC, XynA, ManA, Cthe_1257, Cthe_3163 e Orf2p. Genes envolvidos na mobilidade celular, incluindo proteínas de formação de flagelos, motor e regulação, foram encontrados como a categoria (COG N) de maior regulação positiva em bagaço e palha. Na mesma categoria estão genes envolvidos na quimiotaxia e *quorum sensing*. Durante as análises do transcritoma foi verificada a presença de outra bactéria, *Moorella thermoacetica*, conhecida por não ser celulolítica, porém com a habilidade de fermentar açúcares C5 e C6. A análise de expressão diferencial de genes nos diferentes tratamentos, mostrou *moth_0612* e *moth_0699* regulados positivamente em bagaço, ambos codificando transportadores de ribose (C5). A análise de proteínas ortólogas entre as duas bactérias revelou que *M.thermoacética* não possui ortólogos com potenciais de atuar como enzimas carboidrato ativas de *C.thermocellum*. Relativo à *C.thermocellum*, o gene *cthe_2196*, ausente de caracterização, contendo os domínios GH43, CBM6 e Doquerina tipo I, foi encontrado expresso em bagaço e palha, ausente em celulose, além de ser predito como fazendo parte de um operon juntamente com *cthe_2195*. *Cthe_2196*, foi clonado em sistema pET, expresso em *E.coli*, e a proteína denominada AxB8 foi caracterizada bioquimicamente. AxB8 possui habilidade de degradar arabinoxilana e substratos sintéticos *p*NP-α-L-arabinofuranosidase, e *p*NP-α-L-arabinofuranosidase. AxB8 apresenta domínio Glicosil hidrolase família 43, subfamília 29, e sua sequência mostrou maior similaridade com outras proteínas de termófilos não caracterizadas. Concluindo, os resultados desse trabalho demonstraram proteínas que são essenciais à degradação de bagaço e palha de cana, além de revelar importantes mecanismos regulados nessas condições que facilitam a proximidade da bactéria com o substrato, como os genes envolvidos na motilidade da bactéria. Essas informações podem ser utilizadas para desenvolvimento de novas linhagens e enzimas que aumentem o processo de degradação de biomassas lignocelulósicas, com o concomitante uso de produtos formados na geração de novas tecnologias.

Palavras-chave: *Clostridium thermocellum*, celulossoma, biomassa, expressão diferencial, glicosil hidrolase família 43.

Abstract

Lignocellulosic biomass is in abundance in Brazil, as the residues of the sugar cane industry, which can be converted into products with high aggregated value. *Clostridium thermocellum* is a microorganism with great potential for degradation of cellulose and lignocellulosic biomass due to its ability to secrete a multi-enzymatic complex (cellulosome), formed by an arsenal of enzymes between glycosyl hydrolases, carbohydrate esterases, and polysaccharide lyases, considered Carbohydrate-active. In addition, this bacterium ferments hexoses (C6) and not pentoses (C5), mainly producing ethanol, acetate, and lactate. Metabolism gene expression, regulated by the carbon source, is a subject of continuous study. Thus, the descriptive proteome of *Clostridium thermocellum* was carried out when grown for 96 hours in different carbon sources: cellulose, bagasse and cane straw. Different fractions of protein content were analyzed; Supernatant (FS), bound to the substrate (FES) and partially purified by size exclusion chromatography (CPP). In all the fractions were found proteins related to the degradation of the biomass, in greater number in samples of cellulose CPP, followed by straw FES. The exoglucanases CelS, CelK, CbhA, and structural proteins CipA and OlpB were found in all substrates. In the bagasse sample the endoglucanases CelN, CelA, CelB, CelR, CelJ, CelQ, CelW, CelW, CelG and hemicellulases, Xgh74A, XynC, XynZ, Cthe_0798 were identified, the latter being also present in straw. Transcriptomic (RNAseq) and differential analysis of the genes under the same conditions, but growth at 37 hours, was performed. From the proteins found in the proteome, the encoding genes *cthe_0798* and *xgh74A* were differentially expressed as bagasse when compared to cellulose, followed by the genes encoding CelP, CtMan5A, CelC, XynA, ManA, Cthe_1257, Cthe_3163 and Orf2p. Genes involved in cell motility(COG N), including flagella, motor and regulation proteins, were found in the highest category of up-regulation in bagasse and straw. In the same category were genes involved in chemotaxis and quorum sensing. During transcriptomic analysis, was detected the presence of another bacterium, *Moorella thermoacetica*, knowing to be not cellulolytic, but having the ability to ferment C5 and C6 sugars. Differential gene expression in different treatments showed *moth_0612* and *moth_0699* up-regulated in bagasse, both encoding ribose (C5) transporters. Orthologous proteins analysis between the two bacteria revealed that *M.thermoacética* did not have orthologs with the potential to act as active carbohydrate enzymes from *C.thermocellum*. Regarding *C.thermocellum*, the *cthe_2196* gene, absent from characterization, containing the domains GH43, CBM6 and Dockerin type I, was found expressed in bagasse and straw, but absent in cellulose, besides being predicted as part of an operon together with *cthe_2195*. *Cthe_2196* was cloned into pET system, further expressed in *E. coli*, and the protein named AxB8 was characterized biochemically. AxB8 has the ability to degrade arabinoxylan and synthetic substrates pNP-α-L-arabinofuranosidase, and pNP-α-L-arabinofuranosidase. AxB8 contains Glycosyl hydrolase family domain 43, subfamily 29, and its sequence showed greater similarity with other non-characterized thermophilic proteins. In conclusion, the results of this work demonstrated proteins that are essential for the degradation of sugarcane bagasse and straw, besides revealing important regulated mechanisms due to these conditions that facilitate the proximity of the bacterium with the substrate, as the genes involved in the motility of the bacteria. This information can be used to develop new strains and enzymes that increase the degradation process of lignocellulosic biomasses, with the concomitant use of products formed in the generation of new technologies.

Keywords: *Clostridium thermocellum*, cellulosome, biomass, differential expression, glycoside hydrolase family 43

Sumário

Capítulo 1. Introdução

1. Biomassa vegetal e sua utilização na indústria: Um breve relato

As tecnologias renováveis têm sido alvos de grandes pesquisas e investimentos por parte da academia e empresas ao redor do mundo. Juntamente com a busca por produtos mais “limpos”, uma nova geração de indústrias baseada em utilização de matérias primas renováveis, como a biomassa vegetal, tem sido mundialmente empregada [1,2]. A produção de biorenováveis, como por exemplo o etanol de segunda geração, uma temática amplamente abordada, vem com o intuito de amenizar problemas ambientais como a poluição gerada com a queima de combustíveis fósseis, e fatores político-econômicos associados à segurança energética [3].

As biorefinarias nesse contexto são unidades que integram o processo sustentável de uso da biomassa vegetal, associada às indústrias existentes, nas áreas alimentícias, químicas, e de energia, para obtenção de um produto de valor agregado, como suplementos em rações animais, solventes orgânicos, biocombustíveis, biofertilizantes, produtos químicos diversos e outros [4,5].

A produção mundial de biomassa vegetal total (matéria orgânica seca) é estimada em um número de 100 bilhões de toneladas por ano, fazendo parte desse montante os resíduos industriais, agrícolas, e florestais [6]. O Brasil é um dos países que se encontra em posição privilegiada como um dos maiores produtores de resíduos oriundos de biomassa lignocelulósica [7]. Essa produção se concentra principalmente na agroindústria, que é responsável por gerar resíduos como: bagaço e palha de cana de açúcar, farelo de cevada, sabugo e folhas de milho, palhas de trigo, arroz, e sorgo, casca de aveia, e outros [8]. Os resíduos bagaço e palha de cana, de maior abundância dentro do contexto brasileiro, correspondem juntamente, a 60% em peso, da produção total anual de cana de açúcar. Último boletim da CONAB (Companhia Nacional de abastecimento) relatou a estimativa de produção de 684,77 milhões de toneladas de cana de açúcar para a safra de 2016/17, o que corresponde a aproximadamente 400 milhões de toneladas de resíduos disponíveis [9].

Atualmente, os resíduos da indústria da cana de açúcar, como o bagaço, são principalmente utilizados para geração de energia elétrica na própria indústria (bioeletricidade), sendo que dados da ÚNICA (união da indústria de cana-de-açúcar), apontam uma quantidade

de 1133 MW médios de bioeletricidade produzidos a partir de bagaço de cana-de-açúcar em 2011, totalizando 2% a 3% do total matriz elétrica brasileira [10]. Contudo, a maior parte desse resíduo ainda não é utilizado. Porém, não somente energia elétrica, mas a produção de gêneros como bioplástico e bioetanol, são exemplos de enriquecimento da cadeia energética atualmente em pesquisa no Brasil, que envolvem a indústria da cana de açúcar [11,12].

Todo subproduto gerado na cadeia de produção, como aqui citado, contribui para a sustentabilidade do processo, além de fomentar a receita gerada pela indústria, que incrementa o ciclo de captação de recursos de impostos reaplicados por meio de benefícios à sociedade local [13]. Dessa forma, além do positivo impacto ambiental gerado pela utilização de renováveis, ocorre também um impacto socioeconômico atribuído á incorporação de novas tecnologias nas indústrias brasileiras.

2. Biomassa vegetal: constituição e degradação

A biomassa vegetal, podendo ser também chamada biomassa lignocelulósica, é constituída principalmente de celulose e hemicelulose, e ocasionalmente pectina, estruturas presentes na parede primária da planta, e lignina, constituinte da parede secundária, que confere em geral proteção à planta [14]. De acordo com a espécie e classificação da planta como gimnospermas ou angiospermas e monocotiledôneas ou dicotiledôneas, a composição muda drasticamente. Devido a essa natureza, o teor de hemicelulose e lignina pode ser mais elevado, por exemplo, em gramíneas quando comparados às plantas lenhosas [14,15]. Figura da parede vegetal A composição dos resíduos da cana-de-açúcar, bagaço e palha de cana, apresentam uma porcentagem de aproximadamente 40%, 24% e 25% de celulose, hemicelulose e lignina, respectivamente [16].

A celulose é formada basicamente por unidade de glicose unidas por ligações β-1,4 compondo uma cadeia linear [17]. O arranjo dessas cadeias em microfibrilas, unidas por ligações de hidrogênio, caracteriza diferentes graus de polimerização, que define sua estrutura cristalina [18,19].

A hemicelulose, diferentemente da celulose, possui uma característica de composição mais diversa, composta pelos açúcares xilose e arabinose (pentoses), e manose, galactose, e glucose (hexoses) em ligações β-1,4 [17]. Dessa forma, a hemicelulose é constituída por uma

cadeia principal, como por exemplo a xilana (xilose), arabinana (arabinose), manana (manose), contendo resíduos laterais, como por exemplo resíduos de arabinose presente em ligações α-1,3 e α-1,2, no polímero de xilana (arabinoxilana). Grupos acetil, metil, ácido ferúlico, ácido-metil-glucurônico, e ácido *p*-cumárico são também extensamente encontrados em cadeias laterais nesses polímeros [16]. Além disso, outros açúcares como fucose e ramnose podem ser encontrados, em menor número nessas cadeias [16].

Enquanto isso, a lignina, um polímero não-carboidrato, consiste de uma cadeia basicamente fenólica, formada por principalmente unidades de fenilpropanóides (monolignols), que diferem de acordo com substituição de grupos metóxi, que variam entre diferentes espécies de plantas [19,20]. Essas unidades principais são os álcoois p-hydroxyphenyl, coniferil, e sinapil [19,20]. O ácido ferúlico encontrado em algumas cadeias de hemicelulose, é, em algumas espécies de plantas, responsável pela ligação da hemicelulose com a lignina [21].

A pectina por sua vez é formada principalmente por ácido galacturônico (ligações α-1,4), formando também cadeias heterogêneas de homogalacturona, e ramnogalacturona [22,23].

Assim, para liberação desses monômeros de açúcares, a primeira etapa no processo de degradação da biomassa é o pré-tratamento, que consiste em uma preparação da parede vegetal para que essa se torne acessível às enzimas, consistindo essa etapa em basicamente a retirada de lignina, e diminuição do conteúdo de hemicelulose e pectina [22,24,25]. Essa etapa, considerada limitante, pode envolver tratamentos químicos, físicos ou biológicos. Quando ocorre tratamento biológico, as lacases e peroxidases são as principais enzimas atuantes, liberando unidades de fenilpropanóides [26], e ocasionalmente, as enzimas pectinases atuam na hidrólise de pectina, quando presente [22].

A seguinte etapa do processo é a sacarificação, que envolve a hidrólise da celulose e hemicelulose. Para que isso ocorra, são necessárias diferentes enzimas que possam atuar no interior das cadeias como as endoglucanases, xilanases e arabinanases, e nos terminais das mesmas, como β-glicosidases e xilosidases, além dos resíduos laterais, como arabinofuranosidases e acetil xilana esterases [16,18].

3. Organismo termófilos e *Clostridium thermocellum*

As bactérias apresentam importante relevância ecológica, quando em associações com outros seres vivos e com o ambiente em que estão inseridos [27]. Diversos estudos são realizados para desvendar a diversidade de microogranismos, e a dinâmica existente em determinadas comunidades [28]. A cooperação entre esses organismos desenvolvida em uma relação de sintrofia são necessárias para a estabilidade de uma comunidade [28].

Um interesse particular dentro de algumas comunidades, é direcionado às bactérias termofílicas, que apresentam temperatura de crescimento variado acima de 50° C. Esse interesse é devido a esses organismos apresentarem uma grande vantagem para uso biotecnológico haja vista que os processos industrias ocorrem em sua maioria em altas temperaturas, o que se adequa às enzimas de termófilos que são principalmente termoestáveis [29]. Além disso, a manutenção dessa atividade enzimática em elevadas temperaturas favorece a redução de contaminações no processo industrial [2,29]. Porém, grande parte das espécies presentes nesse grupo são ainda não cultiváveis, ou ainda não foram descobertas, por serem encontradas em ambientes hostis, de difícil coleta. Dados de 2015, estimam um total de aproximadamente 400.000 espécies bacterianas presentes em bancos de dados, e apenas 12.000 de organismos termófilos ou hipertermófilos, e dentre esses, menos de 1000 são descritos formalmente [30]. Portanto, a caracterização de organismos termófilos e elucidação do seu metabolismo possui grande relevância para aumento de informações sobre esse grupo, como conhecimento básico, e culminando com a eventual descoberta de organismos com potencial para utilização em processo industrial.

Organismos que possuem a capacidade de hidrolisar celulose estão distribuídos em diferentes habitats, e desempenham essa função em forma conjunta, envolvendo diferentes espécies [31]. Dentro dos termófílos, *Clostridium thermocellum*, uma bactéria gram-positiva anaeróbica, pertencente ao filo *Firmicutes*, foi descrita primeiramente como possuindo atividade celulolítica (linhagem ATCC27405) em importantes trabalhos desde a década de 50, e tem sido amplamente estudada desde então [32,33].

Diferentes linhagens de *C. thermocellum* já foram descritas, isoladas de diferentes habitats, sendo a primeira, DSM1313 (LG8), isolada de solo em 1926, possuindo genoma sequenciado completo [34–38]. Posteriormente, isolados, ATCC27495, YS, LQR1, JW20, BC1, que possuem genoma sequenciado, foram isolados de fontes termais (YS), algodão (JW20), e como contaminante de cultura pura (LQR1) [39].

Com o aumento de trabalhos de metagenoma, associação simbiótica de organismos com capacidades celulolíticas em intestino de insetos e rúmen de animais têm sido relatados [40,41]. Uma linhagem de *C. thermocellum*, nomeado B8, foi recentemente isolada de rúmen de caprinos e caracterizada perante à sua habilidade de degradação de fontes de carbono diversas, incluindo biomassa lignocelulósica [42–44].

Tem sido relatado em diversos estudos que *C.thermocellum* é capaz de utilizar uma variedade de fontes de carbono, entre monômeros como glicose e frutose, seguido pelo dissacarídeo celobiose, e polímeros como a celulose, CMC e avicel. A xilana e a xilose por sua vez não são utilizadas por *C.thermocellum* como única fonte de carbono [45]. Porém, biomassas complexas, contendo tanto celulose como xilana são utilizadas por *C.thermocellum*, como bagaço de cana, palha de cana, polpa de madeira, piolho de algodão (resíduo de algodão), entre outros [43,46–48].

Trabalhos iniciais de caracterização do crescimento de *C.thermocellum* em celulose levaram à descoberta da maquinaria responsável pela degradação desse polímero, o celulossoma [49,50]. O celulossoma, pode ser considerado por um grande complexo proteico (2 a 10 MDa), constituído por uma proteína estrutural (escafoldina) denominada CipA que possui um domínio de ligação à celulose (CBM3), e domínios para ligação à enzimas com atividade holocelulolítica que podem ser xilanases, endoglucanases, celobiohidrolases, quitinase, entre outras [51,52]. A figura 1 apresenta um esquema que melhor exemplifica as proteínas relacionadas à formação do celulossoma e sua ancoragem. A ligação entre essas enzimas e a proteína estrutural é feita pela interação entre os motivos de sequências de aminoácidos conservados denominados doquerina do tipo I, encontrados nas enzimas, e coesinas do tipo I, encontradas na proteína estrutural [53,54]. Além disso, a ligação entre doquerina e coesina é específica entre espécies, apesar de algumas exceções, portanto enzimas celulossomais de *C.thermocellum* não podem se ligar em proteínas estruturais, contendo coesina, de outros organismos que também apresentam celulossoma, como por exemplo o *Clostridium josui*. Essa especificidade tem sido evidenciada por ter uma maior relação com a estrutura tridimensional formada entre a ligação, do que somente com a presença de determinados aminoácidos conservados nos domínios doquerina/coesina [55].

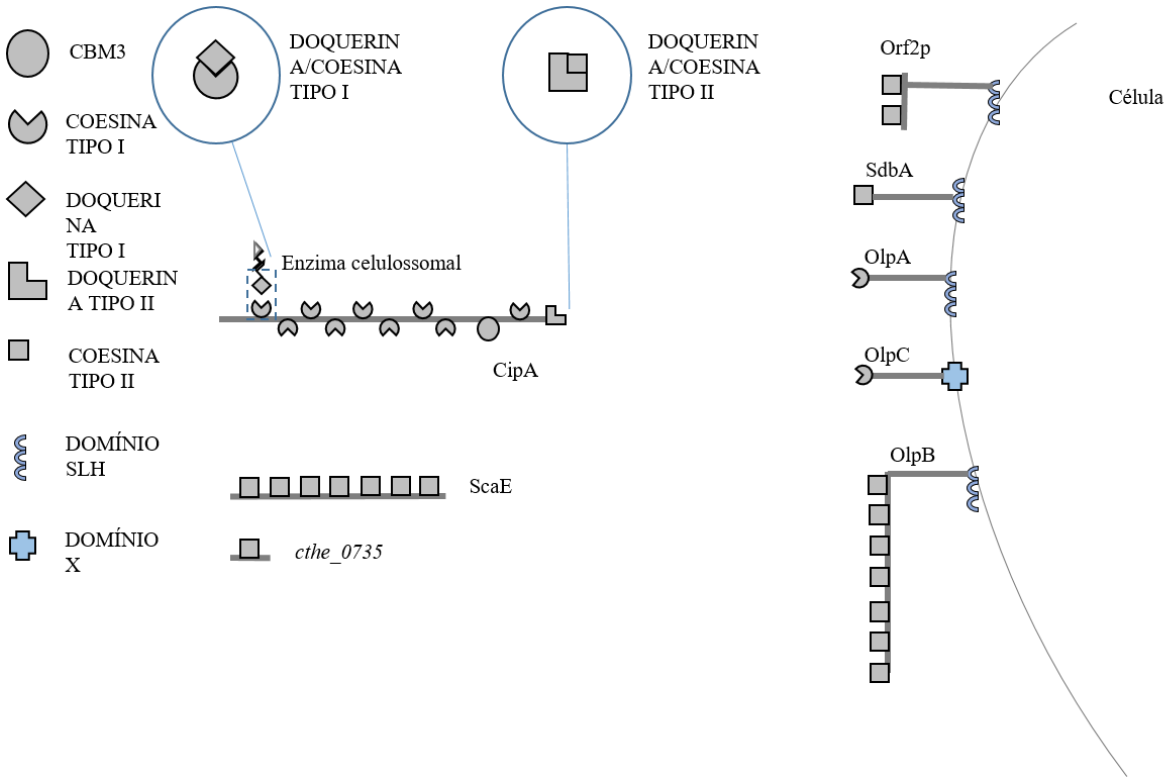


Figura 1. Representação das estruturas que compõem o celulossoma, presentes externamente à célula. Estão identificados as proteínas estruturais CipA, e proteínas estruturais âncora: Orf2p, SdbA, OlpA, OlpC, OlpB, e proteínas estruturais livres *cthe_0735* e ScaE. Do lado esquerdo da figura estão identificados os domínios CBM3, doquerina e coesina do tipo I e II, e os domínios presentes na membrana da célula, SLH e X.

Além disso, a CipA, possui um domínio de ligação denominado doquerina do tipo II, que se liga à coesina do tipo II, presente nas proteínas estruturais âncora como OlpB, Orf2p e SdbA [56]. A diferença entre essas escafoldinas é o número de CipAs que podem ser ancoradas, sendo que isso vai depender pelo número de coesinas do tipo II presentes em sua estrutura. Assim, OlpB possui o maior número de coesinas do tipo II, totalizando 7, o que constituirá um poli-celulossoma, produto da junção de 7 CipAs, resultando em 63 enzimas. As outras duas escafoldinas, possuem apenas uma coesina, porém todas as três possuem um domínio transmembrana, SLH, no C-terminal que as ancoram na superfície da célula [56]. Outras proteínas estruturais, como OlpA e OlpC, apresentam coesinas do tipo I, o que faz enzimas celulossomais se ligarem diretamente à parede da bactéria, uma contendo um domínio SLH de ancoragem e a outra contendo um domínio X, pouco conhecido [57]. Duas recentes proteínas estruturais reveladas foram a ScaE (*cthe_0736*), contendo 7 coesinas do tipo II, e a outra proteína codificada pelo gene (*cthe_0735*), ambas sem domínio de ancoragem à parede da bactéria [58].

4. Proteínas carboidrato-ativas

As enzimas conhecidas como carboidrato ativas, que englobam as citadas acima presente no celulossoma, são responsáveis pela hidrólise de carboidratos diversos, atuando nas cadeias principais e laterais do polímero [59]. Essas enzimas definidas como atuantes na hidrólise ou modificação de polímeros de carboidratos, estão catalogadas e curadas no banco de dados CAZy (Carbohydrate active enzymes), que as agrupa com base nos domínios catalíticos presentes, e que são esses: glicosil hidrolase (GH), carboidrato esterase (CE), glicosil transferase (GT), polissacarídeo liase (PL) e atividade auxiliar (AA), sendo a última classe não encontrada até a presente data em *C.thermocellum* [60,61]. As enzimas que contém esses domínios são comumente denominadas com os nomes dos domínios presentes, como por exemplo, enzimas que possuem domínios GHs são comumente denominadas como glicosil hidrolases. Também presente no CAZy, porém sem atividade catalítica, mas com grande importância, é o domínio denominado CBM (domínio de ligação à carboidrato), intensamente associado às proteínas contendo domínios GHs, possuindo um papel fundamental na ancoragem da enzima ao substrato. Esse domínio é comumente encontrado nas GHs de *C.thermocellum* [62].

Tabela 1. Categoria de domínios carboidrato ativos catalogados no banco CAZy.

Família	**Mecanismo geral**	**Famílias**	**Polímero ativo**
Glicosil hidrolase **(GH)**	Hidrólise da ligação glicosídica	1-135	Celulose/hemicelulose
Carboidrato esterase **(CE)**	Clivagem dos grupos O/N-acetil como substituintes no polissacarídeo	1-16	Hemicelulose
Glicosil transferase **(GH)**	Formação de ligação glicosídica	1-103	
Polissacarídeo liase **(PL)**	Clivagem da ligação ácido urônico-polissacarídeo	1-24	Pectina
Atividade auxiliar **(AA)**	Enzimas oxidativas atuantes em conjunto com as CAZimas	1-13	Lignina/Quitina/Celulose
Domínios de ligação ao carboidrato **(CBM)**	Sem função catalítica/ ligação ao carboidrato	1-81	

As GHs, o grupo mais importante e diverso, atuam principalmente nos polímeros constituintes principais da parede vegetal; celulose e hemicelulose. A celulose, formada unicamente por glicose, é hidrolisada pelas celulases, que se dividem em: Endoglucanases (E.C. 3.2.1.4), que catalisam a hidrólise de ligações β-1,4 internas na cadeia de celulose; as exoglucanases ou celobiohidrolases (E.C. 3.2.1.91 e EC 3.2.1.176), que catalisam a hidrólise a

partir das extremidades redutoras e não-redutoras produzindo celobiose, e as β- glicosidades (E.C. 3.2.1.21) que catalisam a hidrólise de celobiose a glicose [18].

A hemicelulose, formada por uma diversidade maior de açúcares, sendo em sua maioria xilose é hidrolisada principalmente por xilanases : endo-1,4-β-D- xilanases (E.C3.2.1.8), que clivam aleatoriamente cadeias de xilana e arabinoxilana, e as β-xilosidases (EC 3.2.1.37) que degradam nos terminais não redutores os xilooligossacarídeos e xilobiose em xilose [16]. Polímeros também encontrados na hemicelulose, como arabinana e manana são degradados por arabinanases (EC 3.2.1.99), e endo-β-mananases (EC 3.2.1.78), que atuam no interior da cadeias, e α-arabinofuranosidases (EC 3.2.1.55) e β-manosidases (EC 3.2.1.25) que atuam no terminal da cadeia (exo-ativas), resíduos laterais, ou oligossacarídeos [63].

A CEs são enzimas que atuam nas cadeias laterais dos mesmos polímeros citados anteriormente, catalizando O-ou N-deacetilação. Nesse contexto se encontram as enzimas que atuam nos grupos acetil (acetil-xilana-esterases), EC 3.1.1.72, e metil de glucuroxilanas, arabinoglucuroxilanas e galactomananas [64]. Além das acetil esterases, as CEs agrupam enzimas com atividade de feruloil esterase (EC 3.1.1.73), que hidrolisam os grupos de ácido ferúlico encontrados nas hemiceluloses, e pectinesterases, que hidrolisam ácido galacturônico presente nas cadeias laterais de pectinas.

As enzimas pectato liases (PLs), atuam na clivagem do ácido galacturônico, compreendo nesse grupo pectina liase (EC 4.2.2.10), pectato liase (EC 4.2.2.2), e pectato dissacarídeo liase (EC 4.2.2.9) [29]. Além das PLs, atuam também no polímero de pectina, GHs que hidrolisam no interior da cadeia, pectinase (EC 3.2.1.15), e nos terminais, exopoligalacturonase (EC 3.2.1.67), e exopoligalacturanosidase (EC 3.2.182). Além disso, participam na hidrólise de ramnose nos resíduos laterais, as α-L-ramnosidases (EC 3.2.1.40).

Clostridium thermocellum apresenta, portanto, um conjunto de aproximadamente 100 proteínas catalogadas no CAZy, sendo então carboidrato-ativas, contendo domínios catalíticos ou apenas CBM. Desse número, *C.thermocellum* apresenta enzimas livres, que não apresentam domínio doquerina, como por exemplo a enzima arabinofuranosidase CtAraf51 (cthe_2548), e uma das primeiras endoglucanases caracterizadas de *C.thermocellum*, CelC (cthe_2807) [65,66].

CAZy proteínas pertencentes ao celulossoma em *C.thermocellum* totalizam 73 [38]. De acordo com estudos anteriores, como demonstrado na figura 2, 47 proteínas possuem domínios GHs, compreendendo endo-β-glucananases, exo-glucanases, mananases, α-arabinofuranosidases, xilanases e galactanases [38,67]. Contendo domínios CEs são 9 no total, com ativadades acetil xilana esterases, acetil pectina esterases e feruloil esterases. Um total de

três proteínas contendo domínio PL, possuem atividade de pectato liase e ramnogalacturonase liase [48,58]. Algumas proteínas apresentam ambos domínios GH e CE como é o caso da XynZ, que possui os domínios GH10 e CE1, o que faz essa enzima ter atividade de xilanase e feruloil esterase [68]. Outras proteínas contabilizadas nesse montante apresentam apenas domínios CBM anotados, ausentes de domínios catalíticos, ou não apresentam nenhum domínio além de doquerina tipo I, portanto possuem ainda função desconhecida [67]. Esses são os casos de proteínas codificadas pelos genes por exemplo *cthe_2195*, da linhagem ATCC27405, que apresenta domínio CBM6 e doquerina tipo I anotado [48].

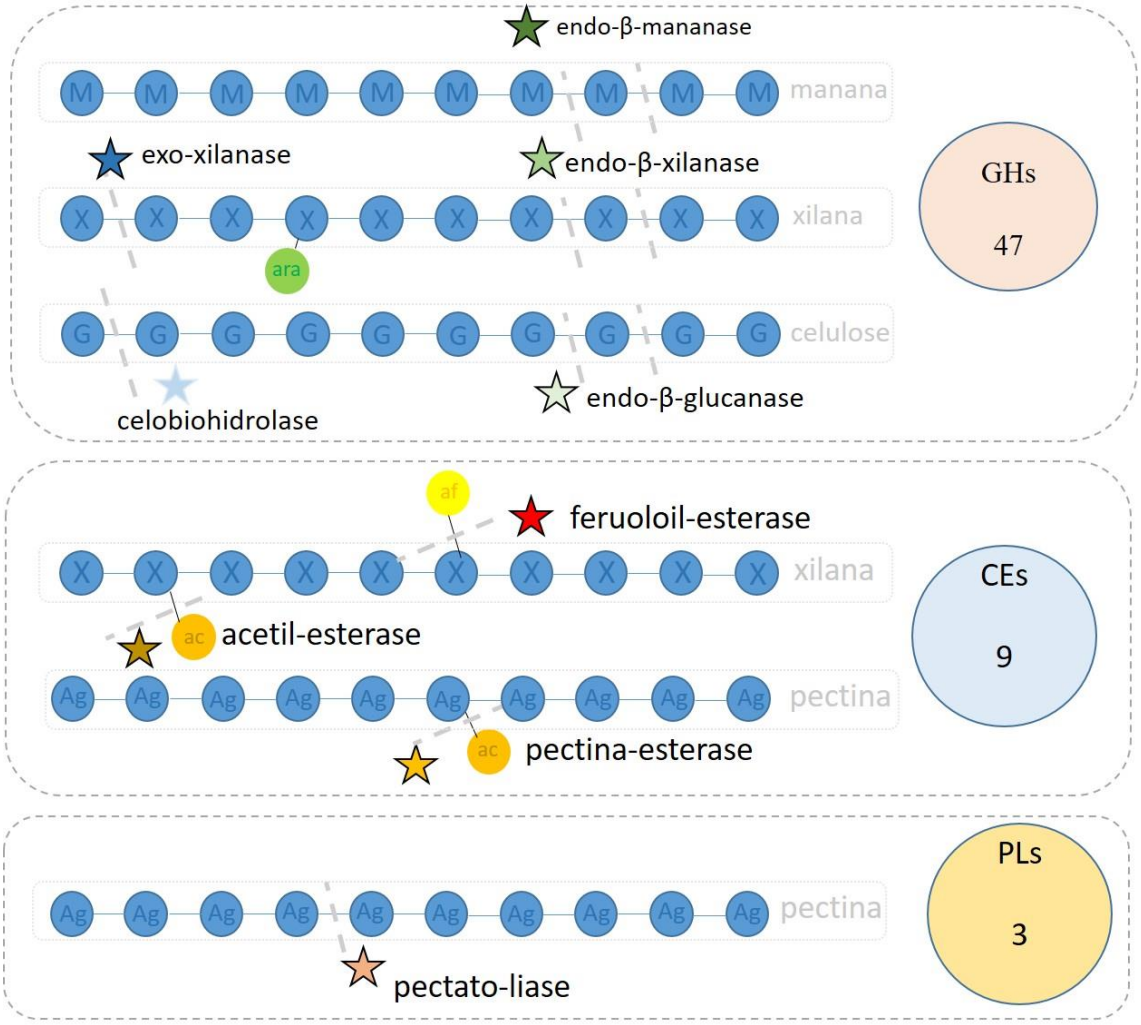


Figura 2. Esquema demonstrando o número de enzimas carboidrato ativas celulossomais de C.thermocellum, das famílias glicosil hidrolases (GHs), carboidrato esterases (CE) e pectato liases (PLs). Estão demonstradas as atividades mais importantes das enzimas celulossomais encontradas nas famílias indicadas nos polímeros, e respectivos monômeros, de manana (manose (M)), xilana (xilose (X)), celulose (glicose (G)) e pectina (ácido galacturônico (Ag)). As estrelas indicam a presença de uma enzima, e as linhas tracejadas o ponto da clivagem. Resíduos laterais são indicados como Ace (acetil), afe (ácido ferúlico) e Ara (arabinana).

A última família adicionada ao CAZy, agrupa as enzimas com domínio de atividade auxiliar (AA), que define um grupo que em sua maioria realiza, por meio de mecanismo de óxido-redução, a clivagem ou modificação do carboidrato [69]. Dentro desse grupo estão presente as lacases, atuantes em lignina, que mesmo não sendo um carboidrato e sim um agrupamento fenólico complexo, foram incluídas pelo fato da lignina estar intimamente ligada às fibras de celulose e hemicelulose. Nos grupos AA9-AA13, encontramos as enzimas que foram recentemente descritas, em 2010, as chamadas monooxigenases polissacarídeo líticas cobre-dependentes (LPMOs), caracterizadas por terem atividade residual em celulose ou quitina, porém atuam como ativadoras na atividade de glicosil hidrolases, quando atuam em sinergia [70]. As atividades auxiliares (AAs) então na presente data não foram anotadas e caracterizadas em *C.thermocellum*.

Além da vantagem de *C.thermocellum* em hidrolisar os diferentes componentes da parede vegetal, esse organismo é fermentativo, sendo capaz de transformar glicose em etanol e em outros produtos, como acetato e lactato [67]. Essa característica pode ser utilizada em processos de obtenção de etanol a partir da biomassa lignocelulósica, o bioetanol de segunda geração. Assim, *C. thermocellum* é um organismo candidato como fonte de enzimas para aplicação em bioprocessos consolidados, em função das características cinéticas de suas enzimas com atividade holocelulolítica em altas temperaturas e de sua capacidade de fermentar hexoses [39] .

5. Metabolismo de carboidratos

Posteriormente à hidrólise de carboidratos pelo celulossoma no meio extracelular, esses açúcares ficam disponíveis para serem utilizados pela célula bacteriana para produção de energia. Desde monômeros de açúcar às dextrinas, os quais são transportados ao interior da célula por meio de transportadores acoplados à ATP, e então sofrem degradação por meio de celodextrinases e celobiase phosporylases, formando a glicose e glicose-1-fosfato que finalmente entram na via glicolítica [71,72]. Dextrinas com grau de polimerização variados (2 a 7) tem sido demonstrados por serem transportados para o interior da célula, sendo a maioria observada na média de n=4 [73,74].

A glicose uma vez dentro da via Embden-Meyerhof, produz fosfoenolpiruvato e piruvato com a concomitante produção de ATP e NADH, e o piruvato é utilizado posteriormente na produção de lactato e etanol [45]. Os principais produtos de fermentação gerados por *C.thermocellum*, são o etanol, acetato e CO_2, e em menor quantidade o lactato e formato [75].

O ciclo do ácido cítrico em *C.themocellum* é incompleto, porém já foi demonstrado que em situações onde a bactéria está em crescimento com altas concentrações de celulose disponível, intermediários da via como malato e fumarato são encontrados entre os metabólitos, além de acúmulo de piruvato [75]. Portanto, estudo relacionados ao entendimento de vias alternativas em *C.thermocellum* e outros organismos anaeróbicos são continuamente realizados com o objetivo de preencher essas " lacunas" presentes no metabolismo de carboidrato e produção de energia, com a criação de modelos metabólicos baseados nos dados de genômica e proteômica existentes [76].

O consumo de produtos finais da fermentação por outros microorganismos, que seria a situação real ocorrida na natureza, uma comunidade, teria grandes efeitos para o metabolismo de *C.thermocellum* [75]. Não somente esse fator, mas por exemplo, o grande requerimento de vitaminas na cultura de crescimento de *C. thermocellum* é um indicativo de que a bactéria existe na natureza em forte associação com outros microorganismos [75]. Estudos com co-culturas de *C. thermocellum* tem sido testadas para o entendimento dessa associação durante a degradação e fermentação de fontes de carbono, além de ter como objetivo o aumento dos produtos de fermentação de interesse [77].

6. Expressão diferencial dos genes de *C.thermocellum*

A expressão diferencial de genes de um organismo pode fornecer informações sobre a função dos genes, além de poder indicar uma função desse gene e seu papel em uma determinada via metabólica, quando expresso em determinada condição experimental que está sendo analisada [78].

Estudos já foram realizados na análise da expressão diferencial de genes de *C. thermocellum* no cultivo diferencial relacionado com a fonte de carbono, comomadeira de *Populus trichocarpa*, *f*olhas de gramínea (*Panicum virgatum*), e álamo amarelo (*Liriodendrontulipifera*) [38,48,79]. Avaliando as mudanças na expressão dessas condições em diferentes tempos de crescimento, concentração de substrato e tipo de substrato em comparação com a celobiose, foram encontrados genes diferencialmente expressos que estão envolvidas nas vias metabólicas de obtenção de carbono, nitrogênio, e energia, além dos genes celulossomais envolvidos na degradação de carboidrato.

Um controle mais refinado descrito em *C.thermocellum*, e relacionado às enzimas diretamente envolvidas na degradação das fontes de carbono é mediado em parte por fatores sigma, genes homólogos aos fatores sigma encontrados em *Bacillus subtillis* [80]. Estes fatores (sigI-rsgI) foram descritos como proteínas sensores que respondem à presença de carboidrato no meio externo e ativam a expressão de genes com papel na metabolização de carboidratos. A proteína denominada anti-sigma (rsgI) é uma proteína transmembrana que em sua porção extracelular, C-terminal, contém uma sequência de aminoácidos correspondente a um CBM (módulo de ligação à celulose) e na porção citoplasmática, N-terminal, apresenta um domínio de ligação ao fator sigma (sigI). Na ausência de carboidratos no meio extracelular, os fatores sigma estão associados à rsgI, na porção citoplasmática, no entanto, na presença do carboidrato no meio externo, este se liga ao CBM (presente na porção externa da membrana), que leva à dissociação de sigI que agora está apto a regular a expressão de genes que codificam enzimas envolvidas na degradação de carboidratos [81,82].

Atualmente, foram descritos 8 pares de fatores sigmas, sendo que a porção encontrada no C-terminal dos fatores anti-sigma, são específicos a diferentes substratos, sendo detalhadas essa especificidade : RsgI1, RsgI2, RsgI4, RsgI6, RsgI24 à celulose, RsgI6, à xilana, e RsgI3, à pectina [81–84]. E ao mesmo tempo que os fatores apresentam uma especificidade, os respectivos domínios de reconhecimento aos carboidratos possuem diferentes composições. Os fatores RsgI1, RsgI2 e RsgI4, apresentam um domínio semelhante ao CBM3, enquanto o RsgI3 apresenta um domínio denominado PA14, e o RsgI6 e RsgI24 apresentam domínios de GH10 e GH5 respectivamente [84,85].

Alguns genes celulossomais foram descritos por serem regulados por fatores sigma, por meio do estudo dos promotores dependentes. São esses: sigI1 (CelS), e sigI6 (XynB, XynD, XynZ, XynY, CelV,CseP, CipA) [80]. Em trabalho anterior, autores compararam o crescimento de *C.thermocellum* em álamo amarelo e celobiose, e observaram dentro dos 60% dos genes de todo metabolismo com expressão aumentada em álamo, os pares SigI-RsgI 4, 3 e 24 [48].

Além do controle da expressão gênica por meio dos fatores sigma, ao longo do genoma de *C.thermocellum* são encontrados alguns clusters envolvendo glicosil hidrolases e genes envolvidos no celulossoma [45,86]. Dentre os clusters encontrados, dois operons foram descritos e caracterizados, um operon contendo genes não celulossomais *celC-glyR3-licA*, e outro contendo dois genes celulossomais, *manB* e *celT* [86–88].

Portanto, o estudo acerca a bactéria *C.thermocellum*, mesmo contendo diversas referências sobre o seu metabolismo e modo de degradação das suas enzimas, muito ainda precisa ser elucidado. Principalmente sobre o metabolismo que envolve todas as bactérias anaeróbicas e enzimas que sejam eficientes na degradação de biomassas complexas, como por exemplo, o bagaço de cana e a palha de cana.

Dessa forma, a junção da análise proteômica e transcritômica de *C.thermocellum* quando em crescimento em resíduos da cana-de-açúcar produzirá informações sobre enzimas e vias metabólicas envolvidas na degradação dessas biomassas. Além disso, a caracterização de enzimas, celulossomais ou não, que atuam na degradação de hemicelulose, fornece uma possibilidade de maior entendimento dos mecanismos de adaptação de bactérias anaeróbicas no uso de biomassas como fonte de carbono, e assim tornando possível o seu uso na indústria biotecnológica.

Objetivo geral

O presente trabalho tem como objetivo geral mapear proteínas e genes envolvidos tanto diretamente na degradação de biomassa lignocelulósica, como indiretamente, no controle da expressão desses genes ou ativação de outros mecanismos que atuam na melhoria do crescimento de *Clostridium thermocellum* quando cultivada em resíduos agroindustriais como fonte de carbono.

Objetivos Específicos

- Analisar o perfil de proteínas de *C.thermocellum* secretadas, e aderidas à biomassa, quando a bactéria é cultivada em meio de cultura contendo celulose, ou bagaço de cana, ou palha de cana de açúcar como fonte de carbono, utilizando espectrometria de massa.
- Analisar o conjunto de transcritos expressos, e a expressão diferencial, por *C.thermocellum* quando cultivada em meio de cultura contendo celulose, ou bagaço de cana, ou palha de cana de açúcar como fonte de carbono utilizando sequenciamento de RNA (RNAseq).
- Utilizar os dados de RNAseq e proteoma para descrição dos mecanismos utilizados pela bactéria para degradação dos substratos complexos bagaço de cana e palha de cana.
- Clonar e expressar um gene cellulossomal de *C.thermocellum, axb8,* não caracterizado, em *E.coli*, visando a obtenção da análise bioquímica e funcional.

Capítulo 2. Análise proteômica descritiva de proteínas de *C.thermocellum* associadas à biomassa e envolvidas na sua degradação

1.Introdução

A desconstrução da parede celular vegetal é um processo complexo que requer a participação de um arsenal de enzimas, e o conjunto de enzimas que apresente maior eficiência na degradação de biomassas complexas de interesse biotecnológico é alvo de constante estudo em misturas enzimáticas [16]. Devido a esse interesse por enzimas eficientes, *C.thermocellum* se apresenta como um importante organismo para estudo, devido à sua habilidade de degradar diferentes biomassas, além de apresentar notável eficiência durante a degradação, devido às suas enzimas se apresentarem em forma de complexo enzimático, o celulossoma [47,50].

Estudos utilizando análise proteômica demonstraram o perfil de proteínas celulossomais secretadas por *C.thermocellum* e a modulação dessas proteínas expressas quando a fonte de carbono é modificada utilizando celulose e celobiose como fonte de carbono [52,89]. Outros trabalhos foram realizados utilizando fontes de carbono mais complexas, como gramínea (*Panicum virgatum*) pré-tratada, celulose, celulose combinada com pectina e celulose combinada com xilana [90]. Ambos estudos demonstraram uma diferença de expressão tanto nas proteínas estruturais, quanto nas enzimas responsáveis pela degradação de carboidrato, como o aumento de glicosil hidrolases da família 9 em meios contendo celulose e gramínea, quando comparado com as outras fontes de carbono.

Somente em 2010, foi demonstrado que havia um controle da expressão de genes em *C.thermocellum* em resposta a fonte de carbono no meio de cultura, e este controle estava sendo mediado por fatores sigma [80]. Outros trabalhos vieram posteriormente, para uma maior elucidação dessa sinalização que ocorre com a expressão dos fatores sigma (sigI) e anti-sigma (rsgI) [81,82]. Assim, estudos de proteômica e transcriptômica identificaram a expressão desses fatores e estudaram a modulação dessa expressão em relação aos tipos de carboidratos presentes no meio extracelular, e a expressão diferencial de genes envolvido na degradação da biomassa e metabolismo como um todo [48,91].

Além disso, *C.thermocellum* utiliza outros mecanismos que aumenta a eficiência de degradação da fonte de carbono, como por exemplo a formação de biofilmes, com o intuito de aproximação da célula ao substrato [92].

Dessa forma, seguindo essa linha de pesquisa relacionada ao controle de expressão de genes de *C.thermocellum* quando modificada a fonte de carbono, o nosso grupo, utilizando um

novo isolado (B8), avaliou o crescimento da bactéria em diferentes fontes de carbono não utilizadas previamente, como bagaço de cana e palha de cana não pré-tratados, ambos resíduos industriais de grande interesse para a economia brasileira [43,44]. Esses estudos demonstraram a capacidade desse novo isolado em crescimento tanto em celulose e outras fontes de carbono, assim como foi realizada a identificação de atividades enzimáticas, e expressão de alguns genes celulossomais identificados por PCR em tempo real. Além disso, foi realizada a identificação de proteínas celulossomais por espectrometria de massas, quando em crescimento em celulose.

Contudo, este é o primeiro trabalho que traz a identificação das proteínas tanto secretadas, quanto aderidas à biomassa, e celulossoma parcialmente purificado, de *C.thermocellum* em crescimento em bagaço de cana e palha de cana. Com isso, visando descrever as proteínas que são possivelmente importantes na desconstrução das biomassas utilizadas nesse trabalho, o que pode trazer perspectivas sobre a utilização dessas enzimas em possíveis aplicações biotecnológicas.

2. Metodologia

2.1. Condições de cultivo de *C. thermocellum*

A linhagem de *C.thermocellum* isolado B8, modelo deste trabalho, foi isolada de amostra líquida de rúmen de caprino da raça Moxotó (Cunha *et al*., 2011). O isolamento foi realizado a partir da amostra líquida utilizando meio líquido redutor, glicose e celulose, como fonte de carbono [43]. O meio redutor (1L) conforme descrito por Freier [49] contém em $g.L^{-}$ 1,5g de NH_2PO_4, 2,5g Na_2HPO_4, 0,5g NH_4Cl, 0,5g $(NH_4)_2SO_4$, 0,5g de $NaHCO_3$, 0,09g de $MgCl_2$, 5,00 mL solução mineral, 0,50 mL solução de vitaminas, 200ml NaOH (0,2M), 100µL $Na_2S_9H_20$, 3g extrato de levedura e 1% de fonte de carbono (g/L). A atmosfera anaeróbica do meio foi mantida pela adição dos agentes redutores cisteína e sulfito de sódio, além da injeção de nitrogênio gasoso.

As biomassas utilizadas no trabalho, bagaço de cana e palha de cana, foram trituradas em água destilada, autoclavadas por 40 minutos, lavadas em água abundante e depois secada em estufa a 70°C. Para a obtenção dos secretomas, *C. thermocellum* foi inoculada em 100 mL de meio de cultura em diferentes condições: suplementado com 1g de celulose, bagaço de cana, ou palha de cana de açúcar como fonte de carbono. Estes foram incubados a 60°C a 200 rpm. O cultivo foi realizado em triplicata biológica, por 96 horas para as amostras contendo bagaço de cana e palha de cana, conforme descrito anteriormente como possuindo maior atividade

enzimática de CMCase e Xilanase nessas mesmas fontes de carbono, e 48h para as amostras de celulose [93].

2.2. Preparo e análise das amostras do secretoma

Com o objetivo de obtenção do maior número de proteínas a serem identificadas, foram realizados três tipos de preparos de amostras. Primeiramente, as culturas obtidas como descrito acima foram filtradas à vácuo em papel filtro n°5 da marca Whatman (GE Healthcare Life Sciences), reservando o filtrado. O filtrado foi centrifugado a 8.000 rpm por 15 minutos e o sobrenadante obtido, nomeado fração do sobrenadante (FS). A fonte de carbono residual foi recuperada do papel e lavada com Tampão Tris-HCl 100 mM pH 8,0. As proteínas foram eluídas do substrato residual após lavagem com volume de 45 mL de água Mili-Q, por frasco de cultura. Posteriormente a amostra eluída foi concentrada por ultrafiltração em membrana com retenção de 200 kDa (PM 200 Millipore Co., MA, USA) e denominada fração eluída do substrato (FES). A amostra FES foi utilizada para purificação por gel filtração como descrito anteriormente (Brenda, 2012), e o purificado denominada como celulossoma parcialmente purificado (CPP). A purificação dos celulossomas foi realizada em sistema cromatográfico automatizado "Akta purifier" (GE, Uppsala, Sweden), utilizando a coluna cromatográfica HiLoad 16/60 Superdex S-200 (GE) equilibrada em tampão TrisHCl 100mM pH 8,0 contendo 150mM NaCl e 2mM de $CaCl_2$. Para a corrida foi utilizado fluxo contínuo de 1 mL/min, pressão de 0,6MPa e frações coletadas de 4 mL. A corrida foi monitorada a 295 nm, e os dados obtidos a partir do programa Unicorn 5.1 (GE, Uppsala, Sweden) acoplado ao sistema Äkta purifier. O primeiro pico obtido de proteína (presente no volume vazio, confirmado com uso de marcadores [42]), concomitante com as atividades de CMCase e xilanase foram coletados e denominado amostra CPP.

O substrato celulose foi utilizado para preparo apenas da amostra FS. Os dados de FES e CPP foram obtidos anteriormente pelo grupo e usados nas análises seguintes de comparação com as amostras obtidas de bagaço e palha de cana como substrato [42].

2.3. Atividade enzimática e quantificação de proteína

As diferentes frações geradas como descrito acima foram quantificadas por método de Bradford [94], para detecção de proteína, e método de DNS (ácido 3,5-dinitrosalicílico) [95], para atividade enzimática de CMCase e xilanase. O método de Bradford foi realizado com a utilização do kit do Quick Start do fabricante BioRad (BioRad, EUA), de acordo com o fabricante: 150 µL de amostra, adicionado de igual quantidade do reagente, seguido por 5 minutos de incubação e leitura a 595 nm. A curva de calibração foi realizada utilizando a proteína BSA nas concentrações de 1 a 10 µg.

As atividade enzimáticas foram quantificadas de acordo com o método de DNS originalmente descrito por Miller [95]. Os substratos utilizados foram carboximetilcelulose (CMC) e xilana oat spelt preparados em concentração de 2% em tampão acetato de sódio pH 5.0 100mM e tampão fosfato de sódio pH 6.0, acrescido de DTT e $CaCl_2$. As curvas de calibração foram realizadas com os açúcares glicose e xilose nas concentrações de 12 a 60 µg. O ensaio enzimático foi realizado com 100 µL de amostra e 50 µL de substrato, incubado a 65°C por 30 minutos, e a reação interrompida com 300 µL de DNS, com posterior fervura por 10 minutos. A leitura foi realizada a 540 nm. Uma unidade de atividade enzimática foi definida como a quantidade de enzima necessária para obter 1 µmol de açúcar (glicose ou xilose), por minuto de reação. Atividade específica foi calculada realizando a divisão da atividade enzimática pela quantidade de proteína.

2.4. Preparo de amostra para aquisição em nanoUPLC-MSE

As amostras FS, FES e CPP foram dialisadas em água mili-Q utilizando membrana de diálise com limite de exclusão de 1 kDa. Após a diálise, 100 µg das amostras foram liofilizadas, para posterior análise em nanoUPLC-MS^E [42]. Primeiramente, as alíquotas foram ressuspendidas em 10µL de bicarbonato de amónio 50 mM, seguido por 25 µL de RapiGest ™ (0,2% v/v), agitadas cuidadosamente, e incubadas em banho seco a 80 ° C durante 15 min. As amostras foram brevemente centrifugadas e 2,5 µl de DTT 100 mM foi adicionado, submetidas ao vortex e incubadas a 60 ° C durante 30 min, seguido por centrifugação a 2000x g por 1 min. Iodoacetamida (2,5 uL de uma solução 300 mM) foi adicionada, e as amostras foram ligeiramente agitadas e incubadas no escuro a temperatura ambiente durante 30 min. Em

seguida, 10 µL de tripsina (com 400 uL de 50 mM de bicarbonato de amônio adicionado em 20µg de tripsina) foi adicionado, e as amostras submetidas ao vortex. A amostra foi digerida a 37 ° C por 12 horas em banho seco. Para precipitar o RapiGest, 10 µL de solução de TFA a 5% foi adicionado, as amostras submetidas ao vortex, incubadas durante 90 min a 37 ° C em um banho seco, e centrifugadas a 10000 x g a 6 ° C durante 30 min. Os sobrenadantes foram transferidos para frascos cilíndricos do tipo Waters (Waters, EUA), liofilizados, e em seguida ressuspendidos em 100 µL de 200 mM de solução de formato de amônio contendo 50 Fmol/µL de phosphorilase B de coelho (PhosB) digerido como padrão (Waters, EUA). A concentração final da proteína foi de 1 µg/µL, e a concentração final foi de 50 PhosB Fmol/µL.

2.5. Aquisição e processamento dos dados

A amostra tripsnizada e preparada no item anterior foi aplicada em um sistema nanoACQUITYTM ™ (Waters Corp., EUA) com a tecnologia de diluição 2D equipado com uma pré-coluna C18 5µm, 5 mm x 300 µm e uma coluna analítica de fase inversa nanoEaseTM BEH130 C18 1,7 mm, 100 mm x 100 mm (Waters , EUA), para separação dos peptídeos gerados. As amostras foram inicialmente transferidas para a pré-coluna utilizando uma solução de ácido fórmico aquoso a 0,1 %. A fase móvel A consistia de ácido fórmico a 0,1 % em água, e a fase móvel B consistiu de ácido fórmico a 0,1 % em acetonitrila. Os peptídeos foram separados utilizando um gradiente de 3-40 % de fase móvel B durante 200 minutos de fluxo de 400 ηL/min ou 600 ηL/min, seguido por 10 minutos de 85 % ou 90 %. Todas as amostras foram analisadas em triplicata técnica. Os peptídeos trípticos foram analisados utilizando espectrômetro de massa SYNAPT G2 HDMSTM (Waters, Manchester, Reino Unido) com um tempo de voo (oa-TOF) geometria híbrida quadrupolo / ion mobilidade / aceleração ortogonal. As análises foram realizadas utilizando os parâmetros descritos na Murad e Rech, 2012 e Murad, Souza et al, 2011. O processamento de dados e identificação de proteínas foram realizadas utilizando ProteinLynx global Server (PLGs) versão 3.0 (Waters, Manchester, Reino Unido). As proteínas foram identificadas com o algoritmo de contagem de íons incorporado do software e uma pesquisa do banco de dados UniProt *Clostridium spp* com os padrões de digestão MassPREP (MPDS) UniProtKB / Swiss-Prot (Fosforilase -PHS2_RABIT) acrescentados ao banco de dados. Modificação fixa de carbamidometil-C foi especificada e modificação variáveis como amidação C-terminal, deamidação N-terminal, deamidação de asparagina, deamidação de glutamina, metilação N-terminal e C-terminal, glicosilação N-

terminal e o-acetilação ou oxidação de metionina. Um sítio de clivagem foi ignorado. Os componentes foram agrupados com uma precisão de massa de 10ppm e uma tolerância de tempo 0,25 min contra as massas de íons peptídeo teórico gerado pelo banco de dados com um mínimo de um peptídeo combinado. Os critérios de identificação de proteínas também incluíram a detecção de pelo menos três íons por peptídeo, 6 fragmentos por proteína e a determinação de pelo menos um peptídeo por proteína; a identificação da proteína foi deixada com uma taxa de detecção máximo 4% falso positivo em pelo menos três replicadas técnicas.

2.6 Anotações funcionais

A partir dos das proteínas identificadas, contendo número de acesso Uniprot, foram extraídos a partir do mesmo banco de dados, dados como o nome do gene, baseado na linhagem de *C.thermocellum* ATCC27405 (NC_009012.1), e outras linhagens, quando ausente o gene, como DSM1313 (NC_017304.1), e AD2 (NZ_CP013828.1), anotação de Pfam e COG [96,97]. Informações adicionais acerca das proteínas carboidrato-ativas foram complementadas com base nas anotações contidas no banco CAZy (carbohidrate-active-enzymes) [61]. O software Blast2GO foi utilizado para gerar as anotações baseadas no Gene Ontology com "cut-off" e-value de 1E-6. Predição de peptídeo sinal foi realizado utilizando os softwares SignalP, SecretomeP e PsortB [98]. O software STRING 10.5 foi utilizado para analisar as interações entre proteínas identificadas [99].

3. Resultados

C.thermocellum foi cultivado como descrito acima, e as seguintes frações foram obtidas: sobrenadante (FS), substrato ligado (FES) e celulossoma parcialmente purificado (CPP). A partir das atividades enzimáticas utilizando os substratos CMC e xilana *oat spelt*, e da quantificação de proteína total, os cálculos de atividade específica foram realizados e os dados podem ser visualizados na tabela1. Na maioria das amostras preparadas é possível observar uma maior atividade específica de xilanase, em relação a CMCase. Valores de atividade específica entre 1 e 7 UI/mg foram encontrados para todos os preparos das atividades de CMCase e xilanase. Além disso, os valores encontrados para CPP de celulose foram maiores do que visualizados em bagaço e palha de cana.

Tabela 1. Atividade específica (UI/mg) de CMCase e xylanase das frações preparadas (FS (fração do sobrenadante), FES (fração eluída do substrato), e CPP (celulossoma parcialmente purificado) para cada substrato (Bagaço B, Palha, P, Celulose, C).

Amostras	Atividade específica (UI/mg)	FS	FES	CPP
B1	CMCase	1.45 ±0.14	4.14 ±0.44	1.69 ±0.01
	Xilanase	4.28 ±0.915	3.22 ±0.27	2.14 ±0.33
B2	CMCase	1.46 ±0.018	2.78 ±0.44	1.66 ±0.31
	Xilanase	4.76 ±0.548	7.44 ±1.05	2.42 ±0.19
B3	CMCase	1.60 ±0.12	2.05 ±0.27	0.82 ±0.03
	Xilanase	5.25 ±0.285	4.13 ±0.593	1.67 ±0.26
P1	CMCase	1.89 ±0.12	0.64 ±0.09	3.17 ±0.81
	Xilanase	6.06 ±0.16	4.05 ±0.09	6.72 ±0.56
P2	CMCase	1.76 ±0.13	1.30 ±0.07	3.37 ±0.46
	Xilanase	5.49 ±0.42	4.53 ±0.35	5.32 ±0.95
P3	CMCase	2.15 ±0.14	2.77 ±0.19	2.48 ±0.88
	Xilanase	6.15 ±0.17	5.77 ±0.276	4.63 ±1.11
C1	CMCase	1.34 ±0.103	1.71*	5.85*
	Xilanase	6.43 ±0.24	2.7*	6.99*
C2	CMCase	1.51 ±0.08		
	Xilanase	6.25 ±0.73		
C3	CMCase	1.37 ±0.05		
	Xilanase	4.99 ±0.49		

* Dados publicados em [42].

A identificação de proteínas utilizando a técnica NanoUPLC-MSe revelou um total de 186 proteínas de acordo com o banco de dados Uniprot, gerando um total de 137 diferentes genes anotados baseados no genoma ATCC27405 (A tabela de proteínas identificadas está presente no Apêndice A). O preparo em celulose obteve o maior número de proteínas identificas, como visualizado do diagrama de Venn (Fig.1). As maiores quantidades de proteínas foram identificadas para o preparo de amostra de celulose, quando somada todas as frações.

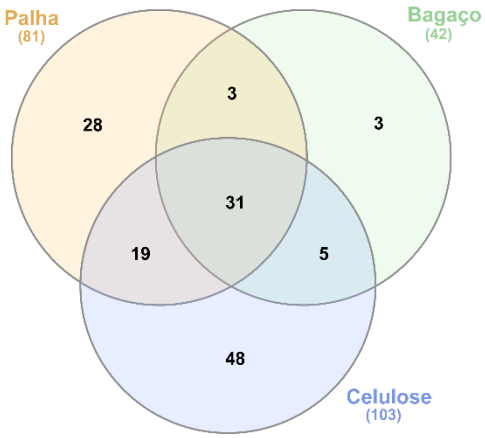


Figura 1. Diagrama de Venn mostrando proteínas em comum e exclusivas de proteínas identificadas em cada fração (FS-Fração do sobrenadante-A), FES- (Fração eluída do substrato-B), CPP-(celulossoma parcialmente purificado-C), de cada amostra utilizando os substratos palha de cana, bagaço de cana e celulose. Diagrama contendo o somatório de proteínas de todas as frações e amostras (D).

3.1. Proteínas identificadas nos meios contendo biomassa

Em geral, com base na classificação funcional (COG), as proteínas identificadas em bagaço e palha de cana foram agrupadas em 14 categorias funcionais (gráfico 2), incluindo predição de função geral (R) e função desconhecida (S). Os grupos funcionais de maior representação entre as proteínas foram: transporte de carboidratos e metabolismo (G); produção e conversão de energia (C). Paralelamente, baseado nas anotações do Gene Ontology (GO) em processos biológicos, dados mostrados em percentagem, a maioria das proteínas contendo essa anotação participam do catabolismo de carboidratos, e processos de oxido redução (Figura 2). Tabela descrevendo todos os GOs para cada proteína pode ser encontrada no Apêndice B.

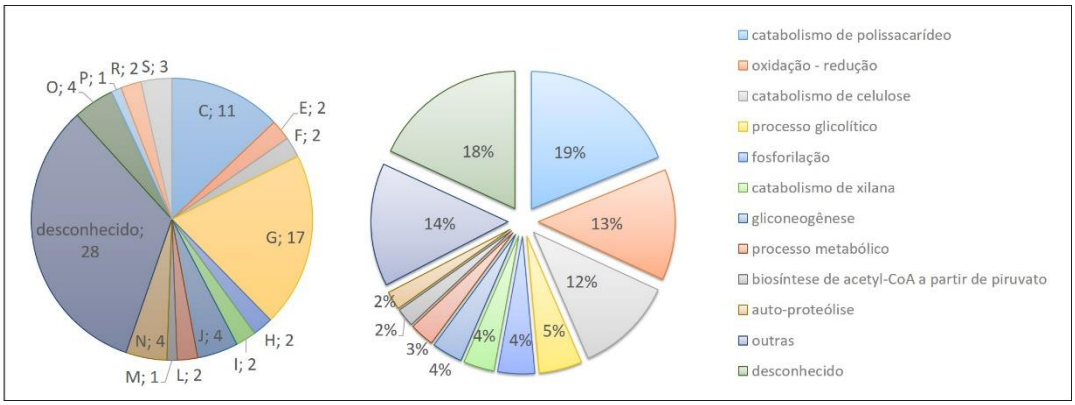


Figura 2. Proteínas identificadas nas amostras preparadas a partir de biomassa, de todas as frações, categorizadas de acordo com o COG (Cluster of Orthologous groups) (A), e Gene Ontology (GO) baseado em processos biológicos. Os COG anotados foram G (metabolismo e transporte de carboidrato), C (conversão de energia), O (modificação pós-traducional, *turnover* protéico, e chaperonas), N (motilidade celular), J (tradução, estrutura do ribossomo e biogêneses), S (função desconhecida), R (somente predição de função geral), E (metabolismo e transporte de aminoácido), F (metabolismo e transporte de nucleotídeo), H (metabolismo e transporte de coenzima), I (metabolismo e transporte de lipídeo), L (replicação, recombinação, e reparo), M (biogêneses da parede celular/membrana/envelope), P (metabolismo e transporte de íon inorgânico).

Entre proteínas agrupadas no transporte de carboidratos e metabolismo, um total de 17, foram encontradas enzimas pertencentes à via de glicólise / gliconeogênese e transportadores ABC. Além disso, nesse mesmo grupo se encontram as glicosil hidrolases e outras enzimas que pertencem à degradação de carboidratos. Proteínas dessa categoria foram identificadas envolvidas na via de Embden-Meyerhof [45]. São essas, gliceraldeído-3-fosfato desidrogenase de tipo I (*cthe_0137*), fosfoglicerato quinase (*cthe_0138*), enolase (*cthe_0143*), glucose-6-phophate isomerase (*cthe_0217*), fosfofrutoquinase (*cthe_0347*), frutose-1,6-bifosfato aldolase classe II (*cthe_0349*) e piruvato fosfato dikinase (*cthe_1308*).

Ainda na categoria G (Tabela 2), transportadores foram encontrados, como transportador de açúcar ABC (*cthe_0393*), responsável pela entrada de celobiose, proteína de ligação ao soluto extracelular CbpB (*cthe_1020*) e proteína relacionada ao transportador ABC (*cthe_1862*), os dois últimos responsáveis pelo transporte de celodextrinas [83]. Outros dois transportadores sem função definida foram caracterizados, RbsD (*cthe_0395*) e subunidade transportador ATP sulfato (*cthe_2534*).

Proteínas agrupadas em conversão de energia, um total de 11 (Tabela 2), sendo a maioria delas também anotadas como envolvidas em processos de oxido-redução (GO), foram identificadas, e muitas são envolvidas na produção de: etanol, como álcool desidrogenase

GroES (*cthe_0388*), álcool ferro desidrogenase (*cthe_0394*); acetato, como álcool desidrogenase (*cthe_0423*) e acetato quinase (*cthe_1028*); lactato, como subunidade gama piruvatoquetoisovelarato oxidoredutase (*cthe_2390*), piruvato flavoodoxinferredoxina oxidoredutase (*cthe_2392*), e proteína contendo ligação de domínio á Tiamina pirofosfato TPP (*cthe_2393*). Além disso também foram encontradas proteínas envolvidas potencialmente na produção de piruvato, a partir de oxaloacetato, por meio da ação da enzima oxaloacetato descarboxilase putativa (*cthe_0701*), e a partir de fosfoenolpiruvato, por meio da ação da fosfoenolpiruvato carboxiquinase (*cthe_2874*). Por último, nessa categoria foi identificada uma subunidade beta da ATPase (*cthe_2608*), presente na membrana da célula [100].

Proteínas envolvidas na motilidade da célula (N) foram também identificadas (Tabela2), uma delas presente na formação do flagelo, a flagelina (*cthe_2237*), e outras duas anotadas nessa categoria, porém sem função definida. Essas proteínas são: transportador ABC (*cthe_2708*), que apresenta anotação Interpro IPR019196, o que indica domínio encontrado em muitos transportadores envolvidos com a flagelina, e uma proteína putativa denominada prepilina (*cthe_1104*). A proteína prepilina já foi encontrada em proteoma de C.thermocellum, em preparos da membrana da célula [100].

Tabela 2. Identificação das proteínas nas amostras de palha e bagaço, de todas as frações (FS, FES e CPP) e sua categoria COG (Cluster of Orthologous groups) anotada. Categorias COG: (C) Produção de energia e conversão; (D) Cromossomo, divisão celular e controle do ciclo celular; (E) Metabolismo e transporte de aminoácido; (F) Metabolismo e transporte de nucleotídeo; (G) Metabolismo e transporte de carboidrato; (H) Metabolismo e transporte coenzima; (I) Metabolismo e transporte de lipídeo; (J) Biogêneses, tradução e estrutura do ribossomo; (K) Transcrição; (L) Replicação, recombinação, e reparo; (M) Parede celular/ biogêneses membrana; (N) Motilidade da célula; (O) Modificação pós-traducional, turnover proteína, chaperonas; (P) Metabolismo e transporte de íons inorgânicos; (Q) Biossíntese do metabolismo secundário, transporte e catabolismo; (R) Predição de função geral somente; (S) Função desconhecida; (T) Mecanismo de transdução do sinal; (U) Secreção e tráfego intracelular; (V) Mecanismos de defesa. Proteínas marcadas com ‡ apresentam peptídeo sinal.

Gene	Proteína	Substratos	COG
Cthe_0388	Alcohol dehydrogenase GroES domain protein	bagaço	C
Cthe_1795	Phospho-2-dehydro-3-deoxyheptonate aldolase	bagaço	E
Cthe_0395	RbsD or FucU transport	bagaço	G
Cthe_0421	Dipicolinate synthase subunit B	bagaço	H
Cthe_2891	10 kDa chaperonin	bagaço	O
Cthe_0637‡	Uncharacterized protein	bagaço	S
Cthe_2379‡	Uncharacterized protein	bagaço	
Cthe_2423‡	Type 3a cellulose-binding domain protein	bagaço	
Cthe_0394	Iron-containing alcohol dehydrogenase	bagaço/palha	C
Cthe_0423	Iron-containing alcohol dehydrogenase	bagaço/palha	C
Cthe_2874	Phosphoenolpyruvate carboxykinase [GTP]	bagaço/palha	C
Cthe_2390	Pyruvateketoisovalerate oxidoreductase gamma subunit	bagaço/palha	C
Cthe_2393	Thiamine pyrophosphate TPP-binding domain-containing protein	bagaço/palha	C

Cthe_0137	Glyceraldehyde-3-phosphate dehydrogenase type I	bagaço/palha	G
Cthe_0138	Phosphoglycerate kinase	bagaço/palha	G
Cthe_0143	Enolase	bagaço/palha	G
Cthe_0217	Glucose-6-phosphate isomerase	bagaço/palha	G
Cthe_0347	Phosphofructokinase	bagaço/palha	G
Cthe_0349	Fructose-1 6-bisphosphate aldolase class II	bagaço/palha	G
Cthe_0393	Sugar ABC transporter (Sugar-binding protein)	bagaço/palha	G
Cthe_1020	Extracellular solute-binding protein family 1	bagaço/palha	G
Cthe_1308	Pyruvate phosphate dikinase	bagaço/palha	G
Cthe_2723	50S ribosomal protein L7L12	bagaço/palha	J
Cthe_2730	Elongation factor Tu	bagaço/palha	J
Cthe_3202	CRISPR-associated protein Csh2 family	bagaço/palha	L
Cthe_2236	Flagellin domain protein	bagaço/palha	G
Cthe_2237	Flagellin domain protein	bagaço/palha	N
Cthe_1104‡	Putative uncharacterized protein	bagaço/palha	N
Cthe_1965	Peroxiredoxin	bagaço/palha	O
Cthe_2892	60 kDa chaperonin	bagaço/palha	O
Clo1313_0638	Histone family protein DNA-binding protein	bagaço/palha	
Cthe_0412‡	Glycoside hydrolase family 9/CelK	bagaço/palha	
Cthe_0413‡	CbhA	bagaço/palha	
Cthe_2089	Glycoside hydrolase family 48	bagaço/palha	
Cthe_2348	Ig domain protein	bagaço/palha	
Cthe_3077	CipA	bagaço/palha	
Cthe_3078	OlpB	bagaço/palha	
Cthe_0624‡	CelJ	palha	R
Cthe_0701	Carboxylase region-containing protein	Palha	C
Cthe_1028	Acetate kinase	Palha	C
Cthe_2238	L-lactate dehydrogenase	Palha	C
Cthe_2608	ATP synthase subunit beta	Palha	C
Cthe_2392	Pyruvate flavodoxinferredoxin oxidoreductase domain protein	palha	C
Cthe_1840	Cysteine synthase	Palha	E
Cthe_2885	Phosphoribosylaminoimidazole-succinocarboxamide synthase	Palha	F
Cthe_2924	Adenylate kinase	Palha	F
Cthe_0269‡	CelA	palha	G
Cthe_0536‡	CelB	palha	G
Cthe_1838‡	XynC	palha	G
Cthe_1862	ABC transporter related protein	Palha	G
Cthe_1963‡	xylanase Z	palha	G
Cthe_2534	Sulfate ABC transporter ATPase subunit	Palha	G
Cthe_2872‡	CelG	palha	G
Cthe_2203	GTP cyclohydrolase 1	Palha	H
Cthe_0933	Acyl carrier protein	Palha	I
Cthe_0699	Carboxyl transferase	Palha	I
Cthe_0418	Polyribonucleotide nucleotidyltransferase	Palha	J
Cthe_1228	Threonine--tRNA ligase	Palha	J
Cthe_2301‡	CRISPR-associated regulatory protein DevR family	Palha	L

Cthe_0424	Aminoglycoside phosphotransferase	palha	M
Cthe_2708‡	ABC-type uncharacterized transport system	Palha	N
Cthe_1322	Chaperone protein DnaK	Palha	O
Cthe_1754‡	ABC-type transporter periplasmic subunit	Palha	P
Cthe_0043‡	CelN	palha	R
Cthe_1503	Uncharacterized protein	palha	S
Cthe_1504	Linocin M18 bacteriocin protein	Palha	S
Clo1313_1399‡	Putative uncharacterized protein	palha	
Cthe_0275	Glycosyltransferase 36	Palha	
Cthe_0285	Isocitrate dehydrogenase [NADP]	Palha	
Cthe_0338	NADH-quinone oxidoreductase E subunit	Palha	
Cthe_0345	L-lactate dehydrogenase	Palha	
Cthe_0374	Glutamate dehydrogenase	Palha	
Cthe_0578‡	CelR	palha	
Cthe_0625‡	CelQ	palha	
Cthe_0745‡	CelW	palha	
Cthe_0798‡	Cellulosome protein dockerin type I	Palha	
Cthe_1082	Uncharacterized protein	Palha	
Cthe_1368‡	S-layer domain protein	palha	
Cthe_1398‡	Dockerin type 1 protein	palha	
Cthe_1970	Uncharacterized protein	Palha	
Cthe_2010	Putative uncharacterized protein	Palha	
Cthe_2383‡	Copper amine oxidase-like domain-containing protein	palha	
Cthe_2422‡	Uncharacterized protein	Palha	
Cthe_2855	Putative uncharacterized protein	Palha	
Cther_2217	Putative uncharacterized protein	Palha	

3.2. Proteínas putativas

Com o intuito de investigar as proteínas que possam estar envolvidas na degradação da biomassa lignocelulósica e que não possuem ainda função definida, foram selecionadas as proteínas identificadas nas amostras de palha de cana e bagaço de cana que apresentam peptídeo sinal, sendo assim provavelmente localizadas e atuantes no meio extracelular. Assim, das proteínas totais identificadas, 25 possuem peptídeo sinal preditos (proteínas destacas na tabela 2 pelo símbolo ‡), e algumas delas apresentam domínios anotados que podem fornecer indícios de função.

Entre as proteínas que apresentam peptídeo sinal e não são caracterizadas temos: proteína codificada pelo gene *cthe_0637*, possui um domínio denominado WXG100, que está envolvido em outras bactérias como fator virulento e imunidade; proteína codificada pelo gene

cthe_2379, apresenta o domínio DUF4446, com função desconhecida; proteína codificada pelo gene c*the_2423*, possui um domínio CBM3 e um domínio denominado von Willebrand factor (vWF) tipo A ; proteína codificada pelo gene *cthe_1104 (prepilin type cleavage/methilation)* possui um domínio denominado PulG (pseudopilina), que está caracterizado como presente em proteínas envolvidas na via secretória; proteína codificada pelo gene *cthe_2383*, contém um domínio cobre amino oxidase; proteína putativa ABC transportador codificada pelo gene *cthe_2708*; e proteína codificada pelo gene *cthe_1368*, contém domínio S de membrana. A última proteína já foi descrita em outros trabalhos utilizando proteômica, mas ainda não caracterizada [101].

3.3. Enzimas carboidrato ativas

Enzimas envolvidas na degradação ou modificação da fonte de carbono foram identificadas totalizando 15 proteínas. Essas proteínas foram localizadas tanto nas frações CPPs (celulossoma parcialmente purificado), quanto nas frações FES, e fração FS. A fração FES foi mais enriquecida por essas proteínas. Na tabela 3 é possível visualizar todas as CAZy proteínas e os substratos e frações identificadas.

Entre as proteínas carboidrato ativas identificadas, apenas quatro estão na fração contendo bagaço de cana, são essas: CelK (GH9), CbhA (GH9), CelS (GH49), e Cthe_0798 (CE3), as três primeiras apresentam domínios GHs com atividade de exo-1,4-β-glucanase e a última apresenta um domínio da família CE3 (carboidrato esterase 3), sendo ativa nos grupos acetil que decoram o polímero de xilana [102,103]. As outras 11 proteínas foram encontradas apenas em amostras de palha de cana, sendo 8 enzimas com atividade de endo-1,4-β-glucanase, CelN (GH3), CelA (GH8), CelB (GH5), CelR (GH9), CelJ (GH9, GH44), CelQ (GH9), CelW (GH9), e CelG (GH5), xiloglucanase, Xgh74A (GH74), duas endo-1,4-β-xilanases, XynZ (GH10, CE1) e XynC (GH10), sendo que a primeira delas é bifuncional, apresentando atividade de feruloil esterase devido à presença do domínio CE1 [68,73,104–111]. .

Uma proteína contendo domínio CBM3 (*cthe_2423*), e não apresentando tanto anotação de domínio catalítico quanto doquerina tipo I, portanto, não sendo celulossomal, foi encontrada no preparo de bagaço de cana. O gene que codifica essa proteína foi encontrada anteriormente em trabalho de transcritoma, com expressão aumentada, quando a bactéria está sob estresse na presença de etanol [112]. As proteínas estruturais CipA e OlpB foram encontradas em ambos

os preparos, bagaço e palha. A CipA é a proteína estrutural principal que forma o celulossoma, contendo nove coesinas do tipo I, um domínio CBM3, e uma doquerina do tipo 2, que fará com que essa proteína se ligue a outras proteínas estruturais [54]. A OlpB enquanto isso, é uma proteína estrutural de membrana, que possui sete coesinas do tipo II, onde se ligam sete CipAs, formando um policelulossoma [56].

Tabela 3. Proteínas carboiodrato ativas e pertencentes ao celulossoma, identificada nas amostras de palha e bagaço. Na tabela estão descritos, o gene correspondente, substrato da fração correspondente, domínio anotados e atividade da enzima.

Gene	Protein	Substrato	Domínios	Atividade
Cthe_0043‡	CelN	palha	GH3, CBM3, DocI	Endo-1,4-β-glucanase
Cthe_0269‡	CelA	palha	GH8, DocI	Endo-1,4-β-glucanase
Cthe_0412‡	Glycoside hydrolase family 9/CelK	bagaço/palha	GH9, CBM4, DocI	Exo-1,4-β-glucanase
Cthe_0413‡	CbhA	bagaço/palha	GH9, CBM4, CBD3b DocI	Exo-1,4-β-glucanase
Cthe_0536‡	CelB	palha	GH5, DocI	Endo-1,4-β-glucanase
Cthe_0578‡	CelR	palha	GH9, CBM3, DocI	Endo-1,4-β-glucanase
Cthe_0624‡	CelJ	palha	GH9, GH44, CBM30, DocI	Endo-1,4-β-glucanase
Cthe_0625‡	CelQ	palha	GH9, CBM3, DocI	Endo-1,4-β-glucanase
Cthe_0745‡	CelW	palha	GH9, CBM3, DocI	Endo-1,4-β-glucanase
Cthe_0798‡	Cellulosome protein dockerin type I	Palha	CE3, CE3, DocI	Acetylxylan esterase
Cthe_1398‡	Xgh74A	palha	GH74, DocI	Xyloglucanase
Cthe_1838‡	XynC	palha	GH10, CBM22, DocI	Endo-1,4-β-xilanase
Cthe_1963‡	Xyn Z	palha	GH10, CE1, CBM6, DocI	Endo-1,4-β-xilanase/ feruloil esterase
Cthe_2089	CelS	bagaço/palha	GH49, DocI	Exo-1,4-β-glucanase
Cthe_2423‡	Type 3a cellulose-binding domain protein	bagaço	CBM3	
Cthe_2872‡	CelG	palha	GH5, DocI	Endo-1,4-β-glucanase
Cthe_3077	CipA	bagaço/palha	CoheI (9), CBM3, DocII	Estrutural
Cthe_3078	OlpB	bagaço/palha	CoheII (7)	Estrutural

3.4. Comparação entre as proteínas identificadas nos meios contendo biomassas e celulose

O maior número de proteínas identificadas expressas por genes únicos, foram obtidas para as amostras provenientes do meio de celulose (Figura 1). Das proteínas exclusivas encontradas em celulose, proteínas anotadas como cobre amino oxidases chamaram a atenção por apresentarem peptídeo sinal, e um domínio de cobre oxidase não caracterizado (Tabela4). São elas: Cthe_3227, Cthe_3226, Cthe_2383, Cthe_1778, e Cthe_2383, sendo a última encontrada também na amostra FES de celulose e palha de cana. Além dos domínios Pfam,

baseado nas anotações do GO, encontradas na tabela complementar, a primeira oxidase, Cthe_1778 possui anotação como contendo atividade de hidrolase.

Tabela 4. Proteínas cobre amino oxidases identificadas, e as anotações do Pfam e GO (Gene Ontology). Na tabela está indicado o gene correspondente, as frações que as mesmas foram identificadas, tamanho da proteína (aa), acesso do Genbank e presença de peptídeo sinal.

Gene	Proteína	Substrato/ Fração	Tamanho (aa)	Pfam	GO	Genbank (n° acesso)	Peptídeo sinal
Cthe_1778	Copper amine oxidase like protein	Cellulose/FS	319	PF07833 (3)	GO:0016787, GO:0008152	ABN52999.1	Sim
Cthe_3226	Copper amine oxidase domain protein	Cellulose/FS	262	PF07833	-	ABN54421.1	Sim
Cthe_3227	Copper amine oxidase domain protein	Cellulose/FS	267	PF07833	-	ABN54422.1	Sim
Cthe_3228	Copper Amine oxidase like protein	Cellulose/FS	266	PF07833	-	ABN54423.1	Sim
Cthe_2383	Copper amine oxidase domain protein	Cellulose/FS Cellulose/Straw/FES	304	PF07833	-	ABN53585.1	Sim

Enquanto isso, a comparação de proteínas carboidrato ativas encontradas nas três condições, bagaço e palha de cana, e celulose, mostra também uma maioria dessas proteínas identificadas nos preparos de celulose, todas detalhadamente descritas em estudo anterior, sendo algumas delas encontradas exclusivamente nesse substrato [42]. Em contrapartida, uma proteína celulossomal, a acetil xilana esterase, codificada pelo gene *cthe_0798*, citada no tópico anterior foi identificada exclusivamente nas amostras de palha e bagaço.

A comparação de proteínas identificadas no preparo do celulossoma parcialmente purificado (CPP) nas três condições, celulose, palha e bagaço de cana, foi realizada utilizando uma análise de correlação através do banco de dados nomeado STRING (Figura 3), com o objetivo de identificar a relação entre as proteínas identificadas, visualizando assim proteínas relacionadas na formação do celulossoma, e proteínas não caracterizadas ou não relacionadas à essa função de degradação de carboidrato. Utilizando esse banco é possível realizar uma predição de associação existente entre proteínas, baseado em dados provenientes dos variados bancos de dados e literatura existentes, consistindo em citações em artigos, proximidade no genoma, e até co-expressão verificada experimentalmente. A análise revelou uma maior quantidade de proteínas celulossomais identificadas em celulose, apesar das amostras terem sido preparadas utilizando a mesma metodologia, cromatografia de exclusão por tamanho. Nos preparos de bagaço e palha de cana, foram identificadas mais do que uma proteína não

celulossomal, como, proteínas flagelina (Cthe_2236, Cthe_2237), duas proteínas envolvidas em processos de oxido-redução e defesa da bactéria, peroxiredoxina (Cthe_1965) e bacteriocina M18 (Cthe_1504). Além disso, foram identificadas o transportador ABC CbpB (Cthe_1020), e outras duas proteínas não caracterizadas (Cthe_1503), e outra proteína de ligação ao soluto extracelular (Cthe_2348). A proteína CbpB, foi encontrada também em CPP de celulose, a única não celulossomal, e ainda foi identificada em todas as frações dos preparos de todos os substratos.

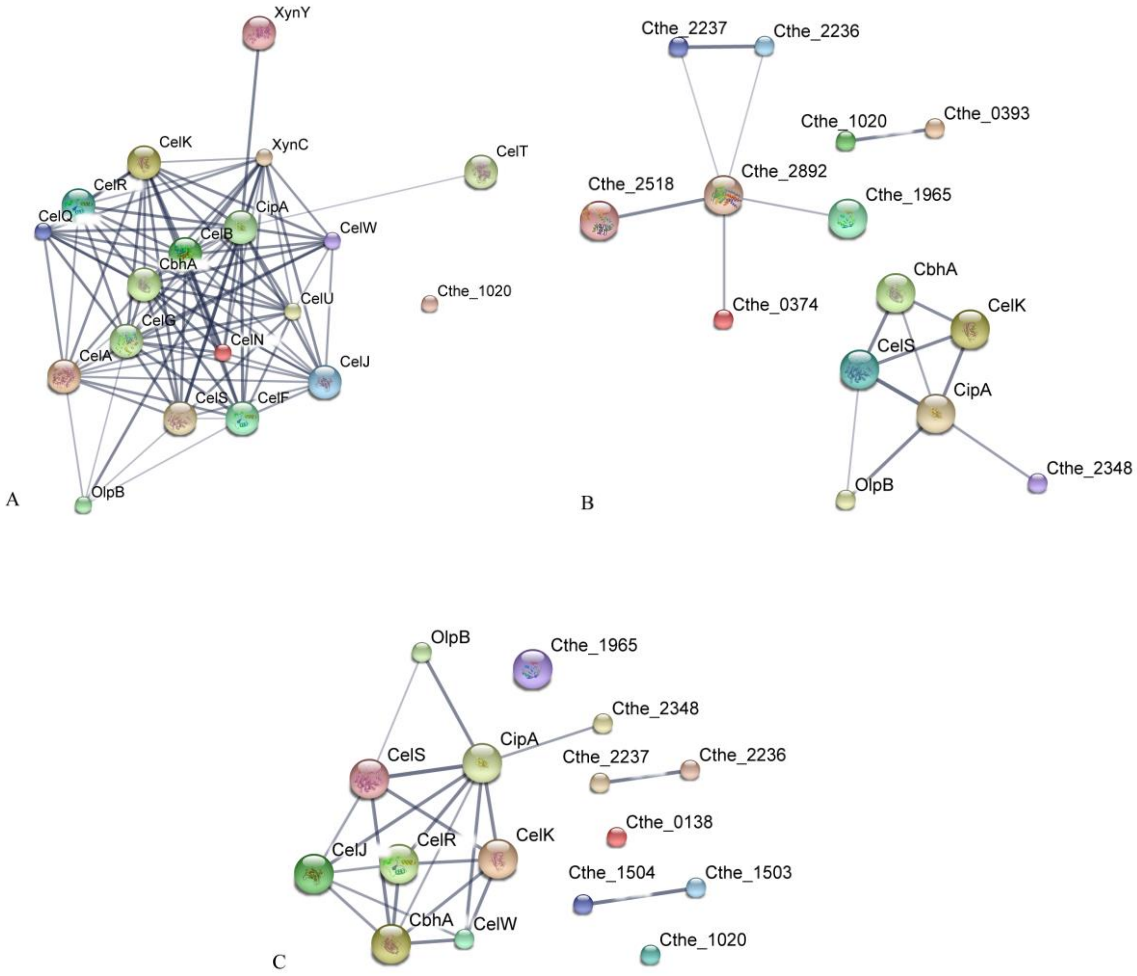


Figura 3. Análise utilizando o programa STRING das proteínas identificadas nas amostras CPP dos substratos celulose (A), e bagaço de cana (B) e palha de cana (C)

4. Discussão

A análise proteômica de *Clostridium thermocellum* quando em crescimento em diferentes substratos foi realizada com o objetivo de identificar quais proteínas carboidrato

ativas, e outras proteínas variadas estariam presentes de acordo com a presença de substratos originados de resíduos de industrias. Nossos resultados mostraram esta análise com alguns substratos complexos não utilizados anteriormente (cana-de-palha e bagaço de cana-de-açúcar), com a concomitante identificação de proteínas envolvidas em vários processos metabólicos e também carboidrato ativas. Em primeiro lugar, os três tipos de amostras de proteínas foram separados com base na afinidade das proteínas pelo substrato (FES), proteínas presentes no celulossoma (CPP) e proteínas livres no meio (FS). Foram encontradas em sua grande maioria proteínas pertencentes à fração citoplasmática, o que poderia significar a lise de células no processo de preparação das amostras, ou a lise natural ocorrida durante o crescimento de *C.thermocellum*, devido ao tempo relativamente longo no cultivo na bactéria (4 dias).

Como foi detectada atividades de CMCase e xilanase em todas as frações preparadas, e não foram encontradas enzimas com essa atividade em todas as frações, foi presumido que no preparo das amostras para a espectrometria de massas, durante principalmente as etapas de limpeza, foram perdidas proteínas no processo. Um pigmento proveniente das biomassas de cana presente em todas as frações quando utilizado esse substrato, podem possivelmente ter interferido no preparo da amostra, prejudicando essa etapa. Dessa forma, as proteínas aqui detectadas são possivelmente de maior abundância nas específicas frações. Além disso, trabalhos anteriores que utilizaram a mesma técnica no preparo da amostra, porém em celulose, obtiveram um maior número de glicosil hidrolases entre as proteínas detectadas [42]. Uma das hipóteses para o menor número de glicosil hidrolases detectadas nas frações CPP e FEC, é a possível diferença na ligação do celulossoma e das proteínas que contem CBM, nos substratos complexos, podendo apresentar uma diferente aderência, desde que esse tipo de ligação não foi ainda caracterizado nesses substratos. Como o número de proteínas celulossomais identificadas foi ainda menor para bagaço de cana, a própria estrutura física desse substrato impediria a melhor aderência das enzimas.

Proteínas envolvidas na formação do celulossoma foram encontradas nos diferentes preparos, e já é demonstrado, que 20% da produção e energia da bactéria é direcionada a produção de proteínas celulossomais, o que corrobora com a presença das mesmas, possivelmente em abundância, no crescimento de *C.thermocellum* [76].

