

Universidade de Brasília - UnB

“Sequenciamento, análise, clonagem e expressão de genes de
celulases e hemicelulases oriundos de bibliotecas metagenômicas de
solo e rúmen”

Brasília

2014

Talita Gabriela Salles Ramos

“Sequenciamento, análise, clonagem e expressão de genes de celulasas e hemicelulasas oriundos de bibliotecas metagenômicas de solo e rúmen”

Orientador: Dr. Ricardo Krüger

Co-orientadora: Dra. Betania Ferraz Quirino

Brasília

2014

AGRADECIMENTOS

- À professora Betania e ao professor Kruger por toda orientação e ajuda nesses dois anos de mestrado
- À Betúlia por toda a ajuda e paciência comigo na bancada e fora dela
- Ao Rodrigo por sempre estar disponível para me ajudar a resolver os mais diversos problemas
- Carol, Barbara, Henrique, Ana e Luiz, meus amigos que ajudaram a tornar o trabalho na Embrapa muito mais divertido
- A todos do grupo de pesquisa da professora Betania pela ajuda sempre que necessária
- À Jessica que me sempre esteve disponível quando eu precisei dos seus conselhos e ajuda
- Aos meus pais que sempre me apoiaram ao longo desses dois anos
- À minha irmã e ao meu irmão por estarem comigo quando eu precisei
- Ao Dr. João Ricardo que me ajudou no trabalho com leveduras
- À Embrapa Agroenergia, Católica e UnB pelo espaço e recursos necessários para a minha pesquisa
- Aos membros da banca presente: Dra. Cristine Chaves Barreto e Dra: Eliane Noronha

RESUMO

A celulose é o polissacarídeo mais abundante da Terra e pode ser encontrada na parede celular das plantas. Para hidrolisar a celulose em resíduos de glicose é necessário um complexo de enzimas hidrolíticas envolvendo as enzimas endoglucanases, celobiohidrolases ou exoglucanases e as β -glicosidases. O uso industrial dessas celulases é bem diverso, abrangendo desde o processamento de algodão, a fabricação de papel, produção de detergentes, extração de polpa de frutas para a fabricação de sucos, o enriquecimento de ração para animais e na produção de bioetanol a partir da biomassa. Devido a essa grande variedade de usos para celulases é que a contínua busca por enzimas que sejam cada vez mais vantajosas na relação custo-benefício e eficiência se faz tão necessária. Uma técnica que pode ser usada na busca por essas novas enzimas é a metagenômica que permite o acesso ao material genético de micro-organismos não cultiváveis. Para a produção industrial dessas proteínas recombinantes, vários sistemas biológicos são empregados, principalmente a bactéria *Escherichia coli* e a levedura *Pichia pastoris*. Nesse trabalho foi feita a triagem de clones provenientes de uma biblioteca metagenômica de rúmen de caprino com fenótipo de degradação de celulases em 4 diferentes substratos: CMC LV, CMC HL, avicel e bagaço de cana pré-tratado. Desses clones 4 foram completamente sequenciados e análises de suas prováveis ORFs e proteínas, foi feita em comparação com bancos de dados existentes. Foi realizada a clonagem do gene *endo 03*, proveniente da biblioteca metagenômica de solo da Amazônia, no vetor de clonagem pJET e posterior clonagem no vetor de expressão pPIC α A, para *P. pastoris*, e expressão em placa. A clonagem dos genes *cbh I* e *cbh II* provenientes da biblioteca metagenômica de rúmen de caprino, no vetor de clonagem pJET e posterior clonagem no vetor de expressão pET21a, para *E. coli*, e expressão das proteínas CBH I e CBH II. Também foi realizada a purificação parcial da proteína CBH II que permitiu a caracterização da

temperatura ótima como 50°C e do pH ótimo na faixa entre 5,0 e 6,0 para a proteína parcialmente purificada.

ABSTRACT

Cellulose is the most abundant polysaccharide on the Earth and it can be found at cell wall of plants. To hydrolyze cellulose into glucose, a hydrolytic enzyme complex is needed. This enzymatic complex includes endoglucanases, celobiohydrolases or exoglucanases and β -glycosidases. The industrial use of cellulases is quite diverse, including cotton processing, paper manufacture, detergent production, extraction of fruits pulp for juice production, enrichment of animal feed and cellulosic ethanol production. Because of its large variety of uses cellulases continued search for even better enzymes with favorable cost-benefit is necessary. Metagenomics is a technique that allows access to genetic material of microorganisms not cultured yet. Different biological systems are used in industrial applications, especially the bacterium *Escherichia coli* and the yeast *Pichia pastoris*. On this work made a screening of 10 clones provinent of a goat rumen metagenomic librarie, with a cellulose degradation phenotype on 4 different substrates: CMC LV, CMC HL, avicel and pre-treated sugar-cane bagasse, of these clones, 4 were completely sequenced and analysis of probable ORFs and proteins comparing with the existing databanks. The cloning of gene *endo 03*, provinent of Amazon soil metagenomic librarie, in the cloning vector pJET and posterior cloning at expression vector pPIC α A, to *P. pastoris*, and expression at plate. The cloning of genes *cbh I* e *cbh II*, provinent of goat rumen metagenomic librarie, at the cloning vector pJET and posterior cloning at expression vector pET21a, to *E. coli*, and expression of CBH I and CBH II proteins. The partial purification of CBH II protein and characterization of ideal temperature of 50°C with an ideal pH between 5,0 and 6,0 to the protein partly purified.

LISTA DE FIGURAS

- Figura 1: Hidrólise enzimática da celulose. Na figura podemos ver um esquema da celulose cristalina sendo hidrolisada pelas endogucanases e pelas exoglucanases ou celobiohidrolases, gerando dímeros de celobiose. Então esses dímeros de celobiose são hidrolisados pelas β -glicosidases gerando glicose. 19
- Figura 2: Mapa do vetor pCF430 que foi usado para clonar as bibliotecas metagenômicas de solo da Amazônia e de rúmen de caprinos (NEWMAN, 1999). 28
- Figura 3: Mapa do vetor comercial pJET da empresa Thermo Scientific que foi usado para clonar os genes de *cbh I*, *cbh II* e *endo 03* (THERMO SCIENTIFIC, 2012) 29
- Figura 4: Mapa do vetor comercial pET21a da empresa Novagen que foi usado para expressar os genes de *cbh I*, *cbh II* (NOVAGEN, 2003). 30
- Figura 5: Mapa do vetor comercial pPICZ α da empresa Invitrogen que foi usado para expressar o gene *endo 03* em *P. pastoris* (INVITROGEN, 2010). 31
- Figura 6: Placa contendo meio LB suplementado com bagaço de cana 1% e arabinose e corada com Congo Red e HCl para visualização dos fenótipos de degradação de celulose. A seta indica o controle negativo. 45
- Figura 7: Gel de agarose 0,8 % com a digestão dos clones A1-7, G8-10;3, E10-15, G8-127, E7-23 e B8-57, com a enzima *BamH I* para poder verificar quais apresentam padrões de digestão iguais M: marcador 1kb plus. clones 48
- Figura 8: Gel de agarose 0,8 % com todos os clones que foram mandados para o sequenciamento. M: marcador 1kb plus. 1 a 11: clones digeridos com a enzima *Pst I*. 50
- Figura 9: Alinhamento dos contigues do clone sequenciado A1-7 montados no programa Geneious. 51
- Figura 10: Alinhamento dos contigues do clone sequenciado C2-16 montados no programa Geneious. 51
- Figura 11: Alinhamento dos contigues do clone sequenciado E10-15 montados no programa Geneious. 51

Figura 12: Alinhamento dos contigues do clone sequenciado G8-127 montados no programa Geneious.....	51
Figura 13: Gel de agarose 0,8% com o resultado das PCRs do gene <i>endo 03</i> com primers para clonagem em bactéria e para clonagem em levedura 1 e 2: PCR do gene <i>endo 03</i> para levedura; 3 e 4: PCR do gene <i>endo 03</i> para bactéria; M: marcador 1kb plus.	54
Figura 14: Clones de <i>P. pastoris</i> transformados e corados com Congo Red e HCl para visualização dos fenótipos de degradação de celulose. A seta azul indica o controle negativo e as setas vermelhas indicam os clones 6 e 27 que foram escolhidos para continuação do trabalho.....	55
Figura 15: : Gel de agarose 0,8% com o resultado das PCRs dos genes <i>cbh I</i> e <i>cbh II</i> com primers para clonagem em bactéria. M: marcador 1kb plus.....	57
Figura 16: Digestão dos clones <i>cbh II</i> no vetor pET21a para a confirmação da clonagem. 1 a 5: clones de <i>cbh II</i> no vetor pET21a digeridos com as enzimas <i>Nde I</i> e <i>Xho I</i> . M: marcador 1kb plus.....	58
Figura 17: Digestão dos clones <i>cbh I</i> no vetor pET21a para a confirmação da clonagem. 1, 3 e 5: clones de <i>cbh I</i> no vetor pET21a intactos; 2, 4 e 6: clones de <i>cbh I</i> no vetor pET21a digeridos com as enzimas <i>Nde I</i> e <i>Xho I</i> . M: marcador 1kb plus.	59
Figura 18: Gel com o resultado da purificação da CBH II	60
Figura 19: Gráfico do pH ótimo da CBH II parcialmente purificada.....	61
Figura 20: Gráfico da temperatura ótima da CBH II parcialmente purificada.	62
Figura 21: Quadro resumo do trabalho de pesquisa realizado e a indicação das próximas etapas na busca de celulasas para a fabricação de biocombustíveis de 2 ^a geração	69

LISTA DE TABELAS

Tabela 1 - Exemplos de algumas das celulases comerciais que podem ser encontradas no mercado.....	20
Tabela 2 - Exemplos de bibliotecas metagenômicas. Na tabela 2 estão alguns exemplos de bibliotecas metagenômicas que foram usadas em estudos para encontrar novas celulases.....	23
Tabela 3: Sistemas heterólogos. Na tabela 3 estão alguns exemplos de sistemas heterólogos de expressão com suas vantagens, desvantagens e exemplos de usos em pesquisas.....	24
Tabela 4 - Digestão com <i>Pst</i> I dos clones da biblioteca de rúmen.....	34
Tabela 5 - Digestão com <i>Bam</i> H I dos clones com padrão de digestão similar ...	35
Tabela 6 - Ligação <i>endo</i> 03 no vetor pPICZαA	38
Tabela 7 - Ligação <i>endo</i> 03 no vetor pET21a	40
Tabela 8 - Ligação <i>cbh</i> I e <i>cbh</i> II no vetor pET21a.....	42
Tabela 9 - Resultados da primeira triagem. Na tabela 10 estão todos os clones que foram triados e os meios de cultura em que cada clone foi triado.	46
Tabela 10 - Resultados da segunda triagem. Na tabela 11 estão todos os clones que foram triados e os meios de cultura em que cada clone foi triado.	46
Tabela 11 - Resultado da terceira triagem. . Na tabela 12 estão todos os clones que foram triados e os meios de cultura em que cada clone foi triado.	49
Tabela 12 - Principais alinhamentos no BLAST P encontrados para cada um dos clones.....	52
Tabela 13 - Principais alinhamentos no Pfam encontrados para cada um dos clones.....	52
Tabela 14 - Primers construídos para o sequenciamento.....	86
Tabela 15 - Clone A1-7. Resultado do alinhamento das ORFs do clone no BLAST P.....	88
Tabela 16 - Clone C2-16. Resultado do alinhamento das ORFs do clone no BLAST P.....	90
Tabela 17 - Clone E10-15. Resultado do alinhamento das ORFs do clone no BLAST P.....	91

Tabela 18 - Clone G8-127. Resultado do alinhamento das ORFs do clone no BLAST P.....	93
---	----

ABREVIACÕES

ORFs Open Reading Frame

CMC Carboximetilcelulose

DNA Ácido Desoxirribonucleico

EC Enzyme Commission Number

EDTA Ethylene-diamine Tetraacetic Acid

g gramas

GH Glicosil Hidrolase

HCl Ácido Clorídrico

HV High Viscosity (Alta viscosidade)

IPTG Isopropyl b-D-1-thiogalactopyranoside

kDa Kilodáltons

Kg Quilogramas

LB Meio de Cultura Luria Bertrani

LV Low Viscosity (Baixa viscosidade)

MgSO₄ Sulfato de Magnésio

KH₂PO₄ Fosfato de Potássio Monobásico

K₂HPO₄ Fosfato de Potássio Dibásico

NaCl Cloreto de Sódio

pb Pares de Bases

pNPbC p-Nitrophenyl-b-D-Cellobioside

SDS Sodium Dodecyl Sulfate

TB Meio de Cultura Terrific

°C Graus Celsius

YNB Base nitrogenada de levedura sem aminoácidos e sulfato de amônio

SUMÁRIO

AGRADECIMENTOS	3
RESUMO	4
ABSTRACT	5
LISTA DE FIGURAS	6
LISTA DE TABELAS	8
ABREVIACÕES	10
SUMÁRIO	12
1 INTRODUÇÃO	16
1.1 PRODUÇÃO DE BIOETANOL NO BRASIL.....	16
1.2 LIGNOCELULOSE E SEUS CONSTITUÍNTES	17
1.3 GLICOSIL HIDROLASES E CELULASES	18
1.4 METAGENÔMICA PARA A DESCOBERTA DE NOVAS ENZIMAS.....	22
1.5 SISTEMAS DE EXPRESSÃO: <i>E. COLI</i> E <i>P. PASTORIS</i>	24
2 OBJETIVOS	26
2.1 GERAL	26
2.2 ESPECÍFICOS	26
3 MATERIAIS E MÉTODOS	27
3.1 CEPAS UTILIZADAS	27
3.2 VETORES UTILIZADOS	27
3.3 BIBLIOTECAS METAGENÔMICAS UTILIZADAS	31
3.4 CLONES DA BIBLIOTECA DE RÚMENS DE CAPRINOS	32
3.4.1 ESCOLHA DOS CLONES E CONFIRMAÇÃO DO FENÓTIPO	32
3.4.2 ANÁLISE DO PERFIL DE RESTRIÇÃO DOS INSERTOS.....	33
3.4.3 TRANSFORMAÇÃO DOS INSERTOS E RECONFIRMAÇÃO DOS FENÓTIPOS.....	35
3.4.4 EXTRAÇÃO DE DNA PLASMIDIAL PARA SEQUENCIAMENTO DOS INSERTOS	35
3.4.5 SEQUENCIAMENTO E MONTAGEM DOS PRIMERS	36
3.4.6 ANÁLISES DAS SEQUÊNCIAS DOS CLONES DA BIBLIOTECA DE RÚMEN DE CAPRINOS	36