As enzimas celulossomais encontradas nos preparos de bagaço e palha de cana, demonstram, como já evidenciado anteriormente, uma maior abundância das mesmas entre as proteínas celulossomais possivelmente secretadas pela bactéria nesses meios, o que pode ser diretamente relacionado nesse caso como sendo possivelmente as enzimas mais importantes

para a degradação desses substratos. As enzimas celulossomais CelS e CelK, contendo atividade de exoglucanase, presentes em todos os preparos, já foram anteriormente descritas como as duas proteínas celulossomais mais abundantes, incluindo em trabalhos utilizando gramínea como fonte de carbono [58,113]. Os dados aqui demonstrados e descritos anteriormente corroboram que a abundância dessas enzimas, demonstradas por serem reguladas de acordo com as etapa de crescimento, são essenciais para a degradação de celulose e biomassas [103]. CelS já foi descrita com sua expressão aumentada em avicel em comparação com celobiose [89].

Dentre as proteínas encontradas para as amostras de bagaço e palha, a proteína exclusiva desses preparos, acetil xilana esterase Cthe_0798, apresenta o domínio CE3, sendo ativa na retirada de grupos acetil encontrados nos polímeros de xilana, presente na hemicelulose, esperado, portanto, nessas biomassas. Essa enzima foi identificada em trabalhos anteriores de *C.thermocellum* cultivados em substratos como celobiose, avicel, celulose, e xilana, e por meio da análise por transcritoma, o gene correspondente foi diferencialmente expresso em amostras de ámalo amarelo em comparação a celobiose, corroborando com o dado aqui encontrado, desde que o álamo também é uma biomassa, porém com o conteúdo mais elevado de celulose comparado à palha e bagaço de cana [48,90,114].

A diversidade de enzimas identificadas em palha e bagaço, caracterizadas por conter atividade em polímeros de celulose e hemicelulose, demonstram a importância das mesmas na degradação das biomassas aqui estudadas [68,73,104–111]. Recente estudo utilizando misturas enzimáticas composta de enzimas celulossomais, revelou que a diversidade das enzimas é um ponto chave para melhor degradação de biomassa [115].

Como citado anteriormente, por meio de estudos que demonstraram que *C.thermocellum* é capaz de crescer em biomassas sem o pré-tratamento, ou seja com um alto conteúdo de hemicelulose, e presença de lignina, questionamentos foram levantados sobre a possível existência de enzimas, ainda não anotadas ou caracterizadas, que possam ter atividade sobre a lignina. Herring e colaboradores investigaram o crescimento de *C.thermocellum* em gramíneas não pré-tratadas e evidenciaram a liberação de ácido cumárico no meio, componente encontrado na lignina, porém não identificaram qual enzima teria esse papel [116]. Assim, proteínas não caracterizadas identificadas nesse trabalho podem ser putativas a essa atividade ainda desconhecida, como é o caso das proteínas que contém o domínio de cobre amino oxidases, além de apresentarem peptídeo sinal, o que indica presença das mesmas no meio extracelular. Curiosamente, OlpC (*cthe_0452*), uma proteína estrutural celulossomal, e uma

quitinase não celulossomal (*cthe_2895*), não caracterizada, apresentam o mesmo domínio adicional de cobre oxidase [117]. Assim, as proteínas Cobre amino oxidases identificadas nesse trabalho poderiam ter alguma função relacionada à degradação ou ancoragem, podendo possivelmente apresentar um domínio adicional, ainda não caracterizado. Porém, experimentos futuros devem ser realizados para elucidar a função dessa e outras proteínas putativas.

Proteínas identificadas e também pouco caracterizadas, participam de outros mecanismos além da degradação da biomassa, tais como na aderência da bactéria à fonte de carbono. Uma delas é a proteína putativa contendo camada S (Cthe_1368), recentemente sugerida por estar envolvida na formação de biofilme, quando encontrada com expressão aumentada quando a bactéria produz biofilme em celulose [92]. A presença dessa proteína sugere a possível formação de biofilme quando *C.thermocellum* cresce em palha de cana e celulose.

Portanto, *C.thermocellum* apresenta proteínas não caracterizadas que podem ter uma ação importante durante o crescimento dessa bactéria em biomassa lignocelulósica, além de apresentar mecanismos pouco estudados sobre aderência da bactéria nesses substratos. As proteínas aqui identificadas podem ser o começo de possíveis pontos a serem melhor estudados e que apresentam relevância no contexto de solubilização de carboidratos complexos.

Capítulo 3. Análise de transcritoma de *C.thermocellum* em diferentes meios de cultivo utilizando resíduos de cana-de-açúcar e celulose

1. Introdução

Clostridium thermocellum é uma bactéria anaeróbica que possui habilidade para degradar celulose e outras fontes de carbono de origem vegetal como, resíduos de algodão, bagaço de cana-de-açúcar, gramínea e fontes lenhosas como, populus e álamo amarelo [46,47]. Este microorganismo possui esta capacidade de degradação devido à expressão de um complexo multienzimático chamado celulossoma, composto por uma variedade de enzimas hidrolíticas [49,51]. O celulossoma é constituído essencialmente por uma proteína estrutural (CipA) que possui um domínio para a ligação de celulose (CBM) e nove enzimas que podem ser combinadas de um total de aproximadamente 70 disponíveis no genoma de *C.thermocellum* [52].

Apesar da existência de enzimas caracterizadas de *C. thermocellum*, envolvidas na degradação de carboidrato, entre essas, endoglucanases, celobiohidrolases, xilanases e mananases, uma porcentagem de aproximadamente 10-15% não possuem função definida [48,67]. Além disso, a habilidade de *C.thermocellum* liberar unidades fenólicas, constituintes da lignina, durante crescimento em biomassas, continua sendo um mistério, uma vez que não foi ainda descoberta a enzima responsável por essa degradação [116]. O estudo dessa bactéria em biomassa lignocelulósicas e, a identificação de genes expressos nessas condições poderiam fornecer uma indicação de enzimas ainda não caracterizadas atuantes na degradação dessas fontes de carbono.

Estudos anteriores discutem a regulação da expressão de enzimas que irão compor o celulossoma sendo por parte liderada pela fonte de carbono disponível no meio de cultura [90,91]. O mecanismo pelo qual essa fonte de carbono poderá influenciar essa expressão foi identificada como sendo mediada por fatores sigma alternativos [80,81]. Estes fatores sigma, que atuam em pares (sigI-rsgI), são proteínas que respondem à presença de carboidratos no ambiente externo (RsgI) e ativam a expressão (SigI) de genes com papel na hidrólise de carboidratos [80–82]. Contudo, apenas alguns pares foram devidamente caracterizados e, pouco

ainda se sabe sobre os genes que esses fatores sigma de fato regulam no genoma de *C. thermocellum.*

Posteriormente, com os avanços da tecnologia de sequenciamento massal, foram realizados trabalhos baseados na análise transcriptômica (RNAseq) de *C.thermocellum* em cultivo contendo substratos complexos como populus e poplar amarelo pré-tratados, avaliando a regulação gênica em diferentes condições [38,48,58,118]. No entanto, apenas dois trabalhos, disponíveis na literatura até o momento, baseados nessa análise, descrevem a expressão diferencial de genes devido à mudança de substrato disponível no crescimento de *C.thermocellum* [38,48].

Além dos mecanismos de regulação da expressão dos genes envolvidos no metabolismo de carboidrato em *C.thermocellum*, poucos dados foram descritos sobre outros mecanismos de ação da bactéria para aumento de sua eficiência na degradação de biomassa [48,118]. Genes envolvidos em vias metabólicas relacionados à motilidade da bactéria, mecanismos *quorum sensing*, e transdução de sinal, possuem funções que facilitam a aproximação da bactéria ao substrato, assim como na sinalização de moléculas presentes no meio externo da célula [119–121]. Alguns desses genes têm sido descritos como sendo regulados na mudança da fonte de carbono, porém pouco ainda tem sido discutido sobre tais mecanismos [119,120].

C.thermocellum é capaz de degradar fontes de carbono para gerar monômeros de açúcares C6 e C5, no entanto, apenas açúcares C6 são utilizados para serem fermentados em etanol, acetato e lactato [122]. Além disso, *C.thermocellum* não é conhecido como um bom organismo fermentativo, portanto, com o objetivo da melhoria de linhagens, engenharia metabólica, e uso de co-culturas têm sido realizados para superar a baixa produção de produtos gerados, de alto valor agregado, por esse microrganismo [39,123]. Organismos que possuam a capacidade de fermentar açúcares C5, e, portanto, contendo habilidades complementares ás de *C.thermocellum*, são de grande interesse. Nesse sentido, a bactéria *Moorella thermoacetica*, é um organismo capaz de degradar ambos os açúcares, C5 e C6 [124]. *M.thermoacetica* é uma bactéria acetogênica, gram positiva, anaeróbica e não celulolítica [124]. Trabalhos de co-cultura dessa bactéria com *C.thermocellum,* com o objetivo de aumentar a produção de fermentação têm sido descritos [125–127].

Este presente trabalho apresenta, pela primeira vez, uma análise transcriptômica de *C.thermocellum*, utilizando resíduos de cana-de-açúcar, sem pré-tratamento, como fonte de carbono, suplementando o meio de cultura, visando uma análise dos genes e vias metabólicas

que podem ser essenciais para o crescimento da bactéria na biomassa vegetal, quando comparados com celulose. Além disso, este estudo traz informações preliminares sobre a regulação de genes de *M. thermoacetica* quando cultivada na presença de *C.thermocellum* em co-cultura, na presença dessas biomassas, com o provável favorecimento de ambos microorganismos.

2. Material e métodos

2.1. Condições de cultivo

A cultura de *C.thermocellum* foi estabelecida em meio de cultura contendo celulose e o crescimento acompanhado pelo método YAS (Yellow affinity substrate), com base na formação do pigmento amarelo [33]. O crescimento bacteriano foi realizado em frascos de 100 mL sob condições anaeróbicas utilizando meio de cultura líquido redutor preparado como descrito no Capítulo 2, seção Materiais e métodos desse trabalho. O meio foi suplementado com 1g de bagaço de cana, palha de cana e celulose. Um pré-inóculo em celulose (OD600 aproximadamente 0.9) foi utilizado para inocular os meios em triplicatas biológicas e realizada uma curva de crescimento, acompanhando a OD a 600 nm. Os meios foram incubados a 60°C.

2.2. Isolamento de RNA e Sequenciamento

Para o isolamento de RNA, foram utilizadas culturas crescidas em meio contendo celulose, palha ou bagaço de cana-de-açúcar (100 mL), a 60 °C, durante 37 horas e cada tratamento em triplicata biológica. Posteriormente, estas culturas (triplicatas biológicas) foram submetidas a extração de RNA total usando Reagente TRizol® (Invitrogen), de acordo com as instruções do fabricante. Porém, uma etapa adicional foi realizada anteriormente ao uso do Reagente TRizol®, 100 mL de cultura foram centrifugadas a 7000 rpm por 5 min, ressuspendidas em tampão TE suplementado com lisozima 10 mg/mL e incubadas durante 1 hora a 37 °C. Para melhorar a qualidade do RNA extraído, foi utilizado o kit de purificação de RNA RiboPureTM (Ambion, Life Technologies, Inc.USA), de acordo com as instruções do fabricante. A qualidade do RNA foi avaliada em gel de agarose (1%) e a integridade e quantificação utilizando o sistema do Agilent 2100 Bioanalyzer e RNA LabChip® kit (Agilent Technologies), conforme instruções do fabricante. A quantidade foi expressa em pg/µl e a qualidade, pelo número de integridade do

RNA (NIR) (Apêndice C). Após certificada a qualidade, as amostras de RNA foram enviadas para a empresa Eurofins Genomics (EUA) que procedeu à construção das bibliotecas e seu sequenciamento no equipamento Hiseq2000 (Illumina).

2.3. Análise RNAseq

Primeiramente, a partir dos *reads* gerados pelo sequenciamento, foi realizada uma limpeza dos dados para retirada dos *reads* correspondentes ao RNA ribossômico (rRNA), utilizando o software SortMeRNA v1.9 [36] e adaptadores provenientes do sequenciamento Illumina pelo programa cutadapt v1.2.1 [37]. Em seguida, o corte de qualidade foi realizado por meio do programa PRINSEQ Lite v0.20.0 [38], utilizando o tamanho de sequência mínimo de 30 pb e uma qualidade mínima de 30 (Phred < 30) em ambas as extremidades da leitura e como qualidade média. Todas as leituras com caracteres não-IUPAC foram descartadas, pois todas as leituras continham mais de três Ns.

Os *reads* oriundos do sequenciamento foram mapeados no genoma de *Clostridium thermocellum*, número de acesso, GenBank: CP000568.1 e *Moorela thermoacetica*, número de acesso, GenBank: CP000232.1 usando o software Rockhopper [39] com configurações de parâmetros padrão para genomas de procariotos. O Rockhopper normaliza as contagens dos fragmentos utilizando quartil superior e calcula valores de expressão semelhantes ao RPKM (Reads per Kilobase per Million mapped reads),porém utilizando o quartil superior ao invés do somatório dos fragmentos [128]. A análise de expressão diferencial foi realizada usando o procedimento Benjamini-Hochberg, e o valor q gerado, se ≤ 0.05, foi considerado estatisticamente significativo. A predição de operons é realizada utilizando uma probabilidade de operons descrita anteriormente, e analisando a distância entre os genes e a co-expressão [129].

2.4. Anotações funcionais de genes diferencialmente expressos

Os genes foram anotados utilizando informações existentes nos bancos de dados Uniprot e JGI. Desses bancos foram retiradas as informações da categorização de genes em COGs (Clusters of Orthologous Groups), e KEGG (Kyoto encyclopedia of Genes and Genomes) [97,130]. Informações adicionais, como domínios em proteínas putativas foram anotadas utilizando o banco de dados Pfam a partir da sequência de aminoácidos da proteína expressa pelo gene de

interesse [96]. O banco de dados CAZy foi utilizado para validar as informações de domínios existentes relacionados a proteínas ativas no carboidrato [61]. Agrupamento hierárquico dos valores de expressão dos genes foi gerado utilizando o software JMP genomics, por meio do método de Ward [131]. A tabela de proteínas ortólogas entre *Clostridium thermocellum* e *Moorella thermoacetica* foi gerada utilizando o software OrthoMCL.

3. Resultados

No presente estudo, o sequenciamento de RNA foi utilizado como ferramenta para identificar a expressão diferencial gênica de *Clostridium thermocellum* B8 quando cultivada em diferentes condições. Com esse objetivo, o meio de cultura foi então modificado devido ao tipo de substrato disponível para o crescimento das bactérias. A celulose, tratada como um controle neste caso, foi comparada aos resíduos não tratados da indústria da cana-de-açúcar, palha e bagaço. Dessa forma, os principais genes destacados nesse trabalho estão diretamente relacionados à desconstrução de biomassa ou indiretamente relacionados, quando atuantes na promoção ou aumento da capacidade das bactérias para realizar esta degradação.

Primeiro, foi realizada a curva de crescimento com a cultura de *C.thermocellum* (Fig. 1), e foi possível visualizar todos os estágios de desenvolvimento de bactérias, fase exponencial e estacionária. Após aproximadamente 48 horas, nas culturas contendo bagaço / palha de cana-de-açúcar (A, C) e celulose (B), a bactéria já estava em fase estacionária. Portanto, às 37 horas de crescimento, a cultura foi encontrada em fase exponencial, quando provavelmente o metabolismo dos microorganismos era mais ativo.

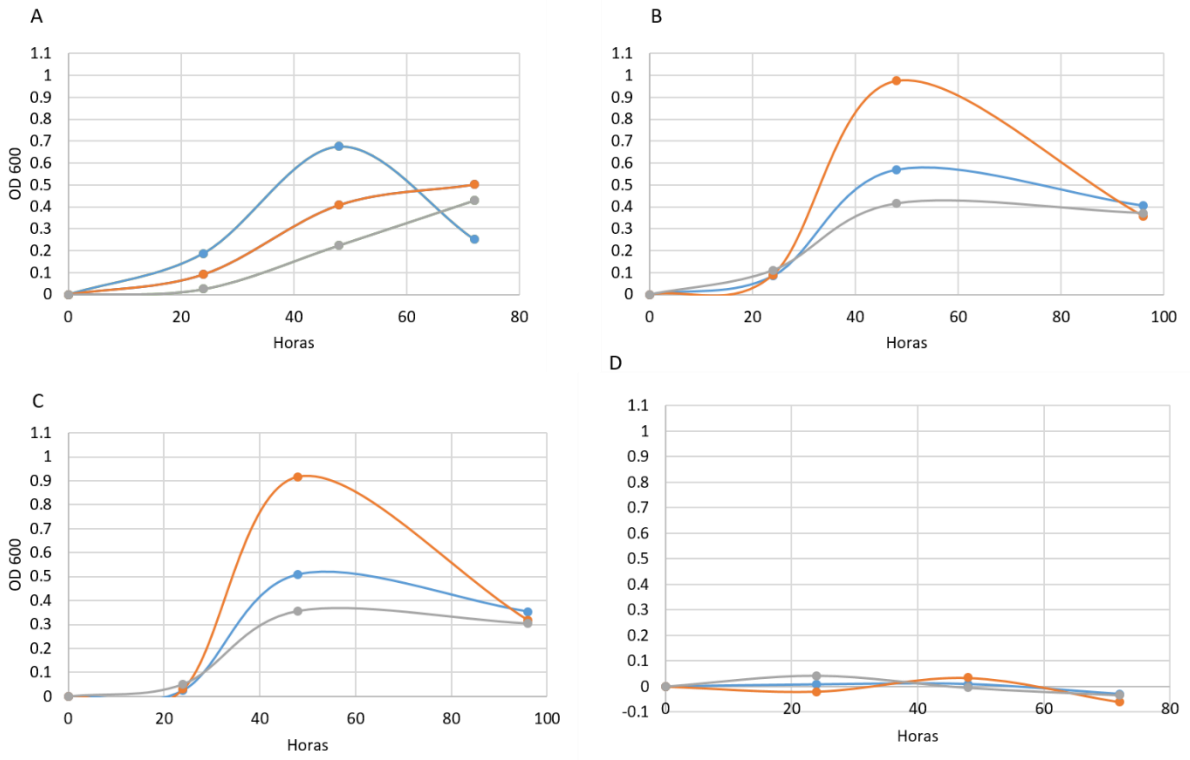


Figura 1. Cinética do crescimento de *Clostridium thermocellum* B8 em meio anaeróbico suplementado com bagaço de cana-de-açúcar (A), celulose (B) e palha de cana-de-açúcar (C). O gráfico D representa o meio com ausência de fonte de carbono. As culturas foram incubadas a 60 ° C até 96 horas. Foram tomadas alíquotas e absorbância medida a 600 nm.

Durante as análises iniciais do transcritoma utilizando o software Rockhopper, ao analisar sequências de r16S nos dados gerados, foi detectada a presença de RNA proveniente de *M.thermoacetica*. Alinhando os *reads* no genoma disponível no banco de dados, foi obtida uma alta cobertura do genoma, e dessa forma confirmada a presença dessa bactéria em co-cultura com *C.thermocellum*.

Os *reads* provenientes do RNAseq produziram uma cobertura de transcritoma para *C.thermocellum* e *M. thermoacetica* de 81x a 700x e 174x a 1400x, respectivamente (Apêndice D), gerando um total de 2878 e 2239 de genes que codificam proteína, representando 90% do genoma de *Clostridium thermocellum* ATCC27495 e 97 % do genoma de *Moorella thermoacetica* ATCC39073. Novos RNAs foram preditos resultando em 422 e 583 sequências respectivamente. Os dados de correlação usando valores de expressão das triplicatas biológicas de cada condição foram avaliados pela análise de componentes principais (PCA) (Apêndice E).

Além da predição de novos RNAs, uma predição de putativos operons foi realizada, gerando um total de 619 e 529 no genoma de *C.thermocellum* e *M.thermoacetica*, respectivamente.

Uma visão geral usando valores de expressão em *C.thermocellum* foi analisada primeiramente por meio de agrupamento hierárquico baseado na semelhança dos perfis de expressão, incorporando os genes em 10 *clusters* distintos (Fig. 2). Em geral, C1 continha genes com baixa expressão para os três substratos, seguidos de C2. Clusters C3-C10 continham genes com expressão abundante, constituídos por 181 genes. No entanto, os valores de expressão foram maiores para celulose, com um máximo de expressão de 20.665, em comparação com bagaço e palha, 2.251 e 4.184, respectivamente. Considerando os 10% dos genes com maiores expressões desses 181 genes, no *cluster* 10 e *cluster* 8 foram encontrados genes que se apresentam com alta expressão em todos os substratos, como os genes *cthe_1020, cthe_2103, cthe_2348* e *cthe_2401*. *Cthe_1020* codifica um transportador ABC, CbpB, *Cthe_2103*, uma ferredoxina, *Cthe_2348*, uma proteína de camada S, não caracterizada e *cthe_2401*, desconhecida.

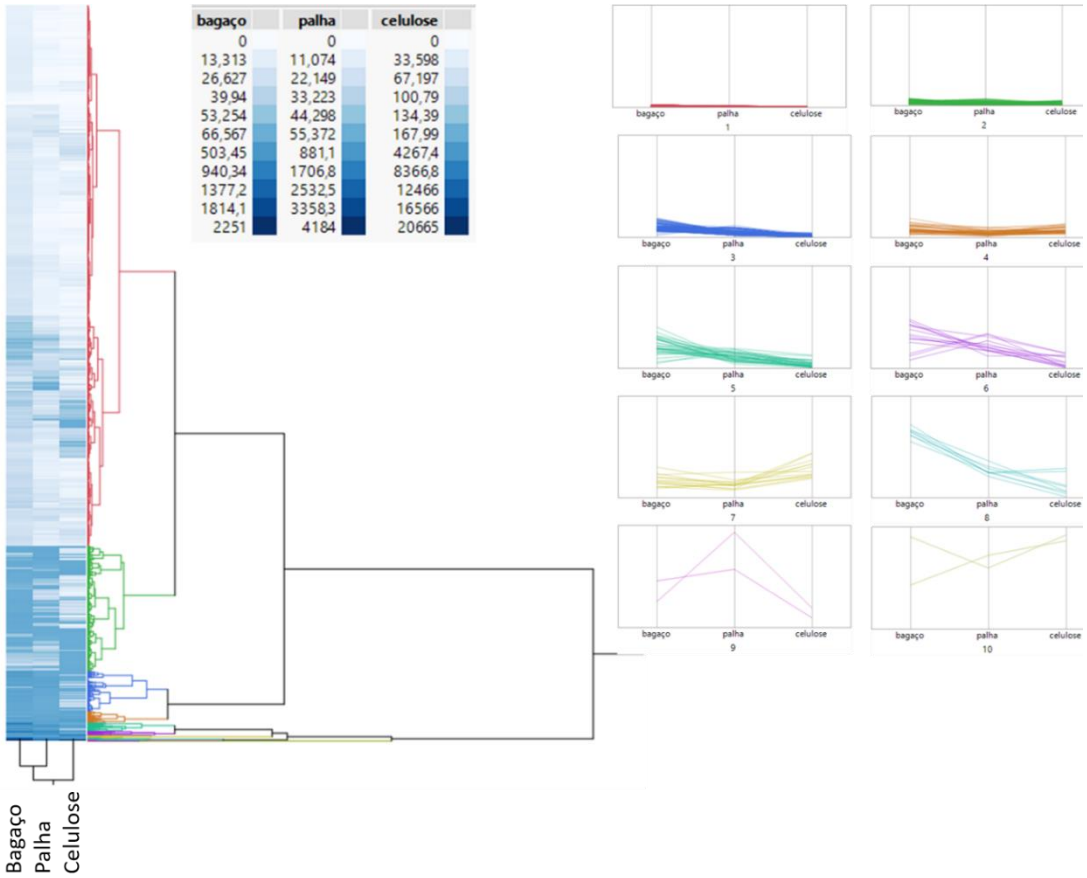


Figura 2.. Agrupamento hierárquico dos valores de expressão dos genes de *C.thermocellum* nas condições de crescimento em cellulose, palha e bagaço de cana. JMP Genomics 6 software foi utilizado para elaboração do gráfico, gerando um total de 10 clusters.

Entre os principais genes com altas expressões em condições de bagaço e palha de cana, presentes no Cluster 8, estão *cthe_0326, cthe_0785, cthe_1965, cthe_3125 e cthe_3348*, e no

Cluster 4, genes *cthe_1104* e *cthe_3383*. Os últimos genes, *cthe_3383* e *cthe_3385* codificam uma enzima putativa AgrD, presente na via de sinalização *quorum sensing*. Os outros genes citados não foram na maioria caracterizados, com exceção de uma proteína de choque térmico Hsp20 (*cthe_3125*) e uma peroxiredoxina (*cthe_1965*).

Outros genes abundantes em bagaço ou palha foram, *cthe_0063*, *cthe_0394*, *cthe_1348*, *cthe_1176*, *cthe_2617* e *cthe_3351*. Dentre esses, estão presentes genes que codificam proteínas que participam de motilidade celular como *cthe_1176* e conversão de energia em etanol (*cthe_0394*).

3.1. Expressão diferencial

Com base na análise estatística (q-value ≤ 0,05), 20 a 22% de genes do genoma de *C.thermocellum* foram diferencialmente expressos entre tratamentos (bagaço x celulose, palha x celulose). Na primeira comparação, bagaço de cana-de-açúcar *versus* celulose, 86 genes foram positivamente regulados e 558 genes negativamente regulados, enquanto isso, na comparação entre palha de cana-de-açúcar *versus* celulose, 40 genes foram positivamente regulados e 657 genes negativamente regulados (Apêndice F (A e B)). Genes positivamente e negativamente regulados das comparações acima estão representados no diagrama de Venn na Fig. 3.

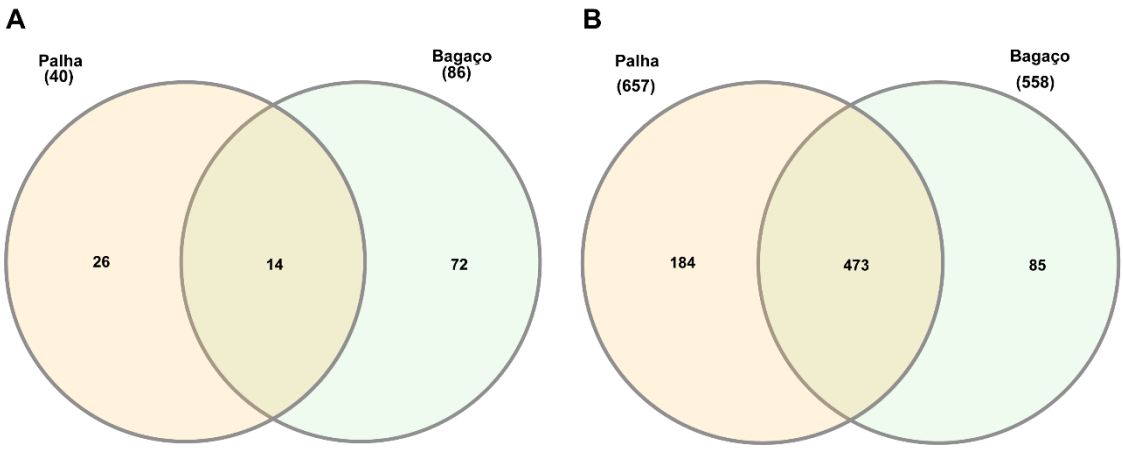


Figura 3. Diagrama de Venn representando o número de genes regulados positivamente (A) e negativamente (B) em bagaço (azul) e palha (rosa) em comparação com celulose.

De acordo com a anotação gerada, baseada em COG, entre os genes significativamente expressos (Fig. 4), a maioria teve expressão diminuída para ambas as condições (Palha (A),

Bagaço (B)) pertencendo a categorias não conhecidas, e pouco caracterizadas (S, R), seguidas da categoria de genes envolvidos em Energia, Produção e conversão (C), Transporte de Aminoácidos e Metabolismo (E), e Tradução, Estrutura ribossômica e Biogênese (J). Por outro lado, a maioria dos genes com expressão aumentada nas mesmas condições também estavam na categoria de pouco caracterizados (S) e envolvidos em funções celulares relacionados à motilidade celular (N), transporte de carboidratos e metabolismo (G) e transdução de sinal (T).

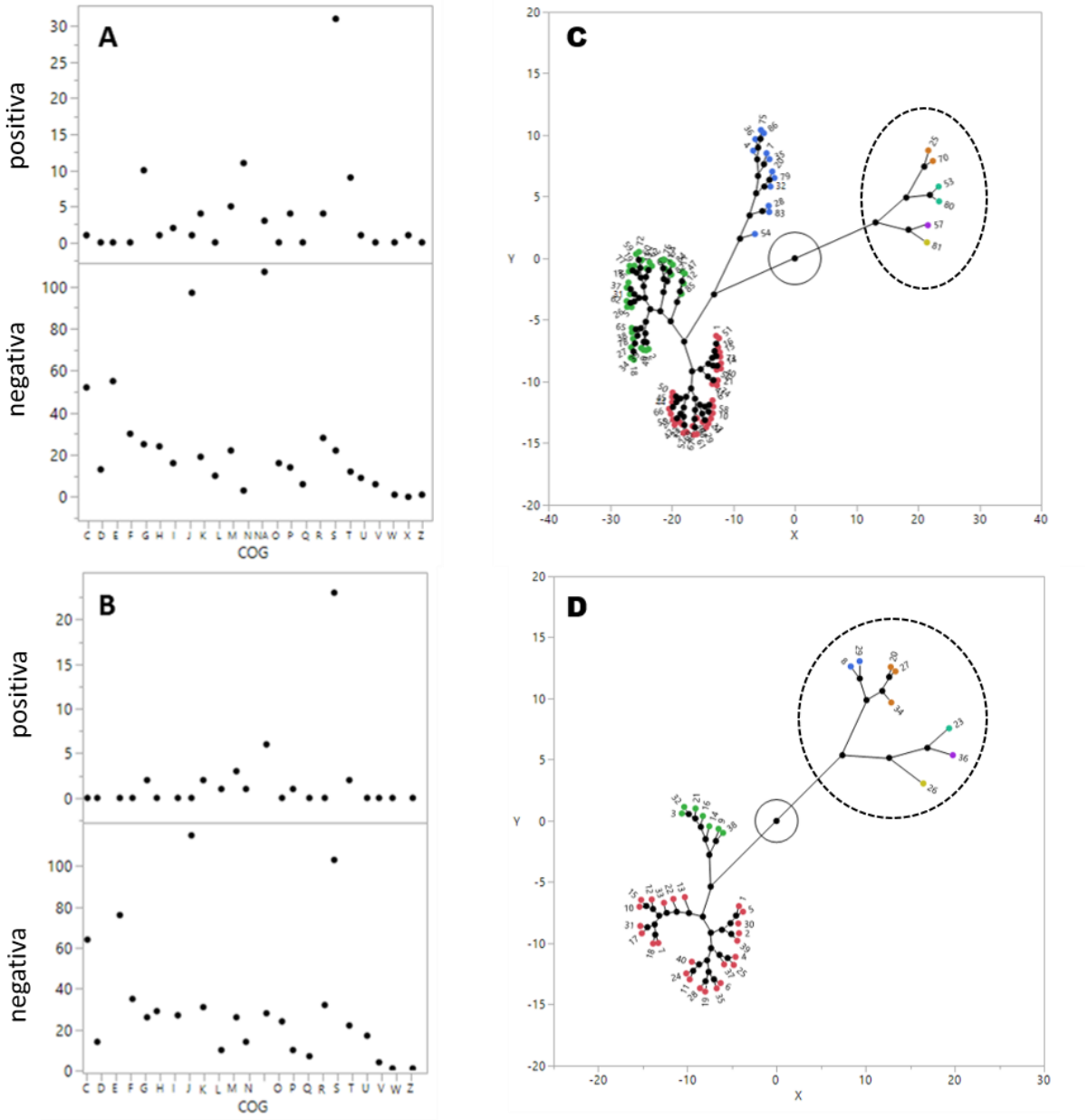


Figure 4. Número de genes diferencialmente expressos em cada categoria COG (Clusters of Orthologous groups) de *C.thermocellum* regulados positivamente e negativamente para as amostras de palha (A), e bagaço (B), em comparação com celulose. Nos painéis C e D estão demonstrados os clusters contendo o agrupamento de genes de acordo com o perfil de expressão. Categorias COG: (C) Produção de energia e conversão; (D) Cromossomo, divisão celular e controle do ciclo celular; (E) Metabolismo e transporte de aminoácido; (F) Metabolismo e transporte de nucleotídeo; (G) Metabolismo e transporte de

carboidrato; (H) Metabolismo e transporte coenzima; (I) Metabolismo e transporte de lipídeo; (J) Biogêneses, tradução e estrutura do ribossomo; (K) Transcrição; (L) Replicação, recombinação, e reparo; (M) Parede celular/ biogêneses membrana; (N) Motilidade da célula; (O) Modificação pós-traducional, turnover proteína, chaperonas; (P) Metabolismo e transporte de íons inorgânicos; (Q) Biossíntese do metabolismo secundário, transporte e catabolismo; (R) Predição de função geral somente; (S) Função desconhecida; (T) Mecanismo de transdução do sinal; (U) Secreção e tráfego intracelular; (V) Mecanismos de defesa.

Os genes diferencialmente expressos, com regulação positiva, foram também analisados por meio de agrupamento hierárquico, incorporando os genes em 7 *clusters* distintos (Fig.4 C e D), onde os clusters C3 a C7 (18 genes), obtiveram valores de expressão mais elevados, baseado em valores de expressão. Dentre esses genes, a maioria possui função associada com motilidade celular, como montagem do pilus (*cthe_0108*, *cthe_0733* e *cthe_2684*) e montagem do flagelo (*cthe_2236*, *cthe*_2237 e *cthe*_2246). Três genes codificam proteínas hipotéticas que apresentam anotações de domínio CBM25 (domínio de ligação ao carboidrato família 25), *cthe_0956*, *cthe_1080* e *cthe_3163*. Além disso, estão presentes proteínas putativas envolvidas no *quorum sensing*, como proteína AgrB (*cthe_1310*). A tabela 1 mostra todos os genes com expressão aumentada em bagaço e palha de cana agrupados de acordo com a função baseada tanto na anotação do KEGG quanto no COG.

Tabela 1. Genes positivamente regulados em bagaço e palha de cana. Na tabela estão demonstrados a regulação em "Fold Change" (FC).

N°	Gene	Descrição	FC bagaço	FC palha
72 genes regulados positivamente em bagaço x celulose				
CARBOIDRATO ATIVAS E RELACIONADAS AO CELULOSSOMA				
1	Cthe_0274	glycoside hydrolase family protein (CelP)‡	4.25	1.4375
2	Cthe_0798	G-D-S-L lipolytic protein‡	22.428	6.857
9	Cthe_0821	coagulation factor 5/8 type-like protein ‡	5.5	2.75
3	Cthe_1257	carbohydrate-binding, CenC-like protein	3.294	1.823
4	Cthe_1398	dockerin type I cellulosome protein (XghA)‡	3.59	2.909
5	Cthe_2807	glycoside hydrolase family protein (CelC)	9.375	3.375
6	Cthe_2972	glycoside hydrolase family protein (XynA)‡	4.228	3
7	Cthe_3163	carbohydrate-binding family 25 protein	3.61	1.9105
8	Cthe_3079	cellulosome anchoring protein cohesin subunit (Orf2p)‡	5.129	1.0967
10	Cthe_2811	glycoside hydrolase family protein (ManA)‡	12.8	5.4
MOTILIDADE DA CÉLULA –quimiotaxia, sistema de dois componentes, quorum sensing				
11	Cthe_0080	CheW protein	5.384	2.653
12	Cthe_0108	type IV pilus assembly PilZ	3.763	0.986
13	Cthe_0475	flagellar protein (FlbD)	2.784	1.117647
14	Cthe_0476	hypothetical protein (FliL)	4.037	1.833
15	Cthe_0479	response regulator receiver protein (CheY)	3	1.222

16	Cthe_0733	type IV pilus assembly PilZ	4.167	1.91
17	Cthe_0808	CheR-type MCP methyltransferase	3.15	1.45
18	Cthe_1065	type IV pilus assembly PilZ	4.656	1.562
19	Cthe_1310	accessory gene regulator B (AgrB)	3.75	1.718
20	Cthe_2217	flagellar protein FliS	5	2.346
21	Cthe_2218	flagellar hook-associated 2-like protein (FliD)	3.304	1.347
22	Cthe_2219	flagellar protein FlaG protein	9.75	4.25
23	Cthe_2236	flagellin-like protein (FliC)	5.128	1.7179
24	Cthe_2237	flagellin-like protein (FliC)	8.253	1.988
25	Cthe_2244	flagellar hook-associated protein FlgK	5.545	1.727
26	Cthe_2245	hypothetical protein (FlgN)	13	3.2
27	Cthe_2425	MotA/TolQ/ExbB proton channel	3.212	1.09
28	Cthe_2426	OmpA/MotB	6.471	2.058
29	Cthe_2620	hypothetical protein (FlgL)	6.444	1.777
30	Cthe_2621	hypothetical protein (FlgL)	4.56	1.72
31	Cthe_3029	CheW protein	3	1.148
		TRANSPORTE DE MEMBRANA		
32	Cthe_0244	heavy metal translocating P-type ATPase	5.578	3.052
33	Cthe_1586	binding-protein-dependent transport systems inner membrane component	18.667	8.666
34	Cthe_2666	P-type HAD superfamily ATPase	5.765	1.647
		METABOLISMO DE AÇÚCAR		
35	Cthe_2562	glucose-1-phosphate cytidylyltransferase	3.684	1.105
		METABOLISMO DE LIPÍDEO		
36	Cthe_0130	3-oxoacyl-(acyl-carrier-protein) synthase	3.304	1.956
37	Cthe_2442	carbohydrate kinase FGGY	4.76	2.04
38	Cthe_1548	enoyl-CoA hydratase/isomerase	6.609	2.347
		BIOSÍNTESE DE PEPTIDEOGLICANO E DEGRADAÇÃO DE PROTEÍNAS		
39	Cthe_1804	cell wall hydrolase/autolysin‡	4.615	1.461
		ENOVELAMENTO, SEPARAÇÃO E DEGRADAÇÃO		
40	Cthe_2174	transcription termination factor Rho	3.68	1.765
		METABOLISMO DE COFATORES E VITAMINAS		
41	Cthe_1839	biotin synthase	3.619	0.714
		FATORES DE TRANSCRIÇÃO		
42	Cthe_0245	ArsR family transcriptional regulator	6.235	1.044
43	Cthe_2441	DeoR family transcriptional regulator	2.916	1.875
44	Cthe_2940	CarD family transcriptional regulator	3.357	1.771
		REGULAÇÃO		
45	Cthe_1805	response regulator receiver modulated diguanylate cyclase	4.13	1.739
46	Cthe_1916	two component transcriptional regulator	3.173	1.043
		ESTRESSE E ESPORULAÇÃO		
47	Cthe_3089	hypothetical protein	3.325	1.918
		OUTROS		
48	Cthe_0079	hypothetical protein	10.286	4.428
49	Cthe_0533	radical SAM family protein	9.125	2.625
50	Cthe_0558	hypothetical protein	12	6
51	Cthe_0603	GCN5-like N-acetyltransferase	4.933	3.2666

52	Cthe_0898	metal dependent phosphohydrolase	6.688	2.937
53	Cthe_1205	putative serine protein kinase, PrkA	7	1.55
54	Cthe_1258	copper amine oxidase-like protein‡	4.323	1.911
55	Cthe_1309	radical SAM family protein	5.982	1.9298
56	Cthe_1463	hypothetical protein	3.742	2.2258
57	Cthe_2247	regulatory protein MerR	4.714	2.4285
58	Cthe_2307	hemerythrin-like metal-binding protein	3.786	1.357
59	Cthe_2506	S-layer-like domain-containing protein	4.313	2.5
60	Cthe_2744	lytic transglycosylase	3.567	1.3
		HIPOTÉTICAS E NÃO-CARACTERIZADAS		
61	Cthe_0303	hypothetical protein	6.833	2.333
62	Cthe_0499	hypothetical protein	14.5	6.5
63	Cthe_0642	hypothetical protein	3.857	1.514
64	Cthe_0704	hypothetical protein	2.827	0.827
65	Cthe_1061	hypothetical protein	2.708	2
66	Cthe_1081	hypothetical protein	5.071	3.428
67	Cthe_1210	hypothetical protein	2.785	2.5357
68	Cthe_1212	hypothetical protein	3.795	1.225
69	Cthe_2175	hypothetical protein	3.539	2.941
70	Cthe_2234	hypothetical protein	5.714	1.714
71	Cthe_3384	hypothetical protein	6.225	1.95
72	Cthe_3385	hypothetical protein	3.126760563	2.042254
		26 genes regulados positivamente em palha x celulose		
		TRANSPORTE DE MEMBRANA		
1	Cthe_3134	hypothetical protein	1.88	3.08
		FATOR SIGMA		
2	Cthe_1272	putative RNA polymerase sigma factor SigI	0.866666667	4
3	Cthe_1273	alpha-L-arabinofuranosidase B	1.625	5.125
		ESPORULAÇÃO		
20	Cthe_2165	hypothetical protein	33	55
21	Cthe_2655	AbrB family transcriptional regulator	2.181818182	4.363636
		OUTROS		
22	Cthe_0209	glycosyltransferase	0.9	5.6
23	Cthe_1334	FHA domain-containing protein	1.625	4.5625
24	Cthe_2402	peptidoglycan-binding LysM	1.9615	5.423
25	Cthe_2617	peptidase M23B	1.924	8.236
26	Cthe_2737	excinuclease ABC subunit C	2.812	3.625
		HIPOTÉTICAS E NÃO-CARACTERIZADAS		
4	Cthe_0084	hypothetical protein	3.138888889	3.055556
5	Cthe_0264	hypothetical protein	2.333333333	4.4
6	Cthe_0453	hypothetical protein	2.424242424	3.727273
7	Cthe_0454	hypothetical protein	2.214285714	3.071429
8	Cthe_0899	hypothetical protein	1.05	3.45
9	Cthe_1779	hypothetical protein	0.767857143	4.785714
10	Cthe_1780	hypothetical protein	1.357142857	4.571429
11	Cthe_2018	hypothetical protein	1.428571429	4.02381

12	Cthe_2019	hypothetical protein	2.333333333	4.555556
13	Cthe_2030	hypothetical protein	1.55	3.35
14	Cthe_2074	hypothetical protein	2.168224299	3.878505
15	Cthe_2094	hypothetical protein	2.354166667	3.604167
16	Cthe_2651	hypothetical protein	1.952830189	3.632075
17	Cthe_2854	hypothetical protein	2.4	4.088889
18	Cthe_2951	hypothetical protein	1.352941176	6.470588
19	Cthe_3337	hypothetical protein	2.107142857	7.428571
		14 genes regulados positivamente em comum em palha e celulose		
		CARBOIDRATO ATIVAS E RELACIONADAS AO CELULOSSOMA		
1	Cthe_0956	hypothetical protein	3.23	5.051282
2	Cthe_1080	carbohydrate-binding family 25 protein	3.3018	3.981132
		MOTILIDADE DA CÉLULA		
3	Cthe_0401	methyl-accepting chemotaxis sensory transducer	6.437	4.125
4	Cthe_2246	anti-sigma-28 factor, FlgM	9.796	5.968
5	Cthe_2684	hypothetical protein (PilZ)	6.08	5.625
		TRANSPORTE DE MEMBRANA		
6	Cthe_1588	extracellular solute-binding protein	24.222	9.333333
7	Cthe_2736	phosphoenolpyruvate--protein phosphotransferase	3.361	4.388
		OUTROS		
8	Cthe_2503	cupin 2 barrel domain-containing protein	4.833	7.055
		ESTRESSE E ESPOPRULAÇÃO		
9	Cthe_2071	hypothetical protein	3.347	3.043478
		HIPOTÉTICAS E NÃO-CARACTERIZADAS		
10	Cthe_0084	hypothetical protein	3.138	3.055556
11	Cthe_3178	hypothetical protein	21.019	16.117
12	Cthe_3194	hypothetical protein	6.083	4.458333
13	Cthe_3328	hypothetical protein	4.807	5.980769
14	Cthe_3345	hypothetical protein	2.888	4.740741

3.2. Genes regulados positivamente em bagaço e palha de cana

3.2.1. Quimiotaxia e transdução de sinal

De acordo com os genes regulados em condições de bagaço, como descrito anteriormente, a mobilidade celular foi a via mais enriquecida, que inclui a quimiotaxia, o sistema de dois componentes e *quorum sensing*, de acordo com a anotação do KEGG, totalizando 25 genes, 29% de todos os genes positivamente regulados. Os genes encontrados pertencentes ao sistema flagelar, estão envolvidos: na montagem dos flagelos, na estrutura de haste, gancho e filamento, fliC (*cthe_2236*, *cthe_2237*), fliD (*cthe_2218*), fliL (*cthe_0476*), flgK (*cthe_2244*), flgL

(*cthe_2620, cthe_2621*), flag (*cthe_2219*) e flbD (*cthe_0475*); motor do flagelo, como motA (*cthe_2425*) e motB (*cthe_2426*); chaperonas , fliS (*cthe_2217*) e flgN (*cthe_2245*); genes reguladores para síntese de genes da flagelina, FlgM (fator anti-sigma-28) (*cthe_2246*) e proteína reguladora MerR (*cthe_2247*), que não possui função determinada. Um diagrama esquemático pode ser visto na Fig. 5a, com os genes envolvidos nessa via.

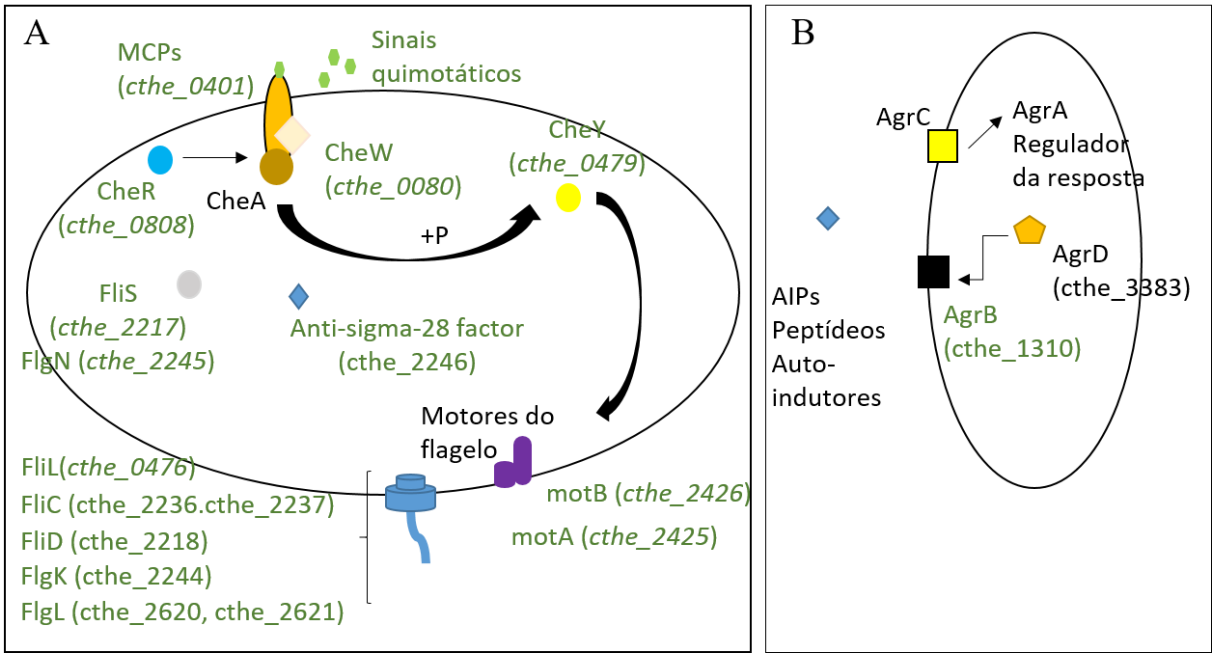


Figura 4. Genes positivamente regulados (destacados em verde) na condição de bagaço, nas vias representadas: montagem do flagelo (ko02040) e sistema de dois-componentes (ko02020) (A); quorum sensing –operon *Agr* (B).

Também participando da mesma via, relacionado à quimiotaxia, estão presentes genes que codificam as proteínas cheR (*cthe_0808*), cheW (*cthe_0080*), cheY (*cthe_0479*) e um transdutor sensorial aceptor-metil MCP (*cthe_0401*). Assim, a sinalização começa com a proteína sensível, MCP (cthe_0401), que ativa todo o sistema flagelar por meio do reconhecimento de diferentes concentrações de solutos no meio extracelular [43]. Após o sinal recebido da MCP, a histidina quinase CheA fosforila a última proteína, CheY, que transmite o sinal para ativar o movimento do flagelo.

Também envolvidos na motilidade da bactéria estão os genes envolvidos na montagem de pilus, especificamente tipo IV da montagem de pilus (pilZ), *cthe_0108*, *cthe_0733*, *cthe_1065*, *cthe_2684*.

Entre os genes regulados na condição de bagaço de cana que estão anotados como envolvidos na transdução de sinal [T], estão presentes genes envolvidos no sistema de dois

componentes, como proteína CheW, e o regulador transcricional de dois componentes (*cthe_1916*), também com função de transcrição [K]. Outros genes com essa mesma anotação, transdução de sinal, não foram, portanto, completamente elucidados, como a proteína receptora do regulador de resposta (*cthe_0479*), AgrB (*cthe_1310*) e proteína de ligação ao metal semelhante à hemeritrina (*cthe_2307*). Proteína AgrB, juntamente com a AgrD, descritos acima como encontrados altamente expressos, atuam no sistema de quorum sensing, que é uma forma de comunicação entre células [46]. Essa via, esquematizada na figura 5b, aponta resumidamente a participação dessa proteína na via de sinalização. Primeiramente, o pentapeptídeo AgrD, considerada uma pré-proteína, se transforma no AIP (peptídeo autoinduzido), que por sua vez reconhece e ativa a proteína quinase AgrC, que fosforila a proteína AgrA, responsável por regular o operon *agr* [46].

Não pertencendo à quimiotaxia, mas com uma regulação oposta, são as proteínas relacionadas à esporulação. Os genes da quimiotaxia são geralmente ativados quando as células estão em fase exponencial, oposta à esporulação, quando geralmente ocorre em fase estacionária, com escassez de nutrientes. Neste caso, encontrou-se apenas um gene regulado positivamente em bagaço, para a esporulação (*cthe_2071*).

3.2.2. Transporte de membrana

Entre os genes regulados para a condição do bagaço envolvidos no transporte de membrana, encontram-se os genes que expressam transportadores de ferro, AfuA, componentes de membrana interna de sistemas de transporte dependentes da proteína de ligação (*cthe_1586*), e AfuB, proteína de ligação ao soluto extracelular (*cthe_1588*). Em seguida, presentes genes como ATPase transporte de Ca^{+} (*cthe_2666*) e ATPase transporte de Zn^{+2} e Cd^{+2} (*cthe_0244*). Em condições de palha, também foi encontrado o gene *cthe_1588* com expressão aumentada e outro gene, *cthe_3134*, expressando uma proteína hipotética como transportador de Ca^{+2}.

Curiosamente, o gene *cthe_2736* anotado como fosfoenolpiruvato-proteína fosfotransferase (PTS), ainda não caracterizado em *C.thermocellum*, foi regulado com expressão aumentada em bagaço de cana. Este tipo de transporte de membrana, caracterizado por *Clostridium beijerinckii* e *Clostridium acetobutylicum*, é capaz de transportar glicose, maltose e lactose e, consequentemente, fosforila o substrato uma vez que esse se encontra dentro da célula [47].

3.2.3. Genes carboidrato ativos e celulossomais

Os dados de RNAseq revelaram um total de 98 genes envolvidos na degradação ou transformação de carboidratos, tendo neste grupo GHs (glicosídeo hidrolase), CEs (carboxil esterase), PLs (pectate lyases), proteínas contendo CBMs (domínio de ligação ao carboidrato) e proteínas estruturais (escafoldinas), que estão envolvidas na formação e anexação do celulossoma na parede da célula. Entre esses genes estão contabilizados aqueles que expressam proteínas putativas com função desconhecida, mas que contém o domínio doquerina tipo I ou CBM. O agrupamento hierárquico que resultou em 6 Clusters mostrou que o cluster C1 continha genes com baixa abundância de expressão para os três substratos, ao contrário dos últimos 5 grupos (53 genes), contendo genes altamente abundantes (Figura 6). O arquivo adicional com todos os genes e clusters está disponível (Apêndice G), assim como as anotações adicionais dos domínios encontrados e função de enzimas já caracterizadas baseadas na literatura [67] .

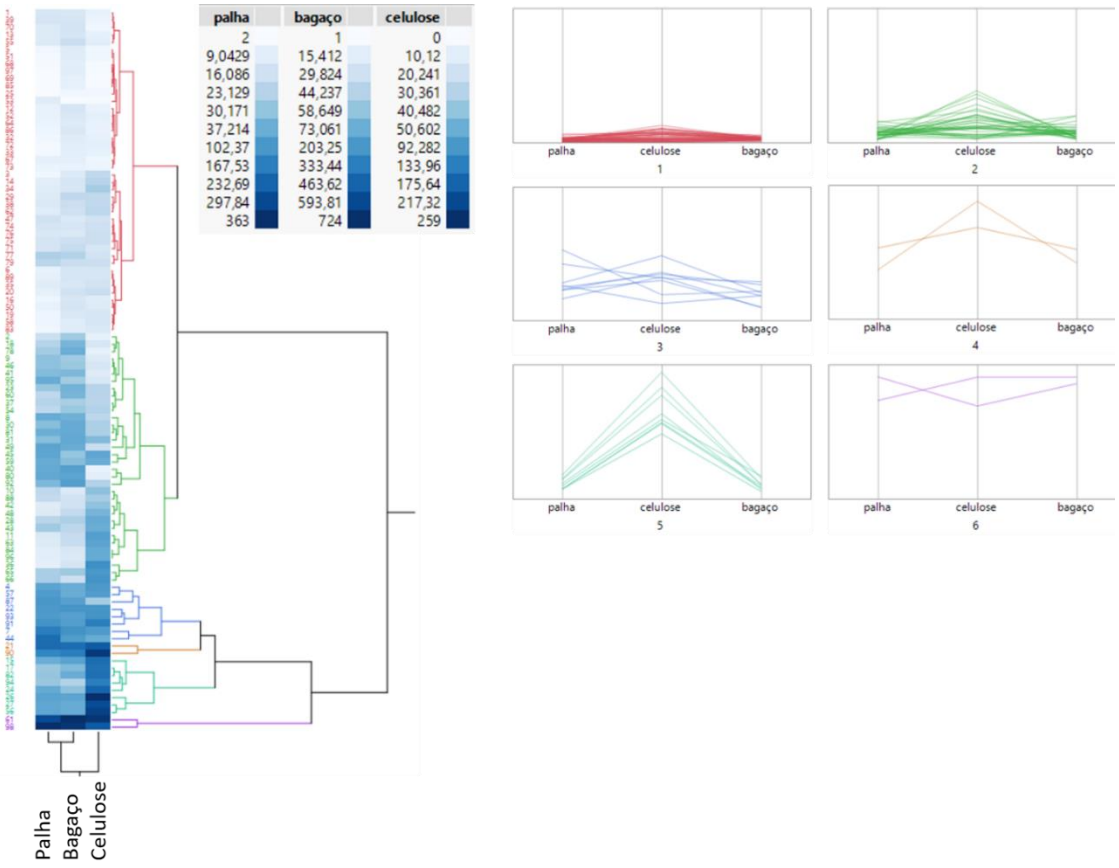


Figure 6. Agrupamento hierárquico, utilizando valores de expressão dos genes que codificam enzimas carboidrato ativas (CAZy- GHs, CBMs), e proteínas celulossomais, para as condições palha, bagaço e cellulose. JMP Genomics 6 software foi utilizado para elaboração do gráfico, gerando um total de 6 clusters.

Entre os 10 genes de maior expressão para cada condição, quatro genes são encontrados em todas três condições, celS (*cthe_2089*), cipA (*cthe_3077*), celK (*cthe_0412*) e uma proteína não caracterizada contendo CBM25 (cthe_3163). CelS e CelK, apresentam domínios GH48 e GH9, são proteínas celulossomais, com atividade de exoglucanase, enquanto CipA é a proteína estrutural principal de formação do celulossoma[54,102,103]. Dentre os genes mais expressos encontrados em ambos os preparos de bagaço e palha, estão CbhA, *cthe_3078* (OlpB), *cthe_0109* (proteína contendo doquerina tipo I) e *cthe_1080* (proteína contendo o domínio de carboidrato da família 25). CbhA é uma celobiohidrolase e OlpB é uma proteína estrutural que não possui ancoragem na célula, possuindo 7 doquerinas do tipo II [114,132].

Os genes celulossomais, como *cthe_2196*, anotados como uma arabinofuranosidase putativa, não apresentaram expressão na condição de celulose, valor igual a 0, e apresentaram maior expressão no bagaço em comparação com a palha. O mesmo gene foi predito por ser expresso em operon com o gene o *cthe_2195*. Além disso, três genes obtiveram valor de expressão de 0 para todas as condições, os genes *cthe_2137, cthe_2138, cthe_2139.* A análise de dados do genoma parcialmente montado de B8 (não publicado), não possui *reads* com cobertura nessa parte do genoma, poderia indicar uma deleção no genoma da linhagem B8, em comparação com a linhagem ATCC27405. Além disso, o gene *cthe_0918*, uma proteína dockerin tipo I, também não detectou nenhum *read*. Portanto, de um total de 81 genes celulossomais, encontrados em estudos anteriores, 77 foram detectados nos dados gerados nesse trabalho [23].

Entre os genes diferencialmente expressos ($p < 0,05$), com expressão aumentada em bagaço, 7 são celulossomais, um gene que codifica CelC (*cthe_2807*), uma endoglucanase não celulossomal, e 4 genes (*cthe_0956*, *cthe_1080*, *cthe_1257*, *cthe_3163*) que expressam proteínas apenas com anotações de CBM[66]. Dentre as celulossomais estão: proteína estrutural Orf2p, endoglucanase CelP (*cthe_0274*), endo-xilanase/esterase XynA (*cthe_2972*), xiloglucanase XghA (*cthe_1398*), endo-mananase ManA (*cthe_2811*), acetil xilana esterase (*cthe_0798*) e fator de coagulação (*cthe_0821*) [111,133,134]. Todos esses genes foram de 3.5 a 22 vezes regulados positivamente para o bagaço e 2.9 a 5 vezes regulados para a palha, porém para a condição palha apenas dois foram estatisticamente significativos (*cthe_0956, cthe_1080*). O gene que codifica a proteína acetil xilana esterase teve a maior regulação entre eles.

3.2.4. Fatores Sigma

Os fatores sigma (SigI-rsgI) já descritos, como atuantes na regulação de genes, sinalizados por carboidrato extracelulares, foram encontrados regulados positivamente em condições de palha. O par de genes *cthe_1272* e *cthe_1273*, correspondendo aos fatores sig5 e rsgi5, tiveram 4 a 5 vezes expressão aumentada em comparação com a celulose. A proteína Rsgi5 contém um domínio de alfa-L-arabinofuranosidase e um domínio CBM42 no C-terminal, do lado externo da célula, domínio predito pela capacidade de ligação ao polímero de arabinoxilana [17]. Na porção N-terminal dessa enzima, contém um fator anti-sigma que se liga ao fator sigma (sig5), que irá possivelmente regular a expressão de determinados genes envolvido na degradação.

3.2.5. Proteínas putativas

Genes expressando proteínas não caracterizadas foram encontrados regulados positivamente em condições de bagaço e palha. Dentre esses genes, putativos no envolvimento das modificações de carboidrato, foram encontrados genes que codificam as proteínas contendo CBM, já citadas acima, porém as proteínas preditas não apresentam peptídeo sinal. Porém, o gene *cthe_1258*, regulado positivamente em bagaço, codifica uma proteína cobre amino oxidativa, que apesenta um peptídeo sinal, podendo ter algum envolvimento na oxidação de compostos, como a lignina por exemplo. O domínio amino oxidativo presente possui código pfam PF07833. Uma análise no genoma de *C.thermocellum* revelou que 35 genes apresentam esse domínio, porém nenhum deles foi caracterizado.

3.3. Contribuição de *Moorella thermoacetica*

De acordo com o banco de dados do CAZy, existem 49 genes em *M.thermoacetica* que expressam genes que apresentam os domínios catalíticos, GHs, CEs e GTs, e domínios de ligação ao carboidrato (CBM) (Tabela 2). Todas essas proteínas não possuem caracterização descrita. A partir dessas proteínas, que podem ser potencialmente ativas em carboidratos presentes no meio, existem dois genes que codificam GHs apresentando peptídeo sinal, como

moth_2073 e *moth_1790*. *Moth_2073* expressa uma proteína que contém um domínio GH26 e um domínio cobre amino oxidase. A família GH26 apresenta proteínas com atividade como, β-mananase, exo-β-1,4-manobiohidrolase, β-1,3-xilanase, liquenase e exo-β-mananase [135]. Outro gene é o *moth_1790* que codifica uma proteína com um domínio GH73. No entanto, a família GH73 apresenta atividades como lisozima e acetilglucosaminidase, o que não indica a possibilidade de ser ativo na biomassa. Os genes *moth_1858*, e outros três que não apresentam peptídeo sinal (*moth_0040*, *moth_0739*, *moth_1057*) possuem um domínio de carboidrato esterase família 4 (CE4), porém o seu domínio anotado PF01476, foi caracterizado em Streptococcus, como contendo atividade de deacetilação de glicanos presentes na superfície da célula [136]. Os outros genes contendo peptídeo sinal apresentam (*moth_0054*, *moth_2104*, *moth_2055)* o domínio CBM50, Pfam PF01476, representando o domínio LysM associado principalmente com enzimas de degradação da parede da célula.

Tabela 2. Genes de *M.thermoacetica* codantes de proteínas que apresentam domínios carboidratos ativos anotados de acordo com o CAZy. Na tabela estpa indicando o gene, o Uniprot ID, a presença de peptídeo sinal, e o Pfam anotado.

Gene	Uniprot ID	Peptídeo sinal	Domínio (CAZy)	Pfam
Moth_1937	Q2RH54		CBM25	PF16760
Moth_1809	Q2RHI0		CBM48;GH13	PF00128;PF02922;PF11941
Moth_2055	Q2RGT8	Sinal	CBM50	PF01476
Moth_2104	Q2RGP3	Sinal	CBM50	PF01476;PF00877
Moth_0054	Q2RME6	Sinal	CBM50	PF07486;PF01476
Moth_1303	Q2RIX3		CBM50	PF01476
Moth_2401	Q2RFV6	Sinal	CBM50	PF01476;PF01551
Moth_1271	Q2RJ04		CBM50	PF01476
Moth_0040	Q2RMG0		CE4	PF01522
Moth_0739	Q2RKI2		CE4	PF01522
Moth_1057	Q2RJL8		CE4	PF01522
Moth_1858	Q2RHD1	Sinal	CE4	PF01522
Moth_1811	Q2RHH8		GH13	PF00128
Moth_1856	Q2RHD3		GH13	PF00128;PF02903;PF16657
Moth_0651	Q2RKR0		GH15	PF00723
Moth_0063	Q2RMD7		GH18	PF00704
Moth_0315	Q2RLP0		GH23	PF01551;PF01464
Moth_1608	Q2RI25		GH23	PF01464
Moth_1837	Q2RHF2		GH23	PF01464
Moth_2073	Q2RGS1	Sinal	GH26	PF07833;PF02156
Moth_2139	Q2RGK8		GH26	PF02156
Moth_1810	Q2RHH9		GH57	PF12055;PF03065
Moth_1854	Q2RHD5		GH57	PF09210;PF03065
Moth_1915	Q2RH74		GH57	PF09210;PF03065
Moth_1790	Q2RHJ8	Sinal	GH73	PF07833;PF01832

Moth_1850	Q2RHD9	GH77	PF02446
Moth_0055	Q2RME5	GT2	
Moth_0453	Q2RLA5	GT2	PF00535;PF02585
Moth_1660	Q2RHX7	GT2	
Moth_1831	Q2RHF8	GT2	PF13632
Moth_2154	Q2RGJ6	GT2	
Moth_2228	Q2RGC4	GT2	PF00535
Moth_2359	Q2RFZ7	GT2	PF00535
Moth_1367	Q2RIQ9	GT20	PF00982
Moth_2365	Q2RFZ2	GT26	PF03808
Moth_0843	Q2RK79	GT28	PF04101;PF03033
Moth_1852	Q2RHD7	GT35	PF11897;PF00343
Moth_0666	Q2RKP7	GT4	PF13439
Moth_0668	Q2RKP6	GT4	PF13579;PF00534
Moth_1133	Q2RJE2	GT4	PF13439;PF00534
Moth_1365	Q2RIR1	GT4	PF13439;PF00534
Moth_1658	Q2RHX9	GT4	PF00534
Moth_1661	Q2RHX6	GT4	PF00534
Moth_1812	Q2RHH7	GT4	PF00534
Moth_1853	Q2RHD6	GT4	PF13439;PF00534
Moth_1934	Q2RH57	GT4	PF13439;PF00534
Moth_0664	Q2RKP9	GT4	PF13439
Moth_1936	Q2RH55	GT5	PF08323;PF00534
Moth_1710	Q2RHS7	GT51	PF00912;PF00905

C.thermocellum e *M.thermoacetica*, apresentaram 1336 e 1257 genes ortólogos entre eles. Para a maioria das glicosil hidrolases de *C.thermocellum* discutidas anteriormente, existe uma única proteína ortóloga em *M.thermoacetica* expressa pelo gene *moth_0241*, contendo apenas um domínio de sialidase, ausente de um domínio ativo em polímeros de carboidratos (CAZy). Assim, para as enzimas correspondentes de *C.thermocellum* presentes no CAZy, não houve enzimas que pudessem ter a mesma função em *M.thermoacetica*,

A expressão diferencial dos genes de *C.thermocellum* mostrou um total de 89 e 253 genes regulados positivamente em palha e bagaço em comparação com a celulose, respectivamente (Apêndice H (A e B)). Esses genes foram na maioria não atribuídos a qualquer categoria de COG, seguidos por função geral (R), como mostrado na Fig.7. No bagaço, as principais categorias de genes encontradas foram transcrição (K), transporte de coenzima e metabolismo (H), tradução, estrutura ribossômica e biogênese (J), transporte de carboidratos e metabolismo (G) e produção e conversão de energia (C). Na palha, foram genes categorizados

em Produção e conversão de energia (C), transporte de coenzima e metabolismo (H) e transporte de carboidratos e metabolismo (G).

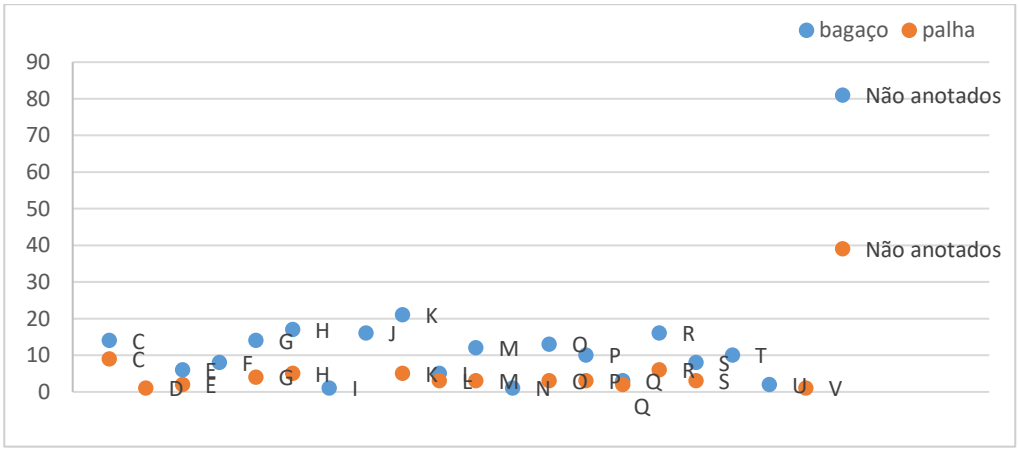


Figura 7. Número de genes diferencialmente expressos em cada categoria COG (Clusters of Orthologous groups) de Morella thermoacetica regulados positivamente para as amostras de palha (vermelho), e bagaço (azul), em comparação com celulose. Categorias COG: (C) Produção de energia e conversão; (D) Cromossomo, divisão celular e controle do ciclo celular; (E) Metabolismo e transporte de aminoácido; (F) Metabolismo e transporte de nucleotídeo; (G) Metabolismo e transporte de carboidrato; (H) Metabolismo e transporte coenzima; (I) Metabolismo e transporte de lipídeo; (J) Biogêneses, tradução e estrutura do ribossomo; (K) Transcrição; (L) Replicação, recombinação, e reparo; (M) Parede celular/ biogêneses membrana; (N) Motilidade da célula; (O) Modificação pós-traducional, turnover proteína, chaperonas; (P) Metabolismo e transporte de íons inorgânicos; (Q) Biossíntese do metabolismo secundário, transporte e catabolismo; (R) Predição de função geral somente; (S) Função desconhecida; (T) Mecanismo de transdução do sinal; (U) Secreção e tráfego intracelular; (V) Mecanismos de defesa.

Relativamente ao transporte de carboidratos e ao metabolismo (G), as proteínas transportadoras ABC ganharam atenção na regulação positiva em bagaço, como *moth_0612* e *moth_0699*. Ambos os transportadores são putativos para o transporte de ribose, porém não foram ainda caracterizados. Além desses transportadores, dois genes que expressam proteínas do sistema PTS fosfotransferase (*moth_0013*, *moth_0014*), estão anotados, como atuantes na fosforilação da ribose uma vez que se encontra dentro da célula, e posteriormente segue para a via glicolítica.

4. Discussão

Primeiramente, genes que estão sendo expressos em abundância nas condições que apresentamos nesse trabalho, podendo não ter aumento significativo nas amostras de palha e bagaço, são genes de relevância para o metabolismo de *C.thermocellum*. Assim, genes que

apresentaram por exemplo, altos valores de expressão em todas as condições, como o gene *cthe_1020* que codifica um transportador ABC, CbpB, que se liga a celodextrinas, evidenciou a absorção de açúcar na célula para todas as condições de crescimento de substratos. Além disso, esse dado corrobora com a proteína CbpB descrita anteriormente por ser altamente abundante entre transportadores de açúcar[100]. Outro gene abundante, *cthe_3383*, também encontrado como altamente expresso em trabalhos anteriores quando *C.thermocellum* cresce em gramínea ou populus, foi caracterizado seu gene homólogo (*clo1313_2818*) anteriormente [38]. Esse gene descrito como sendo sua expressão induzida por xilose como resposta ao estresse da bactéria em presença a esse soluto [122]. Portanto, a presença desse gene, pode indicar a presença de xilose no meio de cultura, resultante da degradação da biomassa.

No geral, os genes do metabolismo total foram ativados na condição de celulose, o que poderia representar este substrato como mais favorável e enérgico ao crescimento de *C.thermocellum*. Entretanto, a bactéria, quando em crescimento em bagaço de cana e palha de cana, ativou diferentes vias envolvidas principalmente envolvidas na motilidade e sinalização da bactéria, além de regular positivamente carboidrato enzimas ativas.

Como *C.thermocellum* é uma bactéria móvel, o movimento flagelar ajuda a proximidade entre a bactéria e o substrato [120]. Estudos anteriores mostraram por exemplo em linhagem de *C.thermocellum* tolerante ao etanol, uma elevação da regulação das proteínas envolvidas na transdução de sinal e na quimiotaxia, o mesmo anteriormente descrito quando a bactéria cresce em celulose [119,120]. Assim, a quimiotaxia pode ser entendida como um mecanismo, ou processos celulares para a sobrevivência das bactérias, e neste caso a bactéria está tentando alcançar a biomassa, a fonte de carbono. Além disso, a presença de regulação positiva de genes que codificam pilus indicam a aderência da bactéria ao substrato, como descrito anteriormente na predominância desses genes em espécies de *Clostridium e Caldicellulosiruptor*, quando crescem em celulose, participando da adesão da bactéria no substrato[26].

De acordo com estudos anteriores de RNAseq, foi demonstrado uma alta expressão de transportadores em geral que são responsáveis pelos transportadores de ânions, como sulfato e fosfato, possivelmente tendo um papel na adaptação ambiental devido à mudança osmótica devido à composição da biomassa [38]. Dessa forma, os transportadores de íons encontrados regulados positivamente nas condições de biomassa, podem estar relacionados também á essa resposta da composição da biomassa.

Outro transportador encontrado e sem caracterização em *C.thermocellum* foi o transportador putativo PTS (*cthe_2736*). O sistema PTS consiste em uma proteína de membrana e acoplador de energia, e sua expressão foi demonstrada como induzida pelo substrato em outras bactérias, como *Clostridium beijerinckii* e *Clostridium acetobutylicum*[137]. No entanto, o transporte de açúcar foi descrito apenas usando transportadores ABC em *C.thermocellum* [83]. Como este gene foi induzido nos substratos (bagaço e palha), poderia indicar um novo transportador de açúcar em *C.thermocellum.*

Analisando todos os genes expressos, relativos à degradação de carboidrato, em todas as condições, CelS e CelK se mostrou abundante, dado corroborado por estudos anteriores que descrevem CelS como sendo a proteína celulossomal mais abundante [48]. De acordo com um estudo anterior, onde *C.thermocellum* cresce em gramínea, CelS e CelK, também estão entre os mais abundantes [58]. Essas enzimas são, portanto, importantes na degradação de quaisquer substratos, desde que são as mais abundantes tanto em celulose, quanto em bagaço e palha de cana, o que sugere a atividade de exoglucanase como sendo essencial na degradação desses polímeros. Por outro lado, entre proteínas abundantes apenas em palha e bagaço de cana, sugerem a participação dessas com extrema importância em biomassas, sendo elas CbhA, uma cellobiohidrolase, demonstrada também abundante em cultivos em gramínea, e outras duas proteínas não caracterizadas (cthe_1080), (cthe_0109), apresentando um domínio CBM25 e um domínio doquerina tipo I[38,58]. Abundante também em bagaço de cana e palha de cana, a proteína estrutural OlpB, é descrita como relacionada à colonização das bactérias em substratos sólidos [58].

Dentre os genes positivamente regulados em bagaço de cana, foram encontrados genes que codificam enzimas ativas nos polímeros de hemicelulose. XynA é bifuncional, sendo uma endo-xilanse e também esterase [133]. O gene (cthe_0798) também expressa uma proteína putativa, com atividade de esterase, XynA e cthe_0798, ambas são capazes, portanto de liberar grupos acetil da cadeia principalmente de hemicelulose. Enquanto isso, a xiloglucanase XghA hidrolisa ligações β-1,4-glucanas no polímero de xiloglucano[111].Todas as três enzimas podem atuar em sinergia durante a degradação do bagaço, atuando em polímeros de xilana, xilana contendo acetilações e xiloglucano. A enzima acetil esterase (cthe_0798) possui a mais alta regulação, demonstrando uma função provavelmente chave na degradação da hemicelulose, sendo uma enzima de grande interesse para caracterizações futuras. A enzima ManA, também regulada positivamente, possui uma atividade de endo-mananase. Dados anteriores demonstraram que resíduos agrícolas, como palha de trigo e bagaço de cana-de-

açúcar, continham manana, mesmo que em pequena quantidade quando em comparação com xilana e arabinana [138]. Além disso, ManA foi relatada por atuar em sinergia com arabinofuranosidases em estudos anteriores, aumentando a liberação de açúcar na degradação de beterraba e bagaço de cana de açúcar pré-tratado [139,140].

Estudos anteriores demonstraram a liberação de ácido cumárico, um componente fenólico de lignina, proveniente de bagaço de cana-de-açúcar não pré-tratado e gramíneas, quando *C.thermocellum* utiliza esses substratos como fonte de carbono para crescimento [116]. No entanto, ainda é desconhecido quais enzimas são responsáveis por esta degradação, mesmo com testes experimentais na deleção de genes que codificam feruloil esterases (homólogos dos genes *cthe_0912*, *cthe_1963*, *cthe_2194*) [55]. Assim, as proteínas hipotéticas encontradas aqui reguladas em bagaço e cana-de-palha, poderiam estar relacionadas a algumas dessas atividades e outras relacionadas à degradação da biomassa.

Os fatores sigma sig5-rsgi5 encontrados regulado positivamente em palha de cana , foram também demonstrados por ter a expressão aumentada quando *C.thermocellum* cresce em xilana e celulose, em comparação com apenas celulose [80]. O fator rsgi5, possui um domínio de AbfB, que tem como principal função, detectar o carboidrato no meio externo, e não degradação, como outros como RsgI6 (cthe_*2119*), tendo o domínio GH10 e Rsgi24C (*cthe_1471*) possuindo o domínio GH5 [82]. Assim, a regulação de rsgi5 provavelmente indica que a porção de xilana ou arabinoxilana do substrato palha de cana estava exposta, disponível para ser detectado por rsgi5, com a consequente ativação dos fatores sigma na regulação da expressão gênica. No entanto, esse par de fatores sigma ainda não está totalmente caracterizado acerca dos genes que regula.

A co-cultura de *C,thermocellum* e outras bactéria tem sido descrita em outros trabalhos [39]. Esta condição pode ser útil como uma ferramenta para melhorar os produtos fermentativos, principalmente devido a *C.thermocellum* possuir a maquinaria para hidrolisar celulose e hemicelulose, liberando xilose (C5) e glicose (C6), portanto possuindo capacidade de fermentar apenas açúcares C6 [67]. *M. thermoacetica*, em co-cultura com *C.thermocellum* nesse trabalho, é capaz de fermentar os açúcares C5 e C6, entre então, xilose, preferencialmente, seguido de frutose e glicose [28]. Por outro lado, essa bactéria não é capaz de hidrolisar celulose e hemicelulose, sendo, portanto, um microrganismo não celulolítico. Embora, esta bactéria seja capaz de crescer em H2 / CO2 ou CO / CO2, chamado crescimento chemo-litoautotrófico.

Dados de cobertura maiores do genoma de *Moorella termoacetica* em condições de palha e bagaço em comparação com celulose, sugere que a maior disponibilidade de açúcar xilose acima de glicose, privilegia a fermentação de *M. thermoacetica*, portanto aumentando seu crescimento na cultura. Além disso, a capacidade de *M.thermoacetica* de utilizar xilose aparentemente pose se apresentar benéfica para o crescimento de *C.thermocellum*, uma vez que a xilose pode prejudicar o crescimento de *C.thermocellum*, como discutido anteriormente [25].

Contudo, como descrito nos Resultados, *M.thermoacetica* não apresenta enzimas potenciais de degradação do substrato, mas demonstras na regulação positiva de transportadores de ribose, o que demonstra a utilização de açúcares liberados por *C.thermocellum*, em uma relação de sintrofia. Dados gerados desse transcritoma, aliados à proteômica e caracterização enzimática deveriam ser realizados para corroborar os dados aqui obtidos.

Capítulo 4. Expressão heteróloga

Este capítulo foi aceito para publicação na revista Enzyme and Microbial technology. Brenda R. de Camargo[a]; Nico J. Claassens[b], Eliane F. Noronha[a*], Servé W.M. Kengen[b] (2017). **Heterologous expression and characterization of a putative glycoside hydrolase family 43 arabinofuranosidase from *Clostridium thermocellum* B8.**

Heterologous expression and characterization of a putative glycoside hydrolase family 43 arabinofuranosidase from *Clostridium thermocellum* B8

Brenda R. de Camargo[a]; Nico J. Claassens[b], Eliane F. Noronha[a*], Servé W.M. Kengen[b]

[a]Laboratory of Enzymology, Department of Cell Biology, University of Brasília, Brasilia, DF, Brazil.
[b]Laboratory of Microbiology, Wageningen University and Research, Wageningen, The Netherlands

*Corresponding author:
PhD. Eliane Ferreira Noronha
Enzymology Laboratory, University of Brasilia (UnB), 70910-900, Brazil
e-mail: enoronha@unb.br. Tel: +55 61 31072952

Abstract

An extensive list of putative cellulosomal enzymes from *C. thermocellum* is now available in the public databanks, however, most of these remain unvalidated with regard to their activity and expression control mechanisms. This is particularly true of those enzymes putatively involved in hemicellulose deconstruction. Our research group has been working on mapping and characterization of glycoside hydrolases produced by *C. thermocellum* B8, that are critical for lignocellulosic biomass deconstruction. One of the identified genes expressed during growth on sugar cane bagasse and straw is *axb8*, which encodes a putative cellulosomal GH43_29 α-arabinofuranosidase (EC 3.2.1.55) that has not previously been characterized at the molecular or kinetic levels. The AxB8 predicted amino acid sequence presented GH43 and dockerin domains, as well as a family 6 carbohydrate-binding module (CBM6). Also, it is a close homologue of *Firmicutes* putatives α-arabinofuranosidases, including cellulosomal proteins. Multiple alignment analysis grouped AxB8 in a cluster with four uncharacterized putative GH43_29 subfamily enzymes, all containing dockerin type I domain and CBM6 modules. Purified heterologously expressed AxB8 showed activity against the synthetic substrates *p*NPX (*p*-nytrophenyl-β-D-xylopyranoside) and *p*NPA (*p*-nytrophenyl-α-L-arabinofuranoside), as well as against the natural substrate wheat arabinoxylan (WAX), with maximal activity at 50°C and pH between 5.0 and 6.0. The WAX degradation profile by AxB8 is different from those typically seen for α-arabinofuranosidases, presenting mainly xylose as a hydrolysis product, instead of arabinose. In addition, unlike other GH43_29 enzymes already characterized, AxB8 did not present activity against arabinan. Kinetic parameters using *p*NPA as a substrate were K_m of 23 ± 3 mM and k_{cat} of 104 ± 7 s^{-1}. Despite its activity against *p*NPX, we did not observe AxB8 saturation with this substrate. AxB8 is the first member in its clade to be characterized regarding kinetic parameters, and together with its closest homologues could represent a large group of glycoside hydrolases with particular properties within the GH43_29 subfamily.

Keywords: *Clostridium thermocellum* B8, Firmicutes, GH43 glycoside hydrolase, arabinofuranosidase.

Introduction

The potential use of plant biomass in production of biofuels, especially second-generation bioethanol, has led to an increase in studies aiming to achieve efficient lignocellulose deconstruction. Filamentous fungi from the genus *Trichoderma* and *Aspergillus* have been studied as sources of enzymes for the complete hydrolysis of lignocellulose feedstocks. In the same way, thermophilic microorganisms have been studied as an attractive source of enzymes for industrial purposes [1]. These organisms produce a wide variety of robust glycoside hydrolases (GHs) [2,3], which are active under harsh bioprocessing conditions, and in general, are more active against recalcitrant substrates, presenting higher hydrolysis rates in comparison to their fungal counterparts [4].

Among thermophiles, *Clostridium thermocellum*, a Gram-positive bacterium belonging to the *Firmicutes* phylum, is one of the most intensively studied microorganisms for biotechnological applications. In particular, it has been highlighted with regard to its ability to hydrolyse crystalline cellulose and secrete holocellulases arranged in a supramolecular structure called the cellulosome [5].

The development of next-generation sequencing platforms has caused a remarkable increase in the available number of sequences for putative bacterial GHs, including members of GH43, a relatively less-well characterized GH family, with a total of 7622 members and 151 characterized proteins, according to the CAZy database. Indeed, the increasing numbers of GH43 members has driven its subdivision into 37 subfamilies [6], organized according to significant structural and biochemical differences between the subfamilies and between these and other GH families. Mewis et al (2016) also suggested that GH43 members probably display an underestimated variety of suitable specificity features that are not readily apparent when these enzymes are assayed with simple synthetic substrates.

In general, GH43 is characterized as a diverse family mainly constituted by extracellular and bacterial debranching and degrading GHs that are active against hemicelluloses including *β*-xylosidases (EC 3.2.1.37), endo-*α*-L-arabinanases (EC 3.2.1.99), *α*-L-arabinofuranosidases (EC 3.2.1.55) and 1,3-*β*-galactosidases (EC 3.2.1.145) [7]. A common feature of members of this family is the arrangement of a GH43 domain in a five-bladed β-propeller fold, containing three active side residues (Asp^{24}, Glu^{225}, Asp^{163}) [8]. As well as this, they contain specific binding sites related to their specificity and

activity against arabinan and arabinoxylan, the triads $W^{183}G^{184}N^{185}$ and $W^{86}A^{87}P^{88}$, respectively, with a central role for the amino acid residues tryptophan, asparagine and proline [9,10]. In addition to the catalytic GH43 domain, members of this family usually harbour carbohydrate-binding modules (CBMs), such as CBM6, CBM35, CBM13 and CBM42, which enables their attachment to xylose polymers, and arabinose polymers in case of CBM42 [6].

Regarding *C. thermocellum*, the latest genomic update shows that a total of 81 genes are annotated as cellulosomal genes and 21 of these potentially encode hemicellulase enzymes [11–14]. Among hemicellulase enzyme genes were found seven genes presenting predicted GH43 domains: *cthe_0015* (subfamily 4), *cthe_0661* (subfamily 24), *cthe_2196* (subfamily 29), *cthe_1271* (subfamily 16), *cthe_2138* (subfamily 16) and *cthe_2139* (subfamily 20) [12]. So far, only the proteins encoded by *cthe_1271* (*Ct*Abf43A) and *cthe_0661*, (1,3Gal43A), displaying a-L-arabinofuranosidase and β-1,3-galactosidase activities, respectively, have been functionally characterized [15–17].

In previous reports, our research group showed that a *C. thermocellum* isolate from goat rumen (isolate B8) produces cellulases and hemicelullases in the presence of cellulose or complex lignocellulosic biomass, such as sugar cane bagasse, straw or cotton waste. We also showed the effect of the carbon source on the production and secretion of non-cellulosomal or cellulosomal enzymes [18,19]. Furthermore, in order to identify novel hydrolases involved in biomass deconstruction, we carried out transcriptomic and proteomic analysis of *C. thermocellum* B8 during growth on sugarcane bagasse or sugarcane straw as a carbon source (unpublished work). We identified a set of genes expressed during growth on sugar cane straw and bagasse, including *axb8*, a homologue of *cthe_2196*, annotated as encoding a putative cellulosomal protein of family GH43 subfamily 29. The gene product ~~would~~ represent a previously uncharacterized enzyme with regard to its sequence and biochemical properties.

Taking into account: a) AxB8 expression during growth in the presence of complexes lignocellulosic biomasses as carbon source; b) The lack of information concerning biochemical properties of enzymes classified as pertaining to family GH 43 subfamily 29, including AxB8 and homologous proteins, the present work aimed to characterize the *axb8*-encoded protein (AxB8), aiming to provide more information about the complex process of hemicellulose deconstruction by *C. thermocellum*. In addition, as GH43 is a poorly characterized but expanding family, the present work will also contribute to increasing understanding of this family.

Materials and Methods

Bacterial strain collection and maintenance

Clostridium thermocellum isolate B8 was isolated from goat rumen samples, as previously described by Hamann et al [20]. Enrichment was carried out by culturing on liquid medium containing 1% (m/v) microcrystalline cellulose as the carbon source. To isolate the bacteria, enriched cultures were streaked on solidified medium containing carboxymethyl cellulose (CMC). One of the isolates, designated B8, showed cellulase and hemicellulase activity after growth on microcrystalline cellulose, sugarcane bagasse/straw and cotton waste [18,19]. The B8 isolate was further characterized based on its 16S rDNA and genome sequence (not published).

Transcriptome analysis of *C. thermocellum* B8 grown on sugarcane bagasse, sugarcane straw, and cellulose was performed previously (unpublished data). One of the genes that was highly expressed on sugarcane/straw and bagasse was *axb8*, which showed 100% identity with *cthe_2196* from *C. thermocellum*, ATCC27405. *Cthe_2196* was annotated as encoding a carbohydrate-binding family 6 protein (CBM6), harbouring a GH43 family domain and a CBM6 family domain.

Sequence alignment and analysis

To identify the closest homologues of protein AxB8 (named after the gene *axb8* from *C. thermocellum* B8) and further place it into one of the subfamilies, an *in silico* analysis was performed.

First, homology studies using the full AxB8 protein sequence, including the dockerin domain, were performed by protein-protein BLAST (blastp) against the non-redundant sequence database (http://www.ncbi.nlm.nih.gov). From homologous sequences, carbohydrate hydrolase activity domains and carbohydrate-binding modules were identified using available information at NCBI, CAZy Database – carbohydrate-active enzyme database ([21]; http://www.cazy.org), and UniProtKB ([22]; http://www.uniprot.org/uniprot/). Genome boundary analysis related to the *axb8* location was performed using the genome of *Clostridium thermocellum* ATCC 27405 (accession code NC_009012.1).

Comparative structural analysis was performed by alignment, using the T-Coffee program [23] with the following sequences: predicted sequence of AxB8 from *C. thermocellum* B8, Xsa43E from *Butyrivibrio proteoclasticus* B316 (ADL35052.1,E0RYY0), Cthe_2196 from *C. thermocellum* ATCC27405 (ABN53398.1,A3DHG9), *Cj*Abf43A from *Cellvibrio japonicus* Ueda107 (ACE83886.1, B3PD60), and *Ct*Abf43A from *C. thermocellum* (ABN52503, A3DEX4). Alignments were visualized using ESPRIPT software, with Xsa43E PDB (4NOV) as a secondary structure protein model.

A phylogenetic tree was compiled using the GH43 domain of AxB8 and other GH43 domain-containing proteins from subfamily 29, retrieved from the CAZy database (Supplementary Table 1). Based on a multiple sequence alignment using MAFFT, highly dissimilar sequences were discarded and the phylogenetic tree was constructed with FastTree, including 181 sequences. The tree was visualized by the program Evolview and domain annotation was carried out with the Pfam database [24,25]. Some sequences from the first tree were suppressed, collapsing in nodes, named INT1 (internal node 1) and INT2 (internal node 2) for a better visualization, generating a second tree (Supplementary Fig.1).