3.5 GENE <i>ENDO 03</i>	36
3.5.1 AMPLIFICAÇÃO DO GENE <i>ENDO 03</i> PARA EXPRESSÃO EM <i>E. COLI</i>	37
3.5.2 LIGAÇÃO DO GENE <i>ENDO 03</i> NO VETOR DE CLONAGEM PJET.....	37
3.5.3 AMPLIFICAÇÃO DO GENE <i>ENDO 03</i> PARA EXPRESSÃO EM <i>P. PASTORIS</i>	38
3.5.4 LIGAÇÃO DO <i>ENDO 03</i> NO VETOR PPICZAA DE <i>P. PASTORIS</i>	38
3.5.5 TRANSFORMAÇÃO DE <i>P. PASTORIS</i> E ESCOLHA DOS CLONES TRANSFORMADOS	39
3.5.6 INDUÇÃO DE PROTEÍNAS EM <i>P. PASTORIS</i>	39
3.5.7 LIGAÇÃO DO GENE NO VETOR DE CLONAGEM PJET DE <i>E. COLI</i>	40
3.5.8 LIGAÇÃO DO GENE E DO <i>ENDO 03</i> NO VETOR PÉT21A DE EXPRESSÃO EM <i>E. COLI</i>	40
3.6 GENES <i>CBH I</i> E <i>CBH II</i>	41
3.6.1 AMPLIFICAÇÃO DOS GENES <i>CBH I</i> E <i>CBH II</i> PARA EXPRESSÃO EM <i>E. COLI</i>	41
3.6.2 LIGAÇÃO DOS GENES <i>CBH I</i> E <i>CBH II</i> NO VETOR DE CLONAGEM PJET	42
3.6.3 LIGAÇÃO DOS GENES <i>CBH I</i> E <i>CBH II</i> NO VETOR PET21A DE EXPRESSÃO EM <i>E. COLI</i>	42
3.6.4 TESTE DE ATIVIDADE DAS PROTEÍNAS COM PNP-CELOBIOSÍDEO EM <i>E. COLI</i>	43
3.6.5 PURIFICAÇÃO DA <i>CBH II</i> PRODUZIDA EM <i>E. COLI</i>	43
3.6.6 EFEITO DA TEMPERATURA E PH NA ATIVIDADE.....	44
4 RESULTADOS.....	45
4.1 CLONES DA BIBLIOTECA DE RÚMEN DE CAPRINOS.....	45
4.1.2 ESCOLHA DOS CLONES E CONFIRMAÇÃO DO FENÓTIPO	45
4.1.3 EXTRAÇÃO DE DNA PLASMIDIAL E ANÁLISE DOS INSERTOS.....	47
4.1.4 TRANSFORMAÇÃO DOS CLONES E RECONFIRMAÇÃO DOS FENÓTIPOS	48
4.1.5 EXTRAÇÃO DE DNA PLASMIDIAL PARA SEQUENCIAMENTO DOS INSERTOS	49
4.1.6 ANÁLISES DAS SEQUÊNCIAS DOS CLONES DA BIBLIOTECA DE RÚMEN DE CAPRINOS	50
CLONE	52
ORFS ENCONTRADAS	52
ORFS DE INTERESSE.....	52
E-VALUE	52
COBERTURA	52
ALINHAMENTO DO BLAST P DE INTERESSE	52
A1-7	52
10	52
168AA.....	52
5.5	52
64%	52
GLYOXYLASE-RELATED ZINC-DEPENDENT HYDROLASE [GEOBACTER METALLIREducENS GS-	
15] 31% *	52

CELLULOSOME-ANCHORING PROTEIN, PARTIAL [PEPTONIPHILUS LACRIMALIS] 36%	52
PREDICTED: GLUCAN ENDO-1,3-BETA-GLUCOSIDASE 8-LIKE [GLYCINE MAX] 28%	52
C2-16	52
7	52
274AA.....	52
7.0	52
50%.....	52
GLYCOSYL HYDROLASE [DYSGNOMONAS GADEI] >GB EGK02235.1 HYPOTHETICAL	
PROTEIN HMPREF9455_01505 [DYSGNOMONAS GADEI ATCC BAA-286] 24%	52
E10-15.....	52
13	52
551AA.....	52
5E-82.....	52
62%.....	52
HYPOTHETICAL PROTEIN HMPREF9332_01082 [ALLOPREVOTELLA RAVA F0323]	52
43%	52
G8-127	52
9	52
821AA.....	52
0.0	52
98%.....	52
HELICASE C-TERMINAL DOMAIN PROTEIN [PREVOTELLA MELANINOGENICA ATCC 25845]	
74%.....	52
4.2 GENE ENDO 03.....	53
4.2.1 AMPLIFICAÇÃO DO GENE <i>ENDO 03</i> PARA EXPRESSÃO EM <i>E. COLI</i>	53
4.2.2 AMPLIFICAÇÃO DO GENE <i>ENDO 03</i> PARA EXPRESSÃO EM <i>P. PASTORIS</i>	54
4.2.3 TRANSFORMAÇÃO DE <i>P. PASTORIS</i> E ESCOLHA DOS CLONES TRANSFORMADOS	54
4.2.4 EXPRESSÃO DE PROTEÍNAS EM <i>P. PASTORIS</i>	55
4.2.5 LIGAÇÃO DO GENE E DO <i>ENDO 03</i> NO VETOR PÉT21A DE EXPRESSÃO EM <i>E. COLI</i>	56
4.3 GENES CBH I E CBH II	56
4.3.1 AMPLIFICAÇÃO DOS GENES <i>CBH I</i> E <i>CBH II</i> PARA EXPRESSÃO EM <i>E. COLI</i>	56
4.3.2 LIGAÇÃO DOS GENES <i>CBH I</i> E <i>CBH II</i> NO VETOR PÉT21A DE EXPRESSÃO EM <i>E. COLI</i>	57
4.3.3 TESTE DE ATIVIDADE DAS PROTEÍNAS COM PNP-C EM <i>E. COLI</i>	59
4.3.4 PURIFICAÇÃO DE PROTEÍNAS PRODUZIDAS EM <i>E. COLI</i>	60
4.3.5 CARACTERIZAÇÃO DE TEMPERATURA E PH IDEAIS DA PROTEÍNA	61
5 DISCUSSÃO	63

6 CONCLUSÕES	68
7 PERSPECTIVAS.....	70
8 REFERÊNCIAS BIBLIOGRÁFICAS	71
9 ANEXOS	77

1 INTRODUÇÃO

1.1 Produção de Bioetanol no Brasil

Biocombustíveis são definidos como hidrocarbonetos de valor energético produzidos a partir de matéria orgânica gerada por seres vivos (biomassa). São assim considerados fontes renováveis de energia (WOOLF et al., 2014) com grande potencial de aproveitamento, sobretudo como substitutos dos combustíveis fósseis no setor de transportes.

Um dos biocombustíveis mais utilizados é o bioetanol ou, simplesmente, etanol. A maior parte do etanol, que é produzido no Brasil, vem da fermentação do caldo de cana açúcar, que é rico em sacarose. A sacarose é fermentada por microrganismos gerando, ao final do processo, um álcool (etanol ou metanol). De acordo com as matérias-primas e as tecnologias de produção, o etanol pode ser classificado da seguinte maneira:

Primeira geração – são produzidos a partir de açúcares amiláceos de uma parte de plantas tradicionalmente cultivadas.

Segunda geração – são produzidos a partir de açúcares mais complexos (polissacarídeos) contidos nos resíduos agroindustriais e de certas gramíneas de alta produção de biomassa. São também conhecido como biocombustíveis celulósicos.

Terceira geração – são produzidos a partir de matéria-prima e microrganismos modificados geneticamente para facilitar a fermentação e demais etapas do processo e, assim, tornar a produção mais eficiente (MACHADO et al., 2013).

Para aumentar a produção de etanol no país, sem aumentar às áreas de cultivo da cana-de-açúcar, deverá ser viabilizado, técnica e economicamente, o aproveitamento dos açúcares ainda presentes no bagaço da cana. O etanol produzido a partir do bagaço da cana é chamado de segunda geração, como já foi dito. Estima-se que o uso do bagaço para a produção de etanol de segunda geração eleve a sua produção em 30 a 40% para a mesma área plantada (PACHECO., 2011). O desafio consiste em desenvolver um processo químico ou

bioquímico que degrade a celulose (polissacarídeos) em açúcares menos elaborados (pentoses ou hexoses), separando-o inclusive da lignina.

1.2 Lignocelulose e seus constituintes

A lignocelulose é formada por fibras de celulose emaranhadas por hemicelulose e lignina. A lignocelulose está associada à parede celular das células vegetais proporcionando-lhes sustentação. A quantidade de celulose, hemicelulose e lignina presente na parede celular varia de acordo com cada planta. O bagaço de cana, por exemplo, é constituído por 10-25% de lignina, 40-60% de celulose e de 30-40% de hemicelulose (CGEE 2008).

A celulose, por estar presente na parede celular dos vegetais, é o polissacarídeo mais abundante na Terra. É um homopolissacarídeo formado por moléculas de D-glicopiranosose ligadas umas às outras por ligações do tipo β -1,4. A celulose pode ser hidrolisada em unidades menores chamadas de celobiose, que por sua vez, pode ser convertida em resíduos de glicose (KUMAR; SINGH; SINGH, 2008).

A hemicelulose é formada por diferentes polímeros de carboidratos formados por diferentes açúcares (pentose e hexoses) ligados uns aos outros por diferentes tipos de ligações (HIMMEL et al .,2010). Hemiceluloses são normalmente mais fáceis de hidrolizar quando comparadas com celulosas, porém certas estruturas oligoméricas são recalcitrantes devido às suas ramificações complexas e aos seus padrões de acetilação (AGGER; VIKSØ-NIELSEN; MEYER, 2010).

A lignina é um heteropolímero hidrofóbico constituído por três monolignóis. Na lignocelulose, a lignina está ligada aos carboidratos por ligações do tipo éter ou éster (CHEN; DIXON, 2007).

A heterogênese da composição das paredes de plantas, a inacessibilidade e a recalcitrância dos seus componentes são fatores limitantes para a produção de etanol. Normalmente, as enzimas possuem uma grande dificuldade em degradar os seus substratos (HORN et al., 2012). A busca ou o desenvolvimento de enzimas eficazes é o objeto de diversas pesquisas.

1.3 Glicosil hidrolases e celulases

As glicosil hidrolases (EC 3.2.1.x) são enzimas chaves no metabolismo de carboidratos e podem ser encontradas nos domínios: Archea, Bactéria e Eucária (HENRISSAT, 1991). Para seu estudo, utiliza-se o CAZy (Enzimas carboidrato ativas) que é um banco de dados com informações sobre enzimas que atuam em carboidratos. Nele se podem fazer análises genômicas, estruturais e bioquímicas e compartilhar dados. O CAZy permitiu, ainda, a elaboração de uma classificação que se aplica às glicosil hidrolases e se baseia nos dados do sequenciamento genético. Inicialmente, estas foram distribuídas em 35 famílias, porém trabalhos posteriores já aumentaram o número e incluíram outros tipos de enzimas. É neste grande grupo que as pesquisas se concentram na busca daquelas que são capazes de hidrolizar a celulose em resíduos de glicose e permitir, assim, o seu aproveitamento na produção de etanol. Essas enzimas de interesse são as celulases, mais especificamente: as endoglucanases (E.C 3.2.1.4), as celobiohidrolases ou exoglucanases (EC 3.2.1.91) e as β -glicosidases (EC 3.2.1.21).

As endoglucanases (E.C 3.2.1.4) hidrolisam regiões amorfas da celulose, e catalisam a hidrólise de sítios internos produzindo celoligossacarídeos. Esses celoligossacarídeos são, a seguir, hidrolisados gerando dímeros de celobiose pela ação das enzimas celobiohidrolases ou exoglucanases (EC 3.2.1.91). Por último, a celobiose é convertida em glicose pela enzima β -glicosidase (EC 3.2.1.21) (PHITSUWAN et al., 2013).

De acordo com a CAZy, as celobiohidrolases (CBH) são classificadas nas famílias 5, 6 e 9 de hidrolases, e endoglucanases são classificadas nas famílias 5, 6, 7, 8, 9, 12, 26, 44, 45, 51, 48, 71 e 124. As β -glicosidases são encontradas nas famílias 1, 3, 5, 9, 30 e 116 (NC-IUBMB/ CAZy).

Nos estudos baseados na classificação CAZy, verificou-se que muitas dessas famílias têm estruturas tridimensionais completamente diferentes, isso indica que as celulases tiveram uma evolução convergente de diversos ancestrais (MBA MEDIE et al., 2012) .

Outra característica que distingue as celulases das demais glicosil hidrolases é a sua habilidade de hidrolisar ligações do tipo β -1, 4 glicosídicas entre resíduos de glicose (Figura 1). A quebra enzimática das ligações β -1 4-glicosídicas na celulose se faz por meio de um mecanismo de hidrólise ácida,

usando um doador de prótons e um nucleófilo ou uma base. Os produtos da hidrólise podem resultar na inversão ou retenção da configuração alomérica do carbono-1 na extremidade redutora (LYND et al., 2002).

O aspecto geral da maioria das celulasas é de uma estrutura modular que inclui os sítios catalíticos e de ligação ao carboidrato conhecido como módulos catalíticos de ligação ao carboidrato (CBMs). O CBM se liga à superfície da celulose para facilitar a hidrólise e trazer o domínio catalítico para perto do substrato. Os CBMs são particularmente importantes para dar início à ação das celobiohidrolases (LYND et al., 2002).

Na natureza são encontradas em fungos, bactérias, plantas e insetos. Fungos filamentosos possuem hifas que ajudam na interação organismo-celulose, e por isso, normalmente produzem celulasas livres. Já microrganismos anaeróbicos normalmente têm o que é chamado de celulosoma: um complexo de enzimas celulolíticas que atuam em conjunto. Neste celulosoma estão os vários tipos de celulasas, bem como outras enzimas como xilanases. As celulasas podem ser produzidas em forma isolada ou em forma de celulosoma pelos micro-organismos (BAYER et al., 1998).

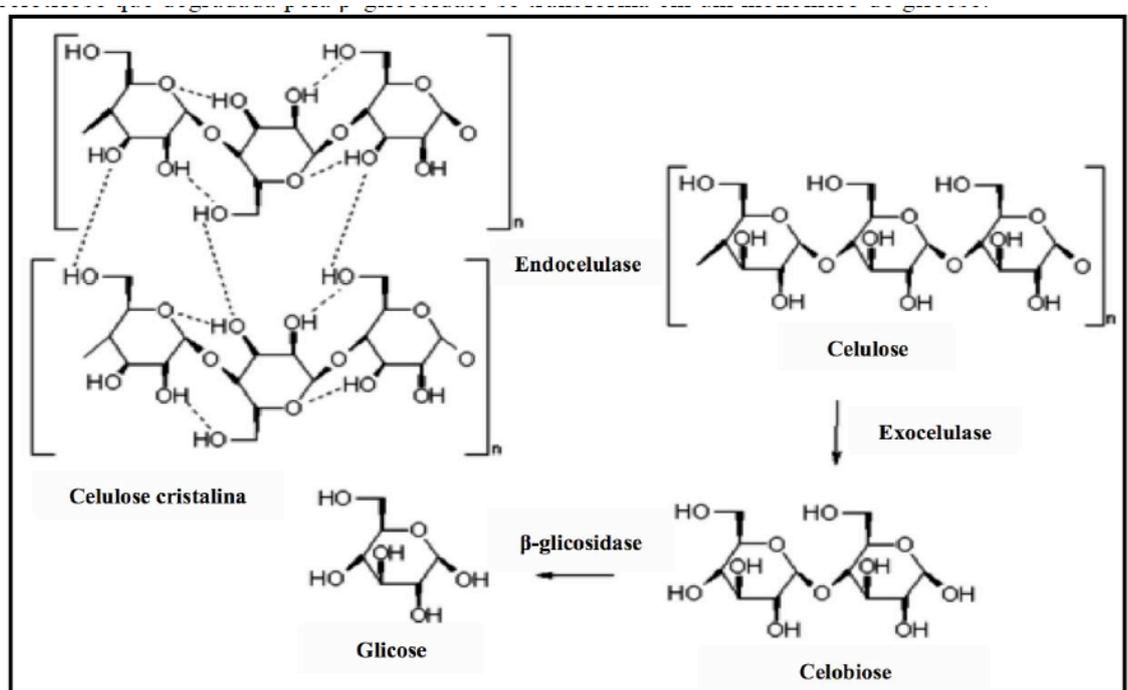


Ilustração: Ohana Costa

Figura 1: Hidrólise enzimática da celulose. Na figura podemos ver um esquema da celulose cristalina sendo hidrolisada pelas endogucanases e pelas exogucanases ou

celobiohidrolases, gerando dímeros de celobiose. Então esses dímeros de celobiose são hidrolisados pelas β -glicosidasases gerando glicose.