DNA extraction

C. thermocellum B8 was cultured using cellulose-medium fiber (Sigma) as the substrate, as described above. Cultures were centrifuged and the cell pellet was used for DNA extraction using the Wizard SV Genomic DNA Purification System (Promega, Madison, WI, USA).

Gene cloning

Axb8 was amplified by PCR using two primer combinations: forward primer (5'-CCTTCCATGGATGAATCCGATAGTACAAACAAT-3') and reverse primers (5'-CCTTCTCGAGTTTATCCGGCATCGGTGT-3') or (5'-CCTTCTCGAGATTTTGAGGAAAATCCGAT-3'), containing NcoI and XhoI restriction sites, respectively. Primers were designed based on the *cthe_2196* sequence (Gene ID: 4811061) to amplify *axb8* with or without the dockerin type I domain (nucleotides 1402 to 1572), with no signal peptide (nucleotides 1 to 72) at the 5' end and no stop codon at 3'end, in order to allow fusion with a C-terminal His_6 affinity tag. The PCR mixture contained 100 ng of genomic DNA, 2.5 mM dNTP, 1.5 units of *Pfu* DNA Polymerase, 0.4µM primers, and 5µL of 10X buffer, completing the final volume with nuclease-free water to 50µL. The PCR protocol consisted of an initial step at 95°C for 1 min, followed by 35 cycles (95°C for 30 s, 60°C for 30

s and 74°C for 4 min); ending with a step at 74°C for 5 min. The amplified product was purified, digested with the corresponding restriction enzymes (Thermo Scientific) and ligated into pET24d (70 ng) (Merck, Millipore, Darmstadt, Germany) using 0.2 µL of T4 DNA ligase (Thermo Scientific). The ligation mixture was used to transform competent *E.coli* DH5α by heat shock (New England Biolabs, Ipswich, MA, USA). Positive clones were selected on kanamycin plates and checked for the expected fragment by colony PCR, using T7 primers: T7 forward (5'-TAATACGACTCACTATAGGG-3') and T7 reverse (5'-GCTAGTTATTGCTCAGCGG-3'). The PCR mixture contained 1 µL of diluted colony and 15µL of ddH_2O, 0.2µM primers, 12.5 µL of One*Taq* Quick-Load 2X Master Mix (NEB), completing the volume to 25 µL. The PCR protocol consisted of an initial step at 94°C for 1 min, followed by 30 cycles (94°C for 30 s, 55°C for 30 s and 68°C for 1.5 min); ending with a step at 68°C for 5 min.

The resulting plasmids, pWURaxb8d (with dockerin) and pWURaxb8 (without dockerin) were sequenced using T7 forward and reverse primers. Chemically competent *E. coli* BL21(DE3) cells were then transformed with the confirmed pWURaxb8 plasmid (no dockerin). Positive clones were selected on kanamycin plates and checked by colony PCR using T7 primers as described above.

Recombinant protein expression

The recombinant *E. coli* BL21(DE3) strain was grown in 1L of Luria-Bertani (LB) medium, containing 50 µg/mL of kanamycin, at 37°C, until the culture reached an OD_{600} of 1.2 to 1.5. Then, 0.5 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) was added and cells were incubated for 20 h at 20°C. Cells were harvested and resuspended in 20 mM sodium phosphate buffer (pH 7.4), containing 500 mM of NaCl and 1 mM of Pefabloc (Sigma-AldrichChemie N.V. Zwijndrecht, NL). Cells were disrupted using a French press treatment at 110 MPa. This treatment was performed twice, and the sample was subsequently centrifuged at 16,000 x *g* for 30 min at 4°C. The resulting supernatant (cell-free extract) was collected and stored at 4°C until further use. The cell-free extract was analysed by SDS-PAGE using precast 10% bis-acrylamide gels (Bio-Rad, Hercules, CA, USA), and proteins were visualized by staining with QC Colloidal Coomassie stain (BioRad, Hercules, CA, USA) [26]. For western blot analysis, proteins were transferred to nitrocellulose membrane using an iBlot® 2 Dry Blotting System (Thermo Scientific, Hudson, NH, USA), and detection was performed using the HisProbe™-HRPdetection kit (Thermo Scientific, Hudson, NH, USA) according to the manufacturer's protocol.

Enzyme assays

Enzymatic activity of the putative arabinofuranosidase (AxB8) was measured using 4-nitrophenyl-β-D-xylopyranoside (*p*NPX) and 4-nitrophenyl-α-L-arabinofuranoside (*p*NPA). The standard reaction mixture contained: 25 µL of 200 mM sodium acetate buffer (pH 5.0) for the *p*NPA-assay or 200 mM sodium phosphate buffer (pH 6.0) for the *p*NPX-assay, 20µL of enzyme preparation at 0.01 mg/mL and 20µL of *p*NP-substrate (5mM). The final mixture of 100 µL was incubated at 50°C for 10 min and then the reaction was stopped by adding 200 µL of 1 M sodium carbonate. Reaction products were measured by their absorbance at 410 nm using a UV-visible plate spectrophotometer, Synergy Mx (BioteK, Winooski, VT). All enzyme assays were performed in triplicate and one unit of enzyme activity (U) was defined as the amount of enzyme that produces one µmol of *p*NP per minute.

Protein purification

Cell-free extract (10 mL) was filtered (0.22 µm) and loaded onto a 1-mL His-trap FF column (GE Healthcare Life Sciences, Eindhoven, NL), equilibrated with buffer A (500 mM of NaCl, 20 mM sodium phosphate buffer, pH 7.4). Unbound proteins were eluted by washing the column with 15 mL of buffer A and bound proteins were eluted by washing with buffer A containing 300 mM imidazole. Fractions (1 mL) were collected and analysed on SDS-PAGE and checked for enzyme activity using *p*NPA. Western blotting was also performed (see "Recombinant protein expression", above). Chromatographic fractions showing enzymatic activity were pooled and desalted using a 5-mL Hi-trap desalting column (GE Healthcare Life Sciences, Eindhoven, NL), equilibrated with 20 mM Tris-HCl buffer (pH 7.8). The eluted sample (10mL) was then used in a second purification step by anion exchange chromatography on a 1mL Mono-Q-HR column (GE Healthcare Life Sciences, Eindhoven, NL), equilibrated with 20 mM Tris-HCl buffer (pH 7.8) (Buffer B). Unbound proteins were eluted by washing the column with 5 mL of buffer B and bound proteins were eluted in 1mL fractions in a 20ml linear gradient of 0.0 to 1.0 M sodium chloride. Fractions containing activity and showing the same profile on SDS-PAGE were pooled and desalted as previously described. Purity was assessed by SDS-PAGE (10%) [26].

Protein concentration was determined by measuring the absorbance at 280 nm in a Nanodrop D-11 Fx spectrophotometer (DeNovix, Wilmington, DE, USA) using a theoretical extinction coefficient of 115 M^{-1} cm $^{-1}$, as calculated using the ProtParam tool (www.expasy.org).

Enzyme characterization

Molecular mass estimation

The molecular mass of native AxB8 was estimated by size exclusion chromatography on a calibrated HiLoad 16/600 Superdex 200 column (GE Healthcare Life Sciences, Eindhoven, NL). The following protein standards were used: ovalbumin (44 kDa), conalbumin (75kDa), aldolase (158 kDa), ferritin (444 kDa) and thyroglobulin (669 kDa). A protein sample of 500 µL (0.01 mg/mL) was applied at a flow rate of 1 ml/min in 20 mM Tris-HCl (pH 8.0), containing 150 mM NaCl. Peak fractions were checked for enzyme activity using *p*NPA as described above.

Substrate specificity

The substrate specificity of AxB8 was analysed using the following *p*NP substrates: *p*NP-β-D-glucopyranoside (*p*NP-β-D-glu), *p*NP-β-D-galactopyranoside (*p*NP-β-D-gal), *p*NP-β-D-mannopyranoside (*p*NP-β-D-man), *p*NP-β-D-xylopyranoside (*p*NPX), *p*NP-β-L-arabinopyranoside (*p*NP-β-L-arap), *p*NP-α-L-arabinopyranoside (*p*NP-α-L-arap), as well as *p*NP-α-L-arabinofuranoside (*p*NPA). Assays were performed as described above, using sodium citrate buffer (pH 5.5) instead of acetate buffer. All assays were performed in triplicate.

The DNS method was used for determining enzyme activity on natural substrates like xylan oat spelt, using xylose as a standard [27]. The reaction mixture contained: 35 µL of 200 mM citrate buffer (pH 5.5), 50 µL of enzyme preparation at 0.01 mg/mL, and 65 µL of substrate (2% w/v). The mixture was incubated at 50°C for 30 min and the reaction was stopped by adding 300 µL of DNS, and boiled for 5 min. All enzyme assays were performed in triplicate. Reaction products were measured by their absorbance at 540 nm. The xylan oat spelt activity assay was also analysed using a combination of a characterized endo-xylanase [28] and AxB8, with the aim of detecting a probable synergistic effect between the two enzymes.

To determine the mode of action of AxB8 on natural substrates we analysed product formation by HPLC. Oligosaccharides and natural substrates used were: xylobiose, xylotriose, xylotetraose, arabinan sugar beet (AA) and wheat arabinoxylan (WAX), all purchased from Sigma-Aldrich (Zwijndrecht, NL) or Megazyme (Bray, County Wicklow, IRL). Hydrolysis assays were carried out at 45°C for 19 h in 270 µL assay mixtures containing 1 mg of substrate diluted in 250 µL of 50 mM sodium citrate buffer (pH 5.5), containing 2 mM $CaCl_2$ and 20 µL of enzyme preparation at 0.01 mg/mL. All enzymatic assays were performed in duplicate. The

mixture was boiled for 5 min at 100°C, and centrifuged for 5 min at 1,758g. The supernatant was used for analysis of hydrolysis products (released sugars) by high-performance anion exchange chromatography **(**HPAEC) as described previously [29]. Xylose and arabinose were used as standards.

Effects of pH and temperature

The effects of temperature and pH on AxB8 activity were analysed as described above with the following changes: the optimal temperature was determined in sodium acetate buffer (pH 5.0) using *p*NPA as substrate, and in sodium phosphate buffer (pH 6.0) for *p*NPX. The temperature was varied between 30° and 70°C. The optimal pH was determined at 55°C, using citrate-phosphate buffer (McIlvaine) buffer, and varying the pH between 3.0 and 8.0. All enzyme assays were performed in triplicate. The residual activity was expressed as a percentage of the highest activity.

Thermostability was determined by incubating AxB8 (0.01mg/ml) at 50°C and 60°C, taking aliquots at different time points and measuring the residual activity with *p*NPA. The assay was performed as described above using sodium acetate buffer (pH 5.0). Samples were taken at 0, 3, 9 and 15 hours. The residual activity was expressed as a percentage of the highest activity.

Salt effect

The effect of mono- and divalent cations on AxB8 was determined in the standard assay described above. First, salt was added to the standard assay mixture containing the enzyme, and pre-incubated at room temperature for 5 min, before starting the reaction by adding the substrate (*p*NPA). Salts and final concentrations evaluated were: NaCl (40mM), $CaCl_2$ (20mM), $MgSO_4$ (4mM), $MnSO_4$ (4mM), $ZnSO_4$ (4mM), $CuSO_4$ (4mM), $NiSO_4$ (4mM), $FeCl_3$ (4mM). All enzyme assays were performed in triplicate. In each case the residual activity was compared to the activity without additional compounds, set as 100%. Concentrations were chosen based on previous experiments on a GH43 protein from *C. thermocellum* [16]

Kinetic parameters

Kinetic parameters, K_m, V_{max} and k_{cat}, were determined using *p*NPX and *p*NPA, as substrates. The standard reaction mixture contained: 10 µL of 200 mM sodium acetate buffer (pH 5.0) for *p*NPA or 200 mM sodium phosphate buffer (pH 6.0) for *p*NPX, 10µL of enzyme sample at 0.01 mg/mL, and *p*NP-substrate (5 to 500mM). The final mixture of 50µL was

incubated at 50°C for 10 min and then the reaction was stopped by adding 100 μL of 1 M sodium carbonate. All enzyme assays were performed in triplicate. Values of K_m, V_{max} and k_{cat} were calculated by nonlinear regression using Michaelis-Menten equation fit and GraphPad Prism6 kinetic analysis package (Graphpad Software, San Diego, CA).

Results

In-silico analysis of axb8

The *axb8* gene, with an open reading frame of 1,602 bp, is 100% identical to *cthe_2196*, a homologous gene from *C. thermocellum* ATCC27405 strain. It is annotated as encoding a putative glycoside hydrolase of 533 amino acid residues classified as pertaining to GH43 family, subfamily 29 [6]. The putative protein sequence, in addition to the typical GH43 domain (amino acids 35 to 311), also presents a signal peptide of 24 amino acids, a carbohydrate-binding module family 6, CBM6 (amino acids 330 to 451), and a dockerin type I domain (amino acids 468-524).

In addition, *axb8* was shown to be clustered on the *C. thermocellum* ATCC27405 genome, as described before by Demain *et al*, with other carbohydrate processing enzymes and uncharacterized proteins [30]. This cluster contains putative cellulosomal genes: *cthe_2193*, which encodes Ctxyl5a, a xylanase; and *cthe_2194*, which encodes a carbohydrate esterase CE1 (*cthe_2194*). The other two genes encode uncharacterized proteins with domains that are likely to be involved in hemicellulose hydrolysis; *cthe_2195*, which encodes a protein containing a CBM6 that lacks a predicted catalytic domain, and *cthe_2197*, which encodes a protein containing a GH2 domain found in enzymes presenting activity of mannosidases, galactosidases, arabinofuranosidases, glucuronidases or glucosaminidases. All proteins contain CBM6 modules and dockerin type I domains (Fig. 1).

The AxB8 predicted amino acid sequence presented the highest identity to bacterial GH43 enzymes (from 65 to 100%), from the *Firmicutes* phylum (Supplementary Table 2), all of which contain a GH43 family motif and a CBM6 domain. Among these, 12 of the top 15 hits also contain dockerin type I domains and are produced by *Clostridium* species such as *Clostridium straminisolvens*, *Clostridium sp.* Bc-is0-3, *Clostridium josui*, *Clostridium papyrosolvens*, *Clostridium sp*. BNL1100, *Clostridium cellulolyticum*, *Clostridium cellobioparum*, as well as one produced by *Acetivibrio cellulolyticus*. However, only five of

these proteins have already been annotated in the CAZy database as pertaining to subfamily 29. In addition, taking into consideration the whole table, only 14 among the 100 proteins are annotated in the CAZy database as pertaining to subfamily 29 (14).

Phylogenetic analysis showed that AxB8 and other GH43 subfamily 29 enzymes cluster mainly in bacterial phyla with the presence of additional domains, such as CBM6 and dockerin (Fig. 2). AxB8 is found in the same branch as *cthe_2196* (A3DHG9) and two putative cellulosomal GH43 proteins from *Clostridium sp.* BNL1100 (H2JDB8) and *Clostridium cellulolyticum* (B8I0L2), all containing the additional domains CBM6 and dockerin type I. Two more enzymes were found in the same node, M1MD76 from *Clostridium saccharoperbutylacetonicum* and A0A0U3AEE6 from a metagenomic library, however, these do not contain CBM and/or dockerin domains.

None of the three previously biochemically characterized GH43 subfamily 29 enzymes, Xsa43E (E0RYY0), *Cj*Abf43A (B3PD60) and Bt_0369 (Q8AAU5) presented CBM6 modules or dockerin type I domains. Xsa43E (E0RYY0) groups with CBM6-containing proteins close to the AxB8 clade, whereas, the other two GH43 subfamily 29 members grouped into a distant node (INT1) (Fig.2).

Alignment of the AxB8 sequence against structures of Xsa43E (E0RYY0), *Cj*Abf43A (B3PD60) and *Ct*Abf43A (A3DEX4) revealed conserved amino acids among GH43 family members (Fig. 3). The catalytic triad, Asp^{11}, Glu^{189} and Asp^{129}, typical of GH43 family members, was found in all sequences aligned, as well as a histidine residue at position 242, responsible for calcium binding, as described previously [10]. Four out of five prolines, which are also related to the catalytic function, were observed in Xsa43E (P^{12}, P^{130}, P^{191} and P^{243}), and in AxB8 [10]. Only the first two were detected in *Cj*Abf43A and *Ct*Abf43A. Another conserved motif among GH43 members is Trp^{73}, Arg^{74} and Pro^{75} (WAP), found in Xsa43E. In AxB8 and *Cj*Abf43A, instead of a proline, glycine and serine residues were detected. Additionally, tryptophan (W^{125}) and an asparagine (N^{146}) already described as responsible for *Cj*Abf43A binding to arabinan were also conserved in AxB8 and Xsa43E [9]. The same tryptophan (W^{125}) and a Phenylalanine residue (F^{36}), conserved in AxB8 and other GH43_29 enzymes, contribute to a curved surface cleft, as described for *Cj*Abf43A [9]. A Phe^{234} residue, present in the *Cj*Abf43A structure that also contributes to this curved surface formation, was absent from AxB8. The predicted secondary structure of AxB8 is typical for GH43 family members, characterized by 20 β-strands. Each β-sheet is composed by 4 β-strands that are arranged in a five-bladed β-propeller fold (Fig.3) [8].

Biochemical Characterization of Axb8

Heterologous AxB8 expression was detected by Western blotting in the soluble fraction of crude cell extract and after the first purification step (nickel -affinity chromatography), presenting activity against two synthetic substrates: *p*NPA and *p*NPX (data not shown) (Fig. 4a). Two chromatographic steps, his-tag affinity and ion exchange, were used for AxB8 purification, detected as a single protein band with an estimated molecular mass of 50 kDa on SDS-PAGE gel (Fig. 4b). Size-exclusion chromatography on a calibrated S200-column resulted in a single protein peak with an estimated molecular mass of ~40 kDa. Purified AxB8 showed activity against the synthetic substrates *p*NPA (89 ± 6 UI/mg) and *p*NPX (184 ± 4 UI/mg) (Table 1). No activity was detected for additional *p*NP-substrates nor against oat spelt xylan (Table 1). Furthermore, we did not detect sugar release from xylan oat spelt in synergy assays containing both AxB8 and a previously characterized endo-β-1,4-xylanase [28] .

Table 1 Substrate specificity of AxB8 on pNP substrates and oat spelt xylan.

Substrates	**Specific Activity (UI/mg)**
*p*NP-β-D-glucopyranoside	ND*
*p*NP-β-D-galactopyranoside	ND
*p*NP- β-D-mannopyranoside	ND
*p*NP-β-D-xylopyranoside (*p*NPX)	184±4
*p*NP-β-L-arabinopyranoside	ND
*p*NP-α-L-arabinopyranoside	ND
*p*NP-α-L-arabinofuranoside (*p*NPA)	89±6
Oat spelt xylan	ND

* ND : not detectable

AxB8 maximal activity was detected at 50ºC and pH 5 or pH 6 according to the substrate tested, *p*NPA or *p*NPX, respectively (Fig. 6). At pH 4.0 and 7.0 activity decreased more than 50%, but at pH 5.0 and pH 6.0 80% of activity was retained, for *p*NPA and *p*NPX respectively (Fig. 6b). Between 75% and 85% activity was maintained at temperatures of 37°C and 40°C, respectively (Fig. 6a). AxB8 thermal stability was higher at 50ºC than at 60ºC. Activity

remained at 100% even after 14 hours of pre-incubation at 50°C. However, at 60°C, it decreased to about 60% after 2 hours of pre-incubation (Fig. 6c). Regarding the effect of ions on AxB8, $CaCl_2$ (20 mM) enhanced its activity by more than 70%. Furthermore, $FeCl_3$ showed a positive effect, leading to an activity increase of more than 20%. $ZnSO_4$ and $MnSO_4$ showed almost no effect (Fig. 7). Inhibitory effects of around 20-30% were detected for other tested ions.

Kinetic parameters of purified AxB8 using *p*NPA as a substrate were K_m of 23 ± 3 mM, k_{cat} of 104 ± 7 s^{-1}, and catalytic efficiency of 4.5 ± 0.7 mM^{-1} s^{-1}. No enzyme saturation was observed using *p*NPX as a substrate, even at substrate concentration of 100 mM, therefore for this substrate we were unable to determine the kinetic parameter values.

Regarding hydrolysis of natural substrates, AxB8 hydrolysed wheat arabinoxylan (WAX), releasing mainly xylose and arabinose (Fig. 5). No products were detected for bioassays carried out using oligossacharides such as xylobiose, xylotriose, xylotetraose, and arabinan sugar beet, as substrates.

Discussion

AxB8 shows the closest homology to GH43 proteins described as lignocellulolytic from *Clostridium* species such as *C. papyrosolvens*, *C. straminisolvens*, *C. josui*, *C. cellulolyticum*, *Clostridium* sp BNL1100. These data reflect the close relationship between the genes in the *Clostridium* genus, and the ability of some of their encoded enzymes to degrade complex holocellulosic biomass [31–35]. Notable also is the number of cellulosomal GH43 enzymes of subfamily 29 (table 2) that have not yet been catalogued in the CAZy database (86 in total). The addition of these proteins to this database will certainly improve the previously proposed categorization of GH43 members into subfamilies, adding information about a novel clade of enzymes involved in adaptation to extreme environments, and with high efficiency in holocellulose deconstruction.

The grouping pattern of proteins onto the phylogenetic tree suggests a probable co-evolution between the catalytic domain (GH43), carbohydrate binding module (CBM) and the specialized conserved cellulosomal domain (dockerin type I), with the CBM6 module being an evolutionary gain, as this module is not seen close to the root of the ancestral tree [6]. The presence of the CBM domain in these enzymes was reported in previous studies to improve performance on cellulose and hemicellulose degradation by increasing enzymatic activity [36].

Gene organization in clusters as described for *axb8* is a conserved and typical feature of bacterial genomes and has been previously reported for hemicellulase encoding genes in *Clostridium* genomes. For example, the previously reported gene cluster *xyl-doc* in *C. cellulolyticum* contains 13 putative cellulosomal enzyme-encoding genes, including B8I0L2, an AxB8-homologous protein. These predicted proteins contain CBM6 modules and catalytic domains potentially involved in hemicellulose degradation [34]. This kind of organization enables the host bacteria to simultaneously express different sets of proteins or enzymes involved in specific process. This may improve degradation efficiency, as well as enabling the bacteria to act on a variety of hemicelluloses. Indeed, conservation of genes and genome organization among *Clostridium* species recognized for their ability to degrade complex lignocellulosic substrates shows the remarkable role of these clusters in determining specific phenotypes and fitness in *C. thermocellum*.

AxB8 activity against both the synthetic substrates *p*NPX and *p*NPA is in agreement with data previously shown for the characterized GH43_29 Xsa43E, and is a common feature among members of this subfamily [6,8]. Despite the activity on both substrates the higher affinity was for *p*NPA, suggesting that the main activity of AxB8 is as an α-L-arabinofuranosidase (EC 3.2.1.55). However, regarding hydrolysis of the natural substrates WAX and sugar beet arabinan, AxB8 presented a different degradation profile to that previously shown for GH43_29 members (Table 3) [10,37]. Indeed, we show that an enzyme of this family can act on the ramified natural substrate WAX, releasing mainly xylose instead of arabinose, an activity not previously described for members of this subfamily. There was no action observed against sugar beet arabinan.

AxB8, together with enzymes derived from the same gene cluster may act in a coordinated manner to remove hemicelluloses from lignocellulosic biomass (Supplementary Fig.2). In our hypothetical model we propose that AxB8 and Ctxyl5A may act on arabinose side chains, and CE1 on acetyl side chains of arabinoxylan or acetylated xylan polymers. In the xylan chain, Ctxyl5A would perform an attack in an endo- mode, and AxB8 in an exo- mode, thereby releasing xylose. Cthe_2197 could also be active on arabinose side chains.

Table 3. Substrate specificity of α-arabinofuranosidases. *C.thermocellum* B8 (AxB8), *C.thermocellum* ATCC 27405 (CtAbf43A), *Butyrivibrio proteoclasticus* B316 (Xsa43E) and *Cellvibrio japonicus ueda* 107 (CjAbf43A).

		Substrates				
Enzyme	Subfamily	*p*NP-substrates	Wheat arabinoxylan	Sugar beet arabinan	Oligosaccharides ($X_{2,3,4}$)	Oligosaccharides ($A_{2,3,4}$)
Xsa43E	29	*p*NPX/*p*NPA	+	Not tested	−	+
*Cj*Abf43A	29	*p*NPA	−	+	Not tested	+
Bt_0369	29	*p*NPA	Not tested	+	Not tested	Not tested
AxB8*	29	*p*NPX/*p*NPA	+	−	−	Not tested

* Enzyme characterized in the present report.

Another distinguishable feature of AxB8 and enzymes in the same clade in comparison to other GH42_29 subfamily members in the phylogenetic tree is the presence of the dockerin type I domain. The co-occurrence of the GH43 domain, CBM 6 and dockerin type I domain in general is not as common a feature among GH43_29 subfamily members as is GH43 and CBM6 co-occurrence.

In fact, the presence of the dockerin domain in celulosomal enzymes together AxB8´s un-usual mode of action compared to other members of GH43 subfamily 29, suggested a revision of the grouping of these enzymes within the same subfamily. The best categorization might be to group AxB8 and its homologues in a separate subfamily or a sub-subfamily, putting together all celulossomal enzymes from thermophilic bacteria. A rearrangement of the categorization proposed by Mewis et al (2016) is expected, since the GH43 group is an expanding family in terms of number of sequences, and has the lowest number of members biochemically characterized, especially concerning mode of action on natural substrates [6]. As previously shown, a strong tendency within a subfamily toward a particular enzymatic activity is observed in subfamilies with multiple characterized members [6]. However, more studies focusing on biochemical characterization of GH43 subfamily 29 is required in order for a reliable and robust classification to be achieved.

Kinetic parameters of AxB8 revealed that the K_m-value is higher than those previously described for α-L-arabinofuranosidases (EC 3.2.1.55) from *C. thermocellum*: *Ct*Abf43A (*cthe_1271*) and *Ct*Araf51A (*cthe_2548*), which are 128- and 92-fold lower, respectively (Table 3) [38,16]. In comparison to GH43_29 members Xsa43e and CjAbf43A, the K_m values are about 6-fold lower. On the other hand, the AxB8 k_{cat} value is 148-fold higher than that

determined for Xsa43e, and 22-fold higher than that described for CtAbf43A from *C. thermocellum* (Table 2) [9]. In addition, AxB8 catalytic efficiency is 22-fold higher than that of Xsa43E. Therefore higher k_{cat} value compensates higher K_m value (Table 4).

The presence of five proline residues in the active site of enzymes from subfamily 29, having a central role in the hydrolysis reaction, has been described for Xsa43E [10]. The presence of four of these residues only in AxB8 may be one of the reasons for the similar activities of both enzymes against the same pNP substrates and the natural substrate arabinoxylan (Table 4) [10]. Three residues (F^{66}, W^{164}, F^{234}) were described in *C*jAbf43A as responsible for forming a curved surface cleft that would block the accommodation of the arabinoxylan backbone, explaining the lack of enzyme activity on this polymer [9]. The absence of this last phenylalanine residue in AxB8 and Xsa43E could explain the ability of these enzymes to accommodate the arabinoxylan backbone and promote its degradation. According to these data, and as previously discussed, GH43_29 appears to be a polyspecific family [7]. Further studies are necessary to characterize the differences between these enzymes that may explain the substrate preference.

Table 4 Kinetic parameters of α-arabinofuranosidases against pNP-a-L-arabinofuranoside (pNPA). *C. thermocellum* ATCC 27405, *Butyrivibrio proteoclasticus* B316 and *Cellvibrio japonicus Ueda* 107.

Strain	Enzyme	K_m (mM)	k_{cat} (s-1)	k_{cat}/K_m (mM-1 s-1)
Clostridium thermocellum B8	AxB8*	23.04	104.20	4.5
Butyrivibrio proteoclasticus B316	Xsa43e	3.60	0.70	0.20
Cellvibrio japonicus Ueda107	*Cj*Abf43A	3.10	1896.00	611.61
Clostridium thermocellum (ATCC27405)	*Ct*Abf43A	0.18	4.72	26.22
Clostridium thermocellum (ATCC27405)	*Ct*Araf51A	0.25	103.00	412.00

* Enzyme characterized in the present report.

The optimum temperature and pH values for AxB8 (50°C; and pH between 5.0 and 6.0) are similar to values described for other GH family 43 members from *C. thermocellum*, especially *Ct*Abf43A [16,39]. However, other previously characterized GH43_29 enzymes, from mesophilic bacteria, have presented activity at temperatures between 25°C and 37 °C, and pH values around 7.0 [9,10]. Thus, AxB8 presented different optimal conditions compared to other GH43_29 subfamily enzymes produced by other bacterial species. Additionally, the $CaCl_2$-positive effect is in agreement with data previously reported for *C. thermocellum* hydrolases [16].

Conclusions

AxB8 is the first GH43_29 enzyme characterized as presenting the additional CBM6 and dockerin type I domains in a thermophilic bacterium. In addition to the different modular organization, this protein presents a different mode of action against the natural substrate WAX, a feature that may be shared with homologous proteins or proteins arranged in the same clade as AxB8. As GH43_29 is a group of proteins with a limited number of biochemically characterized members, the increase in substrate specificity data is crucial for a better understanding of its members' properties, enabling assignment of functions to putative sequences and construction of a robust classification system. This is also important in order to obtain a complete map of GH43_29 contribution to the complex process of microbial hemicellulose deconstruction.

Competing interests

The authors declare that they have no competing interests.

Funding

B.R is the recipient of a postgraduate maintenance scholarship from the Brazilian Research Council (CNPq) including a PDSE (visiting scholarship) at Wageningen University, funded by CAPES. E.F.N, is the recipient of a research fellowship from the Brazilian Research Council (CNPq). This research was supported by a grant from the University of Brasilia—UNB, CAPES, CNPq, FINEP, FAPDF and Wageningen University.

Acknowledgements

We would like to thank Richard Van Kranenburg, Mark Levisson and Tijn Daas from the Laboratory of Microbiology (WUR) for their suggestions and input into the experimental setup. We also thank to Mirjam Kabel and Patricia Murciano from the Laboratory of Food Chemistry (WUR) for their help in HPLC analysis. We thank Fernando Melo for his contribution to the phylogenetic analysis.

References

[1] L. Lin, J. Xu, Dissecting and engineering metabolic and regulatory networks of thermophilic bacteria for biofuel production, Biotechnol. Adv. 31 (2013) 827–837. doi:10.1016/j.biotechadv.2013.03.003.

[2] N. Sarkar, S.K. Ghosh, S. Bannerjee, K. Aikat, Bioethanol production from agricultural wastes: An overview, Renew. Energy. 37 (2012) 19–27. doi:http://dx.doi.org/10.1016/j.renene.2011.06.045.

[3] L.F. Silva, M.K. Taciro, G. Raicher, R.A.M. Piccoli, T.T. Mendonça, M.S.G. Lopes, J.G.C. Gomez, Perspectives on the production of polyhydroxyalkanoates in biorefineries associated with the production of sugar and ethanol, Int. J. Biol. Macromol. 71 (2014) 2–7. doi:http://dx.doi.org/10.1016/j.ijbiomac.2014.06.065.

[4] J.M.D. Paye, A. Guseva, S.K. Hammer, E. Gjersing, M.F. Davis, B.H. Davison, J. Olstad, B.S. Donohoe, T.Y. Nguyen, C.E. Wyman, S. Pattathil, M.G. Hahn, L.R. Lynd, Biological lignocellulose solubilization: comparative evaluation of biocatalysts and enhancement via cotreatment., Biotechnol. Biofuels. 9 (2016) 8. doi:10.1186/s13068-015-0412-y.

[5] V. V Zverlov, J. Kellermann, W.H. Schwarz, Functional subgenomics of Clostridium thermocellum cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes., Proteomics. 5 (2005) 3646–53. doi:10.1002/pmic.200401199.

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

[7] A. Goyal, S. Ahmed, K. Sharma, V. Gupta, P. Bule, V.D. Alves, C.M.G.A. Fontes, S. Najmudin, Molecular determinants of substrate specificity revealed by the structure of Clostridium thermocellum arabinofuranosidase 43A from glycosyl hydrolase family 43 subfamily 16, Acta Crystallogr. Sect. D Struct. Biol. 72 (2016) 1281–1289. doi:10.1107/S205979831601737X.

[8] E. Vandermarliere, T.M. Bourgois, M.D. Winn, S.V.A.N. Campenhout, G. Volckaert, J.A. Delcour, S. V Strelkov, A. Rabijns, C.M. Courtin, Structural analysis of a glycoside hydrolase family 43 arabinoxylan arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family, 47 (2009) 39–47. doi:10.1042/BJ20081256.

[9] A. Cartmell, L.S. Mckee, M.J. Pen, J. Larsbrink, H. Brumer, S. Kaneko, H. Ichinose, R.J. Lewis, A. Viksø-nielsen, H.J. Gilbert, J. Marles-wright, The Structure and Function of an Arabinan-specific α-1 , 2-Arabinofuranosidase Identified from Screening the Activities of Bacterial GH43 Glycoside Hydrolases * □, 286 (2011) 15483–15495. doi:10.1074/jbc.M110.215962.

[10] M. Till, D. Goldstone, Structural analysis of the GH43 enzyme Xsa43E from Butyrivibrio proteoclasticus, (2014) 1193–1198. doi:10.1107/S2053230X14014745.

[11] C.M. Wilson, M.R. Jr, C.M. Johnson, S.L. Martin, T.M. Chu, R.D. Wolfinger, L.J. Hauser, M.L. Land, D.M. Klingeman, M.H. Syed, A.J. Ragauskas, T.J. Tschaplinski, J.R. Mielenz, S.D. Brown, Global transcriptome analysis of Clostridium thermocellum ATCC 27405 during growth on dilute acid pretreated Populus and switchgrass, (2013) 1–18.

[12] H. Wei, Y. Fu, L. Magnusson, J.O. Baker, P.-C. Maness, Q. Xu, S. Yang, A. Bowersox, I. Bogorad, W. Wang, M.P. Tucker, M.E. Himmel, S.-Y. Ding, Comparison of transcriptional profiles of Clostridium thermocellum grown on cellobiose and pretreated yellow poplar using RNA-Seq., Front. Microbiol. 5 (2014) 142. doi:10.3389/fmicb.2014.00142.

[13] N.D. Gold, V.J.J. Martin, Global view of the Clostridium thermocellum cellulosome revealed by quantitative proteomic analysis., J. Bacteriol. 189 (2007) 6787–95. doi:10.1128/JB.00882-07.

[14] J. Krauss, V. V Zverlov, W.H. Schwarz, In vitro reconstitution of the complete Clostridium thermocellum cellulosome and synergistic activity on crystalline cellulose., Appl. Environ. Microbiol. 78 (2012) 4301–7. doi:10.1128/AEM.07959-11.

[15] S.P. Das, R. Ravindran, S. Ahmed, D. Das, D. Goyal, C.M.G.A. Fontes, A. Goyal, Bioethanol production involving recombinant C. thermocellum hydrolytic hemicellulase and fermentative microbes, Appl. Biochem. Biotechnol. 167 (2012) 1475–1488. doi:10.1007/s12010-012-9618-7.

[16] S. Ahmed, A.S. Luis, J.L.A. Bras, A. Ghosh, S. Gautam, M.N. Gupta, C.M.G.A. Fontes, A. Goyal, A Novel a -L-Arabinofuranosidase of Family 43 Glycoside Hydrolase (Ct43Araf) from Clostridium thermocellum, 8 (2013) 1–10. doi:10.1371/journal.pone.0073575.

[17] D. Jiang, J. Fan, X. Wang, Y. Zhao, B. Huang, J. Liu, X.C. Zhang, Crystal structure of 1,3Gal43A, an exo-β-1,3-galactanase from Clostridium thermocellum, J. Struct. Biol. 180 (2012) 447–457. doi:10.1016/j.jsb.2012.08.005.

[18] K.O. Osiro, B.R. de Camargo, R. Satomi, P.R. V. Hamann, J.P. Silva, M.V. de Sousa, B.F. Quirino, E.N. Aquino, C.R. Felix, A.M. Murad, E.F. Noronha, Characterization of Clostridium thermocellum (B8) secretome and purified cellulosomes for lignocellulosic biomass degradation, Enzyme Microb. Technol. 97 (2017) 43–54. doi:10.1016/j.enzmictec.2016.11.002.

[19] P.R. V. Hamann, D.L. Serpa, A.S. Barreto da Cunha, B.R. de Camargo, K.O. Osiro, M. Valle de Sousa, C.R. Felix, R.N.G. Miller, E.F. Noronha, Evaluation of plant cell wall degrading enzyme production by Clostridium thermocellum B8 in the presence of raw agricultural wastes, Int. Biodeterior. Biodegradation. 105 (2015) 97–105. doi:10.1016/j.ibiod.2015.08.013.

[20] P.R. V. Hamann, D.L. Serpa, A.S. Barreto da Cunha, B.R. de Camargo, K.O. Osiro, M. Valle de Sousa, C.R. Felix, R.N.G. Miller, E.F. Noronha, Evaluation of plant cell wall degrading enzyme production by Clostridium thermocellum B8 in the presence of raw agricultural wastes, Int. Biodeterior. Biodegradation. 105 (2015) 97–105. doi:10.1016/j.ibiod.2015.08.013.

[21] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res. 42 (2014) 490–495. doi:10.1093/nar/gkt1178.

[22] A. Bateman, M.J. Martin, C. O'Donovan, M. Magrane, R. Apweiler, E. Alpi, R. Antunes, J. Arganiska, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, G. Chavali, E. Cibrian-Uhalte, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, P. Gane, L.G. Castro, P. Garmiri, E. Hatton-Ellis, R. Hieta, R. Huntley, D. Legge, W. Liu, J. Luo, A. Macdougall, P. Mutowo, A. Nightingale, S. Orchard, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R. Saidi, T. Sawford, A. Shypitsyna, E. Turner, V. Volynkin, T. Wardell, X. Watkins, H. Zellner, A. Cowley, L. Figueira, W. Li, H. McWilliam, R. Lopez, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. Argoud-Puy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M.C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E. Coudert, B. Cuche, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, F. Jungo, G. Keller, V. Lara, P. Lemercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, N. Nouspikel, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, B. Roechert, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, A.L. Veuthey, C.H. Wu, C.N. Arighi, L. Arminski, C. Chen, Y. Chen, J.S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D.A. Natale, B.E. Suzek, C.R. Vinayaka, Q. Wang, Y. Wang, L.S. Yeh, M.S. Yerramalla, J. Zhang, UniProt: A hub for protein information, Nucleic Acids Res. 43 (2015) D204–D212. doi:10.1093/nar/gku989.

[23] C. Notredame, D. Higgins, J. Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment., J. Mol. Biol. 302 (2000) 205–217. doi:10.1006/jmbi.2000.4042.

[24] R.D. Finn, P. Coggill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G.A. Salazar, J. Tate, A. Bateman, The Pfam protein families database: Towards a more sustainable future, Nucleic Acids Res. 44 (2016) D279–D285. doi:10.1093/nar/gkv1344.

[25] Z. He, H. Zhang, S. Gao, M.J. Lercher, W.H. Chen, S. Hu, Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees, Nucleic Acids Res. 44 (2016) W236–W241. doi:10.1093/nar/gkw370.

[26] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4., Nature. 227 (1970) 680–685.

[27] G.L. Miller, Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar, Anal. Chem. 31 (1959) 426–428. doi:10.1021/ac60147a030.

[28] M.J.A. Daas, P.M. Martínez, A.H.P. van de Weijer, J. van der Oost, W.M. de Vos, M.A. Kabel, R. van Kranenburg, Biochemical characterization of the xylan hydrolysis profile of the extracellular endo-xylanase from Geobacillus thermodenitrificans T12, BMC Biotechnol. 17 (2017) 44. doi:10.1186/s12896-017-0357-2.

[29] M. Martínez, Patricia, R. Bakker, P. Harmsen, H. Gruppen, M. Kabel, Importance of acid or alkali concentration on the removal of xylan and lignin for enzymatic cellulose hydrolysis, Ind. Crops Prod. 64 (2015) 88–96. doi:10.1016/j.indcrop.2014.10.031.

[30] A.L. Demain, M. Newcomb, J.H.D. Wu, Cellulase, Clostridia, and Ethanol, Microbiol. Mol. Biol. Rev. 69 (2005) 124–154. doi:10.1128/MMBR.69.1.124–154.2005.

[31] S. Kato, S. Haruta, Z.J. Cui, M. Ishii, A. Yokota, Y. Igarashi, Clostridium straminisolvens sp. nov., a moderately thermophilic, aerotolerant and cellulolytic bacterium isolated from a cellulose-degrading bacterial community, Int. J. Syst. Evol. Microbiol. 54 (2004) 2043–2047. doi:10.1099/ijs.0.63148-0.

[32] J. Sukhumavas, K. Ohmiya, S. Shimizu, K. Uen, K. Ueno, Species from Thai Compost, Int. J. Syst. Bacteriol. 38 (1988) 179–182.

[33] S. Siozios, A. Cestaro, R. Kaur, I. Pertot, O. Rota-Stabelli, G. Anfora, Draft Genome Sequence of the, Genome Annunc. 1 (2013) e00032–13. doi:10.1128/genomeA.00032-13.Copyright.

[34] H. Celik, J.C. Blouzard, B. Voigt, D. Becher, V. Trotter, H.P. Fierobe, C. Tardif, S. Pagès, P. de Philip, A Two-Component System (XydS/R) Controls the Expression of Genes Encoding CBM6-Containing Proteins in Response to Straw in Clostridium cellulolyticum, PLoS One. 8 (2013). doi:10.1371/journal.pone.0056063.

[35] L.L. Li, S. Taghavi, J.A. Izquierdo, D. van der Lelie, Complete genome sequence of Clostridium sp. strain BNL1100, a cellulolytic mesophile isolated from corn stover, J. Bacteriol. 194 (2012) 6982–6983. doi:10.1128/JB.01908-12.

[36] M. Foumani, T. V Vuong, B. MacCormick, E.R. Master, Enhanced Polysaccharide Binding and Activity on Linear beta-Glucans through Addition of Carbohydrate-Binding Modules to Either Terminus of a Glucooligosaccharide Oxidase., PLoS One. 10 (2015) e0125398. doi:10.1371/journal.pone.0125398.

[37] M.A.S. Correia, K. Mazumder, J.L.A. Brás, S.J. Firbank, Y. Zhu, R.J. Lewis, W.S. York, C.M.G.A. Fontes, H.J. Gilbert, Structure and function of an arabinoxylan-specific xylanase, J. Biol. Chem. 286 (2011) 22510–22520. doi:10.1074/jbc.M110.217315.

[38] E.J. Taylor, N.L. Smith, J.P. Turkenburg, S. D'Souza, H.J. Gilbert, G.J. Davies, Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Araf51, from Clostridium thermocellum., Biochem. J. 395 (2006) 31–7. doi:10.1042/BJ20051780.

[39] M. Sajjad, M.I.M. Khan, N.S. Akbar, S. Ahmad, I. Ali, M.W. Akhtar, Enhanced expression and activity yields of Clostridium thermocellum xylanases without non-catalytic domains, J. Biotechnol. 145 (2010) 38–42. doi:10.1016/j.jbiotec.2009.10.013.

Figure captions

Fig.1. Organisation of the gene cluster surrounding *cthe_2196* in *C.thermocellum* ATCC 27405. Gene locus showing *cthe_2193*, *cthe_2194*, *cthe_2195*, *cthe_2196* and *cthe_2197*. The boxes indicate different domains: GH5 (Glycoside hydrolase family 5), CBM6, (carbohydrate binding module 6), D1 (Dockerin type I), Abh (Alpha/Beta Hydrolase), CE1 (Carbohydrate esterase 1), Bhelix (structural domain), GH43 (Glycoside hydrolase family 43), GH2 (Glycoside hydrolase family 2).

Fig.2. Phylogenetic tree of GH43 domains of subfamily 29. Annotated pfam domains of full sequences are represented: Glyco_hydro_43 (PF04616), Big_4 (PF07532), F5_F8_type_C (PF00754), AbfB (PF05270), SLH (PF00395), Big_2 (PF02368), RicinB_lectin_2 (PF14200), Big_3 (PF07523), Laminin_G_3 (PF13385), CBM_6 (PF03422), Dockerin_1 (PF00404), Melibiase_2 (PF16499), CBM_2 (PF00553), CBM_4_9 (PF02018), RicinB_lectin (PF00652), Esterase (PF00756). Proteins are highlighted: AxB8 (dark grey), A3DHG9 corresponding to Cthe_2196 (light grey) and E0RYY0 corresponding to Xsa43E (blue). Internal nodes can be viewed (INT1, INT2) and supressed sequences are in Supplementary Fig.1.

Fig.3. Multiple sequence alignment of AxB8. Predicted AxB8 sequence from GH43_29 subfamily from *C. thermocellum* B8 was aligned with three other GH43_29 members: Xsa43E from *Butyrivibrio proteoclasticus* (PDB entry 4 nov), Cthe_2196 from *C. thermocellum*, *Cj*Abf43A from *Cellvibrio japonicus*, and *Ct*Abf43A from GH43_16, from *C. thermocellum*. Amino acids at equivalent positions are represented in the respective structures. β-strand secondary structures can be viewed.

Fig.4. ***SDS-PAGE and Western blot of protein fractions of AxB8. (a) Western blot of AxB8. Lane1: Marker, Lane2: Crude extract (Soluble). Lane3: Pool of fractions collected after His-trap column chromatographic procedure. (b) Lane1: Molecular mass marker, Lane2: Crude extract (Soluble fraction), Lane3: purified AxB8 after His-trap-Q-FF column.***

Fig. 5. ***HPLC anion exchange chromatogram of wheat arabinoxylan (WAX). WAX and AxB8 after incubation, replicate 1 (a) and 2 (b). WAX incubated without AxB8 (c).***

Fig. 6. Temperature-dependence, pH-dependence and thermostability of AxbB. **a** Effect of temperature using *p*NP-α-L-arabinofuranoside (pH 5.0) (□) and *p*NP- β-xylopyranoside (pH 6.0) (■) as substrates; **b** Effect of pH at 50°C using McIlvaine buffer and *p*NP- α-L-arabinofuranoside (○) and β-xylopyranoside (●), as substrates. **c** Thermostability at 50°C (◊) and 60°C (Δ) using pNP-β-xylopyranoside as substrate.

Fig. 7. Relative activity (%) of AxB8 in the presence of salts. Activities were normalized to enzyme activity without added salts.

Supplementary data

Supplementary Fig. S1. Phylogenetic tree of GH43 domains of subfamily 29. Annotated pfam domains of full sequences are represented: Glyco_hydro_43 (PF04616), Big_4 (PF07532), F5_F8_type_C (PF00754), AbfB (PF05270), SLH (PF00395), Big_2 (PF02368), RicinB_lectin_2(PF14200), Big_3 (PF07523), Laminin_G_3 (PF13385), CBM_6 (PF03422), Dockerin_1 (PF00404), Melibiase_2 (PF16499), CBM_2 (PF00553), CBM_4_9 (PF02018), RicinB_lectin (PF00652), Esterase (PF00756). Numbers in the figure are amino acid positions in the protein structure, according to domain presence. Proteins are highlighted: AxB8 (dark grey), E0RYY0 corresponding to Xsa43E (blue), Q8AAU5 corresponding to Bt_0369 (green) and B3PD60 corresponding to CjAbf43A (green).

Supplementary Fig. S2. Schematic figure showing confirmed characterized activity of AxB8 and CtXyl5A, and probable activity of proteins encoded by the genes *cthe_2194*, *cthe_2195* and *cthe_2197*. The polymer is xylan (X-xylose), with side residues of Arabinose (Ara), and Acetyl (Ac). Stars represent enzymes, and dotted lines represent cleavage by the enzymes.

Supplementary Table. S1. Bacterial sequence list downloaded from the CAZy database, GH43 family, subfamily 29. Uniprot acession numbers were accessed using the GenBank accession number.

Supplementay Table S2. Blast result retrieved from NCBI psi-blast using predicted AxB8 as a query.

Supplementary Table S3. Final sequence list used for construction of the phylogenetic tree. Uniprot acession numbers and Pfam annotations are listed.

Fig.1.

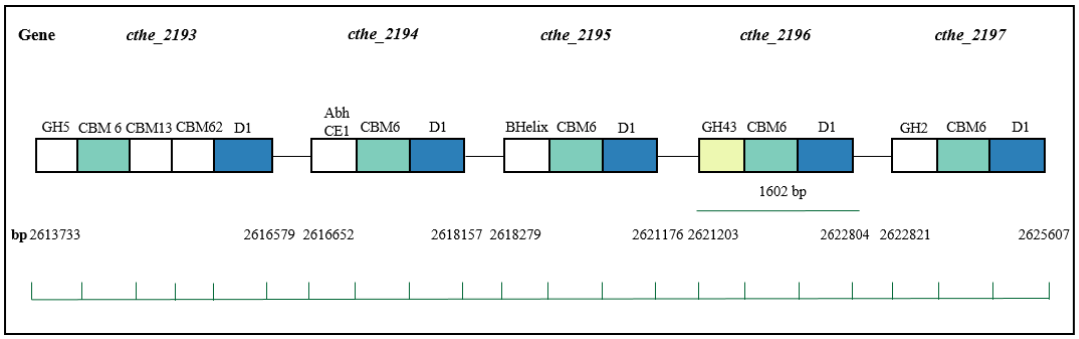

Gene
cthe_2193
cthe_2194
cthe_2195
cthe_2196
cthe_2197
GH5 CBM 6 CBM13 CBM62 D1
Abh CE1 CBM6 D1
BHelix CBM6 D1
GH43 CBM6 D1
GH2 CBM6 D1
1602 bp
bp 2613733
2616579 2616652
2618157 2618279
2621176 2621203
2622804 2622821
2625607

Fig.2.

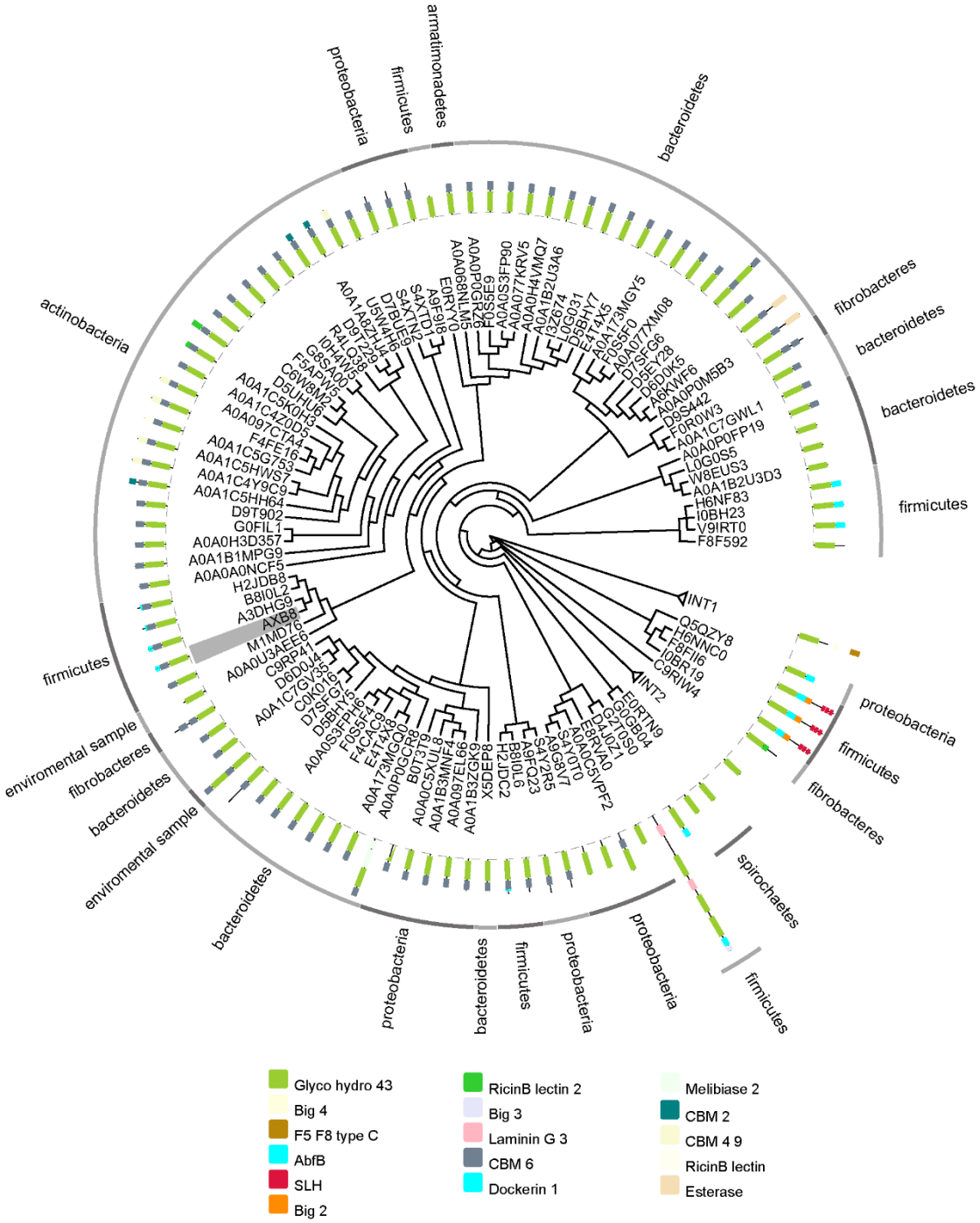

armatimonadetes
bacteroidetes
fibrobacteres
bacteroidetes
bacteroidetes
firmicutes
proteobacteria
firmicutes
actinobacteria
firmicutes
enviromental sample
fibrobacteres
bacteroidetes
enviromental sample
bacteroidetes
proteobacteria
bacteroidetes
firmicutes
proteobacteria
proteobacteria
firmicutes
spirochaetes
proteobacteria
firmicutes
fibrobacteres
Glyco hydro 43
Big 4
F5 F8 type C
AbfB
SLH
Big 2
RicinB lectin 2
Big 3
Laminin G 3
CBM 6
Dockerin 1
Melibiase 2
CBM 2
CBM 4 9
RicinB lectin
Esterase

Fig.3.

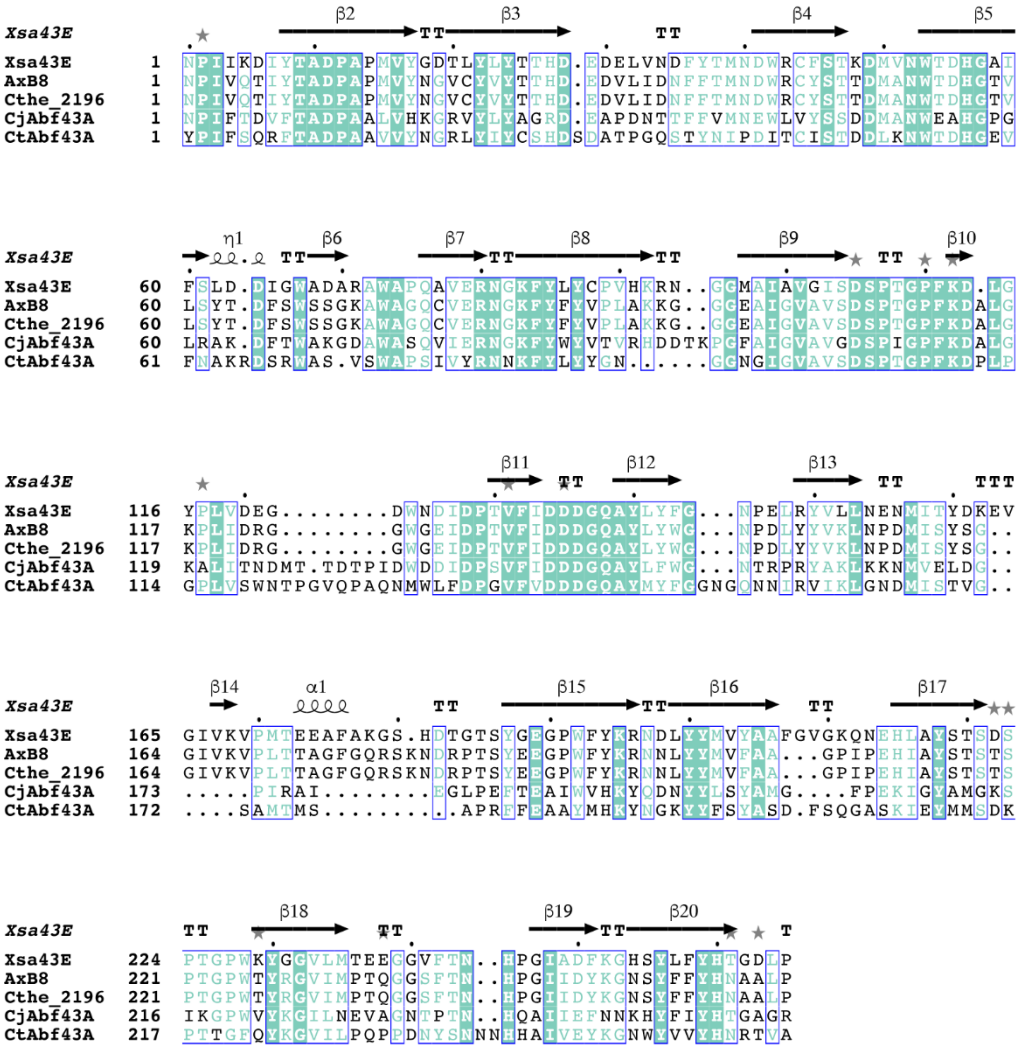

Xsa43E
Xsa43E 1
AxB8 1
Cthe_2196 1
CjAbf43A 1
CtAbf43A 1
Xsa43E 60
AxB8 60
Cthe_2196 60
CjAbf43A 60
CtAbf43A 61
Xsa43E 116
AxB8 117
Cthe_2196 117
CjAbf43A 119
CtAbf43A 114
Xsa43E 165
AxB8 164
Cthe_2196 164
CjAbf43A 173
CtAbf43A 172
Xsa43E 224
AxB8 221
Cthe_2196 221
CjAbf43A 216
CtAbf43A 217

Fig.4.

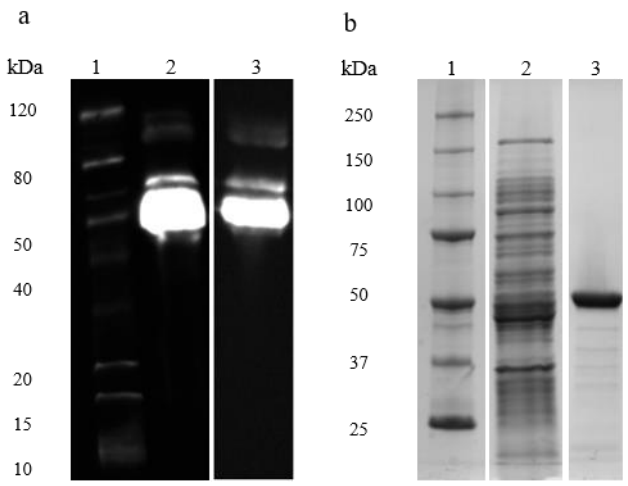

a
kDa
1
2
3
120
80
50
40
20
15
10
b
kDa
1
2
3
250
150
100
75
50
37
25

Fig.5.

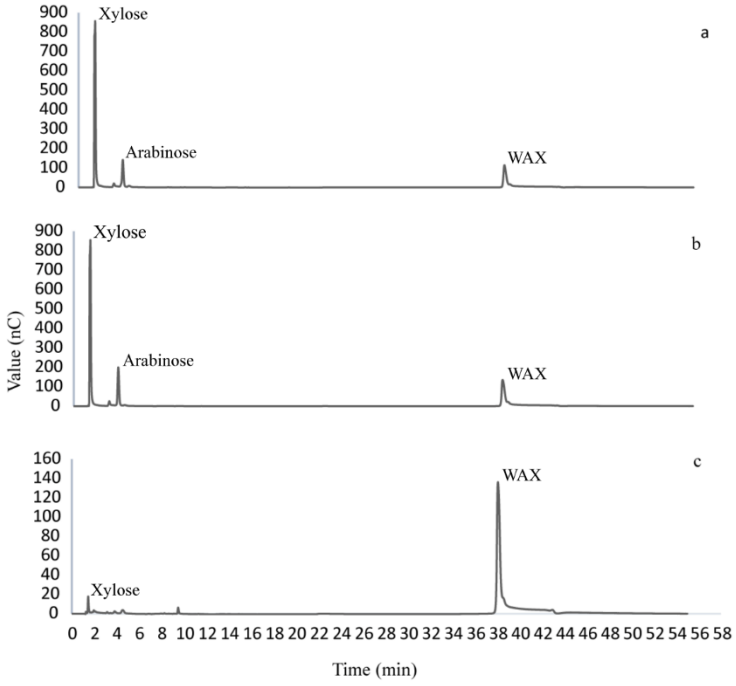

a
Xylose
Arabinose
WAX
b
Xylose
Arabinose
WAX
Value (nC)
c
WAX
Xylose
Time (min)

Fig.6.

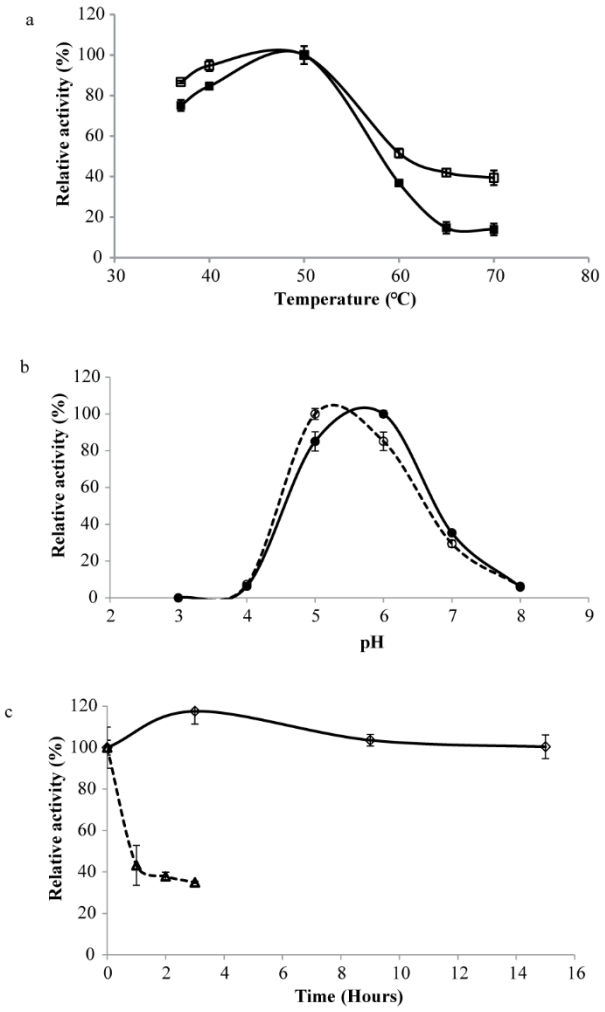
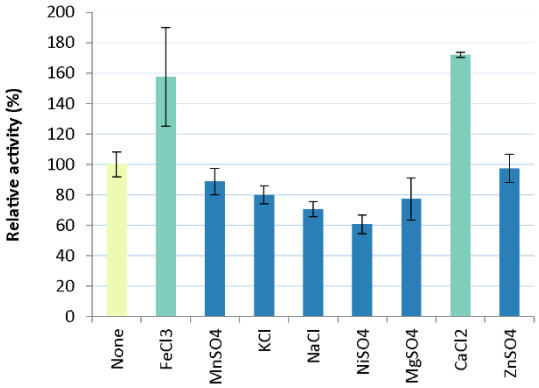
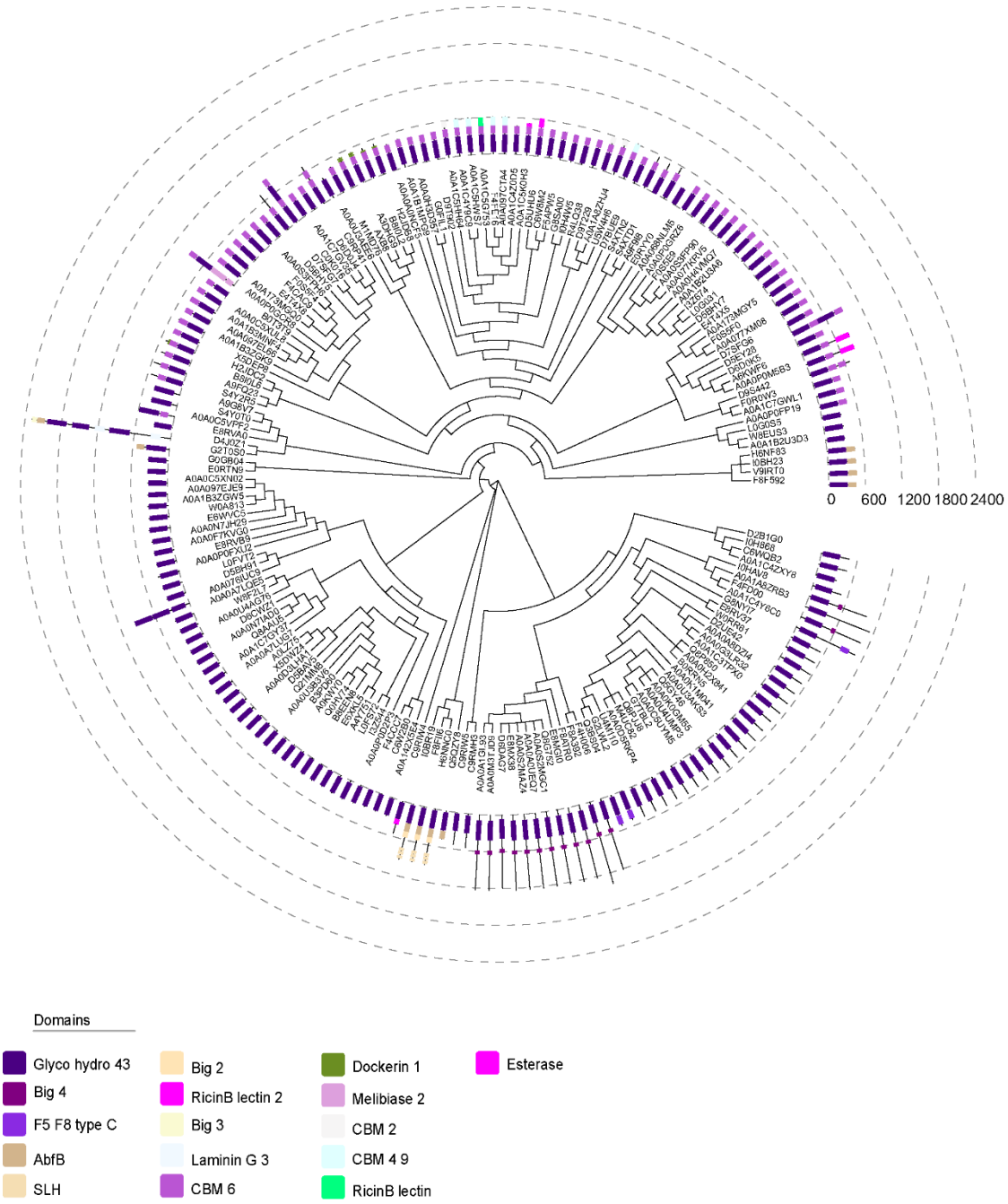

a
Relative activity (%)
Temperature (°C)
b
Relative activity (%)
pH
c
Relative activity (%)
Time (Hours)

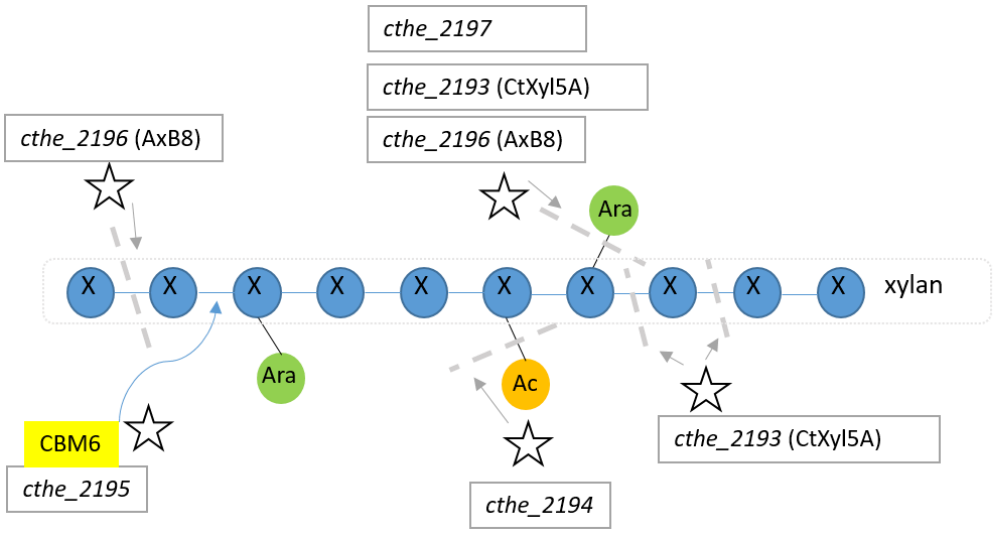
Fig.7.


Relative activity (%)
200
180
160
140
120
100
80
60
40
20
0
None
FeCl3
MnSO4
KCl
NaCl
NiSO4
MgSO4
CaCl2
ZnSO4

S.Fig.1


0
600
1200
1800
2400
Domains
Glyco hydro 43
Big 4
F5 F8 type C
AbfB
SLH
Big 2
RicinB lectin 2
Big 3
Laminin G 3
CBM 6
Dockerin 1
Melibiase 2
CBM 2
CBM 4 9
RicinB lectin
Esterase

S.Fig.2


cthe_2197
cthe_2193 (CtXyl5A)
cthe_2196 (AxB8)
cthe_2196 (AxB8)
Ara
X
xylan
Ara
Ac
CBM6
cthe_2195
cthe_2193 (CtXyl5A)
cthe_2194

Capítulo 5. Conclusão e perspectivas

Clostridium thermocellum é uma bactéria que tem se mostrado em diversos estudos disponíveis na literatura, a capacidade de degradar diferentes fontes de carbono, desde substratos puros, como celulose, até biomassa complexas, como substratos lenhosos, gramínea, e como mostrado nesse trabalho e outros artigos publicados pelo grupo, bagaço e palha de cana.

Por meio das análises proteômica e transcritômica foi possível identificar proteínas secretadas por *C. thermocellum* quando em crescimento em diferentes fontes (celulose, bagaço e palha de cana), e principalmente identificar proteínas carboidrato ativas. Dos 12 genes que codificam proteínas carboidrato ativas e celulossomais, que foram regulados positivamente em palha ou bagaço de cana, dois foram encontrados na análise proteômica em bagaço de cana, os genes *cthe_0798* que codificam um acetil xilana esterase e o gene *cthe_1398*, que codifica XghA, uma xiloglucanase. O gene, *cthe_0798* teve a maior regulação, e a proteína foi exclusivamente encontrada para palha e bagaço, sendo ausente em celulose. Esses dados indicam a importância dessa proteína na degradação de bagaço de cana, e supostamente palha de cana. Além dessas proteínas, outros genes encontrados positivamente regulados, revelam sua importância, na degradação da biomassa, como CelC, CelP, XynA e ManA.

Além disso, por meio de genes abundantes relacionados, como CelS, e CelK, demonstrados no transcritoma para todas as fontes de carbono, e também encontradas no proteoma, podemos também inferir a importância dessas enzimas não somente na degradação de celulose, mas em palha e bagaço de cana. Portanto, em uma futura elaboração de por exemplo um coquetel enzimático para degradação dessas biomassas utilizadas nesse trabalho, e em outras com composições semelhantes, essas enzimas citadas seriam selecionadas como essenciais para o sucesso na degradação.

Além disso, por meio dos dados obtidos aqui, podemos selecionar vários genes, que seriam de grande interesse para expressão heteróloga e sua caracterização bioquímica, devido à sua expressão nessas biomassas, como os genes codificadores das proteínas cobre amino oxidases putativas, e tantos outros genes e proteínas encontradas por somente apresentar domínios de ligação ao carboidrato (CBM) ou doquerina do tipo I.

Outras vias metabólicas pouco discutidas e nesse trabalho foram melhor descritas, são os genes envolvidos na quimiotaxia, motilidade, quorum sensing e formação de biofilme. Todos esses mecanismos apresentaram uma regulação positiva nas biomassas, e são de extrema

importância na sinalização sensorial de nutrientes, para a bactéria, e sua aproximação ao substrato. Estudos mais aprofundados devem ser feitos, já que pouco desses mecanismos são descritos para bactérias termófilas.

Além disso, a bactéria *Moorella thermoacética*, apresentou crescimento juntamente com *C.thermocellum* no meio de cultivo, e apresentou genes diferencialmente expressos nos meios de bagaço e palha, a maioria com regulação positiva. Entre esses genes, estão presentes genes que codificam transportadores de ribose, e enzimas que atuam na fosforilação desse açúcar quando dentro na célula. Essa regulação pode ser um indício da captação de xiloses por *M.thermoacetica*, sendo que estes açúcares foram liberados pela atividade enzimática de *C.thermocellum*. Poucos estudos demonstram a co-cultura desses dois organismos, porém as vias reguladas dessa sintropia não foram descritas, além do pouco material na literatura existente para *M.thermoacetica*, o que torna essa análise de grande interesse, desde que as duas bactérias se beneficiam no meio de cultivo.

Uma proteína anotada como pertencente à família GH43 de *C.thermocellum* foi caracterizada nesse trabalho (AxB8) e descrita por ter maior similaridade com proteínas anotadas por outros termófilos não caracterizados. A localização do gene que codifica essa proteína em operon com o gene *cthe_2195*, além desse mesmo operon estar inserido em um cluster com proteínas que são potencialmente todas hemicelulases, propõe que Cthe_2195 pode também ter alguma função relacionada, e talvez até algum novo domínio catalítico. A construção de plasmídeos pET com o gene *chte_2195*, e outro plasmídeo contendo uma mini CipA com apenas 2 coesinas, foram construídos com o objetivo de construir um mini celulossoma contendo *cthe_2195* e *cthe_2196* (*axb8*), com o intuito de avaliar a atividade das duas e possivelmente identificando a atividade de Cthe_2195. Futuros experimentos serão realizados para expressão dessas proteínas e análise da atividade,

Concluindo, apesar de *C.thermocellum* ser uma bactéria que vem sendo estudada há mais de 80 anos, muitas proteínas e novos genes estão constantemente sendo descobertos, além de novos mecanismos elucidados, tornando essa temática de grande interesse no mundo científico.

Referências

[1] P.L. de A. Coutinho, J.V. Bomtempo, Uso de roadmaps tecnológicos para favorecer o ambiente de inovaçao: uma proposta em matérias primas renováveis, Simpoi. (2010) 1–16.

[2] E.F. Bosma, J. Van Der Oost, W.M. De Vos, R. Van Kranenburg, Sustainable Production of Bio-Based Chemicals by Extremophiles, (2014) 360–379.

[3] M. a Lima, L.D. Gomez, C.G. Steele-King, R. Simister, O.D. Bernardinelli, M. a Carvalho, C. a Rezende, C. a Labate, E.R. Deazevedo, S.J. McQueen-Mason, I. Polikarpov, Evaluating the composition and processing potential of novel sources of Brazilian biomass for sustainable biorenewables production., Biotechnol. Biofuels. 7 (2014) 10. doi:10.1186/1754-6834-7-10.

[4] EMBRAPA Agroenergia, Biorrefinarias, (2011) 1–6.

[5] A.M. Teekens, M.E. Bruins, J.M. van Kasteren, W.H. Hendriks, J.P. Sanders, Synergy between bio-based industry and the feed industry through biorefinery, J. Sci. Food Agric. 96 (2016) 2603–2612. doi:10.1002/jsfa.7596.

[6] H.-T. Tan, K.R. Corbin, G.B. Fincher, Emerging Technologies for the Production of Renewable Liquid Transport Fuels from Biomass Sources Enriched in Plant Cell Walls, Front. Plant Sci. 7 (2016) 1–18. doi:10.3389/fpls.2016.01854.

[7] P. Tecnico, Parte ii, (n.d.) 63–74.

[8] J.M. Cruz, J.M. Domínguez, H. Domínguez, J.C. Parajó, Preparation of fermentation media from agricultural wastes and their bioconversion into xylitol, Food Biotechnol. 14 (2000) 79–97. doi:10.1080/08905430009549981.

[9] CONAB, Companhia Nacional de Abastecimento - Cana-de- açúcar, 3 (2016) 72.

[10] A SUSTENTABILIDADE NO SETOR SUCROENERGÉTICO BRASILEIRO, (n.d.). http://www.unica.com.br/sustentabilidade/.

[11] F.F. Furlan, C.B.B. Costa, G.D.C. Fonseca, R.D.P. Soares, A.R. Secchi, A.J.G. Da Cruz, R.D.C. Giordano, Assessing the production of first and second generation bioethanol from sugarcane through the integration of global optimization and process detailed modeling, Comput. Chem. Eng. 43 (2012) 1–9. doi:10.1016/j.compchemeng.2012.04.002.

[12] M.S.G. Lopes, J.G.C. Gomez, M.K. Taciro, T.T. Mendon??a, L.F. Silva, Polyhydroxyalkanoate biosynthesis and simultaneous remotion of organic inhibitors from sugarcane bagasse hydrolysate by Burkholderia sp., J. Ind. Microbiol. Biotechnol. 41 (2014) 1353–1363. doi:10.1007/s10295-014-1485-5.

[13] cana-de-açúcar, (n.d.). http://www.inee.org.br/biomassa_cana.asp?Cat=biomassa.

[14] Y. Sun, J. Cheng, Hydrolysis of lignocellulosic materials for ethanol production: a review., Bioresour. Technol. 83 (2002) 1–11. http://www.ncbi.nlm.nih.gov/pubmed/12058826.

[15] Y. Wang, C. Fan, H. Hu, Y. Li, D. Sun, Y. Wang, L. Peng, Genetic modification of plant cell walls to enhance biomass yield and biofuel production in bioenergy crops, Biotechnol. Adv. 34 (2016) 997–1017. doi:10.1016/j.biotechadv.2016.06.001.

[16] B.C. Saha, Hemicellulose bioconversion, J. Ind. Microbiol. Biotechnol. 30 (2003) 279–291. doi:10.1007/s10295-003-0049-x.

[17] L.R. Lynd, OVERVIEW AND EVALUATION OF FUEL ETHANOL FROM CELLULOSIC BIOMASS: Technology, Economics, the Environment, and Policy, Annu. Rev. Energy Environ. 21 (1996) 403–465. doi:10.1146/annurev.energy.21.1.403.

[18] L.R. Lynd, OVERVIEW AND EVALUATION OF FUEL ETHANOL FROM CELLULOSIC BIOMASS: Technology, Economics, the Environment, and Policy, Annu. Rev. Energy Environ. 21 (1996) 403–465. doi:10.1146/annurev.energy.21.1.403.

[19] R. Ravindran, A.K. Jaiswal, A comprehensive review on pre-treatment strategy for lignocellulosic food industry waste: Challenges and opportunities., Bioresour. Technol. 199 (2016) 92–102. doi:10.1016/j.biortech.2015.07.106.

[20] A. Duval, M. Lawoko, A review on lignin-based polymeric, micro- and nano-structured materials, React. Funct. Polym. 85 (2014) 78–96. doi:10.1016/j.reactfunctpolym.2014.09.017.

[21] M.A.S. Correia, K. Mazumder, J.L.A. Brás, S.J. Firbank, Y. Zhu, R.J. Lewis, W.S. York, C.M.G.A. Fontes, H.J. Gilbert, Structure and function of an arabinoxylan-specific xylanase, J. Biol. Chem. 286 (2011) 22510–22520. doi:10.1074/jbc.M110.217315.

[22] M.B.G. Latarullo, E.Q.P. Tavares, G.P. Maldonado, D.C.C. Leite, M.S. Buckeridge, Pectins, Endopolygalacturonases, and Bioenergy, Front. Plant Sci. 7 (2016) 1–7. doi:10.3389/fpls.2016.01401.

[23] J.H. Grabber, How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies, Crop Sci. 45 (2005) 820–831. doi:10.2135/cropsci2004.0191.

[24] P. Alvira, E. Tomás-Pejó, M. Ballesteros, M.J. Negro, Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review., Bioresour. Technol. 101 (2010) 4851–61. doi:10.1016/j.biortech.2009.11.093.

[25] Q.Q. Tian, L. Liang, M.J. Zhu, Enhanced biohydrogen production from sugarcane bagasse by Clostridium thermocellum supplemented with CaCO3, Bioresour. Technol. 197 (2015) 422–428. doi:10.1016/j.biortech.2015.08.111.

[26] M. Tuomela, M. Vikman, A. Hatakka, M. It??vaara, Biodegradation of lignin in a compost environment: A review, Bioresour. Technol. 72 (2000) 169–183. doi:10.1016/S0960-8524(99)00104-2.

[27] L.M. Prescott, P.H. Klein, Microbiology, Prescott, 2002.

[28] M. Embree, J.K. Liu, M.M. Al-Bassam, K. Zengler, Networks of energetic and metabolic interactions define dynamics in microbial communities, Proc. Natl. Acad. Sci. 112 (2015) 201506034. doi:10.1073/pnas.1506034112.

[29] P. Turner, G. Mamo, E.N. Karlsson, Potential and utilization of thermophiles and thermostable enzymes in biorefining., Microb. Cell Fact. 6 (2007) 9. doi:10.1186/1475-2859-6-9.

[30] B.P. Hedlund, S.K. Murugapiran, T.W. Alba, A. Levy, J.A. Dodsworth, G.B. Goertz, N. Ivanova, T. Woyke, Uncultivated thermophiles: Current status and spotlight on "Aigarchaeota," Curr. Opin. Microbiol. 25 (2015) 136–145. doi:10.1016/j.mib.2015.06.008.

[31] L. Lin, J. Xu, Dissecting and engineering metabolic and regulatory networks of thermophilic bacteria for biofuel production, Biotechnol. Adv. 31 (2013) 827–837. doi:10.1016/j.biotechadv.2013.03.003.

[32] T.K. Ng, T.K. Weimer, J.G. Zeikus, Cellulolytic and physiological properties of Clostridium thermocellum., Arch. Microbiol. 114 (1977) 1–7.

[33] R.H. McBEE, The anaerobic thermophilic cellulolytic bacteria., Bacteriol. Rev. 14 (1950) 51–63.

[34] J. a Viljoen, E.B. Fred, W.H. Peterson, The fermentation of cellulose by thermophilic bacteria, J. Agric. Sci. 16 (1926) 1–17. doi:10.1017/S0021859600088249.

[35] L. Feinberg, J. Foden, T. Barrett, K.W. Davenport, D. Bruce, C. Detter, R. Tapia, C. Han, A. Lapidus, S. Lucas, J.F. Cheng, S. Pitluck, T. Woyke, N. Ivanova, N. Mikhailova, M. Land, L. Hauser, D.A. Argyros, L. Goodwin, D. Hogsett, N. Caiazza, Complete genome sequence of the cellulolytic thermophile Clostridium thermocellum DSM1313, J. Bacteriol. 193 (2011) 2906–2907. doi:10.1128/JB.00322-11.

[36] D.E. Koeck, D. Wibberg, T. Koellmeier, J. Blom, S. Jaenicke, A. Winkler, A. Albersmeier, V. V. Zverlov, A. Pühler, W.H. Schwarz, A. Schlüter, Draft genome sequence of the cellulolytic Clostridium thermocellum wild-type strain BC1 playing a role in cellulosic biomass degradation, J. Biotechnol. 168 (2013) 62–63. doi:10.1016/j.jbiotec.2013.08.011.

[37] S.D. Brown, R. Lamed, E. Morag, I. Borovok, Y. Shoham, D.M. Klingeman, C.M. Johnson, Z. Yang, M.L. Land, S.M. Utturkar, M. Keller, E.A. Bayer, Draft genome sequences for Clostridium thermocellum wild-type strain YS and derived cellulose adhesion-defective mutant strain AD2, J. Bacteriol. 194 (2012) 3290–3291. doi:10.1128/JB.00473-12.

[38] C.M. Wilson, M.R. Jr, C.M. Johnson, S.L. Martin, T.M. Chu, R.D. Wolfinger, L.J. Hauser, M.L. Land, D.M. Klingeman, M.H. Syed, A.J. Ragauskas, T.J. Tschaplinski, J.R. Mielenz, S.D. Brown, Global transcriptome analysis of Clostridium thermocellum ATCC 27405 during growth on dilute acid pretreated Populus and switchgrass, (2013) 1–18.

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

[40] I.S. Cunha, C.C. Barreto, O.Y. a Costa, M. a Bomfim, A.P. Castro, R.H. Kruger, B.F. Quirino, Bacteria

and Archaea community structure in the rumen microbiome of goats (Capra hircus) from the semiarid region of Brazil., Anaerobe. 17 (2011) 118–24. doi:10.1016/j.anaerobe.2011.04.018.

[41] E.D. Scully, S.M. Geib, K. Hoover, M. Tien, S.G. Tringe, K.W. Barry, T. Glavina del Rio, M. Chovatia, J.R. Herr, J.E. Carlson, Metagenomic Profiling Reveals Lignocellulose Degrading System in a Microbial Community Associated with a Wood-Feeding Beetle, PLoS One. 8 (2013) 1–22. doi:10.1371/journal.pone.0073827.

[42] K.O. Osiro, B.R. de Camargo, R. Satomi, P.R. V. Hamann, J.P. Silva, M.V. de Sousa, B.F. Quirino, E.N. Aquino, C.R. Felix, A.M. Murad, E.F. Noronha, Characterization of Clostridium thermocellum (B8) secretome and purified cellulosomes for lignocellulosic biomass degradation, Enzyme Microb. Technol. 97 (2017) 43–54. doi:10.1016/j.enzmictec.2016.11.002.

[43] P.R. V. Hamann, D.L. Serpa, A.S. Barreto da Cunha, B.R. de Camargo, K.O. Osiro, M. Valle de Sousa, C.R. Felix, R.N.G. Miller, E.F. Noronha, Evaluation of plant cell wall degrading enzyme production by Clostridium thermocellum B8 in the presence of raw agricultural wastes, Int. Biodeterior. Biodegradation. 105 (2015) 97–105. doi:10.1016/j.ibiod.2015.08.013.

[44] V.O. Leitão, E.F. Noronha, B.R. Camargo, P.R.V. Hamann, A.S. Steindorff, B.F. Quirino, M.V. de Sousa, C.J. Ulhoa, C.R. Felix, Growth and expression of relevant metabolic genes of Clostridium thermocellum cultured on lignocellulosic residues, J. Ind. Microbiol. Biotechnol. 44 (2017) 825–834. doi:10.1007/s10295-017-1915-2.

[45] A.L. Demain, M. Newcomb, J.H.D. Wu, Cellulase, Clostridia, and Ethanol, Microbiol. Mol. Biol. Rev. 69 (2005) 124–154. doi:10.1128/MMBR.69.1.124–154.2005.

[46] L.R. Blume, E.F. Noronha, J. Leite, M.V. De Sousa, C.R. Felix, Characterization of Clostridium thermocellum Isolates Grown on Cellulose and Sugarcane Bagasse, (2013) 763–775. doi:10.1007/s12155-013-9295-6.

[47] P. Prawitwong, R. Waeonukul, C. Tachaapaikoon, P. Pason, K. Ratanakhanokchai, L. Deng, J. Sermsathanaswadi, K. Septiningrum, Y. Mori, A. Kosugi, Direct glucose production from lignocellulose using Clostridium thermocellum cultures supplemented with a thermostable β -glucosidase, (2013) 1–11.

[48] H. Wei, Y. Fu, L. Magnusson, J.O. Baker, P.-C. Maness, Q. Xu, S. Yang, A. Bowersox, I. Bogorad, W. Wang, M.P. Tucker, M.E. Himmel, S.-Y. Ding, Comparison of transcriptional profiles of Clostridium thermocellum grown on cellobiose and pretreated yellow poplar using RNA-Seq., Front. Microbiol. 5 (2014) 142. doi:10.3389/fmicb.2014.00142.

[49] D. Freier, C.P. Mothershed, J.W. Wiegel, Characterization of Clostridium thermocellum JW20, Appl. Environ. Microbiol. 54 (1988) 204–211.

[50] C.R. Felix, L.G. Ljungdahl, The cellulosome: the exocellular organelle of Clostridium., Annu. Rev. Microbiol. 47 (1993) 791–819. doi:10.1146/annurev.mi.47.100193.004043.

[51] E.A. Bayer, E. Setter, R. Lamed, Organization and distribution of the cellulosome in Clostridium

thermocellum, J. Bacteriol. 163 (1985) 552–559.

[52] V. V Zverlov, J. Kellermann, W.H. Schwarz, Functional subgenomics of Clostridium thermocellum cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes., Proteomics. 5 (2005) 3646–53. doi:10.1002/pmic.200401199.

[53] K. Tokatlidis, S. Salamitou, P. Béguin, P. Dhurjati, J.P. Aubert, Interaction of the duplicated segment carried by Clostridium thermocellum cellulases with cellulosome components, FEBS Lett. 291 (1991) 185–188. doi:10.1016/0014-5793(91)81279-H.

[54] U.T. Gerngross, M.P.M. Romaniec, T. Kobayashi, N.S. Huskisson, A.L. Demain, Sequencing of a Clostridium thermocellum gene cipA encoding the cellulosomal SL-protein reveals an unusual degree of internal homology, Mol. Microbiol. 8 (1993) 325–334. doi:10.1111/j.1365-2958.1993.tb01576.x.

[55] S. Jindou, A. Soda, S. Karita, T. Kajino, P. Béguin, J.H.D. Wu, M. Inagaki, T. Kimura, K. Sakka, K. Ohmiya, Cohesin-dockerin interactions within and between clostridium josui and clostridium thermocellum: Binding selectivity between cognate dockerin and cohesin domains and species specificity, J. Biol. Chem. 279 (2004) 9867–9874. doi:10.1074/jbc.M308673200.

[56] D.G. Olson, R.J. Giannone, R.L. Hettich, L.R. Lynd, Role of the CipA scaffoldin protein in cellulose solubilization, as determined by targeted gene deletion and complementation in Clostridium thermocellum., J. Bacteriol. 195 (2013) 733–9. doi:10.1128/JB.02014-12.

[57] Q. Xu, M.G. Resch, K. Podkaminer, S. Yang, J.O. Baker, B.S. Donohoe, C. Wilson, D.M. Klingeman, D.G. Olson, S.R. Decker, R.J. Giannone, R.L. Hettich, S.D. Brown, L.R. Lynd, E.A. Bayer, M.E. Himmel, Y.J. Bomble, Dramatic performance of Clostridium thermocellum explained by its wide range of cellulase modalities., Sci. Adv. 2 (2016) e1501254. doi:10.1126/sciadv.1501254.

[58] S. Poudel, R.J. Giannone, M. Rodriguez, B. Raman, M.Z. Martin, N.L. Engle, J.R. Mielenz, I. Nookaew, S.D. Brown, T.J. Tschaplinski, D. Ussery, R.L. Hettich, Integrated omics analyses reveal the details of metabolic adaptation of Clostridium thermocellum to lignocellulose-derived growth inhibitors released during the deconstruction of switchgrass, Biotechnol. Biofuels. 10 (2017) 14. doi:10.1186/s13068-016-0697-5.

[59] B. Henrissat, M. Vegetales, F. Grenoble, A classification of glycosyl hydrolases based sequence similarities amino acid, Biochem. J. 280 (1991) 309–316. doi:10.1007/s007920050009.

[60] D.G. Naumoff, Hierarchical classification of glycoside hydrolases., Biochemistry. (Mosc). 76 (2011) 622–635. doi:10.1134/S0006297911060022.

[61] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res. 42 (2014) 490–495. doi:10.1093/nar/gkt1178.

[62] P. Tomme, A. Boraston, J.M. Kormos, R.A.J. Warren, D.G. Kilburn, Affinity electrophoresis for the identification and characterization of soluble sugar binding by carbohydrate-binding modules, Enzyme Microb. Technol. 27 (2000) 453–458. doi:10.1016/S0141-0229(00)00246-5.

[63] H. Jørgensen, J.B. Kristensen, C. Felby, Enzymatic conversion of lignocellulose into fermentable sugars : challenges and opportunities, (2007) 119–134. doi:10.1002/bbb.

[64] P. Biely, Microbial carbohydrate esterases deacetylating plant polysaccharides, Biotechnol. Adv. 30 (2012) 1575–1588. doi:10.1016/j.biotechadv.2012.04.010.

[65] E.J. Taylor, N.L. Smith, J.P. Turkenburg, S. D'Souza, H.J. Gilbert, G.J. Davies, Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Araf51, from Clostridium thermocellum., Biochem. J. 395 (2006) 31–7. doi:10.1042/BJ20051780.

[66] D. Petre, J. Millet, R. Longin, P. Beguin, H. Girard, J.P. Aubert, Purification and properties of the endoglucanase C of Clostridium thermocellum produced in Escherichia coli., Biochimie. 68 (1986) 687–695.

[67] S.E. Blumer-Schuette, S.D. Brown, K.B. Sander, E.A. Bayer, I. Kataeva, J. V. Zurawski, J.M. Conway, M.W.W. Adams, R.M. Kelly, Thermophilic lignocellulose deconstruction, FEMS Microbiol. Rev. 38 (2014) 393–448. doi:10.1111/1574-6976.12044.

[68] D.L. Blum, I.A. Kataeva, X.L. Li, L.G. Ljungdahl, Feruloyl esterase activity of the Clostridium thermocellum cellulosome can be attributed to previously unknown domains of XynY and XynZ, J. Bacteriol. 182 (2000) 1346–1351. doi:10.1128/JB.182.5.1346-1351.2000.

[69] A. Levasseur, E. Drula, V. Lombard, P.M. Coutinho, B. Henrissat, Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes., Biotechnol. Biofuels. 6 (2013) 41. doi:10.1186/1754-6834-6-41.

[70] K.S. Johansen, Lytic Polysaccharide Monooxygenases: The Microbial Power Tool for Lignocellulose Degradation, Trends Plant Sci. 21 (2016) 926–936. doi:10.1016/j.tplants.2016.07.012.

[71] L.D. Ellis, E.K. Holwerda, D. Hogsett, S. Rogers, X. Shao, T. Tschaplinski, P. Thorne, L.R. Lynd, Closing the carbon balance for fermentation by Clostridium thermocellum (ATCC 27405), Bioresour. Technol. 103 (2012) 293–299. doi:10.1016/j.biortech.2011.09.128.

[72] N.J. Patni, J.K. Alexander, Utilization of glucose by *Clostridium thermocellum*: presence of glucokinase and other glycolytic enzymes in cell extracts., J. Bacteriol. 105 (1971) 220–225.

[73] Y.-H.P. Zhang, L.R. Lynd, Cellulose utilization by Clostridium thermocellum: bioenergetics and hydrolysis product assimilation., Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 7321–5. doi:10.1073/pnas.0408734102.

[74] M. Desvaux, Unravelling carbon metabolism in anaerobic cellulolytic bacteria, Biotechnol. Prog. 22 (2006) 1229–1238. doi:10.1021/bp060016e.

[75] E.K. Holwerda, P.G. Thorne, D.G. Olson, D. Amador-Noguez, N.L. Engle, T.J. Tschaplinski, J.P. van Dijken, L.R. Lynd, The exometabolome of Clostridium thermocellum reveals overflow metabolism at high cellulose loading, Biotechnol. Biofuels. 7 (2014) 155. doi:10.1186/s13068-014-0155-1.

[76] S.B. Roberts, C.M. Gowen, J.P. Brooks, S.S. Fong, Genome-scale metabolic analysis of Clostridium thermocellum for bioethanol production., BMC Syst. Biol. 4 (2010) 31. doi:10.1186/1752-0509-4-31.

[77] J.N. Saddler, M.K.-H. Chan, Conversion of pretreated lignocellulosic substrates to ethanol by Clostridium thermocellum in mono- and co-culture with Clostridium thermosaccharolyticum and Clostridium thermohydrosulphuricurn, Can. J. Microbiol. 30 (1984) 212–220. doi:10.1139/m84-032.

[78] W. Liebermeister, E. Klipp, S. Schuster, R. Heinrich, A theory of optimal differential gene expression, BioSystems. 76 (2004) 261–278. doi:10.1016/j.biosystems.2004.05.022.

[79] J.L. Linville, M. Rodriguez, S.D. Brown, J.R. Mielenz, C.D. Cox, Transcriptomic analysis of Clostridium thermocellum Populus hydrolysate-tolerant mutant strain shows increased cellular efficiency in response to Populus hydrolysate compared to the wild type strain., BMC Microbiol. 14 (2014) 215. doi:10.1186/s12866-014-0215-5.

[80] Y. Nataf, L. Bahari, H. Kahel-Raifer, I. Borovok, R. Lamed, E. a Bayer, A.L. Sonenshein, Y. Shoham, Clostridium thermocellum cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors., Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 18646–51. doi:10.1073/pnas.1012175107.

[81] H. Kahel-Raifer, S. Jindou, L. Bahari, Y. Nataf, Y. Shoham, E. a Bayer, I. Borovok, R. Lamed, The unique set of putative membrane-associated anti-sigma factors in Clostridium thermocellum suggests a novel extracellular carbohydrate-sensing mechanism involved in gene regulation., FEMS Microbiol. Lett. 308 (2010) 84–93. doi:10.1111/j.1574-6968.2010.01997.x.

[82] L. Bahari, Y. Gilad, I. Borovok, H. Kahel-Raifer, B. Dassa, Y. Nataf, Y. Shoham, R. Lamed, E. Bayer, Glycoside hydrolases as components of putative carbohydrate biosensor proteins in Clostridium thermocellum, J. Ind. Microbiol. Biotechnol. 38 (2011) 825–832. doi:10.1007/s10295-010-0848-9.

[83] Y. Nataf, S. Yaron, F. Stahl, R. Lamed, E. a. Bayer, T.H. Scheper, A.L. Sonenshein, Y. Shoham, Cellodextrin and laminaribiose ABC transporters in Clostridium thermocellum, J. Bacteriol. 91 (2009) 203–209. doi:10.1128/JB.01190-08.

[84] I. Muñoz-Gutiérrez, L.O. De Ora, I.R. Grinberg, Y. Garty, E.A. Bayer, Y. Shoham, R. Lamed, I. Borovok, Decoding biomass-sensing regulons of clostridium thermocellum alternative sigma-I factors in a heterologous bacillus subtilis host system, PLoS One. 11 (2016). doi:10.1371/journal.pone.0146316.

[85] O. Yaniv, G. Fichman, I. Borovok, Y. Shoham, E.A. Bayer, R. Lamed, L.J.W. Shimon, F. Frolow, Fine-structural variance of family 3 carbohydrate-binding modules as extracellular biomass-sensing components of Clostridium thermocellum anti-σI factors, Acta Crystallogr. Sect. D Biol. Crystallogr. 70 (2014) 522–534. doi:10.1107/S139900471302926X.

[86] M. Newcomb, J. Millen, C.Y. Chen, J.H.D. Wu, Co-transcription of the celC gene cluster in Clostridium thermocellum, Appl. Microbiol. Biotechnol. 90 (2011) 625–634. doi:10.1007/s00253-011-3121-x.

[87] M. Newcomb, C.-Y. Chen, J.H.D. Wu, Induction of the celC operon of Clostridium thermocellum by

laminaribiose., Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 3747–3752. doi:10.1073/pnas.0700087104.

[88] J. Choi, D.M. Klingeman, S.D. Brown, C.D. Cox, The LacI family protein GlyR3 co-regulates the celC operon and manB in Clostridium thermocellum, Biotechnol. Biofuels. 10 (2017) 163. doi:10.1186/s13068-017-0849-2.

[89] N.D. Gold, V.J.J. Martin, Global view of the Clostridium thermocellum cellulosome revealed by quantitative proteomic analysis., J. Bacteriol. 189 (2007) 6787–95. doi:10.1128/JB.00882-07.

[90] B. Raman, C. Pan, G.B. Hurst, M. Rodriguez, C.K. McKeown, P.K. Lankford, N.F. Samatova, J.R. Mielenz, Impact of pretreated Switchgrass and biomass carbohydrates on Clostridium thermocellum ATCC 27405 cellulosome composition: a quantitative proteomic analysis., PLoS One. 4 (2009) e5271. doi:10.1371/journal.pone.0005271.

[91] A. Riederer, T.E. Takasuka, S.I. Makino, D.M. Stevenson, Y. V. Bukhman, N.L. Elsen, B.G. Fox, Global gene expression patterns in *Clostridium thermocellum* as determined by microarray analysis of chemostat cultures on cellulose or cellobiose, Appl. Environ. Microbiol. 77 (2011) 1243–1253. doi:10.1128/AEM.02008-10.

[92] A. Dumitrache, D.M. Klingeman, J. Natzke, M. Rodriguez Jr, R.J. Giannone, R.L. Hettich, B.H. Davison, S.D. Brown, Specialized activities and expression differences for Clostridium thermocellum biofilm and planktonic cells, Sci. Rep. 7 (2017) 43583. doi:10.1038/srep43583.

[93] B.R. CAMARGO, Produção , caracterização e purificação de complexos enzimáticos holocelulolíticos produzidos por Clostridium thermocellum (Isolado B8) na presença de substratos lignocelulósicos, 2013.

[94] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254. doi:http://dx.doi.org/10.1016/0003-2697(76)90527-3.

[95] G.L. Miller, Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar, Anal. Chem. 31 (1959) 426–428. doi:10.1021/ac60147a030.

[96] R.D. Finn, P. Coggill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G.A. Salazar, J. Tate, A. Bateman, The Pfam protein families database: Towards a more sustainable future, Nucleic Acids Res. 44 (2016) D279–D285. doi:10.1093/nar/gkv1344.

[97] R.L. Tatusov, D.A. Natale, I. V Garkavtsev, T.A. Tatusova, U.T. Shankavaram, B.S. Rao, B. Kiryutin, M.Y. Galperin, N.D. Fedorova, E. V Koonin, The COG database: new developments in phylogenetic classification of proteins from complete genomes., Nucleic Acids Res. 29 (2001) 22–8. doi:10.1093/nar/29.1.22.

[98] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nat. Methods. 8 (2011) 785–786. doi:10.1038/nmeth.1701.

[99] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering, L.J. Jensen, STRING v9.1: protein-protein interaction networks, with increased

coverage and integration., Nucleic Acids Res. 41 (2013) D808–15. doi:10.1093/nar/gks1094.

[100] Y. Peng, Y. Luo, T. Yu, X. Xu, K. Fan, Y. Zhao, K. Yang, A blue native-PAGE analysis of membrane protein complexes in Clostridium thermocellum., BMC Microbiol. 11 (2011) 22. doi:10.1186/1471-2180-11-22.

[101] E. Burton, V.J.J. Martin, Proteomic analysis of Clostridium thermocellum ATCC 27405 reveals the upregulation of an alternative transhydrogenase – malate pathway and nitrogen assimilation in cells grown on cellulose, 1388 (2012) 1378–1388. doi:10.1139/cjm-2012-0412.

[102] I. Kataeva, X. Li, H. Chen, S. Choi, L.G. Ljungdahl, Cloning and Sequence Analysis of a New Cellulase Gene Encoding CelK , a Major Cellulosome Component of Clostridium thermocellum : Evidence for Gene Duplication and Recombination Cloning and Sequence Analysis of a New Cellulase Gene Encoding CelK , a Major, (1999).

[103] T.W. Dror, E. Morag, A. Rolider, E.A. Bayer, R. Lamed, Y. Shoham, Regulation of the Cellulosomal celS (cel48A) Gene of Clostridium thermocellum Is Growth Rate Dependent Regulation of the Cellulosomal celS (cel48A) Gene of Clostridium thermocellum Is Growth Rate Dependent, J. Bacteriol. 185 (2003) 3042–3048. doi:10.1128/JB.185.10.3042.

[104] W.H. Schwarz, F. Grabnitz, W.L. Staudenbauer, Properties of a Clostridium thermocellum endoglucanase produced in Escherichia coli, Appl. Environ. Microbiol. 51 (1986) 1293–1299.

[105] M. Lemaire, P. Beguin, Nucleotide sequence of the celG gene of Clostridium thermocellum and characterization of its product, endoglucanase CelG, J. Bacteriol. 175 (1993) 3353–3360.

[106] T. Kametani, O. Umezawa, NII-Electronic Library Service, PhD Thesis. 14 (1966) 369–375. doi:10.1248/cpb.37.3229.

[107] V. V. Zverlov, G.A. Velikodvorskaya, W.H. Schwarz, Two new cellulosome components encoded downstream of cell in the genome of Clostridium thermocellum: The non-processive endoglucanase CelN and the possibly structural protein CseP, Microbiology. 149 (2003) 515–524. doi:10.1099/mic.0.25959-0.

[108] P. Beguin, P. Cornet, J. Millet, Identification of the endoglucanase encoded by the celB gene of Clostridium thermocellum., Biochimie. 65 (1983) 495–500.

[109] V. V. Zverlov, N. Schantz, W.H. Schwarz, A major new component in the cellulosome of Clostridium thermocellum is a processive endo-β-1,4-glucanase producing cellotetraose, FEMS Microbiol. Lett. 249 (2005) 353–358. doi:10.1016/j.femsle.2005.06.037.

[110] H. Hayashi, K.I. Takagi, M. Fukumura, T. Kimura, S. Karita, K. Sakka, K. Ohmiya, Sequence of xynC and properties of XynC, a major component of the Clostridium thermocellum cellulosome, J. Bacteriol. 179 (1997) 4246–4253. doi:10.1007/s00253-001-0832-4.

[111] V. V. Zverlov, N. Schantz, P. Schmitt-Kopplin, W.H. Schwarz, Two new major subunits in the cellusome of Clostridium thermocellum: Xyloglucanase Xgh74A and endoxylanase Xyn10D, Microbiology. 151 (2005) 3395–3401. doi:10.1099/mic.0.28206-0.

[112] S. Yang, R.J. Giannone, L. Dice, Z.K. Yang, N.L. Engle, T.J. Tschaplinski, R.L. Hettich, S.D. Brown, Clostridium thermocellum ATCC27405 transcriptomic, metabolomic and proteomic profiles after ethanol stress., BMC Genomics. 13 (2012) 336. doi:10.1186/1471-2164-13-336.

[113] T. Rydzak, P.D. McQueen, O. V Krokhin, V. Spicer, P. Ezzati, R.C. Dwivedi, D. Shamshurin, D.B. Levin, J. a Wilkins, R. Sparling, Proteomic analysis of Clostridium thermocellum core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression., BMC Microbiol. 12 (2012) 214. doi:10.1186/1471-2180-12-214.

[114] N.D. Gold, V.J.J. Martin, Global view of the Clostridium thermocellum cellulosome revealed by quantitative proteomic analysis., J. Bacteriol. 189 (2007) 6787–95. doi:10.1128/JB.00882-07.

[115] K. Hirano, M. Kurosaki, S. Nihei, H. Hasegawa, S. Shinoda, M. Haruki, N. Hirano, Enzymatic diversity of the Clostridium thermocellum cellulosome is crucial for the degradation of crystalline cellulose and plant biomass, Sci. Rep. 6 (2016) 35709. doi:10.1038/srep35709.

[116] C.D. Herring, P.G. Thorne, L.R. Lynd, Clostridium thermocellum releases coumaric acid during degradation of untreated grasses by the action of an unknown enzyme., Appl. Microbiol. Biotechnol. 100 (2016) 2907–2915. doi:10.1007/s00253-016-7294-1.

[117] B.A. Pinheiro, H.J. Gilbert, K. Sakka, K. Sakka, V.O. Fernandes, J.A. Prates, V.D. Alves, D.N. Bolam, L.M. Ferreira, C.M. Fontes, Functional insights into the role of novel type I cohesin and dockerin domains from Clostridium thermocellum, Biochem. J. 424 (2009) 375–384. doi:10.1042/BJ20091152.

[118] J.L. Linville, M. Rodriguez, S.D. Brown, J.R. Mielenz, C.D. Cox, Transcriptomic analysis of Clostridium thermocellum Populus hydrolysate-tolerant mutant strain shows increased cellular efficiency in response to Populus hydrolysate compared to the wild type strain., BMC Microbiol. 14 (2014) 215. doi:10.1186/s12866-014-0215-5.

[119] T.I. Williams, J.C. Combs, B.C. Lynn, H.J. Strobel, Proteomic profile changes in membranes of ethanol-tolerant Clostridium thermocellum., Appl. Microbiol. Biotechnol. 74 (2007) 422–32. doi:10.1007/s00253-006-0689-7.

[120] B. Raman, C.K. McKeown, M. Rodriguez, S.D. Brown, J.R. Mielenz, Transcriptomic analysis of Clostridium thermocellum ATCC 27405 cellulose fermentation., BMC Microbiol. 11 (2011) 134. doi:10.1186/1471-2180-11-134.

[121] S. Pitman, K.H. Cho, The mechanisms of virulence regulation by small noncoding RNAs in low GC gram-positive pathogens, Int. J. Mol. Sci. 16 (2015) 29797–29814. doi:10.3390/ijms161226194.

[122] T.J. Verbeke, R.J. Giannone, D.M. Klingeman, N.L. Engle, T. Rydzak, A.M. Guss, T.J. Tschaplinski, S.D. Brown, R.L. Hettich, J.G. Elkins, Pentose sugars inhibit metabolism and increase expression of an AgrD-type cyclic pentapeptide in Clostridium thermocellum, Sci. Rep. 7 (2017) 43355. doi:10.1038/srep43355.

[123] T.J. Verbeke, X. Zhang, B. Henrissat, V. Spicer, T. Rydzak, O. V. Krokhin, B. Fristensky, D.B. Levin, R. Sparling, Genomic Evaluation of Thermoanaerobacter spp. for the Construction of Designer Co-Cultures

to Improve Lignocellulosic Biofuel Production, PLoS One. 8 (2013). doi:10.1371/journal.pone.0059362.

[124] M. Ehsanipour, A.V. Suko, R. Bura, Fermentation of lignocellulosic sugars to acetic acid by Moorella thermoacetica, J. Ind. Microbiol. Biotechnol. 43 (2016) 807–816. doi:10.1007/s10295-016-1756-4.

[125] Y. Nakamura, H. Miyafuji, H. Kawamoto, S. Saka, Acetic acid fermentability with Clostridium thermoaceticum and Clostridium thermocellum of standard compounds found in beech wood as produced in hot-compressed water, J. Wood Sci. 57 (2011) 331–337. doi:10.1007/s10086-010-1169-3.

[126] H. Rabemanolontsoa, G. Kawasaki, S. Saka, Effects of decomposed products from Japanese cedar hydrolyzates on acetic acid fermentation by Clostridium thermocellum and Moorella thermoacetica (C. thermoaceticum), Process Biochem. (n.d.). doi:http://doi.org/10.1016/j.procbio.2017.03.020.

[127] H. Rabemanolontsoa, Y. Kuninori, S. Saka, High conversion efficiency of Japanese cedar hydrolyzates into acetic acid by co-culture of Clostridium thermoaceticum and Clostridium thermocellum, J. Chem. Technol. Biotechnol. 91 (2016) 1040–1047. doi:10.1002/jctb.4679.

[128] J.H. Bullard, E. Purdom, K.D. Hansen, S. Dudoit, Evaluation of Statistical Methods for Normalization and Differential Expression in mRNA-Seq Experiments Evaluation of Statistical Methods for Normalization and Differential Expression in mRNA-Seq Experiments, U.C. Berkeley Div. Biostat. Pap. Ser. 11 (2009) 94. doi:10.1186/1471-2105-11-94.

[129] B.P. Westover, J.D. Buhler, J.L. Sonnenburg, J.I. Gordon, Operon prediction without a training set, Bioinformatics. 21 (2005) 880–888. doi:10.1093/bioinformatics/bti123.

[130] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M. Kanehisa, KEGG: Kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 27 (1999) 29–34. doi:10.1093/nar/27.1.29.

[131] Y. Lian, Y. Yang, J. Guo, Y. Wang, X. Li, Y. Fang, L. Gan, M. Xu, Electron acceptor redox potential globally regulates transcriptomic profiling in Shewanella decolorationis S12, Sci. Rep. 6 (2016) 31143. doi:10.1038/srep31143.

[132] M. Alahuhta, Q. Xu, Y.J. Bomble, R. Brunecky, W.S. Adney, S.Y. Ding, M.E. Himmel, V. V. Lunin, The unique binding mode of cellulosomal CBM4 from clostridium thermocellum cellobiohydrolase A, J. Mol. Biol. 402 (2010) 374–387. doi:10.1016/j.jmb.2010.07.028.

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

[134] H. Hayashi, M. Takehara, T. Hattori, T. Kimura, S. Karita, K. Sakka, K. Ohmiya, Nucleotide sequences of two contiguous and highly homologous xylanase genes xynA and xynB and characterization of XynA from Clostridium thermocellum., Appl. Microbiol. Biotechnol. 51 (1999) 348–357.

[135] P.S. Chauhan, N. Puri, P. Sharma, N. Gupta, Mannanases: Microbial sources, production, properties and potential biotechnological applications, Appl. Microbiol. Biotechnol. 93 (2012) 1817–1830.

doi:10.1007/s00253-012-3887-5.

[136] M.D. Dong, J.E. Urch, J.M. Ten Cate, V.A. Rao, D.M.F. Van Aalten, W. Crielaard, Streptococcus mutans SMU.623c codes for a functional, metal-dependent polysaccharide deacetylase that modulates interactions with salivary agglutinin, J. Bacteriol. 91 (2009) 394–402. doi:10.1128/JB.00838-08.

[137] M. Tangney, W.J. Mitchell, Characterisation of a glucose phosphotransferase system in Clostridium acetobutylicum ATCC 824, Appl. Microbiol. Biotechnol. 74 (2007) 398–405. doi:10.1007/s00253-006-0679-9.

[138] T.A. Gulbrandsen, I.A. Johnsen, M.T. Opedal, K. Toven, K. Oyaas, A. Pranovich, J.P. Mikkola, B.H. Hoff, Extracting Hemicelluloses from Softwood and Bagasse as Oligosaccharides Using Pure Water and Microwave Heating, Cellul. Chem. Technol. 49 (2015) 117–126. <Go to ISI>://000351490400001.

[139] R. Dredge, S.E. Radloff, J.S. Dyk, B.I. Pletschke, Lime pretreatment of sugar beet pulp and evaluation of synergy between ArfA, ManA and XynA from Clostridium cellulovorans on the pretreated substrate, 3 Biotech. 1 (2011) 151–159. doi:10.1007/s13205-011-0019-3.

[140] N. Beukes, B.I. Pletschke, Effect of alkaline pre-treatment on enzyme synergy for efficient hemicellulose hydrolysis in sugarcane bagasse, Bioresour. Technol. 102 (2011) 5207–5213. doi:10.1016/j.biortech.2011.01.090.

[141] N. Sarkar, S.K. Ghosh, S. Bannerjee, K. Aikat, Bioethanol production from agricultural wastes: An overview, Renew. Energy. 37 (2012) 19–27. doi:http://dx.doi.org/10.1016/j.renene.2011.06.045.

[142] L.F. Silva, M.K. Taciro, G. Raicher, R.A.M. Piccoli, T.T. Mendonça, M.S.G. Lopes, J.G.C. Gomez, Perspectives on the production of polyhydroxyalkanoates in biorefineries associated with the production of sugar and ethanol, Int. J. Biol. Macromol. 71 (2014) 2–7. doi:http://dx.doi.org/10.1016/j.ijbiomac.2014.06.065.

[143] J.M.D. Paye, A. Guseva, S.K. Hammer, E. Gjersing, M.F. Davis, B.H. Davison, J. Olstad, B.S. Donohoe, T.Y. Nguyen, C.E. Wyman, S. Pattathil, M.G. Hahn, L.R. Lynd, Biological lignocellulose solubilization: comparative evaluation of biocatalysts and enhancement via cotreatment., Biotechnol. Biofuels. 9 (2016) 8. doi:10.1186/s13068-015-0412-y.

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

[145] A. Goyal, S. Ahmed, K. Sharma, V. Gupta, P. Bule, V.D. Alves, C.M.G.A. Fontes, S. Najmudin, Molecular determinants of substrate specificity revealed by the structure of Clostridium thermocellum arabinofuranosidase 43A from glycosyl hydrolase family 43 subfamily 16, Acta Crystallogr. Sect. D Struct. Biol. 72 (2016) 1281–1289. doi:10.1107/S205979831601737X.

[146] E. Vandermarliere, T.M. Bourgois, M.D. Winn, S.V.A.N. Campenhout, G. Volckaert, J.A. Delcour, S. V Strelkov, A. Rabijns, C.M. Courtin, Structural analysis of a glycoside hydrolase family 43 arabinoxylan

arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family, 47 (2009) 39–47. doi:10.1042/BJ20081256.

[147] A. Cartmell, L.S. Mckee, M.J. Pen, J. Larsbrink, H. Brumer, S. Kaneko, H. Ichinose, R.J. Lewis, A. Viksø-nielsen, H.J. Gilbert, J. Marles-wright, The Structure and Function of an Arabinan-specific α-1 , 2-Arabinofuranosidase Identified from Screening the Activities of Bacterial GH43 Glycoside Hydrolases * □, 286 (2011) 15483–15495. doi:10.1074/jbc.M110.215962.

[148] M. Till, D. Goldstone, Structural analysis of the GH43 enzyme Xsa43E from Butyrivibrio proteoclasticus, (2014) 1193–1198. doi:10.1107/S2053230X14014745.

[149] J. Krauss, V. V Zverlov, W.H. Schwarz, In vitro reconstitution of the complete Clostridium thermocellum cellulosome and synergistic activity on crystalline cellulose., Appl. Environ. Microbiol. 78 (2012) 4301–7. doi:10.1128/AEM.07959-11.

[150] S.P. Das, R. Ravindran, S. Ahmed, D. Das, D. Goyal, C.M.G.A. Fontes, A. Goyal, Bioethanol production involving recombinant C. thermocellum hydrolytic hemicellulase and fermentative microbes, Appl. Biochem. Biotechnol. 167 (2012) 1475–1488. doi:10.1007/s12010-012-9618-7.

[151] S. Ahmed, A.S. Luis, J.L.A. Bras, A. Ghosh, S. Gautam, M.N. Gupta, C.M.G.A. Fontes, A. Goyal, A Novel a -L-Arabinofuranosidase of Family 43 Glycoside Hydrolase (Ct43Araf) from Clostridium thermocellum, 8 (2013) 1–10. doi:10.1371/journal.pone.0073575.

[152] D. Jiang, J. Fan, X. Wang, Y. Zhao, B. Huang, J. Liu, X.C. Zhang, Crystal structure of 1,3Gal43A, an exo-β-1,3-galactanase from Clostridium thermocellum, J. Struct. Biol. 180 (2012) 447–457. doi:10.1016/j.jsb.2012.08.005.

[153] P.R. V. Hamann, D.L. Serpa, A.S. Barreto da Cunha, B.R. de Camargo, K.O. Osiro, M. Valle de Sousa, C.R. Felix, R.N.G. Miller, E.F. Noronha, Evaluation of plant cell wall degrading enzyme production by Clostridium thermocellum B8 in the presence of raw agricultural wastes, Int. Biodeterior. Biodegradation. 105 (2015) 97–105. doi:10.1016/j.ibiod.2015.08.013.

[154] A. Bateman, M.J. Martin, C. O'Donovan, M. Magrane, R. Apweiler, E. Alpi, R. Antunes, J. Arganiska, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, G. Chavali, E. Cibrian-Uhalte, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, P. Gane, L.G. Castro, P. Garmiri, E. Hatton-Ellis, R. Hieta, R. Huntley, D. Legge, W. Liu, J. Luo, A. Macdougall, P. Mutowo, A. Nightingale, S. Orchard, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R. Saidi, T. Sawford, A. Shypitsyna, E. Turner, V. Volynkin, T. Wardell, X. Watkins, H. Zellner, A. Cowley, L. Figueira, W. Li, H. McWilliam, R. Lopez, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. Argoud-Puy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M.C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E. Coudert, B. Cuche, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, F. Jungo, G. Keller, V. Lara, P. Lemercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, N. Nouspikel, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, B. Roechert, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, A.L.

Veuthey, C.H. Wu, C.N. Arighi, L. Arminski, C. Chen, Y. Chen, J.S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D.A. Natale, B.E. Suzek, C.R. Vinayaka, Q. Wang, Y. Wang, L.S. Yeh, M.S. Yerramalla, J. Zhang, UniProt: A hub for protein information, Nucleic Acids Res. 43 (2015) D204–D212. doi:10.1093/nar/gku989.

[155] C. Notredame, D. Higgins, J. Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment., J. Mol. Biol. 302 (2000) 205–217. doi:10.1006/jmbi.2000.4042.

[156] Z. He, H. Zhang, S. Gao, M.J. Lercher, W.H. Chen, S. Hu, Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees, Nucleic Acids Res. 44 (2016) W236–W241. doi:10.1093/nar/gkw370.

[157] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4., Nature. 227 (1970) 680–685.

[158] M. Martínez, Patricia, R. Bakker, P. Harmsen, H. Gruppen, M. Kabel, Importance of acid or alkali concentration on the removal of xylan and lignin for enzymatic cellulose hydrolysis, Ind. Crops Prod. 64 (2015) 88–96. doi:10.1016/j.indcrop.2014.10.031.

[159] M.J.A. Daas, P.M. Martínez, A.H.P. van de Weijer, J. van der Oost, W.M. de Vos, M.A. Kabel, R. van Kranenburg, Biochemical characterization of the xylan hydrolysis profile of the extracellular endo-xylanase from Geobacillus thermodenitrificans T12, BMC Biotechnol. 17 (2017) 44. doi:10.1186/s12896-017-0357-2.

[160] S. Kato, S. Haruta, Z.J. Cui, M. Ishii, A. Yokota, Y. Igarashi, Clostridium straminisolvens sp. nov., a moderately thermophilic, aerotolerant and cellulolytic bacterium isolated from a cellulose-degrading bacterial community, Int. J. Syst. Evol. Microbiol. 54 (2004) 2043–2047. doi:10.1099/ijs.0.63148-0.

[161] J. Sukhumavas, K. Ohmiya, S. Shimizu, K. Uen, K. Ueno, Species from Thai Compost, Int. J. Syst. Bacteriol. 38 (1988) 179–182.

[162] S. Siozios, A. Cestaro, R. Kaur, I. Pertot, O. Rota-Stabelli, G. Anfora, Draft Genome Sequence of the, Genome Annunc. 1 (2013) e00032–13. doi:10.1128/genomeA.00032-13.Copyright.

[163] H. Celik, J.C. Blouzard, B. Voigt, D. Becher, V. Trotter, H.P. Fierobe, C. Tardif, S. Pagès, P. de Philip, A Two-Component System (XydS/R) Controls the Expression of Genes Encoding CBM6-Containing Proteins in Response to Straw in Clostridium cellulolyticum, PLoS One. 8 (2013). doi:10.1371/journal.pone.0056063.

[164] L.L. Li, S. Taghavi, J.A. Izquierdo, D. van der Lelie, Complete genome sequence of Clostridium sp. strain BNL1100, a cellulolytic mesophile isolated from corn stover, J. Bacteriol. 194 (2012) 6982–6983. doi:10.1128/JB.01908-12.

[165] M. Foumani, T. V Vuong, B. MacCormick, E.R. Master, Enhanced Polysaccharide Binding and Activity on Linear beta-Glucans through Addition of Carbohydrate-Binding Modules to Either Terminus of a Glucooligosaccharide Oxidase., PLoS One. 10 (2015) e0125398. doi:10.1371/journal.pone.0125398.

[166] M. Sajjad, M.I.M. Khan, N.S. Akbar, S. Ahmad, I. Ali, M.W. Akhtar, Enhanced expression and activity yields of Clostridium thermocellum xylanases without non-catalytic domains, J. Biotechnol. 145 (2010) 38–42. doi:10.1016/j.jbiotec.2009.10.013.

APÊNDICE

Apêndice A. 1. Proteínas identificadas na amostra FS_Celulose.

Locus gene	protein.Accession	protein.Description	protein.MW	ScoreAVG	ProductsAVG	PeptidesAVG		FmolAVG	FmolCV	Log10FmolAVG		ngramAVG	Repeate.rate	
TOP3AVG	CoverageAVG	TOP3CV	NgramVariance											
Cthe_2089C7HDZ8_CLOTM	Glycoside hydrolase family 48	83702.5726	623.10	72.33	16.00	224.86	0.03	2.35	18.82	3.00	117030.33	31.13	0.02	0.36
Cthe_0412C7HHU0_CLOTM	Glycoside hydrolase family 9	101249.3736	749.60	87.33	19.00	26.34	1.67	1.42	2.67	3.00	14388.67	30.02	1.67	19.84
Cthe_3052C7HJJ2_CLOTM	YD repeat protein (Fragment)	344451.004	21.54	43.50	31.00	10.84	1.41	1.03	3.73	2.00	5556.50	4.36	1.41	27.87
Cthe_3078C7HJX6_CLOTM	Cellulosome anchoring protein cohesin region (Fragment)		19238.6245			290.61	13.33	3.00	8.61	1.73	0.93	0.17	3.00	4329.33
24.34	1.73	0.08												
Cthe_3078D1NHZ1_CLOTM	Cellulosome anchoring protein cohesin region PE=4 SV=1	192072.7446	354.60	17.33	7.00	9.43	1.73	0.97	1.81	3.00	4833.00	4.71	1.73	9.83
Clo1313_1399	D1NIC8_CLOTM	Putative uncharacterized protein	14884.0635		2673.11	34.33	3.67	207.08	0.03	2.32	3.08	3.00	107831.33	35.86
0.04	0.01													
Cthe_1368D1NJG7_CLOTM	S-layer domain protein	77935.3585	163.84	45.50	13.00	53.46	0.17	1.73	4.17	2.00	27922.50	21.98	0.11	0.51
Cthe_1373D1NJH3_CLOTM	YD repeat protein	218400.6805	51.83	28.50	22.00	21.69	0.90	1.34	4.74	2.00	11254.50	4.39	0.87	18.08
Cthe_3226D1NKK6_CLOTM	Copper amine oxidase domain protein		29794.1092		429.28	22.67	4.00	19.70	1.73	1.29	0.59	3.00	10099.33	17.56
1.73	1.03													
Cthe_3227D1NKK7_CLOTM	Copper amine oxidase domain protein		30444.8175		465.44	25.67	5.67	17.59	1.73	1.25	0.54	3.00	8847.33	22.10
1.73	0.86													
Cthe_3227D1NKK8_CLOTM	Copper amine oxidase domain protein	30261.6806	459.23	24.67	5.67	23.21	1.27	1.37	0.70	3.00	12543.00	22.31	1.29	0.79
Cthe_3232D1NKL1_CLOTM	Rhs family protein-like protein	222046.854	46.25	24.00	22.00	14.76	0.11	1.17	3.28	2.00	7499.22	6.08	0.12	0.12
Cthe_0412D1NLS2_CLOTM	Glycoside hydrolase family 9	101339.3908	749.46	88.33	19.00	53.71	0.87	1.73	5.44	3.00	27277.67	26.70	0.87	22.22
Cthe_0423D1NLU9_CLOTM	Iron-containing alcohol dehydrogenase	96537.2986	239.26	52.67	15.00	50.54	0.07	1.70	4.88	3.00	26307.67	20.81	0.06	0.10

Cthe_1104D1NMR1_CLOTM 0.21	Putative uncharacterized protein	19049.5963		1980.46	52.00	8.00	290.37	0.08	2.46	5.53	3.00	151116.00	69.73	0.08	
Cthe_2423D1NNW4_CLOTM 0.23	Type 3a cellulose-binding domain protein	112443.7952		182.06	41.00	12.33	61.34	0.07	1.79	6.90	3.00	31895.33	12.75	0.04	
Cthe_2348D1NP50_CLOTM	Ig domain protein	113170.1435	5525.07	389.67	43.67	1208.25	0.05	3.08	136.74	3.00	629669.33	47.76	0.00	52.64	
Cthe_3077E6UU82_CLOTL 0.87	Cellulosome anchoring protein cohesin region	143945.1691		821.26	92.33	19.33	74.37	0.87	1.87	10.71	3.00	39080.67	20.24	86.06	
Cthe_2390H8EAA6_CLOTM 24291.00	Pyruvateketoisovalerate oxidoreductase gamma subunit	21260.5494		363.01	19.67	5.00	46.67	0.03	1.67	0.99	3.00	27.95	0.02	0.00	
Cthe_2383H8EAB3_CLOTM 0.58	Copper amine oxidase-like domain-containing protein	33203.5375		133.11	16.00	5.00	105.84	0.54	2.02	3.51	3.00	55967.00	18.31	3.60	
Cthe_2168H8EB12_CLOTM 0.22	Propeptide PepSY amd peptidase M4	83182.3578		695.44	117.67	21.00	129.05	0.04	2.11	10.73	3.00	67289.67	36.43	0.08	
Cthe_1192H8EC35_CLOTM	Uncharacterized protein	39085.352	429.64	27.00	8.67	58.68	0.06	1.77	2.29	3.00	30518.00	22.55	0.01	0.02	
Cthe_0137H8EC95_CLOTM 0.07	Glyceraldehyde-3-phosphate dehydrogenase type I	36367.4154		895.76	41.00	7.67	183.83	0.03	2.26	6.69	3.00	95822.33	29.86	0.04	
Cthe_0138H8EC96_CLOTM	Phosphoglycerate kinase	42876.5735		370.16	38.33	11.33	91.21	0.07	1.96	3.91	3.00	47529.00	27.46	0.09	0.08
Cthe_0139H8EC97_CLOTM	Triosephosphate isomerase	27362.2549		485.70	29.67	5.00	81.06	0.14	1.91	2.22	3.00	42246.33	28.29	0.15	0.10
Cthe_0143H8ECA0_CLOTM	Enolase	47159.6652	1067.89	65.67	11.33	155.89	0.03	2.19	7.35	3.00	81205.00	33.87	0.05	0.06	
Cthe_0347H8ECN6_CLOTM	Phosphofructokinase	46023.3183		346.98	37.00	7.67	77.26	0.02	1.89	3.56	3.00	40252.33	23.13	0.05	0.01
Cthe_0349H8ECN7_CLOTM 0.18	Fructose-1 6-bisphosphate aldolase class II	33917.7896		412.95	30.67	6.67	90.88	0.14	1.96	3.08	3.00	47233.00	29.66	0.12	
Cthe_1965H8ECT1_CLOTM	Peroxiredoxin	20918.5684	4079.41	77.33	9.67	223.36	0.05	2.35	4.67	3.00	116214.33	58.64	0.02	0.05	
Cthe_0838H8ECV2_CLOTM	Uncharacterized protein	24090.8118		153.52	18.00	3.50	34.53	0.23	1.54	0.83	2.00	18251.25	20.93	0.29	0.04
Cthe_1308H8ED52_CLOTM 2.09	Pyruvate phosphate dikinase	99363.6689		126.67	28.50	10.50	46.66	0.31	1.67	4.64	2.00	24272.00	9.91	0.25	
Cthe_0037H8EDF4_CLOTM	Heat shock protein Hsp20	16449.5714		442.80	17.00	4.00	26.65	0.26	1.43	0.44	2.00	14036.50	34.86	0.21	0.01
Cthe_1778H8EDM0_CLOTM 0.00	Copper amine oxidase-like domain-containing protein	35472.921	226.08	13.00	5.67	19.53	0.06	1.29	0.69	3.00	10166.33	15.88	0.05		

Cthe_0217	H8EED4_CLOTM	Glucose-6-phosphate isomerase	50401.4304		190.18	36.00	12.00	39.37	0.02	1.60	1.98	2.00	20655.00	26.46	0.04	0.00	
Cthe_0421	H8EEE9_CLOTM	Dipicolinate synthase subunit B	20868.4402		179.14	11.00	2.33	30.95	0.32	1.49	0.65	3.00	16256.17	14.44	0.36	0.04	
Cthe_2730	H8EEP5_CLOTM	Elongation factor Tu	44276.5466	1448.01	68.33	14.33	126.04	0.03	2.10	5.58	3.00	65663.33	45.67	0.05	0.02		
Cthe_2723	H8EEQ2_CLOTM	50S ribosomal protein L7L12	13295.4486	3039.62	34.00	4.00	95.80	0.16	1.98	1.27	3.00	50126.17	49.36	0.21	0.04		
Cthe_0637	H8EFI9_CLOTM	Uncharacterized protein	11114.6504		258.44	5.50	1.00	103.97	0.00	2.02	1.16	2.00	52803.00	19.00	0.02	0.00	
Cthe_2518	H8EGB1_CLOTM	Ketol-acid reductoisomerase	36393.6598		154.99	17.00	5.00	21.39	0.23	1.33	0.78	2.00	11276.50	20.09	0.19	0.03	
Clo1313_0638	H8EGI8_CLOTM	Histone family protein DNA-binding protein	9728.3308		183.54	6.50	2.00	26.96	0.29	1.43	0.26	2.00	13663.50	15.39	0.28	0.01	
Cthe_0394	H8EGU6_CLOTM	Iron-containing alcohol dehydrogenase	42472.4604		239.56	30.67	7.67	56.75	0.07	1.75	2.41	3.00	29563.33	23.57	0.09	0.03	
Cthe_0393	H8EGU7_CLOTM	Sugar ABC transporter (Sugar-binding protein)	34182.3493		5941.96	88.67	7.33	438.69	0.07	2.64	15.00	3.00	228427.67	35.83	0.08	1.16	
Cthe_0360	H8EGY0_CLOTM	Thioredoxin	12295.184	252.59	7.50	2.00	33.46	0.44	1.52	0.41	2.00	17337.00	20.19	0.39	0.03		
Cthe_1020	H8EH05_CLOTM	Extracellular solute-binding protein family 1	50085.6863		6847.15	230.67	25.00	1741.71	0.12	3.24	87.23	3.00	907022.67	58.75	0.12	106.70	
AD2-0592	H8EH70_CLOTM	Uncharacterized protein	9725.8088	973.39	22.00	4.50	45.10	0.36	1.65	0.44	2.00	23705.00	29.41	0.31	0.02		
Cthe_3194	H8EHQ1_CLOTM	Sporulation lipoprotein YhcNYlaJ-like protein	22954.703	182.03	14.50	4.00	33.98	0.09	1.53	0.78	2.00	17248.00	20.60	0.08	0.01		
Cthe_2892	H8EI52_CLOTM	60 kDa chaperonin	57531.2224		199.30	51.00	14.50	51.93	0.13	1.72	2.99	2.00	27371.00	32.53	0.19	0.15	
Cthe_2891	H8EI53_CLOTM	10 kDa chaperonin	10155.8185		2035.24	31.33	4.00	82.87	0.08	1.92	0.84	3.00	43124.67	45.74	0.08	0.01	
Cthe_2877	H8EI67_CLOTM	S-layer domain-containing protein	67228.5538		142.51	38.50	14.50	37.38	0.04	1.57	2.51	2.00	19603.50	22.23	0.02	0.01	
Cthe_2874	H8EI70_CLOTM	Phosphoenolpyruvate carboxykinase [GTP]	68140.7596		394.33	56.33	15.00	80.76	0.05	1.91	5.50	3.00	42041.67	30.69	0.05	0.09	
Cthe_1368	H8EI82_CLOTM	S-layer domain-containing protein	78164.551	128.29	37.50	11.50	0.52	1.41	-0.28	0.04	2.00	264.00	19.86	1.41	0.00		
Cthe_3078	H8ERY1_CLOTM	Cellulosome anchoring protein cohesin region (Fragment)	91214.438	350.36	16.33	5.67	9.53	1.73	0.98	0.87	3.00	5215.00	8.76	1.73	2.27		

Cthe_3077N1JV01_CLOTM	Cellulosome anchoring protein CipA (Fragment)					178671.2766		719.77	82.33	18.00	30.78	1.73	1.49	5.50	3.00			
16835.67	14.90	1.73	90.71															
Cthe_3052N1K0C2_CLOTM	Cellulosome anchoring protein CipA			195913.6538		826.51	93.00	20.00	36.42	1.73	1.56	7.14	3.00	18675.33	16.56	1.73		
152.76																		
Cthe_2348O86999_CLOTM	S-layer protein		113387.3653	5395.52	344.00	35.33	0.39	1.73	-0.41	0.04	3.00	198.33	38.64	1.73	0.01			

Apêndice A.2. Proteínas identificadas na amostra FS_Bagaço.

Locus gene	protein.Accession	protein.Description	protein.MW	ScoreAVG	ProductsAVG		PeptidesAVG		FmolAVG	FmolCV	Log10FmolAVG		ngramAVG	Repeate.rate	
TOP3AVG	CoverageAVG	TOP3CV	NgramVariance												
Cthe_2089C7HDZ8_CLOTM	Glycoside hydrolase family 48	83702.57	555.65	46.33	11.33	114.55	0.02	2.06	9.59	3	49018.67	22.72	0.01	0.04	
Cthe_3078C7HJX6_CLOTM	Cellulosome anchoring protein cohesin region			19238.62	292.07	12.00	4.00	18.00	0.03	1.26	0.35	2	7662.50	44.39	0.06
0.00															
Cthe_0798D1NIA6_CLOTM	Lipolytic protein G-D-S-L family	58243.20	366.20	15.00	4.33	67.69	0.07	1.83	3.94	3	28945.67	14.96	0.05	0.08	
Cther_2217	D1NIC8_CLOTM	Putative uncharacterized protein		14884.06	358.98	16.00	3.00	36.67	0.02	1.56	0.55	3	15693.33	34.09	0.01
0.00															
Cthe_0412D1NLS2_CLOTM	Glycoside hydrolase family 9	101339.39	157.26	24.00	9.33	36.82	0.02	1.57	3.73	3	15755.33	10.73	0.01	0.00	
Cthe_1104D1NMR1_CLOTM	Putative uncharacterized protein	19049.60	571.78	26.67	5.00	139.39	0.05	2.14	2.66	3	59618.67	30.46	0.03	0.02	
Cthe_2423D1NNW4_CLOTM	Type 3a cellulose-binding domain protein	112443.80	60.54	20.50	10.00	31.94	0.29	1.50	3.59	2	13883.50	14.89	0.29	1.05	
Cthe_3077E6UU82_CLOTL	Cellulosome anchoring protein cohesin region			143945.17	789.71	72.00	15.00	120.27	0.02	2.08	17.31	3	51493.00	17.75	0.04
0.14															
Cthe_0137H8EC95_CLOTM	Glyceraldehyde-3-phosphate dehydrogenase type I			36367.42	100.66	11.00	3.00	59.23	0.24	1.77	2.15	2	25713.50	11.46	0.23
0.26															
Cthe_1965H8ECT1_CLOTM	Peroxiredoxin	20918.57	140.42	8.67	2.67	22.57	0.06	1.35	0.47	3	9652.67	22.81	0.04	0.00	
Cthe_2348H8EDW1_CLOTM	Ig domain protein	113214.20	2184.85	155.33	26.33	367.90	0.07	2.57	41.65	3	157484.33	33.40	0.08	9.23	
Cthe_0421H8EEE9_CLOTM	Dipicolinate synthase subunit	20868.44	121.05	6.00	2.00	19.72	0.33	1.29	0.41	3	8458.00	18.73	0.34	0.02	
Cthe_0393H8EGU7_CLOTM	Sugar ABC transporter (Sugar-binding protein)			34182.35	186.47	11.67	3.33	59.81	0.20	1.78	2.04	3	25658.83	14.33	0.22
0.17															

Cthe_1020H8EH05_CLOTM 0.28	Extracellular solute-binding protein family 1	50085.69	2443.27	94.67	12.67	364.81	0.03	2.56	18.27	3	156172.33	39.00	0.04		

Apêndice A.3. Proteínas identificadas na amostra FS_Palha.

Locus gene	protein.Accession	protein.Description	protein.MW	ScoreAVG	ProductsAVG	PeptidesAVG	FmolAVG	FmolCV	Log10FmolAVG	ngramAVG	Repeate.rate		
TOP3AVG	CoverageAVG	TOP3CV	NgramVariance										
Cthe_2089C7HDZ8_CLOTM	Glycoside hydrolase family 48	83702.5726		293.48	31.67	8.67	31.04	0.87	1.49	2.60	3	14700.00	19.21	0.87	5.06
Cthe_3078C7HJX6_CLOTM 0.87	0.14	Cellulosome anchoring protein cohesin region	19238.6245		305.69	10.33	3.67	22.60	0.87	1.35	0.43	3	10586.00	20.60	
Cthe_0798D1NIA6_CLOTM	Lipolytic protein G-D-S-L family	58243.196	385.13	19.33	4.00	53.43	0.02	1.73	3.11	3	25205.67	8.97	0.01	0.01	
Cther_2217	D1NIC8_CLOTM	Putative uncharacterized protein	14884.0635		979.70	22.00	3.33	55.56	0.06	1.74	0.83	3	26226.33	36.62	
0.07	0.00														
Cthe_1368D1NJG7_CLOTM	S-layer domain protein	77935.3585		99.29	16.00	7.00	8.00	1.41	0.90	0.62	2	3830.00	11.62	1.41	0.78
Cthe_1104D1NMR1_CLOTM 0.00	Putative uncharacterized protein	19049.5963		979.38	31.00	6.33	110.49	0.03	2.04	2.10	3	52127.67	43.68	0.02	
Cthe_2089D1NQ73_CLOTM	Glycoside hydrolase family 48	84105.0909		293.48	31.67	9.00	15.25	1.73	1.18	1.28	3	7142.33	20.83	1.73	4.93
Cthe_3077E6UU82_CLOTL 0.48	54.28	Cellulosome anchoring protein cohesin region	143945.1691		441.03	50.00	13.00	105.42	0.49	2.02	15.17	3	49638.33	14.62	
Cthe_3078E6UU83_CLOTL 1.73	7.29	Cellulosome anchoring protein cohesin region	173102.9606		307.63	12.33	5.00	9.01	1.73	0.95	1.56	3	4313.00	3.63	
Cthe_1965H8ECT1_CLOTM	Peroxiredoxin	20918.5684	646.71	18.33	3.33	42.11	0.12	1.62	0.88	3	19881.00	22.82	0.13	0.01	
Cthe_2348H8EDW1_CLOTM	Ig domain protein	113214.1967	2729.35	227.67	33.67	453.30	0.03	2.66	51.32	3	213943.00	39.26	0.04	2.79	
Cthe_0217H8EED4_CLOTM	Glucose-6-phosphate isomerase	50401.4304		122.93	10.00	7.00	11.47	0.14	1.06	0.58	2	5441.00	23.33	0.15	0.01
Cthe_0393H8EGU7_CLOTM	Sugar ABC transporter	34182.3493		178.40	12.67	3.00	53.36	0.38	1.73	1.82	3	25257.67	15.16	0.40	0.49
Cthe_1020H8EH05_CLOTM 12.49	Extracellular solute-binding protein family	50085.6863		3340.46	128.67	13.67	429.10	0.16	2.63	21.49	3	202751.33	37.76	0.18	
Cthe_1368H8EI82_CLOTM	S-layer domain-containing protein	78164.551	96.36	15.50	6.50	8.70	1.41	0.94	0.68	2	4076.00	11.48	1.41	0.93	

Apêndice A.4. Proteínas identificadas na amostra FES_Celulose.

Locus gene	protein.Accession	protein.Description	protein.MW	ScoreAVG	ProductsAVG	PeptidesAVG	FmolCV	ngramAVGRepeate.rate	% of TSP
Cthe_2193D1NNT8_CLOTM	Xylanase	103411.1 2262.5	171.0	26.7	0.6	11.4	9	0.8	
Cthe_2872D1NQT1_CLOTM	Endo-β-1,4-glucanase	63655.5 7141.9	189.2	20.3	1.2	14.9	9	1.1	
Cthe_2193H8EB41_CLOTM	Xylanase	103351.0 2126.4	163.8	25.9	2.1	2.9	9	0.2	
Cthe_0536H8EBB4_CLOTM	Endo-β-1,4-glucanase	64157.2 3641.8	161.2	19.6	0.2	18.7	9	1.3	
Cthe_0821H8ECW7_CLOTM	Coagulation factor 58 type domain protein	63338.2 1125.0	63.0	12.8	0.4	2.6	9	0.2	
Cthe_0405H8EGT5_CLOTM	Endo-β-1,4-glucanase	60338.2 2591.3	94.0	16.7	0.2	5.8	9	0.4	
Cthe_2872H8EI72_CLOTM	Endo-β-1,4-glucanase	63641.5 6742.0	182.1	19.9	1.0	18.0	9	1.3	
Cthe_0269C7HGK4_CLOTM	Endo-β-1,4-glucanase	51657.5 32209.7	352.9	27.4	0.2	61.6	9	4.4	
Cthe_2360A3DHY3_CLOTM	Endo-β-1,4-glucanase	105277.5 1439.3	148.4	29.8	0.6	5.9	9	0.4	
Cthe_0433B0JFE7_CLOTM	Endo-β-1,4-glucanase	90230.6 1209.7	114.3	21.8	3.0	0.8	9	0.1	
Cthe_0625C7HDM7_CLOTM	Endo-β-1,4-glucanase	80224.8 36561.6	401.0	38.6	0.2	93.9	9	6.7	
Cthe_0412C7HHU0_CLOTM	Cellulose 1,4-beta-cellobiosidase - cellulase K			101249.4	69096.6	675.7	60.4	1.3	0.8 9 0.1
Cthe_0413C7HIJ7_CLOTM	Endo-β-1,4-glucanase	138625.6 40640.3	705.1	85.0	3.0	3.0	9	0.2	
Cthe_0274D1NLD7_CLOTM	Endo-β-1,4-glucanase	63176.1 972.2	77.4	14.4	3.0	0.7	9	0.1	
Cthe_0412D1NLS2_CLOTM	Cellulose 1,4-beta-cellobiosidase - cellulase K			101339.4	83013.5	837.0	75.7	0.2	112.2 9 8.0
Cthe_0413D1NLS3_CLOTM	Cellobiohydrolase	137914.9 40639.9	705.6	85.1	0.4	21.9	9	1.6	
Cthe_0578D1NM86_CLOTM	Endo-β-1,4-glucanase	82834.8 10326.8	373.1	45.1	1.2	17.3	9	1.2	
Cthe_2812D1NQQ3_CLOTM	Endo-β-1,4-glucanase	69154.8 5267.6	211.8	28.4	1.2	5.7	9	0.4	
Cthe_2360H8EAD5_CLOTM	Endo-β-1,4-glucanase	107498.4 1441.9	148.9	30.0	2.0	1.9	9	0.1	
Cthe_0578H8EB72_CLOTM	Endo-β-1,4-glucanase	82735.7 10314.1	370.4	44.4	1.0	23.5	9	1.7	

Cthe_0543H8EBA7_CLOTM	Endo-β-1,4-glucanase	82431.2	2409.2	152.9	23.1	0.2	6.9	9	0.5	
Cthe_0274H8EBK7_CLOTM	Endo-β-1,4-glucanase	63190.2	971.8	76.2	14.2	0.4	5.3	9	0.4	
Cthe_0825H8ECW5_CLOTM	Endo-β-1,4-glucanase	72700.6	650.4	60.9	15.7	0.2	3.7	9	0.3	
Cthe_0433H8EEG1_CLOTM	Endo-β-1,4-glucanase	90229.7	1211.5	114.1	21.9	0.5	3.6	9	0.3	
Cthe_2761H8EEL5_CLOTM	Endo-β-1,4-glucanase	81022.5	760.9	93.0	20.9	0.7	3.5	9	0.2	
Cthe_0745H8EFV0_CLOTM	Endo-β-1,4-glucanase	82640.4	11296.6	356.7	40.3	0.2	31.4	9	2.2	
Cthe_0043 H8EI14_CLOTM	Endo-β-1,4-glucanase	82538.4	1428.9	102.3	18.4	0.2	4.2	9	0.3	
Cthe_2360Q1H8P9_CLOTM	Endo-β-1,4-glucanase	105759.2	1327.8	136.6	28.0	3.0	0.7	9	0.1	
Cthe_0413Q59325_CLOTM	Endo-β-1,4-glucanase	138933.9	40324.6	651.3	77.1	3.0	16.8	9	1.2	
Cthe_2812Q8VV73_CLOTM	Endo-β-1,4-glucanase	69024.6	5192.1	208.6	27.8	1.0	6.1	9	0.4	
Cthe_0043 Q9L3J5_CLOTM	Endo-β-1,4-glucanase	82580.5	1230.2	84.3	16.4	3.0	0.6	9	0.0	
Cthe_1963D1NPW2_CLOTM	xylanase Z	92776.5	2308.5	134.6	24.8	0.2	5.9	9	0.4	
Cthe_0270H8EBK3_CLOTM	Glycoside hydrolase family 18	55571.6	334.0	32.2	9.1	1.3	1.6	9	0.1	
Cthe_2590H8EF39_CLOTM	β-1,4-xylanase	71812.6	684.1	63.4	14.0	0.2	2.5	9	0.2	
Cthe_0912H8EFD3_CLOTM	Xylanase Y	120406.2	2810.6	234.9	34.6	0.2	27.7	9	2.0	
Cthe_1838H8EHK9_CLOTM	Xylanase C	69689.2	16480.3	341.8	37.0	0.1	46.4	9	3.3	
Cthe_0270Q59326_CLOTM	Glycoside hydrolase family 18	55199.4	330.9	30.6	8.6	1.4	0.6	9	0.0	
Cthe_2972A3DJP0_CLOTM	acetylxylan esterase - CE4	74641.8	1764.1	71.6	16.9	3.0	0.0	9	0.0	
Cthe_2972D1NR31_CLOTM	acetylxylan esterase - CE4	74640.9	1805.4	77.4	17.7	0.4	2.8	9	0.2	
Cthe_2972D1NRP9_CLOTM	acetylxylan esterase - CE4	48337.9	959.8	41.1	9.9	2.6	0.3	9	0.0	
Cthe_2972O52780_CLOTM	acetylxylan esterase - CE4	74700.9	1713.2	69.6	16.3	2.6	0.4	9	0.0	
Cthe_2811H8EGS1_CLOTM	β-mannanase	67551.0	1605.9	99.6	18.4	0.8	3.2	9	0.2	
Cthe_2811Q9REC7_CLOTM	β-mannanase	67443.8	1208.0	90.6	17.3	1.6	1.2	9	0.1	
Cthe_3012H8ECH6_CLOTM	Carbohydrate binding family 6	70556.4	549.2	54.4	13.6	0.4	2.1	8	0.2	

Cthe_0661A3DD67_CLOTM	Ricin B lectin	64237.1	615.7	64.4	17.0	3.0	0.2	9	0.0						
Cthe_1271H8ED90_CLOTM	β-L-arabinofuranosidase	75017.2	646.8	76.3	20.4	0.5	1.8	9	0.1						
Cthe_0661H8EFL6_CLOTM	Ricin B lectin	64249.1	615.4	64.2	17.1	0.4	2.3	9	0.2						
Cthe_2089C7HDZ8_CLOTM	Glycoside hydrolase family 48	83702.6	35973.6	685.3	59.7	1.2	0.3	9	0.0						
Cthe_2089D1NQ73_CLOTM	Endo-β-1,4-glucanase	84105.1	36154.0	709.2	61.7	0.2	354.6	9	25.4						
Cthe_1400H8EIB0_CLOTM	Arabinogalactan endo-beta-1,4-galactanase			47466.2	447.4	41.1	12.8	0.3	0.9	8	0.1				
Cthe_1890A3DGN2_CLOTM	Putative uncharacterized protein		80490.5	3083.1	84.1	19.6	3.0	0.0	9	0.0					
Cthe_2236	A3DHK9_CLOTM	Flagellin domain protein	29567.3	562.2	17.3	3.7	3.0	0.1	9	0.0					
Clo1313_1399	D1NIC8_CLOTM	Putative uncharacterized protein		14884.1	2520.4	42.2	6.7	0.3	0.8	9	0.1				
Cthe_3228D1NKK8_CLOTM	Copper amine oxidase domain protein		30261.7	371.2	30.0	8.5	0.5	0.6	8	0.0					
Cthe_0423D1NLU9_CLOTM	Iron-containing alcohol dehydrogenase		96537.3	839.3	109.7	24.3	1.6	1.4	9	0.1					
Cthe_1104D1NMR1_CLOTM	Putative uncharacterized protein		19049.6	607.6	21.1	6.1	0.4	0.3	8	0.0					
Cthe_2267D1NNL6_CLOTM	V-type ATP synthase alpha chain		65535.0	305.3	46.7	17.5	0.6	2.5	6	0.2					
Cthe_2237D1NNP6_CLOTM	Flagellin domain protein	29431.5	654.8	22.8	5.0	1.5	0.2	9	0.0						
Cthe_2236	D1NNP7_CLOTM	Flagellin domain protein	29595.4	616.6	19.8	4.4	2.3	0.0	8	0.0					
Cthe_1890D1NPP3_CLOTM	Dockerin type 1	71820.2	3264.1	90.7	20.6	0.3	6.1	9	0.4						
Cthe_2834D1NRJ8_CLOTM	Putative esterase	64157.5	967.9	34.4	5.9	1.6	0.9	9	0.1						
Cthe_0190H8EA87_CLOTM	PspAIM30 family protein	27756.6	675.0	27.0	8.6	0.2	0.7	8	0.0						
Cthe_2393H8EAA3_CLOTM	Thiamine pyrophosphate TPP-binding domain-containing protein					34991.4	1097.5	42.3	8.5	0.4	1.4	8	0.1		
Cthe_2392H8EAA4_CLOTM	Pyruvate flavodoxinferredoxin oxidoreductase domain protein				43632.8	701.6	55.0	13.6	0.3	1.6	7	0.1			
Cthe_2390H8EAA6_CLOTM	Pyruvateketoisovalerate oxidoreductase gamma subunit				21260.5	1758.4	30.8	5.6	0.4	0.5	8	0.0			
Cthe_2383H8EAB3_CLOTM	Copper amine oxidase-like domain-containing protein			33203.5	395.9	30.4	8.9	0.4	0.5	7	0.0				
Cthe_2090H8EAT4_CLOTM	Uncharacterized protein	22050.4	374.4	29.3	7.6	0.2	0.3	8	0.0						
Cthe_0137H8EC95_CLOTM	Glyceraldehyde-3-phosphate dehydrogenase type I			36367.4	6048.2	125.0	17.1	0.4	7.4	9	0.5				

Cthe_0138	H8EC96_CLOTM	Phosphoglycerate kinase	42876.6	289.6	23.1	8.6	0.3	0.6	9	0.0
Cthe_0143	H8ECA0_CLOTM	Enolase	47159.7	728.7	45.9	9.9	0.4	1.9	8	0.1
Cthe_2236	H8ECD9_CLOTM	Flagellin domain protein	26368.7	667.6	19.3	4.0	1.3	0.1	7	0.0
Cthe_3035	H8ECJ9_CLOTM	D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding protein	43147.8	323.3	22.3	8.2	0.3	0.7	6	0.0
Cthe_0347	H8ECN6_CLOTM	Phosphofructokinase	46023.3	714.8	44.2	11.0	0.3	1.8	6	0.1
Cthe_0349	H8ECN7_CLOTM	Fructose-1 6-bisphosphate aldolase class II	33917.8	308.9	25.0	9.8	0.8	0.7	9	0.0
Cthe_1965	H8ECT1_CLOTM	Peroxiredoxin	20918.6	1655.5	31.3	5.8	0.5	0.7	9	0.0
Cthe_1308	H8ED52_CLOTM	Pyruvate phosphate dikinase	99363.7	280.5	55.0	19.8	0.2	1.6	6	0.1
Cthe_2348	H8EDW1_CLOTM	Ig domain protein	113214.2	309.4	73.0	23.3	1.5	1.1	6	0.1
Cthe_0190	H8EEA5_CLOTM	Proteinase inhibitor I4 serpin	68045.4	366.2	40.0	12.3	0.2	1.2	9	0.1
Cthe_0423	H8EEF1_CLOTM	Iron-containing alcohol dehydrogenase	96523.3	833.2	108.6	24.0	0.8	2.7	9	0.2
Cthe_0424	H8EEF2_CLOTM	Aminoglycoside phosphotransferase	28701.0	1266.7	33.9	8.0	0.3	0.8	7	0.1
Cthe_2730	H8EEP5_CLOTM	Elongation factor Tu	44276.5	435.7	41.0	12.4	0.2	0.8	9	0.1
Cthe_2723	H8EEQ2_CLOTM	50S ribosomal protein L7L12 OS=Clostridium thermocellum AD2 GN=rplL PE=3 SV=1	13295.4	1992.6	32.4	5.6	0.6	0.4	9	0.0
Cthe_2657	H8EEX1_CLOTM	Histone family protein DNA-binding protein	10087.7	982.7	18.0	4.9	0.4	0.3	7	0.0
Cthe_0640	H8EFJ2_CLOTM	Dockerin type 1 protein	64784.5	282.6	39.5	13.3	0.4	0.9	8	0.1
Cthe_2383	H8EGG2_CLOTM	Copper amine oxidase-like domain-containing protein	59309.6	697.0	35.9	9.3	0.2	2.4	9	0.2
Clo1313_0638	H8EGI8_CLOTM	Histone family protein DNA-binding protein	9728.3	5267.4	43.9	5.6	0.4	0.8	9	0.1
Cthe_1503	H8EGN3_CLOTM	Uncharacterized protein	13231.1	646.5	13.9	3.6	0.4	0.2	7	0.0
Cthe_0393	H8EGU7_CLOTM	Sugar ABC transporter (Sugar-binding protein)	34182.3	3814.6	63.6	9.4	0.6	3.1	9	0.2
Cthe_0388	H8EGV2_CLOTM	Alcohol dehydrogenase GroES domain protein	38508.2	554.3	40.5	13.8	0.4	1.0	6	0.1
Cthe_1020	H8EH05_CLOTM	ABC transporter substrate-binding protein	50085.7	23636.1	426.0	43.2	0.4	44.5	9	3.2
Cthe_2038	H8EHE5_CLOTM	Dockerin type 1 protein	93232.5	858.8	111.6	27.2	0.1	3.8	9	0.3
Cthe_3169	H8EHS4_CLOTM	Enoyl-[acyl-carrier-protein] reductase [NADH]	27101.0	508.5	22.9	5.1	0.5	0.5	7	0.0

Cthe_2237H8EI41_CLOTM	Flagellin domain protein (Fragment)	26204.8	643.9	21.6	4.6	0.8	0.4	9	0.0					
Cthe_2892H8EI52_CLOTM	60 kDa chaperonin OS=Clostridium thermocellum AD2 GN=groL PE=3 SV=1	57531.2	3076.6	134.8	21.5	0.5	5.7	8	0.4					
Cthe_1398H8EIA8_CLOTM	Dockerin type 1 protein	92707.5	966.0	108.1	24.7	0.2	6.9	9	0.5					
YSBL_2970	H8ERX6_CLOTM	S-layer domain-containing protein (Fragment)	33738.1	820.7	25.6	4.6	2.8	0.0	8	0.0				
Cthe_2348O86999_CLOTM	S-layer protein	113387.4	300.5	69.9	23.1	0.7	1.9	7	0.1					
Cthe_0624H8EFH5_CLOTM	Endo-β-1,4-glucanase	178543.3	6844.3	427.8	50.1	0.1	65.7	9	4.7					
Cthe_3077D1NHZ0_CLOTM	cellulosomal-scaffolding protein A	178367.9	75477.8	568.3	46.3	1.5	86.1	9	6.2					
Cthe_3078D1NHZ1_CLOTM	Cellulosome anchoring protein cohesin region	192072.7	2220.0	72.4	13.3	3.0	1.8	9	0.1					
Cthe_1307D1NJA2_CLOTM	Cellulosome anchoring protein cohesin region	68676.5	447.6	70.4	16.1	3.0	0.3	9	0.0					
Cthe_0736D1NMI8_CLOTM	Cellulosome anchoring protein cohesin region	140645.7	1440.5	214.3	39.8	3.0	1.1	9	0.1					
Cthe_3077E6UU82_CLOTM	cellulosomal-scaffolding protein A	143945.2	75473.9	566.8	46.2	0.8	129.8	9	9.3					
Cthe_3078E6UU83_CLOTM	Cellulosome anchoring protein cohesin region	173103.0	2219.5	72.2	13.1	0.5	8.4	9	0.6					
Cthe_3077H8EAE8_CLOTM	cellulosomal-scaffolding protein A	36765.6	18454.9	178.9	14.7	3.0	3.7	9	0.3					
Cthe_1307H8ED53_CLOTM	Cellulosome anchoring protein cohesin region	68732.6	445.6	70.3	16.2	0.5	1.6	9	0.1					
Cthe_0736H8EFU0_CLOTM	Cellulosome anchoring protein cohesin region	140645.7	1440.5	214.3	39.8	0.4	8.0	9	0.6					
Cthe_3079H8EGH9_CLOTM	Cellulosome anchoring protein cohesin region	75028.2	918.0	44.5	9.3	0.6	2.8	8	0.2					
Cthe_3080H8EGI0_CLOTM	Cellulosome anchoring protein cohesin region	48331.8	402.5	37.1	11.3	0.5	1.0	8	0.1					
Cthe_3202D1NKI2_CLOTM	CRISPR-associated protein Csh2 family	34882.8	329.5	20.2	6.5	0.2	0.4	6	0.0					
Cthe_2179D1NNV1_CLOTM	Pectate lyaseAmb allergen	100782.7	4522.4	256.9	42.9	0.2	12.1	9	0.9					
Cthe_2179H8EB28_CLOTM	Pectate lyaseAmb allergen	62099.6	2633.4	146.0	21.4	2.1	0.0	9	0.0					

Apêndice A.5. Proteínas identificadas na amostra FES_Bagaço.

Locus gene	protein.Accession	protein.Description	protein.MW	ScoreAVG	ProductsAVG	PeptidesAVG	FmolAVG	FmolCV	Log10FmolAVG	ngramAVGRepeate.rate	TOP3AVG	CoverageAVG	TOP3CV	NgramVariance

Cthe_2089C7ED31_CLOTL 0.75	Glycosyl hydrolase family 48 protein CelS (Fragment)	43068.50	206.83	21.33	5.33	11.59	1.73	1.06	0.50	3.00	7116.67	19.42	1.73		
Cthe_2089C7HDZ8_CLOTM	Glycoside hydrolase family 48	83702.57	220.90	28.00	8.67	38.88	0.88	1.59	3.25	3.00	24039.67	22.09	0.90	8.13	
Cthe_0412C7HHU0_CLOTM	Glycoside hydrolase family 9	101249.37	342.31	44.00	14.00	0.65	1.73	-0.18	0.07	3.00	431.33	26.52	1.73	0.01	
Cthe_3202D1NKI2_CLOTM 0.02	CRISPR-associated protein Csh2 family	34882.80	140.41	19.67	6.67	21.66	0.20	1.34	0.76	3.00	13276.33	28.63	0.19		
Cthe_0412D1NLS2_CLOTM	Glycoside hydrolase family 9	101339.39	355.69	45.67	14.00	30.93	0.07	1.49	3.13	3.00	18936.67	27.75	0.03	0.05	
Cthe_1104D1NMR1_CLOTM	Putative uncharacterized protein	19049.60	522.47	20.00	5.00	60.10	0.10	1.78	1.14	3.00	37038.67	49.04	0.16	0.01	
Cthe_3077E6UU82_CLOTL 0.09 0.20	Cellulosome anchoring protein cohesin region	143945.17	201.87	45.00	17.67	38.24	0.08	1.58	5.50	3.00	23456.33	21.38			
Cthe_2393H8EAA3_CLOTM 0.27 0.04	Thiamine pyrophosphate TPP-binding domain-containing protein	34991.36	166.93	13.33	3.67	17.66	0.32	1.25	0.62	3.00	10692.00	17.15			
Cthe_2390H8EAA6_CLOTM 0.04 0.00	Pyruvateketoisovalerate oxidoreductase gamma subunit	21260.55	277.01	12.00	3.50	21.97	0.02	1.34	0.47	2.00	12987.50	18.23			
Cthe_2379H8EAB7_CLOTM	Uncharacterized protein	19064.97	101.24	6.00	2.50	37.87	0.01	1.58	0.72	2.00	22402.50	12.13	0.07	0.00	
Cthe_0137H8EC95_CLOTM 0.01	Glyceraldehyde-3-phosphate dehydrogenase type I	36367.42	1520.32	39.67	4.33	145.22	0.02	2.16	5.28	3.00	89184.00	20.24	0.08		
Cthe_0138H8EC96_CLOTM	Phosphoglycerate kinase	42876.57	324.67	29.67	9.67	40.28	0.04	1.61	1.73	3.00	24738.67	31.15	0.09	0.01	
Cthe_0139H8EC97_CLOTM	Triosephosphate isomerase	27362.25	385.87	20.33	5.67	36.06	0.18	1.56	0.99	3.00	21970.67	37.05	0.13	0.03	
Cthe_0143H8ECA0_CLOTM	Enolase	47159.67	601.86	41.33	8.33	59.58	0.15	1.78	2.81	3.00	36777.33	33.56	0.21	0.18	
Cthe_2236H8ECD9_CLOTM	Flagellin domain protein (Fragment)	26368.71	655.11	19.33	5.33	46.13	0.90	1.66	1.22	3.00	29564.00	29.24	0.92	1.21	
Cthe_0347H8ECN6_CLOTM	Phosphofructokinase	46023.32	326.52	35.67	11.33	42.44	0.05	1.63	1.95	3.00	26000.33	37.27	0.02	0.01	
Cthe_0349H8ECN7_CLOTM	Fructose-1 6-bisphosphate aldolase class II	33917.79	207.06	21.00	5.33	52.63	0.19	1.72	1.78	3.00	32583.50	24.49	0.26	0.11	
Cthe_1965H8ECT1_CLOTM	Peroxiredoxin	20918.57	3593.12	51.33	7.00	112.27	0.02	2.05	2.35	3.00	69016.33	45.63	0.09	0.00	
Cthe_1308H8ED52_CLOTM	Pyruvate phosphate dikinase	99363.67	187.87	25.67	10.00	42.42	0.06	1.63	4.22	3.00	26021.00	18.69	0.07	0.06	
Cthe_1795H8EDK3_CLOTM 0.06 0.00	Phospho-2-dehydro-3-deoxyheptonate aldolase	37088.88	147.50	16.67	8.33	13.98	0.02	1.15	0.52	3.00	8581.67	30.11			

Cthe_2348H8EDW1_CLOTM	Ig domain protein	113214.20	457.87	83.67	23.67	110.62	0.09	2.04	12.52	3.00	68141.33	35.62	0.14	1.16		
Cthe_0217H8EED4_CLOTM	Glucose-6-phosphate isomerase	50401.43	177.12	24.33	11.00	20.15	0.09	1.30	1.02	3.00	12414.67	33.78	0.15	0.01		
Cthe_0423H8EEF1_CLOTM	Iron-containing alcohol dehydrogenase	96523.27	177.95	37.33	13.00	33.10	0.05	1.52	3.19	3.00	20328.67	23.44	0.09	0.02		
Cthe_2730H8EEP5_CLOTM	Elongation factor Tu	44276.55	1818.18	65.67	10.33	156.12	0.01	2.19	6.91	3.00	95847.67	38.75	0.07	0.01		
Cthe_2723H8EEQ2_CLOTM	50S ribosomal protein L7L12	13295.45	913.52	25.00	5.33	37.26	0.26	1.57	0.50	3.00	22609.67	62.79	0.20	0.02		
Cthe_0637H8EFI9_CLOTM	Uncharacterized protein	11114.65	298.50	6.67	1.67	49.84	0.03	1.70	0.55	3.00	30644.00	26.67	0.09	0.00		
Cthe_2518H8EGB1_CLOTM	Ketol-acid reductoisomerase	36393.66	279.90	23.00	8.67	29.06	0.05	1.46	1.06	3.00	17866.67	35.95	0.11	0.00		
Clo1313_0638	H8EGI8_CLOTM	Histone family protein DNA-binding protein	9728.33	344.50	6.67	2.00	37.72	0.44	1.58	0.37	3.00	23562.50	23.81	0.51	0.03	
Cthe_0395H8EGU5_CLOTM	RbsD or FucU transport	15674.93	505.22	13.33	3.67	20.46	0.27	1.31	0.32	3.00	12703.33	40.43	0.34	0.01		
Cthe_0394H8EGU6_CLOTM	Iron-containing alcohol dehydrogenase	42472.46	341.06	28.00	7.67	57.37	0.05	1.76	2.44	3.00	35205.33	27.16	0.07	0.01		
Cthe_0393H8EGU7_CLOTM	Sugar ABC transporter (Sugar-binding protein)	34182.35	1158.01	45.00	7.67	101.43	0.02	2.01	3.47	3.00	62229.67	38.63	0.05	0.01		
Cthe_0388H8EGV2_CLOTM	Alcohol dehydrogenase GroES domain protein	38508.20	228.96	21.33	7.00	85.85	0.22	1.93	3.31	3.00	52200.17	27.52	0.17	0.55		
Cthe_1020H8EH05_CLOTM	Extracellular solute-binding protein family 1	50085.69	4826.65	141.33	14.00	589.44	0.10	2.77	29.52	3.00	360573.67	42.55	0.06	8.75		
Cthe_2237H8EI41_CLOTM	Flagellin domain protein (Fragment)	26204.79	726.89	23.67	6.00	15.03	1.33	1.18	0.39	3.00	8661.17	29.79	1.31	0.27		
Cthe_2892H8EI52_CLOTM	60 kDa chaperonin	57531.22	818.53	69.33	15.67	102.56	0.04	2.01	5.90	3.00	62878.00	37.28	0.04	0.06		
Cthe_2891H8EI53_CLOTM	10 kDa chaperonin	10155.82	696.82	22.67	4.00	30.69	0.13	1.49	0.31	3.00	18916.67	52.84	0.19	0.00		
Cthe_2874H8EI70_CLOTM	Phosphoenolpyruvate carboxykinase [GTP]	68140.76	657.06	48.67	14.67	62.81	0.07	1.80	4.28	3.00	38701.67	33.44	0.14	0.09		

Apêndice A.6. Proteínas identificadas na amostra FES_Palha.

Locus	gene	protein.Accession	protein.Description	protein.MW	ScoreAVG	ProductsAVG	PeptidesAVG	FmolAVG	FmolCV	Log10FmolAVG	ngramAVG	Repeate.rate	TOP3AVG	CoverageAVG	TOP3CV	NgramVariance
Cthe_0798C7HD86_CLOTM	Cellulosome protein dockerin type I	22069.20	277.60	9.00	3.00	9.70	0.35	0.99	0.21	3	11142.00	16.08	0.40	0.01		

Cthe_0625C7HDM7_CLOTM	Glycoside hydrolase family 9		80224.77	426.91	38.33	12.33	20.58	0.12	1.31	1.65	3	23317.00	28.02	0.09	0.04	
Cthe_2089C7HDZ8_CLOTM	Glycoside hydrolase family 48	83702.57	6161.68	126.00	17.67	196.09	0.87	2.29	16.41	3	226398.33	33.74	0.87	202.04		
Cthe_0269C7HGK4_CLOTM	Glycoside hydrolase family 8		51657.52	166.16	18.00	9.00	10.87	0.30	1.04	0.56	3	12372.00	17.24	0.30	0.03	
Clo1313_1399	D1NIC8_CLOTM	Putative uncharacterized protein		14884.06	2179.32	20.00	3.00	23.56	0.04	1.37	0.35	3	26788.00	34.09	0.09	0.00
Cthe_1368D1NJG7_CLOTM	S-layer domain protein	77935.36	363.11	39.00	13.00	41.14	0.87	1.61	3.21	3	44811.00	27.89	0.87	7.72		
Cthe_3202D1NKI2_CLOTM	CRISPR-associated protein Csh2 family			34882.80	346.42	36.00	11.00	48.44	1.02	1.69	1.69	3	57749.33	46.12	1.08	2.95
Cthe_0043D1NKP9_CLOTM	Glycoside hydrolase family 9		59355.37	201.35	14.67	5.67	6.87	0.87	0.84	0.41	3	7489.00	15.95	0.87	0.13	
Cthe_0275D1NLD9_CLOTM	Glycosyltransferase 36	93507.18	437.76	43.33	13.67	33.76	0.05	1.53	3.16	3	38376.67	26.59	0.09	0.03		
Cthe_0338D1NLK2_CLOTM	NADH-quinone oxidoreductase E subunit		18640.31	1299.12	12.33	2.00	24.26	0.08	1.38	0.45	3	27657.00	21.21	0.14	0.00	
Cthe_2238D1NLK9_CLOTM	L-lactate dehydrogenase	35076.44	822.13	15.33	3.33	5.51	1.73	0.74	0.19	3	6070.33	13.63	1.73	0.11		
Cthe_0412D1NLS2_CLOTM	Glycoside hydrolase family 9		101339.39	8794.18	181.33	28.33	124.24	0.02	2.09	12.59	3	141327.67	38.92	0.09	0.05	
Cthe_0413D1NLS3_CLOTM	Glycoside hydrolase family 9		137914.91	2341.14	88.33	24.67	18.10	0.01	1.26	2.50	3	20586.00	22.90	0.08	0.00	
Cthe_0423D1NLU9_CLOTM	Iron-containing alcohol dehydrogenase		96537.30	1417.43	83.00	17.33	42.52	0.06	1.63	4.10	3	48221.67	28.98	0.05	0.07	
Cthe_1104D1NMR1_CLOTM	Putative uncharacterized protein		19049.60	5772.21	63.33	9.33	83.37	0.03	1.92	1.59	3	94824.67	51.72	0.09	0.00	
Cthe_2348D1NP50_CLOTM	Ig domain protein	113170.14	4754.80	174.00	31.00	51.77	1.73	1.71	5.86	3	63747.00	38.45	1.73	102.98		
Cthe_2010D1NQ12_CLOTM	Putative uncharacterized protein		48724.20	132.19	10.50	5.00	2.51	1.41	0.40	0.12	2	2708.00	12.62	1.41	0.03	
Cthe_2855D1NQ15_CLOTM	Putative uncharacterized protein		49016.67	148.66	12.00	10.00	6.16	0.37	0.79	0.30	2	6690.00	25.18	0.35	0.01	
Cthe_2089D1NQ73_CLOTM	Glycoside hydrolase family 48	84105.09	6161.79	127.00	18.67	196.24	0.87	2.29	16.50	3	213792.00	34.90	0.87	204.31		
Cthe_3077E6UU82_CLOTL	Cellulosome anchoring protein cohesin region			143945.17	17705.83	243.00	24.33	152.28	0.87	2.18	21.92	3	176522.67	27.32	0.89	364.43
Cthe_3078E6UU83_CLOTL	Cellulosome anchoring protein cohesin region			173102.96	608.77	20.67	9.33	24.35	0.43	1.39	4.22	3	27572.00	6.89	0.41	3.33
Cthe_2422H8EA72_CLOTM	Uncharacterized protein	40608.41	138.08	13.33	6.00	8.98	0.25	0.95	0.36	3	10309.67	16.34	0.31	0.01		
Cthe_2393H8EAA3_CLOTM	Thiamine pyrophosphate TPP-binding domain-containing protein			34991.36	537.12	14.67	4.00	12.03	0.39	1.08	0.42	3	13590.00	14.90	0.37	0.03

Cthe_2392H8EAA4_CLOTM 0.10 0.00	Pyruvate flavodoxinferredoxin oxidoreductase domain protein				43632.82	513.08	26.67	6.67	26.99	0.06	1.43	1.18	3	30706.67	18.28	
Cthe_2390H8EAA6_CLOTM 0.07 0.00	Pyruvateketoisovalerate oxidoreductase gamma subunit				21260.55	983.85	16.33	4.67	14.12	0.00	1.15	0.30	3	16048.00	27.60	
Cthe_2855H8EAE5_CLOTM	Uncharacterized protein	49124.55	133.65	12.00	7.50	0.30	1.41	-0.52	0.01	2	332.50	13.84	1.41	0.00		
Cthe_1228H8EAN1_CLOTM	Threonine--tRNA ligase	73850.33	144.35	19.00	8.00	6.43	0.04	0.81	0.47	2	7486.50	11.97	0.04	0.00		
Cthe_0578H8EB72_CLOTM	Glycoside hydrolase family 9		82735.71	114.28	18.50	9.50	14.19	0.06	1.15	1.17	2	16585.50	14.20	0.14	0.01	
Cthe_0536H8EBB4_CLOTM	Glycoside hydrolase family 5		64157.22	180.76	11.00	8.50	7.66	0.29	0.88	0.49	2	8832.75	11.99	0.21	0.02	
Cthe_0285H8EBL8_CLOTM	Isocitrate dehydrogenase [NADP]		45953.47	126.65	13.67	7.00	7.35	0.03	0.87	0.34	3	8368.67	16.67	0.10	0.00	
Cthe_0933H8EBW2_CLOTM	Acyl carrier protein	8367.37	267.71	4.50	1.00	14.18	0.11	1.15	0.12	2	16611.00	16.22	0.18	0.00		
Cthe_0137H8EC95_CLOTM 0.00	Glyceraldehyde-3-phosphate dehydrogenase type I				36367.42	2168.34	43.67	5.67	65.63	0.03	1.82	2.39	3	74693.33	27.78	0.10
Cthe_0138H8EC96_CLOTM	Phosphoglycerate kinase	42876.57	1021.21	32.33	7.00	31.75	0.01	1.50	1.36	3	36103.33	17.97	0.08	0.00		
Cthe_0139H8EC97_CLOTM	Triosephosphate isomerase		27362.25	1934.47	27.67	5.00	39.44	0.05	1.60	1.08	3	44821.00	29.48	0.08	0.00	
Cthe_0143H8ECA0_CLOTM	Enolase	47159.67	1028.31	30.67	10.00	27.52	0.02	1.44	1.30	3	31266.00	34.72	0.06	0.00		
Cthe_2203H8ECA6_CLOTM	GTP cyclohydrolase	20981.72	343.42	9.00	2.67	5.50	0.22	0.74	0.12	3	6241.50	19.43	0.21	0.00		
Cthe_2236H8ECD9_CLOTM	Flagellin domain protein (Fragment)		26368.71	870.36	15.33	4.33	9.95	1.24	1.00	0.26	3	11973.50	24.32	1.28	0.11	
Cthe_0275H8ECF3_CLOTM	Glycosyltransferase 36	112286.51	163.19	32.67	18.33	28.43	0.86	1.45	3.19	3	31270.33	17.45	0.82	7.47		
Cthe_3035H8ECJ9_CLOTM 16835.00 19.78	D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding protein					43147.77	511.21	17.67	6.67	14.81	0.01	1.17	0.64	3		
	0.08 0.00															
Cthe_0345H8ECN4_CLOTM	L-lactate dehydrogenase	35004.37	758.93	13.67	3.00	18.90	1.06	1.28	0.66	3	21229.33	9.43	1.01	0.49		
Cthe_0347H8ECN6_CLOTM	Phosphofructokinase	46023.32	1077.34	46.00	11.33	35.48	0.03	1.55	1.63	3	40366.67	42.89	0.10	0.00		
Cthe_0349H8ECN7_CLOTM	Fructose-1 6-bisphosphate aldolase class II		33917.79	777.57	33.67	7.00	32.42	0.07	1.51	1.10	3	36884.67	21.58	0.11	0.01	
Cthe_1963H8ECS9_CLOTM	Glycoside hydrolase family 10	92790.48	130.70	19.00	11.00	7.73	0.22	0.89	0.72	3	8832.67	14.38	0.26	0.03		
Cthe_1965H8ECT1_CLOTM	Peroxiredoxin	20918.57	9570.72	52.67	6.67	60.58	0.02	1.78	1.27	3	68928.67	45.63	0.09	0.00		
Cthe_1970H8ECT7_CLOTM	Uncharacterized protein	11122.40	317.17	8.00	2.00	9.45	0.04	0.98	0.11	3	10764.50	21.78	0.11	0.00		

Cthe_1322H8ED36_CLOTM	Chaperone protein DnaK	65795.51	154.18	19.33	8.00	7.40	0.07	0.87	0.49	3	8419.00	17.05	0.12	0.00	
Cthe_1308H8ED52_CLOTM	Pyruvate phosphate dikinase		99363.67	294.24	25.00	8.33	24.65	0.07	1.39	2.45	3	28093.67	11.78	0.14	0.03
Cthe_1754H8EDP3_CLOTM	ABC-type transporter periplasmic subunit		35912.04	269.25	15.67	9.33	10.06	0.04	1.00	0.36	3	11436.33	24.74	0.09	0.00
Cthe_2348H8EDW1_CLOTM	Ig domain protein	113214.20	4738.37	172.67	30.67	101.70	0.87	2.01	11.51	3	110738.00	37.58	0.87	100.03	
Cthe_2301H8EDZ8_CLOTM 0.00	CRISPR-associated regulatory protein DevR family		32757.06	461.46	15.00	6.50	7.25	0.21	0.86	0.24	2	8524.25	18.71	0.29	
Cthe_0217H8EED4_CLOTM	Glucose-6-phosphate isomerase	50401.43	680.54	31.67	9.33	20.98	0.01	1.32	1.06	3	23854.33	25.22	0.08	0.00	
Cthe_0418H8EEE6_CLOTM	Polyribonucleotide nucleotidyltransferase		77403.07	170.85	12.50	5.50	5.50	0.36	0.74	0.43	2	6507.00	9.79	0.44	0.02
Cthe_0423H8EEF1_CLOTM	Iron-containing alcohol dehydrogenase		96523.27	1394.62	78.00	17.67	0.16	1.73	-0.80	0.02	3	193.00	29.21	1.73	0.00
Cthe_0424H8EEF2_CLOTM 0.00	Aminoglycoside phosphotransferase			28701.00	386.66	8.50	4.00	12.81	0.17	1.11	0.37	2	15042.25	19.48	0.24
Cthe_2730H8EEP5_CLOTM	Elongation factor Tu	44276.55	9781.53	94.33	10.67	147.97	0.03	2.17	6.55	3	168433.67	38.67	0.10	0.05	
Cthe_2723H8EEQ2_CLOTM	50S ribosomal protein L7L12	13295.45	1940.74	18.67	4.00	15.85	0.08	1.20	0.21	3	17970.67	46.77	0.06	0.00	
Cthe_2708H8EER5_CLOTM	ABC-type uncharacterized transport system		53650.37	179.18	12.50	6.50	3.58	0.08	0.55	0.19	2	3906.50	16.81	0.09	0.00
Cthe_2608H8EF21_CLOTM	ATP synthase subunit beta	51069.71	378.35	21.00	10.50	6.24	0.02	0.80	0.32	2	7212.00	25.76	0.12	0.00	
Cthe_0624H8EFH5_CLOTM	Glycoside hydrolase family 9		178543.32	777.67	123.33	33.33	37.03	0.09	1.57	6.61	3	42115.33	30.89	0.12	0.37
Cthe_0699H8EFQ1_CLOTM	Carboxyl transferase	56219.46	155.73	20.50	7.00	12.16	0.37	1.09	0.68	2	14394.50	17.25	0.44	0.06	
Cthe_0701H8EFQ4_CLOTM 0.00	Carboxylase region-containing protein			52770.89	159.66	23.50	11.00	6.50	0.18	0.81	0.34	2	7629.50	23.98	0.26
Cthe_2924H8EG36_CLOTM	Adenylate kinase	24754.16	410.76	7.67	4.00	2.78	0.14	0.44	0.07	3	3180.67	19.97	0.22	0.00	
Cthe_2518H8EGB1_CLOTM	Ketol-acid reductoisomerase		36393.66	798.60	22.00	4.67	25.43	0.33	1.41	0.93	3	28746.00	23.06	0.30	0.09
Cthe_2534H8EGC7_CLOTM	Sulfate ABC transporter ATPase subunit		39956.99	95.20	8.50	7.50	8.05	0.04	0.91	0.32	2	9317.50	23.87	0.14	0.00
Cthe_2383H8EGG2_CLOTM 0.00	Copper amine oxidase-like domain-containing protein		59309.58	128.48	9.00	7.00	5.35	0.10	0.73	0.32	2	6210.50	10.84	0.02	
Cthe_3080H8EGI0_CLOTM 0.00	Cellulosome anchoring protein cohesin region			48331.81	137.74	14.50	9.00	3.72	0.19	0.57	0.18	2	4368.00	13.94	0.27

Clo1313_0638 0.00	H8EGI8_CLOTM	Histone family protein DNA-binding protein	9728.33	1051.56	7.67	1.33	28.05	0.02	1.45	0.27	3	31902.00	17.58	0.09			
Cthe_0394	H8EGU6_CLOTM	Iron-containing alcohol dehydrogenase	42472.46	216.98	13.00	7.33	7.73	0.17	0.89	0.33	3	8817.00	21.34	0.21	0.00		
Cthe_0393 0.04	H8EGU7_CLOTM	Sugar ABC transporter (Sugar-binding protein)	34182.35	337.99	14.33	6.33	14.54	0.42	1.16	0.50	3	16755.00	30.84	0.46			
Cthe_0374	H8EGW6_CLOTM	Glutamate dehydrogenase	49002.98	227.86	34.00	14.00	13.05	0.46	1.12	0.64	2	14261.50	36.26	0.48	0.09		
Cthe_1020 0.07	H8EH05_CLOTM 1.04	Extracellular solute-binding protein family 1	50085.69	20774.47	213.00	13.00	2495.42	0.01	3.40	124.98	3	2836780.67		41.03			
Cthe_1028	H8EH13_CLOTM	Acetate kinase	44365.13	165.94	12.33	7.00	5.12	0.16	0.71	0.23	3	5804.33	18.88	0.15	0.00		
Cthe_1082	H8EH70_CLOTM	Uncharacterized protein	9725.81	788.35	12.67	4.33	16.18	0.04	1.21	0.16	3	18399.00	36.86	0.09	0.00		
Cthe_1862	H8EHI5_CLOTM	ABC transporter related protein	42177.95	366.13	14.50	4.50	41.99	0.03	1.62	1.77	2	45762.00	16.09	0.05	0.00		
Cthe_1840	H8EHK7_CLOTM	Cysteine synthase	33350.23	234.92	9.00	5.50	7.21	0.24	0.86	0.24	2	8413.50	20.10	0.33	0.00		
Cthe_1838	H8EHK9_CLOTM	Glycoside hydrolase family 10	69689.21	297.71	28.67	9.00	20.35	0.06	1.31	1.42	3	23083.00	17.02	0.05	0.01		
Cthe_0043	H8EI14_CLOTM	Glycoside hydrolase family 9	82538.42	205.09	18.67	7.33	3.49	1.73	0.54	0.29	3	4301.00	13.57	1.73	0.25		
Cthe_2237	H8EI41_CLOTM	Flagellin domain protein (Fragment)	26204.79	1620.72	19.67	4.00	18.13	0.87	1.26	0.48	3	19745.50	22.31	0.87	0.17		
Cthe_2892	H8EI52_CLOTM	60 kDa chaperonin	57531.22	1607.72	63.00	12.00	49.23	0.01	1.69	2.83	3	55982.33	27.36	0.08	0.00		
Cthe_2885 0.52	H8EI59_CLOTM 0.01	Phosphoribosylaminoimidazole-succinocarboxamide synthase	33343.13	108.67	11.00	7.67	6.04	0.50	0.78	0.20	3	6961.50	21.43				
Cthe_2874	H8EI70_CLOTM	Phosphoenolpyruvate carboxykinase [GTP]	68140.76	972.77	47.67	13.33	30.03	0.05	1.48	2.05	3	34198.33	28.21	0.12	0.01		
Cthe_2872	H8EI72_CLOTM	Glycoside hydrolase family 5	63641.51	143.29	13.00	6.50	6.31	0.34	0.80	0.40	2	7255.00	8.92	0.27	0.02		
Cthe_1368	H8EI82_CLOTM	S-layer domain-containing protein	78164.55	359.08	40.00	13.00	12.73	1.73	1.10	0.99	3	15669.00	30.14	1.73	2.97		
Cthe_1398	H8EIA8_CLOTM	Dockerin type 1 protein	92707.46	217.30	28.50	12.00	10.97	0.04	1.04	1.02	2	12774.50	18.35	0.04	0.00		
Cthe_3077 783.39	N1K0X3_CLOTM	Cellulosome anchoring protein CipA (Fragment)	196680.39	17703.90	244.33	25.00	82.16	1.73	1.91	16.16	3	90471.67	20.69	1.73			
Cthe_2348	O86999_CLOTM	S-layer protein	113387.37	4635.45	154.00	27.33	0.16	1.73	-0.80	0.02	3	173.67	35.91	1.73	0.00		

Apêndice A. 7. Proteínas identificadas na amostra CPP_Celulose.