Atualmente existem no mercado celulases comerciais que são utilizadas para a hidrólise da celulose. As duas maiores empresas que atualmente produzem celulases comerciais são a Genencor e a Novozyme, tendo sido as principais responsáveis pela diminuição nos custos das celulases em razão do contínuo investimento em novas técnicas biotecnológicas (SINGHANIA et al., 2010). A Tabela 1 mostra algumas das celulases existentes no mercado

Tabela 1 - Exemplos de algumas das celulases comerciais que podem ser encontradas no mercado.

Enzima	Empresa	Organismo
Carezyme 1000L	Sigma-aldrich/ Novozyme	<i>A. Níger</i>
Novozyme 188	Novozyme	<i>A. Níger</i>
Cellulase 2000L	Rhodia-Danisco	<i>T. longibrachiatu m/ T. reesei</i>
Rohament CL	Rohm-AB Enzymes	<i>T. longibrachiatu m/ T. reesei</i>
Viscostar 150L	Dyadic	<i>T. longibrachiatu m/ T. reesei</i>
Multifect CL	Genecor Intl	<i>T. reesei</i>
Bio- feed β L	Novozyme	<i>T. longibrachiatu m/ T. reesei</i>
Energex L	Novozyme	<i>T. longibrachiatu m/ T. reesei</i>
Ultraflo L	Novozyme	<i>T. longibrachiatu m/ T. reesei</i>
Viscozyme L	Novozyme	<i>T. longibrachiatu m/ T. reesei</i>
Cellulyve 50L	Lyven	<i>T. longibrachiatu</i>

		<i>m/ T. reesei</i>
GC 440	Genencor-Danisco	<i>T. longibrachiatu</i> <i>m/ T. reesei</i>
GC 880	Genencor	<i>T. longibrachiatu</i> <i>m/ T. reesei</i>
Spezyme CP	Genencor	<i>T. longibrachiatu</i> <i>m/ T. reesei</i>
GC 220	Genencor	<i>T. longibrachiatu</i> <i>m/ T. reesei</i>
Cellulase AP30K	Amano Enzyme	<i>A. Níger</i>
Cellulase TRL	Solvay Enzymes	<i>T. longibrachiatu</i> <i>m/ T. reesei</i>
Econase CE	Alko-EDC	<i>T. longibrachiatu</i> <i>m/ T. reesei</i>
Cellulase TAP106	Amano Enzyme	<i>T. viride</i>
Cellucast 1.5L	Sigma-aldrich/ Novozyme	<i>T. reesei</i>
Biocellulase TRI	Quest Intl.	<i>T. longibrachiatu</i> <i>m/ T. reesei</i>
Biocellulase A	Quest Intl.	<i>A. Níger</i>
Ultra-low microbial (ULM)	logen	<i>T. longibrachiatu</i> <i>m/ T. reesei</i>
Accellerase	Genencor	<i>T. reesei</i>
Cellic CTec2	Novozyme	*
Cellic CTec3	Novozyme	*

Adaptado (SINGHANIA et al., 2010).*Como é um coquetel de enzimas não tem um organismo específico.

O uso industrial das celulasas é bem diverso, abrangendo desde o processamento de algodão, a fabricação de papel, produção de detergentes, extração de polpa de frutas para a fabricação de sucos, até o enriquecimento de ração para animais e na produção de bioetanol a partir da biomassa (BHAT, 2000). Devido a essa grande variedade de usos para as celulasas, há uma

busca contínua por enzimas que tenham uma melhor relação custo-benefício com elevada eficiência.

Para fins industriais, as celulases de fungos filamentosos, em especial da espécie *Trichoderma reesei*, são as mais estudadas. Entretanto *T. reesei* possui uma baixa produção de β -glicosidases. Assim, no processo industrial, normalmente recorre-se a uma suplementação de β -glicosidases oriundas de *Aspergillus*. β -glicosidases de *Aspergillus* são conhecidas por serem mais tolerantes à inibição por glicose (LYND et al., 2002).

Uma das grandes dificuldades no uso de celulases, com ações de endo e exoglucanases, é a inibição, em um sistema de feedback, por celobiose, que essas enzimas sofrem. Assim, a glicose, que é o produto final esperado, não se acumula rapidamente (KUMAR; SINGH; SINGH, 2008). Uma das soluções para esse problema seria a utilização de enzimas recombinantes com uma baixa resposta à inibição por celobiose. Essa seria uma importante motivação para a descoberta e o desenvolvimento de novas enzimas recombinantes.

1.4 Metagenômica para a descoberta de novas enzimas

A metagenômica é uma técnica que permite o acesso ao material genético de microrganismos não cultiváveis, criando uma biblioteca com esses dados. Essa técnica apresenta grande valor tendo em vista que somente 1% dos organismos são cultiváveis. Ou seja, 99% dos microrganismos ficariam de fora de estudos caso não se utilizasse a metagenômica (XING; ZHANG; HUANG, 2012).

Pode se utilizar a metagenômica na busca de novas enzimas que degradam a celulose (CUNHA et al., 2011). Para tanto, a busca deverá ser feita em ambientes com altos índices do substrato em questão. Por isso, normalmente são feitas bibliotecas de solos, onde a concentração de material vegetal é alta, e bibliotecas de rúmen de herbívoros, que possuem uma microbiota altamente especializada na degradação de biomassa.

Para a criação de bibliotecas metagenômicas, a primeira etapa consiste na extração total do DNA de amostras do ambiente escolhido. Depois da extração, o DNA é fragmentado com o auxílio de enzimas de restrição e subclonado em um vetor que é transformado em um hospedeiro. As células do hospedeiro, com os insertos, são então cultivadas e a biblioteca metagenômica é

montada (HANDELSMAN et al., 1998). Na Tabela 2 estão alguns exemplos de bibliotecas metagenômicas que foram usadas em estudos para encontrar novas celulases.

Tabela 2 - Exemplos de bibliotecas metagenômicas. Na tabela 2 estão alguns exemplos de bibliotecas metagenômicas que foram usadas em estudos para encontrar novas celulases.

Ambiente usado	Tipo de vetor	Tamanho dos Fragmentos	Referências
Intestino de porcos	pCR4Blunt-TOPO	4.2 Kb	(WANG et al., 2012)
Rúmen de bovinos	Fosmídeo pCC2FOS	60 Kb	(RASHAMUSE et al., 2013)
Rúmen de vacas leiteiras	pSMART BAC	50 Kb	(GONG et al., 2012)
Rúmen de búfalos	Fosmídeo pCC1FOS	-	(NGUYEN et al., 2012)
Solo das pradarias na Alemanha	Fosmídeo pCC1FOS	16 - 36 Kb	(NACKE et al., 2012)
Intestino de Microcerotermes SP	Fosmídeo pCC1FOS	20 - 40 Kb	(NIMCHUA et al., 2012)
Rúmen de búfalos	Cosmídeo pWEB-TNC	20-50 Kb	(DUAN et al., 2009)
Rúmen de vacas leiteiras	Fosmídeo pCCFOS	-	(FERRER et al., 2012)
Solo Ladakh	pUC19	3,8 Kb	(BHAT et al., 2013)
Solo da Coréia	pUC118/HincII/BAP	-	(KIM et al., 2008)
Solo vermelho de China	pCUGIBAC _I	200Mb	(LIU et al., 2011)
Rúmen de iaque	Cosmídeo pWEB-TNC	30-40 Kb	(CHANG et al., 2011)
Estrume de porco e resíduos de cogumelos	Fosmídeo pCC2FOS	40 Kb	(KWON et al., 2010)
Amostras de bioreator	Cosmídeo pWEB-TNC	40-50 Kb	(JIANG et al., 2010)

Solo poluído do suldeste da China	pGEM-3Zf	1-15 Kb	(JIANG et al., 2009)
Solo da Alemanha	Cosmídio pWE15	20-40 Kb	(VOGET et al., 2003)
Amostras de bioreator	Cosmídio	23-55 Kb	(GENG et al., 2012)
Intestino de cupins	Fosmídio pCC2FOS	-	(LIU et al., 2011)
Ceco de coelhos	Cosmídio pWEB-TNC	30-50 Kb	(FENG et al., 2007)

Na pesquisa feita neste trabalho por novas celulasas de interesse comercial foram utilizadas duas bibliotecas metagenômicas: a biblioteca de pequenos insertos de rúmen de caprinos (CUNHA et al., 2011) e a biblioteca de pequenos insertos de solo da Amazônia (PAULAN, 2011).

1.5 Sistemas de expressão: *E. coli* e *P. pastoris*

Para a produção enzimas recombinantes, em escala industrial, podem ser usadas uma vasta gama de organismos ou células hospedeiras. A escolha normalmente é feita baseada em diversos critérios: a relação custo-benefício, se a proteína tem que estar na forma solúvel, sua facilidade de purificação, e se, ao final, ela possui atividade ou será cristalizada (HARTLEY, 2006). Por exemplo, enzimas farmacoterápicas, que visam o uso em seres humanos ou animais, são normalmente expressas em cultura de células de mamíferos, pois modificações pós-traducionais erradas podem acarretar em rejeição ao fármaco. A Tabela 3 mostra alguns exemplos de sistemas heterólogos usados para expressar proteínas recombinantes.

Tabela 3: Sistemas heterólogos. Na tabela 3 estão alguns exemplos de sistemas heterólogos de expressão com suas vantagens, desvantagens e exemplos de usos em pesquisas.

Sistema de expressão	Vantagens	Desvantagens	Exemplos de usos
Bactérias	Baixo custo e rápido crescimento	Não faz modificações pós-traducionais e formação de	Expressão de β -glicosidase (LI et al., 2012) .

		corpos de inclusão	
Levedura e fungos filamentosos	Forma solúvel da proteína e faz algumas modificações pós-traducionais	Hiperglicosilação das proteínas	Expressão de glicose oxidase (GUO et al., 2010).
Células de plantas	Modificações pós-traducionais mais parecidas com mamíferos	Incorporação de açúcares potencialmente tóxicos e alto custo	Expressão de enzimas lignocelulolíticas (MIR et al., 2014)
Células de mamíferos	Modificações pós-traducionais de mamíferos	Alto custo e altas taxas de contaminação	Expressão de neuroglobinas (QIN et al., 2012).

Para a produção industrial de proteínas recombinantes, vários sistemas biológicos são empregados, principalmente a bactéria *Escherichia coli* e a levedura *Pichia pastoris* (HARTLEY, 2006). *E. coli* é amplamente utilizada para clonagem molecular em pesquisas na área de biologia molecular. Seu uso se deve, principalmente, ao rápido crescimento, ao baixo custo e por causar o mínimo de modificações no produto final (HARTLEY, 2006). Já *P. pastoris* é uma levedura largamente utilizada como sistema para expressão de proteínas heterólogas (CREGG et al., 2000). Seu uso é indicado quando se deseja produzir proteínas em grandes quantidades na forma solúvel, daí sua grande aceitação. Uma das grandes vantagens da expressão de proteínas recombinantes em *P. pastoris* é que ela possui a capacidade de fazer modificações pós traducionais tipicamente associadas a organismos eucariotos como processamento de sequências sinais, dobramento, formação de pontes dissulfeto, N-glicosilações e O-glicosilações (CROGNALÉ 2006 apud CRUEGER 1990).

2 OBJETIVOS

2.1 Geral

Sequenciar, analisar, clonar e expressar genes de celulases e hemicelulases oriundos de bibliotecas metagenômicas de pequenos insertos de solo da Amazônia e de rúmen de caprino.

2.2 Específicos

- Confirmar fenótipos observados anteriormente em clones metagenômicos com atividade positiva para celulase provenientes de biblioteca de rúmen caprino;
- Identificar se alguns dos clones são repetidos através de análise eletroforética do padrão de digestão com enzimas de restrição;
- Escolher clones provenientes da biblioteca de rúmen de caprinos para o sequenciamento;
- Sequenciar clones metagenômicos com atividade celulolítica;
- Analisar o resultado do sequenciamento, identificando qual é a provável ORF responsável pela atividade enzimática do clone;
- Realizar análise filogenética e de resíduos conservados com a sequência da provável proteína com atividade celulase;
- Clonagem e expressão em *E.coli* os genes *cbh I* e *cbh II*, proveniente da biblioteca metagenômica de rúmen caprino;
- Clonar e expressar em *P. pastoris* o gene *endo 03*, proveniente da biblioteca metagenômica de solo de Amazônia;
- Purificação de CBH II em coluna de afinidade e caracterização enzimática de pH e temperaturas ideais.

3 MATERIAIS E MÉTODOS

Em termos gerais, a pesquisa se desenvolveu em duas frentes de trabalho, cada qual com seus métodos e materiais aplicáveis: a busca por genes responsáveis pela produção de celulases em biblioteca metagenômica de pequenos insertos de rúmen de caprino; e o estudo da possibilidade de aproveitamento dos genes *cbh I*, *cbh II* e *endo 03* para a produção, em escala comercial, de celulase através da expressão em *E coli* e *P pastoris*.

3.1 CEPAS UTILIZADAS

E. coli:

EPI300 para clonagem e BL21 para a expressão.

P. pastoris:

X-33 para expressão em leveduras.

3.2 VETORES UTILIZADOS

Vetor da biblioteca de rúmen de caprino pCF430:

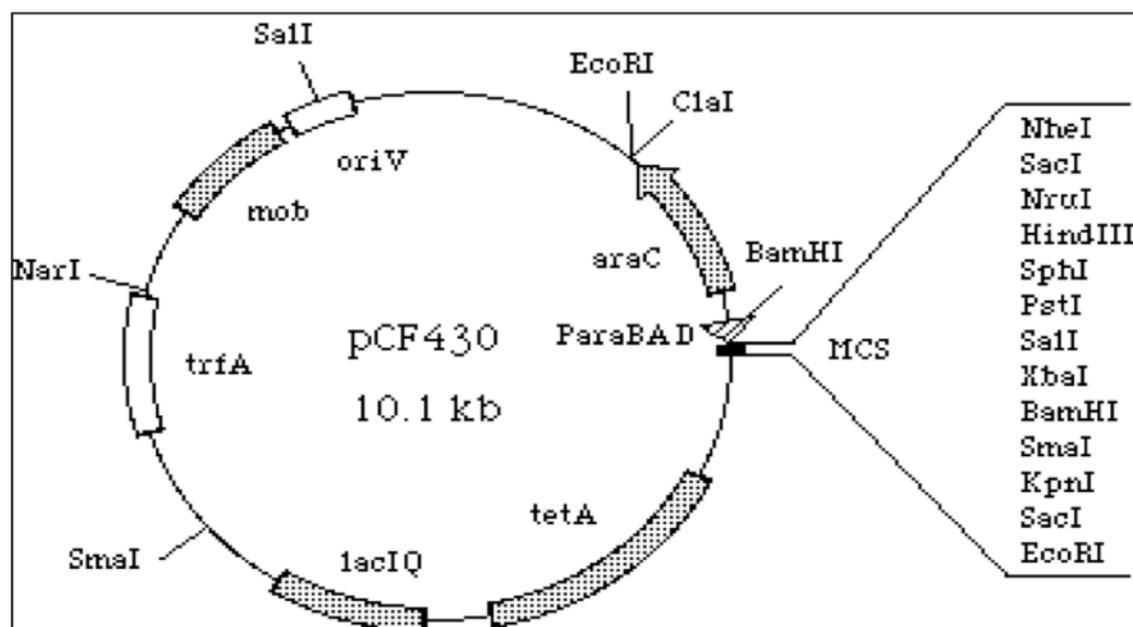


Figura 2: Mapa do vetor pCF430 que foi usado para clonar as bibliotecas metagenômicas de solo da Amazônia e de rúmen de caprinos (NEWMAN, 1999).

Vetor de clonagem pJET (Thermo Scientific):

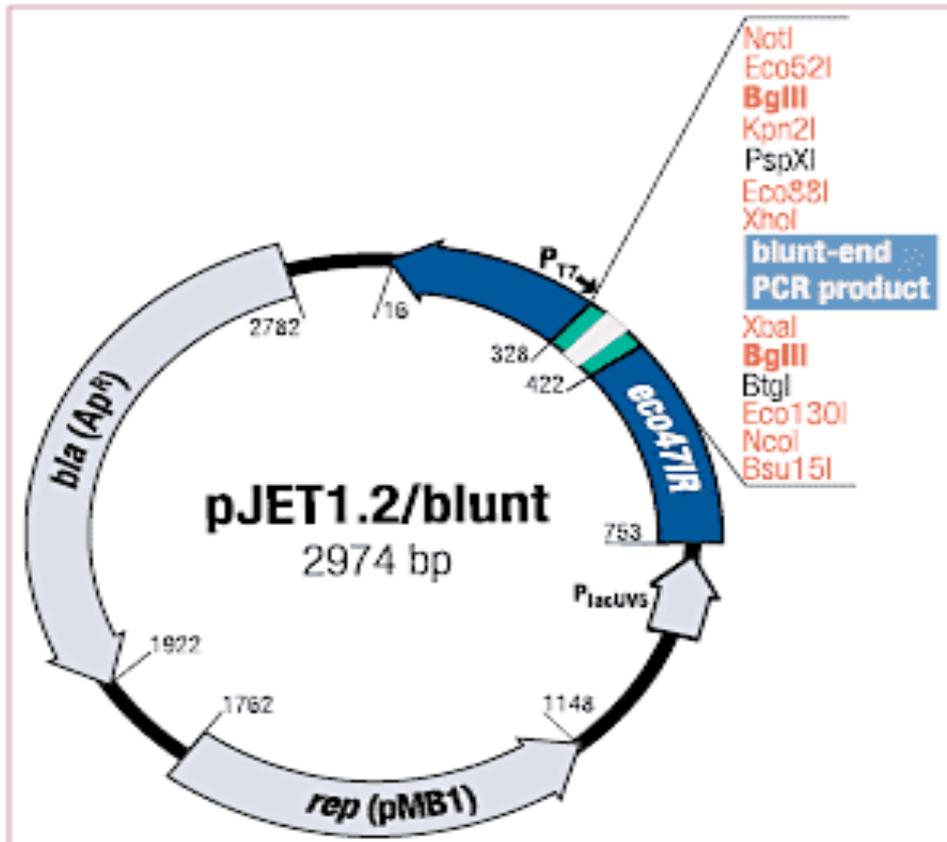


Figura 3: Mapa do vetor comercial pJET da empresa Thermo Scientific que foi usado para clonar os genes de *cbh I*, *cbh II* e *endo 03* (THERMO SCIENTIFIC, 2012)

Vetor de expressão em *E. coli* pET21a:

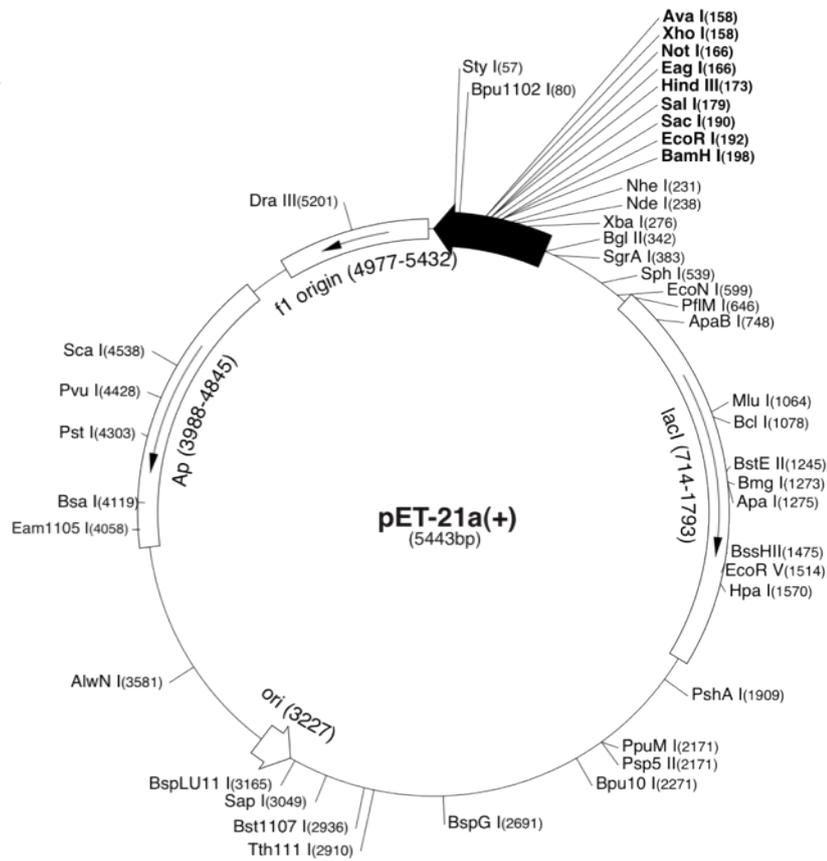


Figura 4: Mapa do vetor comercial pET21a da empresa Novagen que foi usado para expressar os genes de *cbh I*, *cbh II* (NOVAGEN, 2003).

Vetor de expressão em *P. pastoris* pPICZ α :

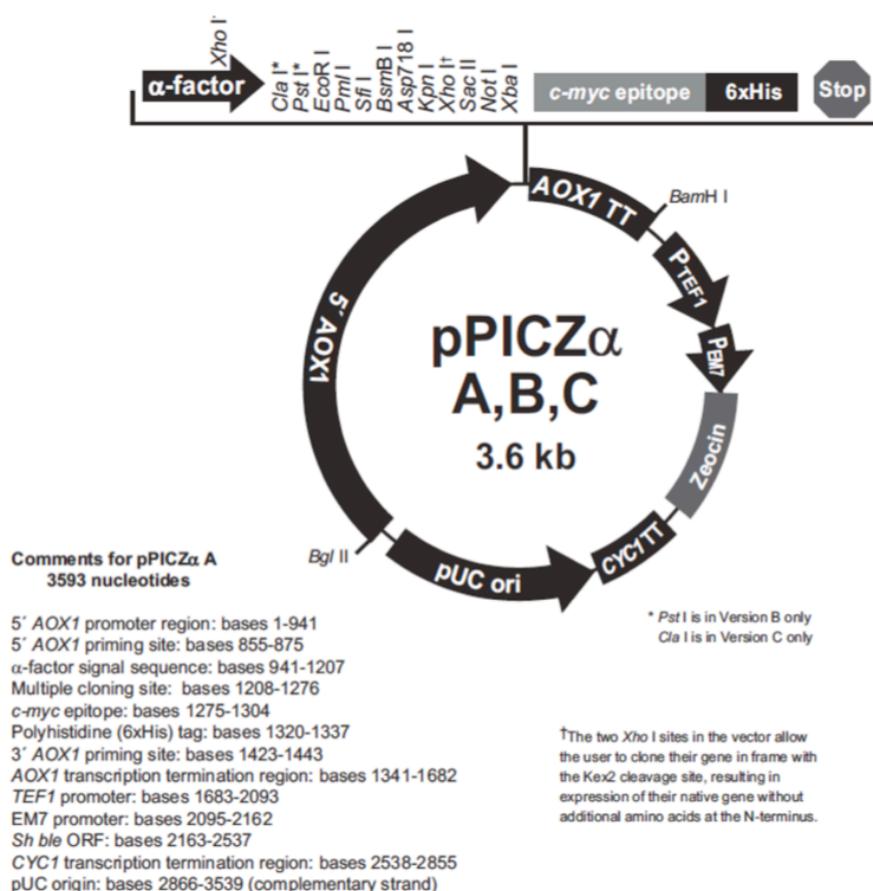


Figura 5: Mapa do vetor comercial pPICZ α da empresa Invitrogen que foi usado para expressar o gene *endo 03* em *P. pastoris* (INVITROGEN, 2010).

3.3 BIBLIOTECAS METAGENÔNICAS UTILIZADAS

A biblioteca metagenômica de pequenos insertos de rúmen de caprinos utilizada nesse trabalho foi construída pela aluna Isabel Cunha (CUNHA et al., 2011), (CUNHA et al., 2009) a partir da porção sólida coletada do rúmen de caprinos da raça Moxotó. Essa raça de caprinos vive no clima semi-árido da região Nordeste e possui uma dieta generalista se alimentando de qualquer tipo de vegetação disponível. O DNA total foi extraído e então digerido com a enzima de restrição *Pst* I e selecionados fragmentos com tamanhos de 2 a 8 Kb. Esses fragmentos foram então clonados no vetor pCF430 linearizado e então

transformados em cepas de *E.coli* EPI 300 constituindo a biblioteca metagenômica de rúmen de caprinos.

A biblioteca de pequenos insertos de solo da Amazônia utilizada nesse trabalho foi construída pela aluna Shelly Paulan (PAULAN, 2011) a partir de amostras coletadas na floresta nativa do município de Manaus. A biblioteca foi construída a partir da microbiota existente na porção mais superficial do solo (5 cm). É na superfície do solo que se acumula a matéria orgânica rica em celulose proveniente da vegetação e, portanto, é também ali que se encontram a maioria dos microrganismos capazes de produzir as enzimas necessárias para a sua hidrólise. O DNA total da amostra foi extraído e depois digerido com a enzima de restrição *Pst* I. Foram então selecionados fragmentos de 3 a 8 Kb e clonados no vetor pCF430 linearizado. O vetor, ligado aos fragmentos foi transformado em cepas de *E. coli* EPI 300 constituindo assim uma biblioteca metagenômica da microbiota encontrada no solo da Amazônia.

3.4 CLONES DA BIBLIOTECA DE RÚMENS DE CAPRINOS

3.4.1 Escolha dos clones e confirmação do fenótipo

Foram escolhidos 12 clones da biblioteca de rúmen de caprinos para se fazer o trabalho: A1-7; G2-8; G8-10; E10-15; C2-16; H11-35; G8-127; B8-57; G10-36; B3-24; E2-23; A7-93.

Os clones foram escolhidos a partir de uma primeira triagem feita pelos alunos do Laboratório de Ciências Genômicas e Biotecnologias da Universidade Pontifícia Católica de Brasília na biblioteca metagenômica de rúmen de caprinos. O critério de seleção utilizado foi o grau de atividade nos seguintes substratos: CMC alta viscosidade; CMC baixa viscosidade; xilana; avicel e bagaço de cana pré-tratado. Outro aspecto considerado foi a presença de atividade sem a adição de arabinose, indicando assim que o inserto possui um promotor próprio.

Após a escolha dos clones, uma nova análise dos fenótipos para a atividade de celulase confirmou a atividade positiva dos mesmos. O ensaio foi feito com 5 colônias isoladas, escolhidas de cada clone, e passadas para placas de petri contendo 5 meios diferentes. Os meios utilizados foram: LB ágar (peptona 1%, NaCl 1%, extrato de levedura 0,5% e ágar 1,5%) suplementado com CMC baixa viscosidade (1%); LB ágar suplementado com CMC alta

viscosidade (1%); LB ágar suplementado com xilana (1%); LB ágar suplementado com avicel (1%) e LB ágar suplementado com bagaço de cana pré-tratado (1%).

As placas com as colônias foram colocadas na estufa 37°C por 18 horas e depois armazenadas à temperatura ambiente por 12 dias. Após 12 dias, as placas foram lavadas com água destilada, retirando-se, assim, as colônias de bactérias, para evitar que o corante se ligasse aos restos celulares e prejudicasse a análise dos halos, gerando um possível falso positivo. Em seguida, uma solução de Congo Red (0,1%) foi colocada até cobrir completamente a placa e deixada por 10 minutos. Após o descarte dessa solução, uma nova solução de NaCl 0,1 M foi usada para lavar as placas para retirar o excesso de corante por 1h e 30 minutos. Por último, uma solução de HCl (1 M) foi colocada por 3 minutos para intensificar a coloração.

O pré-tratamento do bagaço de cana foi realizado por colaboradores da Embrapa Agroenergia. Para tal foram pesados 200 g de bagaço-de-cana e feita uma solução Organosolv 70% água e 30% etanol. Como a diluição desses 200 g foi de 7,5 vezes, resultou em uma mistura para pré-tratamento com 1.500 mL, sendo 450 mL de Etanol e 1.050 mL de água e mais 22,5 mL de ácido sulfúrico (1,5%). A mistura foi autoclavada durante 30 minutos (121°C e 1 atm) e, em seguida, filtrada, resultando em uma fração líquida e uma fração sólida. A fração sólida foi macerada com nitrogênio líquido e depois usada para fazer os meios.

Todo o procedimento foi repetido uma segunda vez para reconfirmar os resultados obtidos.

3.4.2 Análise do perfil de restrição dos insertos

O DNA plasmidial dos clones contendo os insertos foi extraído usando o protocolo de extração de DNA plasmidial adaptado (SAMBROOK AND RUSSEL; 2001). Do inóculo, 3 mL foram centrifugados a 10.000 g por 2 minutos. O sedimentado foi então lavado em 500 µL de STE (Tris-Cl pH 8,0 10 mM; NaCl 0,1 mM; EDTA pH 8,0 1 mM) gelado e centrifugado novamente nas mesmas condições anteriores. O sedimentado foi ressuspenso em 250 µL de solução I (Tris -HCl pH 8,0 25 mM; EDTA 20 mM; Glicose 50 mM). Foi adicionado 250 µL de solução II (NaOH 0,2 M; SDS 1%) e deixado por 5 minutos na bancada. Adicionou-se 350 µL de solução III (Acetato de potássio 3 M; Ácido acético 2 M) e refrigerado a 4° C por 10 minutos. Centrifugou-se por 6 minutos à 10.000 g. O

sobrenadante foi transferido para um novo tubo, acrescentando-se 0,7 V de isopropanol e sendo invertido a cada 3 minutos por 20 minutos. Centrifugou-se por 15 minutos à 10.000 g. O sobrenadante foi descartado e o sedimentado foi lavado com 300 µL de etanol 70% gelado. Centrifugou-se por 3 minutos à 10000 g . Após, o sedimentado foi deixado secando e depois ressuspensão em 30 µL de água ultrapura com RNase (2 µL/mL).

Os DNAs obtidos foram usados para fazer uma análise de restrição com a enzima de restrição *Pst* I (Fermentas). Os insertos foram digeridos com a enzima *Pst* I (Fermentas) pois já haviam sido clonados no vetor, nos sítios de *Pst* I. Assim, ao usar a mesma enzima, tem-se a liberação do inserto do vetor e, se o mesmo tiver sítios internos de *Pst* I, tem se a fragmentação nos sítios internos.