Locus gene	protein.Accession	protein.Description	protein.MW	ScoreAVG	ProductsAVG		PeptidesAVG		FmolAVG	FmolCV	Log10FmolAVG		ngramAVG	Repeate.rate	
TOP3AVG	CoverageAVG	TOP3CV	NgramVariance												
Cthe_2360A3DHY3_CLOTH	Glycoside hydrolase family 9	105277.528	231.03	33.00	15.67	20.95	0.10	1.32	2.21	3	17599.67	18.03	0.16	0.05	
Cthe_0625C7HDM7_CLOTM	Glycoside hydrolase family 9	80224.7687	17837.33	220.00	31.00	207.45	0.06	2.32	16.64	3	173720.67	52.42	0.09	0.86	
Cthe_0269C7HGK4_CLOTM	Glycoside hydrolase family 8	51657.5153	4985.52	85.67	12.67	149.36	0.03	2.17	7.72	3	125387.67	40.91	0.11	0.07	
Cthe_0412C7HHU0_CLOTM	Glycoside hydrolase family 9	101249.3736	39152.23	329.00	38.67	4.55	1.05	0.66	0.46	3	3982.67	52.89	1.04	0.23	
Cthe_3077D1NHZ0_CLOTM	Cellulosome anchoring protein cohesin region		178367.9272		61133.53	413.00	35.33	179.99	1.73	2.26	32.10	3	160167.33	31.96	
1.73	3091.92														
Cthe_3078D1NHZ1_CLOTM	Cellulosome anchoring protein cohesin region		192072.7446		8034.49	62.00	15.00	73.44	0.03	1.87	14.11	3	61522.00	11.77	
0.08	0.18														
Cthe_0412D1NLS2_CLOTM	Glycoside hydrolase family 9	101339.3908	47941.25	407.33	46.00	398.29	0.04	2.60	40.36	3	333095.00	59.03	0.00	2.77	
Cthe_0413D1NLS3_CLOTM	Glycoside hydrolase family 9	137914.9087	17768.07	232.00	41.00	59.96	0.09	1.78	8.27	3	50178.67	40.52	0.11	0.54	
Cthe_2089D1NQ73_CLOTM	Glycoside hydrolase family 48	84105.0909	35288.97	390.00	44.00	1611.31	0.08	3.21	135.52	3	1345909.33		61.34	0.07	
128.83															
Cthe_2812D1NQQ3_CLOTM	Glycoside hydrolase family 9	69154.7733	776.12	60.33	16.67	15.97	1.73	1.20	1.10	3	14214.33	40.74	1.73	3.66	
Cthe_2872D1NQT1_CLOTM	Glycoside hydrolase family 5	63655.5399	2801.81	60.67	11.00	61.37	0.87	1.79	3.91	3	49633.67	29.39	0.87	11.59	
Cthe_3077E6UU82_CLOTL	Cellulosome anchoring protein cohesin region		143945.1691		61133.14	411.33	34.33	372.47	0.87	2.57	53.62	3	302551.67	34.69	
0.87	2155.97														
Cthe_3077H8EAE8_CLOTM	Type 3a cellulose-binding domain protein (Fragment)		36765.5957		16756.88	112.33	10.00	447.02	0.02	2.65	16.43	3	374562.67	57.17	
0.08	0.14														
Cthe_0578H8EB72_CLOTM	Glycoside hydrolase family 9	82735.7086	2335.72	107.67	22.67	86.66	0.07	1.94	7.17	3	72469.00	44.11	0.08	0.27	
Cthe_0543H8EBA7_CLOTM	Glycoside hydrolase family 9	82431.1558	634.22	28.00	9.00	11.21	0.05	1.05	0.92	3	9429.33	18.49	0.14	0.00	
Cthe_0536H8EBB4_CLOTM	Glycoside hydrolase family 5	64157.2217	601.21	35.67	10.67	49.70	0.03	1.70	3.19	3	41589.67	24.28	0.06	0.01	
Cthe_0912H8EFD3_CLOTM	Glycoside hydrolase family 10	120406.1918	193.16	41.33	19.67	32.52	0.11	1.51	3.92	3	27275.33	19.68	0.15	0.18	
Cthe_0624H8EFH5_CLOTM	Glycoside hydrolase family 9	178543.3171	1890.71	181.67	40.00	90.73	0.01	1.96	16.20	3	76057.00	35.46	0.09	0.02	
Cthe_0745H8EFV0_CLOTM	Glycoside hydrolase family 9	82640.4133	3927.58	135.33	23.33	88.79	0.02	1.95	7.34	3	74340.33	40.04	0.07	0.02	

Cthe_1020H8EH05_CLOTM	Extracellular solute-binding protein family 1	50085.6863		1253.86	51.00	11.00	57.93	0.04	1.76	2.90	3	48604.67	37.40	0.11	0.01
Cthe_1838H8EHK9_CLOTM	Glycoside hydrolase family 10	69689.2061	3158.55	119.67	19.67	121.14	0.05	2.08	8.44	3	101859.33	45.93	0.14	0.21	
Cthe_0043H8EI14_CLOTM	Glycoside hydrolase family 9	82538.423	316.28	19.33	10.33	12.11	0.20	1.08	1.00	3	10160.00	20.13	0.24	0.04	
Cthe_2872H8EI72_CLOTM	Glycoside hydrolase family 5	63641.513	2761.52	57.00	10.00	29.04	1.73	1.46	1.85	3	25844.67	25.85	1.73	10.25	
Cthe_2812Q8VV73_CLOTM	Endoglucanase	69024.5594	762.16	61.00	16.67	41.30	0.90	1.62	2.85	3	33143.67	42.12	0.88	6.64	

Apêndice A.8. Proteínas identificadas na amostra CPP_Bagaço.

Locus gene	protein.Accession	protein.Description	protein.MW	ScoreAVG	ProductsAVG	PeptidesAVG	FmolAVG	FmolCV	Log10FmolAVG	ngramAVG	Repeate.rate			
TOP3AVG	CoverageAVG	TOP3CV	NgramVariance											
Cthe_2089C7HDZ8_CLOTM	Glycoside hydrolase family 48	83702.5726	459.04	40.67	10.33	15.64	1.73	1.19	1.31	3	14310.67	27.84	1.73	5.14	
Cthe_3078C7HJX6_CLOTM	Cellulosome anchoring protein cohesin region (Fragment)	19238.6245		141.99	6.50	4.00	8.06	0.07	0.91	0.16	2	7311.00			
32.30	0.06	0.00													
Cthe_0412D1NLS2_CLOTM	Glycoside hydrolase family 9	101339.3908	910.67	49.67	13.00	34.39	0.23	1.54	3.49	3	30713.33	23.31	0.20	0.63	
Cthe_0413D1NLS3_CLOTM	Glycoside hydrolase family 9	137914.9087	732.93	27.67	7.33	0.49	1.73	-0.31	0.07	3	452.33	10.08	1.73	0.01	
Cthe_2089D1NQ73_CLOTM	Glycoside hydrolase family 48	84105.0909	453.12	41.67	10.67	36.68	0.88	1.56	3.09	3	32472.67	28.45	0.87	7.32	
Cthe_3077E6UU82_CLOTL	Cellulosome anchoring protein cohesin region	143945.1691		351.53	44.33	17.67	30.97	0.26	1.49	4.46	3	27638.33	21.97		
0.24	1.39														
Cthe_2236H8ECD9_CLOTM	Flagellin domain protein (Fragment)	26368.7071		96.63	11.50	4.00	7.85	0.93	0.90	0.21	2	7087.00	14.14	0.93	
0.04															
Cthe_1965H8ECT1_CLOTM	Peroxiredoxin	20918.5684	98.50	6.00	3.50	4.73	0.16	0.67	0.10	2	4239.50	25.67	0.19	0.00	
Cthe_2348H8EDW1_CLOTM	Ig domain protein	113214.1967	228.65	50.00	16.00	44.96	0.27	1.65	5.09	3	40122.33	20.03	0.25	1.92	
Cthe_2518H8EGB1_CLOTM	Ketol-acid reductoisomerase	36393.6598	71.17	6.00	3.00	4.92	0.00	0.69	0.18	2	4358.25	14.05	0.03	0.00	
Cthe_0393H8EGU7_CLOTM	Sugar ABC transporter (Sugar-binding protein)	34182.3493		130.02	6.50	3.50	7.00	0.10	0.84	0.24	2	6195.50	18.69		
0.08	0.00														
Cthe_0374H8EGW6_CLOTM	Glutamate dehydrogenase	49002.9773	70.01	7.50	5.00	4.20	0.28	0.62	0.21	2	3772.25	9.91	0.31	0.00	

Cthe_1020H8EH05_CLOTM	Extracellular solute-binding protein family 1	50085.6863	2891.53	138.67	18.67	170.16	0.05	2.23	8.52	3	152342.67	46.33	0.03	0.17	
Cthe_2237H8EI41_CLOTM	Flagellin domain protein (Fragment)	26204.787	113.43	13.00	3.33	14.21	0.26	1.15	0.37	3	12745.83	16.46	0.27	0.01	
Cthe_2892H8EI52_CLOTM	60 kDa chaperonin	57531.2224	115.29	22.33	10.33	15.27	0.13	1.18	0.88	3	13701.33	23.54	0.15	0.01	
Cthe_0413Q6RSN8_CLOTM	Cellobiohydrolase (Fragment)	68582.3967	731.34	26.00	5.67	0.21	1.73	-0.69	0.01	3	180.00	14.72	1.73	0.00	

Apêndice A.9. Proteínas identificadas na amostra CPP_Palha.

Locus gene	protein.Accession	protein.Description	protein.MW	ScoreAVG	ProductsAVG	PeptidesAVG	FmolAVG	FmolCV	Log10FmolAVG	ngramAVG	Repeate.rate			
Cthe_2089C7HDZ8_CLOTM	Glycoside hydrolase family 48	83702.5726	701.66	49.00	11.33	61.22	0.94	1.79	5.12	3	55924.67	22.99	0.95	23.02	
Cthe_3078C7HJU0_CLOTM	Cellulosome anchoring protein cohesin region (Fragment)	22375.7228	129.13	5.50	2.00	7.49	1.41	0.87	0.17	2	7567.50	13.11	1.41	0.06	
Cthe_3077D1NHZ0_CLOTM	Cellulosome anchoring protein cohesin region	178367.9272	831.77	60.00	15.33	17.65	1.73	1.25	3.15	3	16470.33	15.25	1.73	29.73	
Cthe_3078D1NHZ1_CLOTM	Cellulosome anchoring protein cohesin region	192072.7446	194.41	8.00	5.00	4.09	1.73	0.61	0.78	3	3601.33	3.67	1.73	1.85	
Cthe_0412D1NLS2_CLOTM	Glycoside hydrolase family 9	101339.3908	1131.49	75.00	16.33	42.57	0.11	1.63	4.31	3	40294.67	25.92	0.18	0.23	
Cthe_0413D1NLS3_CLOTM	Glycoside hydrolase family 9	137914.9087	214.03	29.00	15.00	8.53	0.42	0.93	1.18	3	8048.33	18.06	0.43	0.24	
Cthe_2089D1NQ73_CLOTM	Glycoside hydrolase family 48	84105.0909	701.66	49.33	11.33	40.33	1.73	1.61	3.39	3	40763.33	22.80	1.73	34.52	
Cthe_3077E6UU82_CLOTL	Cellulosome anchoring protein cohesin region	143945.1691	831.15	59.67	15.00	36.76	0.87	1.57	5.29	3	34741.33	17.26	0.87	21.01	
Cthe_3078E6UU83_CLOTL	Cellulosome anchoring protein cohesin region	173102.9606	194.37	8.00	4.67	4.59	1.73	0.66	0.79	3	4282.67	4.07	1.73	1.89	
Cthe_0578H8EB72_CLOTM	Glycoside hydrolase family 9	82735.7086	81.49	6.67	3.33	31.65	0.05	1.50	2.62	3	29750.00	3.39	0.05	0.02	
Cthe_0138H8EC96_CLOTM	Phosphoglycerate kinase	42876.5735	51.09	7.67	6.33	3.20	0.16	0.50	0.14	3	2992.67	15.11	0.10	0.00	
Cthe_2236H8ECD9_CLOTM	Flagellin domain protein (Fragment)	26368.7071	138.26	15.67	6.67	5.48	0.70	0.74	0.14	3	5309.33	29.92	0.77	0.01	

Cthe_1965H8ECT1_CLOTM	Peroxiredoxin	20918.5684	129.64	9.67	3.00	9.11	0.06	0.96	0.19	3	8603.50	20.86	0.13	0.00	
Cthe_2348H8EDW1_CLOTM	Ig domain protein	113214.1967	215.33	40.33	18.00	37.52	0.19	1.57	4.25	3	35029.33	21.90	0.13	0.67	
Cthe_0624H8EFH5_CLOTM	Glycoside hydrolase family 9	178543.3171	84.38	41.00	15.33	30.06	0.23	1.48	5.37	3	28016.67	13.24	0.18	1.57	
Cthe_0745H8EFV0_CLOTM	Glycoside hydrolase family 9	82640.4133	46.58	11.00	6.50	7.75	0.18	0.89	0.64	2	7267.75	5.14	0.08	0.01	
Cthe_1503H8EGN3_CLOTM	Uncharacterized protein	13231.076	198.14	11.00	3.50	12.34	0.01	1.09	0.16	2	12000.00	29.83	0.07	0.00	
Cthe_1504H8EGN4_CLOTM	Linocin M18 bacteriocin protein	30314.6852	191.45	6.33	3.33	9.75	0.15	0.99	0.30	3	9147.33	19.38	0.13	0.00	
Cthe_1020H8EH05_CLOTM	Extracellular solute-binding protein family 1	50085.6863	1282.85	76.33	14.00	127.33	0.18	2.10	6.38	3	119101.00	36.96	0.12	1.30	
Cthe_2237H8EI41_CLOTM	Flagellin domain protein (Fragment)	26204.787	335.81	18.67	3.67	22.21	0.06	1.35	0.58	3	20924.83	16.87	0.10	0.00	

Apêndice B. Lista de GO (Gene Ontology) das proteínas identificadas utilizando o software Blast2GO.

Seq. Name	Gene	Seq. Description	Seq. Length	#Hits	min. eValue	mean Similarity	#GOs	GOs
gi|125712786|gb|ABN51278.1|	Cthe_0037	heat-shock protein hsp20	142	20	8.59E-98	74.20%	2	"P:protein stabilization; P:response to heat"
gi|125712792|gb|ABN51284.1|	Cthe_0043	glycoside hydrolase family 9	742	20	0	81.35%	11	"F:carbohydrate binding; P:cellulose catabolic process; F:hydrolase activity; P:polysaccharide catabolic process; P:carbohydrate metabolic process; F:catalytic activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:metabolic process; F:cellulase activity; F:hydrolase activity, acting on glycosyl bonds; F:cellulose binding"
gi|125712886|gb|ABN51378.1|	Cthe_0137	glyceraldehyde-3-phosphate dehydrogenase	336	20	0	91.45%	6	"P:oxidation-reduction process; P:glucose metabolic process; F:oxidoreductase activity; F:oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor; F:NAD binding; F:NADP binding"
gi|125712887|gb|ABN51379.1|	pgk Cthe_0138		397	1	0	100%	9	"P:small molecule metabolic process; P:carbohydrate metabolic process; P:biological_process; P:generation of precursor metabolites and energy; P:catabolic process; P:cofactor metabolic process; P:cellular nitrogen compound metabolic process; P:biosynthetic process; F:kinase activity"
gi|125712888|gb|ABN51380.1|	tpiA Cthe_0139	triosephosphate isomerase	251	20	0	88.35%	14	"C:cytoplasm; P:metabolic process; P:pentose-phosphate shunt; F:catalytic activity; F:triose-phosphate isomerase activity; P:glycolytic process; F:isomerase activity; P:gluconeogenesis; F:transferase activity; F:phosphoglycerate kinase activity; P:phosphorylation; F:nucleotide binding; F:ATP binding; F:kinase activity"
gi|125712939|gb|ABN51431.1|	Cthe_0190	proteinase inhibitor i4 serpin	600	20	0	68.90%	9	"F:metal ion binding; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:dephosphorylation; P:carbohydrate metabolic process; C:extracellular space; F:acid phosphatase activity; P:polysaccharide catabolic process; P:proteolysis; F:peptidase activity"

gi|125713016|gb|ABN51508.1| celA Cthe_0269 endoglucanase 477 20 0 75.70% 9 "P:cellulose catabolic process; F:hydrolase activity; P:polysaccharide catabolic process; P:carbohydrate metabolic process; F:catalytic activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:metabolic process; F:cellulase activity; F:hydrolase activity, acting on glycosyl bonds"

gi|125713017|gb|ABN51509.1| Cthe_0270glycosyl hydrolase family 18 484 20 0 70.85% 8 "P:chitin catabolic process; F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:hydrolase activity; P:metabolic process; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds; P:polysaccharide catabolic process; F:chitinase activity"

gi|125713021|gb|ABN51513.1| Cthe_0274glycoside hydrolase 563 20 0 78.50% 10 "P:cellulose catabolic process; F:hydrolase activity; P:polysaccharide catabolic process; P:carbohydrate metabolic process; F:catalytic activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:metabolic process; F:cellulase activity; F:hydrolase activity, acting on glycosyl bonds; F:metal ion binding"

gi|125713022|gb|ABN51514.1| Cthe_0275glycosyl transferase 811 20 0 90.15% 8 "F:carbohydrate binding; F:transferase activity; P:carbohydrate metabolic process; F:catalytic activity; F:cellobiose phosphorylase activity; F:transferase activity, transferring glycosyl groups; P:cellulose catabolic process; P:polysaccharide catabolic process"

gi|125713032|gb|ABN51524.1| Cthe_0285isocitrate dehydrogenase 402 20 0 92.95% 10 "P:isocitrate metabolic process; F:NAD binding; F:nucleotide binding; P:tricarboxylic acid cycle; P:oxidation-reduction process; F:oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor; F:oxidoreductase activity; F:isocitrate dehydrogenase (NADP+) activity; F:metal ion binding; F:magnesium ion binding"

gi|125713084|gb|ABN51576.1| Cthe_0338nadh dehydrogenase 165 20 2.61E-113 87.65% 7 "F:metal ion binding; P:oxidation-reduction process; F:oxidoreductase activity; F:2 iron, 2 sulfur cluster binding; F:iron-sulfur cluster binding; F:DNA binding; F:NADH dehydrogenase (ubiquinone) activity"

gi|125713091|gb|ABN51583.1| ldh Cthe_0345 lactate dehydrogenase 318 20 0 81.90% 9 "F:L-lactate dehydrogenase activity; C:cytoplasm; P:oxidation-reduction process; P:carbohydrate metabolic process; P:glycolytic process; F:oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor; F:oxidoreductase activity; F:catalytic activity; P:carboxylic acid metabolic process"

gi|125713095|gb|ABN51587.1| Cthe_0349fructose-bisphosphate aldolase 309 20 0 91.90% 9 "F:fructose-bisphosphate aldolase activity; F:zinc ion binding; F:aldehyde-lyase activity; F:lyase activity; P:carbohydrate metabolic process; P:glycolytic process; F:catalytic activity; P:fructose 1,6-bisphosphate metabolic process; F:metal ion binding"

gi|125713106|gb|ABN51598.1| Cthe_0360thioredoxin 109 20 2.22E-72 83.50% 5 "P:oxidation-reduction process; P:cell redox homeostasis; P:glycerol ether metabolic process; F:protein disulfide oxidoreductase activity; C:cell"

gi|125713120|gb|ABN51612.1| Cthe_0374glutamate dehydrogenase 444 20 0 91.10% 4 "P:oxidation-reduction process; P:cellular amino acid metabolic process; F:oxidoreductase activity; F:oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor"

gi|125713134|gb|ABN51626.1| Cthe_0388alcohol dehydrogenase 344 20 0 81.25% 4 "F:metal ion binding; P:oxidation-reduction process; F:oxidoreductase activity; F:zinc ion binding"

gi|125713139|gb|ABN51631.1| Cthe_0393 321

gi|125713140|gb|ABN51632.1| Cthe_0394alcohol dehydrogenase 389 20 0 82.35% 4 "F:metal ion binding; P:oxidation-reduction process; F:oxidoreductase activity; F:alcohol dehydrogenase (NAD) activity"

gi|125713141|gb|ABN51633.1| Cthe_0395ribose pyranase 151 20 8.24E-105 71.25% 7 "P:monosaccharide metabolic process; F:monosaccharide binding; F:isomerase activity; P:D-ribose catabolic process; C:cytoplasm; F:intramolecular lyase activity; P:carbohydrate metabolic process"

gi|125713151|gb|ABN51643.1| Cthe_0405glycoside hydrolase family 5 526 20 0 77.75% 14 "F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:cellulase activity; F:hydrolase activity; P:metabolic process; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds; P:polysaccharide catabolic process; F:carbohydrate binding; P:cellulose catabolic process; F:catalytic activity; F:cellulose 1,4-beta-cellobiosidase activity; F:cellulose binding; F:calcium ion binding; C:cellulosome"

gi|125713158|gb|ABN51650.1| celK Cthe_0412 CtheDRAFT_2165 895 1 0 100% 3 "P:carbohydrate metabolic process; P:catabolic process; F:hydrolase activity, acting on glycosyl bonds"

gi|125713159|gb|ABN51651.1| Cthe_0413glycoside hydrolase family 9 1224 20 0 91.15% 15 "F:carbohydrate binding; F:hydrolase activity; P:polysaccharide catabolic process; P:carbohydrate metabolic process; F:catalytic activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:metal ion binding; P:metabolic process; F:cellulase activity; F:hydrolase activity, acting on glycosyl bonds; F:cellulose binding; C:extracellular region; F:cellulose 1,4-beta-cellobiosidase activity; C:extracellular space; P:cellulose catabolic process"

gi|125713164|gb|ABN51656.1| pnp Cthe_0418 polyribonucleotide nucleotidyltransferase 700 20 0 89.25% 12 "F:transferase activity; F:3'-5'-exoribonuclease activity; P:RNA processing; F:nucleic acid binding; C:cytoplasm; F:RNA binding; F:nucleotidyltransferase activity; P:RNA phosphodiester bond hydrolysis, exonucleolytic; P:mRNA catabolic process; F:metal ion binding; F:polyribonucleotide nucleotidyltransferase activity; F:magnesium ion binding"

gi|125713167|gb|ABN51659.1| Cthe_0421dipicolinic acid b subunit 194 20 3.45E-135 83.70% 5 "P:metabolic process; F:catalytic activity; P:oxidation-reduction process; F:oxidoreductase activity; P:sporulation resulting in formation of a cellular spore"

gi|125713169|gb|ABN51661.1| Cthe_0423iron-containing alcohol dehydrogenase 873 20 0 92.85% 9 "F:acetaldehyde dehydrogenase (acetylating) activity; P:carbon utilization; P:oxidation-reduction process; F:oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor; F:oxidoreductase activity; P:alcohol metabolic process; F:alcohol dehydrogenase (NAD) activity; P:metabolic process; F:metal ion binding"

gi|125713170|gb|ABN51662.1| Cthe_0424aminoglycoside phosphotransferase 249 20 0 87.05% 6 "F:ATP binding; F:protein kinase activity; F:transferase activity; P:protein phosphorylation; F:transferase activity, transferring phosphorus-containing groups; P:metabolic process"

gi|125713179|gb|ABN51671.1| Cthe_0433glycoside hydrolase 789 20 0 82.40% 10 "F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:cellulose binding; F:carbohydrate binding; F:hydrolase activity; P:carbohydrate metabolic process; F:catalytic activity; P:polysaccharide catabolic process; P:metabolic process; F:cellulase activity; F:hydrolase activity, acting on glycosyl bonds"

gi|125713280|gb|ABN51772.1| celB Cthe_0536 glycoside hydrolase family 5 563 20 0 80.60% 9 "F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:cellulase activity; P:cellulose catabolic process; F:hydrolase activity; P:metabolic process; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds; P:polysaccharide catabolic process; C:cellulosome"

gi|125713287|gb|ABN51779.1| celF Cthe_0543 glycoside hydrolase family 9 739 20 0 75.65% 11 "F:carbohydrate binding; P:cellulose catabolic process; F:hydrolase activity; P:polysaccharide catabolic process; P:carbohydrate metabolic process; F:catalytic activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:metabolic process; F:cellulase activity; F:hydrolase activity, acting on glycosyl bonds; F:cellulose binding"

gi|125713322|gb|ABN51814.1| Cthe_0578glycoside hydrolase family 9 736 20 0 82.30% 12 "F:carbohydrate binding; F:hydrolase activity; P:polysaccharide catabolic process; P:carbohydrate metabolic process; F:catalytic activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:metabolic process; F:hydrolase activity, acting on glycosyl bonds; F:cellulose binding; P:cellulose catabolic process; F:cellulase activity; F:metal ion binding"

gi|125713367|gb|ABN51859.1| Cthe_0624glycoside hydrolase 1601 20 0 79.10% 8 "F:metal ion binding; F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:cellulase activity; F:hydrolase activity; P:carbohydrate metabolic process; F:catalytic activity; P:polysaccharide catabolic process; C:cell wall"

gi|125713368|gb|ABN51860.1| Cthe_0625glycoside hydrolase family 9 710 20 0 71.70% 11 "F:carbohydrate binding; F:hydrolase activity; P:polysaccharide catabolic process; P:carbohydrate metabolic process; F:catalytic activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:metabolic process; F:hydrolase activity, acting on glycosyl bonds; F:cellulose binding; P:cellulose catabolic process; F:cellulase activity"

gi|125713380|gb|ABN51872.1| Cthe_0637virulence factor esxa	100	20	2.71E-64	65.75%	0	-
gi|125713383|gb|ABN51875.1| Cthe_0640dockerin type 1	582	20	0	56.50%	4	"F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:carbohydrate metabolic process; P:polysaccharide catabolic process; F:hydrolase activity"
gi|125713404|gb|ABN51896.1| Cthe_0661ricin b lectin	571	20	0	77.25%	7	"F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:carbohydrate binding; F:hydrolase activity; P:metabolic process; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds; P:polysaccharide catabolic process"
gi|125713442|gb|ABN51934.1| Cthe_0699methylmalonyl- carboxyltransferase	516	20	0	88.50%	4	"F:ligase activity; F:transferase activity; P:metabolic process; F:propionyl-CoA carboxylase activity"
gi|125713444|gb|ABN51936.1| Cthe_0701oxaloacetate decarboxylase	465	20	0	91.75%	9	"P:metabolic process; F:catalytic activity; F:DNA binding; F:transferase activity; F:ligase activity; P:sodium ion transport; F:pyruvate carboxylase activity; F:oxaloacetate decarboxylase activity; F:lyase activity"
gi|125713479|gb|ABN51971.1| Cthe_0736cellulosome anchoring protein cohesin region	1305	20	0	66.35%	8	"F:carbohydrate binding; P:polysaccharide catabolic process; C:cell wall; C:S-layer; C:extracellular region; C:extrachromosomal circular DNA; F:cellulose binding; P:carbohydrate metabolic process"
gi|125713488|gb|ABN51980.1| Cthe_0745glycoside hydrolase family 9	730	20	0	77.05%	11	"F:carbohydrate binding; F:hydrolase activity; P:polysaccharide catabolic process; P:carbohydrate metabolic process; F:catalytic activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:metabolic process; F:hydrolase activity, acting on glycosyl bonds; F:cellulose binding; P:cellulose catabolic process; F:cellulase activity"
gi|125713541|gb|ABN52033.1| Cthe_0798	528					
gi|125713564|gb|ABN52056.1| Cthe_0821coagulation factor 5 8 type domain protein	558	20	0	75.40%	10	"F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:hydrolase activity; P:metabolic process; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds; P:polysaccharide catabolic process; F:carbohydrate binding; P:cellulose catabolic process; F:cellulase activity; F:cellulose binding"
gi|125713568|gb|ABN52060.1| celD Cthe_0825 endoglucanase	649	20	0	82.50%	9	"P:cellulose catabolic process; F:hydrolase activity; P:polysaccharide catabolic process; P:carbohydrate metabolic process; F:catalytic activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:metabolic process; F:cellulase activity; F:hydrolase activity, acting on glycosyl bonds"
gi|125713581|gb|ABN52073.1| Cthe_0838stage iii sporulation protein ah	215	20	1.16E-82	66.80%	0	-
gi|125713654|gb|ABN52146.1| Cthe_0912glycoside hydrolase	1077	20	0	73.75%	10	"F:endo-1,4-beta-xylanase activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:hydrolase activity; P:metabolic process; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds; P:polysaccharide catabolic process; P:xylan catabolic process; C:cellulosome; F:carbohydrate binding"
gi|125713675|gb|ABN52167.1| acpP Cthe_0933 acyl carrier protein	74	20	1.28E-41	94.70%	6	"P:lipid metabolic process; C:cytoplasm; F:ACP phosphopantetheine attachment site binding involved in fatty acid biosynthetic process; P:fatty acid biosynthetic process; P:fatty acid metabolic process; F:phosphopantetheine binding"
gi|125713760|gb|ABN52252.1| Cthe_1020abc transporter substrate-binding protein	459	20	0	72.35%	0	-
gi|125713768|gb|ABN52260.1| ackA ack Cthe_1028 acetate kinase	399	20	0	86.70%	14	"F:transferase activity; P:organic acid metabolic process; P:phosphorylation; F:acetate kinase activity; F:nucleotide binding; C:cytoplasm; F:ATP binding; F:kinase activity; C:intracellular; F:phosphotransferase activity, carboxyl group as acceptor; F:metal ion binding; P:metabolic process; F:magnesium ion binding; P:acetyl-CoA biosynthetic process"
gi|125713822|gb|ABN52314.1| Cthe_1082transcription initiation factor iie	85	4	5.63E-43	85.00%	2	"P:translational initiation; F:translation initiation factor activity"

gi|125713844|gb|ABN52336.1| Cthe_1104	prepilin cleavage protein	176	20	4.20E-110	66.15%	3	"C:type II protein secretion system complex; P:protein secretion by the type II secretion system; F:protein transporter activity"
gi|125713932|gb|ABN52424.1| Cthe_1192	proteinase inhibitor	343	20	0	68.40%	0	-
gi|125713968|gb|ABN52460.1| thrS Cthe_1228	threonyl-trna synthetase	635	20	0	89.15%	13	"F:ligase activity; F:threonine-tRNA ligase activity; P:tRNA aminoacylation for protein translation; F:nucleotide binding; C:cytoplasm; P:translation; F:ATP binding; P:threonyl-tRNA aminoacylation; F:aminoacyl-tRNA ligase activity; F:metal ion binding; F:ligase activity, forming aminoacyl-tRNA and related compounds; P:tRNA aminoacylation; F:hydrolase activity"
gi|125714011|gb|ABN52503.1| Cthe_1271	sugar-binding protein	679	20	0	76.25%	8	"F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:carbohydrate binding; F:hydrolase activity; P:metabolic process; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds; P:polysaccharide catabolic process; F:alpha-L-arabinofuranosidase activity"
gi|125714047|gb|ABN52539.1| Cthe_1307	cellulosome anchoring protein cohesin region	631	20	0	74.75%	2	"F:carbohydrate binding; P:polysaccharide catabolic process"
gi|125714048|gb|ABN52540.1| Cthe_1308	phosphate dikinase	883	20	0	91.25%	9	"F:transferase activity; F:pyruvate, phosphate dikinase activity; P:phosphorylation; F:ATP binding; F:kinase activity; F:catalytic activity; P:pyruvate metabolic process; F:metal ion binding; F:transferase activity, transferring phosphorus-containing groups"
gi|125714062|gb|ABN52554.1| dnaK Cthe_1322	molecular chaperone	608	20	0	89.35%	5	"F:unfolded protein binding; F:ATP binding; P:protein folding; F:nucleotide binding; F:hydrolase activity"
gi|125714108|gb|ABN52600.1| Cthe_1368	s-layer domain protein	710	20	0	73.60%	8	"P:xylan catabolic process; F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:carbohydrate binding; F:hydrolase activity; P:metabolic process; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds; P:carbohydrate catabolic process"
gi|125714113|gb|ABN52605.1| Cthe_1373	yd repeat protein	1917	20	0	92.65%	1	P:self proteolysis
gi|125714137|gb|ABN52629.1| xghA Cthe_1398	xyloglucanase	842	20	0	79.80%	12	"F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:cellulose catabolic process; F:hydrolase activity; P:metabolic process; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds; P:polysaccharide catabolic process; C:cellulosome; F:xyloglucan-specific endo-beta-1,4-glucanase activity; P:xyloglucan catabolic process; F:cellulose binding; F:carbohydrate binding"
gi|125714139|gb|ABN52631.1| Cthe_1400	arabinogalactan endo- -beta-galactosidase	415	20	0	80.90%	8	"F:glucosidase activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:arabinogalactan endo-1,4-beta-galactosidase activity; F:hydrolase activity; P:metabolic process; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds; P:polysaccharide catabolic process"
gi|125714239|gb|ABN52731.1| Cthe_1503	ubiquinone biosynthesis protein coq7	119	20	1.47E-68	83.40%	2	"P:oxidation-reduction process; P:ubiquinone biosynthetic process"
gi|125714240|gb|ABN52732.1| Cthe_1504	bacteriocin	270	20	0	72.75%	5	"P:proteolysis; F:peptidase activity; P:defense response to bacterium; P:oxidation-reduction process; F:peroxidase activity"
gi|125714483|gb|ABN52975.1| Cthe_1754	---NA---	318	0		0	-	
gi|125714507|gb|ABN52999.1| Cthe_1778	copper amine oxidase	319	20	0	65.60%	2	"F:hydrolase activity; P:metabolic process"
gi|125714524|gb|ABN53016.1| Cthe_1795	3-deoxy-7-phosphoheptulonate synthase	341	20	0	90.00%	6	"P:biosynthetic process; F:transferase activity; F:3-deoxy-7-phosphoheptulonate synthase activity; F:catalytic activity; F:aldehyde-lyase activity; P:aromatic amino acid family biosynthetic process"

gi|125714567|gb|ABN53059.1| Cthe_1838	glycoside hydrolase family 10	619	20	0	74.70%	12	"F:xylan endo-1,3-beta-xylosidase activity; F:hydrolase activity; P:polysaccharide catabolic process; P:xylan catabolic process; P:carbohydrate metabolic process; F:endo-1,4-beta-xylanase activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:metabolic process; F:hydrolase activity, acting on glycosyl bonds; F:carbohydrate binding; F:cellulose binding; P:carbohydrate catabolic process"
gi|125714569|gb|ABN53061.1| Cthe_1840	cysteine synthase	311	20	0	94.05%	5	"F:transferase activity; P:cysteine biosynthetic process; P:cellular amino acid biosynthetic process; F:cysteine synthase activity; P:cysteine biosynthetic process from serine"
gi|125714591|gb|ABN53083.1| Cthe_1862	sugar abc transporter atp-binding protein	370	20	0	92.50%	10	"F:transporter activity; P:transport; F:nucleotide binding; F:ATP binding; C:ATP-binding cassette (ABC) transporter complex; F:ATPase activity; P:transmembrane transport; P:metabolic process; F:hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances; P:carbohydrate transport"
gi|125714619|gb|ABN53111.1| Cthe_1890	dockerin type 1	710	20	0	70.90%	10	"F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:carbohydrate metabolic process; P:polysaccharide catabolic process; P:xylan catabolic process; F:hydrolase activity; P:metabolic process; F:hydrolase activity, acting on glycosyl bonds; F:chitin binding; P:chitin metabolic process; C:extracellular region"
gi|125714689|gb|ABN53181.1| xynZ Cthe_1963	carbohydrate-binding protein	837	20	0	82.10%	10	"F:carbohydrate binding; F:xylan endo-1,3-beta-xylosidase activity; F:hydrolase activity; P:polysaccharide catabolic process; P:xylan catabolic process; P:carbohydrate metabolic process; F:endo-1,4-beta-xylanase activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:metabolic process; F:hydrolase activity, acting on glycosyl bonds"
gi|125714691|gb|ABN53183.1| Cthe_1965	187	1	0	100%	3	"P:biological_process; F:oxidoreductase activity; P:response to stress"
gi|125714696|gb|ABN53188.1| Cthe_1970	hypothetical protein	101	10	7.79E-66	70.20%	0	-
gi|125714732|gb|ABN53224.1| Cthe_2010	hypothetical protein	412	20	0	91.15%	0	-
gi|125714755|gb|ABN53247.1| Cthe_2038	carbohydrate binding family 6	807	20	0	73.00%	16	"F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:carbohydrate metabolic process; P:polysaccharide catabolic process; P:sphingolipid metabolic process; F:glucosylceramidase activity; F:alpha-L-arabinofuranosidase activity; P:L-arabinose metabolic process; F:carbohydrate binding; F:xylan endo-1,3-beta-xylosidase activity; F:hydrolase activity; P:xylan catabolic process; F:endo-1,4-beta-xylanase activity; P:metabolic process; F:hydrolase activity, acting on glycosyl bonds; F:lyase activity; C:extracellular region"
gi|125714804|gb|ABN53296.1| celS Cthe_2089	741	1	0	100%	3	"P:carbohydrate metabolic process; P:catabolic process; F:hydrolase activity, acting on glycosyl bonds"
gi|125714805|gb|ABN53297.1| Cthe_2090	zinc-ribbon domain-containing protein	192	20	9.12E-122	67.65%	1	F:zinc ion binding
gi|125714878|gb|ABN53370.1| Cthe_2168	peptidase m4	731	20	0	62.45%	3	"P:proteolysis; F:metallopeptidase activity; F:peptidase activity"
gi|125714889|gb|ABN53381.1| Cthe_2179	pectate lyase	922	20	0	80.45%	7	"F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:carbohydrate binding; P:carbohydrate metabolic process; P:polysaccharide catabolic process; F:lyase activity; C:extracellular region; P:metabolic process"
gi|125714903|gb|ABN53395.1| Cthe_2193	carbohydrate-binding protein	948	20	0	74.70%	16	"F:metal ion binding; F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:carbohydrate binding; F:hydrolase activity; P:metabolic process; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds; P:polysaccharide catabolic process; P:oxidation-reduction process; F:quinone binding; F:calcium ion binding; C:membrane; F:catalytic activity; F:oxidoreductase activity, acting on the CH-OH group of donors, quinone or similar compound as acceptor; C:periplasmic space; F:lyase activity"

gi|125714913|gb|ABN53405.1| folE Cthe_2203	gtp cyclohydrolase i	187	20	2.65E-130	92.00%	10	"P:tetrahydrofolate biosynthetic process; F:hydrolase activity; F:nucleotide binding; C:cytoplasm; F:GTP binding; P:one-carbon metabolic process; F:zinc ion binding; P:7,8-dihydroneopterin 3'-triphosphate biosynthetic process; F:GTP cyclohydrolase I activity; F:metal ion binding"		
gi|125714946|gb|ABN53438.1| Cthe_2236	flagellin	272	20	0	88.75%	4	"F:structural molecule activity; P:bacterial-type flagellum-dependent cell motility; C:bacterial-type flagellum filament; C:bacterial-type flagellum"		
gi|125714947|gb|ABN53439.1| Cthe_2237	flagellin	273	20	0	89.15%	4	"F:structural molecule activity; P:bacterial-type flagellum-dependent cell motility; C:bacterial-type flagellum filament; C:bacterial-type flagellum"		
gi|125714948|gb|ABN53440.1| Cthe_2238	aldehyde dehydrogenase	472	20	0	76.75%	6	"P:cellular aldehyde metabolic process; P:oxidation-reduction process; F:oxidoreductase activity; F:oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor; P:metabolic process; F:aldehyde dehydrogenase [NAD(P)+] activity"		
gi|125714977|gb|ABN53469.1| atpA Cthe_2267	v-type atp synthase subunit a	589	20	0	90.10%	14	"P:ion transport; F:hydrolase activity; P:transport; F:proton-transporting ATP synthase activity, rotational mechanism; P:ATP metabolic process; F:nucleotide binding; P:plasma membrane ATP synthesis coupled proton transport; F:ATP binding; P:ATP biosynthetic process; P:proton transport; C:proton-transporting two-sector ATPase complex, catalytic domain; F:hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances; P:ATP hydrolysis coupled proton transport; P:ATP synthesis coupled proton transport"		
gi|125715011|gb|ABN53503.1| Cthe_2301	crispr-associated protein	294	20	0	72.40%	0	-		
gi|125715058|gb|ABN53550.1| Cthe_2348	1036								
gi|125715070|gb|ABN53562.1| Cthe_2360	glycoside hydrolase family 9	928	20	0	80.30%	11	"F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:cellulose binding; F:carbohydrate binding; F:hydrolase activity; P:carbohydrate metabolic process; F:catalytic activity; P:polysaccharide catabolic process; P:metabolic process; F:cellulase activity; F:hydrolase activity, acting on glycosyl bonds; F:metal ion binding"		
gi|125715089|gb|ABN53581.1| Cthe_2379	conserved protein	169	20	1.31E-102	78.45%	0	-		
gi|125715093|gb|ABN53585.1| Cthe_2383	copper amine oxidase-like domain-containing protein	304	20	0	66.65%	0	-		
gi|125715100|gb|ABN53592.1| Cthe_2390	pyruvate synthase	192	20	4.50E-124	91.85%	6	"P:oxidation-reduction process; F:oxidoreductase activity; P:acetyl-CoA biosynthetic process from pyruvate; F:oxidoreductase activity, acting on the aldehyde or oxo group of donors; F:pyruvate synthase activity; F:oxidoreductase activity, acting on the aldehyde or oxo group of donors, iron-sulfur protein as acceptor"		
gi|125715102|gb|ABN53594.1| Cthe_2392	pyruvate flavodoxin ferredoxin oxidoreductase domain protein	394	20	0	88.65%	6	"P:oxidation-reduction process; F:oxidoreductase activity; P:metabolic process; F:catalytic activity; P:acetyl-CoA biosynthetic process from pyruvate; F:pyruvate synthase activity"		
gi|125715103|gb|ABN53595.1| Cthe_2393	thiamine pyrophosphate tpp-binding domain-containing protein	311	20	0	90.10%	7	"F:thiamine pyrophosphate binding; P:metabolic process; F:catalytic activity; P:oxidation-reduction process; F:oxidoreductase activity; P:acetyl-CoA biosynthetic process from pyruvate; F:pyruvate synthase activity"		
gi|125715118|gb|ABN53610.1| Cthe_2408	242								
gi|125715132|gb|ABN53624.1| Cthe_2422	hypothetical protein	357	3	0	91.33%	0	-		
gi|125715133|gb|ABN53625.1| Cthe_2423	type 3a cellulose-binding domain protein	998	18	0	58.72%	4	"F:cellulose binding; F:carbohydrate binding; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds"		

gi|125715228|gb|ABN53720.1| ilvC Cthe_2518 331 1 0 100% 7 "P:small molecule metabolic process; P:biological_process; F:oxidoreductase activity; P:cofactor metabolic process; P:cellular amino acid metabolic process; P:biosynthetic process; P:cellular nitrogen compound metabolic process"

gi|125715244|gb|ABN53736.1| cysA Cthe_2534 sulfate abc transporter atp-binding protein 354 20 0 89.05% 14 "P:sulfate transmembrane transport; F:ATPase activity, coupled to transmembrane movement of substances; C:membrane; F:sulfate transmembrane-transporting ATPase activity; F:hydrolase activity; P:transport; P:sulfate transport; F:nucleotide binding; F:ATP binding; C:ATP-binding cassette (ABC) transporter complex; F:ATPase activity; P:transmembrane transport; P:metabolic process; C:plasma membrane"

gi|125715298|gb|ABN53790.1| Cthe_2590glycoside hydrolase 639 20 0 76.50% 8 "F:endo-1,4-beta-xylanase activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:hydrolase activity; P:metabolic process; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds; P:polysaccharide catabolic process; P:xylan catabolic process"

gi|125715316|gb|ABN53808.1| atpD Cthe_2608 atp synthase beta subunit 464 20 0 91.45% 17 "P:ATP synthesis coupled proton transport; C:proton-transporting ATP synthase complex, catalytic core F(1); C:membrane; F:hydrolase activity; P:ion transport; P:transport; F:proton-transporting ATP synthase activity, rotational mechanism; P:ATP metabolic process; F:nucleotide binding; P:plasma membrane ATP synthesis coupled proton transport; F:ATP binding; P:ATP biosynthetic process; P:proton transport; C:proton-transporting two-sector ATPase complex, catalytic domain; F:hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances; C:plasma membrane; P:ATP hydrolysis coupled proton transport"

gi|125715364|gb|ABN53856.1| Cthe_2657histone family protein dna-binding protein 91 20 9.72E-56 90.20% 2 "P:chromosome condensation; F:DNA binding"

gi|125715415|gb|ABN53907.1| Cthe_2708abc transporter 473 20 0 60.45% 0 -

gi|125715430|gb|ABN53922.1| rplL Cthe_2723 50s ribosomal protein l7 l12 129 20 2.14E-57 87.00% 5 "F:structural constituent of ribosome; P:translation; C:ribonucleoprotein complex; C:ribosome; C:intracellular"

gi|125715468|gb|ABN53960.1| Cthe_2761glycoside hydrolase 707 20 0 82.90% 11 "F:carbohydrate binding; F:hydrolase activity; P:polysaccharide catabolic process; P:carbohydrate metabolic process; F:catalytic activity; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:metabolic process; F:hydrolase activity, acting on glycosyl bonds; F:cellulose binding; P:cellulose catabolic process; F:cellulase activity"

gi|125715518|gb|ABN54010.1| Cthe_2811glycoside hydrolase 591 20 0 83.10% 9 "F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:substituted mannan metabolic process; P:carbohydrate metabolic process; P:polysaccharide catabolic process; F:mannan endo-1,4-beta-mannosidase activity; F:hydrolase activity; P:metabolic process; F:hydrolase activity, acting on glycosyl bonds; F:carbohydrate binding"

gi|125715519|gb|ABN54011.1| Cthe_2812glycoside hydrolase 611 20 0 83.85% 11 "F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:cellulase activity; F:hydrolase activity; P:metabolic process; P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds; F:catalytic activity; P:polysaccharide catabolic process; F:metal ion binding; F:cellulose binding; F:carbohydrate binding"

gi|125715541|gb|ABN54033.1| Cthe_2834hedgehog intein hint domain protein 561 20 0 98.25% 3 "P:intein-mediated protein splicing; P:nucleic acid phosphodiester bond hydrolysis; F:endonuclease activity"

gi|125715562|gb|ABN54054.1| Cthe_2855hypothetical protein 412 20 0 91.10% 0 -

gi|125715578|gb|ABN54070.1| celG Cthe_2872 endoglucanase 566 20 0 84.55% 9 "P:cellulose catabolic process; F:hydrolase activity; P:polysaccharide catabolic process; P:carbohydrate metabolic process; C:cellulosome; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:metabolic process; F:cellulase activity; F:hydrolase activity, acting on glycosyl bonds"

gi|125715583|gb|ABN54075.1| Cthe_2877s-layer protein 585 20 0 64.35% 0 -

gi|125715591|gb|ABN54083.1| purC Cthe_2885 phosphoribosylaminoimidazole-succinocarboxamide synthase 294 20 0 88.50% 6 "P:purine nucleotide biosynthetic process; F:ATP binding; F:ligase activity; F:nucleotide binding; F:phosphoribosylaminoimidazolesuccinocarboxamide synthase activity; P:'de novo' IMP biosynthetic process"

gi|125715598|gb|ABN54090.1| groL groEL mopA Cthe_2892		541						
gi|125715630|gb|ABN54122.1| adk Cthe_2924	adenylate kinase	217	20	8.31E-156	82.85%	15	"F:transferase activity; P:nucleotide biosynthetic process; P:nucleotide phosphorylation; P:AMP salvage; P:phosphorylation; F:nucleobase-containing compound kinase activity; F:nucleotide binding; C:cytoplasm; F:zinc ion binding; F:ATP binding; F:kinase activity; F:phosphotransferase activity, phosphate group as acceptor; P:nucleobase-containing compound metabolic process; F:metal ion binding; F:adenylate kinase activity"	
gi|125715677|gb|ABN54169.1| Cthe_2972glycoside hydrolase	683	20	0	96.35%	11	"F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:carbohydrate binding; F:hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds; F:hydrolase activity; P:carbohydrate metabolic process; F:catalytic activity; P:polysaccharide catabolic process; P:xylan catabolic process; F:metal ion binding; P:metabolic process; F:hydrolase activity, acting on glycosyl bonds"			
gi|125715716|gb|ABN54208.1| Cthe_3012carbohydrate-binding protein	630	20	0	82.75%	10	"F:carbohydrate binding; F:hydrolase activity; F:glucosylceramidase activity; P:polysaccharide catabolic process; P:carbohydrate metabolic process; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:metabolic process; F:hydrolase activity, acting on glycosyl bonds; P:sphingolipid metabolic process; F:endo-1,4-beta-xylanase activity"			
gi|125715739|gb|ABN54231.1| Cthe_30353-phosphoglycerate dehydrogenase	391	20	0	81.35%	7	"P:oxidation-reduction process; F:amino acid binding; P:metabolic process; F:oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor; F:NAD binding; F:oxidoreductase activity; F:phosphoglycerate dehydrogenase activity"		
gi|125715756|gb|ABN54248.1| Cthe_3052rhs repeat-associated core domain-containing partial	2942	20	0	70.45%	3	"P:self proteolysis; F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:polysaccharide catabolic process"		
gi|125715781|gb|ABN54273.1| cipA Cthe_3077	1853	1	0	100%	4	"P:carbohydrate metabolic process; P:catabolic process; F:molecular_function; F:hydrolase activity, acting on glycosyl bonds"		
gi|125715782|gb|ABN54274.1| olpB Cthe_3078	2313	1	0	100%	3	"P:carbohydrate metabolic process; P:catabolic process; F:molecular_function"		
gi|125715783|gb|ABN54275.1| Cthe_3079cellulosome anchoring protein cohesin region	688	20	0	73.05%	5	"C:cell wall; C:S-layer; F:carbohydrate binding; P:polysaccharide catabolic process; C:extracellular region"		
gi|125715784|gb|ABN54276.1| ancA Cthe_3080	cellulosome anchoring protein cohesin region	447	20	0	75.55%	7	"C:cell wall; C:S-layer; F:carbohydrate binding; P:cellulose catabolic process; P:carbohydrate metabolic process; P:polysaccharide catabolic process; C:extracellular region"	
gi|125715872|gb|ABN54364.1| Cthe_3169enoyl-acp reductase	252	20	0	79.15%	5	"P:oxidation-reduction process; F:oxidoreductase activity; F:enoyl-[acyl-carrier-protein] reductase (NADH) activity; P:fatty acid biosynthetic process; F:enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) activity"		
gi|125715897|gb|ABN54389.1| Cthe_3194sporulation lipoprotein -like protein	216	10	4.67E-91	67.90%	0	-		
gi|125715905|gb|ABN54397.1| Cthe_3202crispr-associated protein csh2	305	20	0	77.40%	0	-		
gi|125715929|gb|ABN54421.1| Cthe_3226copper amine oxidase-like domain-containing protein	262	20	0	87.00%	0	-		
gi|125715930|gb|ABN54422.1| Cthe_3227copper amine oxidase-like domain-containing protein	267	20	0	87.95%	0	-		
gi|125715931|gb|ABN54423.1| Cthe_3228copper amine oxidase-like domain-containing protein	266	20	0	87.40%	0	-		
gi|125715935|gb|ABN54427.1| Cthe_3232yd repeat protein	1669	20	0	89.70%	1	P:self proteolysis		
gi|316939684|gb|ADU73718.1|Clo1313_0638	91	1	0	100%	1	F:DNA binding		

gi|316940428|gb|ADU74462.1|Clo1313_1399		hypothetical protein	132	3	1.54E-90	98.00%	0	-	
gi|1345745|sp|P48223.1|CH10_CLOTH	Cthe_2891	molecular chaperone	94	20	1.89E-46	94.60%	3	"F:ATP binding; C:cytoplasm; P:protein folding"	
gi|281408942|gb|EFB39200.1|	Cther_2217	hypothetical protein	132	3	1.06E-90	97.67%	0	-	
gi|166222856|sp|A3DJ00.1|EFTU_CLOTH	Cthe_2730	translation elongation factor tu	400	20	0	93.80%	11	"P:translational elongation; F:GTP binding; F:translation elongation factor activity; P:translation; F:GTPase activity; C:cytoplasm; F:nucleotide binding; C:intracellular; F:hydrolase activity; F:transferase activity; F:nucleotidyltransferase activity"	
gi|380769454|gb|EIC03370.1|	Cthe_3078	cellulosome anchoring protein cohesin region	316	20	2.82E-163	87.90%	7	"C:cell wall; C:S-layer; F:carbohydrate binding; P:polysaccharide catabolic process; C:extracellular region; P:cellulose catabolic process; P:carbohydrate metabolic process"	
gi|166232188|sp|A3DBQ5.1|ENO_CLOTH	Cthe_0143	enolase	433	20	0	91.85%	9	"C:cell surface; C:cytoplasm; F:lyase activity; P:glycolytic process; F:phosphopyruvate hydratase activity; C:extracellular region; F:metal ion binding; F:magnesium ion binding; C:phosphopyruvate hydratase complex"	
gi|166216920|sp|A3DBX9.1|G6PI_CLOTH	Cthe_0217	glucose-6-phosphate isomerase	448	20	0	89.90%	5	"F:glucose-6-phosphate isomerase activity; C:cytoplasm; P:glycolytic process; F:isomerase activity; P:gluconeogenesis"	
gi|189037327|sp|A3DJE3.1|PCKG_CLOTH	Cthe_2874	phosphoenolpyruvate carboxykinase	605	20	0	89.00%	13	"F:phosphoenolpyruvate carboxykinase (GTP) activity; F:nucleotide binding; C:cytoplasm; F:phosphoenolpyruvate carboxykinase activity; F:GTP binding; F:lyase activity; F:carboxy-lyase activity; P:gluconeogenesis; F:metal ion binding; F:manganese ion binding; F:purine nucleotide binding; P:phosphorylation; F:kinase activity"	
gi|160185042|sp|A3DBQ1.1|TPIS_CLOTH	Cthe_0139	triosephosphate isomerase	251	20	0	88.35%	14	"C:cytoplasm; P:metabolic process; P:pentose-phosphate shunt; F:catalytic activity; F:triose-phosphate isomerase activity; P:glycolytic process; F:isomerase activity; P:gluconeogenesis; F:transferase activity; F:phosphoglycerate kinase activity; P:phosphorylation; F:nucleotide binding; F:ATP binding; F:kinase activity"	

Apêndice C. Valores de concentração de RNA em ng/µL e valores de RIN (integridade do RNA) para as amostras preparadas de palha (P1,P2,P3), Bagaço (B1,B2,B3) e celulose(C1,C2,C3).

Sample	Qubit Conc (ng/ul)	RIN
P1	256.0	8.4
P2	57	8.2
P3	108	7.3
B1	60.0	8
B2	69.0	8
B3	56.0	7.3
C1	179	7.9
C2	240	8.8

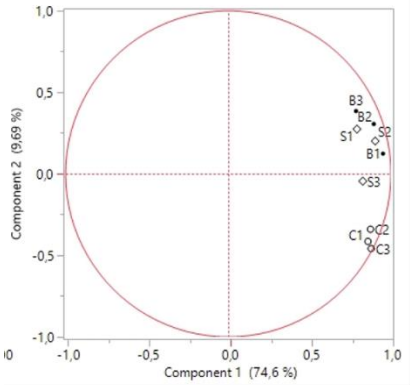
C3	66.2	7.7

Apêndice D. Cobertura dos reads no genoma de *C.thermocellum* e *M.thermoacetica*.

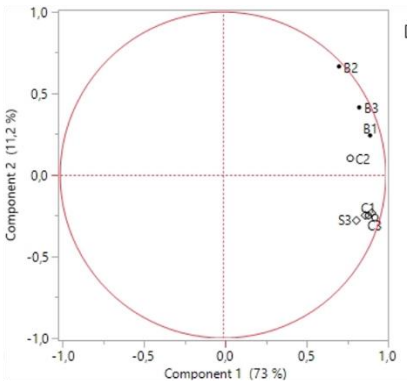
Library	Aligned *Reads*	Coverage *C.thermocellum* (reads x 100bp/ 3843301c)	Aligned *Reads*	Coverage *M.thermoacetica* (reads x 100bp/ 2628784c)
P1	3094521	81	36449661	1387
P2	16894408	440	34086461	1297
P3	6824386	178	18488253	703
B1	10277764	267	30018282	1142
B2	8274725	215	37015422	1408
B3	11636195	303	25555675	972
C1	26608664	692	11105421	422
C2	3298127	86	4561988	174
C3	21395867	557	22371276	851

Apêndice E. PCA (análise do componente principal) do número de fragmentos mapeados, gerados das triplicatas biológicas das amostras de palha (S), bagaço (B) e celulose (C), nos genomas de *C.thermocellum* (A) e *M.thermoacetica* (B).

A


1,0
0,5
0,0
-0,5
-1,0
Component 2 (9,69 %)
B3
B2
S1
S2
B1
S3
C1
C2
C3
-1,0
-0,5
0,0
0,5
1,0
Component 1 (74,6 %)

B


1,0
0,5
0,0
-0,5
-1,0
Component 2 (11,2 %)
B2
B3
B1
C2
S3
C1
C3
-1,0
-0,5
0,0
0,5
1,0
Component 1 (73 %)

Apêndice F.A. Tabela dos genes diferencialmente expressos de *C.thermocellum* em palha em comparação com celulose. Na tabela estão presentes os genes, descrição, número de fragmentos de cada triplicata biológica para bagaço (B1, B2, B3), palha (P1, P2, P3) e celulose (C1, C2, C3), números de expressão calculados e *q-value* para cada comparação de cada gene. FC, indica o número de vezes que o gene foi aumentado ou diminuído. Para valores maiores que 1, os genes são regulados positivamente para a condição palha.

Gene	Descrição	B1	B2	B3	Expressão	P1	P2	P3	Expressão	C1	C2	C3	Expressão	qValue B vs C		qValue P vs C		FC	
Cthe_0042	small GTP-binding protein			740	686	1321	32	57	509	164	11	3260	133	1161	82	5.4E-02	4.8E-10	0.13	
Cthe_0049	XRE family transcriptional regulator			354	138	652	24	40	623	234	27	1440	170	1593	110	1.4E-04	7.6E-03	0.25	
Cthe_0052	hypothetical protein	2321	1699	2554	74	686	2968	660	58	10884	519	3838	267	5.3E-04	8.3E-03	0.22			
Cthe_0053	ribonucleotide-diphosphate reductase subunit alpha					2500	3440	5011	66	333	2221	729	26	19283	1196	7534	265	2.5E-03	1.2E-16
	0.10																		
Cthe_0054	hypothetical protein	851	364	1961	20	166	878	253	10	4084	273	1484	61	1.0E-02	1.3E-05	0.16			
Cthe_0060	hypothetical protein	306	220	585	31	9	376	52	13	1718	56	350	91	5.1E-02	2.6E-08	0.14			
Cthe_0067	NAD-dependent deacetylase			2092	1135	1813	99	822	3302	734	114	6925	1453	6810	537	5.6E-10	5.6E-03	0.21	
Cthe_0068	peptidylprolyl isomerase			1192	690	1072	69	248	1349	360	57	6829	394	5264	440	1.5E-13	4.8E-10	0.13	
Cthe_0069	asparagine synthetase AsnA			431	352	1259	28	22	324	90	7	2161	59	711	60	1.6E-01	6.1E-12	0.12	
Cthe_0070	asparaginyl-tRNA synthetase			820	589	1598	30	142	1006	314	19	6923	366	3430	168	1.2E-09	7.3E-14	0.11	
Cthe_0081	N-acetylglutamate synthase / glutamate N-acetyltransferase					848	603	1413	33	165	1009	196	18	8623	172	2209	191	3.5E-10	
	3.1E-21	0.09																	
Cthe_0084	hypothetical protein	1017	1548	1681	113	29	1388	1018	110	299	69	439	36	1.8E-02	3.1E-02	3.06			
Cthe_0093	septum site-determining protein MinD			1103	1061	1516	66	137	793	139	21	4869	123	1628	174	3.1E-02	1.8E-12	0.12	
Cthe_0094	cell division topological specificity factor MinE					464	396	556	71	57	520	83	37	2001	103	847	228	1.6E-02	1.6E-07
	0.16																		
Cthe_0095	methylglyoxal synthase			525	436	638	58	100	637	222	47	2460	201	1849	255	6.0E-05	3.1E-05	0.18	
Cthe_0100	hypothetical protein	530	511	892	35	32	447	171	16	2288	102	1053	95	4.5E-02	3.5E-06	0.17			
Cthe_0101	iron-containing alcohol dehydrogenase			1366	1081	1768	51	196	1814	564	41	8086	673	6465	291	1.3E-10	1.4E-07	0.14	
Cthe_0102	tRNA (guanine-N(7)-)-methyltransferase			296	157	705	24	15	276	140	14	2532	45	607	101	1.1E-03	3.6E-10	0.14	
Cthe_0103	hypothetical protein	227	77	558	30	25	226	78	16	726	72	1021	101	2.5E-02	3.5E-07	0.16			

Cthe_0110HPr kinase	328	131	1108	23	34	308	84	8	1698	55	630	53	1.5E-01	4.1E-08	0.15		
Cthe_0116hypothetical protein	594	217	635	21	22	457	137	12	2134	149	1662	91	2.4E-04	1.8E-11	0.13		
Cthe_0117HPrNtr domain-containing protein			357	120	207	35	28	290	218	48	1489	290	2446	366	6.9E-28	1.2E-11	0.13
Cthe_0129metal dependent phosphohydrolase			845	758	1762	31	112	918	158	12	2708	117	1424	60	2.2E-01	7.4E-04	0.20
Cthe_0137glyceraldehyde-3-phosphate dehydrogenase				23590	9463	8178	583	4783	19364	5274	521	216355	12028	113322	7434	2.4E-28	4.9E-22
0.07																	
Cthe_0138phosphoglycerate kinase		3138	1881	2849	94	180	3491	807	65	43029	566	6923	881	8.7E-26	4.2E-31	0.07	
Cthe_0139triosephosphate isomerase		4330	1221	2131	144	430	4185	1510	159	30041	1720	19320	1473	2.5E-33	2.9E-10	0.11	
Cthe_0143phosphopyruvate hydratase		6040	2946	3229	134	596	5217	928	84	27986	956	6698	595	1.1E-03	1.3E-05	0.14	
Cthe_0144preprotein translocase subunit SecG		2876	1990	1595	370	1178	2278	1969	513	18633	3529	31923	5029	4.4E-87	6.8E-12	0.10	
Cthe_0151hypothetical protein	2243	1409	1485	270	906	2256	882	286	4270	1105	7769	1187	2.9E-06	2.6E-02	0.24		
Cthe_0154hypothetical protein	232	176	448	42	8	607	104	40	1347	193	1201	226	2.8E-06	4.5E-06	0.18		
Cthe_0156radical SAM family protein		1173	846	2066	31	306	1232	216	15	3862	253	2061	75	7.1E-02	6.7E-04	0.20	
Cthe_016050S ribosomal protein L21		1814	941	902	168	670	1763	987	235	16307	2428	18675	2829	4.1E-139	2.8E-24	0.08	
Cthe_0161hypothetical protein	1964	1016	1001	167	178	1934	586	148	17286	2023	16554	2424	1.9E-98	6.2E-53	0.06		
Cthe_016250S ribosomal protein L27		2805	1052	1371	267	436	3157	2666	567	22210	4627	28894	4941	1.9E-180	7.0E-08	0.11	
Cthe_0174sulfatase	1116	698	1921	28	134	952	332	14	7071	245	3099	117	1.7E-05	4.9E-12	0.12		
Cthe_0178argininosuccinate lyase		558	295	1493	24	61	460	146	9	5643	93	903	101	3.9E-05	4.3E-28	0.09	
Cthe_0179argininosuccinate synthase		351	149	1025	17	18	230	71	4	3937	39	769	80	4.0E-05	1.2E-67	0.05	
Cthe_0186UDP-galactose 4-epimerase		468	273	924	22	61	505	98	10	2231	77	1035	66	2.3E-02	3.7E-08	0.15	
Cthe_0188ribonuclease PH	382	212	922	27	60	595	67	13	1811	65	524	65	1.2E-01	2.2E-04	0.20		
Cthe_0196glutamine synthetase, catalytic region			1345	1018	2141	30	191	978	550	18	4383	220	1812	67	1.1E-01	4.0E-02	0.27
Cthe_0209glycosyltransferase	108	95	384	27	10	487	1077	168	117	35	215	30	1.0E+00	7.7E-08	5.60		
Cthe_0212beta-glucosidase	935	982	1676	36	111	814	160	12	4148	104	945	78	1.2E-01	1.1E-06	0.15		

Cthe_0214phenylalanyl-tRNA synthetase subunit alpha 0.10			675	269	1120	28	24	374	105	9	3246	61	1051	88	1.1E-02	2.7E-21
Cthe_0215phenylalanyl-tRNA synthetase subunit beta	1056	482	2135	21	102	880	332	10	8783	192	2844	103	2.4E-07	9.9E-18	0.10	
Cthe_0217glucose-6-phosphate isomerase	1512	843	1923	45	113	1337	349	24	15058	326	4469	309	1.6E-16	4.4E-34	0.08	
Cthe_0218metallophosphoesterase	447	216	703	24	32	389	80	10	1815	69	868	72	2.9E-02	2.1E-09	0.14	
Cthe_0230hypothetical protein	967	832	1902	31	302	855	192	13	2795	216	1320	58	2.2E-01	3.0E-03	0.22	
Cthe_0231radical SAM family protein	749	425	1383	27	245	815	221	17	1991	289	1167	64	7.7E-02	1.7E-02	0.27	
Cthe_0234AMP-dependent synthetase and ligase	900	341	2249	19	147	734	357	10	6677	310	2587	82	1.0E-05	5.6E-12	0.12	
Cthe_0235YruB family glutaredoxin-like protein	2514	934	656	254	736	2243	868	324	1558	2459	4051	1370	1.2E-09	2.1E-02	0.24	
Cthe_0243copper amine oxidase-like protein	970	535	1163	44	26	895	102	18	1833	48	587	61	5.0E-01	4.1E-02	0.30	
Cthe_0262gamma-glutamyl phosphate reductase	2171	2335	2756	81	777	3558	345	52	9545	399	3725	229	3.6E-02	1.5E-02	0.23	
Cthe_0264hypothetical protein	968	919	1219	105	3621	1526	264	198	379	65	330	45	1.3E-01	1.1E-04	4.40	
Cthe_0269glycoside hydrolase family protein	3497	4818	5217	137	365	3651	191	39	9394	181	1428	162	1.0E+00	2.3E-02	0.24	
Cthe_0271type 3a, cellulose-binding	975	1225	2015	65	132	1114	288	29	4705	186	1719	154	7.4E-02	1.4E-04	0.19	
Cthe_0275cellobiose phosphorylase	2628	2004	3326	46	309	2545	680	26	12077	545	4828	156	1.2E-02	5.1E-05	0.17	
Cthe_0285isocitrate dehydrogenase	3304	2232	3092	102	553	3654	877	73	31764	908	11729	777	4.5E-15	9.0E-17	0.09	
Cthe_0290homoserine dehydrogenase	1108	342	1570	33	163	1151	608	32	5462	535	3371	170	3.0E-08	4.7E-04	0.19	
Cthe_0295phosphoserine aminotransferase	3380	1942	2960	105	263	3044	450	52	18831	742	5672	492	4.7E-05	6.8E-14	0.11	
Cthe_0307hypothetical protein	3124	1145	1305	210	687	2678	550	173	7763	523	9812	1013	6.6E-08	9.2E-05	0.17	
Cthe_0308hypothetical protein	2415	953	1144	189	650	2524	610	194	5816	565	8097	930	9.7E-08	3.9E-03	0.21	
Cthe_0312ATPase AAA-2	2146	2118	3664	47	292	2473	544	23	6870	536	2927	99	2.7E-01	1.8E-02	0.23	
Cthe_0317hypothetical protein	557	122	399	22	53	531	90	16	2083	119	2256	140	7.7E-09	5.3E-16	0.11	
Cthe_0323hypothetical protein	105	60	199	21	6	44	24	6	628	34	221	81	2.4E-04	3.8E-70	0.07	
Cthe_0324valyl-tRNA synthetase	1359	1024	2881	28	142	1017	491	13	16619	365	3260	161	1.6E-11	1.8E-29	0.08	
Cthe_0325NAD synthetase	1047	798	2360	30	211	941	410	16	11443	460	3202	170	3.7E-10	7.5E-21	0.09	

Cthe_0328peptide chain release factor 3	1070	342	1673	27	128	829	177	12	8884	299	2728	160	2.4E-11	9.6E-40	0.08	
Cthe_0338NADH-quinone oxidoreductase subunit E	1258	1007	1076	96	72	810	202	38	6422	297	3295	435	2.3E-06	7.6E-27	0.09	
Cthe_0339histidine kinase	1112	1100	1340	90	77	858	167	32	6043	130	1780	294	2.8E-03	2.3E-16	0.11	
Cthe_0340ferredoxin 428	481	819	67	27	415	108	27	2889	72	648	208	1.5E-02	1.3E-12	0.13		
Cthe_0341NADH dehydrogenase (quinone)	2446	2478	4116	72	264	1930	576	28	18539	468	3869	272	3.5E-03	8.5E-15	0.10	
Cthe_0342hydrogenase, Fe-only	5867	3880	4857	119	759	3855	1223	62	30833	1555	11242	553	4.4E-03	1.9E-09	0.11	
Cthe_0343FMN-binding flavin reductase-like protein	355	207	570	26	50	380	149	19	2129	202	2146	166	4.1E-09	8.2E-16	0.11	
Cthe_0344malate dehydrogenase	7424	5677	5768	231	878	4940	1351	110	66350	2108	22472	1657	1.9E-08	6.1E-32	0.07	
Cthe_0345malate dehydrogenase (NAD)	3749	2900	3088	146	249	2200	880	70	39498	1046	14710	1216	9.3E-18	9.0E-57	0.06	
Cthe_0346hypothetical protein	431	416	546	56	118	425	219	45	1099	152	1571	181	1.7E-02	9.3E-03	0.25	
Cthe_03476-phosphofructokinase	3029	2789	3833	111	382	2895	1309	77	52039	690	8177	1016	2.4E-22	7.2E-26	0.08	
Cthe_0349fructose-bisphosphate aldolase	11480	5492	6754	364	2440	13509	3703	382	130986	4438	60243	4466	8.3E-30	3.0E-11	0.09	
Cthe_0350signal peptidase I	284	92	561	22	6	156	72	8	1603	100	858	98	6.6E-04	2.2E-34	0.08	
Cthe_0362AsnC family transcriptional regulator	725	1002	851	78	14	520	59	18	1836	97	597	118	4.2E-01	6.1E-08	0.15	
Cthe_0363aminotransferase	1636	1991	2149	71	263	1301	255	25	4900	389	1964	144	1.5E-01	3.5E-05	0.17	
Cthe_0364XRE family transcriptional regulator	582	485	972	41	72	580	175	22	819	113	1401	74	3.1E-01	4.6E-02	0.30	
Cthe_0365peptide chain release factor 2	2179	1639	2341	82	195	2514	601	54	16335	667	7512	491	4.5E-13	6.1E-13	0.11	
Cthe_0372sulfide dehydrogenase (flavoprotein) subunit SudA 0.12	532	321	1435	23	35	433	110	7	3168	95	570	59	5.2E-02	1.1E-13		
Cthe_0373ferredoxin-NADP(+) reductase subunit alpha 0.07	280	180	798	21	35	189	52	5	2376	39	509	72	1.0E-02	5.7E-41		
Cthe_0374glutamate dehydrogenase	5432	963	1901	87	409	3531	1208	74	20337	972	10873	522	8.0E-09	7.3E-06	0.14	
Cthe_0375GMP synthase	2077	944	1737	44	361	1870	517	32	20256	663	6316	381	8.1E-29	1.4E-24	0.08	
Cthe_0398hypothetical protein	137	90	312	23	19	283	40	16	742	78	814	114	1.9E-05	3.5E-10	0.14	
Cthe_0401methyl-accepting chemotaxis sensory transducer 4.13	4889	4831	6625	103	227	9220	900	66	670	127	714	16	3.2E-09	1.8E-03		

Cthe_0402copper amine oxidase-like protein		3791	3303	3651	70	349	3171	873	36	12525	599	8067	206	6.7E-02	2.2E-04	0.17
Cthe_0409hypothetical protein	143	137	260	28	19	185	53	17	958	24	316	101	6.2E-03	1.0E-07	0.17	
Cthe_0410hypothetical protein	1566	1242	2484	55	408	2696	659	48	9573	547	3084	219	5.3E-05	9.1E-03	0.22	
Cthe_041730S ribosomal protein S15		3119	1942	1860	377	186	2376	814	249	8549	1768	13088	2142	4.1E-10	2.6E-11	0.12
Cthe_0418polynucleotide phosphorylase/polyadenylase 0.11			1504	1072	2761	36	204	1258	394	16	9947	306	3633	140	7.2E-05	1.0E-12
Cthe_0435dockerin type I cellulosome protein			1524	1166	1721	60	788	2012	310	46	3778	484	3558	174	1.3E-02	4.7E-02	0.26
Cthe_0436hypothetical protein	3237	2302	3696	41	265	2742	741	21	12813	579	5078	127	3.8E-02	7.0E-05	0.17	
Cthe_0437hypothetical protein	815	499	697	62	68	928	367	60	3002	427	3090	336	2.5E-08	2.8E-05	0.18	
Cthe_0442cell division protein FtsQ		827	557	922	38	157	1046	205	27	1861	189	1937	104	3.3E-02	1.7E-02	0.26
Cthe_0453hypothetical protein	813	748	902	80	2145	1288	159	123	323	49	219	33	8.1E-02	1.3E-03	3.73	
Cthe_0454hypothetical protein	590	482	689	62	1408	755	112	86	264	31	181	28	1.5E-01	2.1E-02	3.07	
Cthe_0461tRNA (uracil-5-)-methyltransferase Gid			868	443	1295	28	57	729	306	17	2203	165	1094	60	1.4E-01	4.0E-02	0.28
Cthe_0504poly-gamma-glutamate biosynthesis protein 0.16			537	281	1056	21	26	573	109	9	2210	69	950	55	6.4E-02	1.9E-06
Cthe_0509sodium ion-translocating decarboxylase subunit beta 0.11			1049	925	1786	49	167	988	325	26	9050	244	2483	232	5.5E-07	4.7E-15
Cthe_0549hypothetical protein	547	290	1104	32	184	683	279	26	2092	241	2339	124	4.9E-04	1.0E-03	0.21	
Cthe_0551AMP-dependent synthetase and ligase			916	856	1900	31	63	802	247	12	4583	114	1436	77	5.2E-02	8.7E-07	0.16
Cthe_0552transcriptional regulator		312	308	546	30	25	208	49	9	1054	22	483	58	3.3E-01	1.0E-09	0.16
Cthe_0554phosphoribosylformylglycinamidine synthase 0.14			2743	1527	3864	30	299	1672	330	10	9883	248	2961	73	1.6E-01	1.8E-08
Cthe_0569hypothetical protein	666	260	541	27	76	419	122	15	996	189	666	64	1.5E-01	2.0E-03	0.23	
Cthe_0572ribosomal RNA large subunit methyltransferase N 0.13			528	391	1085	27	48	609	162	14	2958	191	1946	107	2.9E-04	2.8E-11
Cthe_0573protein serine/threonine phosphatases			477	451	1041	38	14	505	85	13	2934	107	1077	120	9.7E-03	4.4E-18	0.11
Cthe_0574serine/threonine protein kinase	1519	905	2790	35	135	1690	473	20	8794	375	3085	126	3.7E-04	3.1E-06	0.16	

Cthe_0575	ribosome small subunit-dependent GTPase A			744	402	917	33	74	806	285	25	3291	300	1450	134	1.3E-04	9.6E-05	0.19
Cthe_0576	ribulose-5-phosphate 3-epimerase		1670	1043	979	80	309	797	179	32	2287	335	1355	151	2.2E-01	1.3E-03	0.21	
Cthe_0577	thiamine pyrophosphokinase	1260	879	953	69	52	644	155	23	2599	292	1081	153	1.2E-01	4.6E-08	0.15		
Cthe_0580	aspartate aminotransferase	676	314	1480	29	40	616	214	14	2144	70	806	54	2.6E-01	1.7E-02	0.26		
Cthe_0584	hypothetical protein	128	75	151	17	12	116	57	13	604	57	766	107	9.0E-13	4.3E-17	0.12		
Cthe_0596	ribosome biogenesis GTP-binding protein YsxC			179	89	612	19	2	182	183	16	929	65	638	58	6.0E-02	1.7E-02	0.28
Cthe_0597	thiamine biosynthesis protein ThiS		70	5	275	24	3	63	43	13	266	31	164	55	7.9E-02	2.9E-05	0.24	
Cthe_0598	thiazole synthase	236	46	734	18	21	331	202	16	1285	110	773	64	1.2E-02	9.2E-03	0.25		
Cthe_0613	thiamine pyrophosphate enzyme-like TPP-binding protein				983	1236	2279	35	129	822	269	12	7795	145	1718	112	5.1E-03	4.7E-15 0.11
Cthe_0614	indolepyruvate oxidoreductase subunit beta			220	210	735	28	9	301	95	13	1639	55	370	78	8.0E-02	4.3E-06	0.17
Cthe_0615	phenylacetate-CoA ligase	675	689	1696	33	130	985	471	25	5349	298	2277	135	6.7E-05	2.4E-04	0.19		
Cthe_0616	ACT domain-containing protein	245	261	652	38	28	352	145	25	1880	56	608	125	1.8E-02	7.9E-05	0.20		
Cthe_0620	DtxR family iron dependent repressor		1545	1345	968	142	100	910	119	44	1247	247	2173	220	3.9E-01	2.4E-04	0.20	
Cthe_0621	translation initiation factor 2B subunit I		929	854	1526	46	129	1133	260	25	5556	163	1666	153	2.3E-03	3.6E-06	0.16	
Cthe_0622	methylthioadenosine phosphorylase		510	364	958	32	45	603	114	15	4285	80	854	136	9.9E-05	7.1E-16	0.11	
Cthe_0626	hypothetical protein	629	314	1121	20	17	375	134	7	3317	76	1262	68	4.5E-03	3.7E-19	0.10		
Cthe_0627	hypothetical protein	650	592	1137	38	86	684	230	22	3251	184	1583	126	4.4E-03	6.5E-06	0.17		
Cthe_0628	hypothetical protein	506	489	1039	34	112	491	169	17	2828	155	1305	113	6.0E-03	7.4E-08	0.15		
Cthe_0629	type II secretion system protein E		931	854	1879	36	142	779	214	14	4401	229	1905	99	2.0E-02	3.8E-09	0.14	
Cthe_0632	hypothetical protein	940	531	820	54	96	957	381	48	4730	268	2893	285	2.3E-08	5.9E-06	0.17		
Cthe_0637	hypothetical protein	8270	4346	3089	745	2137	5528	3227	780	20457	3872	44710	5177	2.0E-07	1.3E-03	0.15		
Cthe_0657	hypothetical protein	607	386	765	41	25	702	496	50	2698	150	3549	221	1.2E-07	3.4E-03	0.23		

Cthe_0672pyrroline-5-carboxylate reductase	665	347	819	31	57	819	150	20	1406	155	1226	78	8.5E-02	1.6E-02	0.26	
Cthe_0681inosine 5-monophosphate dehydrogenase	1447	1133	2313	46	267	1596	470	28	30146	557	6308	520	1.6E-52	1.1E-69	0.05	
Cthe_0683diaminopimelate decarboxylase	1755	522	1366	39	102	774	162	13	7124	174	2030	150	1.4E-04	7.6E-27	0.09	
Cthe_0686tryptophanyl-tRNA synthetase	640	750	1125	36	93	415	103	10	3599	119	1213	106	1.5E-02	6.8E-22	0.09	
Cthe_0689RDD domain-containing protein	273	241	518	25	69	227	131	16	853	105	635	66	1.2E-01	4.1E-03	0.24	
Cthe_0694spermidine synthase	518	258	678	24	48	467	242	20	8732	201	2964	301	1.8E-47	5.1E-54	0.07	
Cthe_0695agmatinase	690	437	1295	39	158	818	334	30	9794	291	2016	300	9.2E-20	1.1E-19	0.10	
Cthe_0696putative rRNA methylase	362	249	673	32	35	343	157	20	5191	137	1381	251	3.6E-16	1.7E-38	0.08	
Cthe_0699carboxyl transferase	1450	2250	2571	58	347	1256	288	20	9866	269	3589	186	2.9E-03	1.2E-14	0.11	
Cthe_0700biotin/lipoyl attachment protein	220	507	572	47	22	271	29	12	2470	37	472	155	1.4E-02	1.4E-50	0.08	
Cthe_0701oxaloacetate decarboxylase	7202	10176	5599	240	855	5193	721	75	19935	1608	9920	519	4.6E-01	1.6E-05	0.14	
Cthe_0703hypothetical protein	598	876	1155	156	20	570	77	42	1235	43	702	171	9.6E-01	4.3E-03	0.25	
Cthe_0706RpiR family transcriptional regulator	664	1328	3014	79	61	973	181	22	2523	54	729	77	9.2E-01	4.6E-02	0.29	
Cthe_0707phosphoglycerate mutase	469	745	1852	69	53	716	245	31	1976	115	1008	108	3.9E-01	4.3E-02	0.29	
Cthe_0710hypothetical protein	1076	1196	1410	122	93	925	105	38	3761	154	1787	281	8.1E-02	4.8E-10	0.14	
Cthe_0711chorismate mutase	540	687	895	86	18	381	33	17	1945	41	388	140	3.7E-01	2.8E-13	0.12	
Cthe_0712cytidylate kinase	955	1039	1398	71	131	837	156	27	2931	94	915	123	3.0E-01	1.4E-03	0.22	
Cthe_07131-acyl-sn-glycerol-3-phosphate acyltransferase			634	602	1134	55	88	581	183	26	1659	116	1020	102	2.7E-01	1.2E-02
0.25																
Cthe_07144-hydroxy-3-methylbut-2-enyl diphosphate reductase/S1 RNA-binding domain-containing protein								3530	3494	5455	86	158	4218	1005	45	18156	975
7825	286	5.0E-02	8.6E-05	0.16												
Cthe_0715S-adenosylmethionine decarboxylase proenzyme			1763	653	852	124	739	1394	780	163	10194	1147	19194	1754	2.6E-90	3.1E-19
0.09																
Cthe_0723tyrosyl-tRNA synthetase	1916	1096	1700	54	128	1271	225	21	10069	323	2202	222	4.0E-05	2.2E-21	0.09	
Cthe_0724HisJ family histidinol phosphate phosphatase			356	132	514	17	2	272	85	8	1471	67	704	60	1.5E-02	4.2E-09
0.13																