Para a digestão foram usados: 0,5 µL da enzima *Pst* I (Fermentas), 1 µL do tampão 10X O (Fermentas), 3 µL do DNA plasmidial e 5,5 µL de água ultrapura e incubada por 2 horas à 37°C (Tabela 4). Um gel de agarose 0,8% foi usado para visualizar a digestão.

Tabela 4 - Digestão com *Pst* I dos clones da biblioteca de rúmen.

Solução	Quantidade	Marca
Enzima <i>Pst</i> I	0,5 µL	Fermentas
Tampão 10X O	2 µL	Fermentas
DNA plasmidial	3 µL	-
Água ultrapura	14,5 µL	-
Total	20 µL	-

Digestão incubada por 2 horas à 37°C.

Para os clones em duplicata, foi feita uma segunda digestão usando 0,5 µL da enzima *BamH* I (Fermentas), 2 µL do tampão 10X BamH I (Fermentas), 3 µL do DNA plasmidial e 14,5 µL de água ultrapura e incubada por 2 horas à 37°C (Tabela 5). Um gel de agarose 0,8% foi usado para visualizar a digestão.

Tabela 5 - Digestão com *BamH* I dos clones com padrão de digestão similar

Solução	Quantidade	Marca
Enzima <i>BamH</i> I	0,5 µL	Fermentas
Tampão 10X <i>BamH</i> I	2 µL	Fermentas
DNA plasmidial	3 µL	-
Água ultrapura	14,5 µL	-
Total	20 µL	-

Digestão incubada por 2 horas à 37°C.

3.4.3 Transformação dos insertos e reconfirmação dos fenótipos

A transformação dos clones, que apresentaram padrões distintos de digestão com as duas enzimas de restrição, foi feita pela técnica de eletroporação. Utilizou-se uma adaptação do protocolo (SAMBROOK AND RUSSEL; 2001) em que se emprega 2 µL de DNA em 100 µL de *E. coli* EPI 300 competentes. A eletroporação foi realizada em cubeta com os seguintes parâmetros elétricos: 2,5 kV, 25 µF e 200 Ω. A seguir, as células foram ressuspensas em 1 mL de meio LB e incubadas à 37°C por 50 minutos. Foram, ainda, plaqueadas em placas com LB ágar e tetraciclina (20 µg/mL) e deixadas na estufa 37°C por 16 horas. Todos os ensaios, para confirmar o fenótipo dos insertos para degradação de celulose, foram feitos em duplicidade para assegurar os resultados.

3.4.4 Extração de DNA plasmidial para sequenciamento dos insertos

A extração do DNA plasmidial de 2 clones (A1-7 e H11-35) empregou, primeiramente, o mesmo protocolo de extração adaptado (SAMBROOK AND RUSSEL; 2001) usado nas extrações anteriores e, depois, a purificação foi feita com o kit de purificação da QIAGEN.

Na segunda estratégia, o DNA plasmidial dos outros 9 clones (G8-10; E10-15; C2-16; G8-127; B8-57; G10-36; B3-24; E2-23 e A7-93) foi extraído usando o kit de maxiprep da QIAGEN.

O protocolo adaptado e utilizado na extração do DNA dos clones A1-7 e H11-35 consistiu em 100 mL do inóculo de cada um dos clones com

centrifugação a 8.000 g por 8 minutos. O sedimentado foi então lavado em 10 mL de STE (Tris-Cl pH 8,0 10 mM; NaCl 0,1 mM; EDTA pH 8,0 1 mM) resfriado a 4° C e centrifugado novamente nas mesmas condições anteriores. O sedimentado foi ressuspenso em 5 mL de solução I (Tris -HCl pH 8 25 mM; EDTA 20 mM; Glicose 50 mM), adicionou-se 5 mL de solução II (NaOH 0,2 M; SDS 1%) e guardado por 10 minutos no gelo. Foi feita nova adição de 7,5 mL da solução III (Acetato de potássio 3 M; Ácido acético 2 M) e guardado no gelo por 20 minutos. Centrifugou-se por 30 minutos à 10.000 g. Depois o sobrenadante foi transferido para um novo tubo, acrescentou-se 0,6 V de isopropanol e o tubo foi invertido a cada 3 minutos por 10 minutos. Centrifugou-se por 10 minutos à 10.000 g. O sobrenadante foi descartado e o sedimentado foi lavado com 5 mL de etanol 70% a 4° C. Centrifugou-se por 10 minutos à 10.000 g. Após a secagem do sedimentado, foi ressuspenso em 300 µl de água ultrapura com RNase (2 µL/mL). A purificação foi feita com o kit de purificação GeneJET da Fermentas.

3.4.5 Sequenciamento e montagem dos primers

O sequenciamento dos clones foi do tipo Sanger usando a metodologia primer walking. Nos anexos se encontra uma tabela com todos os primers que foram construídos para cada clone.

3.4.6 Análises das sequências dos clones da biblioteca de rúmen de caprinos

As sequências foram analisadas nos programas ORFinder para a detecção de ORFs (Open Reading Frames). As ORFs detectadas, com tamanho mínimo de 300pb, foram analisadas e um blastp delas foi feito para uma comparação com sequências já presentes nos bancos de sequências do NCBI e Pfam.

3.5 GENE *ENDO 03*

O gene *endo 03*, utilizado ao longo desse trabalho, foi selecionado por meio de uma triagem em busca de celulases. O isolamento, o sequenciamento do clone e a identificação do gene de interesse foram realizados pela aluna Shelly Paluan, que utilizou a biblioteca de solo da Amazônia e confirmou a codificação de uma beta-manosidase, possivelmente, proveniente de uma bactéria do filo Acidobactérias (PALUAN, 2011).

3.5.1 Amplificação do gene *endo 03* para expressão em *E. coli*

A amplificação do gene *endo 03* foi realizada por reação em cadeia da polimerase (PCR) usando os seguintes primers: Endo03petHind_R_2 (5' AAG CTT TAA TGC GTC GGC CAC GAG GAC 3') e Endo03petNde_F (5' CAT ATG TTT CGC GAG TGG CGC G 3'). Estes primers possuem sítios para as enzimas de restrição *Hind* III e *Nde* I, respectivamente (sublinhadas nas sequências), para facilitar a clonagem no vetor de expressão em *E. coli* pET21a.

A reação de PCR foi realizada com os seguintes parâmetros: 1ª fase a 98°C por 3 minutos, seguidos de 30 ciclos de 95°C por 30 segundos, 60°C por 30 segundos e 72°C por 4 minutos, e a finalização a 72°C por 15 minutos. Foi utilizado 1 µL da enzima Hotstart 5 U/µL (QIAGEN), 10 µL do tampão 5 X HotStart (QIAGEN), 5 µL do primer R (10 ng/µL), 5 µL do primer F (10 ng/µL), 2 µL de DNA (10 ng/ µL) e 27 µL de água ultrapura. Um gel de agarose 0,8% foi feito para visualizar o resultado da PCR com o auxílio do marcador 1Kb plus (Invitrogen). Para a purificação da banda amplificada por PCR foi feito um gel de agarose 1% e usado o kit de eluição da QIAGEN.

3.5.2 Ligação do gene *endo 03* no vetor de clonagem pJET

Para a clonagem do *endo 03* foi utilizado o vetor pJET (Thermo Scientific) e a enzima T4 DNA ligase que vem junto com o vetor. A ligação foi deixada a 4°C por 16 horas e depois 2 µL foram usados para transformar 100 µl de *E.coli* EPI300 competentes, por eletroporação. As células transformadas foram plaqueadas em placas de petri com LB ágar mais ampicilina (200 µg/mL), como antibiótico

3.5.3 Amplificação do gene *endo 03* para expressão em *P. pastoris*

O gene *endo 03* foi amplificado na reação em cadeia da polimerase (PCR) usando os seguintes primers: Endo03ppicXba_R (5' CAG TCT AGA CGT AAT GCG TCG GCC AC.3') e Endo03ppicEcoR_F (5' GTC GAA TTC TTT CGC GAG TGG CGC GTG.3'). Estes primers possuem sítios para as enzimas de restrição *Xba* I e *EcoR* I, respectivamente (sublinhadas nas sequências), para facilitar a clonagem no vetor de expressão em *P. pastoris* pPICZαA.

A reação de PCR foi realizada com os seguintes parâmetros: 1ª fase a 98°C por 3 minutos, seguidos, 30 ciclos de 95°C por 30 segundos, 57°C por 30 segundos e 72°C por 4 minutos, e a finalização a 72°C por 15 minutos. Foi utilizado 1 µL da enzima Hotstart 5 U/µL (QIAGEN), 10 µL do tampão 5 X HotStart (QIAGEN), 5 µL do primer R (10 ng/µL), 5 µL do primer F (10 ng/µL), 2 µL de DNA (10 ng/µL) e 27 µL de água ultrapura. Um gel de agarose 0,8% foi feito para visualizar o resultado da PCR com o auxílio do marcador 1kb plus (Invitrogen). Para a purificação da banda amplificada por PCR foi feito um gel de agarose 1% e usado o kit Purelink da Invitrogen.

3.5.4 Ligação do *endo 03* no vetor pPICZαA de *P. pastoris*

Para a clonagem do *endo 03* foi utilizado 1 µL do vetor pPICZαA (60 ng/µL) (Invitrogen), 14 µL de DNA *endo 03* (6 ng/µL), 2 µL de tampão 10X T4 ligase (Thermo Scientific), 1 µL da ligase T4 (Thermo Scientific) e 2 µL de água ultrapura (Tabela 6). A ligação foi deixada a 22°C por 1 hora e depois 5 µL foram usados na transformação de 100 µL de *E.coli* EPI300 competentes, por eletroporação. As células transformadas foram plaqueadas em placas de LB low salt (peptona 1%, NaCl 0,5%, extrato de levedura 0,5% e ágar 1,5%) ágar mais zeocina (50 µg/mL), como antibiótico, e deixadas na estufa 37°C por 16 horas.

Tabela 6 - Ligação *endo 03* no vetor pPICZαA

Solução	Quantidade	Marca
Enzima T4 DNA ligase	1 µL	Thermo Scientific
Tampão 10X T4	2 µL	Thermo Scientific
Vetor pPICZαA (60 ng/µL)	1 µL	Invitrogen
DNA <i>endo 03</i> (6 ng/µL)	14 µL	-

Água ultrapura	2 µL	-
Total	20 µL	-
Ligação incubada por 1 hora à 22°C.		

3.5.5 Transformação de *P. pastoris* e escolha dos clones transformados

Foi feita a transformação por eletroporação de *P. pastoris* da cepa X-33 competentes, usando o vetor ENDO3+pPICZαA e seguindo o protocolo abaixo descrito que foi adaptado do manual Invitrogen (Cat. no. V195-20): Foram colocadas no gelo as células competentes, o DNA e as curvetas de eletroporação. Depois foi pipetado 5 µL do DNA linearizado em 80 µL células e deixado por 5 minutos. A alíquota das células foi passada, com o DNA, para a curveta de eletroporação com o eletroporador ajustado para 1.500 volts, 25µF, 400Ω. As células foram eletroporadas, 1 ml de sorbitol 1 M foi adicionado e as células eletroporadas foram passadas para um frasco de 1.5 mL e deixadas a 30°C por 2 hora. Por último, as células transformadas foram plaqueadas em meio YPDS (1% extrato de levedura; 2% peptona; 2% glicose; sorbitol 1 M; 2% ágar) ágar com zeocina 100 µg/mL e incubadas na estufa 30°C por 3 dias.

Após o crescimento, foram escolhidas 49 colônias isoladas e 1 colônia controle para plaqueamento em placas de petri com meio BMMH (100mM fosfato de potássio; 1,34% YNB; 0,0004% biotina; 1% de metanol; 2% ágar) +CMC LV ágar. As placas foram colocadas na estufa a 30 °C por 1 dia, retiradas, incubadas por uma semana, coradas com Congo Red (0,1%) e o resultado intensificado com HCl (1 M).

3.5.6 Indução de proteínas em *P. pastoris*

A expressão das proteínas, dos clones escolhidos, foi induzida por metanol. Um pré-inóculo de 50 mL de meio BMGY (1% extrato de levedura; 2% peptona; 100 mM fosfato de potássio; 1,34% YNB; 0,0004% biotina; 1% de glicerol) com zeocina dos clones 6, 27 e 50 foram feitos e colocados no shaker a 30°C sob agitação de 180 rpm por 48 horas. Esses pré-inóculos foram centrifugados a 3.000 g por 5 minutos e o sedimentado foi então ressuscitado em 120 mL de meio BMMY (1% extrato de levedura; 2% peptona; 100mM fosfato de potássio; 1,34% YNB; 0,0004% biotina; 1% de metanol). Os frascos

com os inóculos foram, então, colocados no shaker a 30 °C e 180 rpm. Alíquotas de 1 mL foram retiradas dos clones no tempo 0 hora e armazenadas a - 80 °C. Foram acrescentados 600 µL de metanol em cada inóculo de 24 em 24 horas, e retiradas alíquotas de 1 ml dos clones nos tempos 24, 48, 72, 96 e 120 horas. Essas alíquotas foram centrifugadas com o sobrenadante e o sedimentado armazenados, separadamente, a - 80°C. Um gel de poliacrilamida 12% foi feito para a visualização das proteínas. O teste de atividade para CMC 4% com DNS foi feito com o sobrenadante e o sedimentado para verificar a atividade da endoglucanases seguindo o protocolo adaptado (XIAO; STORMS; TSANG, 2005): acréscimo de 50 µL de CMC LV 4%, 5 µl de tampão citrato pH 5.0 e 45 µl do sedimentado lisado com o kit masterpure (Epicenter). A reação foi incubada por 30 minutos a 50°C no termociclador, depois foi adicionado 120 µL de DNS, incubada novamente por 5 minutos a 95°C e, por fim, foi feita a leitura de absorbância em espectrofotômetro a 540 nm.

3.5.7 Ligação do gene no vetor de clonagem pJET de *E. coli*

Para a clonagem do gene, foi utilizado o vetor pJET (Thermo Scientific) e a enzima T4 DNA ligase que acompanha o vetor. A ligação foi deixada a 4°C por 16 horas e depois 2 µL foram usados para transformar 100 µL de *E.coli* EPI300 competentes, por eletroporação. As células transformadas foram plaqueadas em placas de petri com LB ágar mais ampicilina (100 µg/mL), como antibiótico.

3.5.8 Ligação do gene e do *endo 03* no vetor pET21a de expressão em *E. coli*

Para a clonagem do gene *endo 03* no vetor de expressão pET21a foi utilizado 1 µL do vetor pET21a (60 ng/µL), 14 µL de DNA *endo 03* (6 ng/µL), 2 µL de tampão 10X T4 ligase (Thermo Scientific), 1 µL da ligase T4 (Thermo Scientific) e 2 µl de água ultrapura (Tabela 7). A ligação foi deixada a 4°C por 16 horas e depois 2 µL foram usados para transformar 100 µL de *E.coli* BL21 competentes, por eletroporação. As células transformadas foram plaqueadas em placas de LB ágar mais ampicilina (100 µg/mL), como antibiótico, e deixadas na estufa 37°C por 16 horas.