Cthe_0726putative aminopeptidase 1	1432	1778	2616	55	159	1358	304	20	3640	174	1374	76	4.8E-01	3.9E-02	0.26	
Cthe_0727cell envelope-related transcriptional attenuator			1365	627	1690	56	129	1473	1080	70	9227	1214	12935	570	2.8E-42	8.9E-10
0.12																
Cthe_0730hypothetical protein	574	262	660	31	66	621	269	29	2386	74	1392	117	1.7E-03	8.1E-03	0.25	
Cthe_0731shikimate kinase	321	206	518	29	22	490	96	20	1526	27	404	79	8.1E-02	5.0E-03	0.25	
Cthe_0735cellulosome anchoring protein cohesin subunit			363	225	893	26	16	276	82	8	2942	90	1015	106	4.2E-04	8.1E-33
0.08																
Cthe_0738copper ion binding protein	1442	454	548	163	75	1471	211	129	1487	1124	2036	769	1.1E-06	4.3E-06	0.17	
Cthe_0741adenylosuccinate lyase	1070	695	1767	35	85	1030	496	24	10961	261	4017	223	5.1E-14	2.4E-15	0.11	
Cthe_0743MATE efflux family protein	757	432	1491	26	222	542	194	12	2288	293	1932	73	2.1E-02	2.8E-06	0.16	
Cthe_0747extracellular solute-binding protein	739	180	828	23	62	490	226	15	4237	214	1951	131	5.5E-09	1.3E-14	0.11	
Cthe_0748binding-protein-dependent transport systems inner membrane component					339	71	516	16	45	207	125	10	1846	107	831	78
9.0E-05	1.3E-10	0.13														
Cthe_0749binding-protein-dependent transport systems inner membrane component					377	60	634	18	52	237	79	8	2013	75	728	73
1.5E-03	8.7E-15	0.11														
Cthe_0750spermidine/putrescine ABC transporter ATPase subunit				475	92	890	19	12	326	104	8	3194	76	647	78	5.2E-04
7.7E-21	0.10															
Cthe_0751XRE family transcriptional regulator	295	90	415	20	98	224	106	16	2313	117	1478	156	7.0E-13	1.2E-20	0.10	
Cthe_0755aspartate aminotransferase	2829	770	1254	57	392	1290	775	45	18421	888	10623	543	2.6E-35	1.1E-24	0.08	
Cthe_0764signal peptidase I	482	336	665	31	41	403	86	13	1040	68	666	58	3.3E-01	1.7E-03	0.22	
Cthe_076550S ribosomal protein L19	2894	1161	1399	225	455	4209	3457	594	30376	7507	58220	6752	0.0E+00	1.8E-11	0.09	
Cthe_0766tRNA (Guanine37-N1) methyltransferase	1058	417	940	46	136	787	309	33	3107	676	2299	219	8.5E-07	5.3E-08	0.15	
Cthe_076716S rRNA-processing protein RimM	221	91	602	24	2	163	86	10	1784	63	470	95	3.4E-03	1.6E-17	0.11	
Cthe_0768nucleic acid binding protein	752	275	833	114	20	996	289	106	8018	754	4526	1346	3.1E-50	1.2E-35	0.08	
Cthe_076930S ribosomal protein S16	446	153	445	60	8	480	126	45	5149	254	1898	656	1.0E-32	2.0E-53	0.07	
Cthe_0770signal recognition particle subunit FFH/SRP54 (srp54)			973	511	2249	39	93	977	324	20	3326	176	1780	86	1.1E-01	4.4E-03
0.23																

Cthe_0771	putative helix-turn-helix protein, YlxM/p13-like protein				420	227	816	58	13	303	72	19	1290	91	801	137	1.4E-01	2.7E-09	0.14
Cthe_0773	SOS-response transcriptional repressor LexA				775	1244	1511	79	72	818	188	29	2649	74	850	118	4.2E-01	8.2E-03	0.25
Cthe_0778	hypothetical protein	1003	878	895	102	21	734	35	26	1731	30	248	107	1.0E+00	4.2E-03	0.24			
Cthe_0786	3-dehydroquinate synthase	725	648	1288	35	67	549	272	17	7331	423	3862	236	3.6E-15	1.6E-40	0.07			
Cthe_0787	isoleucyl-tRNA synthetase	1070	580	2655	21	109	811	317	8	21659	337	3715	192	4.2E-30	3.2E-124	0.04			
Cthe_0790	hypothetical protein	650	711	540	97	60	479	67	33	1213	45	534	134	5.2E-01	4.0E-03	0.25			
Cthe_0794	aluminium resistance protein	757	392	1234	26	111	1144	259	20	3219	186	1606	85	4.0E-03	5.4E-03	0.24			
Cthe_0828	1-deoxy-D-xylulose-5-phosphate synthase	1109	1232	2685	38	215	1130	286	15	6009	234	2115	96	4.4E-02	9.0E-07	0.16			
Cthe_0829	hypothetical protein	735	627	1152	49	110	557	118	18	3886	252	2337	197	1.1E-04	2.1E-26	0.09			
Cthe_0836	hypothetical protein	1474	1147	1324	103	106	1111	194	42	2301	219	2098	189	2.5E-01	1.7E-03	0.22			
Cthe_0846	hypothetical protein	1131	371	739	77	175	1827	402	101	6361	536	6580	720	4.0E-31	3.5E-08	0.14			
Cthe_0847	elongation factor P	3661	1446	2160	185	333	3893	862	157	15399	699	12359	1075	9.5E-10	2.5E-06	0.15			
Cthe_0852	type 4 prepilin peptidase 1	846	723	1093	45	204	795	125	21	1830	100	1118	78	3.1E-01	2.1E-02	0.27			
Cthe_0856	branched chain amino acid aminotransferase				1675	1844	2128	76	204	1731	752	52	19470	586	7973	553	1.6E-20	5.7E-19	0.09
Cthe_0857	hypothetical protein	576	571	764	56	110	446	100	22	1517	65	958	109	2.4E-01	2.3E-04	0.20			
Cthe_0858	hypothetical protein	764	1108	1953	55	114	837	204	20	3490	139	948	101	2.5E-01	3.4E-04	0.20			
Cthe_0859	hypothetical protein	563	805	1119	65	43	631	239	33	2152	79	1019	126	2.2E-01	1.8E-02	0.26			
Cthe_0860	rubrerythrin	656	201	612	35	20	698	109	22	5692	244	3482	337	5.4E-28	2.6E-59	0.07			
Cthe_0864	2-oxoglutarate ferredoxin oxidoreductase subunit gamma				759	511	894	57	257	849	437	60	8466	441	4249	533	8.1E-31	3.1E-14	0.11
Cthe_0865	3-methyl-2-oxobutanoate dehydrogenase (ferredoxin)				579	537	1069	41	94	613	189	22	8706	159	1536	293	8.1E-16	4.1E-39	0.08
Cthe_0866	pyruvate flavodoxin/ferredoxin oxidoreductase-like protein				505	551	1511	35	57	704	223	18	6432	121	1287	158	4.2E-06	2.3E-15	0.11

Cthe_08672-oxoglutarate ferredoxin oxidoreductase subunit delta					126	83	263	32	5	106	44	15	814	34	262	117	2.5E-03
4.8E-17	0.13																
Cthe_0869hypothetical protein	605	407	823	39	66	521	97	16	1990	83	711	90	1.2E-01	1.8E-05	0.18		
Cthe_0870NADPH-dependent FMN reductase			763	330	989	50	113	1325	256	48	4217	203	2654	258	1.0E-07	7.8E-05	0.19
Cthe_0877GTP-dependent nucleic acid-binding protein EngD					902	1401	2270	60	169	1184	455	32	10655	221	1973	248	2.7E-05	1.2E-09
0.13																	
Cthe_08803-deoxy-D-arabinoheptulosonate-7-phosphate synthase					1224	1133	1690	57	290	684	288	24	3531	162	1691	114	1.8E-01
9.0E-04	0.21																
Cthe_0899hypothetical protein	104	36	276	21	8	184	386	69	98	14	124	20	1.0E+00	3.6E-04	3.45		
Cthe_0903protein translocase subunit secF	1994		1200	1472	69	382	1674	742	60	13966	506	8115	498	2.8E-18	1.0E-10	0.12	
Cthe_0904protein-export membrane protein SecD			1429	973	1934	46	235	1209	354	24	13744	223	2481	257	2.9E-10	7.1E-21	0.09
Cthe_0907hypothetical protein	454	187	263	91	76	354	239	110	3554	463	5177	1468	1.3E-77	3.3E-41	0.07		
Cthe_0916hypothetical protein	1028	594	965	41	122	884	252	26	5013	244	2963	200	5.1E-07	3.5E-11	0.13		
Cthe_0917glutaminyl-tRNA synthetase	1365		1051	2322	38	264	1804	418	24	8383	276	2274	135	9.6E-04	1.8E-04	0.18	
Cthe_0922diaminopimelate dehydrogenase			514	196	930	23	71	584	164	14	4506	159	1571	133	4.7E-09	1.8E-16	0.11
Cthe_0925hypothetical protein	272	64	596	20	30	195	111	11	683	75	741	52	1.2E-01	1.4E-03	0.21		
Cthe_0926signal recognition particle-docking protein FtsY					446	235	920	24	58	556	210	18	2664	149	1559	105	1.7E-04	4.1E-06
0.17																	
Cthe_09323-oxoacyl-[acyl-carrier-protein] synthase II			1738	1409	1931	59	452	1843	382	36	12618	644	7491	363	1.9E-13	2.3E-17	0.10
Cthe_0933acyl carrier protein	3014	1929	1660	422	1242	3109	2319	689	24089	5353	49408	8246	7.4E-199	8.6E-17	0.08		
Cthe_09343-oxoacyl-[acyl-carrier-protein] reductase			422	596	942	38	49	684	87	16	3994	115	1425	157	9.7E-05	4.1E-19	0.10
Cthe_0935[Acyl-carrier-protein] S-malonyltransferase			522	657	1118	35	182	634	130	16	3908	165	2184	143	1.0E-04	2.5E-14	0.11
Cthe_09363-oxoacyl-[acyl-carrier-protein] synthase III			610	641	1235	35	121	639	93	13	2662	127	1347	90	5.5E-02	6.5E-09	0.14
Cthe_0937phosphate:acyl-[acyl carrier protein] acyltransferase					897	1183	1504	50	337	1049	249	27	4030	300	3153	162	3.6E-03	6.9E-06
0.17																	
Cthe_0938fatty acid biosynthesis transcriptional regulator					952	1419	957	85	49	700	29	18	1826	111	903	111	5.3E-01	1.5E-07
0.16																	

Cthe_0942	MiaB-like tRNA modifying protein YliG	1009	1039	2132	44	65	1459	141	17	3710	93	877	73	3.4E-01	6.9E-03	0.23
Cthe_0948	dihydroorotate oxidase B, electron transfer subunit	157	26	626	14	9	182	77	7	1173	39	776	52	7.6E-03	2.0E-11	0.13
Cthe_0949	carbamoyl phosphate synthase large subunit	634	198	2430	14	36	489	234	5	7507	124	2401	64	2.0E-06	1.6E-36	0.08
Cthe_0950	carbamoyl-phosphate synthase small subunit	264	51	1040	17	9	175	86	5	2040	46	891	57	1.5E-02	2.9E-25	0.09
Cthe_0952	dihydroorotase	344	86	1148	17	12	203	135	6	3029	113	1390	75	8.2E-05	1.6E-30	0.08
Cthe_0953	aspartate carbamoyltransferase catalytic subunit	220	137	749	16	5	147	45	4	2293	26	757	67	1.5E-03	9.7E-108	0.06
Cthe_0954	Uracil phosphoribosyltransferase	1048	1092	909	79	105	971	171	36	5641	114	3582	343	1.2E-05	1.7E-17	0.10
Cthe_0956	hypothetical protein	5103	4269	3005	378	7524	12375	472	591	1040	163	1109	117	4.3E-02	7.7E-05	5.05
Cthe_0957	protein translocase subunit yajC	206	108	282	28	10	181	34	12	955	73	561	120	5.5E-04	7.3E-29	0.10
Cthe_0959	S-adenosylmethionine:tRNA ribosyltransferase-isomerase	465	133	1047	22	26	369	88	8	2043	76	885	62	4.5E-02	1.6E-11	0.13
Cthe_0961	aspartate semialdehyde dehydrogenase	953	969	1485	49	149	672	167	17	4212	71	1332	116	7.0E-02	2.5E-08	0.15
Cthe_0962	dihydrodipicolinate synthase	780	865	1321	48	89	728	222	22	3219	87	1138	106	1.2E-01	5.3E-04	0.21
Cthe_0963	dihydrodipicolinate reductase	826	646	1238	51	74	565	173	20	2197	53	560	78	4.1E-01	9.7E-03	0.26
Cthe_0965	cob(I)yrinic acid a,c-diamide adenosyltransferase	439	274	536	33	20	331	118	17	2503	143	1550	172	3.2E-06	5.3E-22	0.10
Cthe_0976	phospho-N-acetylmuramoyl-pentapeptide-transferase	494	221	1005	24	133	804	185	19	2023	156	1243	77	1.2E-02	1.0E-02	0.25
Cthe_0983	cell cycle protein	923	556	1242	34	211	751	138	15	2431	164	1775	85	6.1E-02	5.0E-05	0.18
Cthe_0987	riboflavin kinase / FMN adenylyltransferase	459	176	796	21	46	358	141	11	1214	217	1199	69	1.5E-02	6.0E-07	0.16
Cthe_0989	phosphoesterase, RecJ-like protein	504	353	1062	28	232	572	148	17	1533	198	1064	70	7.9E-02	6.6E-03	0.24
Cthe_0994	NusA antitermination factor	862	871	1520	38	70	976	157	15	2486	149	1003	67	2.9E-01	4.3E-03	0.22
Cthe_0995	hypothetical protein	485	424	441	41	7	283	32	10	786	43	360	56	7.1E-01	8.8E-07	0.18

Cthe_0999	1-deoxy-D-xylulose 5-phosphate reductoisomerase		496	321	1062	23	92	414	84	8	2042	89	601	52	1.4E-01	1.9E-07
	0.15															
Cthe_1000	phosphatidate cytidylyltransferase		552	414	974	33	229	544	135	18	2073	108	1004	83	7.0E-02	1.7E-03	0.22
Cthe_1001	undecaprenyl pyrophosphate synthetase		579	465	883	36	117	693	131	20	2454	82	857	95	5.2E-02	7.0E-04	0.21
Cthe_1005	elongation factor Ts	1311	558	1122	65	53	1805	494	66	10185	636	5658	564	3.3E-27	6.8E-12	0.12	
Cthe_1006	30S ribosomal protein S2		1583	789	1535	73	150	2114	1042	94	23339	864	9507	953	6.3E-76	5.0E-16	0.10
Cthe_1007	GTP-binding protein TypA		734	250	1648	20	85	584	182	8	4625	153	2091	78	1.5E-04	4.3E-17	0.10
Cthe_1008	aminodeoxychorismate lyase		747	495	948	27	62	860	202	17	3238	165	1610	97	1.8E-03	2.0E-05	0.18
Cthe_1020	extracellular solute-binding protein		115884	56365	44308	2251	12140	154216	39477	2681	957331	27550	363001	20665	1.9E-03	2.4E-03	0.13
Cthe_1022	glycerol-3-phosphate dehydrogenase (NAD(P)(+))		362	154	966	20	31	316	97	8	1915	111	1195	69	8.6E-03	1.1E-14	
	0.12															
Cthe_1026	50S ribosomal protein L32P		1805	1013	892	265	474	1483	950	326	12510	2634	20112	4225	1.9E-122	2.6E-29	0.08
Cthe_1027	hypothetical protein	3310	2226	1804	210	611	2542	762	142	26441	3415	37913	3029	2.9E-79	1.5E-91	0.05	
Cthe_1028	acetate kinase	1144	365	1213	32	51	1060	448	27	19121	480	4957	434	1.5E-78	1.2E-57	0.06	
Cthe_1029	phosphotransacetylase		469	287	1025	23	28	641	231	16	12111	203	2074	280	1.1E-50	1.2E-70	0.06
Cthe_1037	cell wall hydrolase/autolysin		948	587	1127	44	138	648	106	16	3373	130	1336	123	2.7E-02	1.5E-10	0.13
Cthe_1039	ribosomal protein S20		5350	2542	1885	403	502	3574	3224	541	73844	8581	64400	9828	9.1E-223	4.5E-35	0.06
Cthe_1047	hypothetical protein	4114	2933	2208	244	477	4219	751	163	7340	1497	8897	807	1.9E-02	4.1E-03	0.20	
Cthe_1050	recA protein	1703	1643	2919	85	429	2398	488	53	9958	415	3128	281	1.6E-03	5.6E-04	0.19	
Cthe_1053	L-lactate dehydrogenase		434	237	846	22	20	313	116	9	3135	110	990	95	1.8E-04	3.2E-23	0.09
Cthe_1057	phage shock protein PspC		340	146	206	50	290	448	172	87	1204	150	1616	344	1.8E-10	1.1E-02	0.25
Cthe_1058	serine hydroxymethyltransferase		2338	1107	1867	61	220	1531	428	32	13168	492	4252	313	5.8E-09	5.6E-17	0.10
Cthe_1063	thiamine biosynthesis/tRNA modification protein ThiI		582	401	1254	26	31	531	134	10	2847	75	948	69	5.2E-02	1.6E-08	
	0.14															
Cthe_1064	class V aminotransferase		778	677	1284	33	41	701	132	12	2329	74	698	57	3.3E-01	9.3E-04	0.21
Cthe_1075	hypothetical protein	797	820	1458	62	122	870	130	25	2276	116	995	106	3.3E-01	4.5E-03	0.24	

Cthe_1080carbohydrate-binding family 25 protein	1458	1313	1151	175	1655	3307	100	211	280	41	433	53	6.7E-03	5.3E-04	3.98	
Cthe_1093methenyltetrahydrofolate cyclohydrolase	586	318	1101	33	65	717	145	18	1768	87	758	66	2.1E-01	2.2E-02	0.27	
Cthe_1164FAD dependent oxidoreductase	919	804	2051	33	217	951	394	19	2898	444	2174	86	4.1E-02	4.8E-03	0.22	
Cthe_1166hypothetical protein	1365	1332	1748	72	146	1057	279	30	3594	222	1888	144	1.9E-01	8.6E-04	0.21	
Cthe_1167radical SAM family protein	367	110	1016	23	73	355	118	11	1452	67	594	52	1.9E-01	9.2E-04	0.21	
Cthe_1168hypothetical protein	709	287	1929	29	58	811	191	13	3507	113	938	70	7.6E-02	6.9E-05	0.19	
Cthe_1172hypothetical protein	657	439	1252	29	91	658	79	10	3082	56	463	66	1.1E-01	4.2E-07	0.15	
Cthe_1183ATPase	549	557	1231	36	67	573	131	14	1577	75	658	54	4.3E-01	1.2E-02	0.26	
Cthe_1187hypothetical protein	411	135	680	20	100	321	92	10	4075	167	1642	149	9.2E-14	3.5E-48	0.07	
Cthe_1188hypothetical protein	259	79	717	18	47	162	71	7	1959	86	442	70	4.8E-03	3.1E-21	0.10	
Cthe_1192hypothetical protein	968	715	1491	44	98	1923	372	38	12045	573	3411	345	1.6E-22	1.4E-13	0.11	
Cthe_1199amidohydrolase	1180	911	1501	39	365	1032	170	19	2611	366	1448	85	1.2E-01	2.4E-03	0.22	
Cthe_1200adenosylhomocysteinase	847	558	1308	31	223	761	266	19	3164	244	1616	92	1.2E-02	5.0E-04	0.21	
Cthe_1201purine nucleoside phosphorylase I	535	423	887	31	83	408	103	12	2006	100	741	76	1.0E-01	3.9E-07	0.16	
Cthe_1207hypothetical protein	1286	1165	2211	43	369	1315	301	22	4714	191	2009	97	1.0E-01	5.0E-03	0.23	
Cthe_1211tryptophan synthase subunit beta	5478	4718	4429	154	1271	4774	767	79	37910	2663	12563	901	5.7E-05	7.4E-17	0.09	
Cthe_1220hypothetical protein	1250	645	979	45	705	1078	389	44	3758	381	4514	214	1.1E-06	1.9E-03	0.21	
Cthe_122350S ribosomal protein L20	1706	964	1356	162	53	2088	663	152	8944	1029	5486	1047	1.8E-14	4.4E-07	0.15	
Cthe_122450S ribosomal protein L35P	767	472	707	140	1	856	96	72	2782	78	449	361	6.2E-02	1.3E-04	0.20	
Cthe_1225translation initiation factor 3	1340	746	1285	97	34	1415	117	44	3830	147	1300	231	7.0E-02	1.3E-04	0.19	
Cthe_1226dihydrofolate reductase	576	348	593	43	25	392	47	14	853	49	301	53	9.2E-01	1.4E-02	0.26	
Cthe_1227thymidylate synthase	1103	869	1171	54	91	831	52	16	1583	69	617	59	7.9E-01	1.7E-02	0.27	
Cthe_1228threonyl-tRNA synthetase	2203	2797	4028	68	159	1525	381	19	21924	454	4916	301	2.1E-04	7.9E-52	0.06	
Cthe_1229spore germination protein-like protein	549	492	883	44	37	544	46	14	3248	71	816	141	9.7E-03	3.8E-21	0.10	

Cthe_1232	AMP-dependent synthetase and ligase	703	769	1863	27	52	637	125	7	3682	110	1185	61	1.0E-01	3.2E-12	0.11
Cthe_1236	fibronectin, type III	244	65	529	16	67	231	94	11	913	36	825	52	3.7E-02	5.5E-04	0.21
Cthe_1237	leucyl-tRNA synthetase	3707	2248	4189	58	843	4325	1647	54	41953	2252	18507	559	2.1E-24	2.4E-13	0.10
Cthe_1238	iojap-like protein	219	157	315	28	11	214	29	11	1449	37	431	118	1.4E-03	6.3E-33	0.09
Cthe_1240	metal dependent phosphohydrolase	273	138	505	22	59	310	160	20	1281	135	1118	102	3.2E-04	6.5E-05	0.20
Cthe_1241	nicotinate-nucleotide adenylyltransferase	226	124	569	20	17	204	69	9	1058	39	522	55	1.1E-01	1.5E-07	0.16
Cthe_1245	phosphoribosylamine--glycine ligase	1109	522	1861	39	127	903	395	23	8050	414	3791	213	8.4E-10	1.3E-15	0.11
Cthe_1246	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	795	317	1602	24	27	504	170	8	6450	109	1531	109	8.0E-06	3.5E-36	0.07
Cthe_1247	phosphoribosylglycinamide formyltransferase	323	119	650	24	71	217	167	17	2272	123	1072	121	1.7E-05	3.2E-10	0.14
Cthe_1248	phosphoribosylaminoimidazole synthetase	395	172	1200	24	23	258	85	6	3254	58	664	81	6.0E-03	3.4E-30	0.07
Cthe_1249	amidophosphoribosyltransferase	469	286	1352	20	24	337	88	5	5499	78	879	91	8.4E-06	1.4E-70	0.05
Cthe_1250	phosphoribosylaminoimidazole carboxylase, catalytic subunit	222	151	589	26	13	131	64	8	2299	50	325	109	1.2E-03	4.2E-50	0.07
Cthe_1251	xanthine/uracil/vitamin C permease	4740	2165	2677	97	411	2387	563	41	25391	896	7222	519	1.9E-06	5.8E-27	0.08
Cthe_1252	auxin efflux carrier	413	288	1071	23	58	531	132	11	3931	71	1566	105	1.5E-05	3.1E-17	0.10
Cthe_1255	hypothetical protein	348	135	399	27	127	463	379	51	1229	248	2061	184	2.9E-10	3.5E-02	0.28
Cthe_1259	CDP-alcohol phosphatidyltransferase	414	171	415	26	127	369	267	32	1686	320	1949	179	6.9E-11	2.4E-05	0.18
Cthe_1260	4-hydroxybenzoyl-CoA thioesterase	338	240	473	36	36	351	150	26	1486	134	1001	144	1.8E-03	1.3E-05	0.18
Cthe_1261	6-phosphofructokinase	740	581	1202	37	49	792	160	17	6734	178	1547	185	1.0E-07	3.5E-24	0.09
Cthe_1262	hypothetical protein	1152	773	1011	213	183	2095	349	214	3336	755	4742	1109	1.6E-08	5.2E-04	0.19
Cthe_1263	transcription attenuation protein MtrB	1074	700	808	158	115	1712	335	155	3815	731	5378	1021	1.2E-13	4.6E-07	0.15
Cthe_1265	alpha-phosphoglucomutase	2261	1489	2344	50	162	2685	858	40	12433	534	6791	243	3.4E-08	6.5E-05	0.16
Cthe_1269	hypothetical protein	442	228	1363	22	28	449	212	11	2575	86	669	56	6.9E-02	2.9E-04	0.20
Cthe_1272	putative RNA polymerase sigma factor SigI	220	103	393	13	9	355	936	60	82	58	253	15	1.0E+00	6.5E-04	4.00

Cthe_1273alpha-L-arabinofuranosidase B	863	790	2312	39	97	1466	3728	123	379	143	842	24	1.0E+00	8.6E-06	5.13	
Cthe_1276phosphopantetheine adenylyltransferase	468	262	542	37	51	604	128	27	1224	78	680	91	1.4E-01	4.9E-02	0.30	
Cthe_127950S ribosomal protein L28	2548	1107	752	334	112	1510	1228	382	37145	4275	38994	9725	0.0E+00	7.6E-140	0.04	
Cthe_1286peptidase S1 and S6, chymotrypsin/Hap	1118	1020	2404	42	91	2304	1231	53	27347	2800	16050	713	1.8E-137	2.5E-31	0.07	
Cthe_1288two component transcriptional regulator	465	260	925	34	67	551	98	17	2218	98	657	95	4.2E-02	5.2E-06	0.18	
Cthe_1289hypothetical protein	152	121	301	42	4	157	35	17	779	38	250	121	5.0E-02	6.4E-14	0.14	
Cthe_1302hypothetical protein	1186	496	1475	26	63	940	365	16	3742	243	1848	78	1.3E-02	1.0E-03	0.21	
Cthe_1303group 1 glycosyl transferase	1167	563	1329	35	235	1012	194	19	1975	313	1929	83	8.0E-02	3.7E-03	0.23	
Cthe_1306hypothetical protein	1075	1704	2128	177	535	1541	1965	275	7204	1560	13086	1294	8.1E-20	2.0E-02	0.21	
Cthe_1312glycyl-tRNA synthetase	2179	1257	2255	58	257	2194	501	36	10602	465	3517	230	4.0E-05	9.1E-06	0.16	
Cthe_1313phosphopantothenoylcysteine decarboxylase / phosphopantothenate-cysteine ligase	646	345	1245	26	175	521	258	16	2968	342	2125	107	1.1E-04	1.3E-07	0.15	
Cthe_1314DNA-directed RNA polymerase subunit omega	371	176	335	53	25	432	315	80	2015	333	2275	469	2.2E-19	4.2E-06	0.17	
Cthe_1315guanylate kinase	904	443	1212	59	128	1193	337	49	3478	302	2109	221	4.3E-04	3.4E-03	0.22	
Cthe_1317hypothetical protein	904	450	1099	39	53	1012	174	22	3236	174	1389	120	9.7E-03	7.7E-05	0.18	
Cthe_1321chaperone protein DnaJ	465	186	1222	22	65	536	205	13	1805	95	1007	54	1.0E-01	1.0E-02	0.24	
Cthe_1326GTP-binding protein LepA	889	392	2079	26	95	1066	292	14	7466	301	2764	126	2.1E-07	3.5E-14	0.11	
Cthe_1330thylakoidal processing peptidase	520	253	639	38	17	361	138	20	1683	88	931	112	3.7E-02	3.5E-06	0.18	
Cthe_1331aspartyl-tRNA synthetase	1549	1058	2652	41	186	1099	248	14	13830	322	2275	196	2.8E-07	7.4E-39	0.07	
Cthe_1332histidyl-tRNA synthetase	778	666	1637	35	66	491	95	8	5706	80	1027	113	4.0E-03	2.1E-42	0.07	
Cthe_1334FHA domain-containing protein	533	409	1952	26	57	1440	2244	73	477	73	539	16	1.0E+00	1.5E-04	4.56	
Cthe_1337type II secretion system protein	408	240	801	23	70	346	146	13	996	126	877	54	1.6E-01	3.4E-03	0.24	
Cthe_1345adenine phosphoribosyltransferase	412	311	695	39	25	325	95	16	1305	53	478	77	2.9E-01	3.1E-04	0.21	
Cthe_134830S ribosomal protein S21	4913	5280	3875	1156	133	6391	433	542	17921	831	3574	2834	3.6E-01	3.1E-03	0.19	

Cthe_1350single-strand binding protein	1108	745	1263	110	522	1193	508	108	2230	685	4669	468	1.6E-05	9.1E-03	0.23	
Cthe_1363lipopolysaccharide biosynthesis protein	1313	792	1454	36	88	1224	334	21	4427	279	2502	114	4.4E-03	1.2E-04	0.18	
Cthe_1367PHP-like protein	668	526	696	35	27	431	127	14	1774	97	1435	91	5.7E-02	2.6E-08	0.15	
Cthe_1375aspartate kinase	1649	1010	1678	45	232	1893	644	39	8484	683	5220	242	4.1E-09	1.9E-05	0.16	
Cthe_1376homoserine dehydrogenase	539	289	1252	23	22	278	82	5	3043	61	748	64	3.5E-02	2.5E-27	0.08	
Cthe_1377hypothetical protein	230	134	289	20	8	141	28	6	1278	38	547	88	8.0E-04	7.8E-65	0.07	
Cthe_1380response regulator receiver protein	1166	553	739	93	266	854	495	94	679	1360	2295	459	2.5E-07	1.0E-03	0.20	
Cthe_1385preprotein translocase subunit SecA	1978	1915	3799	40	298	2824	1003	29	27301	604	5726	260	3.1E-11	8.0E-12	0.11	
Cthe_1386hypothetical protein	692	589	927	52	285	694	233	37	7989	383	3760	440	6.2E-24	6.0E-30	0.08	
Cthe_1387hypothetical protein	557	456	539	56	125	531	136	35	2109	168	1725	224	5.8E-04	9.8E-08	0.16	
Cthe_1388TrpR like protein, YerC/YecD	426	356	408	58	22	400	57	25	1533	113	1170	212	5.2E-03	1.4E-13	0.12	
Cthe_1401hypothetical protein	447	306	355	47	130	508	146	42	1036	109	1524	175	5.3E-03	4.7E-03	0.24	
Cthe_1411tryptophan synthase, alpha chain	513	356	944	33	101	438	178	18	2216	223	1362	115	3.7E-03	2.5E-07	0.16	
Cthe_1412tryptophan synthase subunit beta	1066	1042	1686	46	89	616	100	10	2778	163	854	72	3.7E-01	7.9E-09	0.14	
Cthe_1421signal peptide peptidase SppA, 36K type	859	465	1218	36	69	967	196	21	4835	109	1112	130	1.1E-03	8.9E-07	0.16	
Cthe_1422RDD domain-containing protein	946	629	1405	54	161	1298	447	47	4370	130	1810	172	6.0E-03	4.9E-02	0.27	
Cthe_1424hypothetical protein	610	444	732	27	43	397	106	10	1349	92	1478	66	9.4E-02	4.5E-08	0.15	
Cthe_1431hypothetical protein	932	1029	2149	30	225	654	166	9	3771	146	1888	63	1.4E-01	1.5E-08	0.14	
Cthe_1451GCN5-like N-acetyltransferase	208	70	376	19	9	183	45	8	450	67	584	54	7.7E-02	2.9E-09	0.15	
Cthe_1456ABC transporter-like protein	257	58	569	15	41	130	29	4	1614	193	740	80	1.7E-05	7.6E-137	0.05	
Cthe_1457polar amino acid ABC transporter inner membrane subunit	355	59	797	25	37	177	49	7	1822	136	679	92	5.9E-03	6.5E-50	0.08	
Cthe_1458extracellular solute-binding protein	253	103	531	15	6	170	54	5	2859	152	1555	123	4.4E-15	1.4E-194	0.04	
Cthe_1465redoxin	661	249	572	46	24	551	47	19	384	156	556	69	5.1E-01	2.3E-02	0.28	
Cthe_1478TetR family transcriptional regulator	358	200	360	23	62	402	277	31	2281	353	3387	247	4.1E-29	4.5E-12	0.13	

Cthe_1479hypothetical protein	107	26	383	11	67	217	183	18	839	81	1069	71	2.2E-09	9.8E-03	0.25	
Cthe_1481hypothetical protein	449	107	1654	13	32	687	438	11	3325	189	2236	54	1.3E-04	1.8E-03	0.20	
Cthe_1482Cof-like hydrolase	414	178	826	24	59	432	216	18	1931	242	1583	109	9.5E-05	2.9E-06	0.17	
Cthe_1485beta-lactamase-like protein	1216	618	1498	37	54	1295	160	18	2102	231	1094	65	2.9E-01	3.0E-02	0.28	
Cthe_1495pyridoxamine 5'-phosphate oxidase-like FMN-binding protein	803	574	1117	90	167	1163	407	85	4477	464	5638	596	2.9E-14	3.5E-08	0.14	
Cthe_1503hypothetical protein	4747	4156	2319	453	150	2127	446	129	6971	743	3160	732	5.9E-01	8.9E-05	0.18	
Cthe_1509hypothetical protein	4128	2658	3263	116	230	12076	529	136	1009	240	582	41	1.0E-01	4.9E-02	3.32	
Cthe_1519SAM-dependent methyltransferase	334	82	739	13	74	251	191	10	1533	132	1568	58	2.5E-04	1.7E-06	0.17	
Cthe_1540aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	259	192	901	13	180	238	147	8	2832	140	626	57	3.1E-04	1.9E-09	0.14	
Cthe_1541aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit A	290	189	844	12	191	233	147	8	2645	146	594	52	5.7E-04	2.7E-08	0.15	
Cthe_1542aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit C	57	57	229	16	27	65	17	6	770	26	144	70	1.3E-05	5.3E-52	0.09	
Cthe_1543aspartyl-tRNA synthetase	578	494	1376	20	264	492	210	10	6089	307	1416	101	8.5E-08	1.1E-17	0.10	
Cthe_1588extracellular solute-binding protein	9726	4913	2185	218	1623	5525	75	84	283	11	189	9	1.7E-235	5.1E-25	9.33	
Cthe_1756putative virion core protein (lumpy skin disease virus)-like protein	1324	1250	1658	52	274	1063	274	24	7651	363	3439	214	3.3E-05	1.5E-14	0.11	
Cthe_1757peptidase M23B	415	325	960	26	17	291	62	6	2767	46	501	74	3.3E-02	4.8E-26	0.08	
Cthe_1758hypothetical protein	339	185	643	37	21	303	85	17	1512	116	702	122	2.0E-02	1.4E-09	0.14	
Cthe_17601A family penicillin-binding protein	1141	744	2378	24	314	974	258	11	9279	421	3552	117	2.4E-07	8.4E-23	0.09	
Cthe_1765hypothetical protein	569	454	1550	31	101	896	112	14	2531	120	996	69	1.2E-01	6.3E-04	0.20	
Cthe_1766gamma-glutamyl kinase	615	530	1082	38	65	464	77	11	3164	61	790	100	4.3E-02	2.8E-16	0.11	
Cthe_1767hypothetical protein	1064	464	1252	39	372	1057	569	42	6262	521	4894	261	1.1E-14	7.5E-06	0.16	
Cthe_1768NifU-like domain-containing protein	430	188	656	26	27	408	115	14	3438	67	944	134	2.7E-06	1.1E-18	0.10	
Cthe_1771rubrerythrin	2482	1358	1743	144	392	3325	740	138	18107	1429	13761	1340	1.9E-36	1.4E-14	0.10	

Cthe_1775peptidase M22, glycoprotease	332	218	889	28	120	400	180	20	1156	114	795	67	1.4E-01	4.4E-02	0.30	
Cthe_1776hypothetical protein	246	174	749	34	23	287	103	16	1086	25	288	60	3.9E-01	2.1E-02	0.27	
Cthe_1778copper amine oxidase-like protein		1313	620	1586	52	37	1705	660	51	13484	144	3479	365	1.2E-17	6.8E-08	0.14
Cthe_1779hypothetical protein	171	200	835	43	15	1301	2090	268	281	68	579	56	9.6E-01	1.6E-05	4.79	
Cthe_1780hypothetical protein	184	58	514	19	4	505	665	64	167	22	137	14	1.0E+00	4.5E-05	4.57	
Cthe_178230S ribosomal protein S9	1215	409	916	91	180	1252	575	107	6483	589	5198	707	8.2E-21	8.0E-07	0.15	
Cthe_178350S ribosomal protein L13	632	335	789	57	21	604	234	40	6553	196	2454	445	1.3E-18	7.6E-27	0.09	
Cthe_17953-deoxy-D-arabinoheptulosonate-7-phosphate synthase				2845	2345	2664	110	649	3384	585	73	19060	822	7321	577	8.9E-07
3.0E-09 0.13																
Cthe_1796prephenate dehydrogenase	3213	2980	2709	117	683	2810	883	74	16313	837	7137	487	6.8E-04	4.5E-06	0.15	
Cthe_17973-phosphoshikimate 1-carboxyvinyltransferase			3188	2612	3032	100	605	3179	698	60	13041	520	5276	319	2.4E-02	8.0E-04
0.19																
Cthe_1801ABC transporter-like protein	831	494	1279	48	145	1118	358	40	3187	296	1713	159	3.5E-03	1.5E-02	0.25	
Cthe_1837hypothetical protein	1090	692	1170	82	293	2450	739	127	5215	770	8858	656	1.9E-22	1.2E-03	0.19	
Cthe_1840cysteine synthase	993	645	1059	41	211	1242	401	37	6282	586	3881	266	8.1E-14	1.6E-08	0.14	
Cthe_1844BadM/Rrf2 family transcriptional regulator		253	182	573	31	32	254	50	12	1381	59	525	92	4.9E-02	4.7E-12	0.13
Cthe_1845homoserine O-succinyltransferase		1115	654	1253	47	123	1140	168	24	4488	214	1703	153	3.4E-03	4.7E-07	0.16
Cthe_1853cold-shock DNA-binding protein family protein				1984	1296	1050	310	509	2842	1093	439	6569	1905	15111	2856	3.0E-33	1.1E-05
0.15																
Cthe_1856small acid-soluble spore protein beta		78	48	156	23	7	137	31	17	401	41	348	104	2.7E-07	3.4E-10	0.16
Cthe_1857carboxyl-terminal protease	475	282	1032	19	39	569	110	9	2392	89	980	57	2.5E-02	1.8E-07	0.16	
Cthe_1858peptidase M23B	602	433	1826	35	63	702	191	15	2559	168	1239	79	1.2E-01	1.4E-04	0.19	
Cthe_1859cell division protein FtsX	386	236	1044	26	49	291	132	11	1897	93	1011	73	4.6E-02	6.9E-09	0.15	
Cthe_1860cell division ATP-binding protein FtsE		335	226	600	23	49	309	63	10	2123	92	1256	104	1.6E-04	9.1E-23	0.10
Cthe_1863N-acetyl-gamma-glutamyl-phosphate reductase				224	64	546	11	3	84	28	2	4149	42	694	96	6.2E-18	0.0E+00
0.02																

Cthe_1864acetylglutamate kinase	269	38	756	16	10	131	46	4	2344	25	468	63	2.4E-03	8.9E-95	0.06	
Cthe_1866acetylornithine aminotransferase	515	403	1388	27	33	368	108	7	7690	52	1006	149	3.6E-09	4.1E-102	0.05	
Cthe_1867carbamoyl-phosphate synthase small subunit	432	285	1308	26	21	286	85	6	5493	44	497	117	3.0E-05	8.5E-72	0.05	
Cthe_1868carbamoyl-phosphate synthase large subunit	1054	570	3540	22	54	736	330	7	15024	140	1444	107	2.3E-07	2.2E-49	0.07	
Cthe_1869ornithine carbamoyltransferase	602	355	1220	33	35	458	222	16	7862	164	1595	220	9.6E-14	2.4E-40	0.07	
Cthe_1870GCN5-like N-acetyltransferase	280	127	547	28	31	311	276	33	3837	226	2090	284	8.1E-26	1.3E-14	0.12	
Cthe_1912copper amine oxidase-like protein	6507	4127	4449	134	435	5578	1421	83	27048	983	10282	512	3.9E-02	3.1E-04	0.16	
Cthe_1914TPR repeat-containing protein	859	752	1852	38	113	925	190	16	2309	162	1358	67	2.9E-01	6.3E-03	0.24	
Cthe_1918arginine decarboxylase	1722	1106	2027	47	356	1363	315	23	17165	759	5341	350	4.5E-20	2.3E-45	0.07	
Cthe_1923CTP synthetase	1763	792	2506	44	178	1308	327	19	11886	352	3074	201	8.1E-07	4.8E-20	0.09	
Cthe_1933beta-lactamase fold Zn-dependent hydrolase	489	334	1068	42	25	714	101	20	1820	62	864	91	1.7E-01	1.3E-03	0.22	
Cthe_1944hypothetical protein	1202	693	776	69	496	1268	384	71	3724	541	4672	378	3.0E-09	2.3E-04	0.19	
Cthe_1945thioredoxin-disulfide reductase	1789	639	1148	56	226	1506	553	50	3191	1191	2200	231	3.3E-05	4.0E-03	0.22	
Cthe_1960peptidoglycan binding domain-containing protein	2100	1451	2300	77	254	2526	237	40	13065	492	2684	330	3.9E-06	6.2E-11	0.12	
Cthe_1970hypothetical protein	757	481	862	98	50	1134	363	97	3073	340	2646	458	3.0E-06	1.2E-03	0.21	
Cthe_1971hypothetical protein	650	392	749	77	22	834	116	45	2226	90	864	209	5.0E-02	1.0E-03	0.22	
Cthe_1988hypothetical protein	814	566	942	119	63	1060	475	121	2406	314	3028	471	1.6E-04	2.2E-02	0.26	
Cthe_2018hypothetical protein	1357	590	3369	60	120	5221	3414	169	571	234	1148	42	1.0E+00	5.5E-03	4.02	
Cthe_2019hypothetical protein	1686	817	3072	42	107	5752	1918	82	756	94	631	18	1.0E+00	3.5E-04	4.56	
Cthe_2022hypothetical protein	815	566	947	119	63	1060	476	121	2406	315	3028	471	1.7E-04	2.2E-02	0.26	
Cthe_2030hypothetical protein	1108	497	2093	31	98	3167	1747	67	755	80	697	20	1.0E+00	2.4E-02	3.35	
Cthe_2070hypothetical protein	646	488	1136	37	196	638	232	23	1693	258	2098	110	1.6E-02	1.1E-03	0.21	

Cthe_2071hypothetical protein	1616	1646	1674	77	290	2943	601	70	384	87	347	23	9.0E-03	3.6E-02	3.04	
Cthe_2074hypothetical protein	1321	1112	1093	232	2874	3335	186	415	348	71	564	107	2.0E-01	1.6E-03	3.88	
Cthe_2079thylakoidal processing peptidase	1170	776	953	73	273	1382	415	67	3009	382	3987	294	8.3E-05	6.6E-03	0.23	
Cthe_2092dimethyladenosine transferase	386	112	787	21	49	331	77	9	1180	103	726	54	1.1E-01	3.5E-06	0.17	
Cthe_2093hypothetical protein	777	422	1153	31	25	514	66	8	4987	137	1701	136	1.8E-05	1.9E-64	0.06	
Cthe_2094hypothetical protein	1571	1533	1229	113	2835	2756	475	173	579	90	388	48	1.3E-01	5.4E-03	3.60	
Cthe_2095TatD family hydrolase	468	281	954	31	115	567	340	30	3293	320	2964	193	1.4E-10	1.3E-07	0.16	
Cthe_2096methionyl-tRNA synthetase	891	623	2359	27	126	732	199	9	5961	178	1851	86	5.3E-03	6.9E-16	0.10	
Cthe_2102methyltransferase small	374	210	823	26	11	308	95	10	1541	74	539	63	1.4E-01	9.6E-07	0.16	
Cthe_21034Fe-4S ferredoxin, iron-sulfur binding		7265	3599	2812	1072	4278	7824	10139	3240	40505	14656	80006	19386	8.0E-72	3.0E-03	0.17
Cthe_2104hypothetical protein	945	488	1229	43	165	941	234	27	3329	225	2238	146	2.3E-03	4.1E-05	0.18	
Cthe_2105DNA polymerase III subunit delta'		822	254	1161	31	105	632	244	19	1584	306	1312	83	4.7E-02	4.5E-03	0.23
Cthe_2106hypothetical protein	378	184	570	35	35	427	220	33	1588	186	1228	156	2.5E-04	5.6E-04	0.21	
Cthe_2107thymidylate kinase	505	212	883	33	23	386	134	15	1929	125	943	98	2.7E-02	5.9E-08	0.15	
Cthe_2108arginine decarboxylase	902	582	1629	30	177	1171	1952	68	13918	2199	17206	540	5.1E-153	1.5E-07	0.13	
Cthe_2109copper amine oxidase-like protein		1506	1260	3073	32	193	1676	588	18	16485	768	4945	193	9.2E-13	3.8E-18	0.09
Cthe_2115hypothetical protein	194	55	282	26	51	146	86	21	1492	185	1533	258	7.2E-29	3.5E-33	0.08	
Cthe_2116binding-protein-dependent transport systems inner membrane component						362	85	588	19	74	290	146	14	2421	201	1498	123
1.3E-09	2.1E-16	0.11														
Cthe_2117putative sulfonate transport system substrate-binding protein					322	85	1020	19	36	268	153	10	2604	78	491	66	7.6E-03
3.9E-09	0.15															
Cthe_2118ABC transporter-like protein	508	283	980	33	165	592	202	24	1947	115	898	87	6.1E-02	2.7E-02	0.28	
Cthe_2142recombination protein RecR	813	364	770	46	180	1095	263	44	5689	419	2626	333	3.0E-16	4.5E-10	0.13	
Cthe_2143hypothetical protein	415	169	739	54	47	518	69	28	3338	117	822	272	7.0E-06	1.4E-20	0.10	
Cthe_2148carbohydrate-binding, CenC-like protein		544	286	1089	35	23	385	90	11	1352	100	914	69	2.5E-01	4.8E-07	0.16

Cthe_2150integral hypothetical protein	1095	661	1774	37	210	903	303	19	3671	209	2328	100	2.5E-02	2.8E-04	0.19	
Cthe_2152hypothetical protein 1118	1393	1029	99	426	818	342	59	2338	275	3933	276	1.9E-02	1.6E-03	0.21		
Cthe_2164rubredoxin-type Fe(Cys)4 protein		1957	805	530	296	758	1949	667	397	3011	2617	6555	2551	4.3E-28	6.2E-06	0.16
Cthe_2165hypothetical protein 557	260	591	33	166	2504	27	55	25	3	13	1	1.0E+00	0.0E+00	55.00		
Cthe_2168peptidase 1327 1260	2941	36	133	2305	1663	45	27086	1486	16320	445	1.7E-71	2.8E-12	0.10			
Cthe_2171type III restriction enzyme, res subunit	1712	910	4337	33	206	2030	508	16	7828	481	2772	84	5.1E-02	7.6E-04	0.19	
Cthe_217250S ribosomal protein L31P 3576	1831	2097	504	309	3133	2359	671	25366	4436	36470	7350	3.2E-78	3.8E-15	0.09		
Cthe_2182hypothetical protein 1104	1089	2793	25	249	917	285	9	5511	178	1914	58	8.6E-02	2.2E-06	0.16		
Cthe_2183UDP-glucose pyrophosphorylase 1440	1413	1758	76	80	1546	165	30	6720	158	1687	208	2.4E-02	2.2E-08	0.14		
Cthe_218530S ribosomal protein S18 1226	469	922	130	195	1468	1335	272	14040	1630	12558	2301	3.9E-146	1.1E-10	0.12		
Cthe_2186single-strand binding protein 1379	533	990	99	32	1083	159	48	21750	332	4124	1312	8.4E-75	3.2E-213	0.04		
Cthe_218730S ribosomal protein S6P 779	289	448	75	86	686	248	69	11444	347	4457	1196	1.7E-93	6.4E-73	0.06		
Cthe_2202hypothetical protein 424	203	794	27	238	453	175	22	1856	188	1330	106	1.3E-03	5.7E-04	0.21		
Cthe_2203GTP cyclohydrolase I 572	228	812	40	184	964	237	43	4141	387	3512	322	1.1E-18	1.8E-10	0.13		
Cthe_22093-isopropylmalate dehydrogenase	765	264	1035	26	138	820	172	17	1834	228	926	67	6.4E-02	1.5E-02	0.25	
Cthe_2246anti-sigma-28 factor, FlgM 16933	9818	4583	1538	481	22221	144	937	510	292	589	157	1.4E-23	1.7E-07	5.97		
Cthe_2251methionine adenosyltransferase 862	209	1235	27	61	688	175	13	5191	125	1347	118	1.7E-05	1.6E-14	0.11		
Cthe_2252thioesterase superfamily protein	295	129	548	34	19	191	56	12	1214	36	492	91	1.0E-01	1.9E-13	0.13	
Cthe_2253ATP-dependent metalloprotease FtsH	4671	5850	5499	129	714	4875	1185	66	23993	1162	7789	406	1.3E-01	3.9E-04	0.16	
Cthe_2256primary replicative DNA helicase 874	692	1813	36	130	921	143	14	2599	159	692	59	3.4E-01	5.4E-03	0.24		
Cthe_225750S ribosomal protein L9 366	316	760	46	42	354	61	16	1183	74	481	88	3.1E-01	6.1E-05	0.18		
Cthe_2259hypothetical protein 1408	1016	722	133	61	602	156	43	868	336	1272	204	4.0E-01	4.9E-04	0.21		
Cthe_2260prephenate dehydratase 632	352	1350	39	51	531	87	12	1653	43	562	57	4.7E-01	1.1E-03	0.21		
Cthe_2261hypothetical protein 359	572	640	69	25	289	22	14	668	29	211	61	1.0E+00	4.6E-04	0.23		

Cthe_2262V-type ATPase, 116 kDa subunit	1823	2578	2815	53	386	1808	281	20	3779	306	2162	73	6.0E-01	4.9E-02	0.27	
Cthe_2263V-type ATP synthase subunit K	449	495	835	54	83	501	106	25	1031	89	727	89	4.0E-01	2.0E-02	0.28	
Cthe_2266V-type ATP synthase subunit F	159	192	373	32	17	165	43	12	440	45	272	55	4.5E-01	1.1E-04	0.22	
Cthe_2268V-type ATP synthase subunit B	1187	1131	1672	41	259	1079	186	17	3133	239	1144	76	2.5E-01	4.2E-03	0.22	
Cthe_2269V-type ATP synthase subunit D	945	502	915	50	109	823	303	36	2226	342	2488	179	1.1E-03	3.5E-04	0.20	
Cthe_2276AAA ATPase, central region	647	540	1042	27	54	551	109	10	2559	115	1143	70	4.9E-02	1.2E-09	0.14	
Cthe_2278extracellular solute-binding protein	586	395	730	28	110	412	39	9	2236	70	921	79	3.9E-02	3.7E-14	0.11	
Cthe_2305undecaprenyl pyrophosphate phosphatase	696	378	790	32	335	678	387	36	5380	427	5239	295	2.0E-27	3.9E-12	0.12	
Cthe_2306MATE efflux family protein	827	422	1552	28	197	622	264	15	7023	409	3242	168	1.4E-11	1.4E-25	0.09	
Cthe_2328UDP-N-acetylglucosamine 1-carboxyvinyltransferase	1280	884	2205	49	408	1375	236	25	3491	481	2873	131	3.2E-02	3.9E-04	0.19	
Cthe_2330hypothetical protein	1528	1462	2341	52	78	882	191	13	2217	153	1396	57	6.7E-01	4.4E-03	0.23	
Cthe_2331hypothetical protein	6194	4000	2962	357	229	4095	300	121	7588	715	7732	673	7.4E-01	3.0E-04	0.18	
Cthe_2333two component transcriptional regulator	490	294	794	32	79	449	145	18	1813	191	1533	116	2.7E-03	4.7E-08	0.16	
Cthe_2334polysaccharide biosynthesis protein CapD	1344	1290	2589	40	300	1448	688	27	5833	308	3046	108	2.3E-02	3.3E-02	0.25	
Cthe_2348hypothetical protein	129807	115039	101048	1609	12912	149001	34991	1111	678932	21039	238296	6429	1.0E+00	4.6E-02	0.17	
Cthe_2355L-aspartate oxidase	852	464	2405	32	51	602	151	9	3465	101	867	59	2.6E-01	1.9E-08	0.15	
Cthe_2356quinolinate synthetase A	2245	1409	2022	88	305	3136	1214	102	20464	1982	15650	954	2.2E-51	7.3E-12	0.11	
Cthe_2357hypothetical protein	1283	772	1472	42	91	1613	883	49	24729	482	7346	579	1.9E-82	2.5E-23	0.08	
Cthe_2366single-stranded nucleic acid binding R3H	499	226	892	36	5	467	69	13	1402	55	227	60	4.1E-01	1.7E-03	0.22	
Cthe_2369ribonuclease P	500	344	596	54	5	635	148	36	1806	75	567	140	8.1E-02	1.0E-02	0.26	
Cthe_237050S ribosomal protein L34	328	180	232	78	14	348	54	49	989	106	469	280	1.0E-02	1.1E-05	0.18	
Cthe_2375hypothetical protein	256	122	354	33	8	202	66	17	665	27	213	62	3.7E-01	9.7E-03	0.27	
Cthe_2381seryl-tRNA synthetase	1060	738	2002	42	39	1220	348	23	3910	188	1395	95	9.7E-02	1.1E-02	0.24	
Cthe_2383hypothetical protein	5196	3695	3782	199	605	5462	1969	172	69176	1668	25754	2212	2.2E-26	5.8E-19	0.08	

Cthe_2384	S-layer-like domain-containing protein	3874	3298	2507	245	660	2152	127	71	5815	383	3562	380	5.9E-01	2.9E-04	0.19
Cthe_2390	pyruvate ferredoxin oxidoreductase subunit gamma	1684	1474	1737	121	244	1243	607	76	18086	553	7425	953	4.3E-23	5.8E-29	0.08
Cthe_2391	pyruvate ferredoxin oxidoreductase subunit delta	645	669	973	107	45	531	147	43	9582	184	1702	791	4.9E-18	9.1E-91	0.05
Cthe_2392	pyruvate ferredoxin oxidoreductase subunit alpha	1625	1757	2702	73	245	1534	848	49	30340	529	5234	640	1.7E-31	1.4E-30	0.08
Cthe_2393	thiamine pyrophosphate enzyme-like TPP-binding protein	2458	2080	2900	114	653	2600	845	82	31914	939	10374	983	6.8E-23	6.3E-24	0.08
Cthe_2399	formate-tetrahydrofolate ligase	1217	690	2371	36	144	1214	252	16	5859	181	1751	99	2.2E-02	1.8E-06	0.16
Cthe_2401	hypothetical protein	12734	11930	8361	1795	638	14591	2482	1080	19248	6068	34518	5786	1.1E-01	4.6E-02	0.19
Cthe_2402	peptidoglycan-binding LysM	1605	1122	2948	51	667	15933	540	141	670	204	540	26	1.0E+00	2.6E-06	5.42
Cthe_2403	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	302	82	775	19	31	328	86	9	2060	172	1232	93	2.3E-05	7.7E-21	0.10
Cthe_2404	GntR family transcriptional regulator	431	141	943	30	76	529	132	18	2164	128	1055	107	5.4E-03	3.5E-06	0.17
Cthe_2405	heavy metal transport/detoxification protein	189	69	171	29	70	247	111	42	559	85	556	139	4.1E-06	4.4E-02	0.30
Cthe_2408	phage shock protein PspA	2161	1100	1387	91	107	2197	402	61	8091	471	3413	367	4.8E-05	2.1E-05	0.17
Cthe_2419	DNA replication protein DnaD	742	785	1398	43	54	560	114	12	3043	99	1469	99	9.0E-02	4.2E-13	0.12
Cthe_2430	RnfABCDGE type electron transport complex subunit C	3864	1518	2109	80	530	2460	576	46	4799	1643	3556	230	3.9E-02	2.4E-03	0.20
Cthe_2431	RnfABCDGE type electron transport complex subunit D	1547	580	1389	51	253	1010	262	28	2926	384	1279	120	7.9E-02	4.5E-03	0.23
Cthe_2432	RnfABCDGE type electron transport complex subunit G	851	274	869	47	50	575	145	23	2055	270	909	137	2.3E-02	1.1E-06	0.17
Cthe_2433	RnfABCDGE type electron transport complex subunit E	807	206	794	38	166	501	146	22	1763	323	1006	124	9.4E-03	9.3E-06	0.18
Cthe_2434	RnfABCDGE type electron transport complex subunit A	776	201	968	46	318	593	182	33	1927	460	1387	177	6.7E-04	6.4E-05	0.19

Cthe_2435RnfABCDGE type electron transport complex subunit B	936	207	1336	43	233	720	260	28	1953	958	1825	193	6.4E-06	3.3E-08	0.15	
Cthe_2438hypothetical protein	170	161	173	39	38	157	33	21	225	35	263	68	4.0E-01	1.2E-02	0.31	
Cthe_2502tRNA-adenosine deaminase	696	280	741	50	253	759	258	49	3096	363	4733	372	2.3E-16	7.6E-11	0.13	
Cthe_2503cupin 2 barrel domain-containing protein	765	1249	1648	87	331	3180	817	127	279	31	165	18	9.9E-06	3.3E-13	7.06	
Cthe_2505LacI family transcription regulator	610	424	1046	28	90	495	159	13	1938	141	1423	75	5.0E-02	1.1E-05	0.17	
Cthe_2510hypothetical protein	878	424	890	36	289	597	181	22	8205	481	6189	375	2.4E-39	6.8E-69	0.06	
Cthe_2516acetolactate synthase, large subunit	865	583	2312	32	191	766	182	12	3135	208	1389	65	1.7E-01	3.7E-05	0.18	
Cthe_2517acetolactate synthase, small subunit	233	161	705	30	16	221	71	11	1445	42	476	81	9.0E-02	5.6E-10	0.14	
Cthe_2518ketol-acid reductoisomerase	1092	660	1691	49	129	1226	455	36	8088	442	3632	270	5.3E-10	1.9E-09	0.13	
Cthe_2519putative alpha-isopropylmalate/homocitrate synthase family transferase	599	392	1646	23	86	560	177	9	4046	148	1327	73	5.1E-03	1.3E-11	0.12	
Cthe_2579iron-containing alcohol dehydrogenase	434	217	1115	22	31	472	139	10	2106	83	857	59	6.0E-02	1.0E-05	0.17	
Cthe_2593peptide chain release factor 1	1153	636	1806	47	228	1838	504	43	5972	588	3322	215	1.2E-06	1.3E-03	0.20	
Cthe_2597ribose-5-phosphate isomerase	926	680	1168	88	65	1417	82	47	2906	90	849	183	1.6E-01	1.3E-02	0.26	
Cthe_2600glycosyl transferase family protein	2197	1307	1944	68	447	2259	391	44	7271	383	2663	203	8.2E-03	5.4E-03	0.22	
Cthe_2601UDP-N-acetylglucosamine 2-epimerase	1102	650	1736	43	76	1123	233	21	4584	218	1335	118	2.1E-02	1.7E-05	0.18	
Cthe_2603ATP synthase F0 subunit C	981	572	813	152	283	1083	666	205	8647	1729	12620	2503	1.5E-120	3.1E-27	0.08	
Cthe_2604ATP synthase F0 subunit B	708	722	1073	66	25	808	114	27	6729	128	1033	305	1.8E-06	1.4E-28	0.09	
Cthe_2605ATP synthase F1 subunit delta	334	382	789	38	16	378	110	16	2880	111	680	142	2.0E-03	5.9E-16	0.11	
Cthe_2606F0F1 ATP synthase subunit alpha	799	864	2296	37	166	1097	379	20	7109	456	2796	154	2.1E-05	1.3E-09	0.13	
Cthe_2607F0F1 ATP synthase subunit gamma	476	438	1243	34	21	547	158	15	3151	120	944	103	1.5E-02	5.1E-09	0.15	
Cthe_2608F0F1 ATP synthase subunit beta	1567	1087	2452	52	380	1939	860	47	13108	1205	6780	354	7.5E-17	2.1E-08	0.13	
Cthe_2609ATP synthase F1 subunit epsilon	869	503	863	78	22	847	284	55	5076	454	1956	436	1.8E-09	5.2E-12	0.13	
Cthe_2617peptidase M23B	10683	4893	7564	358	6775	99998	2990	1532	1510	1040	2832	186	8.9E-01	1.8E-09	8.24	

Cthe_2619rod shape-determining protein Mbl	1758	1209	1866	66	269	1937	301	38	6202	511	3673	229	1.2E-03	1.6E-05	0.17		
Cthe_2625(3R)-hydroxymyristoyl-ACP dehydratase	341	157	556	33	125	279	106	22	2026	192	3149	262	1.4E-14	3.8E-34	0.08		
Cthe_2627pur operon repressor	779	416	1110	40	374	709	317	33	3074	379	3615	197	3.1E-07	7.7E-06	0.17		
Cthe_2628hypothetical protein	5884	2109	1335	467	92	4488	517	268	9899	1127	9353	1648	8.7E-03	3.1E-05	0.16		
Cthe_2630ribose-phosphate pyrophosphokinase	1070	358	1458	42	103	1318	356	33	4554	257	2079	158	3.7E-04	1.4E-03	0.21		
Cthe_2633hypothetical protein	1056	912	1464	45	31	678	142	13	3360	125	1058	90	1.8E-01	8.3E-09	0.14		
Cthe_2651hypothetical protein	1639	2034	1044	207	3831	3025	804	385	656	115	657	106	3.3E-01	6.4E-03	3.63		
Cthe_2655AbrB family transcriptional regulator	765	383	734	48	327	3284	221	96	163	46	268	22	1.0E+00	6.0E-05	4.36		
Cthe_2657nucleoid protein Hbs	3164	1091	1334	288	36	3492	445	220	9320	747	2113	1060	2.1E-04	3.3E-03	0.21		
Cthe_2658RNA-binding S4 protein	235	70	134	26	5	180	26	13	412	45	260	71	4.2E-02	4.9E-07	0.18		
Cthe_2662RNA binding S1	4210	2729	1918	302	688	4259	701	214	12203	1126	16855	1546	4.2E-06	6.9E-07	0.14		
Cthe_2668hypothetical protein	182	179	438	30	11	238	122	22	620	87	578	82	1.0E-01	1.5E-02	0.27		
Cthe_2669hypothetical protein	425	368	800	34	11	409	135	16	1047	96	727	65	3.0E-01	5.6E-03	0.25		
Cthe_2676GumN	1184	1101	2270	45	73	1718	344	25	4218	214	1407	90	1.8E-01	4.9E-02	0.28		
Cthe_2684hypothetical protein	7477	6040	5612	535	752	15417	1446	495	833	139	923	88	9.1E-08	1.7E-06	5.63		
Cthe_2687thioesterase superfamily protein	191	144	611	26	8	261	93	14	825	45	412	56	2.9E-01	1.1E-02	0.25		
Cthe_2689GntR family transcriptional regulator	894	792	1728	32	133	607	363	16	2287	259	2051	71	1.2E-01	3.8E-03	0.23		
Cthe_2693hypothetical protein	498	203	570	36	41	490	103	21	2957	97	1594	196	5.0E-07	6.1E-18	0.11		
Cthe_2694O-antigen polymerase	1261	659	2026	35	284	1558	362	23	6168	213	2351	116	2.0E-03	1.2E-03	0.20		
Cthe_2706ABC transporter-like protein	440	301	1063	23	32	447	198	12	5048	203	1951	140	1.4E-10	7.3E-27	0.09		
Cthe_2707ABC-type transport system involved in multi-copper enzyme maturation permease component	261	116	739	21	36	224	171	14	2623	108	1461	123	2.1E-07	1.3E-15	0.11		
Cthe_2708hypothetical protein	543	302	1639	24	41	437	223	10	5865	164	1907	118	5.6E-07	1.1E-26	0.08		
Cthe_2709hypothetical protein	579	328	1528	23	38	638	225	12	5875	191	2188	119	5.5E-08	2.6E-20	0.10		
Cthe_2711hypothetical protein	688	507	1524	32	132	591	98	11	1877	61	1022	52	3.7E-01	4.1E-04	0.21		

Cthe_2712hypothetical protein	374	204	743	26	23	321	33	7	1366	44	424	54	2.4E-01	1.9E-10	0.13	
Cthe_2713dihydroxy-acid dehydratase	1135	489	1995	30	200	1194	194	15	4466	336	1362	86	1.6E-02	1.3E-05	0.17	
Cthe_2714acetolactate synthase, large subunit		811	464	1723	25	118	905	124	10	3986	154	1059	68	3.2E-02	1.7E-07	0.15
Cthe_271750S ribosomal protein L33	161	107	264	50	0	146	47	25	944	8	73	141	5.3E-02	1.9E-08	0.18	
Cthe_2718preprotein translocase subunit SecE		350	242	577	68	20	325	78	30	2186	77	554	252	4.8E-03	3.4E-13	0.12
Cthe_2719transcription antitermination protein nusG		669	524	990	58	80	700	164	30	4576	123	1482	245	5.3E-05	2.5E-12	0.12
Cthe_272050S ribosomal protein L11	1167	1170	1652	134	147	1038	227	56	12778	339	2338	779	2.8E-11	4.5E-43	0.07	
Cthe_272150S ribosomal protein L1	1983	1455	2218	116	178	1702	779	80	19791	610	4825	781	5.0E-17	3.3E-15	0.10	
Cthe_272250S ribosomal protein L10P	1942	663	1314	103	101	1741	713	98	20681	525	6908	1104	1.9E-47	7.2E-22	0.09	
Cthe_272350S ribosomal protein L12P	1959	761	1341	148	186	2255	1207	208	19517	1860	16885	2222	5.6E-106	2.8E-18	0.09	
Cthe_2724DNA-directed RNA polymerase subunit beta 0.10			2549	2061	5216	37	165	1550	734	13	16114	757	6003	134	7.9E-03	3.4E-15
Cthe_2725DNA-directed RNA polymerase subunit beta' 0.09			2211	1774	4124	33	210	1662	845	16	18248	1067	7219	170	2.2E-06	1.0E-16
Cthe_272650S ribosomal protein L7AE	447	369	417	74	126	340	192	59	4146	312	3110	701	5.2E-25	2.2E-32	0.08	
Cthe_272730S ribosomal protein S12	970	869	1133	99	238	749	165	45	10824	651	3710	802	2.9E-23	1.2E-77	0.06	
Cthe_272830S ribosomal protein S7	577	681	1084	71	15	553	122	25	10746	137	1288	539	2.8E-19	2.1E-130	0.05	
Cthe_2729elongation factor G	5045	3424	4675	90	778	4333	2970	91	66779	2552	23888	963	1.7E-22	5.1E-10	0.09	
Cthe_2730elongation factor Tu	20922	9235	11143	490	3995	21558	9544	602	151034	8809	93229	4606	6.2E-14	2.3E-05	0.13	
Cthe_2736phosphoenolpyruvate--protein phosphotransferase 4.39			6705	3887	3809	121	339	21029	417	158	1563	226	647	36	4.8E-02	2.8E-03
Cthe_2737excinuclease ABC subunit C	1939	1210	2860	45	150	7874	331	58	722	107	355	16	1.0E+00	5.8E-03	3.63	
Cthe_2743putative DNA-binding/iron metalloprotein/AP endonuclease 1.7E-02 0.24				915	363	1475	38	534	1511	585	50	3797	698	4168	209	1.7E-09
Cthe_2746hypothetical protein	278	142	432	25	60	304	90	17	807	100	671	79	4.0E-02	1.2E-03	0.22	
Cthe_2748SsrA-binding protein	1072	658	961	82	79	1091	219	50	2656	288	2587	271	3.7E-03	4.7E-05	0.18	

Cthe_2759AraC family transcriptional regulator	385	188	759	21	69	506	151	15	1132	94	1024	56	8.2E-02	1.6E-02	0.27	
Cthe_2764TROVE domain-containing protein	893	661	1301	27	94	733	111	10	2871	184	1541	70	5.5E-02	3.1E-09	0.14	
Cthe_2815lysyl-tRNA synthetase	1720	1183	2347	46	144	1900	301	23	11930	277	4002	212	8.8E-07	1.6E-14	0.11	
Cthe_2854hypothetical protein	4961	2511	6682	108	213	17585	3041	184	1788	237	1532	45	3.1E-01	1.2E-02	4.09	
Cthe_2870hypothetical protein	430	96	1168	18	43	324	120	7	3469	122	1213	79	7.0E-05	5.7E-27	0.09	
Cthe_2872glycoside hydrolase family protein	2520	2755	4043	78	385	2858	266	30	10316	225	1996	156	3.5E-01	6.5E-04	0.19	
Cthe_2874phosphoenolpyruvate carboxykinase	12032	5249	7805	196	1662	9301	1682	114	65788	2218	21507	1058	1.2E-04	2.9E-07	0.11	
Cthe_2877S-layer-like domain-containing protein	1488	1320	2433	42	123	2454	582	31	13181	613	4995	234	1.9E-10	2.7E-08	0.13	
Cthe_2884imidazoleglycerol-phosphate dehydratase	655	435	636	42	110	430	119	20	2686	136	1790	168	3.5E-04	7.8E-15	0.12	
Cthe_2885phosphoribosylaminoimidazole-succinocarboxamide synthase	1414	717	1220	54	353	1298	477	46	9082	782	8268	456	6.2E-27	6.1E-17	0.10	
Cthe_2886imidazole glycerol phosphate synthase subunit HisH	312	129	662	25	39	246	77	11	1764	46	624	83	1.9E-02	1.5E-10	0.13	
Cthe_2891co-chaperonin GroES	273	139	482	44	8	287	67	22	4710	62	861	410	2.5E-21	5.8E-87	0.05	
Cthe_2892chaperonin GroEL	3941	1998	5815	102	495	6476	5624	200	62220	3706	39040	1415	3.9E-53	1.4E-03	0.14	
Cthe_2895glycoside hydrolase family protein	719	467	1651	23	190	597	158	9	2738	165	1682	57	5.6E-02	1.2E-06	0.16	
Cthe_2897transcription elongation factor GreA	5998	5060	3900	451	190	4710	797	193	15150	927	7004	1076	4.0E-01	7.7E-04	0.18	
Cthe_2898anti-sigma-factor antagonist	346	267	445	48	21	297	62	20	982	74	712	125	1.1E-01	1.1E-06	0.16	
Cthe_2899putative anti-sigma regulatory factor, serine/threonine protein kinase	350	310	516	43	1	280	48	13	1025	28	339	75	4.0E-01	3.4E-06	0.17	
Cthe_290230S ribosomal protein S10	379	230	628	56	37	253	159	33	4012	211	3239	510	2.1E-22	1.2E-61	0.06	
Cthe_290350S ribosomal protein L3	395	263	987	36	13	315	79	11	4723	95	1151	196	1.1E-07	2.0E-73	0.06	
Cthe_290450S ribosomal protein L4	377	246	932	35	8	241	64	9	3202	101	1060	148	1.8E-04	3.1E-71	0.06	
Cthe_290550S ribosomal protein L23	252	152	526	37	5	188	51	12	2018	94	872	184	3.9E-05	1.3E-68	0.07	
Cthe_290650S ribosomal protein L2	441	292	1193	33	18	302	66	8	3074	81	819	102	1.2E-02	3.7E-35	0.08	
Cthe_290730S ribosomal protein S19	140	99	463	34	4	88	21	7	755	22	200	73	2.8E-01	2.7E-46	0.10	

Cthe_290850S ribosomal protein L22	173	122	518	30	7	164	73	13	1125	60	528	99	2.5E-02	6.8E-12	0.13	
Cthe_290930S ribosomal protein S3P	369	284	944	33	22	324	92	11	2102	97	1030	102	1.9E-02	2.6E-15	0.11	
Cthe_291050S ribosomal protein L16	301	199	704	39	18	263	79	15	1673	78	932	132	1.2E-02	4.5E-14	0.11	
Cthe_291150S ribosomal protein L29	144	98	276	36	8	107	29	13	791	32	360	124	7.2E-03	1.2E-33	0.10	
Cthe_291230S ribosomal protein S17	180	113	373	36	13	165	68	19	1122	57	799	164	2.6E-04	1.3E-16	0.12	
Cthe_291350S ribosomal protein L14	237	149	482	33	19	245	84	18	1569	75	776	142	7.9E-04	2.1E-11	0.13	
Cthe_291450S ribosomal protein L24	177	125	383	27	7	179	48	12	1335	54	357	109	2.7E-03	1.8E-22	0.11	
Cthe_291550S ribosomal protein L5	429	340	703	38	26	457	187	25	3305	203	1714	211	9.3E-08	1.2E-14	0.12	
Cthe_291630S ribosomal protein S14	155	118	234	39	12	129	73	26	1210	97	788	255	3.9E-11	1.7E-25	0.10	
Cthe_291730S ribosomal protein S8	274	232	568	38	15	292	142	24	2528	119	1409	219	2.0E-07	7.8E-18	0.11	
Cthe_291850S ribosomal protein L6P	232	241	670	29	14	299	102	14	2787	91	935	148	1.1E-05	9.3E-22	0.09	
Cthe_291950S ribosomal protein L18P	189	139	566	34	19	168	79	15	1856	41	532	138	2.1E-03	1.0E-17	0.11	
Cthe_292030S ribosomal protein S5	429	236	722	39	35	426	139	23	3558	105	1528	218	1.2E-07	5.8E-20	0.11	
Cthe_292150S ribosomal protein L30	192	144	298	50	26	189	68	31	1820	82	837	330	6.7E-10	5.3E-27	0.09	
Cthe_292250S ribosomal protein L15	416	298	730	46	43	467	177	31	4520	221	2081	338	2.5E-14	1.3E-26	0.09	
Cthe_2923protein translocase subunit secY/sec61 alpha 0.12			1146	866	1695	41	312	1365	686	39	12461	607	6362	329	5.7E-24	2.9E-11
Cthe_2924adenylate kinase	531	449	988	43	43	639	318	33	6578	336	3163	337	6.7E-20	4.5E-21	0.10	
Cthe_2925methionine aminopeptidase, type I		1300	967	1362	67	364	1348	1729	124	8625	1223	6376	515	1.1E-21	4.6E-02	0.24
Cthe_2926hypothetical protein	82	84	361	27	4	88	33	9	1048	24	198	99	4.0E-03	2.0E-45	0.09	
Cthe_2927translation initiation factor IF-1	155	102	298	36	26	203	170	46	1429	176	1323	312	4.8E-20	2.4E-09	0.15	
Cthe_292850S ribosomal protein L36P	99	51	130	35	11	194	382	165	1679	313	2595	953	0.0E+00	7.0E-06	0.17	
Cthe_292930S ribosomal protein S13	240	212	566	39	9	195	38	11	1697	52	510	130	1.9E-02	3.7E-45	0.08	
Cthe_293030S ribosomal protein S11	275	250	666	42	43	227	100	18	2501	81	914	184	2.3E-04	5.7E-20	0.10	
Cthe_293130S ribosomal protein S4P	722	623	1095	55	93	779	291	36	6627	336	2948	347	1.5E-12	5.0E-18	0.10	

Cthe_2932DNA-directed RNA polymerase subunit alpha 0.08				1163	1031	2102	64	39	1074	351	29	13953	204	2190	360	3.1E-10	1.6E-30
Cthe_293350S ribosomal protein L17		2670	1175	1662	149	323	2849	2486	276	21219	2406	21585	1970	3.1E-82	1.3E-05	0.14	
Cthe_2938glucokinase	850	504	1288	39	71	867	130	17	3841	159	1159	118	1.1E-02	2.2E-08	0.14		
Cthe_2947prolyl-tRNA synthetase		868	430	1816	25	121	733	201	11	3366	215	1113	62	6.4E-02	1.4E-05	0.18	
Cthe_2951hypothetical protein	249	69	881	23	13	597	1694	110	196	35	298	17	1.0E+00	6.8E-12	6.47		
Cthe_2989glycosyltransferase	4864	3650	4679	64	753	4640	1905	50	68357	2419	27706	713	3.8E-24	7.0E-27	0.07		
Cthe_2996PfpI family intracellular peptidase			1295	704	1259	80	1085	1277	503	91	5680	1339	10265	723	4.5E-31	1.1E-09	0.13
Cthe_2998ABC transporter-like protein		743	308	1892	24	149	723	218	11	6711	253	2047	113	8.2E-07	3.1E-19	0.10	
Cthe_3000phosphate transporter		1232	801	1544	48	467	979	288	29	8838	634	5700	321	1.4E-15	1.3E-23	0.09	
Cthe_3001hypothetical protein	775	505	940	50	11	753	77	19	4098	59	523	157	9.6E-03	3.6E-13	0.12		
Cthe_3002hypothetical protein	652	368	757	46	12	507	113	20	1970	154	1024	131	3.3E-02	1.3E-08	0.15		
Cthe_3003hydrogenase, Fe-only	1522	1043	2327	36	270	1797	420	22	12766	362	3278	181	2.7E-08	8.3E-11	0.12		
Cthe_3004ferredoxin 1136	868	1785	36	189	1399	220	19	10256	224	2227	181	4.9E-08	9.4E-17	0.10			
Cthe_3013hydrogenase expression/formation protein HypE 0.15				309	109	1002	20	50	263	104	8	1592	119	655	55	6.2E-02	2.7E-08
Cthe_3017hydrogenase accessory protein HypB			392	216	500	24	68	302	151	16	1707	218	1122	113	6.8E-05	5.2E-09	0.14
Cthe_3018hydrogenase expression/synthesis, HypA			137	82	227	18	20	104	27	8	723	24	245	62	5.7E-03	7.0E-21	0.13
Cthe_30194Fe-4S ferredoxin, iron-sulfur binding			181	109	297	21	25	149	34	9	736	27	314	59	7.1E-02	2.6E-11	0.15
Cthe_3020ech hydrogenase subunit E		663	497	1108	30	75	430	131	11	2716	125	999	78	4.7E-02	2.7E-10	0.14	
Cthe_3021ech hydrogenase subunit D		424	275	410	44	57	206	93	20	1272	142	742	143	2.1E-02	2.4E-09	0.14	
Cthe_3022ech hydrogenase subunit C		130	78	366	18	16	66	16	4	869	29	216	55	4.0E-02	5.7E-96	0.07	
Cthe_3023ech hydrogenase subunit B		426	177	1048	26	70	266	103	9	2677	126	1022	96	2.6E-03	7.4E-21	0.09	
Cthe_3024ech hydrogenase subunit A		1715	869	1701	31	661	1023	925	32	12203	1095	12297	297	4.3E-34	2.6E-13	0.11	
Cthe_3026GreA/GreB family elongation factor			285	379	589	37	17	265	42	10	993	27	348	60	4.7E-01	1.3E-06	0.17

Cthe_3027citrate synthase	6523	4836	5247	174	1443	7141	1216	115	25835	1545	13176	656	3.8E-02	4.2E-03	0.18		
Cthe_3035D-3-phosphoglycerate dehydrogenase		2505	2546	3880	109	213	2737	426	45	13680	411	3834	326	4.3E-02	1.1E-07	0.14	
Cthe_3048DNA repair protein RadC		782	497	1399	57	146	771	154	26	1457	243	1228	109	2.3E-01	7.1E-03	0.24	
Cthe_3058hypothetical protein	219	171	1232	47	18	1455	5401	519	510	98	1432	94	2.7E-01	2.1E-06	5.52		
Cthe_3064polysaccharide biosynthesis protein		555	308	1245	18	136	735	178	11	2675	148	1483	58	8.2E-03	1.3E-04	0.19	
Cthe_3072acyl-ACP thioesterase	360	205	721	23	17	276	31	6	1927	36	541	68	5.1E-02	1.1E-31	0.09		
Cthe_3075von Willebrand factor, type A		716	673	1688	27	92	793	508	20	15796	900	6395	320	2.5E-57	2.0E-55	0.06	
Cthe_3082hypothetical protein	79	93	242	24	6	87	50	13	534	49	214	77	7.7E-03	2.1E-09	0.17		
Cthe_3093adenylosuccinate synthetase		1965	2016	2592	74	140	1847	276	28	7433	238	2212	166	9.9E-02	1.2E-05	0.17	
Cthe_3096hypothetical protein	996	1888	2098	77	49	886	123	17	3473	108	965	104	5.0E-01	2.1E-06	0.16		
Cthe_3105exsB protein	627	269	668	33	57	497	137	18	1944	124	969	101	2.1E-02	1.6E-05	0.18		
Cthe_3106putative 6-pyruvoyl tetrahydropterin synthase				418	206	427	39	62	391	105	26	1397	102	778	134	1.3E-02	5.0E-05
0.19																	
Cthe_3110UBA/THIF-type NAD/FAD binding fold protein				615	233	1004	34	121	623	168	21	1565	296	1222	104	1.6E-02	3.8E-04
0.20																	
Cthe_3112glycosidase	840	652	1706	45	223	943	348	29	2374	251	3406	133	1.4E-02	2.7E-03	0.22		
Cthe_3113nucleotidyl transferase		488	412	1566	33	70	550	250	17	1953	173	1640	80	7.2E-02	1.0E-03	0.21	
Cthe_3118hemerythrin-like metal-binding protein			561	426	389	50	19	437	138	29	1031	125	721	115	1.8E-01	6.3E-03	0.25
Cthe_3120pyruvate flavodoxin/ferredoxin oxidoreductase-like protein					4204	2873	4687	47	949	5095	2129	46	19431	2697	9456	226	1.6E-04
4.2E-02	0.20																
Cthe_3124AMP-dependent synthetase and ligase			2968	2426	3720	79	370	3089	494	38	20533	569	4991	341	4.3E-04	4.5E-12	0.11
Cthe_3134hypothetical protein	205	148	312	47	30	552	197	77	54	17	133	25	1.0E+00	1.2E-03	3.08		
Cthe_3136peptidase S8/S53 subtilisin kexin sedolisin			334	280	1088	21	13	336	145	9	2292	73	647	57	5.5E-02	9.9E-08	0.16
Cthe_3137hypothetical protein	188	164	630	23	1	191	187	17	1681	19	224	68	6.7E-02	7.1E-03	0.25		
Cthe_3142hypothetical protein	260	164	462	25	25	322	189	25	2159	333	3270	271	6.7E-29	9.4E-27	0.09		
Cthe_3154hypothetical protein	150	56	494	17	19	224	55	9	1373	38	297	66	5.8E-03	1.8E-10	0.14		

Cthe_3155beta-lactamase-like protein	524	176	919	31	50	927	279	32	3841	122	1250	152	3.5E-06	8.0E-04	0.21	
Cthe_3157pyruvate carboxyltransferase	633	623	1698	30	83	758	159	12	3561	140	1481	81	3.5E-02	5.8E-08	0.15	
Cthe_3158aconitate hydratase	1609	1039	2931	41	528	2016	744	33	13303	852	6276	238	5.8E-12	1.3E-07	0.14	
Cthe_3159GntR family transcriptional regulator		918	583	1446	62	123	1330	471	56	6638	588	4067	387	1.4E-12	9.7E-08	0.14
Cthe_3169enoyl-[acyl-carrier-protein] reductase [NADH] 0.13			834	378	1165	44	136	1033	216	31	5713	197	2901	245	1.9E-09	5.3E-12
Cthe_3171S-layer-like domain-containing protein		535	140	2143	16	3	377	232	5	4212	200	2053	57	1.0E-03	2.7E-20	0.09
Cthe_3172hypothetical protein	461	177	1479	21	95	636	467	20	4763	634	4118	160	4.9E-19	4.0E-11	0.13	
Cthe_3177monogalactosyldiacylglycerol synthase		726	895	1800	38	44	636	129	10	9507	149	1713	187	1.6E-07	4.9E-80	0.05
Cthe_3178hypothetical protein	52741	44748	19710	1072	16022	85227	1370	822	1126	322	1740	51	7.3E-111	1.1E-58	16.12	
Cthe_3188copper amine oxidase-like protein		1212	648	1346	30	92	1092	149	13	3340	111	556	58	2.2E-01	4.3E-03	0.22
Cthe_3189hypothetical protein	343	319	665	30	71	389	84	15	1170	95	841	76	1.1E-01	5.4E-05	0.20	
Cthe_3191chromosome partitioning ATPase		1865	1281	2002	60	386	1873	333	34	6281	291	3617	179	8.1E-03	4.8E-04	0.19
Cthe_3194hypothetical protein	2220	1970	2426	146	939	4176	152	107	350	52	224	24	1.3E-10	2.9E-05	4.46	
Cthe_3196hypothetical protein	681	292	1298	23	85	491	120	9	6263	213	2586	141	3.0E-11	1.2E-65	0.06	
Cthe_3200alanyl-tRNA synthetase	952	361	2904	22	163	881	260	9	5073	331	2463	66	9.4E-03	3.9E-10	0.14	
Cthe_3306hypothetical protein	36	26	65	16	6	42	26	15	180	9	133	63	5.8E-07	6.0E-06	0.24	
Cthe_3321hypothetical protein	239	135	215	39	12	200	59	23	918	45	347	134	1.2E-02	4.4E-07	0.17	
Cthe_3326sodium pump decarboxylase gamma subunit 0.07			100	225	276	51	4	100	12	10	994	10	170	141	7.5E-02	1.9E-94
Cthe_3328hypothetical protein	745	1041	361	250	1261	925	212	311	106	27	114	52	2.5E-07	9.3E-12	5.98	
Cthe_3332helicase-like protein	100	75	104	28	0	111	18	14	473	26	179	105	5.5E-05	7.9E-21	0.13	
Cthe_3333hypothetical protein	1494	1010	699	302	90	986	265	159	4059	585	5707	1552	1.9E-08	2.4E-19	0.10	
Cthe_3337hypothetical protein	88	136	375	59	3	270	622	208	85	11	73	28	4.8E-02	5.7E-28	7.43	
Cthe_3345hypothetical protein	222	221	337	78	35	1364	42	128	78	17	45	27	1.7E-03	4.3E-10	4.74	

Cthe_3376	hypothetical protein	85	34	157	27	11	87	48	23	156	28	216	68	3.0E-02	2.8E-02	0.34
Cthe_3380	hypothetical protein	292	122	341	62	107	515	1180	341	3151	637	6668	1410	2.3E-169	2.2E-02	0.24
Cthe_3419	hypothetical protein	137	134	186	41	22	101	84	32	1397	129	1627	436	2.4E-36	1.1E-50	0.07

Apêndice F.B. Tabela dos genes diferencialmente expressos de *C.thermocellum* em bagaço em comparação com celulose. Na tabela estão presentes os genes, descrição, número de fragmentos de cada triplicata biológica para bagaço (B1, B2, B3), palha (P1, P2, P3) e celulose (C1, C2, C3), números de expressão calculados e *q-value* para cada comparação de cada gene. FC, indica o número de vezes que o gene foi aumentado ou diminuído. Valores maiores que 1, os genes são regulados positivamente para a condição palha.