Tabela 7 - Ligação *endo 03* no vetor pET21a

Solução	Quantidade	Marca
Enzima T4 DNA ligase	1 µL	Thermo Scientific
Tampão 10X T4	2 µL	Thermo Scientific
Vetor pET21a (60 ng/µL)	1 µL	-
DNA <i>endo 03</i> (6 ng/µL)	14 µL	-
Água ultrapura	2 µL	-
Total	20 µL	-

Ligação incubada por 16 horas à 4°C.

3.6 GENES *CBH I* E *CBH II*

Os genes *cbh I* e *cbh II*, utilizados ao longo desse trabalho, foram selecionados por meio de uma triagem em busca de celulases. O isolamento, o sequenciamento do clone e a identificação do gene de interesse foram realizados pelo aluno Lucas Carvalho, que utilizou a biblioteca de rúmen de caprinos e confirmou a codificação de celobiohidrolases.

3.6.1 Amplificação dos genes *cbh I* e *cbh II* para expressão em *E. coli*

A amplificação dos genes *cbh I* e *cbh II* foram realizadas por reação em cadeia da polimerase (PCR) usando os seguintes primers: CBH1petNde_F (3' CAT ATG AGT AAA AAA GTC ATT ACT TTG GC 5') e CBH1petXho_R (3' CTC GAG CGG GCT CAT AGA CAC CCC ACT CGC CGC AG 5') para o gene *cbh I* e CBH2petNde_F (3' CAT ATG AGA AAG AAT GTT TTT AGA TTA GC 5') e CBH2petXho_R (3' CTC GAG TTT CTT CTC CAG AGG CTT GCC 5') para o gene *cbh II*. Estes primers possuem sítios para as enzimas de restrição *Nde I* e *Xho I*, respectivamente (sublinhadas nas sequências), para facilitar a clonagem no vetor de *E. coli*.

A reação de PCR foi realizada com os seguintes parâmetros: 1ª fase a 98°C por 5 minutos, seguidos de 35 ciclos de 94°C por 15 segundos, 57°C por 1 minuto e 72°C por 2 minutos, e a finalização a 72°C por 10 minutos. Foi utilizado 1 µL da enzima Hotstart 5 U/µL (QIAGEN), 10 µL do tampão 5 X HotStart (QIAGEN), 5 µL do primer R (10 ng/µL), 5 µL do primer F (10 ng/µL), 2 µL de DNA (10 ng/ µL) e 27 µL de água ultrapura. Um gel de agarose 0,8% foi feito para

visualizar o resultado da PCR com o auxílio do marcador 1kb plus (Invitrogen). Para a purificação da banda amplificada por PCR foi feito um gel de agarose 1% e usado o kit de eluição da QIAGEN.

3.6.2 Ligação dos genes *cbh I* e *cbh II* no vetor de clonagem pJET

Para a clonagem dos genes, foi utilizado o vetor pJET (Thermo Scientific) e a enzima T4 DNA ligase que acompanha o vetor. A ligação foi deixada a 4°C por 16 horas e depois 2 µL foram usados para transformar 100 µL de *E.coli* EPI300 competentes, por eletroporação. As células transformadas foram plaqueadas em placas de petri com LB ágar mais ampicilina (100 µg/mL), como antibiótico.

3.6.3 Ligação dos genes *cbh I* e *cbh II* no vetor pET21a de expressão em *E. coli*

Para a clonagem dos genes *cbh I* e *cbh II* no vetor de expressão pET21A pET21a foi utilizado 1 µL do vetor pET21a (60 ng/µL), 4 µL de DNA *cbh I* (18 ng/µL) e 4 µL de DNA *cbh II* (20 ng/µL), 2 µL de tampão 10X T4 ligase (Thermo Scientific), 1 µL da ligase T4 (Thermo Scientific) e 2 µL de água ultrapura (Tabela 8). As ligações foram deixadas a 4°C por 16 horas e depois 2 µL foram usados para transformar 100 µL de *E.coli* BL21 competentes, por eletroporação. As células transformadas foram plaqueadas em placas de LB ágar mais ampicilina (100 µg/mL), como antibiótico, e deixadas na estufa 37°C por 16 horas.

Tabela 8 - Ligação *cbh I* e *cbh II* no vetor pET21a

Solução	Quantidade	Marca
Enzima T4 DNA ligase	1 µl	Thermo Scientific
Tampão 10X T4	2 µl	Thermo Scientific
Vetor pET21a (60 ng/µL)	1 µl	-
DNA <i>cbh II</i> (20 ng/µL)	4 µl	-
DNA <i>cbh I</i> (18 ng/µL)	4 µl	-
Água ultrapura	2 µl	-
Total	20 µl	-

Ligação incubada por 16 horas à 4°C.

3.6.4 Teste de atividade das proteínas com pNP-Celobiosídeo em *E.coli*

A expressão das proteínas foi induzida por IPTG. Um pré-inóculo com os clones das proteínas CBH I e de CBH II de 5 mL de meio LB, mais ampicilina (100 µg/mL), foi feito e ficou incubado a 37°C por 16 horas. Depois 20 µL de cada um dos pré-inóculos foi inoculado em 100 ml de meio TB (1,2% triptona; 2,4% extrato de levedura; 0,4% glicerol; solução KH_2PO_4 0,17 M K_2HPO_4 0,72 M) com ampicilina (100 µg/mL) até atingir uma densidade celular entre 1,4 - 1,6 lida em um espectrofotômetro no comprimento de onda de 600 nm. IPTG (10 mM) foi acrescentado para induzir a expressão e inóculos foram então incubados por 20 horas a 28°C. Depois as células foram centrifugadas a 5.000 g por 10 minutos a 4°C, ressuspendidas e lavadas com STE (Tris-Cl pH 8,0 10 mM; NaCl 0,1 mM; EDTA pH 8,0 1 mM) duas vezes. 1 g de sedimentado foi ressuspendido em 1 volume de tampão MES 100 mM pH 6,0 e as células foram, então, sonicadas no ultrassonicador a uma amplitude de 40%, tempo de pulso de 40 segundos, tempo de espera de 20 segundos e um tempo total de 10 minutos. As células foram centrifugadas por 30 minutos a 3.250 g a 4°C, e a fração líquida foi passada para um tubo novo e chamada de lisado.

O teste de atividade do lisado foi feito usando pNP-C. Todos os ensaios foram realizados em triplicada, conforme descrito a seguir: 10 µL de pNP-C (50 mM), 10 µL de enzima (lisada) e 80 µL de tampão (20 mM HEPES, 50 mM de NaCl) em placas de PCR. As reações foram realizadas em termocicladores por um período de 15 minutos. As reações foram interrompidas adicionando-se 100 µL de Na_2CO_3 (2%) e lidas em espectrofotômetro no comprimento de onda de 405 nm. Os valores de pH variaram de 4,0 até 8,5 e a temperatura variou de 30°C até 85°C. A enzima teve que ser diluída em 1:32.

3.6.5 Purificação da CBH II produzida em *E.coli*

As proteínas recombinantes produzidas têm uma cauda de histidina (His-tag). A purificação foi feita em colunas GE His-trap HP 1 mL. Primeiro foi adicionado 1 volume de binding buffer (Tris-HCl pH 7,4 20 mM; NaCl 0,5 M; Imidazole 20 mM) para 1 volume de amostra. A coluna foi preparada e equilibrada com 10 volumes da quantidade de resina. A amostra passou pela coluna e, do material passado, foi coletada uma fração. A seguir a coluna foi lavada com 10 volumes de binding buffer e outra fração, do material lavado, foi

coletada. Para se eluir a proteína foram usados com 5 mL de Elution Buffer (Tris-HCl pH 7,4 20 mM; NaCl 0,5 M; Imidazole 50 mM) e coletada a fração do eluído 50; eluído novamente com 5 mL de Elution Buffer (Tris-HCl pH 7,4 20 mM; NaCl 0,5 M; Imidazole 100 mM) e a fração do eluído 100 foi coletada. A última eluição foi feita com 5 mL de Elution Buffer (Tris-HCl pH 7,4 20 mM; NaCl 0,5 M; Imidazole 500 mM) e coletada a fração do eluído 500. As frações foram armazenadas a 4°C e um gel de SDS-PAGE 12% foi feito para visualizar a proteína eluída.

3.6.6 Efeito da temperatura e pH na atividade

Para a caracterização bioquímica foram determinados os valores de pH ótimo e temperatura ótima da proteína lisada e parcialmente purificada com ensaios de atividade com o substrato pNP-Celobiosídeo (pNP-C).

Um teste de atividade usando pNP-C foi feito para ver a atividade do lisado e da proteína parcialmente purificada (DESHPANDE; ERIKSSON; PETTERSSON, 1984) Todos os ensaios foram realizados em triplicada, conforme descrito a seguir: 20 µL de pNP-C (25 mM), 10 µL de enzima (lisada ou semi-purificada) e 80 µL de tampão (20 mM HEPES, 50 mM de NaCl) em placas de PCR. As reações foram realizadas em termocicladores por um período de 15 minutos. As reações foram interrompidas adicionando-se 100 µL de Na₂CO₃ (2%) e lidas em espectrofotômetro no comprimento de onda de 405 nm. Os valores de pH variaram de 4,0 até 8,5 e a temperatura variou de 35°C até 80°C. A fração lisada foi diluída em 1:32 e a fração semi-purificada, em 1:128.

4 RESULTADOS

4.1 CLONES DA BIBLIOTECA DE RÚMEN DE CAPRINOS

4.1.2 Escolha dos clones e confirmação do fenótipo

Foram escolhidos 12 clones da biblioteca de rúmen de caprinos para serem usados ao longo do trabalho.

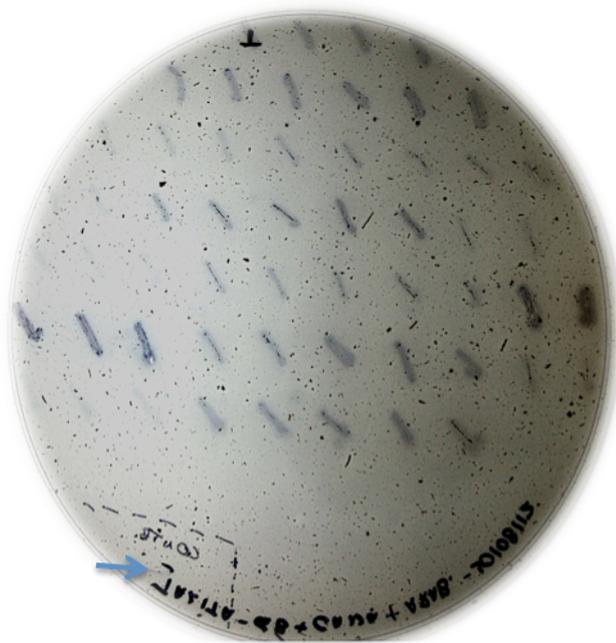


Figura 6: Placa contendo meio LB suplementado com bagaço de cana 1% e arabinose, corada com Congo Red e HCl para visualização dos fenótipos de degradação de celulose. A seta indica o controle negativo.

Na tabela 10 estão os resultados obtidos quando as placas foram coradas com Congo Red e o resultado foi intensificado com HCl para visualização dos fenótipos de degradação de celulose. A Figura 6 mostra um exemplo de uma placa contendo meio LB ágar com bagaço de cana e o indutor do promotor arabinose. A placa foi corada com Congo red e o resultado intensificado com HCl para o fenótipo de degradação de celulose. A seta indica o controle negativo. Esse ensaio foi feito para se confirmar os resultados obtidos na primeira triagem feita na biblioteca metagenômica de rúmen de caprinos.

Tabela 9 - Resultados da primeira triagem. Na tabela 10 estão todos os clones que foram triados e os meios de cultura em que cada clone foi triado.

	Bagaço-de-Cana**	CMC HV	CMC LV	Xilana
A1-7	+	+	++	+
G2-8	++	+	++	++
G8-10	++	++	++	+
E10-15	++	++	++	+
C2-10	+	+	+	+
H11-35	+	+	+	+
G8-127	+	++	+	++
B8-57	+	+	++	+
G10-36	+	+	+	+
B3-24	+	+	+	+
E7-23	++	+	++	++
A7-93	++	+	++	++

*As cruzes, na tabela, representam a intensidade da coloração formada por cada clone. Quanto mais forte mais cruzes ele ganha.** Os substrato em que foram testados cada um dos clones

Na tabela 11 estão os resultados do segundo ensaio feito para a confirmação dos fenótipos de celulose. Um segundo ensaio foi necessário pois precisávamos ter uma certeza maior de que os clones escolhidos apresentariam um fenótipo de degradação de celulose.

Tabela 10 - Resultados da segunda triagem. Na tabela 11 estão todos os clones que foram triados e os meios de cultura em que cada clone foi triado.

	Bagaço-de- Avicel	CMC HV	CMC LV	Xilana
A1-7	+	++	++	+
G2-8	+	++	+	+

G8-10	+	+	++	++	+
E10-15	+	+	++	+	++
C2-10	+	+	++	+	+
G8-127	++	+	++	++	++
B8-57	+	+	+	+	++
B8-57	+	+	+	+	++
G10-36	++	+	+	+	+
B3-24	+++	++	+++	+++	+++
E7-23	++	+	++	++	++
A7-93	+	+	++	++	+

*As cruces, na tabela, representam a intensidade da coloração formada por cada clone. Quanto mais forte mais cruces ele ganha.** Os substrato em que foram testados cada um dos clones

4.1.3 Extração de DNA plasmidial e análise dos insertos

Após a segunda análise dos fenótipos dos clones foi extraído o DNA plasmidial, seguido de uma digestão dos mesmos com a enzima de restrição *Pst* I.

Com a digestão foi possível fazer uma comparação do padrão de bandas liberadas por cada um dos clones. A primeira seleção permitiu saber se existia algum dos insertos que estava em duplicata e, dessa maneira, já abandoná-lo. Para se confirmar a duplicata de clones foi feita também uma digestão com a enzima *Bam*H I (Fermentas) pois, se com duas enzimas diferentes o padrão de bandas continuasse o mesmo, a probabilidade do clone estar em duplicata era grande e seria abandonado.

O padrão eletroforético da digestão dos clones com *Pst* I revelou que os clones A1-7 e G8-10 exibiam o mesmo padrão de bandas. Os clones E10-15; G8-127 e E7-23 também se pareciam e que o clone B8-57 foi selecionado duas vezes. Digestão com a enzima de restrição *Bam*H I foi então feita para confirmar a existência de clones em duplicata (Figura 6). Ao final do experimento foi possível concluir que os clones A1-7 e G8-10 são provavelmente o mesmo clone; e que os clones E10-15, G8-127 e E7-23 possuem padrões diferentes e portanto são clones diferentes e que o clone B8-57 estava em duplicata.