Gene	Descrição	B1	B2	B3	Expressão	P1	P2	P3	Expressão	C1	C2	C3	Expressão	qValue B vs C	qValue P vs C	FC	
Cthe_0019	P-II family regulatory protein	852	669	974	109	583	1466	625	162	1687	462	3004	394	8.49E-04	4.15E-01	0.28	
Cthe_0049	XRE family transcriptional regulator	354	138	652	24	40	623	234	27	1440	170	1593	110	1.42E-04	7.57E-03	0.22	
Cthe_0052	hypothetical protein	2321	1699	2554	74	686	2968	660	58	10884	519	3838	267	5.27E-04	8.34E-03	0.28	
Cthe_0053	ribonucleotide-diphosphate reductase subunit alpha	2500	3440	5011	66	333	2221	729	26	19283	1196	7534	265	2.54E-03	1.22E-16	0.25	
Cthe_0054	hypothetical protein	851	364	1961	20	166	878	253	10	4084	273	1484	61	1.04E-02	1.35E-05	0.33	
Cthe_0067	NAD-dependent deacetylase	2092	1135	1813	99	822	3302	734	114	6925	1453	6810	537	5.64E-10	5.57E-03	0.18	
Cthe_0068	peptidylprolyl isomerase	1192	690	1072	69	248	1349	360	57	6829	394	5264	440	1.50E-13	4.80E-10	0.16	
Cthe_0070	asparaginyl-tRNA synthetase	820	589	1598	30	142	1006	314	19	6923	366	3430	168	1.22E-09	7.28E-14	0.18	
Cthe_0072	phage shock protein PspC	442	140	489	31	119	1112	119	41	1539	116	1187	129	8.82E-04	8.82E-02	0.24	
Cthe_0073	hypothetical protein	919	416	1293	38	301	2795	231	49	3227	236	1946	125	3.83E-03	3.46E-01	0.30	
Cthe_0077	hypothetical protein	1626	1075	1442	154	82	2689	641	157	3014	513	3603	446	1.38E-02	2.69E-01	0.35	
Cthe_0079	hypothetical protein	687	672	1246	72	106	870	104	31	75	12	62	7	6.39E-60	1.00E+00	10.29	
Cthe_0080	CheW protein	1663	1290	1262	140	233	1438	217	69	227	37	203	26	1.80E-07	9.00E-02	5.38	
Cthe_0081	N-acetylglutamate synthase / glutamate N-acetyltransferase	848	603	1413	33	165	1009	196	18	8623	172	2209	191	3.52E-10	3.13E-21	0.17	
Cthe_0084	hypothetical protein	1017	1548	1681	113	29	1388	1018	110	299	69	439	36	1.75E-02	3.12E-02	3.14	

Cthe_0093	septum site-determining protein MinD	1103	1061	1516	66	137	793	139	21	4869	123	1628	174	3.09E-02	1.80E-12	0.38
Cthe_0094	cell division topological specificity factor MinE	464	396	556	71	57	520	83	37	2001	103	847	228	1.63E-02	1.63E-07	0.31
Cthe_0095	methylglyoxal synthase	525	436	638	58	100	637	222	47	2460	201	1849	255	5.95E-05	3.10E-05	0.23
Cthe_0100	hypothetical protein	530	511	892	35	32	447	171	16	2288	102	1053	95	4.46E-02	3.54E-06	0.37
Cthe_0101	iron-containing alcohol dehydrogenase	1366	1081	1768	51	196	1814	564	41	8086	673	6465	291	1.31E-10	1.44E-07	0.18
Cthe_0102	tRNA (guanine-N(7)-)-methyltransferase	296	157	705	24	15	276	140	14	2532	45	607	101	1.10E-03	3.59E-10	0.24
Cthe_0103	hypothetical protein	227	77	558	30	25	226	78	16	726	72	1021	101	2.49E-02	3.48E-07	0.30
Cthe_0107	6,7-dimethyl-8-ribityllumazine synthase	1184	687	1134	91	310	1557	532	98	2163	856	2581	352	2.13E-04	6.34E-02	0.26
Cthe_0108	type IV pilus assembly PilZ	3810	4364	3844	271	98	2965	213	71	1173	89	824	72	7.34E-03	1.00E+00	3.76
Cthe_0116	hypothetical protein	594	217	635	21	22	457	137	12	2134	149	1662	91	2.40E-04	1.83E-11	0.23
Cthe_0117	HPrNtr domain-containing protein	357	120	207	35	28	290	218	48	1489	290	2446	366	6.90E-28	1.23E-11	0.10
Cthe_0130	3-oxoacyl-(acyl-carrier-protein) synthase	1777	1639	2095	76	461	2506	233	45	545	61	436	23	1.09E-02	1.00E+00	3.30
Cthe_0137	glyceraldehyde-3-phosphate dehydrogenase	23590	9463	8178	583	4783	19364	5274	521	216355	12028	113322	7434	2.39E-28	4.89E-22	0.08
Cthe_0138	phosphoglycerate kinase	3138	1881	2849	94	180	3491	807	65	43029	566	6923	881	8.70E-26	4.19E-31	0.11
Cthe_0139	triosephosphate isomerase	4330	1221	2131	144	430	4185	1510	159	30041	1720	19320	1473	2.47E-33	2.92E-10	0.10
Cthe_0143	phosphopyruvate hydratase	6040	2946	3229	134	596	5217	928	84	27986	956	6698	595	1.15E-03	1.26E-05	0.23
Cthe_0144	preprotein translocase subunit SecG	2876	1990	1595	370	1178	2278	1969	513	18633	3529	31923	5029	4.37E-87	6.78E-12	0.07
Cthe_0145	metal dependent phosphohydrolase	2315	1891	2684	179	227	4407	2125	265	5776	624	5853	504	2.35E-02	1.00E+00	0.36
Cthe_0149	HPrNtr domain-containing protein	3456	2785	3471	542	580	3983	3899	865	12064	1752	18119	2822	3.81E-06	4.47E-01	0.19
Cthe_0151	hypothetical protein	2243	1409	1485	270	906	2256	882	286	4270	1105	7769	1187	2.88E-06	2.58E-02	0.23
Cthe_0154	hypothetical protein	232	176	448	42	8	607	104	40	1347	193	1201	226	2.82E-06	4.52E-06	0.19
Cthe_0160	50S ribosomal protein L21	1814	941	902	168	670	1763	987	235	16307	2428	18675	2829	4.10E-139	2.76E-24	0.06
Cthe_0161	hypothetical protein	1964	1016	1001	167	178	1934	586	148	17286	2023	16554	2424	1.88E-98	6.15E-53	0.07

Cthe_0162	50S ribosomal protein L27	2805	1052	1371	267	436	3157	2666	567	22210	4627	28894	4941	1.87E-180	7.02E-08	0.05	
Cthe_0163	GTP1/OBG subdomain-containing protein	891	511	1781	35	144	2585	318	36	4429	180	1605	106	9.95E-03	2.14E-01	0.33	
Cthe_0174	sulfatase	1116	698	1921	28	134	952	332	14	7071	245	3099	117	1.66E-05	4.90E-12	0.24	
Cthe_0178	argininosuccinate lyase	558	295	1493	24	61	460	146	9	5643	93	903	101	3.90E-05	4.26E-28	0.24	
Cthe_0179	argininosuccinate synthase	351	149	1025	17	18	230	71	4	3937	39	769	80	3.97E-05	1.18E-67	0.21	
Cthe_0186	UDP-galactose 4-epimerase	468	273	924	22	61	505	98	10	2231	77	1035	66	2.27E-02	3.70E-08	0.33	
Cthe_0189	RdgB/HAM1 family non-canonical purine NTP pyrophosphatase	785	339	942	48	279	1731	428	71	2403	287	2015	184	6.07E-04	3.30E-01	0.26	
Cthe_0214	phenylalanyl-tRNA synthetase subunit alpha	675	269	1120	28	24	374	105	9	3246	61	1051	88	1.06E-02	2.69E-21	0.32	
Cthe_0215	phenylalanyl-tRNA synthetase subunit beta	1056	482	2135	21	102	880	332	10	8783	192	2844	103	2.38E-07	9.87E-18	0.20	
Cthe_0217	glucose-6-phosphate isomerase	1512	843	1923	45	113	1337	349	24	15058	326	4469	309	1.62E-16	4.38E-34	0.15	
Cthe_0218	metallophosphoesterase	447	216	703	24	32	389	80	10	1815	69	868	72	2.95E-02	2.07E-09	0.33	
Cthe_0220	transposase, IS4	1065	599	1691	50	616	1243	1446	90	5054	754	5626	285	1.10E-10	1.99E-01	0.18	
Cthe_0234	AMP-dependent synthetase and ligase	900	341	2249	19	147	734	357	10	6677	310	2587	82	9.97E-06	5.55E-12	0.23	
Cthe_0235	YruB family glutaredoxin-like protein	2514	934	656	254	736	2243	868	324	1558	2459	4051	1370	1.19E-09	2.08E-02	0.19	
Cthe_0237	response regulator receiver protein	907	594	902	87	486	1719	1838	271	2521	685	3870	460	1.73E-08	1.00E+00	0.19	
Cthe_0244	heavy metal translocating P-type ATPase	3198	8296	3818	106	1489	7401	313	58	1107	140	401	19	2.31E-06	7.60E-02	5.58	
Cthe_0245	ArsR family transcriptional regulator	4775	4524	1681	424	128	1751	100	71	267	163	360	68	1.25E-09	9.77E-01	6.24	
Cthe_0248	transposase, IS4	1089	602	1732	51	617	1257	1448	91	5114	756	5638	287	2.08E-10	1.98E-01	0.18	
Cthe_0262	gamma-glutamyl phosphate reductase	2171	2335	2756	81	777	3558	345	52	9545	399	3725	229	3.61E-02	1.52E-02	0.35	
Cthe_0274	glycoside hydrolase family protein	2269	2369	3405	68	369	2407	139	23	752	54	402	16	2.82E-04	1.00E+00	4.25	
Cthe_0275	cellobiose phosphorylase	2628	2004	3326	46	309	2545	680	26	12077	545	4828	156	1.20E-02	5.07E-05	0.29	
Cthe_0285	isocitrate dehydrogenase	3304	2232	3092	102	553	3654	877	73	31764	908	11729	777	4.51E-15	8.96E-17	0.13	
Cthe_0290	homoserine dehydrogenase	1108	342	1570	33	163	1151	608	32	5462	535	3371	170	3.04E-08	4.72E-04	0.19	

Cthe_0295phosphoserine aminotransferase	3380	1942	2960	105	263	3044	450	52	18831	742	5672	492	4.70E-05	6.79E-14	0.21		
Cthe_0303hypothetical protein	1501	2062	1464	164	18	634	408	56	108	40	300	24	2.96E-12	1.88E-01	6.83		
Cthe_0307hypothetical protein	3124	1145	1305	210	687	2678	550	173	7763	523	9812	1013	6.56E-08	9.17E-05	0.21		
Cthe_0308hypothetical protein	2415	953	1144	189	650	2524	610	194	5816	565	8097	930	9.72E-08	3.86E-03	0.20		
Cthe_0317hypothetical protein	557	122	399	22	53	531	90	16	2083	119	2256	140	7.66E-09	5.28E-16	0.16		
Cthe_0323hypothetical protein	105	60	199	21	6	44	24	6	628	34	221	81	2.42E-04	3.84E-70	0.26		
Cthe_0324valyl-tRNA synthetase	1359	1024	2881	28	142	1017	491	13	16619	365	3260	161	1.59E-11	1.76E-29	0.17		
Cthe_0325NAD synthetase	1047	798	2360	30	211	941	410	16	11443	460	3202	170	3.72E-10	7.50E-21	0.18		
Cthe_0328peptide chain release factor 3	1070	342	1673	27	128	829	177	12	8884	299	2728	160	2.44E-11	9.60E-40	0.17		
Cthe_0338NADH-quinone oxidoreductase subunit E	1258	1007	1076	96	72	810	202	38	6422	297	3295	435	2.27E-06	7.59E-27	0.22		
Cthe_0339histidine kinase	1112	1100	1340	90	77	858	167	32	6043	130	1780	294	2.80E-03	2.28E-16	0.31		
Cthe_0340ferredoxin 428	481	819	67	27	415	108	27	2889	72	648	208	1.51E-02	1.29E-12	0.32			
Cthe_0341NADH dehydrogenase (quinone)	2446	2478	4116	72	264	1930	576	28	18539	468	3869	272	3.49E-03	8.47E-15	0.26		
Cthe_0342hydrogenase, Fe-only	5867	3880	4857	119	759	3855	1223	62	30833	1555	11242	553	4.45E-03	1.88E-09	0.22		
Cthe_0343FMN-binding flavin reductase-like protein	355	207	570	26	50	380	149	19	2129	202	2146	166	4.12E-09	8.20E-16	0.16		
Cthe_0344malate dehydrogenase	7424	5677	5768	231	878	4940	1351	110	66350	2108	22472	1657	1.89E-08	6.07E-32	0.14		
Cthe_0345malate dehydrogenase (NAD)	3749	2900	3088	146	249	2200	880	70	39498	1046	14710	1216	9.32E-18	8.97E-57	0.12		
Cthe_0346hypothetical protein	431	416	546	56	118	425	219	45	1099	152	1571	181	1.68E-02	9.27E-03	0.31		
Cthe_03476-phosphofructokinase	3029	2789	3833	111	382	2895	1309	77	52039	690	8177	1016	2.41E-22	7.25E-26	0.11		
Cthe_0349fructose-bisphosphate aldolase	11480	5492	6754	364	2440	13509	3703	382	130986	4438	60243	4466	8.26E-30	3.02E-11	0.08		
Cthe_0350signal peptidase I	284	92	561	22	6	156	72	8	1603	100	858	98	6.58E-04	2.17E-34	0.22		
Cthe_0365peptide chain release factor 2	2179	1639	2341	82	195	2514	601	54	16335	667	7512	491	4.50E-13	6.10E-13	0.17		
Cthe_0373ferredoxin-NADP(+) reductase subunit alpha			280	180	798	21	35	189	52	5	2376	39	509	72	1.01E-02	5.72E-41	0.29
Cthe_0374glutamate dehydrogenase	5432	963	1901	87	409	3531	1208	74	20337	972	10873	522	8.01E-09	7.27E-06	0.17		

Cthe_0375	GMP synthase	2077	944	1737	44	361	1870	517	32	20256	663	6316	381	8.07E-29	1.44E-24	0.12		
Cthe_0387	transposase, IS4	1060	597	1682	50	616	1240	1444	90	5049	751	5621	285	8.84E-11	1.99E-01	0.18		
Cthe_0388	alcohol dehydrogenase GroES-like protein	6243	2440	2614	155	2376	26008	2464	433	31376	604	10011	845	2.89E-06	1.00E+00	0.18		
Cthe_0389	PfkB	2971	1003	2085	88	2182	20575	3588	442	17085	464	6225	516	1.72E-12	1.00E+00	0.17		
Cthe_0390	ROK domain-containing protein	3652	1781	4144	112	1421	47087	9350	805	15110	1552	11785	540	3.02E-05	1.00E+00	0.21		
Cthe_0392	inner-membrane translocator	1650	1476	2396	75	804	13785	1477	226	9937	195	3259	265	5.20E-04	1.00E+00	0.28		
Cthe_0393	sugar ABC transporter sugar-binding protein	1838	4021	4149	151	836	12474	2250	265	38339	504	6988	985	5.44E-10	2.98E-01	0.15		
Cthe_0394	iron-containing alcohol dehydrogenase	3268	4244	8855	200	1539	25071	25771	1219	57852	2158	13033	1369	3.15E-07	1.00E+00	0.15		
Cthe_0395	RbsD or FucU transport	5849	7060	6811	627	9990	68633	22037	4184	37513	6351	35834	4309	8.32E-08	1.00E+00	0.15		
Cthe_0398	hypothetical protein	137	90	312	23	19	283	40	16	742	78	814	114	1.88E-05	3.51E-10	0.20		
Cthe_0401	methyl-accepting chemotaxis sensory transducer	4889	4831	6625	103	227	9220	900	66	670	127	714	16	3.18E-09	1.80E-03	6.44		
Cthe_0409	hypothetical protein	143	137	260	28	19	185	53	17	958	24	316	101	6.19E-03	1.02E-07	0.28		
Cthe_0410	hypothetical protein	1566	1242	2484	55	408	2696	659	48	9573	547	3084	219	5.31E-05	9.12E-03	0.25		
Cthe_0417	30S ribosomal protein S15	3119	1942	1860	377	186	2376	814	249	8549	1768	13088	2142	4.08E-10	2.58E-11	0.18		
Cthe_0418	polynucleotide phosphorylase/polyadenylase	1504	1072	2761	36	204	1258	394	16	9947	306	3633	140	7.21E-05	9.98E-13	0.26		
Cthe_0435	dockerin type I cellulosome protein	1524	1166	1721	60	788	2012	310	46	3778	484	3558	174	1.31E-02	4.72E-02	0.34		
Cthe_0436	hypothetical protein	3237	2302	3696	41	265	2742	741	21	12813	579	5078	127	3.79E-02	7.01E-05	0.32		
Cthe_0437	hypothetical protein	815	499	697	62	68	928	367	60	3002	427	3090	336	2.48E-08	2.82E-05	0.18		
Cthe_0442	cell division protein FtsQ	827	557	922	38	157	1046	205	27	1861	189	1937	104	3.29E-02	1.70E-02	0.37		
Cthe_0458	exodeoxyribonuclease III Xth	2076	891	1187	78	368	1847	649	73	3331	808	3573	268	1.40E-03	6.30E-02	0.29		
Cthe_0459	DNA protecting protein DprA	782	371	1001	27	102	997	748	40	2084	259	1757	87	7.65E-03	5.21E-01	0.31		
Cthe_0475	flagellar protein	1193	955	709	142	131	942	82	57	198	64	298	51	3.32E-02	8.33E-01	2.78		
Cthe_0476	hypothetical protein	2918	2497	2192	218	348	3232	161	99	411	109	520	54	7.16E-04	4.36E-01	4.04		

Cthe_0479	response regulator receiver protein		697	581	776	81	17	839	42	33	174	47	134	27	2.84E-02	1.00E+00	3.00
Cthe_0499	hypothetical protein	336	332	610	58	3	631	17	26	54	2	12	4	1.67E-168	1.00E+00	14.50	
Cthe_0509	sodium ion-translocating decarboxylase subunit beta		1049	925	1786	49	167	988	325	26	9050	244	2483	232	5.54E-07	4.70E-15	0.21
Cthe_0533	radical SAM family protein	1980	2772	3002	73	97	1999	185	21	222	41	256	8	2.78E-25	1.00E+00	9.13	
Cthe_0549	hypothetical protein	547	290	1104	32	184	683	279	26	2092	241	2339	124	4.91E-04	1.04E-03	0.26	
Cthe_0558	hypothetical protein	1743	876	678	72	96	1665	68	36	29	26	67	6	3.92E-83	1.00E+00	12.00	
Cthe_0572	ribosomal RNA large subunit methyltransferase N		528	391	1085	27	48	609	162	14	2958	191	1946	107	2.92E-04	2.76E-11	0.25
Cthe_0573	protein serine/threonine phosphatases		477	451	1041	38	14	505	85	13	2934	107	1077	120	9.67E-03	4.37E-18	0.32
Cthe_0574	serine/threonine protein kinase	1519	905	2790	35	135	1690	473	20	8794	375	3085	126	3.74E-04	3.13E-06	0.28	
Cthe_0575	ribosome small subunit-dependent GTPase A		744	402	917	33	74	806	285	25	3291	300	1450	134	1.35E-04	9.56E-05	0.25
Cthe_0584	hypothetical protein	128	75	151	17	12	116	57	13	604	57	766	107	9.02E-13	4.32E-17	0.16	
Cthe_0598	thiazole synthase	236	46	734	18	21	331	202	16	1285	110	773	64	1.19E-02	9.17E-03	0.28	
Cthe_0603	GCN5-like N-acetyltransferase	1426	789	494	74	107	1865	54	49	164	24	135	15	1.94E-06	1.00E+00	4.93	
Cthe_0613	thiamine pyrophosphate enzyme-like TPP-binding protein		983	1236	2279	35	129	822	269	12	7795	145	1718	112	5.09E-03	4.68E-15	0.31
Cthe_0615	phenylacetate-CoA ligase	675	689	1696	33	130	985	471	25	5349	298	2277	135	6.70E-05	2.42E-04	0.24	
Cthe_0616	ACT domain-containing protein	245	261	652	38	28	352	145	25	1880	56	608	125	1.82E-02	7.91E-05	0.30	
Cthe_0617	recombination factor protein RarA		1079	848	2554	49	419	2265	1326	69	5467	469	3126	160	2.80E-03	5.33E-01	0.31
Cthe_0621	translation initiation factor 2B subunit I		929	854	1526	46	129	1133	260	25	5556	163	1666	153	2.28E-03	3.61E-06	0.30
Cthe_0622	methylthioadenosine phosphorylase		510	364	958	32	45	603	114	15	4285	80	854	136	9.86E-05	7.13E-16	0.24
Cthe_0626	hypothetical protein	629	314	1121	20	17	375	134	7	3317	76	1262	68	4.50E-03	3.72E-19	0.29	
Cthe_0627	hypothetical protein	650	592	1137	38	86	684	230	22	3251	184	1583	126	4.39E-03	6.48E-06	0.30	
Cthe_0628	hypothetical protein	506	489	1039	34	112	491	169	17	2828	155	1305	113	6.00E-03	7.41E-08	0.30	

Cthe_0629	type II secretion system protein E	931	854	1879	36	142	779	214	14	4401	229	1905	99	2.05E-02	3.75E-09	0.36
Cthe_0632	hypothetical protein	940	531	820	54	96	957	381	48	4730	268	2893	285	2.33E-08	5.93E-06	0.19
Cthe_0633	hypothetical protein	1006	558	1099	78	120	1418	2906	295	4760	852	5757	566	4.78E-18	1.00E+00	0.14
Cthe_0637	hypothetical protein	8270	4346	3089	745	2137	5528	3227	780	20457	3872	44710	5177	1.97E-07	1.33E-03	0.14
Cthe_0642	hypothetical protein	854	511	389	135	6	735	27	53	74	39	95	35	1.36E-04	1.00E+00	3.86
Cthe_0648	glutamyl-tRNA synthetase	1885	1186	2640	49	174	1510	2211	70	12489	588	6212	255	3.14E-09	1.60E-01	0.19
Cthe_0657	hypothetical protein	607	386	765	41	25	702	496	50	2698	150	3549	221	1.17E-07	3.40E-03	0.19
Cthe_0681	inosine 5-monophosphate dehydrogenase	1447	1133	2313	46	267	1596	470	28	30146	557	6308	520	1.61E-52	1.13E-69	0.09
Cthe_0683	diaminopimelate decarboxylase	1755	522	1366	39	102	774	162	13	7124	174	2030	150	1.44E-04	7.59E-27	0.26
Cthe_0686	tryptophanyl-tRNA synthetase	640	750	1125	36	93	415	103	10	3599	119	1213	106	1.50E-02	6.81E-22	0.34
Cthe_0694	spermidine synthase	518	258	678	24	48	467	242	20	8732	201	2964	301	1.78E-47	5.09E-54	0.08
Cthe_0695	agmatinase	690	437	1295	39	158	818	334	30	9794	291	2016	300	9.21E-20	1.14E-19	0.13
Cthe_0696	putative rRNA methylase	362	249	673	32	35	343	157	20	5191	137	1381	251	3.64E-16	1.72E-38	0.13
Cthe_0699	carboxyl transferase	1450	2250	2571	58	347	1256	288	20	9866	269	3589	186	2.87E-03	1.24E-14	0.31
Cthe_0700	biotin/lipoyl attachment protein	220	507	572	47	22	271	29	12	2470	37	472	155	1.45E-02	1.43E-50	0.30
Cthe_0704	hypothetical protein	488	1173	2035	82	71	791	126	24	582	32	254	29	4.18E-02	1.00E+00	2.83
Cthe_0714	4-hydroxy-3-methylbut-2-enyl diphosphate reductase/S1 RNA-binding domain-containing protein	3530	3494	5455	86	158	4218	1005	45	18156	975	7825	286	4.97E-02	8.57E-05	0.30
Cthe_0715	S-adenosylmethionine decarboxylase proenzyme	1763	653	852	124	739	1394	780	163	10194	1147	19194	1754	2.60E-90	3.14E-19	0.07
Cthe_0723	tyrosyl-tRNA synthetase	1916	1096	1700	54	128	1271	225	21	10069	323	2202	222	4.04E-05	2.16E-21	0.24
Cthe_0724	HisJ family histidinol phosphate phosphatase	356	132	514	17	2	272	85	8	1471	67	704	60	1.48E-02	4.18E-09	0.28
Cthe_0727	cell envelope-related transcriptional attenuator	1365	627	1690	56	129	1473	1080	70	9227	1214	12935	570	2.79E-42	8.87E-10	0.10
Cthe_0730	hypothetical protein	574	262	660	31	66	621	269	29	2386	74	1392	117	1.66E-03	8.06E-03	0.26

Cthe_0733	type IV pilus assembly PilZ	5458	4441	4938	325	58	4616	990	149	832	196	966	78	2.30E-03	5.13E-01	4.17	
Cthe_0735	cellulosome anchoring protein cohesin subunit			363	225	893	26	16	276	82	8	2942	90	1015	106	4.16E-04	8.10E-33	0.25
Cthe_0738	copper ion binding protein	1442	454	548	163	75	1471	211	129	1487	1124	2036	769	1.07E-06	4.26E-06	0.21	
Cthe_0741	adenylosuccinate lyase	1070	695	1767	35	85	1030	496	24	10961	261	4017	223	5.08E-14	2.41E-15	0.16	
Cthe_0743	MATE efflux family protein	757	432	1491	26	222	542	194	12	2288	293	1932	73	2.10E-02	2.75E-06	0.36	
Cthe_0747	extracellular solute-binding protein		739	180	828	23	62	490	226	15	4237	214	1951	131	5.50E-09	1.26E-14	0.18	
Cthe_0748	binding-protein-dependent transport systems inner membrane component					339	71	516	16	45	207	125	10	1846	107	831	78	8.97E-05 1.26E-10 0.21
Cthe_0749	binding-protein-dependent transport systems inner membrane component					377	60	634	18	52	237	79	8	2013	75	728	73	1.46E-03 8.71E-15 0.25
Cthe_0750	spermidine/putrescine ABC transporter ATPase subunit				475	92	890	19	12	326	104	8	3194	76	647	78	5.20E-04	7.68E-21 0.24
Cthe_0751	XRE family transcriptional regulator		295	90	415	20	98	224	106	16	2313	117	1478	156	6.98E-13	1.16E-20	0.13	
Cthe_0755	aspartate aminotransferase	2829	770	1254	57	392	1290	775	45	18421	888	10623	543	2.55E-35	1.14E-24	0.10		
Cthe_0765	50S ribosomal protein L19	2894	1161	1399	225	455	4209	3457	594	30376	7507	58220	6752	0.00E+00	1.82E-11	0.03		
Cthe_0766	tRNA (Guanine37-N1) methyltransferase		1058	417	940	46	136	787	309	33	3107	676	2299	219	8.53E-07	5.32E-08	0.21	
Cthe_0767	16S rRNA-processing protein RimM		221	91	602	24	2	163	86	10	1784	63	470	95	3.38E-03	1.56E-17	0.25	
Cthe_0768	nucleic acid binding protein	752	275	833	114	20	996	289	106	8018	754	4526	1346	3.14E-50	1.24E-35	0.08		
Cthe_0769	30S ribosomal protein S16	446	153	445	60	8	480	126	45	5149	254	1898	656	1.04E-32	1.97E-53	0.09		
Cthe_0772	peptidoglycan-binding LysM	895	673	953	125	212	1115	1068	216	1910	836	4183	650	2.80E-08	2.15E-01	0.19		
Cthe_0786	3-dehydroquinate synthase	725	648	1288	35	67	549	272	17	7331	423	3862	236	3.58E-15	1.62E-40	0.15		
Cthe_0787	isoleucyl-tRNA synthetase	1070	580	2655	21	109	811	317	8	21659	337	3715	192	4.23E-30	3.17E-124	0.11		
Cthe_0794	aluminium resistance protein	757	392	1234	26	111	1144	259	20	3219	186	1606	85	3.99E-03	5.41E-03	0.31		
Cthe_0798	G-D-S-L lipolytic protein	3105	7802	6078	157	327	3485	708	48	345	27	156	7	6.32E-176	1.00E+00	22.43		
Cthe_0808	CheR-type MCP methyltransferase		1068	1324	1208	63	175	1283	140	29	229	79	301	20	1.40E-02	1.00E+00	3.15	

Cthe_0818	hypothetical protein	2212	1002	2182	113	460	6891	1607	233	5819	336	3804	321	1.35E-02	1.00E+00	0.35	
Cthe_0819	ABC transporter-like protein	1613	852	2545	82	230	6715	1309	160	4711	406	2988	214	3.33E-02	1.00E+00	0.38	
Cthe_0821	coagulation factor 5/8 type-like protein	2141	2206	3414	66	211	2479	546	33	372	50	457	12	6.80E-08	1.00E+00	5.50	
Cthe_0828	1-deoxy-D-xylulose-5-phosphate synthase	1109	1232	2685	38	215	1130	286	15	6009	234	2115	96	4.37E-02	8.97E-07	0.40	
Cthe_0829	hypothetical protein	735	627	1152	49	110	557	118	18	3886	252	2337	197	1.05E-04	2.13E-26	0.25	
Cthe_0837	hypothetical protein	308	107	572	34	35	1106	69	40	1053	108	906	117	1.68E-02	1.44E-01	0.29	
Cthe_0846	hypothetical protein	1131	371	739	77	175	1827	402	101	6361	536	6580	720	4.02E-31	3.50E-08	0.11	
Cthe_0847	elongation factor P	3661	1446	2160	185	333	3893	862	157	15399	699	12359	1075	9.49E-10	2.55E-06	0.17	
Cthe_0856	branched chain amino acid aminotransferase				1675	1844	2128	76	204	1731	752	52	19470	586	7973	553	1.57E-20	5.68E-19	0.14
Cthe_0860	rubrerythrin	656	201	612	35	20	698	109	22	5692	244	3482	337	5.44E-28	2.65E-59	0.10		
Cthe_0864	2-oxoglutarate ferredoxin oxidoreductase subunit gamma				759	511	894	57	257	849	437	60	8466	441	4249	533	8.11E-31	3.08E-14	0.11
Cthe_0865	3-methyl-2-oxobutanoate dehydrogenase (ferredoxin)				579	537	1069	41	94	613	189	22	8706	159	1536	293	8.14E-16	4.10E-39	0.14
Cthe_0866	pyruvate flavodoxin/ferredoxin oxidoreductase-like protein				505	551	1511	35	57	704	223	18	6432	121	1287	158	4.23E-06	2.26E-15	0.22
Cthe_0867	2-oxoglutarate ferredoxin oxidoreductase subunit delta				126	83	263	32	5	106	44	15	814	34	262	117	2.51E-03	4.80E-17	0.27
Cthe_0870	NADPH-dependent FMN reductase			763	330	989	50	113	1325	256	48	4217	203	2654	258	1.04E-07	7.81E-05	0.19	
Cthe_0877	GTP-dependent nucleic acid-binding protein EngD				902	1401	2270	60	169	1184	455	32	10655	221	1973	248	2.73E-05	1.16E-09	0.24
Cthe_0895	RNA polymerase sigma factor RpoD			4792	2210	3073	133	3007	7303	1837	195	14938	2032	12707	660	2.16E-05	4.14E-01	0.20	
Cthe_0898	metal dependent phosphohydrolase			3525	2369	1807	107	102	2610	394	47	338	66	249	16	7.69E-12	1.00E+00	6.69	
Cthe_0903	protein translocase subunit secF	1994	1200	1472	69	382	1674	742	60	13966	506	8115	498	2.83E-18	1.04E-10	0.14		
Cthe_0904	protein-export membrane protein SecD			1429	973	1934	46	235	1209	354	24	13744	223	2481	257	2.89E-10	7.11E-21	0.18	
Cthe_0907	hypothetical protein	454	187	263	91	76	354	239	110	3554	463	5177	1468	1.31E-77	3.28E-41	0.06		

Cthe_0916	hypothetical protein	1028	594	965	41	122	884	252	26	5013	244	2963	200	5.09E-07	3.51E-11	0.21		
Cthe_0917	glutaminyl-tRNA synthetase	1365	1051	2322	38	264	1804	418	24	8383	276	2274	135	9.59E-04	1.76E-04	0.28		
Cthe_0922	diaminopimelate dehydrogenase	514	196	930	23	71	584	164	14	4506	159	1571	133	4.68E-09	1.84E-16	0.17		
Cthe_0926	signal recognition particle-docking protein FtsY	446	235	920	24	58	556	210	18	2664	149	1559	105	1.75E-04	4.07E-06	0.23		
Cthe_0932	3-oxoacyl-[acyl-carrier-protein] synthase II	1738	1409	1931	59	452	1843	382	36	12618	644	7491	363	1.87E-13	2.35E-17	0.16		
Cthe_0933	acyl carrier protein	3014	1929	1660	422	1242	3109	2319	689	24089	5353	49408	8246	7.39E-199	8.61E-17	0.05		
Cthe_0934	3-oxoacyl-[acyl-carrier-protein] reductase	422	596	942	38	49	684	87	16	3994	115	1425	157	9.73E-05	4.13E-19	0.24		
Cthe_0935	[Acyl-carrier-protein] S-malonyltransferase	522	657	1118	35	182	634	130	16	3908	165	2184	143	1.01E-04	2.51E-14	0.24		
Cthe_0937	phosphate:acyl-[acyl carrier protein] acyltransferase	897	1183	1504	50	337	1049	249	27	4030	300	3153	162	3.63E-03	6.89E-06	0.31		
Cthe_0941	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	1164	669	1063	67	593	1951	664	98	3004	548	4447	307	2.27E-06	1.76E-01	0.22		
Cthe_0948	dihydroorotate oxidase B, electron transfer subunit	157	26	626	14	9	182	77	7	1173	39	776	52	7.61E-03	2.01E-11	0.27		
Cthe_0949	carbamoyl phosphate synthase large subunit	634	198	2430	14	36	489	234	5	7507	124	2401	64	2.00E-06	1.60E-36	0.22		
Cthe_0950	carbamoyl-phosphate synthase small subunit	264	51	1040	17	9	175	86	5	2040	46	891	57	1.54E-02	2.86E-25	0.30		
Cthe_0952	dihydroorotase	344	86	1148	17	12	203	135	6	3029	113	1390	75	8.17E-05	1.57E-30	0.23		
Cthe_0953	aspartate carbamoyltransferase catalytic subunit	220	137	749	16	5	147	45	4	2293	26	757	67	1.47E-03	9.69E-108	0.24		
Cthe_0954	Uracil phosphoribosyltransferase	1048	1092	909	79	105	971	171	36	5641	114	3582	343	1.21E-05	1.66E-17	0.23		
Cthe_0956	hypothetical protein	5103	4269	3005	378	7524	12375	472	591	1040	163	1109	117	4.31E-02	7.70E-05	3.23		
Cthe_0957	protein translocase subunit yajC	206	108	282	28	10	181	34	12	955	73	561	120	5.46E-04	7.32E-29	0.23		
Cthe_0959	S-adenosylmethionine:tRNA ribosyltransferase-isomerase	465	133	1047	22	26	369	88	8	2043	76	885	62	4.49E-02	1.56E-11	0.35		
Cthe_0964	amino acid-binding ACT	968	599	841	75	264	762	375	63	2388	189	1835	214	1.96E-02	6.30E-02	0.35		

Cthe_0965	cob(I)yrinic acid a,c-diamide adenosyltransferase	439	274	536	33	20	331	118	17	2503	143	1550	172	3.22E-06	5.26E-22	0.19	
Cthe_0967	transposase, IS4	1068	602	1686	50	616	1239	1445	90	5057	755	5625	285	1.00E-10	1.99E-01	0.18	
Cthe_0976	phospho-N-acetylmuramoyl-pentapeptide-transferase	494	221	1005	24	133	804	185	19	2023	156	1243	77	1.19E-02	9.97E-03	0.31	
Cthe_0982	cell division protein MraZ	2461	2029	2374	229	817	4226	1003	243	5419	805	4754	617	3.47E-02	5.17E-01	0.37	
Cthe_0987	riboflavin kinase / FMN adenylyltransferase	459	176	796	21	46	358	141	11	1214	217	1199	69	1.49E-02	6.03E-07	0.30	
Cthe_1002	hypothetical protein	356	211	368	66	25	515	178	70	977	102	956	230	1.49E-02	5.23E-02	0.29	
Cthe_1005	elongation factor Ts	1311	558	1122	65	53	1805	494	66	10185	636	5658	564	3.29E-27	6.85E-12	0.12	
Cthe_1006	30S ribosomal protein S2	1583	789	1535	73	150	2114	1042	94	23339	864	9507	953	6.35E-76	5.01E-16	0.08	
Cthe_1007	GTP-binding protein TypA	734	250	1648	20	85	584	182	8	4625	153	2091	78	1.46E-04	4.28E-17	0.26	
Cthe_1008	aminodeoxychorismate lyase	747	495	948	27	62	860	202	17	3238	165	1610	97	1.83E-03	1.98E-05	0.28	
Cthe_1013	hypothetical protein	638	430	1091	28	105	738	734	38	2113	174	2010	87	1.05E-02	4.53E-01	0.32	
Cthe_1018	binding-protein-dependent transport systems inner membrane component	14398	6882	7282	470	5483	14859	4166	502	32965	3990	31015	1818	2.02E-02	3.13E-01	0.26	
Cthe_1020	extracellular solute-binding protein	115884	56365	44308	2251	12140	154216	39477	2681	957331	27550	363001	20665	1.85E-03	2.36E-03	0.11	
Cthe_1022	glycerol-3-phosphate dehydrogenase (NAD(P)(+))	362	154	966	20	31	316	97	8	1915	111	1195	69	8.55E-03	1.11E-14	0.29	
Cthe_1026	50S ribosomal protein L32P	1805	1013	892	265	474	1483	950	326	12510	2634	20112	4225	1.86E-122	2.56E-29	0.06	
Cthe_1027	hypothetical protein	3310	2226	1804	210	611	2542	762	142	26441	3415	37913	3029	2.87E-79	1.55E-91	0.07	
Cthe_1028	acetate kinase	1144	365	1213	32	51	1060	448	27	19121	480	4957	434	1.49E-78	1.19E-57	0.07	
Cthe_1029	phosphotransacetylase	469	287	1025	23	28	641	231	16	12111	203	2074	280	1.13E-50	1.19E-70	0.08	
Cthe_1037	cell wall hydrolase/autolysin	948	587	1127	44	138	648	106	16	3373	130	1336	123	2.68E-02	1.50E-10	0.36	
Cthe_1039	ribosomal protein S20	5350	2542	1885	403	502	3574	3224	541	73844	8581	64400	9828	9.09E-223	4.55E-35	0.04	
Cthe_1047	hypothetical protein	4114	2933	2208	244	477	4219	751	163	7340	1497	8897	807	1.85E-02	4.11E-03	0.30	
Cthe_1050	recA protein	1703	1643	2919	85	429	2398	488	53	9958	415	3128	281	1.62E-03	5.61E-04	0.30	

Cthe_1053L-lactate dehydrogenase	434	237	846	22	20	313	116	9	3135	110	990	95	1.79E-04	3.17E-23	0.23	
Cthe_1057phage shock protein PspC	340	146	206	50	290	448	172	87	1204	150	1616	344	1.77E-10	1.07E-02	0.15	
Cthe_1058serine hydroxymethyltransferase		2338	1107	1867	61	220	1531	428	32	13168	492	4252	313	5.80E-09	5.65E-17	0.19
Cthe_1065type IV pilus assembly PilZ	2533	2234	2108	149	167	2036	157	50	396	65	444	32	3.50E-05	1.00E+00	4.66	
Cthe_1078hypothetical protein	4812	2022	1147	359	986	4153	1749	437	2053	2524	5924	1152	1.88E-02	4.91E-01	0.31	
Cthe_1080carbohydrate-binding family 25 protein		1458	1313	1151	175	1655	3307	100	211	280	41	433	53	6.67E-03	5.34E-04	3.30
Cthe_1081hypothetical protein	692	998	604	71	105	1270	142	48	120	32	95	14	3.41E-07	1.00E+00	5.07	
Cthe_1164FAD dependent oxidoreductase	919	804	2051	33	217	951	394	19	2898	444	2174	86	4.14E-02	4.77E-03	0.38	
Cthe_1187hypothetical protein	411	135	680	20	100	321	92	10	4075	167	1642	149	9.20E-14	3.46E-48	0.13	
Cthe_1188hypothetical protein	259	79	717	18	47	162	71	7	1959	86	442	70	4.85E-03	3.08E-21	0.26	
Cthe_1192hypothetical protein	968	715	1491	44	98	1923	372	38	12045	573	3411	345	1.57E-22	1.45E-13	0.13	
Cthe_1200adenosylhomocysteinase	847	558	1308	31	223	761	266	19	3164	244	1616	92	1.22E-02	4.99E-04	0.34	
Cthe_1205putative serine protein kinase, PrkA		4297	7795	6453	140	306	3940	236	31	589	206	582	20	1.77E-10	1.00E+00	7.00
Cthe_1210hypothetical protein	937	1143	1026	78	1678	1005	93	71	456	40	193	28	3.47E-02	9.71E-02	2.79	
Cthe_1211tryptophan synthase subunit beta		5478	4718	4429	154	1271	4774	767	79	37910	2663	12563	901	5.72E-05	7.45E-17	0.17
Cthe_1212hypothetical protein	7191	8163	7546	353	225	6831	544	114	1786	287	1452	93	1.88E-02	1.00E+00	3.80	
Cthe_1220hypothetical protein	1250	645	979	45	705	1078	389	44	3758	381	4514	214	1.14E-06	1.86E-03	0.21	
Cthe_122350S ribosomal protein L20	1706	964	1356	162	53	2088	663	152	8944	1029	5486	1047	1.80E-14	4.41E-07	0.15	
Cthe_1228threonyl-tRNA synthetase	2203	2797	4028	68	159	1525	381	19	21924	454	4916	301	2.07E-04	7.85E-52	0.23	
Cthe_1229spore germination protein-like protein		549	492	883	44	37	544	46	14	3248	71	816	141	9.75E-03	3.82E-21	0.31
Cthe_1236fibronectin, type III	244	65	529	16	67	231	94	11	913	36	825	52	3.68E-02	5.50E-04	0.31	
Cthe_1237leucyl-tRNA synthetase	3707	2248	4189	58	843	4325	1647	54	41953	2252	18507	559	2.14E-24	2.40E-13	0.10	
Cthe_1238iojap-like protein	219	157	315	28	11	214	29	11	1449	37	431	118	1.43E-03	6.30E-33	0.24	
Cthe_1240metal dependent phosphohydrolase		273	138	505	22	59	310	160	20	1281	135	1118	102	3.24E-04	6.47E-05	0.22

Cthe_1245	phosphoribosylamine--glycine ligase	1109	522	1861	39	127	903	395	23	8050	414	3791	213	8.37E-10	1.33E-15	0.18
Cthe_1246	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	795	317	1602	24	27	504	170	8	6450	109	1531	109	7.99E-06	3.47E-36	0.22
Cthe_1247	phosphoribosylglycinamide formyltransferase	323	119	650	24	71	217	167	17	2272	123	1072	121	1.71E-05	3.16E-10	0.20
Cthe_1248	phosphoribosylaminoimidazole synthetase	395	172	1200	24	23	258	85	6	3254	58	664	81	5.98E-03	3.43E-30	0.30
Cthe_1249	amidophosphoribosyltransferase	469	286	1352	20	24	337	88	5	5499	78	879	91	8.42E-06	1.43E-70	0.22
Cthe_1250	phosphoribosylaminoimidazole carboxylase, catalytic subunit	222	151	589	26	13	131	64	8	2299	50	325	109	1.18E-03	4.20E-50	0.24
Cthe_1251	xanthine/uracil/vitamin C permease	4740	2165	2677	97	411	2387	563	41	25391	896	7222	519	1.92E-06	5.80E-27	0.19
Cthe_1252	auxin efflux carrier	413	288	1071	23	58	531	132	11	3931	71	1566	105	1.52E-05	3.09E-17	0.22
Cthe_1255	hypothetical protein	348	135	399	27	127	463	379	51	1229	248	2061	184	2.90E-10	3.47E-02	0.15
Cthe_1257	carbohydrate-binding, CenC-like protein	2289	3252	6761	56	114	6108	566	31	1073	168	982	17	3.90E-02	1.00E+00	3.29
Cthe_1258	copper amine oxidase-like protein	2952	2546	2466	147	229	3234	212	65	479	89	499	34	1.74E-04	4.03E-01	4.32
Cthe_1259	CDP-alcohol phosphatidyltransferase	414	171	415	26	127	369	267	32	1686	320	1949	179	6.94E-11	2.37E-05	0.15
Cthe_1260	4-hydroxybenzoyl-CoA thioesterase	338	240	473	36	36	351	150	26	1486	134	1001	144	1.81E-03	1.31E-05	0.25
Cthe_1261	6-phosphofructokinase	740	581	1202	37	49	792	160	17	6734	178	1547	185	1.04E-07	3.46E-24	0.20
Cthe_1262	hypothetical protein	1152	773	1011	213	183	2095	349	214	3336	755	4742	1109	1.62E-08	5.21E-04	0.19
Cthe_1263	transcription attenuation protein MtrB	1074	700	808	158	115	1712	335	155	3815	731	5378	1021	1.19E-13	4.62E-07	0.15
Cthe_1265	alpha-phosphoglucomutase	2261	1489	2344	50	162	2685	858	40	12433	534	6791	243	3.38E-08	6.48E-05	0.21
Cthe_1270	proteinase inhibitor I4, serpin	458	403	1265	24	41	699	873	37	2545	172	1164	71	1.98E-02	6.40E-01	0.34
Cthe_1279	50S ribosomal protein L28	2548	1107	752	334	112	1510	1228	382	37145	4275	38994	9725	0.00E+00	7.59E-140	0.03
Cthe_1286	peptidase S1 and S6, chymotrypsin/Hap	1118	1020	2404	42	91	2304	1231	53	27347	2800	16050	713	1.79E-137	2.48E-31	0.06
Cthe_1288	two component transcriptional regulator	465	260	925	34	67	551	98	17	2218	98	657	95	4.17E-02	5.19E-06	0.36
Cthe_1290	hypothetical protein	715	459	654	76	444	711	296	81	1341	196	2271	250	7.26E-03	1.13E-01	0.30
Cthe_1292	phosphoglycerate mutase	2459	1512	2387	75	746	3553	1783	107	12104	1026	7189	385	4.23E-09	1.68E-01	0.19

Cthe_1302	hypothetical protein	1186	496	1475	26	63	940	365	16	3742	243	1848	78	1.30E-02	1.03E-03	0.33		
Cthe_1306	hypothetical protein	1075	1704	2128	177	535	1541	1965	275	7204	1560	13086	1294	8.11E-20	2.00E-02	0.14		
Cthe_1309	radical SAM family protein		5608	10491	8985	341	717	7455	456	110	1413	182	936	57	9.83E-07	5.41E-01	5.98	
Cthe_1310	accessory gene regulator B		2605	3605	3702	240	304	4083	311	110	592	157	703	64	4.03E-03	5.41E-01	3.75	
Cthe_1312	glycyl-tRNA synthetase		2179	1257	2255	58	257	2194	501	36	10602	465	3517	230	4.04E-05	9.07E-06	0.25	
Cthe_1313	phosphopantothenoylcysteine decarboxylase / phosphopantothenate-cysteine ligase							646	345	1245	26	175	521	258	16	2968	342	2125
	107	1.07E-04	1.26E-07	0.24														
Cthe_1314	DNA-directed RNA polymerase subunit omega				371	176	335	53	25	432	315	80	2015	333	2275	469	2.18E-19	4.22E-
06	0.11																	
Cthe_1315	guanylate kinase	904	443	1212	59	128	1193	337	49	3478	302	2109	221	4.32E-04	3.41E-03	0.27		
Cthe_1316	hypothetical protein	821	405	975	111	219	1153	438	126	1995	298	1572	337	1.21E-02	2.68E-01	0.33		
Cthe_1317	hypothetical protein	904	450	1099	39	53	1012	174	22	3236	174	1389	120	9.73E-03	7.68E-05	0.33		
Cthe_1326	GTP-binding protein LepA		889	392	2079	26	95	1066	292	14	7466	301	2764	126	2.09E-07	3.52E-14	0.21	
Cthe_1329	putative CoA-substrate-specific enzyme activase				9626	5988	7677	77	1962	10350	4123	76	27507	4298	16032	284	3.79E-02	3.65E-
01	0.27																	
Cthe_1330	thylakoidal processing peptidase		520	253	639	38	17	361	138	20	1683	88	931	112	3.71E-02	3.49E-06	0.34	
Cthe_1331	aspartyl-tRNA synthetase		1549	1058	2652	41	186	1099	248	14	13830	322	2275	196	2.76E-07	7.40E-39	0.21	
Cthe_1332	histidyl-tRNA synthetase		778	666	1637	35	66	491	95	8	5706	80	1027	113	3.97E-03	2.09E-42	0.31	
Cthe_1336	hypothetical protein	1029	627	1032	54	289	1928	1066	103	1963	276	1877	139	4.96E-02	1.00E+00	0.39		
Cthe_1350	single-strand binding protein		1108	745	1263	110	522	1193	508	108	2230	685	4669	468	1.56E-05	9.05E-03	0.24	
Cthe_1363	lipopolysaccharide biosynthesis protein		1313	792	1454	36	88	1224	334	21	4427	279	2502	114	4.44E-03	1.17E-04	0.32	
Cthe_1375	aspartate kinase	1649	1010	1678	45	232	1893	644	39	8484	683	5220	242	4.11E-09	1.93E-05	0.19		
Cthe_1376	homoserine dehydrogenase		539	289	1252	23	22	278	82	5	3043	61	748	64	3.46E-02	2.54E-27	0.36	
Cthe_1377	hypothetical protein	230	134	289	20	8	141	28	6	1278	38	547	88	8.04E-04	7.76E-65	0.23		
Cthe_1380	response regulator receiver protein				1166	553	739	93	266	854	495	94	679	1360	2295	459	2.53E-07	1.03E-03	0.20
Cthe_1385	preprotein translocase subunit SecA				1978	1915	3799	40	298	2824	1003	29	27301	604	5726	260	3.11E-11	7.99E-12	0.15

Cthe_1386	hypothetical protein	692	589	927	52	285	694	233	37	7989	383	3760	440	6.16E-24	5.97E-30	0.12		
Cthe_1387	hypothetical protein	557	456	539	56	125	531	136	35	2109	168	1725	224	5.82E-04	9.77E-08	0.25		
Cthe_1388	TrpR like protein, YerC/YecD	426	356	408	58	22	400	57	25	1533	113	1170	212	5.23E-03	1.43E-13	0.27		
Cthe_1398	dockerin type I cellulosome protein		5664	3004	5381	79	1528	9947	518	64	1021	199	1022	22	2.07E-02	1.31E-01	3.59	
Cthe_1401	hypothetical protein	447	306	355	47	130	508	146	42	1036	109	1524	175	5.32E-03	4.72E-03	0.27		
Cthe_1411	tryptophan synthase, alpha chain		513	356	944	33	101	438	178	18	2216	223	1362	115	3.69E-03	2.47E-07	0.29	
Cthe_1421	signal peptide peptidase SppA, 36K type		859	465	1218	36	69	967	196	21	4835	109	1112	130	1.08E-03	8.93E-07	0.28	
Cthe_1422	RDD domain-containing protein	946	629	1405	54	161	1298	447	47	4370	130	1810	172	6.04E-03	4.86E-02	0.31		
Cthe_1456	ABC transporter-like protein	257	58	569	15	41	130	29	4	1614	193	740	80	1.70E-05	7.62E-137	0.19		
Cthe_1457	polar amino acid ABC transporter inner membrane subunit				355	59	797	25	37	177	49	7	1822	136	679	92	5.87E-03	6.47E-50 0.27
Cthe_1458	extracellular solute-binding protein		253	103	531	15	6	170	54	5	2859	152	1555	123	4.37E-15	1.43E-194	0.12	
Cthe_1463	hypothetical protein	1253	609	987	116	7	1679	96	69	308	34	85	31	1.22E-03	2.07E-01	3.74		
Cthe_1478	TetR family transcriptional regulator		358	200	360	23	62	402	277	31	2281	353	3387	247	4.08E-29	4.48E-12	0.09	
Cthe_1479	hypothetical protein	107	26	383	11	67	217	183	18	839	81	1069	71	2.18E-09	9.83E-03	0.15		
Cthe_1481	hypothetical protein	449	107	1654	13	32	687	438	11	3325	189	2236	54	1.27E-04	1.81E-03	0.24		
Cthe_1482	Cof-like hydrolase	414	178	826	24	59	432	216	18	1931	242	1583	109	9.49E-05	2.86E-06	0.22		
Cthe_1484	transcriptional regulator	851	342	966	43	250	1523	385	53	1611	564	1596	155	1.19E-03	2.02E-01	0.28		
Cthe_1495	pyridoxamine 5'-phosphate oxidase-like FMN-binding protein				803	574	1117	90	167	1163	407	85	4477	464	5638	596	2.94E-14	3.47E-08 0.15
Cthe_1519	SAM-dependent methyltransferase		334	82	739	13	74	251	191	10	1533	132	1568	58	2.55E-04	1.68E-06	0.22	
Cthe_1540	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B				259	192	901	13	180	238	147	8	2832	140	626	57	3.06E-04	1.87E-09 0.23
Cthe_1541	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit A				290	189	844	12	191	233	147	8	2645	146	594	52	5.66E-04	2.74E-08 0.23
Cthe_1542	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit C				57	57	229	16	27	65	17	6	770	26	144	70	1.34E-05	5.30E-52 0.23

Cthe_1543aspartyl-tRNA synthetase	578	494	1376	20	264	492	210	10	6089	307	1416	101	8.48E-08	1.09E-17	0.20	
Cthe_1548enoyl-CoA hydratase/isomerase	2994	3072	1766	152	535	1671	301	54	203	93	279	23	3.03E-12	1.00E+00	6.61	
Cthe_1574hypothetical protein	698	464	1019	41	340	1238	309	44	2255	179	1587	121	1.91E-02	2.29E-01	0.34	
Cthe_1586binding-protein-dependent transport systems inner membrane component	4226	1181	1228	56	1326	1747	68	26	104	21	114	3	1.16E-122	1.00E+00	18.67	
Cthe_1588extracellular solute-binding protein	9726	4913	2185	218	1623	5525	75	84	283	11	189	9	1.66E-235	5.07E-25	24.22	
Cthe_1660transposase, IS4	1060	597	1678	50	616	1237	1441	90	5045	751	5616	285	8.24E-11	1.98E-01	0.18	
Cthe_1666transposase, IS4	1060	597	1681	50	616	1240	1443	90	5050	753	5621	285	7.98E-11	1.98E-01	0.18	
Cthe_1756putative virion core protein (lumpy skin disease virus)-like protein	1324	1250	1658	52	274	1063	274	24	7651	363	3439	214	3.28E-05	1.50E-14	0.24	
Cthe_1757peptidase M23B	415	325	960	26	17	291	62	6	2767	46	501	74	3.27E-02	4.79E-26	0.35	
Cthe_1758hypothetical protein	339	185	643	37	21	303	85	17	1512	116	702	122	2.04E-02	1.40E-09	0.30	
Cthe_17601A family penicillin-binding protein	1141	744	2378	24	314	974	258	11	9279	421	3552	117	2.44E-07	8.42E-23	0.21	
Cthe_1766gamma-glutamyl kinase	615	530	1082	38	65	464	77	11	3164	61	790	100	4.30E-02	2.79E-16	0.38	
Cthe_1767hypothetical protein	1064	464	1252	39	372	1057	569	42	6262	521	4894	261	1.07E-14	7.47E-06	0.15	
Cthe_1768NifU-like domain-containing protein	430	188	656	26	27	408	115	14	3438	67	944	134	2.75E-06	1.12E-18	0.19	
Cthe_1771rubrerythrin	2482	1358	1743	144	392	3325	740	138	18107	1429	13761	1340	1.87E-36	1.35E-14	0.11	
Cthe_1777amidohydrolase	783	485	1630	34	180	1107	619	35	3579	217	1781	103	1.02E-02	2.00E-01	0.33	
Cthe_1778copper amine oxidase-like protein	1313	620	1586	52	37	1705	660	51	13484	144	3479	365	1.15E-17	6.77E-08	0.14	
Cthe_178230S ribosomal protein S9	1215	409	916	91	180	1252	575	107	6483	589	5198	707	8.16E-21	8.02E-07	0.13	
Cthe_178350S ribosomal protein L13	632	335	789	57	21	604	234	40	6553	196	2454	445	1.29E-18	7.63E-27	0.13	
Cthe_1791UvrB/UvrC protein	2146	1895	7853	333	211	6317	6828	732	9255	2320	15671	1330	3.64E-03	1.00E+00	0.25	
Cthe_17953-deoxy-D-arabinoheptulosonate-7-phosphate synthase	2845	2345	2664	110	649	3384	585	73	19060	822	7321	577	8.90E-07	2.98E-09	0.19	
Cthe_1796prephenate dehydrogenase	3213	2980	2709	117	683	2810	883	74	16313	837	7137	487	6.77E-04	4.51E-06	0.24	

Cthe_1797	3-phosphoshikimate 1-carboxyvinyltransferase			3188	2612	3032	100	605	3179	698	60	13041	520	5276	319	2.35E-02	8.04E-04	0.31
Cthe_1799	ABC transporter-like protein	678	543	2551	27	135	843	880	26	2981	279	3471	77	1.88E-02	2.03E-01	0.35		
Cthe_1801	ABC transporter-like protein	831	494	1279	48	145	1118	358	40	3187	296	1713	159	3.54E-03	1.54E-02	0.30		
Cthe_1802	cobalt ABC transporter inner membrane subunit CbiQ			1607	884	1678	73	674	2181	700	81	5363	457	3677	266	4.89E-04	1.47E-01	0.27
Cthe_1803	cobalamin (vitamin B12) biosynthesis CbiM protein			3452	2101	2948	115	861	4109	1186	106	11352	1051	7321	430	3.49E-03	7.91E-02	0.27
Cthe_1804	cell wall hydrolase/autolysin	627	869	1561	60	26	1007	50	19	229	30	129	13	3.31E-05	1.00E+00	4.62		
Cthe_1805	response regulator receiver modulated diguanylate cyclase			2549	1843	1787	95	167	2492	126	40	470	59	403	23	4.55E-04	1.00E+00	4.13
Cthe_1837	hypothetical protein	1090	692	1170	82	293	2450	739	127	5215	770	8858	656	1.86E-22	1.21E-03	0.13		
Cthe_1839	biotin synthase	1545	1933	2154	76	139	919	85	15	413	46	570	21	3.37E-03	1.00E+00	3.62		
Cthe_1840	cysteine synthase	993	645	1059	41	211	1242	401	37	6282	586	3881	266	8.10E-14	1.61E-08	0.15		
Cthe_1844	BadM/Rrf2 family transcriptional regulator			253	182	573	31	32	254	50	12	1381	59	525	92	4.85E-02	4.65E-12	0.34
Cthe_1845	homoserine O-succinyltransferase			1115	654	1253	47	123	1140	168	24	4488	214	1703	153	3.42E-03	4.68E-07	0.31
Cthe_1853	cold-shock DNA-binding protein family protein			1984	1296	1050	310	509	2842	1093	439	6569	1905	15111	2856	3.05E-33	1.06E-05	0.11
Cthe_1856	small acid-soluble spore protein beta			78	48	156	23	7	137	31	17	401	41	348	104	2.67E-07	3.39E-10	0.22
Cthe_1857	carboxyl-terminal protease	475	282	1032	19	39	569	110	9	2392	89	980	57	2.49E-02	1.83E-07	0.33		
Cthe_1859	cell division protein FtsX	386	236	1044	26	49	291	132	11	1897	93	1011	73	4.63E-02	6.92E-09	0.36		
Cthe_1860	cell division ATP-binding protein FtsE			335	226	600	23	49	309	63	10	2123	92	1256	104	1.59E-04	9.14E-23	0.22
Cthe_1863	N-acetyl-gamma-glutamyl-phosphate reductase			224	64	546	11	3	84	28	2	4149	42	694	96	6.23E-18	0.00E+00	0.11
Cthe_1864	acetylglutamate kinase	269	38	756	16	10	131	46	4	2344	25	468	63	2.39E-03	8.86E-95	0.25		
Cthe_1866	acetylornithine aminotransferase			515	403	1388	27	33	368	108	7	7690	52	1006	149	3.63E-09	4.14E-102	0.18

Cthe_1867	carbamoyl-phosphate synthase small subunit			432	285	1308	26	21	286	85	6	5493	44	497	117	2.96E-05	8.53E-72	0.22
Cthe_1868	carbamoyl-phosphate synthase large subunit			1054	570	3540	22	54	736	330	7	15024	140	1444	107	2.32E-07	2.25E-49	0.21
Cthe_1869	ornithine carbamoyltransferase	602	355	1220	33	35	458	222	16	7862	164	1595	220	9.64E-14	2.41E-40	0.15		
Cthe_1870	GCN5-like N-acetyltransferase	280	127	547	28	31	311	276	33	3837	226	2090	284	8.11E-26	1.31E-14	0.10		
Cthe_1912	copper amine oxidase-like protein		6507	4127	4449	134	435	5578	1421	83	27048	983	10282	512	3.87E-02	3.06E-04	0.26	
Cthe_1916	two component transcriptional regulator		791	892	1845	73	54	940	110	24	386	49	207	23	8.50E-03	1.00E+00	3.17	
Cthe_1918	arginine decarboxylase	1722	1106	2027	47	356	1363	315	23	17165	759	5341	350	4.53E-20	2.29E-45	0.13		
Cthe_1923	CTP synthetase	1763	792	2506	44	178	1308	327	19	11886	352	3074	201	8.10E-07	4.83E-20	0.22		
Cthe_1939	magnesium transporter	742	441	849	21	48	765	366	19	1775	210	1271	58	4.43E-02	1.14E-01	0.36		
Cthe_1944	hypothetical protein	1202	693	776	69	496	1268	384	71	3724	541	4672	378	3.03E-09	2.26E-04	0.18		
Cthe_1945	thioredoxin-disulfide reductase	1789	639	1148	56	226	1506	553	50	3191	1191	2200	231	3.26E-05	4.03E-03	0.24		
Cthe_1960	peptidoglycan binding domain-containing protein			2100	1451	2300	77	254	2526	237	40	13065	492	2684	330	3.86E-06	6.23E-11	0.23
Cthe_1967	DNA segregation ATPase FtsK/SpoIIIE and-like proteins				818	369	1355	20	58	1591	767	30	2134	527	2561	78	3.55E-04	3.54E-01	0.26
Cthe_1970	hypothetical protein	757	481	862	98	50	1134	363	97	3073	340	2646	458	3.00E-06	1.21E-03	0.21		
Cthe_1971	hypothetical protein	650	392	749	77	22	834	116	45	2226	90	864	209	4.96E-02	1.04E-03	0.37		
Cthe_1988	hypothetical protein	814	566	942	119	63	1060	475	121	2406	314	3028	471	1.60E-04	2.17E-02	0.25		
Cthe_2022	hypothetical protein	815	566	947	119	63	1060	476	121	2406	315	3028	471	1.72E-04	2.17E-02	0.25		
Cthe_2061	hypothetical protein	1047	600	922	86	639	1288	1016	162	3042	961	5749	590	1.74E-15	7.52E-02	0.15		
Cthe_2070	hypothetical protein	646	488	1136	37	196	638	232	23	1693	258	2098	110	1.63E-02	1.06E-03	0.34		
Cthe_2071	hypothetical protein	1616	1646	1674	77	290	2943	601	70	384	87	347	23	9.02E-03	3.62E-02	3.35		
Cthe_2075	hypothetical protein	1475	663	1851	77	206	2942	749	96	4678	361	2777	243	5.00E-03	4.05E-01	0.32		
Cthe_2079	thylakoidal processing peptidase	1170	776	953	73	273	1382	415	67	3009	382	3987	294	8.27E-05	6.59E-03	0.25		

Cthe_2093hypothetical protein	777	422	1153	31	25	514	66	8	4987	137	1701	136	1.84E-05	1.95E-64	0.23	
Cthe_2095TatD family hydrolase	468	281	954	31	115	567	340	30	3293	320	2964	193	1.38E-10	1.26E-07	0.16	
Cthe_2096methionyl-tRNA synthetase	891	623	2359	27	126	732	199	9	5961	178	1851	86	5.30E-03	6.90E-16	0.31	
Cthe_2100AbrB family transcriptional regulator		4869	2625	1862	560	568	5000	4687	1120	23621	3255	34008	5766	6.50E-30	5.10E-02	0.10
Cthe_21034Fe-4S ferredoxin, iron-sulfur binding		7265	3599	2812	1072	4278	7824	10139	3240	40505	14656	80006	19386	7.99E-72	3.03E-03	0.06
Cthe_2104hypothetical protein	945	488	1229	43	165	941	234	27	3329	225	2238	146	2.29E-03	4.10E-05	0.29	
Cthe_2105DNA polymerase III subunit delta'		822	254	1161	31	105	632	244	19	1584	306	1312	83	4.65E-02	4.47E-03	0.37
Cthe_2106hypothetical protein	378	184	570	35	35	427	220	33	1588	186	1228	156	2.54E-04	5.60E-04	0.22	
Cthe_2107thymidylate kinase	505	212	883	33	23	386	134	15	1929	125	943	98	2.66E-02	5.87E-08	0.34	
Cthe_2108arginine decarboxylase		902	582	1629	30	177	1171	1952	68	13918	2199	17206	540	5.12E-153	1.50E-07	0.06
Cthe_2109copper amine oxidase-like protein		1506	1260	3073	32	193	1676	588	18	16485	768	4945	193	9.18E-13	3.76E-18	0.17
Cthe_2115hypothetical protein	194	55	282	26	51	146	86	21	1492	185	1533	258	7.16E-29	3.52E-33	0.10	
Cthe_2116binding-protein-dependent transport systems inner membrane component						362	85	588	19	74	290	146	14	2421	201	1498	123
1.32E-09	2.12E-16	0.15															
Cthe_2117putative sulfonate transport system substrate-binding protein						322	85	1020	19	36	268	153	10	2604	78	491	66	7.57E-03
3.88E-09	0.29																	
Cthe_2142recombination protein RecR		813	364	770	46	180	1095	263	44	5689	419	2626	333	2.96E-16	4.47E-10	0.14	
Cthe_2143hypothetical protein	415	169	739	54	47	518	69	28	3338	117	822	272	6.97E-06	1.44E-20	0.20	
Cthe_2150integral hypothetical protein		1095	661	1774	37	210	903	303	19	3671	209	2328	100	2.53E-02	2.81E-04	0.37
Cthe_2152hypothetical protein	1118	1393	1029	99	426	818	342	59	2338	275	3933	276	1.87E-02	1.64E-03	0.36	
Cthe_2164rubredoxin-type Fe(Cys)4 protein		1957	805	530	296	758	1949	667	397	3011	2617	6555	2551	4.29E-28	6.25E-06	0.12
Cthe_2168peptidase	1327	1260	2941	36	133	2305	1663	45	27086	1486	16320	445	1.72E-71	2.78E-12	0.08	
Cthe_217250S ribosomal protein L31P		3576	1831	2097	504	309	3133	2359	671	25366	4436	36470	7350	3.20E-78	3.83E-15	0.07
Cthe_2174transcription termination factor Rho		4798	8132	10527	173	360	9930	975	83	2670	187	1247	47	3.02E-02	1.00E+00	3.68
Cthe_2175hypothetical protein	1191	909	676	60	122	2392	79	50	209	45	196	17	4.24E-03	1.00E+00	3.53	

Cthe_2183UDP-glucose pyrophosphorylase	1440	1413	1758	76	80	1546	165	30	6720	158	1687	208	2.37E-02	2.23E-08	0.37	
Cthe_218530S ribosomal protein S18	1226	469	922	130	195	1468	1335	272	14040	1630	12558	2301	3.88E-146	1.06E-10	0.06	
Cthe_2186single-strand binding protein	1379	533	990	99	32	1083	159	48	21750	332	4124	1312	8.42E-75	3.18E-213	0.08	
Cthe_218730S ribosomal protein S6P	779	289	448	75	86	686	248	69	11444	347	4457	1196	1.68E-93	6.39E-73	0.06	
Cthe_2193carbohydrate-binding family 6 protein	1844	1250	3395	32	671	2337	898	26	8092	610	3632	99	5.20E-03	6.72E-02	0.32	
Cthe_2202hypothetical protein	424	203	794	27	238	453	175	22	1856	188	1330	106	1.27E-03	5.67E-04	0.25	
Cthe_2203GTP cyclohydrolase I	572	228	812	40	184	964	237	43	4141	387	3512	322	1.10E-18	1.81E-10	0.12	
Cthe_2217flagellar protein FliS	1643	1150	864	130	20	1870	60	61	230	42	132	26	2.57E-06	1.72E-01	5.00	
Cthe_2218flagellar hook-associated 2-like protein	5156	3998	4182	76	310	5284	283	31	752	325	863	23	4.52E-02	1.00E+00	3.30	
Cthe_2219flagellar protein FlaG protein	879	694	588	78	7	1056	21	34	69	15	41	8	4.99E-47	1.00E+00	9.75	
Cthe_2234hypothetical protein	804	631	716	80	7	692	25	24	156	15	64	14	1.04E-11	1.00E+00	5.71	
Cthe_2236flagellin-like protein	22667	21587	12057	1000	752	20859	396	335	3226	402	3254	195	1.40E-03	1.00E+00	5.13	
Cthe_2237flagellin-like protein	14054	17342	8955	718	493	10979	126	173	1711	192	1137	87	5.62E-13	5.42E-01	8.25	
Cthe_2244flagellar hook-associated protein FlgK	1545	2132	2551	61	175	1922	80	19	412	48	216	11	3.39E-08	1.00E+00	5.55	
Cthe_2245hypothetical protein	686	891	671	65	3	668	9	16	73	7	27	5	4.36E-105	1.00E+00	13.00	
Cthe_2246anti-sigma-28 factor, FlgM	16933	9818	4583	1538	481	22221	144	937	510	292	589	157	1.37E-23	1.68E-07	9.80	
Cthe_2247regulatory protein MerR	1393	910	772	99	79	1737	31	51	155	41	145	21	1.44E-05	1.00E+00	4.71	
Cthe_2251methionine adenosyltransferase	862	209	1235	27	61	688	175	13	5191	125	1347	118	1.73E-05	1.63E-14	0.23	
Cthe_2269V-type ATP synthase subunit D	945	502	915	50	109	823	303	36	2226	342	2488	179	1.15E-03	3.54E-04	0.28	
Cthe_2276AAA ATPase, central region	647	540	1042	27	54	551	109	10	2559	115	1143	70	4.85E-02	1.16E-09	0.39	
Cthe_2278extracellular solute-binding protein	586	395	730	28	110	412	39	9	2236	70	921	79	3.95E-02	3.74E-14	0.35	
Cthe_2305undecaprenyl pyrophosphate phosphatase	696	378	790	32	335	678	387	36	5380	427	5239	295	2.01E-27	3.94E-12	0.11	
Cthe_2306MATE efflux family protein	827	422	1552	28	197	622	264	15	7023	409	3242	168	1.41E-11	1.35E-25	0.17	
Cthe_2307hemerythrin-like metal-binding protein	1322	1027	777	106	7	1154	60	38	315	27	178	28	9.01E-04	1.00E+00	3.79	

Cthe_2323	transposase, IS4	1060	597	1676	50	616	1237	1440	90	5044	751	5614	284	9.18E-11	1.97E-01	0.18		
Cthe_2328	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	1280	884	2205	49	408	1375	236	25	3491	481	2873	131	3.16E-02	3.90E-04	0.37		
Cthe_2333	two component transcriptional regulator	490	294	794	32	79	449	145	18	1813	191	1533	116	2.74E-03	4.71E-08	0.28		
Cthe_2334	polysaccharide biosynthesis protein CapD	1344	1290	2589	40	300	1448	688	27	5833	308	3046	108	2.34E-02	3.26E-02	0.37		
Cthe_2345	DegT/DnrJ/EryC1/StrS aminotransferase	1014	452	1773	39	220	2251	1599	85	2542	833	5493	194	1.46E-07	8.68E-01	0.20		
Cthe_2356	quinolinate synthetase A	2245	1409	2022	88	305	3136	1214	102	20464	1982	15650	954	2.24E-51	7.35E-12	0.09		
Cthe_2357	hypothetical protein	1283	772	1472	42	91	1613	883	49	24729	482	7346	579	1.89E-82	2.45E-23	0.07		
Cthe_2368	hypothetical protein	204	122	226	36	22	276	92	35	645	53	333	110	3.43E-02	7.53E-02	0.33		
Cthe_2370	50S ribosomal protein L34	328	180	232	78	14	348	54	49	989	106	469	280	1.01E-02	1.06E-05	0.28		
Cthe_2376	DNA gyrase subunit B	3279	1714	3798	65	297	4302	1740	67	10233	1612	9657	271	5.58E-04	8.20E-02	0.24		
Cthe_2383	hypothetical protein	5196	3695	3782	199	605	5462	1969	172	69176	1668	25754	2212	2.22E-26	5.84E-19	0.09		
Cthe_2390	pyruvate ferredoxin oxidoreductase subunit gamma	1684	1474	1737	121	244	1243	607	76	18086	553	7425	953	4.27E-23	5.81E-29	0.13		
Cthe_2391	pyruvate ferredoxin oxidoreductase subunit delta	645	669	973	107	45	531	147	43	9582	184	1702	791	4.85E-18	9.12E-91	0.14		
Cthe_2392	pyruvate ferredoxin oxidoreductase subunit alpha	1625	1757	2702	73	245	1534	848	49	30340	529	5234	640	1.68E-31	1.42E-30	0.11		
Cthe_2393	thiamine pyrophosphate enzyme-like TPP-binding protein	2458	2080	2900	114	653	2600	845	82	31914	939	10374	983	6.83E-23	6.25E-24	0.12		
Cthe_2399	formate-tetrahydrofolate ligase	1217	690	2371	36	144	1214	252	16	5859	181	1751	99	2.21E-02	1.84E-06	0.36		
Cthe_2403	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	302	82	775	19	31	328	86	9	2060	172	1232	93	2.35E-05	7.74E-21	0.20		
Cthe_2404	GntR family transcriptional regulator	431	141	943	30	76	529	132	18	2164	128	1055	107	5.37E-03	3.53E-06	0.28		
Cthe_2405	heavy metal transport/detoxification protein	189	69	171	29	70	247	111	42	559	85	556	139	4.06E-06	4.37E-02	0.21		
Cthe_2408	phage shock protein PspA	2161	1100	1387	91	107	2197	402	61	8091	471	3413	367	4.83E-05	2.07E-05	0.25		

Cthe_2425	MotA/TolQ/ExbB proton channel			2532	2060	1617	106	191	1983	104	36	481	115	472	33	1.62E-02	1.00E+00	3.21
Cthe_2426	OmpA/MotB	2204	1978	1925	110	101	2123	39	35	305	61	139	17	4.72E-12	1.00E+00	6.47		
Cthe_2430	RnfABCDGE type electron transport complex subunit C					3864	1518	2109	80	530	2460	576	46	4799	1643	3556	230	3.89E-02
	2.39E-03	0.35																
Cthe_2432	RnfABCDGE type electron transport complex subunit G					851	274	869	47	50	575	145	23	2055	270	909	137	2.33E-02
	1.14E-06	0.34																
Cthe_2433	RnfABCDGE type electron transport complex subunit E					807	206	794	38	166	501	146	22	1763	323	1006	124	9.44E-03	9.30E-06
	0.31																	
Cthe_2434	RnfABCDGE type electron transport complex subunit A					776	201	968	46	318	593	182	33	1927	460	1387	177	6.65E-04
	6.43E-05	0.26																
Cthe_2435	RnfABCDGE type electron transport complex subunit B					936	207	1336	43	233	720	260	28	1953	958	1825	193	6.36E-06
	3.27E-08	0.22																
Cthe_2441	DeoR family transcriptional regulator			778	1164	1977	70	62	1613	377	45	413	49	370	24	2.80E-02	1.00E+00	2.92
Cthe_2442	carbohydrate kinase FGGY	2349	5522	4228	119	214	4857	364	51	885	126	508	25	8.50E-05	1.00E+00	4.76		
Cthe_2502	tRNA-adenosine deaminase	696	280	741	50	253	759	258	49	3096	363	4733	372	2.35E-16	7.62E-11	0.13		
Cthe_2503	cupin 2 barrel domain-containing protein			765	1249	1648	87	331	3180	817	127	279	31	165	18	9.85E-06	3.30E-13	4.83
Cthe_2506	S-layer-like domain-containing protein			9407	8647	11111	138	1103	14270	1368	80	1173	511	1641	32	5.26E-03	3.48E-01	4.31
Cthe_2510	hypothetical protein	878	424	890	36	289	597	181	22	8205	481	6189	375	2.44E-39	6.81E-69	0.10		
Cthe_2518	ketol-acid reductoisomerase	1092	660	1691	49	129	1226	455	36	8088	442	3632	270	5.34E-10	1.87E-09	0.18		
Cthe_2519	putative alpha-isopropylmalate/homocitrate synthase family transferase					599	392	1646	23	86	560	177	9	4046	148	1327	73	
	5.09E-03	1.26E-11	0.32															
Cthe_2520	hypothetical protein	574	590	1590	152	354	4247	2345	605	3192	1116	5397	1006	1.84E-14	1.00E+00	0.15		
Cthe_2523	hypothetical protein	5701	2282	962	336	1317	2854	1646	326	1200	3207	6234	1084	2.19E-02	2.44E-01	0.31		
Cthe_2562	glucose-1-phosphate cytidylyltransferase			1210	1336	1225	70	79	920	117	21	247	80	196	19	2.17E-03	1.00E+00	3.68
Cthe_2589	hypothetical protein	561	228	1435	23	76	860	620	27	2936	112	2147	80	2.08E-03	2.05E-01	0.29		
Cthe_2593	peptide chain release factor 1	1153	636	1806	47	228	1838	504	43	5972	588	3322	215	1.18E-06	1.29E-03	0.22		
Cthe_2594	zinc/iron permease	497	200	1003	32	161	590	299	30	2035	176	1026	101	1.42E-02	5.07E-02	0.32		

Cthe_2598uracil phosphoribosyltransferase		1822	1169	1683	106	281	3054	433	93	5878	422	2572	322	7.63E-03	9.94E-02	0.33
Cthe_2600glycosyl transferase family protein		2197	1307	1944	68	447	2259	391	44	7271	383	2663	203	8.20E-03	5.37E-03	0.33
Cthe_2601UDP-N-acetylglucosamine 2-epimerase		1102	650	1736	43	76	1123	233	21	4584	218	1335	118	2.13E-02	1.72E-05	0.36
Cthe_2602F0F1 ATP synthase subunit A	4963	3571	3963	237	3936	9962	7587	665	26035	3776	25594	1750	1.84E-10	1.00E+00	0.14	
Cthe_2603ATP synthase F0 subunit C	981	572	813	152	283	1083	666	205	8647	1729	12620	2503	1.46E-120	3.08E-27	0.06	
Cthe_2604ATP synthase F0 subunit B	708	722	1073	66	25	808	114	27	6729	128	1033	305	1.78E-06	1.38E-28	0.22	
Cthe_2605ATP synthase F1 subunit delta	334	382	789	38	16	378	110	16	2880	111	680	142	1.96E-03	5.85E-16	0.27	
Cthe_2606F0F1 ATP synthase subunit alpha		799	864	2296	37	166	1097	379	20	7109	456	2796	154	2.09E-05	1.26E-09	0.24
Cthe_2607F0F1 ATP synthase subunit gamma		476	438	1243	34	21	547	158	15	3151	120	944	103	1.50E-02	5.05E-09	0.33
Cthe_2608F0F1 ATP synthase subunit beta	1567	1087	2452	52	380	1939	860	47	13108	1205	6780	354	7.51E-17	2.10E-08	0.15	
Cthe_2609ATP synthase F1 subunit epsilon	869	503	863	78	22	847	284	55	5076	454	1956	436	1.77E-09	5.17E-12	0.18	
Cthe_2619rod shape-determining protein Mbl		1758	1209	1866	66	269	1937	301	38	6202	511	3673	229	1.23E-03	1.60E-05	0.29
Cthe_2620hypothetical protein	2137	2289	1832	116	196	1716	51	32	319	55	156	18	9.06E-13	1.00E+00	6.44	
Cthe_2621hypothetical protein	2133	2314	1841	114	280	2239	88	43	394	76	309	25	3.45E-05	1.00E+00	4.56	
Cthe_2625(3R)-hydroxymyristoyl-ACP dehydratase		341	157	556	33	125	279	106	22	2026	192	3149	262	1.36E-14	3.83E-34	0.13
Cthe_2627pur operon repressor	779	416	1110	40	374	709	317	33	3074	379	3615	197	3.13E-07	7.68E-06	0.20	
Cthe_2628hypothetical protein	5884	2109	1335	467	92	4488	517	268	9899	1127	9353	1648	8.70E-03	3.15E-05	0.28	
Cthe_2630ribose-phosphate pyrophosphokinase		1070	358	1458	42	103	1318	356	33	4554	257	2079	158	3.73E-04	1.44E-03	0.27
Cthe_2657nucleoid protein Hbs	3164	1091	1334	288	36	3492	445	220	9320	747	2113	1060	2.11E-04	3.31E-03	0.27	
Cthe_2658RNA-binding S4 protein		235	70	134	26	5	180	26	13	412	45	260	71	4.21E-02	4.87E-07	0.37
Cthe_2662RNA binding S1	4210	2729	1918	302	688	4259	701	214	12203	1126	16855	1546	4.20E-06	6.90E-07	0.20	
Cthe_2666P-type HAD superfamily ATPase	4223	6766	6638	98	583	3751	459	28	1174	92	724	17	7.19E-07	1.00E+00	5.76	
Cthe_2675hypothetical protein	608	469	814	40	87	970	1989	146	1116	246	2045	121	1.77E-02	1.00E+00	0.33	
Cthe_2684hypothetical protein	7477	6040	5612	535	752	15417	1446	495	833	139	923	88	9.06E-08	1.75E-06	6.08	

Cthe_2693	hypothetical protein	498	203	570	36	41	490	103	21	2957	97	1594	196	4.97E-07	6.07E-18	0.18	
Cthe_2694	O-antigen polymerase	1261	659	2026	35	284	1558	362	23	6168	213	2351	116	1.96E-03	1.17E-03	0.30	
Cthe_2696	carbohydrate kinase, YjeF-like protein	1175	557	1881	33	299	1396	697	32	2378	417	3277	93	1.66E-02	2.53E-01	0.35	
Cthe_2706	ABC transporter-like protein	440	301	1063	23	32	447	198	12	5048	203	1951	140	1.44E-10	7.28E-27	0.16	
Cthe_2707	ABC-type transport system involved in multi-copper enzyme maturation permease component	261	116	739	21	36	224	171	14	2623	108	1461	123	2.08E-07	1.33E-15	0.17	
Cthe_2708	hypothetical protein	543	302	1639	24	41	437	223	10	5865	164	1907	118	5.61E-07	1.12E-26	0.20	
Cthe_2709	hypothetical protein	579	328	1528	23	38	638	225	12	5875	191	2188	119	5.46E-08	2.64E-20	0.19	
Cthe_2713	dihydroxy-acid dehydratase	1135	489	1995	30	200	1194	194	15	4466	336	1362	86	1.59E-02	1.25E-05	0.35	
Cthe_2714	acetolactate synthase, large subunit	811	464	1723	25	118	905	124	10	3986	154	1059	68	3.24E-02	1.70E-07	0.37	
Cthe_2718	preprotein translocase subunit SecE	350	242	577	68	20	325	78	30	2186	77	554	252	4.79E-03	3.44E-13	0.27	
Cthe_2719	transcription antitermination protein nusG	669	524	990	58	80	700	164	30	4576	123	1482	245	5.25E-05	2.47E-12	0.24	
Cthe_2720	50S ribosomal protein L11	1167	1170	1652	134	147	1038	227	56	12778	339	2338	779	2.75E-11	4.48E-43	0.17	
Cthe_2721	50S ribosomal protein L1	1983	1455	2218	116	178	1702	779	80	19791	610	4825	781	5.01E-17	3.29E-15	0.15	
Cthe_2722	50S ribosomal protein L10P	1942	663	1314	103	101	1741	713	98	20681	525	6908	1104	1.93E-47	7.23E-22	0.09	
Cthe_2723	50S ribosomal protein L12P	1959	761	1341	148	186	2255	1207	208	19517	1860	16885	2222	5.59E-106	2.84E-18	0.07	
Cthe_2724	DNA-directed RNA polymerase subunit beta	2549	2061	5216	37	165	1550	734	13	16114	757	6003	134	7.85E-03	3.37E-15	0.28	
Cthe_2725	DNA-directed RNA polymerase subunit beta'	2211	1774	4124	33	210	1662	845	16	18248	1067	7219	170	2.23E-06	1.05E-16	0.19	
Cthe_2726	50S ribosomal protein L7AE	447	369	417	74	126	340	192	59	4146	312	3110	701	5.25E-25	2.20E-32	0.11	
Cthe_2727	30S ribosomal protein S12	970	869	1133	99	238	749	165	45	10824	651	3710	802	2.89E-23	1.19E-77	0.12	
Cthe_2728	30S ribosomal protein S7	577	681	1084	71	15	553	122	25	10746	137	1288	539	2.83E-19	2.12E-130	0.13	
Cthe_2729	elongation factor G	5045	3424	4675	90	778	4333	2970	91	66779	2552	23888	963	1.75E-22	5.09E-10	0.09	
Cthe_2730	elongation factor Tu	20922	9235	11143	490	3995	21558	9544	602	151034	8809	93229	4606	6.16E-14	2.31E-05	0.11	

Cthe_2736	phosphoenolpyruvate--protein phosphotransferase	6705	3887	3809	121	339	21029	417	158	1563	226	647	36	4.82E-02	2.84E-03	3.36	
Cthe_2743	putative DNA-binding/iron metalloprotein/AP endonuclease	915	363	1475	38	534	1511	585	50	3797	698	4168	209	1.68E-09	1.70E-02	0.18	
Cthe_2744	lytic transglycosylase	1815	1751	1505	107	175	1915	46	39	527	71	204	30	3.17E-03	1.00E+00	3.57	
Cthe_2746	hypothetical protein	278	142	432	25	60	304	90	17	807	100	671	79	4.01E-02	1.21E-03	0.32	
Cthe_2748	SsrA-binding protein	1072	658	961	82	79	1091	219	50	2656	288	2587	271	3.68E-03	4.69E-05	0.30	
Cthe_2807	glycoside hydrolase family protein	1617	1572	2236	75	81	1665	165	27	133	31	158	8	1.48E-26	1.00E+00	9.38	
Cthe_2811	glycoside hydrolase family protein	5407	4979	5373	128	614	5448	523	54	228	62	430	10	3.87E-56	1.00E+00	12.80	
Cthe_2815	lysyl-tRNA synthetase	1720	1183	2347	46	144	1900	301	23	11930	277	4002	212	8.76E-07	1.60E-14	0.22	
Cthe_2870	hypothetical protein	430	96	1168	18	43	324	120	7	3469	122	1213	79	7.04E-05	5.70E-27	0.23	
Cthe_2874	phosphoenolpyruvate carboxykinase	12032	5249	7805	196	1662	9301	1682	114	65788	2218	21507	1058	1.24E-04	2.85E-07	0.19	
Cthe_2877	S-layer-like domain-containing protein	1488	1320	2433	42	123	2454	582	31	13181	613	4995	234	1.91E-10	2.71E-08	0.18	
Cthe_2884	imidazoleglycerol-phosphate dehydratase	655	435	636	42	110	430	119	20	2686	136	1790	168	3.48E-04	7.80E-15	0.25	
Cthe_2885	phosphoribosylaminoimidazole-succinocarboxamide synthase	1414	717	1220	54	353	1298	477	46	9082	782	8268	456	6.21E-27	6.09E-17	0.12	
Cthe_2886	imidazole glycerol phosphate synthase subunit HisH	312	129	662	25	39	246	77	11	1764	46	624	83	1.88E-02	1.47E-10	0.30	
Cthe_2891	co-chaperonin GroES	273	139	482	44	8	287	67	22	4710	62	861	410	2.49E-21	5.84E-87	0.11	
Cthe_2892	chaperonin GroEL	3941	1998	5815	102	495	6476	5624	200	62220	3706	39040	1415	3.87E-53	1.45E-03	0.07	
Cthe_2902	30S ribosomal protein S10	379	230	628	56	37	253	159	33	4012	211	3239	510	2.13E-22	1.18E-61	0.11	
Cthe_2903	50S ribosomal protein L3	395	263	987	36	13	315	79	11	4723	95	1151	196	1.09E-07	2.05E-73	0.18	
Cthe_2904	50S ribosomal protein L4	377	246	932	35	8	241	64	9	3202	101	1060	148	1.79E-04	3.09E-71	0.24	
Cthe_2905	50S ribosomal protein L23	252	152	526	37	5	188	51	12	2018	94	872	184	3.90E-05	1.31E-68	0.20	
Cthe_2906	50S ribosomal protein L2	441	292	1193	33	18	302	66	8	3074	81	819	102	1.25E-02	3.73E-35	0.32	
Cthe_2908	50S ribosomal protein L22	173	122	518	30	7	164	73	13	1125	60	528	99	2.50E-02	6.78E-12	0.30	

Cthe_2909	30S ribosomal protein S3P	369	284	944	33	22	324	92	11	2102	97	1030	102	1.89E-02	2.63E-15	0.32	
Cthe_2910	50S ribosomal protein L16	301	199	704	39	18	263	79	15	1673	78	932	132	1.23E-02	4.53E-14	0.30	
Cthe_2911	50S ribosomal protein L29	144	98	276	36	8	107	29	13	791	32	360	124	7.23E-03	1.23E-33	0.29	
Cthe_2912	30S ribosomal protein S17	180	113	373	36	13	165	68	19	1122	57	799	164	2.56E-04	1.27E-16	0.22	
Cthe_2913	50S ribosomal protein L14	237	149	482	33	19	245	84	18	1569	75	776	142	7.91E-04	2.08E-11	0.23	
Cthe_2914	50S ribosomal protein L24	177	125	383	27	7	179	48	12	1335	54	357	109	2.69E-03	1.79E-22	0.25	
Cthe_2915	50S ribosomal protein L5	429	340	703	38	26	457	187	25	3305	203	1714	211	9.34E-08	1.18E-14	0.18	
Cthe_2916	30S ribosomal protein S14	155	118	234	39	12	129	73	26	1210	97	788	255	3.87E-11	1.74E-25	0.15	
Cthe_2917	30S ribosomal protein S8	274	232	568	38	15	292	142	24	2528	119	1409	219	1.99E-07	7.83E-18	0.17	
Cthe_2918	50S ribosomal protein L6P	232	241	670	29	14	299	102	14	2787	91	935	148	1.12E-05	9.31E-22	0.20	
Cthe_2919	50S ribosomal protein L18P	189	139	566	34	19	168	79	15	1856	41	532	138	2.09E-03	9.97E-18	0.25	
Cthe_2920	30S ribosomal protein S5	429	236	722	39	35	426	139	23	3558	105	1528	218	1.20E-07	5.81E-20	0.18	
Cthe_2921	50S ribosomal protein L30	192	144	298	50	26	189	68	31	1820	82	837	330	6.72E-10	5.29E-27	0.15	
Cthe_2922	50S ribosomal protein L15	416	298	730	46	43	467	177	31	4520	221	2081	338	2.54E-14	1.34E-26	0.14	
Cthe_2923	protein translocase subunit secY/sec61 alpha			1146	866	1695	41	312	1365	686	39	12461	607	6362	329	5.72E-24	2.88E-11	0.12
Cthe_2924	adenylate kinase	531	449	988	43	43	639	318	33	6578	336	3163	337	6.69E-20	4.51E-21	0.13		
Cthe_2925	methionine aminopeptidase, type I			1300	967	1362	67	364	1348	1729	124	8625	1223	6376	515	1.06E-21	4.61E-02	0.13
Cthe_2926	hypothetical protein	82	84	361	27	4	88	33	9	1048	24	198	99	4.03E-03	1.96E-45	0.27		
Cthe_2927	translation initiation factor IF-1	155	102	298	36	26	203	170	46	1429	176	1323	312	4.84E-20	2.36E-09	0.12		
Cthe_2928	50S ribosomal protein L36P	99	51	130	35	11	194	382	165	1679	313	2595	953	0.00E+00	7.03E-06	0.04		
Cthe_2929	30S ribosomal protein S13	240	212	566	39	9	195	38	11	1697	52	510	130	1.88E-02	3.67E-45	0.30		
Cthe_2930	30S ribosomal protein S11	275	250	666	42	43	227	100	18	2501	81	914	184	2.27E-04	5.65E-20	0.23		
Cthe_2931	30S ribosomal protein S4P	722	623	1095	55	93	779	291	36	6627	336	2948	347	1.51E-12	5.02E-18	0.16		

Cthe_2932	DNA-directed RNA polymerase subunit alpha			1163	1031	2102	64	39	1074	351	29	13953	204	2190	360	3.11E-10	1.56E-30	0.18
Cthe_2933	50S ribosomal protein L17	2670	1175	1662	149	323	2849	2486	276	21219	2406	21585	1970	3.08E-82	1.29E-05	0.08		
Cthe_2938	glucokinase	850	504	1288	39	71	867	130	17	3841	159	1159	118	1.07E-02	2.23E-08	0.33		
Cthe_2940	CarD family transcriptional regulator			2886	2877	1961	235	215	3772	275	124	894	63	535	70	1.17E-02	4.85E-01	3.36
Cthe_2972	glycoside hydrolase family protein			5086	5671	10483	148	1308	10303	1748	105	1619	189	1323	35	5.40E-03	1.50E-01	4.23
Cthe_2989	glycosyltransferase	4864	3650	4679	64	753	4640	1905	50	68357	2419	27706	713	3.78E-24	7.04E-27	0.09		
Cthe_2996	PfpI family intracellular peptidase			1295	704	1259	80	1085	1277	503	91	5680	1339	10265	723	4.51E-31	1.10E-09	0.11
Cthe_2998	ABC transporter-like protein	743	308	1892	24	149	723	218	11	6711	253	2047	113	8.19E-07	3.14E-19	0.21		
Cthe_3000	phosphate transporter	1232	801	1544	48	467	979	288	29	8838	634	5700	321	1.43E-15	1.28E-23	0.15		
Cthe_3001	hypothetical protein	775	505	940	50	11	753	77	19	4098	59	523	157	9.64E-03	3.62E-13	0.32		
Cthe_3002	hypothetical protein	652	368	757	46	12	507	113	20	1970	154	1024	131	3.28E-02	1.32E-08	0.35		
Cthe_3003	hydrogenase, Fe-only	1522	1043	2327	36	270	1797	420	22	12766	362	3278	181	2.70E-08	8.34E-11	0.20		
Cthe_3004	ferredoxin 1136	868	1785	36	189	1399	220	19	10256	224	2227	181	4.92E-08	9.36E-17	0.20			
Cthe_3017	hydrogenase accessory protein HypB			392	216	500	24	68	302	151	16	1707	218	1122	113	6.84E-05	5.25E-09	0.21
Cthe_3018	hydrogenase expression/synthesis, HypA			137	82	227	18	20	104	27	8	723	24	245	62	5.66E-03	6.96E-21	0.29
Cthe_3020	ech hydrogenase subunit E	663	497	1108	30	75	430	131	11	2716	125	999	78	4.70E-02	2.72E-10	0.38		
Cthe_3021	ech hydrogenase subunit D	424	275	410	44	57	206	93	20	1272	142	742	143	2.08E-02	2.39E-09	0.31		
Cthe_3022	ech hydrogenase subunit C	130	78	366	18	16	66	16	4	869	29	216	55	4.02E-02	5.72E-96	0.33		
Cthe_3023	ech hydrogenase subunit B	426	177	1048	26	70	266	103	9	2677	126	1022	96	2.60E-03	7.37E-21	0.27		
Cthe_3024	ech hydrogenase subunit A	1715	869	1701	31	661	1023	925	32	12203	1095	12297	297	4.28E-34	2.62E-13	0.10		
Cthe_3027	citrate synthase	6523	4836	5247	174	1443	7141	1216	115	25835	1545	13176	656	3.77E-02	4.23E-03	0.27		
Cthe_3028	histidine decarboxylase	899	532	1828	38	103	1961	1647	79	5454	405	3879	181	6.52E-07	7.26E-01	0.21		
Cthe_3029	CheW protein	1124	887	772	81	46	1041	56	31	190	75	179	27	1.63E-02	1.00E+00	3.00		
Cthe_3035	D-3-phosphoglycerate dehydrogenase			2505	2546	3880	109	213	2737	426	45	13680	411	3834	326	4.33E-02	1.07E-07	0.33

Cthe_3064	polysaccharide biosynthesis protein	555	308	1245	18	136	735	178	11	2675	148	1483	58	8.20E-03	1.30E-04	0.31	
Cthe_3075	von Willebrand factor, type A	716	673	1688	27	92	793	508	20	15796	900	6395	320	2.51E-57	2.01E-55	0.08	
Cthe_3079	cellulosome anchoring protein cohesin subunit	3376	9028	10170	159	183	4369	381	34	1864	139	850	31	1.49E-04	1.00E+00	5.13	
Cthe_3082	hypothetical protein	79	93	242	24	6	87	50	13	534	49	214	77	7.69E-03	2.10E-09	0.31	
Cthe_3083	hypothetical protein	215	218	479	46	38	317	250	53	931	88	783	145	3.73E-02	1.76E-01	0.32	
Cthe_3089	hypothetical protein	3115	2280	2428	286	164	3844	401	165	705	84	673	86	1.34E-02	4.15E-01	3.33	
Cthe_3092	hypothetical protein	6337	3310	1789	795	1180	5193	1467	687	5565	3215	8899	2665	2.45E-02	1.09E-01	0.30	
Cthe_3105	exsB protein	627	269	668	33	57	497	137	18	1944	124	969	101	2.10E-02	1.61E-05	0.33	
Cthe_3106	putative 6-pyruvoyl tetrahydropterin synthase	418	206	427	39	62	391	105	26	1397	102	778	134	1.30E-02	5.02E-05	0.29	
Cthe_3110	UBA/THIF-type NAD/FAD binding fold protein	615	233	1004	34	121	623	168	21	1565	296	1222	104	1.64E-02	3.76E-04	0.33	
Cthe_3112	glycosidase	840	652	1706	45	223	943	348	29	2374	251	3406	133	1.38E-02	2.71E-03	0.34	
Cthe_3114	group 1 glycosyl transferase	833	527	1945	39	83	982	924	44	3690	551	3752	162	4.01E-05	6.35E-02	0.24	
Cthe_3115	hypothetical protein	705	420	1803	40	157	928	1036	57	2676	744	3602	181	5.75E-06	1.59E-01	0.22	
Cthe_3116	mannose-6-phosphate isomerase, class I	260	175	1107	19	21	370	439	21	585	248	1508	55	4.73E-02	2.43E-01	0.35	
Cthe_3117	zinc/iron permease	2455	1588	1121	98	624	1278	946	86	2428	823	3605	246	4.42E-02	2.68E-01	0.40	
Cthe_3119	FMN-binding flavin reductase-like protein	611	466	633	37	41	601	446	40	2010	218	1243	124	6.18E-03	1.14E-01	0.30	
Cthe_3120	pyruvate flavodoxin/ferredoxin oxidoreductase-like protein	4204	2873	4687	47	949	5095	2129	46	19431	2697	9456	226	1.60E-04	4.23E-02	0.21	
Cthe_3124	AMP-dependent synthetase and ligase	2968	2426	3720	79	370	3089	494	38	20533	569	4991	341	4.27E-04	4.52E-12	0.23	
Cthe_3142	hypothetical protein	260	164	462	25	25	322	189	25	2159	333	3270	271	6.72E-29	9.40E-27	0.09	
Cthe_3150	adenosylcobinamide-phosphate synthase	756	287	1286	31	126	1448	291	29	2253	167	1515	83	3.76E-02	2.03E-01	0.37	
Cthe_3154	hypothetical protein	150	56	494	17	19	224	55	9	1373	38	297	66	5.85E-03	1.81E-10	0.26	
Cthe_3155	beta-lactamase-like protein	524	176	919	31	50	927	279	32	3841	122	1250	152	3.50E-06	7.99E-04	0.20	

Cthe_3157	pyruvate carboxyltransferase	633	623	1698	30	83	758	159	12	3561	140	1481	81	3.53E-02	5.79E-08	0.37	
Cthe_3158	aconitate hydratase	1609	1039	2931	41	528	2016	744	33	13303	852	6276	238	5.84E-12	1.26E-07	0.17	
Cthe_3159	GntR family transcriptional regulator	918	583	1446	62	123	1330	471	56	6638	588	4067	387	1.37E-12	9.66E-08	0.16	
Cthe_3163	carbohydrate-binding family 25 protein	12795	6506	4209	686	93	14398	179	363	2710	185	1229	190	3.30E-02	7.48E-01	3.61	
Cthe_3169	enoyl-[acyl-carrier-protein] reductase [NADH]	834	378	1165	44	136	1033	216	31	5713	197	2901	245	1.92E-09	5.25E-12	0.18	
Cthe_3171	S-layer-like domain-containing protein	535	140	2143	16	3	377	232	5	4212	200	2053	57	1.01E-03	2.67E-20	0.28	
Cthe_3172	hypothetical protein	461	177	1479	21	95	636	467	20	4763	634	4118	160	4.90E-19	4.05E-11	0.13	
Cthe_3177	monogalactosyldiacylglycerol synthase	726	895	1800	38	44	636	129	10	9507	149	1713	187	1.55E-07	4.93E-80	0.20	
Cthe_3178	hypothetical protein	52741	44748	19710	1072	16022	85227	1370	822	1126	322	1740	51	7.32E-111	1.12E-58	21.02	
Cthe_3183	potassium transporter peripheral membrane component	538	187	1002	17	10	612	496	20	2728	124	1428	66	1.03E-03	8.75E-02	0.26	
Cthe_3191	chromosome partitioning ATPase	1865	1281	2002	60	386	1873	333	34	6281	291	3617	179	8.14E-03	4.76E-04	0.34	
Cthe_3194	hypothetical protein	2220	1970	2426	146	939	4176	152	107	350	52	224	24	1.30E-10	2.91E-05	6.08	
Cthe_3196	hypothetical protein	681	292	1298	23	85	491	120	9	6263	213	2586	141	3.00E-11	1.18E-65	0.16	
Cthe_3200	alanyl-tRNA synthetase	952	361	2904	22	163	881	260	9	5073	331	2463	66	9.44E-03	3.93E-10	0.33	
Cthe_3306	hypothetical protein	36	26	65	16	6	42	26	15	180	9	133	63	5.85E-07	5.99E-06	0.25	
Cthe_3321	hypothetical protein	239	135	215	39	12	200	59	23	918	45	347	134	1.15E-02	4.44E-07	0.29	
Cthe_3324	hypothetical protein	69	74	137	23	8	154	477	129	655	112	967	233	9.72E-47	5.68E-01	0.10	
Cthe_3328	hypothetical protein	745	1041	361	250	1261	925	212	311	106	27	114	52	2.46E-07	9.28E-12	4.81	
Cthe_3332	helicase-like protein	100	75	104	28	0	111	18	14	473	26	179	105	5.51E-05	7.87E-21	0.27	
Cthe_3333	hypothetical protein	1494	1010	699	302	90	986	265	159	4059	585	5707	1552	1.91E-08	2.40E-19	0.19	
Cthe_3337	hypothetical protein	88	136	375	59	3	270	622	208	85	11	73	28	4.83E-02	5.66E-28	2.11	
Cthe_3345	hypothetical protein	222	221	337	78	35	1364	42	128	78	17	45	27	1.68E-03	4.30E-10	2.89	
Cthe_3350	hypothetical protein	749	1011	740	173	221	834	504	166	546	437	2635	461	3.60E-02	2.21E-01	0.38	

Cthe_3351hypothetical protein	2197	941	1220	300	650	3479	6017	1495	7001	2617	17773	3346	3.85E-52	1.00E+00	0.09
Cthe_3376hypothetical protein	85	34	157	27	11	87	48	23	156	28	216	68	3.02E-02	2.81E-02	0.40
Cthe_3380hypothetical protein	292	122	341	62	107	515	1180	341	3151	637	6668	1410	2.28E-169	2.21E-02	0.04
Cthe_3383putative autoinducer prepeptide			3964	2211	2714	1035	1797	4315	2170	1394	4106	2072	7954	3309	2.36E-02	1.00E+00	0.31
Cthe_3384hypothetical protein	1250	855	323	249	116	474	85	78	48	30	118	40	7.46E-15	1.84E-01	6.23
Cthe_3385hypothetical protein	602	387	451	222	14	580	149	145	101	29	114	71	4.82E-03	1.95E-01	3.13
Cthe_3408hypothetical protein	433	391	300	95	59	266	273	92	671	205	977	287	3.88E-02	8.22E-02	0.33
Cthe_3419hypothetical protein	137	134	186	41	22	101	84	32	1397	129	1627	436	2.39E-36	1.11E-50	0.09

Apêndice G. Valores de expressão dos genes de *C. thermocellum* que expressam proteínas carboidrato ativas e relacionadas ao celulossoma. Cluster hierárquico em um total de 6. Lista apresente o gene, fontes de carbono, cluster, posição do cluster, nome da enzima, domínios presentes, celulossomal ou não, e atividade.

gene	palha	celulose	bagaço	Cluster	Cluster Ordem	nome	domínio	celulossomal	atividade
Cthe_001513	7	22	1	1	/	GH43 GH54 D1		S	α-L-arabinofuranosidase B
Cthe_003217	8	53	2	45	Ctman	GH 26 CBM6 D1		S	β-mananase
Cthe_00404	27	20	1	23	CelI	GH9 CBM3		N	cellulose 1,4-beta-cellobiosidase
Cthe_004365	90	78	3	79	CelN	GH 9 CBM3 d1		S	endo-1,4-β-glucanase
Cthe_00444	3	18	1	6	CseP	Coth D1	S		
Cthe_00717	17	25	1	36	CelY	GH48 CBM3		N	cellulose 1,4-beta-cellobiosidase
Cthe_0109169	87	231	3	85	/	D1	S		
Cthe_019042	28	71	2	56	/	Serpin FN3 D1		S	
Cthe_019131	13	54	2	48	proteinase inhibitor I4			D1	S
Cthe_021121	42	20	2	66	LicB	GH16 D1	S	endo-1,3-1,4-β-glucanase	
Cthe_021212	78	36	2	72	BglA	GH1	N	beta-glucosidase	

Cthe_02397	6	25	1	14	/	D1	S		
Cthe_024611	9	25	1	4	PL11	PL11 CBM6 D1	S	Ramnogalacturona liase	
Cthe_02589	30	25	1	24	Doc258	D1	S		
Cthe_026939	162	137	5	89	CelA	GH8 D1	S	endo-1,4-β-glucanase	
Cthe_02705	12	23	1	40	ChiA	GH18 D1	S	endo-quitinase	
Cthe_027129	154	65	5	90	/	CBM 3	N		
Cthe_027423	16	68	2	46	CelP	GH9 D1	S	endo-1,4-β-glucanase	
Cthe_03225	17	22	1	42	/	GH3	N		
Cthe_04058	23	29	1	39	CelL	GH5 D1	S	endo-1,4-β-glucanase	
Cthe_0412212	188	416	4	87	CelK	GH9 CBM4 D1	S	exo-1,4-β-glucanase	
Cthe_041399	95	215	3	82	CbhA	GH9 CBM4 CBD3b D1	S	exo-1,4-β-glucanase	
Cthe_043329	30	70	2	52	/	GH9 CBM3c D1	S	endo-1,4-β-glucanase	
Cthe_043546	174	60	5	93	Cel124A	GH124	S	endo-1,4-β-glucanase	
Cthe_04382	3	5	1	12	/	D1	S		
Cthe_045271	259	124	5	94	OlpC	1 coesina tipo I	S	estrutural	
Cthe_053657	212	87	5	95	CelB	GH5 D1	S	endo-1,4-β-glucanase	
Cthe_054322	56	48	2	70	CelF	GH9 CBM3 D1	S	endo-1,4-β-glucanase	
Cthe_057811	27	40	1	26	CelR	GH9 CBM3 D1	S	endo-1,4-β-glucanase	
Cthe_062432	33	90	2	57	CelJ	GH9 GH44 CBM 30 D1	S	endo-1,4-β-glucanase	
Cthe_062532	45	103	2	59	CelQ	GH9 CBM3 D1	S	endo-1,4-β-glucanase	
Cthe_06404	8	17	1	18	/	D1	S		
Cthe_06603	7	14	1	20	/	GH81 D1	S	endo-1,3-β-glucanase	
Cthe_066110	35	23	1	25	Ct1,3Gal43A	GH43 CBM 13	S	exo-1,3-β-galactanase	
Cthe_072910	19	22	1	38	/	D1	S		

Cthe_07358	106	26	2	76	ScaE	7 coesinas tipo II	S	estrutural
Cthe_073623	29	50	2	54	/	1 coesina tipo II	S	estrutural
Cthe_074512	27	33	1	27	CelW	GH9 CBM 3 D1	S	endo-1,4-β-glucanase
Cthe_079714	5	25	1	2	CelE, CtCE2	GH5 CE2 D1	S	endo-1,4-β-glucanase/acetilxilan esterase
Cthe_079848	7	157	2	63	/	CE3 CE3 D1	S	acetilxilana esterase
Cthe_082133	12	66	2	50	Ctman5A	GH5 CBM32 D1	S	β-mannanase
Cthe_082510	41	29	2	68	CelD	GH9 D1	S	endo-1,4-β-glucanase
Cthe_091228	52	41	2	71	XynY	GH10 CE1 CBM22 CBM22	S	endo-1,4-β-xilanase/ feruloil esterase
Cthe_1080211	53	175	3	86	/	CBM25	N	
Cthe_125660	65	58	2	61	BglB	GH3	N	beta glucosidase B
Cthe_125731	17	56	2	49	/	CBM3 CBM4_9	N	
Cthe_127116	19	21	1	29	CtAbf43A	GH43 CBM6 D1	S	
Cthe_130711	48	38	2	69	SdbA	1 coesina tipo II	S	estrutural
Cthe_139864	22	79	2	60	XghA	GH74 D1	S	xiloglucanase
Cthe_14007	13	26	1	41	/	GH53 D1	S	endo-1,4-β-galactanase
Cthe_14283	4	17	1	7	bgl	GH1	N	beta glucosidase
Cthe_14727	9	21	1	15	CelH	GH5 GH26 CBM11 D1	S	endo-1,4-β-glucanase
Cthe_161310	1	1	1	13	/	GH18	N	
Cthe_178721	31	56	2	55	/	GH15	N	
Cthe_180611	10	33	1	5	/	D1	S	
Cthe_183860	228	87	5	96	XynC	GH10 CBM22 D1	S	endo-1,4-β-xilanase
Cthe_189091	82	80	3	80	/	D1	S	
Cthe_19115	17	17	1	43	/	CBM6	N	
Cthe_196344	52	63	2	62	XynZ	GH10 CE1 CBM6 D1	S	endo-1,4-β-xilanase/ feruloil esterase

Cthe_203821	29	68	2	53	/	D1	S			
Cthe_2089293	249	724	6	97	CelS	GH48 D1	S	exo-1,4-β-glucanase		
Cthe_214710	28	30	1	28	CelO	GH5 CBM3 D1	S	exo-1,4-β-glucanase		
Cthe_214811	69	35	2	73	/	CBM_NC	N			
Cthe_21676	7	18	1	16	/	GH126	N			
Cthe_217926	91	55	2	77	/	PL1 PL9 CBM 6 D1	S	Pectate lyase		
Cthe_219326	99	32	2	78	CtXyl5A	GH5 CBM6 CBM13 CBM62 D1	S	arabinoxilanase		
Cthe_21948	4	14	1	21	/	CE1 CBM6 D1	S	acetilxilana esterase		
Cthe_21953	1	15	1	8	/	CBM6- GH141	S			
Cthe_21962	0	13	1	10	/	GH43 CBM6 D1	S	α-L-arabinofuranosidase		
Cthe_227111	3	27	1	3	/	D1	S			
Cthe_236014	16	36	1	33	CelU	GH8 CBM3b CBM3c D1	S	endo-1,4-β-glucanase		
Cthe_25483	9	15	1	19	CtArf51A	GH51	N	α-N-arabinofuranosidase		
Cthe_25496	5	10	1	22	/	D1	S			
Cthe_259014	21	25	1	30	XynD	GH10 CBM 22 D1	S	endo-1,4-β-xilanase		
Cthe_276014	22	31	1	32	CelV	GH9 CBM3b CBM3c D1	S	endo-1,4-β-glucanase		
Cthe_276112	22	21	1	31	LecA	GH9 CBM3 D1	S	endo-1,4-β-glucanase		
Cthe_280525	19	45	1	34	/	CBM16	N			
Cthe_280727	8	75	2	47	CelC	GH5	N	endo-1,4-β-glucanase		
Cthe_280922	22	34	1	35	Lic16A	Lic A GH16 CBM 4	N			
Cthe_281154	10	128	2	64	ManA	GH26 CBM D1	S	β-mananase		
Cthe_281237	33	91	2	58	CelT	GH9 D1	S	endo-1,4-β-glucanase		
Cthe_287230	156	78	5	91	CelG	GH5 D1	S	endo-1,4-β-glucanase		
Cthe_28794	14	17	1	44	/	CE_NC D1	S			

Cthe_28959	57	23	2	74	/	GH18	N		
Cthe_29492	1	12	1	11	CtPMe	CE8 D1	S	pectinesterase	
Cthe_29505	6	22	1	17	pelB2	PL1 CBM 35 D1	S	Pectate lyase	
Cthe_2972105	35	148	3	81	XynA/U	GH11 CE4 CBM6 D1	S	endo-1,4-β-xilanase	
Cthe_301220	35	28	2	67	/	GH30 CBM6 D1	S		
Cthe_30638	17	25	1	37	/	CBM 4_9 CE7	N	acetyl xylan esterase	
Cthe_3077147	242	335	4	88	CipA	9 coesinas tipo I	S		
Cthe_3078112	133	167	3	84	OlpB	7 coesinas tipo II	S		
Cthe_307934	31	159	2	65	Orf2p	2 cosinas tipo II	S		
Cthe_308092	99	148	3	83	OlpA	1 coesina tipo I	S		
Cthe_311229	133	45	5	92	/	GH130 DUF 377	N		
Cthe_313242	15	52	2	51	/	D1	S		
Cthe_31369	57	21	2	75	/	D1	S		
Cthe_31412	1	16	1	9	/	CE12 CBM6 D1	S		
Cthe_3163363	190	686	6	98	/	CBM25	N		

Apêndice H.A. Tabela dos genes diferencialmente expressos de *M. thermoacetica* em palha em comparação com celulose. Na tabela estão presentes os genes, descrição, número de fragmentos de cada triplicata biológica para bagaço (B1, B2, B3), palha (P1, P2, P3) e celulose (C1, C2, C3), números de expressão calculados e *qvalue* para cada comparação de cada gene. FC, indica o número de vezes que o gene foi aumentado ou diminuído. Valores maiores que 1, os genes são regulados positivamente para a condição palha.

Gene	Descrição	B1	B2	B3	Expressão	P1	P2	P3	Expressão	C1	C2	C3	Expressão	qValue B vs C		qValue P vs C		FC	
Moth_0011	DeoR family transcriptional regulator					8084	11142	7612	504	514	1109	158	35	967	1875	839	251	8.80E-01	5.98E-09
0.14																			
Moth_0013	PTS fructose IIC component				9305	41728	6553	627	285	1481	158	20	2468	599	871	84	5.62E-09	8.92E-03	0.24
Moth_0014	putative PTS IIA-like nitrogen-regulatory protein PtsN					2936	14498	1871	629	43	532	69	21	924	163	327	83	6.28E-14	
6.26E-03	0.25																		

Moth_0028	hypothetical protein	2533	3046	1721	338	3892	2792	662	382	257	173	728	106	1.92E-02	6.70E-03	3.60	
Moth_0029	DNA replication and repair protein RecR	3690	4437	2599	261	6502	4169	867	308	352	306	1064	88	7.27E-02	2.60E-02	3.50	
Moth_0030	hypothetical protein	1703	1671	1162	301	2790	2224	333	378	269	95	649	116	7.80E-02	1.87E-02	3.26	
Moth_0031	SigmaK-factor processing regulatory BofA	2288	1973	1621	308	4690	3097	534	473	430	141	1152	152	3.54E-01	5.00E-02	3.11	
Moth_0038	FAD linked oxidase-like protein	10051	5946	3326	205	13335	16613	446	307	1956	145	1656	60	5.06E-02	2.81E-04	5.12	
Moth_0039	hypothetical protein	10836	6149	3408	226	14916	18085	574	358	2344	202	2258	82	1.96E-01	5.86E-03	4.37	
Moth_0144	BioY protein	1181	3017	1028	136	2372	950	177	96	130	105	320	30	3.85E-05	1.88E-02	3.20	
Moth_0145	hypothetical protein	660	1548	589	114	1533	659	174	106	76	49	227	27	2.44E-04	1.29E-03	3.93	
Moth_0197	radical SAM family protein	7457	5987	3357	123	8312	4332	1381	120	761	192	1474	29	3.58E-03	6.32E-03	4.14	
Moth_0221	phosphoglucomutase	4543	4511	2449	118	8846	4564	988	162	792	254	1405	44	1.70E-01	2.56E-02	3.68	
Moth_0222	hypothetical protein	3051	3246	620	235	1695	1083	339	123	199	48	361	34	2.35E-14	2.14E-03	3.62	
Moth_0374	NADPH-dependent FMN reductase	5092	5043	4283	368	1134	11037	548	301	1043	153	699	83	6.12E-04	1.44E-02	3.63	
Moth_0387	5-methyltetrahydrofolate--homocysteine methyltransferase	836	332	365	35	683	4412	101	107	189	52	238	20	1.00E+00	1.15E-07	5.35	
Moth_0443	biotin/lipoate A/B protein ligase	1008	1050	468	23	2926	2199	453	56	214	93	550	14	1.00E+00	7.99E-04	4.00	
Moth_0444	geranylgeranyl reductase	581	634	333	22	1877	1714	362	62	178	50	416	15	1.00E+00	1.81E-04	4.13	
Moth_0445	radical SAM family protein	503	569	392	20	1674	1867	384	60	227	86	512	20	1.00E+00	3.66E-02	3.00	
Moth_0479	metallophosphoesterase	4733	4967	2707	160	10524	4646	1094	232	495	383	1368	60	1.77E-01	1.50E-02	3.87	
Moth_0591	hypothetical protein	2080	1743	2277	289	6777	2592	570	521	409	45	644	79	2.69E-03	3.16E-11	6.59	
Moth_0611	tagatose-6-phosphate kinase	2647	1294	723	71	807	3406	46	58	208	115	262	20	3.23E-03	4.70E-02	2.90	
Moth_0612	periplasmic binding protein/LacI transcriptional regulator	5289	2674	1523	139	1653	6141	30	100	519	150	517	33	5.61E-04	4.70E-02	3.03	
Moth_0616	hypothetical protein	698	452	234	42	570	848	275	64	141	29	279	22	1.00E+00	4.28E-02	2.91	
Moth_0723	hypothetical protein	1945	1309	1737	265	8724	3525	1565	893	726	207	1750	239	1.00E+00	1.90E-02	3.74	

Moth_0851	sporulation factor SpoIIGA	16679	8947	4337	468	28633	17200	1591	766	2180	188	5336	172	2.76E-01	1.16E-02	4.45	
Moth_0852	sporulation sigma factor SigE	12367	5293	2687	398	17459	12557	626	601	1626	148	3602	152	2.58E-01	2.33E-02	3.95	
Moth_0873	hypothetical protein	15487	6279	4205	250	50495	44073	873	891	4236	459	6041	159	1.00E+00	2.34E-03	5.60	
Moth_0887	ATPase	20376	11875	4594	190	20297	9505	1612	174	1978	155	3616	42	5.44E-03	1.80E-02	4.14	
Moth_0919	phosphodiesterase	4247	6999	3077	372	6129	3033	999	303	392	298	1023	92	3.74E-03	4.63E-02	3.29	
Moth_0929	ABC transporter	6193	5228	642	240	3986	5058	1231	235	386	305	943	69	2.22E-02	3.59E-02	3.41	
Moth_0931	ferric uptake regulator family protein	8705	13071	723	734	2144	2257	448	172	156	157	357	50	2.61E-71	1.07E-02	3.44	
Moth_0958	RNA polymerase sigma factor SigX	82217	150605	104467	10115	26880	8907	3234	1333	1440	617	3383	295	2.40E-292	7.66E-03	4.52	
Moth_0961	ABC-type transport system involved in multi-copper enzyme maturation, permease component	5592	10330	6769	378	4709	1385	540	126	224	149	674	33	2.27E-38	3.29E-03	3.82	
Moth_1010	acyltransferase 3	923	1418	2085	53	3696	1474	273	70	223	139	563	22	1.00E+00	2.34E-02	3.18	
Moth_1129	hypothetical protein	3966	8059	10353	547	21129	6584	1020	744	1142	571	2506	198	2.18E-01	4.26E-02	3.76	
Moth_1144	stationary phase survival protein SurE	6677	4471	2692	255	20970	13742	1264	683	1522	385	3519	164	1.00E+00	1.72E-02	4.16	
Moth_1145	hypothetical protein	1400	954	550	269	5036	3286	190	791	394	92	855	203	1.00E+00	1.85E-03	3.90	
Moth_1153	hypothetical protein	2925	2443	1215	203	2610	1568	411	161	302	81	635	52	1.23E-03	3.33E-02	3.10	
Moth_1239	hypothetical protein	3806	3891	151	488	374	696	89	74	86	14	69	18	0.00E+00	2.28E-07	4.11	
Moth_1240	hypothetical protein	5302	5172	245	559	471	922	72	75	99	27	94	21	0.00E+00	5.28E-04	3.57	
Moth_1269	hypothetical protein	19420	6558	600	609	1459	3034	204	106	446	74	343	34	4.59E-114	2.85E-02	3.12	
Moth_1292	glutamate synthase (NADPH) GltB2 subunit	2513	1536	1525	53	4856	2324	646	85	401	144	997	25	1.00E+00	2.34E-02	3.40	
Moth_1316	5-methyltetrahydrofolate--homocysteine methyltransferase	487	165	169	18	560	3190	50	77	99	22	115	9	1.00E+00	7.65E-26	8.56	
Moth_1317	hypothetical protein	1114	429	390	23	709	7097	66	82	208	28	154	8	1.00E+00	5.82E-32	10.25	

Moth_1330	1-acyl-sn-glycerol-3-phosphate acyltransferase	2290	2205	1767	156	6653	3033	664	284	333	222	1022	76	3.50E-01	9.29E-03	3.74		
Moth_1356	nucleoside recognition protein	2273	1392	855	107	5157	3233	347	216	442	120	823	56	3.66E-01	3.11E-03	3.86		
Moth_1401	thiamine biosynthesis protein ThiC	3054	2014	1727	75	7628	3834	1694	180	743	282	1530	51	1.00E+00	4.27E-02	3.53		
Moth_1402	hypothetical protein	1154	715	613	135	1880	1097	708	272	308	68	642	90	7.81E-01	3.88E-02	3.02		
Moth_1404	manganese transport transcriptional regulator	2755	1364	657	145	2150	1749	107	122	308	25	623	42	4.79E-03	4.40E-02	2.90		
Moth_1405	nucleoside recognition protein	7206	6827	1936	167	5639	4491	362	110	659	96	1349	33	3.50E-05	3.38E-02	3.33		
Moth_1417	CoA enzyme activase	10558	6152	9683	482	30273	17353	1451	928	1110	423	2825	141	6.12E-02	5.09E-07	6.58		
Moth_1418	2-hydroxyglutaryl-CoA dehydratase subunit D	15123	8392	13437	403	45739	24558	1809	810	1530	714	4120	128	1.68E-01	2.39E-05	6.33		
Moth_1439	SirA-like protein	1909	1393	1112	291	2930	2752	67	364	181	44	381	66	1.66E-05	7.67E-09	5.52		
Moth_1440	sugar kinase	5976	4066	3158	216	7288	7187	182	231	599	137	1122	50	9.34E-04	3.24E-04	4.62		
Moth_1441	hypothetical protein	2173	1577	1207	145	2306	2636	38	137	199	47	322	28	1.47E-07	1.42E-06	4.89		
Moth_1442	hypothetical protein	6465	5593	3664	214	8992	7848	376	231	608	153	1107	43	4.25E-05	6.66E-06	5.37		
Moth_1446	2-hydroxyglutaryl-CoA dehydratase subunit D	1613	2100	3499	90	5228	2504	95	98	256	124	536	22	6.37E-04	1.55E-04	4.45		
Moth_1447	cobalamin B12-binding	2916	4496	5927	115	9477	3933	378	123	493	246	891	28	2.53E-03	1.02E-03	4.39		
Moth_1491	hypothetical protein	538	1028	459	120	1055	442	169	114	51	49	196	35	7.87E-03	1.81E-02	3.26		
Moth_1532	hypothetical protein	3918	2622	2089	100	9976	5563	415	189	844	92	1875	47	3.47E-01	8.06E-03	4.02		
Moth_1533	stage III sporulation protein AD	1528	1125	705	124	3420	1750	183	206	275	37	603	50	9.55E-02	5.20E-04	4.12		
Moth_1534	hypothetical protein	675	586	356	113	1483	883	103	180	141	18	307	48	1.34E-01	2.64E-03	3.75		
Moth_1535	stage III sporulation protein SpoAB	2491	2142	1331	166	6786	3179	539	315	635	76	1332	84	4.01E-01	8.32E-03	3.75		
Moth_1536	hypothetical protein	10278	9845	5035	348	25011	10917	1715	553	1809	220	3786	118	1.74E-01	3.69E-03	4.69		
Moth_1585	hypothetical protein	792	807	413	115	1240	1516	60	154	134	53	304	50	1.16E-01	1.84E-02	3.08		

Moth_1624	redox-active disulfide protein 2	3695	3780	2790	606	5672	3388	452	591	309	123	608	113	5.84E-07	1.59E-06	5.23
Moth_1625	selenophosphate synthase	10553	11773	8675	457	23442	10596	1620	557	938	435	2280	99	2.05E-03	4.03E-05	5.63
Moth_1659	hypothetical protein	4868	7465	251	405	921	828	102	60	108	23	163	15	0.00E+00	2.70E-04	4.00
Moth_1765	hypothetical protein	3512	682	733	49	3404	2425	389	68	351	32	591	14	1.00E+00	2.36E-05	4.86
Moth_1821	hypothetical protein	624	514	556	79	1764	739	178	139	73	75	217	38	2.05E-01	2.39E-03	3.66
Moth_1826	lactate dehydrogenase	5460	5390	1344	187	4070	3213	846	142	474	154	769	38	3.05E-05	6.34E-03	3.74
Moth_1828	hypothetical protein	4807	2699	2933	369	8507	8152	2449	816	721	289	1720	175	3.73E-01	1.04E-03	4.66
Moth_1829	peroxiredoxin-like	1262	1913	753	114	1253	870	135	69	61	85	162	22	8.75E-07	2.21E-02	3.14
Moth_1842	hypothetical protein	11902	9468	5680	387	33165	27811	3946	990	1862	594	5681	190	7.90E-01	3.05E-03	5.21
Moth_1857	copper amine oxidase-like	24865	8357	6844	885	48974	57002	3702	2442	3040	733	8421	432	1.00E+00	3.38E-03	5.65
Moth_1873	hypothetical protein	1650	943	531	139	1428	826	271	133	258	16	466	47	1.73E-02	3.90E-02	2.83
Moth_2084	ArsR family transcriptional regulator	1774	1971	882	216	12894	4476	1049	941	117	84	489	59	2.17E-03	3.34E-91	15.95
Moth_2085	hypothetical protein	2675	3059	1522	157	20689	7811	2003	732	314	171	979	58	7.92E-02	5.81E-46	12.62
Moth_2086	nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	1828	1536	1115	87	11923	5978	1426	424	353	204	1083	60	1.00E+00	5.01E-11	7.07
Moth_2170	acetyl-coenzyme A synthetase	2264	1943	730	72	27146	2852	153	457	92	48	201	9	2.56E-15	0.00E+00	50.78
Moth_2224	NLP/P60	2017	2012	776	75	4261	4007	620	149	249	147	854	35	2.49E-01	6.73E-04	4.26
Moth_2225	hypothetical protein	4784	5804	1253	145	8335	9058	1308	241	574	242	1969	57	2.13E-01	4.61E-03	4.23
Moth_2226	putative small multidrug resistance transmembrane protein	2542	3749	435	294	3301	3054	272	288	164	84	570	62	1.04E-05	2.43E-05	4.65
Moth_2227	hypothetical protein	3387	4477	478	337	3652	4043	243	310	237	102	765	75	8.70E-05	6.40E-04	4.13
Moth_2228	glycosyl transferase family protein	10475	20725	2556	730	8063	8548	1315	405	495	196	1640	82	4.00E-17	1.65E-04	4.94
Moth_2276	indolepyruvate oxidoreductase subunit beta	992	1269	1004	82	3069	1063	439	134	134	129	406	36	1.62E-01	3.11E-03	3.72

Moth_2296	N-acetylmuramoyl-L-alanine amidase	4715	2818	1132	167	2191	2043	880	126	508	80	810	42	1.43E-03	4.92E-02	3.00
Moth_2368	TetR family transcriptional regulator	680	1728	1346	92	4493	1971	1003	226	207	197	589	53	5.09E-01	4.57E-04	4.26
Moth_2369	acriflavin resistance protein	7123	12240	9075	132	38071	9402	4801	284	580	711	1750	33	1.47E-02	5.38E-13	8.61
Moth_2370	secretion protein HlyD	5718	14795	6316	326	19365	5811	2609	389	325	385	746	43	6.93E-12	1.03E-17	9.05
Moth_2374	sporulation protein and related proteins	2387	1462	751	67	2882	1609	129	69	261	60	486	19	5.18E-03	5.18E-03	3.63

Apêndice H.B. Tabela dos genes diferencialmente expressos de *M.thermoacetica* em bagaço em comparação com celulose. Na tabela estão presentes os genes, descrição, número de fragmentos de cada triplicata biológica para bagaço (B1, B2, B3), palha (P1, P2, P3) e celulose (C1, C2, C3), números de expressão calculados e *qvalue* para cada comparação de cada gene. FC, indica o número de vezes que o gene foi aumentado ou diminuído. Valores maiores que 1, os genes são regulados positivamente para a condição palha.

Gene	Descrição	B1	B2	B3	Expressão	P1	P2	P3	Expressão	C1	C2	C3	Expressão	qValue B vs C	qValue P vs C	FC
Moth_0012	1-phosphofructokinase	4451	17325	3284	407	41	570	272	20	681	274	281	45	1.56E-19	1.00E+00	9.04
Moth_0013	PTS fructose IIC component	9305	41728	6553	627	285	1481	158	20	2468	599	871	84	5.62E-09	8.92E-03	7.46
Moth_0014	putative PTS IIA-like nitrogen-regulatory protein PtsN	2936	14498	1871	629	43	532	69	21	924	163	327	83	6.28E-14	6.26E-03	7.58
Moth_0015	hypothetical protein	1876	10488	1260	734	37	347	67	26	520	105	166	80	2.28E-24	1.04E-01	9.18
Moth_0016	phosphoenolpyruvate--protein phosphotransferase	7825	34183	4320	410	398	2034	299	25	1938	979	1420	87	4.14E-03	7.35E-02	4.71
Moth_0026	DNA adenine methylase	4451	6168	1761	161	1158	1074	248	36	365	405	716	49	3.62E-02	1.00E+00	3.29
Moth_0028	hypothetical protein	2533	3046	1721	338	3892	2792	662	382	257	173	728	106	1.92E-02	6.70E-03	3.19
Moth_0032	hypothetical protein	2152	1426	422	121	872	900	294	76	333	24	399	33	2.00E-03	1.86E-01	3.67

Moth_0035	2-oxoacid:acceptor oxidoreductase subunit gamma	1435	796	280	63	662	601	220	46	276	5	389	23	3.41E-02	1.00E+00	2.74		
Moth_0041	arginine decarboxylase	19096	38390	15970	770	15011	8054	4212	357	1565	943	4757	141	1.38E-03	4.91E-01	5.46		
Moth_0042	thymidylate kinase	6412	13059	5053	581	4720	2598	1055	236	501	320	1458	103	7.38E-06	3.71E-01	5.64		
Moth_0043	DNA-directed DNA polymerase	7079	14802	4958	465	6134	3722	1121	216	653	460	1902	102	1.85E-03	5.07E-01	4.56		
Moth_0044	Signal peptidase II	9183	17310	6142	670	10461	5858	1404	400	961	647	2799	175	3.15E-02	4.94E-01	3.83		
Moth_0045	hypothetical protein	3433	5810	3764	599	4080	2153	677	366	336	256	1107	154	5.22E-03	2.77E-01	3.89		
Moth_0046	hypothetical protein	11410	18060	15974	717	10432	6444	3002	393	1367	643	3560	160	7.75E-03	4.47E-01	4.48		
Moth_0056	hypothetical protein	2194	1309	229	61	63	390	16	7	37	21	48	3	1.49E-170	1.00E+00	20.33		
Moth_0057	hypothetical protein	31991	140022	36542	11363	15203	17480	8021	2832	3503	3830	9813	2098	2.99E-02	1.00E+00	5.42		
Moth_0058	electron transfer flavoprotein subunit beta	12845	21594	4357	716	4019	2823	1428	193	1383	356	1532	107	9.38E-08	8.71E-01	6.69		
Moth_0059	electron transfer flavoprotein alpha and beta-subunits	14566	23242	4607	612	4383	3012	1731	172	1492	202	1761	77	1.36E-11	4.52E-01	7.95		
Moth_0060	FAD dependent oxidoreductase	8243	16235	2959	314	2688	1699	880	75	668	116	956	31	6.36E-26	2.30E-01	10.13		
Moth_0061	ferredoxin 370	694	209	67	210	114	42	22	40	13	58	10	5.64E-20	1.00E+00	6.70			
Moth_0062	Ferritin and Dps	2924	2633	527	207	439	594	128	44	105	38	249	23	1.77E-23	1.00E+00	9.00		
Moth_0081	transcription-repair coupling factor	28689	21013	4760	223	14015	9198	3482	132	1415	692	3042	40	2.45E-04	1.43E-01	5.58		
Moth_0082	MscS mechanosensitive ion channel	2115	1375	460	62	1098	914	172	38	115	62	229	13	3.16E-06	1.00E+00	4.77		
Moth_0094	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	1595	2384	1097	131	1184	447	266	61	147	188	348	45	3.26E-02	7.22E-01	2.91		
Moth_0117	hypothetical protein	1183	607	319	110	703	597	46	71	145	10	297	33	1.11E-02	2.71E-01	3.33		
Moth_0118	phosphosulfolactate synthase	12957	4000	1151	316	2863	3276	299	116	727	67	1069	49	1.42E-09	2.39E-01	6.45		
Moth_0121	3-methyl-2-oxobutanoate hydroxymethyltransferase	11103	20139	15383	842	13248	9989	1773	487	1450	628	3922	191	1.08E-02	4.14E-01	4.41		

Moth_0122	hypothetical protein	8848	15559	10572	721	11262	8259	1366	460	1080	441	3123	164	6.05E-03	2.60E-01	4.40	
Moth_0140	NADP oxidoreductase, coenzyme F420-dependent				2067	7106	1055	166	857	404	383	37	161	138	230	20	2.24E-18
1.00E+00	8.30																
Moth_0141	pantothenate synthetase	1445	4350	722	114	630	369	195	26	126	103	193	17	5.73E-13	1.00E+00	6.71	
Moth_0142	aspartate alpha-decarboxylase	536	1789	343	103	327	184	142	34	50	33	89	14	1.62E-22	1.00E+00	7.36	
Moth_0143	biotin--acetyl-CoA-carboxylase ligase			16118	31758	15547	1102	16290	12105	3889	656	2045	1194	4235	261	4.20E-02	5.50E-01
4.22																	
Moth_0144	BioY protein	1181	3017	1028	136	2372	950	177	96	130	105	320	30	3.85E-05	1.88E-02	4.53	
Moth_0145	hypothetical protein	660	1548	589	114	1533	659	174	106	76	49	227	27	2.44E-04	1.29E-03	4.22	
Moth_0148	putative transcriptional regulator			15715	34244	5421	1567	3881	1828	1121	243	601	155	1384	96	2.36E-74	2.23E-01
16.32																	
Moth_0149	hypothetical protein	20546	37479	14755	1190	6208	5610	4286	381	2134	483	5544	206	3.74E-04	1.00E+00	5.78	
Moth_0150	shikimate/quinate 5-dehydrogenase			3854	6835	2863	184	3081	2084	906	101	751	196	1554	56	4.31E-02	5.93E-01
3.29																	
Moth_0155	hypothetical protein	971	2277	1208	233	2586	936	318	227	133	152	433	84	4.65E-02	8.67E-02	2.77	
Moth_0158	polyprenyl synthetase		2741	2521	493	87	1703	1395	458	64	510	28	642	25	6.26E-03	1.33E-01	3.48
Moth_0160	UvrB/UvrC protein	2179	1438	3466	198	3957	3363	2434	400	1309	2133	3171	524	4.70E-02	1.00E+00	0.38	
Moth_0175	CopG-like DNA-binding protein		468	682	1247	142	182	1036	32	67	166	67	215	53	4.61E-02	7.14E-01	2.68
Moth_0176	hypothetical protein	959	1287	2621	160	355	1903	32	67	286	109	416	52	1.96E-02	7.62E-01	3.08	
Moth_0177	hypothetical protein	386	514	1061	102	122	894	21	48	120	52	152	35	3.46E-02	1.00E+00	2.91	
Moth_0196	AbrB family transcriptional regulator			879	1130	1718	216	266	1148	16	73	182	112	217	69	1.38E-02	9.09E-01
3.13																	
Moth_0197	radical SAM family protein		7457	5987	3357	123	8312	4332	1381	120	761	192	1474	29	3.58E-03	6.32E-03	4.24
Moth_0200	hypothetical protein	6642	4355	403	165	856	1315	249	39	285	23	324	13	2.93E-54	1.00E+00	12.69	
Moth_0204	hypothetical protein	2168	1656	89	244	147	440	34	38	61	12	101	17	3.55E-124	1.00E+00	14.35	
Moth_0205	hypothetical protein	1062	978	83	75	154	241	33	16	32	9	51	5	1.10E-125	1.00E+00	15.00	
Moth_0222	hypothetical protein	3051	3246	620	235	1695	1083	339	123	199	48	361	34	2.35E-14	2.14E-03	6.91	

Moth_0244	cell division ATP-binding protein FtsE			762	617	1180	48	2587	1169	4457	326	365	817	1667	143	1.35E-02	4.64E-01	0.34
Moth_0246	peptidase M23B	1522	1343	3266	77	5135	2474	7986	397	761	2053	3201	220	1.05E-02	1.00E+00	0.35		
Moth_0274	hypothetical protein	2057	2889	692	311	1001	1026	443	173	372	48	643	88	4.00E-03	3.57E-01	3.53		
Moth_0279	XRE family transcriptional regulator			187	507	627	77	37	214	21	16	44	13	67	13	2.77E-17	1.00E+00	5.92
Moth_0306	hypothetical protein	555	747	173	66	296	219	111	36	44	38	60	16	1.43E-06	1.00E+00	4.13		
Moth_0367	hypothetical protein	12038	21282	1759	1312	141	2170	41	73	857	33	203	60	5.97E-186	8.14E-01	21.87		
Moth_0371	hypothetical protein	1812	4258	1110	131	1075	706	354	48	145	200	478	35	2.70E-03	1.00E+00	3.74		
Moth_0372	hypothetical protein	1025	2530	618	167	667	409	270	71	100	110	320	46	2.25E-03	5.54E-01	3.63		
Moth_0373	putative transcriptional regulator			1359	1644	1490	175	335	845	41	45	314	65	160	40	2.71E-05	1.00E+00	4.38
Moth_0374	NADPH-dependent FMN reductase			5092	5043	4283	368	1134	11037	548	301	1043	153	699	83	6.12E-04	1.44E-02	4.43
Moth_0412	2-oxoisovalerate dehydrogenase, E1 component subunit beta			1782	2809	1541	428	1866	2118	360	327	226	142	695	140	2.30E-02	2.07E-01	3.06
Moth_0431	sugar fermentation stimulation protein			2113	2287	1580	124	1356	820	307	62	164	76	415	25	1.50E-06	1.21E-01	4.96
Moth_0434	DsrE-like protein	331	977	149	62	470	281	53	36	47	72	36	22	3.77E-02	1.00E+00	2.82		
Moth_0462	amidohydrolase	2399	4365	2672	114	2621	986	755	69	274	224	659	30	3.61E-03	2.57E-01	3.80		
Moth_0464	Iron-containing alcohol dehydrogenase			1631	326	545	32	1103	1744	1186	80	596	678	1092	82	4.94E-02	1.00E+00	0.39
Moth_0469	hypothetical protein	1367	912	127	135	265	367	20	35	57	6	113	14	2.43E-50	1.00E+00	9.64		
Moth_0470	hypothetical protein	2332	1388	240	129	567	641	44	40	104	21	182	16	2.16E-20	1.00E+00	8.06		
Moth_0471	heat shock protein Hsp20	2397	1383	272	125	425	692	123	42	178	20	292	23	7.25E-08	1.00E+00	5.43		
Moth_0478	DeoR family transcriptional regulator			3864	7271	2925	271	5367	2916	592	184	362	353	1127	76	2.03E-02	2.88E-01	3.57
Moth_0485	hypothetical protein	478	562	420	81	124	671	23	41	146	20	163	29	4.17E-02	1.00E+00	2.79		

Moth_0518	glutamate racemase	7556	7274	3496	321	3700	2373	1256	166	554	395	1650	90	3.16E-02	6.26E-01	3.57	
Moth_0519	ribonuclease PH	8124	7659	3217	368	3816	2439	965	170	520	331	1523	88	4.72E-03	5.58E-01	4.18	
Moth_0520	Ham1-like protein	5520	5269	2179	312	2606	1834	691	150	326	208	929	68	1.94E-04	3.21E-01	4.59	
Moth_0521	diguanylate cyclase	3305	5969	3534	215	654	2049	71	42	540	281	603	52	2.29E-03	1.00E+00	4.13	
Moth_0522	hypothetical protein	1615	3840	2293	314	403	1053	14	52	293	185	207	70	9.24E-05	1.00E+00	4.49	
Moth_0545	chaperonin Cpn10	20571	59275	16705	5061	6997	10765	896	953	1675	2006	3938	976	9.40E-03	1.00E+00	5.19	
Moth_0553	NH(3)-dependent NAD(+) synthetase			14688	17420	6634	327	3587	3081	1220	82	529	452	1435	43	1.70E-10	5.94E-01 7.60
Moth_0576	DNA polymerase III subunit delta			3808	4130	1928	145	3312	2293	742	110	560	159	911	42	1.38E-02	1.59E-01 3.45
Moth_0584	GrpE protein HSP-70 cofactor	5842	10648	5903	485	1592	3622	415	125	449	456	1189	103	7.03E-04	1.00E+00	4.71	
Moth_0587	hypothetical protein	2489	4454	1795	174	1581	1148	459	77	280	146	616	39	1.53E-04	3.87E-01	4.46	
Moth_0588	MiaB-like tRNA modifying enzyme			3162	6746	2462	135	2349	1637	642	61	369	177	912	29	1.04E-04	3.55E-01 4.66
Moth_0591	hypothetical protein	2080	1743	2277	289	6777	2592	570	521	409	45	644	79	2.69E-03	3.16E-11	3.66	
Moth_0611	tagatose-6-phosphate kinase	2647	1294	723	71	807	3406	46	58	208	115	262	20	3.23E-03	4.70E-02	3.55	
Moth_0612	periplasmic binding protein/LacI transcriptional regulator					5289	2674	1523	139	1653	6141	30	100	519	150	517	33 5.61E-04 4.70E-02 4.21
Moth_0654	hypothetical protein	1773	1033	563	131	214	1018	24	44	97	34	166	21	4.13E-10	1.00E+00	6.24	
Moth_0655	hypothetical protein	2579	1369	557	176	360	1508	90	73	146	56	391	40	3.17E-05	3.70E-01	4.40	
Moth_0692	response regulator receiver modulated diguanylate cyclase					1501	1715	2886	108	2645	2006	6055	398	914	1787	4513	313 8.80E-03 1.00E+00 0.35
Moth_0693	hypothetical protein	399	378	782	81	658	506	1660	319	264	476	1279	255	1.77E-02	1.00E+00	0.32	
Moth_0698	DeoR family transcriptional regulator			3054	9240	938	249	108	1542	20	25	111	29	48	7	0.00E+00	1.00E+00 35.57
Moth_0699	ABC transporter substrate-binding protein			2582	9492	612	178	618	3975	32	53	115	76	73	9	6.89E-138	1.00E+00 19.78

Moth_0705	5'-methylthioadenosine phosphorylase	12978	18271	4504	674	9298	6095	1472	353	823	626	2378	145	2.64E-03	4.07E-01	4.65		
Moth_0711	peptidase U61, LD-carboxypeptidase A	12124	30246	4781	738	8979	9840	1864	347	1130	263	3290	110	2.05E-07	1.32E-01	6.71		
Moth_0712	hypothetical protein	6164	4755	1443	555	2197	3261	954	348	882	86	1426	159	1.93E-02	3.72E-01	3.49		
Moth_0725	sigma-54 dependent trancsriptional regulator	5938	8644	2190	129	3534	3737	748	67	371	268	1080	26	8.45E-05	2.01E-01	4.96		
Moth_0727	4-hydroxythreonine-4-phosphate dehydrogenase	1235	310	5601	97	156	193	83	8	83	219	181	23	2.49E-04	1.00E+00	4.22		
Moth_0737	amino acid permease	4549	3335	1046	85	3049	2506	665	68	572	72	1071	24	1.02E-02	9.51E-02	3.54		
Moth_0768	flagellar basal-body rod protein FlgB	134	94	215	15	173	212	291	44	196	200	653	83	2.02E-07	5.24E-01	0.18		
Moth_0769	flagellar basal-body rod protein FlgC	181	133	427	24	235	380	920	108	410	313	1179	129	3.19E-06	1.00E+00	0.19		
Moth_0770	flagellar hook-basal body complex protein (FliE)	74	38	120	10	46	126	117	23	131	128	334	64	3.63E-15	1.77E-01	0.16		
Moth_0771	flagellar MS-ring protein	412	303	879	14	638	779	1903	63	666	775	2090	75	4.57E-07	1.00E+00	0.19		
Moth_0772	flagellar motor switch protein G	410	276	767	20	847	726	1774	96	566	781	1998	113	7.50E-08	1.00E+00	0.18		
Moth_0773	flagellar assembly protein FliH	330	229	612	22	784	709	1436	108	361	463	1108	88	1.22E-03	1.00E+00	0.25		
Moth_0774	ATPase FliI/YscN	602	491	1260	25	1837	1160	3678	152	633	851	2370	98	2.83E-04	1.00E+00	0.26		
Moth_0775	flagellar export FliJ	116	88	243	14	349	246	561	74	131	209	491	68	1.06E-05	9.41E-01	0.21		
Moth_0776	flagellar hook-length control protein	494	273	823	14	1208	954	2466	84	522	581	1822	58	2.44E-04	1.00E+00	0.24		
Moth_0780	flagellar protein	102	42	122	17	112	172	104	39	82	97	185	63	1.62E-04	7.48E-01	0.27		
Moth_0785	hypothetical protein	364	229	421	35	583	519	909	132	288	312	860	118	1.85E-02	1.00E+00	0.30		
Moth_0789	flagellar biosynthetic protein FlhB	549	408	830	22	952	873	1947	96	413	470	1687	70	1.35E-02	1.00E+00	0.31		
Moth_0791	flagellar biosynthesis regulator FlhF	683	514	1055	27	1647	933	1899	99	580	555	1450	72	3.90E-02	1.00E+00	0.38		

Moth_0829	hypothetical protein	1174	1833	457	208	711	752	181	111	119	110	216	64	7.31E-03	4.30E-01	3.25	
Moth_0844	UDP-N-acetylmuramate--L-alanine ligase	11097	16308	1683	306	4549	3163	1124	109	673	274	1205	42	8.74E-11	2.20E-01	7.29	
Moth_0845	UDP-N-acetylmuramate dehydrogenase	7971	6195	1449	250	2871	2852	1210	142	707	218	1203	60	3.09E-03	3.13E-01	4.17	
Moth_0846	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	8363	6237	1653	188	2889	2820	1119	99	519	271	1122	43	1.25E-03	3.29E-01	4.37	
Moth_0874	ABC transporter	4038	4625	2165	168	1872	1431	289	61	257	156	558	30	2.11E-07	3.56E-01	5.60	
Moth_0875	hypothetical protein	1659	1624	857	78	1023	791	130	39	124	87	255	19	1.17E-04	1.00E+00	4.11	
Moth_0878	bifunctional pyrimidine regulatory protein PyrR/uracil phosphoribosyltransferase	4276	8283	4639	446	2351	895	215	97	304	144	529	49	1.18E-22	3.74E-01	9.10	
Moth_0879	aspartate carbamoyltransferase catalytic subunit	1461	2186	2286	93	1776	517	164	43	269	114	511	26	4.89E-03	1.00E+00	3.58	
Moth_0880	dihydroorotase	1501	2595	2582	75	2107	726	304	41	369	125	663	23	1.60E-02	1.00E+00	3.26	
Moth_0887	ATPase	20376	11875	4594	190	20297	9505	1612	174	1978	155	3616	42	5.44E-03	1.80E-02	4.52	
Moth_0915	5-amino-6-(5-phosphoribosylamino)uracil reductase / diaminohydroxyphosphoribosylaminopyrimidine deaminase	13822	18603	6980	512	11589	6374	1821	291	846	584	2024	92	5.51E-05	1.25E-01	5.57	
Moth_0916	riboflavin synthase subunit alpha	7511	10090	3365	471	6024	3656	684	250	387	313	993	80	7.98E-07	7.36E-02	5.89	
Moth_0917	GTP cyclohydrolase II	11602	17438	5753	423	12392	6619	1193	262	708	556	1828	78	9.64E-05	8.61E-02	5.42	
Moth_0918	6,7-dimethyl-8-ribityllumazine synthase	3558	5556	1976	349	5092	2459	523	275	298	264	744	90	3.68E-03	6.11E-02	3.88	
Moth_0919	phosphodiesterase	4247	6999	3077	372	6129	3033	999	303	392	298	1023	92	3.74E-03	4.63E-02	4.04	
Moth_0929	ABC transporter	6193	5228	642	240	3986	5058	1231	235	386	305	943	69	2.22E-02	3.59E-02	3.48	
Moth_0930	periplasmic solute binding protein	11105	12242	951	407	4422	5318	1673	229	498	424	1239	78	9.48E-05	1.52E-01	5.22	
Moth_0931	ferric uptake regulator family protein	8705	13071	723	734	2144	2257	448	172	156	157	357	50	2.61E-71	1.07E-02	14.68	

Moth_0943	fatty acid biosynthesis transcriptional regulator	36834	84812	43562	4172	34559	6969	5315	1428	2265	2346	8344	683	3.38E-03	1.00E+00	6.11		
Moth_0944	phosphate:acyl-[acyl carrier protein] acyltransferase	37715	84693	54586	2554	46747	9911	5597	1046	2830	2902	11732	511	4.91E-02	1.00E+00	5.00		
Moth_0958	RNA polymerase sigma factor SigX	82217	150605	104467	10115	26880	8907	3234	1333	1440	617	3383	295	2.40E-292	7.66E-03	34.29		
Moth_0959	putative transmembrane transcriptional regulator (anti-sigma factor)	126468	216215	170282	6528	49126	16755	5670	1033	3153	1340	6777	263	2.84E-92	1.33E-01	24.82		
Moth_0960	ABC transporter	17888	26994	25992	1123	8662	3621	1800	274	937	384	2211	100	4.28E-25	2.46E-01	11.23		
Moth_0961	ABC-type transport system involved in multi-copper enzyme maturation, permease component	5592	10330	6769	378	4709	1385	540	126	224	149	674	33	2.27E-38	3.29E-03	11.45		
Moth_0962	signal recognition particle-docking protein FtsY	5332	9260	5389	313	7454	3720	1324	228	633	463	1805	91	4.58E-02	3.24E-01	3.44		
Moth_0972	signal peptidase I	3120	5357	1247	259	2573	1547	347	130	163	171	513	49	8.75E-07	1.07E-01	5.29		
Moth_0973	GTP-binding protein YlqF	3789	5961	1674	207	3909	2575	753	150	318	235	745	49	8.74E-04	5.82E-02	4.22		
Moth_0974	histidine kinase	1969	4228	959	92	1310	919	159	32	136	108	295	15	2.51E-10	1.00E+00	6.13		
Moth_0998	MarR family transcriptional regulator	804	1303	707	88	310	280	48	21	85	48	170	19	3.40E-05	1.00E+00	4.63		
Moth_1012	spore germination protein	493	2060	3267	78	234	376	23	8	462	90	141	18	1.15E-04	1.00E+00	4.33		
Moth_1015	hypothetical protein	2219	1085	487	213	502	1061	110	97	307	21	407	60	2.30E-03	4.95E-01	3.55		
Moth_1016	catalase	7106	8578	1154	361	1789	2257	664	119	751	35	812	48	1.62E-14	1.64E-01	7.52		
Moth_1020	transcriptional regulator/antitoxin MazE	341	857	399	96	60	345	10	22	60	8	90	14	4.72E-22	1.00E+00	6.86		
Moth_1021	transcriptional modulator of MazE/toxin MazF	599	1513	672	118	93	573	11	24	104	32	147	21	2.58E-08	1.00E+00	5.62		
Moth_1046	hypothetical protein	4474	11637	4178	632	3052	1758	381	175	379	514	1058	149	3.81E-03	1.00E+00	4.24		
Moth_1047	NusA antitermination factor	13788	30941	10544	753	8165	5912	1510	237	1177	1129	2810	158	5.90E-03	1.00E+00	4.77		
Moth_1048	hypothetical protein	1798	4523	1599	417	1380	893	134	132	146	163	409	86	1.11E-05	6.03E-01	4.85		

Moth_1049	50S ribosomal protein L7AE	1732	4026	1325	340	1204	842	103	105	136	140	375	70	6.63E-06	6.01E-01	4.86
Moth_1050	translation initiation factor 2	17914	38538	11827	374	16596	10023	2055	169	1728	1303	3818	79	1.46E-02	1.00E+00	4.73
Moth_1051	ribosome-binding factor A	3280	7856	2055	533	3685	1902	422	261	357	230	765	111	6.11E-05	2.49E-01	4.80
Moth_1052	phosphoesterase, RecJ-like	7370	17197	4496	447	7976	3812	1373	232	645	448	1799	87	1.81E-04	2.55E-01	5.14
Moth_1053	tRNA pseudouridine synthase B	5990	13658	3610	377	6142	3293	1599	218	720	395	1870	90	5.71E-03	3.54E-01	4.19
Moth_1054	riboflavin kinase/FMN adenylyltransferase	2448	6815	1618	169	2059	1489	311	64	274	199	613	35	2.38E-05	4.58E-01	4.83
Moth_1058	Serine-type D-Ala-D-Ala carboxypeptidase	7955	9292	2647	252	6119	3823	792	148	927	234	2133	67	1.91E-02	4.50E-01	3.76
Moth_1069	hypothetical protein	23762	29679	6093	526	5471	12171	801	156	1267	1429	2459	115	1.48E-02	1.00E+00	4.57
Moth_1070	hypothetical protein	1044	751	234	110	394	467	76	55	112	22	164	26	7.23E-05	2.19E-01	4.23
Moth_1071	ATP-dependent Clp protease proteolytic subunit ClpP	2667	1964	566	96	979	1146	195	47	341	58	381	24	4.78E-04	1.00E+00	4.00
Moth_1121	hypothetical protein	1237	4840	2336	304	892	1040	339	98	251	247	795	99	3.90E-02	1.00E+00	3.07
Moth_1122	citrate synthase	6107	19910	10019	397	1546	3712	299	59	997	746	2376	94	1.31E-02	1.00E+00	4.22
Moth_1123	hypothetical protein	8870	11506	6653	962	1743	810	119	100	229	258	649	94	1.76E-28	1.00E+00	10.23
Moth_1125	hypothetical protein	916	1463	1475	63	1017	573	110	30	96	84	301	16	5.12E-04	1.00E+00	3.94
Moth_1153	hypothetical protein	2925	2443	1215	203	2610	1568	411	161	302	81	635	52	1.23E-03	3.33E-02	3.90
Moth_1165	GntR family transcriptional regulator	1765	2735	1476	224	1663	1297	200	126	231	179	596	79	4.49E-02	5.82E-01	2.84
Moth_1173	hypothetical protein	1408	1973	458	69	518	728	72	24	153	78	326	19	1.85E-03	1.00E+00	3.63
Moth_1174	ABC transporter	1202	1929	378	74	421	503	70	21	108	68	245	17	1.13E-04	1.00E+00	4.35
Moth_1175	GntR family transcriptional regulator	839	1513	253	103	279	285	56	26	84	54	176	25	5.09E-04	1.00E+00	4.12
Moth_1179	ferredoxin-dependent glutamate synthase	4413	3799	873	80	2883	4187	111	59	429	141	732	20	1.88E-03	7.16E-02	4.00

Moth_1180	IclR family transcriptional regulator	901	412	354	29	690	2226	443	69	782	254	1978	92	1.63E-02			
1.00E+00	0.32																
Moth_1209	methanol:corrinoid methyltransferase	964	744	664	23	1142	1641	332	35	424	1330	715	94	1.43E-04	3.16E-		
01	0.24																
Moth_1211	EmrB/QacA family drug resistance transporter	2932	3138	2027	82	2269	2293	593	61	320	217	709	26	3.24E-			
02	2.61E-01	3.15															
Moth_1224	N-ethylammeline chlorohydrolase	534	648	1313	27	2283	1710	733	65	936	440	2163	72	3.26E-02			
1.00E+00	0.38																
Moth_1225	molybdopterin binding aldehyde oxidase and xanthine dehydrogenase	379	352	747	21	1533	1046	490	56	595	262	1412					
60	4.20E-02	1.00E+00	0.35														
Moth_1226	aldehyde oxidase and xanthine dehydrogenase	763	705	1251	30	3227	2222	1932	123	868	471	2381	79	4.21E-			
02	1.00E+00	0.38															
Moth_1236	phosphate transport regulator	2063	1041	457	77	390	685	57	24	191	43	372	21	1.43E-03	1.00E+00	3.67	
Moth_1237	phosphate transporter	3947	5486	1316	161	990	1077	298	41	329	118	606	28	5.71E-08	1.00E+00	5.75	
Moth_1239	hypothetical protein	3806	3891	151	488	374	696	89	74	86	14	69	18	0.00E+00	2.28E-07	27.11	
Moth_1240	hypothetical protein	5302	5172	245	559	471	922	72	75	99	27	94	21	0.00E+00	5.28E-04	26.62	
Moth_1262	LacI family transcriptional regulator	1968	7995	4080	200	912	1777	109	37	952	405	381	58	3.27E-02	8.28E-		
01	3.45																
Moth_1266	hypothetical protein	5341	2391	1151	425	1264	1887	215	169	489	44	729	90	2.70E-05	4.24E-01	4.72	
Moth_1267	hypothetical protein	5519	2285	1087	461	1510	2229	223	213	603	68	756	116	1.06E-03	4.50E-01	3.97	
Moth_1269	hypothetical protein	19420	6558	600	609	1459	3034	204	106	446	74	343	34	4.59E-114	2.85E-02	17.91	
Moth_1271	peptidoglycan-binding LysM	11484	3305	497	236	828	1754	114	41	265	67	223	16	2.29E-78	1.00E+00	14.75	
Moth_1282	S-adenosylmethionine decarboxylase	6473	21410	22719	1958	7570	3632	1685	609	843	1216	2515	442	1.32E-02			
1.00E+00	4.43																
Moth_1311	cobalamin B12-binding	983	404	766	46	2252	3512	459	143	785	321	1997	120	4.79E-02	1.00E+00	0.38	
Moth_1324	hypothetical protein	848	1089	390	67	560	432	170	41	127	27	246	18	3.20E-03	1.00E+00	3.72	
Moth_1325	hypothetical protein	1111	1209	327	83	603	441	145	44	121	28	206	18	7.36E-05	1.00E+00	4.61	
Moth_1327	hypothetical protein	3847	5060	1144	195	2309	1588	682	109	516	94	697	41	2.73E-05	1.08E-01	4.76	

Moth_1358	hypothetical protein	4872	5003	1910	359	3072	2373	877	234	794	130	1359	108	3.32E-02	3.97E-01	3.32	
Moth_1365	group 1 glycosyl transferase	7865	9484	1780	236	5034	4176	996	141	646	321	1630	59	8.41E-03	3.50E-01	4.00	
Moth_1376	SpoVA protein	2144	1810	851	136	1721	1538	412	120	455	33	700	48	4.52E-02	1.47E-01	2.83	
Moth_1404	manganese transport transcriptional regulator				2755	1364	657	145	2150	1749	107	122	308	25	623	42	4.79E-03
4.40E-02	3.45																
Moth_1405	nucleoside recognition protein	7206	6827	1936	167	5639	4491	362	110	659	96	1349	33	3.50E-05	3.38E-02	5.06	
Moth_1406	HSR1-like GTP-binding protein	3967	7853	1365	247	2487	1722	257	87	306	100	541	31	2.78E-17	7.07E-02	7.97	
Moth_1407	hypothetical protein	618	1970	320	175	360	227	38	39	56	21	74	17	5.67E-59	1.00E+00	10.29	
Moth_1439	SirA-like protein	1909	1393	1112	291	2930	2752	67	364	181	44	381	66	1.66E-05	7.67E-09	4.41	
Moth_1440	sugar kinase	5976	4066	3158	216	7288	7187	182	231	599	137	1122	50	9.34E-04	3.24E-04	4.32	
Moth_1441	hypothetical protein	2173	1577	1207	145	2306	2636	38	137	199	47	322	28	1.47E-07	1.42E-06	5.18	
Moth_1442	hypothetical protein	6465	5593	3664	214	8992	7848	376	231	608	153	1107	43	4.25E-05	6.66E-06	4.98	
Moth_1446	2-hydroxyglutaryl-CoA dehydratase subunit D				1613	2100	3499	90	5228	2504	95	98	256	124	536	22	6.37E-04
1.55E-04	4.09																
Moth_1447	cobalamin B12-binding	2916	4496	5927	115	9477	3933	378	123	493	246	891	28	2.53E-03	1.02E-03	4.11	
Moth_1487	methyl-accepting chemotaxis sensory transducer				3613	6843	2098	94	444	905	194	13	259	178	563	15	4.67E-09
1.00E+00	6.27																
Moth_1489	DNA end-binding protein Ku	2734	2650	797	108	2150	1595	385	81	563	69	773	37	4.14E-02	2.85E-01	2.92	
Moth_1491	hypothetical protein	538	1028	459	120	1055	442	169	114	51	49	196	35	7.87E-03	1.81E-02	3.43	
Moth_1506	SpoIVB peptidase	7847	1843	516	109	586	1254	359	29	357	24	254	10	6.13E-38	1.00E+00	10.90	
Moth_1560	DNA-3-methyladenine glycosylase III				2807	2963	1012	128	1374	1429	536	78	315	129	550	35	3.78E-03	2.57E-01
3.66																	
Moth_1570	hypothetical protein	922	1191	682	160	318	747	62	64	124	48	261	45	2.42E-03	6.16E-01	3.56	
Moth_1571	hypothetical protein	940	1341	799	147	99	567	10	28	143	55	130	32	4.17E-05	1.00E+00	4.59	
Moth_1582	sporulation sigma factor SigK	1797	2223	171	91	152	418	24	12	71	21	102	7	5.09E-78	1.00E+00	13.00	
Moth_1587	patatin	4191	3958	910	161	715	864	156	33	209	93	462	25	6.94E-11	1.00E+00	6.44	

Moth_1590	major facilitator transporter	4642	631	558	63	787	2579	1518	83	4219	335	5174	161	3.29E-02	1.00E+00	0.39
Moth_1596	hypothetical protein	1112	2242	696	244	350	515	59	56	168	149	313	89	4.96E-02	6.65E-01	2.74
Moth_1599	dissimilatory sulfite reductase B	1636	1334	792	233	1118	1005	91	140	250	102	329	85	4.61E-02	5.12E-01	2.74
Moth_1624	redox-active disulfide protein 2	3695	3780	2790	606	5672	3388	452	591	309	123	608	113	5.84E-07	1.59E-06	5.36
Moth_1625	selenophosphate synthase	10553	11773	8675	457	23442	10596	1620	557	938	435	2280	99	2.05E-03	4.03E-05	4.62
Moth_1647	hypothetical protein	3908	3852	999	123	2503	1928	579	82	518	67	815	30	5.95E-04	8.73E-02	4.10
Moth_1648	hypothetical protein	2021	1638	447	356	973	1013	231	218	319	40	430	106	6.05E-03	2.85E-01	3.36
Moth_1649	hypothetical protein	4702	4217	1079	318	2346	2195	704	200	696	85	998	86	5.98E-03	2.61E-01	3.70
Moth_1659	hypothetical protein	4868	7465	251	405	921	828	102	60	108	23	163	15	0.00E+00	2.70E-04	27.00
Moth_1679	D-tyrosyl-tRNA(Tyr) deacylase	2375	4073	2850	302	4237	1743	898	274	263	219	921	91	1.91E-02	6.24E-02	3.32
Moth_1708	hypothetical protein	5982	7552	3242	369	6936	3424	1291	302	547	341	1696	107	3.90E-02	1.87E-01	3.45
Moth_1760	hypothetical protein	2684	6139	3154	297	2294	1144	638	127	322	351	813	86	2.25E-02	7.92E-01	3.45
Moth_1781	hypothetical protein	2142	1663	310	72	460	492	91	20	154	21	172	9	3.90E-15	1.00E+00	8.00
Moth_1787	hypothetical protein	40570	14337	3998	3603	3324	6736	5981	1653	3062	226	2010	506	3.72E-08	1.19E-01	7.12
Moth_1825	LysR family transcriptional regulator					3049	3058	969	113	2399	2209	721	102	329	151	807	37	3.30E-02	1.02E-01	3.05
Moth_1826	lactate dehydrogenase	5460	5390	1344	187	4070	3213	846	142	474	154	769	38	3.05E-05	6.34E-03	4.92
Moth_1827	XRE family transcriptional regulator					3402	3350	918	147	1944	1679	433	89	289	126	588	36	5.52E-04	1.56E-01	4.08
Moth_1829	peroxiredoxin-like	1262	1913	753	114	1253	870	135	69	61	85	162	22	8.75E-07	2.21E-02	5.18
Moth_1830	cytochrome c biogenesis protein, transmembrane region					3746	6168	2267	262	1646	2054	758	119	653	114	802	55	4.76E-05	3.09E-01	4.76
Moth_1833	hypothetical protein	1007	717	426	112	723	695	98	84	122	38	280	39	3.27E-02	2.37E-01	2.87
Moth_1873	hypothetical protein	1650	943	531	139	1428	826	271	133	258	16	466	47	1.73E-02	3.90E-02	2.96
Moth_1875	small acid-soluble spore protein	464	259	128	74	233	246	35	48	83	8	120	28	7.72E-03	1.00E+00	2.64
Moth_1892	OsmC-like protein	2608	5608	3406	382	3286	2717	503	231	261	302	625	95	2.17E-03	2.28E-01	4.02

Moth_1896	hypothetical protein	12941	28879	7784	676	2409	3950	250	86	618	139	1295	44	2.46E-66	4.78E-01	15.36		
Moth_1914	hypothetical protein	2078	1868	451	96	552	567	155	32	220	19	245	16	7.72E-10	1.00E+00	6.00		
Moth_1920	amino acid permease	8232	13645	6006	293	5942	2637	718	108	733	576	2382	77	2.38E-02	1.00E+00	3.81		
Moth_1923	2-oxoacid:acceptor oxidoreductase subunit delta	12719	28311	8425	2527	19562	2321	2508	1485	1257	719	4709	596	2.29E-02	4.74E-01	4.24		
Moth_1924	pyruvate ferredoxin oxidoreductase subunit gamma	61847	151240	39861	6415	56801	9708	12171	2551	5244	2220	20639	1149	2.86E-02	1.00E+00	5.58		
Moth_1935	galactose-1-phosphate uridylyltransferase	2500	2247	482	75	1281	1215	266	44	381	31	461	18	3.48E-04	1.00E+00	4.17		
Moth_1936	glycogen/starch synthase	3925	2722	929	75	2610	1995	495	57	746	59	939	25	4.72E-02	2.69E-01	3.00		
Moth_1939	biotin/lipoate A/B protein ligase	6781	13258	3844	431	2044	4541	673	138	967	432	2368	121	4.18E-02	1.00E+00	3.56		
Moth_1940	hypothetical protein	2153	4699	1175	463	426	1540	66	106	299	170	636	125	3.35E-03	1.00E+00	3.70		
Moth_1941	hypothetical protein	2045	5423	1237	421	313	1557	47	81	303	171	555	100	3.74E-04	1.00E+00	4.21		
Moth_1960	xanthine dehydrogenase subunit XdhA	1368	1490	3169	37	5442	2997	2456	97	1528	1227	4361	94	3.73E-02	1.00E+00	0.39		
Moth_1965	histidine kinase internal region	10032	18785	3455	277	4014	4718	1509	104	667	467	1921	51	6.05E-05	5.41E-01	5.43		
Moth_1966	5-oxoprolinase (ATP-hydrolyzing)	1051	630	837	20	2886	1876	1697	78	483	699	1455	54	3.44E-02	1.00E+00	0.37		
Moth_1977	hypothetical protein	879	989	1016	58	1317	1530	46	55	153	65	355	21	3.34E-02	6.86E-02	2.76		
Moth_2004	MerR family transcriptional regulator	564	363	608	40	1808	1800	893	162	502	338	1276	116	4.06E-02	1.00E+00	0.34		
Moth_2015	hypothetical protein	5867	3516	234	244	434	888	84	36	133	25	111	12	9.73E-201	1.00E+00	20.33		
Moth_2027	aerolysin	9954	4509	820	183	1369	2608	292	53	450	95	600	23	2.73E-16	1.00E+00	7.96		
Moth_2030	phosphoribosyl-ATP pyrophosphatase / phosphoribosyl-AMP cyclohydrolase	2318	479	782	59	4398	3632	3385	300	800	1071	1584	161	2.40E-02	1.00E+00	0.37		
Moth_2031	imidazole glycerol phosphate synthase subunit hisF	1754	375	557	50	3339	3066	3019	285	732	819	1330	144	1.68E-02	7.49E-01	0.35		

Moth_2032	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino] imidazole-4-carboxamide isomerase	1563	375	504	47	3035	3134	3005	289	828	806	1561	156	3.35E-03	1.00E+00	0.30			
Moth_2033	imidazole glycerol phosphate synthase subunit hisH	1301	318	386	44	2540	2677	2313	267	682	658	1481	153	3.23E-03	1.00E+00	0.29			
Moth_2034	imidazoleglycerol-phosphate dehydratase	1014	205	301	36	2216	2258	1913	241	584	590	1272	145	6.41E-04	1.00E+00	0.25			
Moth_2035	histidinol dehydrogenase	3157	822	1033	55	6120	7276	5681	331	1815	1379	3543	172	5.30E-03	1.00E+00	0.32			
Moth_2036	ATP phosphoribosyltransferase catalytic subunit	1721	632	804	65	2002	3947	5240	439	1870	872	3622	260	8.26E-05	1.00E+00	0.25			
Moth_2037	ATP phosphoribosyltransferase regulatory subunit	2532	518	657	44	6801	8031	5076	355	2225	1231	5306	212	3.08E-07	1.00E+00	0.21			
Moth_2057	hypothetical protein	3647	3077	3083	255	4064	2468	211	180	359	241	887	79	3.16E-02	3.08E-01	3.23			
Moth_2058	cobalt transport protein	1920	1274	1158	89	1611	1154	39	56	157	79	312	22	2.91E-04	9.71E-02	4.05			
Moth_2059	ABC transporter	3421	2336	2144	74	2573	2144	106	45	219	117	586	16	6.69E-05	1.00E+00	4.63			
Moth_2060	hypothetical protein	1197	1040	936	82	831	943	53	46	93	67	262	22	1.24E-03	1.00E+00	3.73			
Moth_2067	hypothetical protein	1423	2263	967	73	983	858	290	39	189	100	496	22	1.35E-02	1.00E+00	3.32			
Moth_2084	ArsR family transcriptional regulator	1774	1971	882	216	12894	4476	1049	941	117	84	489	59	2.17E-03	3.34E-91	3.66			
Moth_2096	cytochrome d ubiquinol oxidase, subunit II	1906	1736	749	63	7606	1929	549	161	246	4050	397	357	1.77E-10	1.00E+00	0.18			
Moth_2097	ABC transporter	793	1311	1025	23	3048	920	835	48	245	2549	533	121	2.94E-08	4.78E-01	0.19			
Moth_2130	chaperonin Cpn10	1297	1738	862	198	1946	1253	144	175	162	120	275	64	1.68E-02	7.17E-02	3.09			
Moth_2170	acetyl-coenzyme A synthetase	2264	1943	730	72	27146	2852	153	457	92	48	201	9	2.56E-15	0.00E+00	8.00			
Moth_2209	ArsR family transcriptional regulator	1281	5074	1300	306	795	775	133	71	161	287	356	91	1.30E-02	1.00E+00	3.36			
Moth_2210	PadR family transcriptional regulator	4798	7650	3672	576	5323	3182	1621	457	497	391	1276	163	2.79E-02	1.81E-01	3.53			
Moth_2226	putative small multidrug resistance transmembrane protein	2542	3749	435	294	3301	3054	272	288	164	84	570	62	1.04E-05	2.43E-05	4.74			

Moth_2227	hypothetical protein	3387	4477	478	337	3652	4043	243	310	237	102	765	75	8.70E-05	6.40E-04	4.49	
Moth_2228	glycosyl transferase family protein			10475	20725	2556	730	8063	8548	1315	405	495	196	1640	82	4.00E-17	1.65E-04
8.90																	
Moth_2261	hypothetical protein	832	966	455	146	431	377	111	70	170	25	226	43	6.99E-03	5.22E-01	3.40	
Moth_2273	LacI family transcriptional regulator			2566	4455	1788	129	998	1099	245	38	246	187	374	27	3.15E-05	
1.00E+00	4.78																
Moth_2296	N-acetylmuramoyl-L-alanine amidase			4715	2818	1132	167	2191	2043	880	126	508	80	810	42	1.43E-03	4.92E-02
3.98																	
Moth_2300	aldehyde ferredoxin oxidoreductase			4173	2883	834	66	1931	1845	767	48	762	60	945	22	3.90E-02	
1.00E+00	3.00																
Moth_2305	hypothetical protein	8083	7568	9562	165	54954	45480	19954	1016	12696	7280	28671	642	2.10E-02	1.00E+00	0.26	
Moth_2306	hypothetical protein	1797	2078	2224	156	14787	11020	4455	968	2770	1639	6275	552	4.14E-04	1.00E+00	0.28	
Moth_2309	adenylosuccinate synthetase	14017	25915	10384	579	11116	5220	1467	228	580	609	1516	71	5.32E-12	1.06E-01	8.15	
Moth_2310	FAD dependent oxidoreductase	14749	25179	12497	550	9976	4291	1181	180	665	781	1787	80	1.55E-07	4.91E-01	6.88	
Moth_2352	DNA polymerase, beta-like region			2133	3074	1792	229	843	3038	53	114	558	100	694	73	2.12E-02	6.06E-01
3.14																	
Moth_2362	sigma-24 (FecI-like protein)	66764	90011	33981	4651	15186	30390	8592	1589	3328	2267	13801	854	2.23E-02	1.00E+00	5.45	
Moth_2369	acriflavin resistance protein	7123	12240	9075	132	38071	9402	4801	284	580	711	1750	33	1.47E-02	5.38E-13	4.00	
Moth_2370	secretion protein HlyD	5718	14795	6316	326	19365	5811	2609	389	325	385	746	43	6.93E-12	1.03E-17	7.58	
Moth_2374	sporulation protein and related proteins			2387	1462	751	67	2882	1609	129	69	261	60	486	19	5.18E-03	5.18E-03
3.53																	
Moth_2414	putative manganese-dependent inorganic pyrophosphatase					7942	13367	4269	288	4906	3912	857	120	758	368	1984	63
1.76E-03	6.09E-01	4.57															
Moth_2417	peptidase S1 and S6, chymotrypsin/Hap			2405	2304	721	67	1205	1221	375	41	317	58	505	17	1.08E-03	
1.00E+00	3.94																
Moth_2464	30S ribosomal protein S7	17981	39003	14564	2252	9851	7873	563	566	1142	1567	3208	459	8.20E-03	1.00E+00	4.91	
Moth_2465	30S ribosomal protein S12	20371	45528	15416	3220	10410	8557	1199	828	1299	1711	3595	637	8.89E-03	1.00E+00	5.05	
Moth_2466	50S ribosomal protein L7AE	11814	26687	10136	2862	5471	4534	902	695	822	1092	2365	609	4.45E-03	1.00E+00	4.70	

Moth_2476	transporter	738	950	462	89	583	411	72	47	93	32	179	22	5.89E-04	1.00E+00	4.05	
Moth_2490	DNA repair protein RadA	14613	31796	11885	588	10948	9880	1043	222	1534	882	3657	118	3.42E-03	1.00E+00	4.98	
Moth_2494	multi-sensor signal transduction histidine kinase	2971	3251	1469	76	3545	2737	383	69	258	211	623	23	1.83E-02	5.82E-02	3.30	
Moth_2518	tRNA uridine 5-carboxymethylaminomethyl modification protein GidA	6120	11212	4608	171	4436	3183	570	68	410	317	1309	31	1.30E-05	4.10E-01	5.52	
Moth_2519	tRNA modification GTPase TrmE	5486	12555	6348	259	3245	2063	737	75	558	455	1522	56	1.24E-03	1.00E+00	4.63	

Anexo

Artigos publicados

1. OSIRO, KAREN O.; DE CAMARGO, BRENDA R.; SATOMI, RACHEL; HAMANN, PEDRO RICARDO V.; SILVA, JÉSSICA PINHEIRO; DE SOUSA, MARCELO VALLE; QUIRINO, BETANIA F.; AQUINO, ELAINE N.; FELIX, CARLOS R.; MURAD, ANDRÉ MELRO; NORONHA, ELIANE F. Characterization of Clostridium thermocellum (B8) secretome and purified cellulosomes for lignocellulosic biomass degradation. Enzyme and Microbial Technology. Fator de Impacto(2016 JCR): 2,5020, v.97, p.43 - 54, 2017.

2. LEITÃO, VANESSA O.; NORONHA, ELIANE F.; CAMARGO, BRENDA R.; HAMANN, PEDRO R. V.; STEINDORFF, ANDREI S.; QUIRINO, BETANIA F.; DE SOUSA, MARCELO VALLE; ULHOA, CIRANO J.; FELIX, CARLOS R. Growth and expression of relevant metabolic genes of Clostridium thermocellum cultured on lignocellulosic residues. Journal of Industrial Microbiology & Biotechnology. Fator de Impacto(2016 JCR): 2,8100, v.1, p.1 - 10, 2017.

3. HAMANN, PEDRO RICARDO V.; SERPA, DAYANE L.; BARRETO DA CUNHA, AMANDA SOUZA; DE CAMARGO, BRENDA R.; OSIRO, KAREN OFUJI; VALLE DE SOUSA, MARCELO; FELIX, CARLOS R.; MILLER, ROBERT N.G.; NORONHA, ELIANE F. Evaluation of plant cell wall degrading enzyme production by Clostridium thermocellum B8 in the presence of raw agricultural wastes. International Biodeterioration & Biodegradation. Fator de Impacto(2016 JCR): 2,9620, v.105, p.97 - 105, 2015.

4.LAMOUNIER, THAIS ALVES DA COSTA; OLIVEIRA, LAYSSA MIRANDA DE; CAMARGO, BRENDA RABELLO DE; RODRIGUES, KELLY BARRETO; NORONHA, ELIANE FERREIRA; RIBEIRO, BERGMANN MORAIS; NAGATA, TATSUYA. Production of Brazilian human norovirus VLPs and comparison of purification methods. Brazilian Journal of Microbiology (Online). , v.46, p.1265 - 1268, 2015.