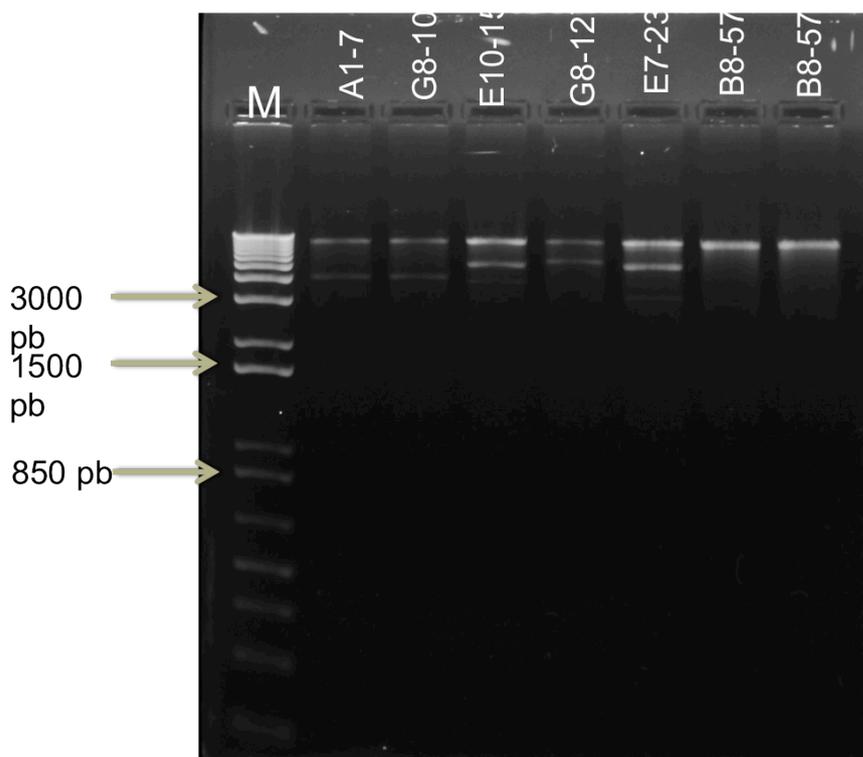


Figura 7: Gel de agarose 0,8 % com a digestão dos clones A1-7, G8-10; E10-15, G8-127, E7-23 e B8-57, com a enzima *BamH* I I para poder verificar os padrões de digestão e eliminar os clones com padrões similares. M: marcador 1kb plus. clones

4.1.4 Transformação dos clones e reconfirmação dos fenótipos

Uma nova transformação foi feita com os clones diferentes e posteriormente foi feita uma nova confirmação dos fenótipos de degradação de celulose pois precisava se ter certeza de que o fenótipo vinha mesmo dos insertos e não da bactéria. A tabela 12 mostra os resultados do ensaio. Pode-se observar que nenhum clone perdeu a sua atividade, demonstrando a estabilidade do fenótipo celulolítico destes clones.

Nessa segunda revelação dos fenótipos de degradação de celulose com Congo red HCl foi feita um da placa de LB ágar contendo bagaço de cana pré-tratado acrescido de arabinose, usado como indutor do promotor AraC presente no vetor, para ver se existia uma diferença de resultados e um possível aumento do potencial de degradação dos clones.

Tabela 11 - Resultado da terceira triagem. Na tabela 12 estão todos os clones que foram triados e os meios de cultura em que cada clone foi triado.

	Bagaço- de-Cana+ Arab.**	Bagaço- de-Cana	Avicel	CMC H	CMC L	Xilana
1-A1-7	+++	++	++	+++	+++	++
2- G2-8	++	+	+	+	++	+
3- E10-15	+	+	+	+	++	+
4- C2-10	+	+	+	+	++	+
5- G8-127	++	++	++	++	++	++
6- B8-57	+	+	+	+	+	+
7- G10-36	++	+	+	+	+	+
8- B3-24	+++	+++	+++	+++	+++	+++
9- E7-23	++	+++	++	+++	+++	+++
10- A7-93	+	+	+	+	+	+
11- H11-	++	+++	++	++	+++	++

*As cruces, na tabela, representam a intensidade da coloração formada por cada clone. Quanto mais forte mais cruces ele ganha.** Os substrato em que foram testados cada um dos clones

4.1.5 Extração de DNA plasmidial para sequenciamento dos insertos

Dos 11 clones, 9 foram extraídos o DNA com o kit maxiprep da QIAGEN e dois deles foi feita uma extração caseira que depois foi purificada usando o kit de purificação da Fermentas.

A figura 7 mostra o padrão de digestão de todos os clones utilizando a enzima *Pst* I.

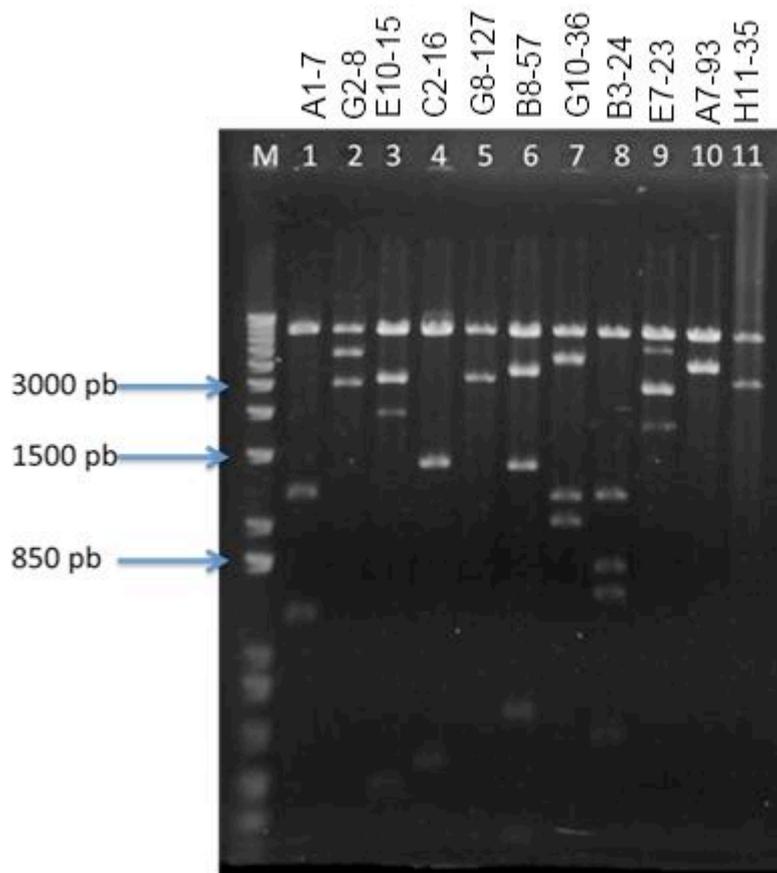


Figura 8: Gel de agarose 0,8 % com todos os clones que foram mandados para o sequenciamento. M: marcador 1kb plus. 1 a 11: clones digeridos com a enzima Pst I.

4.1.6 Análises das sequências dos clones da biblioteca de rúmen de caprinos

Foi descoberto ao longo do sequenciamento que o clone H11-35 estava em duplicata com o clone G8-127 por isso o sequenciamento do mesmo foi abandonado.

Os clones A1-7, C2-16, E10-15 e G8-127 foram completamente sequenciados. Abaixo se encontram as Figuras 8, 9, 10 e 11 com o alinhamento de todas as sequências de cada clone foi feito programa Geneious e as tabelas 12 e 13 com os principais alinhamentos encontrados no BLAST P para cada um dos clones e no Pfam. Nos anexos se encontram as tabelas com todas as ORFs e suas análises para esses clones no BLAST P.

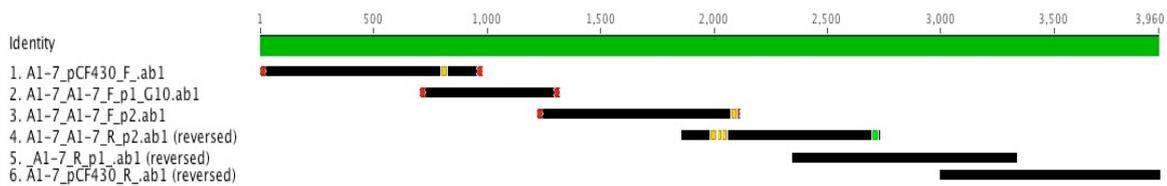


Figura 9: Alinhamento dos contigues do clone sequenciado A1-7 montados no programa Geneious.

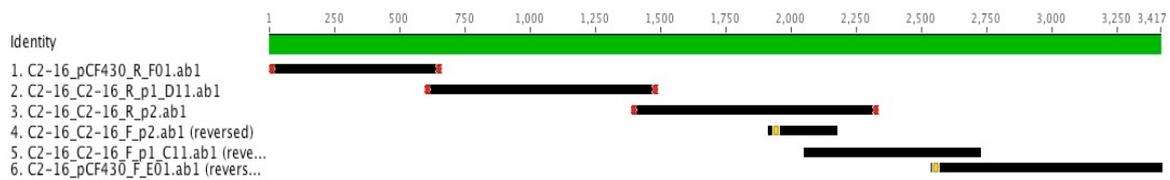


Figura 10: Alinhamento dos contigues do clone sequenciado C2-16 montados no programa Geneious.



Figura 11: Alinhamento dos contigues do clone sequenciado E10-15 montados no programa Geneious.

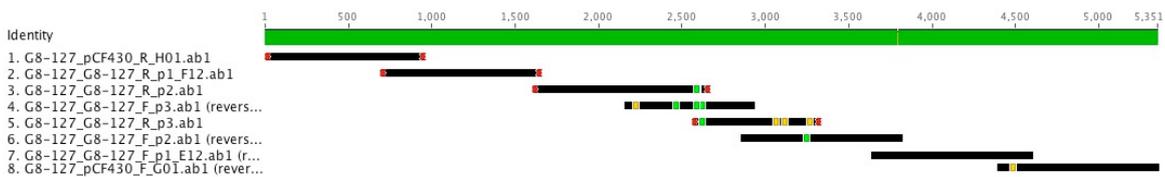


Figura 12: Alinhamento dos contigues do clone sequenciado G8-127 montados no programa Geneious.

Tabela 12 - Principais alinhamentos no BLAST P encontrados para cada um dos clones.

Clone	ORFs encontradas	ORFs de interesse	e-value	Cobertura	Alinhamento do BLAST P de interesse
A1-7	10	168aa	5.5	64%	glyoxylase-related zinc-dependent
		119aa	0.002	32%	hydrolase [Geobacter metallireducens GS-15] 31% *
		318aa	4.2	42%	cellulosome-anchoring protein, partial [Peptoniphilus lacrimalis] 36% PREDICTED: glucan endo-1,3-beta-glucosidase 8-like [Glycine max] 28%
C2-16	7	274aa	7.0	50%	glycosyl hydrolase [Dysgonomonas gadei] >gb EGK02235.1 hypothetical protein HMPREF9455_01505 [Dysgonomonas gadei ATCC BAA-286] 24%
E10-15	13	551aa	5e-82	62%	hypothetical protein HMPREF9332_01082 [Alloprevotella rava F0323] 43%
G8-127	9	821aa	0.0	98%	helicase C-terminal domain protein [Prevotella melaninogenica ATCC 25845] 74%

* valores de similaridade.

Tabela 13 - Principais alinhamentos no Pfam encontrados para cada um dos clones.

Clone	ORFS encontradas	ORFs Interessantes	Cobertura	e-value	Alinhamento no Pfam
A1-7	10	305aa	47.6	1.7e-12	Methyltransf 11 Methyltransferase domain Domain CL0063
		404aa	21.1	0.00028	Trm112p Trm112p-like protein Domain CL0167
			359.2	3.1e-107	Trm112p is required for tRNA methylation in <i>S. cerevisiae</i> and is found in complexes with 2 tRNA methylases (TRM9 and TRM11) also with putative methyltransferase YDR140W Peptidase M32 Carboxypeptidase Taq (M32) metallopeptidase Family CL0126
C2-16	7	236aa	68.4	2.5e-19	Zn Tnp IS1 InsA N-terminal domain Domain CL0167
		274aa	81.8	1.6e-2	HTH Tnp IS1 InsA C-terminal domain Domain CL0123
			32.6	8.3e-08	DDE Tnp IS1 IS1 transposase Family CL0219 256.2 5e-77 Laminin G 3 Concanavalin A-like lectin/glucanases superfamily Domain CL0004
E10-15	13	244aa	188.0	1.3e-55	PDDEXK 2 PD-(D/E)XK nuclease family transposase Domain CL0236
		177aa	61.4	6.7e-17	EPN HEPN domain Family CL0291
		109aa	91.9	4.6e-10	NTP transf 2 Nucleotidyltransferase domain Family CL0260 39.6
		551aa		2e-26	DUF285 Mycoplasma protein of unknown function, DUF285 Family

G8-127	9	821aa 118aa	47.2 26.6 68.4	1.6e-12 2.3e-06 4e-19	CL0022 PLDc 2 PLD-like domain Domain CL0479 SNF2 N SNF2 family N-terminal domain Family CL0023 Peptidase M15 3 Peptidase M15 Domain CL0170
--------	---	----------------	----------------------	-----------------------------	--

4.2 GENE ENDO 03

4.2.1 Amplificação do gene *endo 03* para expressão em *E. Coli*

Para a amplificação do gene *endo 03* e seu posterior uso no sistema de expressão de *E. coli* foram construídos primers e adicionados sítios para as enzimas de restrição *Hind* III e *Nde* I, facilitando assim a clonagem e posterior expressão do inserto no vetor pET21a.

Como na PCR foi amplificado não só o fragmento desejado de 1866 pb, o fragmento teve que ser eluído do gel de agarose (Figura 13).

4.2.2 Amplificação do gene *endo 03* para expressão em *P. pastoris*

Para a amplificação do gene *endo 03* e seu posterior uso no sistema de expressão de *P. pastoris* foram construídos primers e adicionados sítios de restrição para as enzimas *EcoR* I e *Xba* I, facilitando assim sua clonagem no vetor de expressão pPICZ α A.

Como na PCR não foi amplificado só o fragmento desejado como pode ser visto na figura 13, o mesmo foi eluído do gel de agarose para depois ser usado para na reação de ligação com o vetor de expressão em *P. pastoris* PICZ α A.

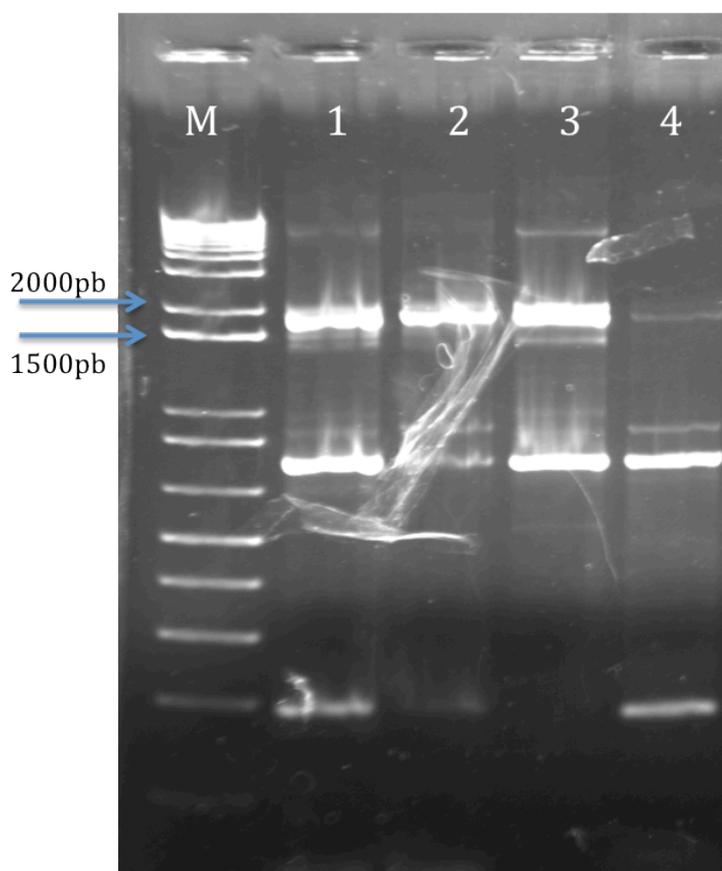


Figura 13: Gel de agarose 0,8% com o resultado das PCRs do gene *endo 03* com primers para clonagem em bactéria e para clonagem em levedura. 1 e 2: PCR do gene *endo 03* para levedura; 3 e 4: PCR do gene *endo 03* para bactéria; M: marcador 1kb plus.

4.2.3 Transformação de *P. pastoris* e escolha dos clones transformados

Para poder escolher quais clones de *P. pastoris* seriam usados na expressão da proteína, foi antes feita uma análise do fenótipo de degradação de celulose

usando Congo Red e HCl para intensificar os resultados. A partir desse resultado (Figura 14) foram escolhidos os clones 6 e 27 (setas vermelhas) para serem expressos, assim como um clone de *P. pastoris* transformado só com o vetor pPICZ α como controle negativo (seta azul).

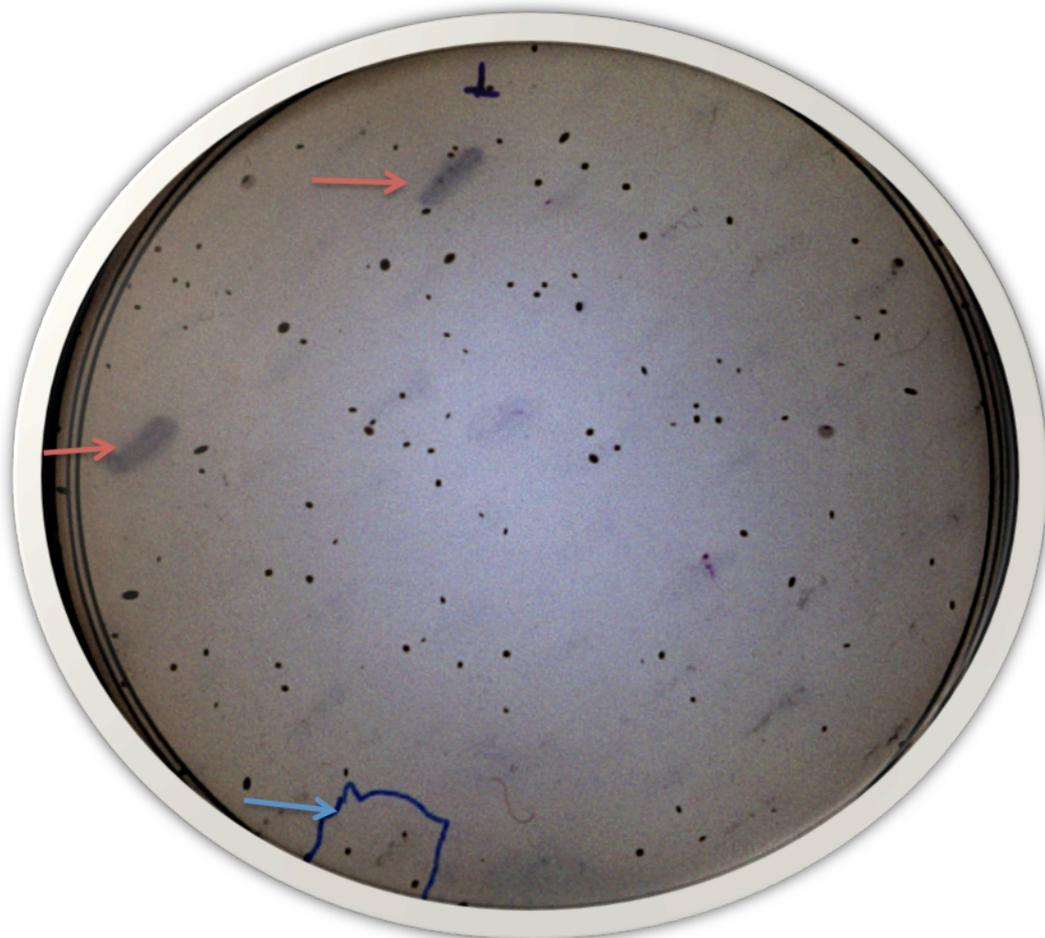


Figura 14: Clones de *P. pastoris* transformados e corados com Congo Red e HCl para visualização dos fenótipos de degradação de celulose. A seta azul indica o controle negativo e as setas vermelhas indicam os clones 6 e 27 que foram escolhidos para continuação do trabalho.

4.2.4 Expressão de proteínas em *P. pastoris*

Foi feita a expressão de proteínas em *P. pastoris* com a indução por metanol dos clones selecionados e do controle negativo e alíquotas foram retiradas nos tempos 0, 24, 48, 72, 96 e 120 horas do sedimentado e do sobrenadante separados.

O vetor pPICZ α A usado para clonar o gene *endo 03* possui uma sequência de peptídeo sinal e por isso um gel de poliacrilamida 12% foi feito

com o sobrenadante em todos os tempos. Mas nenhuma banda de indução foi detectada em nenhum dos clones.

Foi feito também um teste de atividade com CMC 4% e DNS para confirmar a atividade de endoglucanase dos clones. O teste foi feito inicialmente, com o sobrenadante dos clones. Como não foi detectada atividade o teste passou a ser feito também com o as células lisadas e então foi possível visualizar uma baixa atividade em um dos clones.

4.2.5 Ligação do gene e do *endo 03* no vetor pET21a de expressão em *E. coli*

Como não foi confirmado nenhum clone transformado com a ligação do vetor pET21 com o gene *endo 03* a expressão em *E.coli* teve que ser abandonada.

4.3 GENES CBH I E CBH II

4.3.1 Amplificação dos genes *cbh I* e *cbh II* para expressão em *E. coli*

Para a amplificação dos genes *cbh I* e *cbh II* e seu posterior uso no sistema de expressão de *E. coli* foram construídos primers e adicionados sítios de restrição para as enzimas *Nde I* e *Xho I*, facilitando assim sua clonagem no vetor de expressão pET21a.

Como na PCR não foi amplificado só os fragmentos desejados de 1035 pb da *cbh I* e de 1194 pb da *cbh II*, os mesmos tiveram que ser eluídos do gel de agarose para depois serem usados para na reação de ligação com o vetor de *E. coli* (Figura 15).

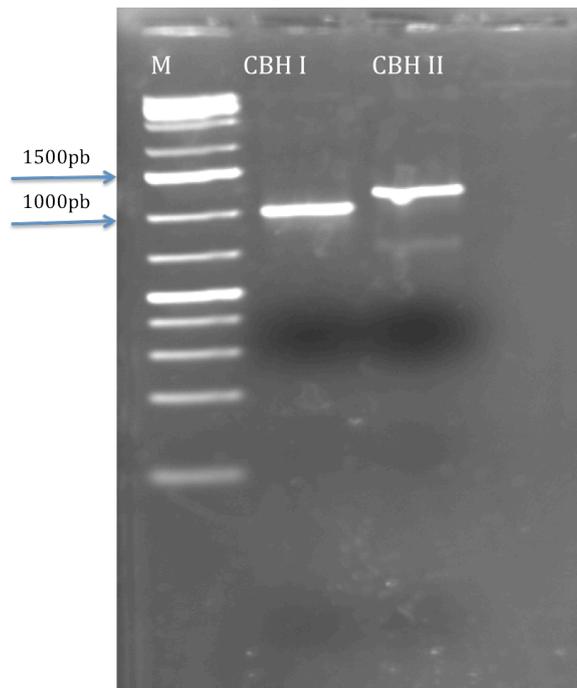


Figura 15: : Gel de agarose 0,8% com o resultado das PCRs dos genes *cbh I* e *cbh II* com primers para clonagem em bactéria. M: marcador 1kb plus

4.3.2 Ligação dos genes *cbh I* e *cbh II* no vetor pET21a de expressão em *E. coli*

A ligação dos genes *cbh I* e *cbh II* com o vetor de expressão pET21a foi realizada e uma digestão dos clones selecionados, com as enzimas de restrição *Nde I* e *Xho I* foi feita para confirmar as ligações (Figuras 16 e 17).

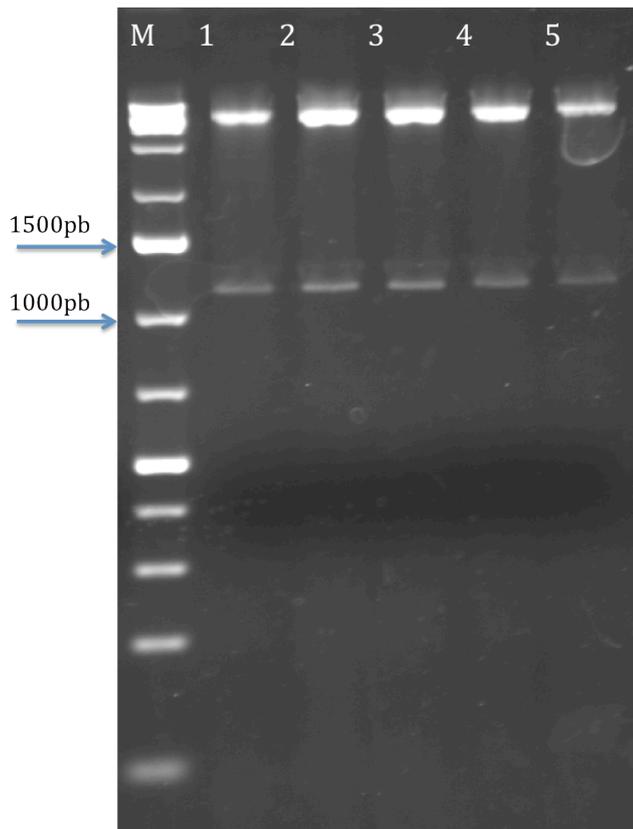


Figura 16: Digestão dos clones *cbh II* no vetor pET21a para a confirmação da clonagem. 1 a 5: clones de *cbh II* no vetor pET21a digeridos com as enzimas *Nde I* e *Xho I*. M: marcador 1kb plus.

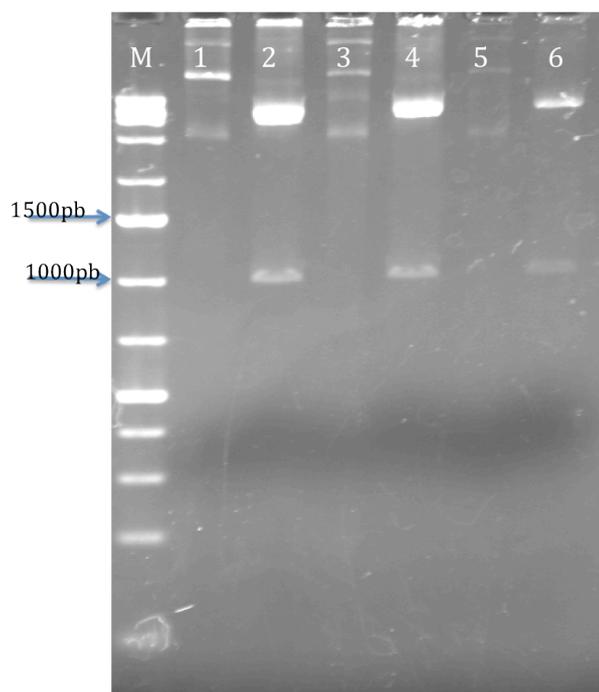


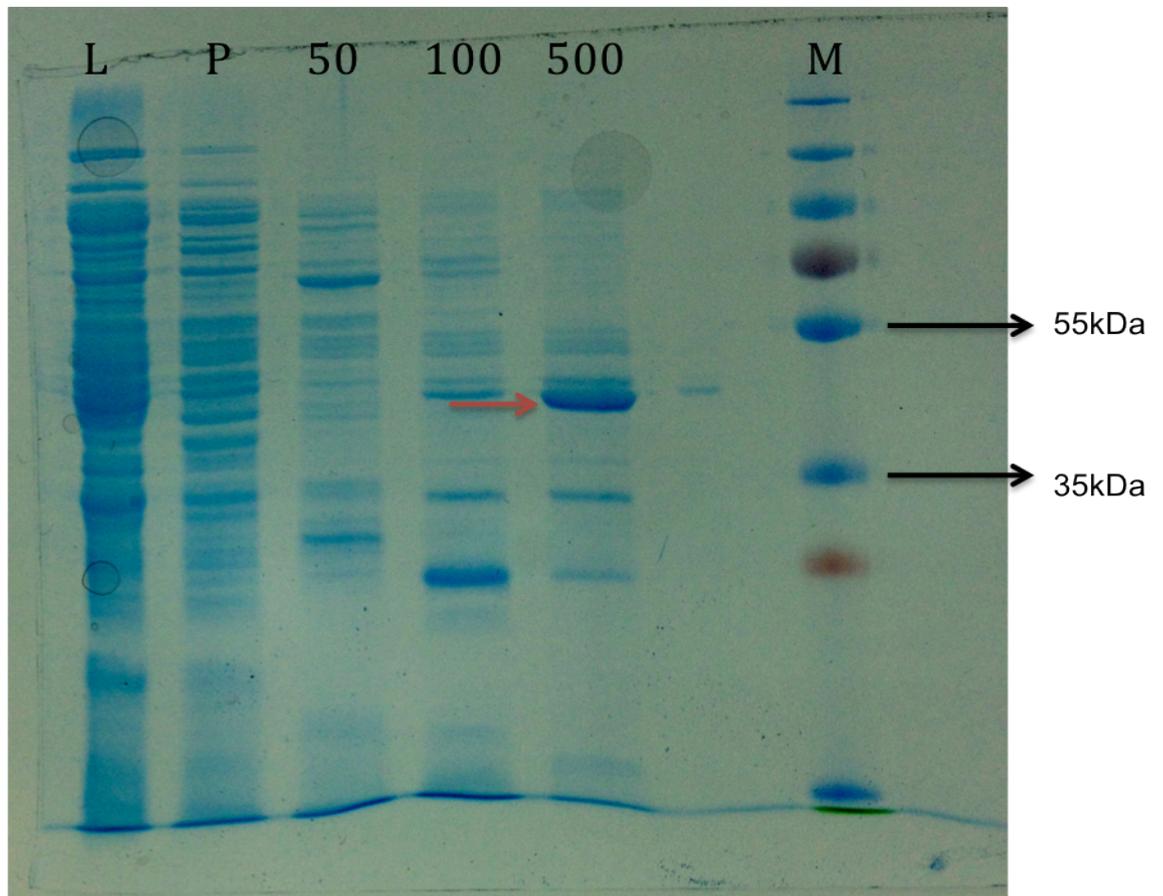
Figura 17: Digestão dos clones *cbh I* no vetor pET21a para a confirmação da clonagem. 1, 3 e 5: clones de *cbh I* no vetor pET21a intactos; 2, 4 e 6: clones de *cbh I* no vetor pET21a digeridos com as enzimas *Nde I* e *Xho I*. M: marcador 1kb plus.

4.3.3 Teste de atividade das proteínas com pNP-C em *E.coli*

Foi feito um teste com o substrato pNP-C para testar a atividade dos 3 clones de CBH I e dos 3 CBH II. Foi visto que todos os clones de CBH I e de CBH II tinham atividade, mas que o clone 3 de CBH II teve atividade melhor que os outros e por isso esse clone foi escolhido para se fazer à expressão da proteína e a tentativa de purificação da mesma.

4.3.4 Purificação de proteínas produzidas em *E.coli*

Foi feita a purificação parcial da proteína CBH II com o uso da coluna GE His-trap HP 1 mL e o resultado pode ser visualizado na figura 18 (gel de poliacrilamida 12%). Uma análise usando o programa ImageJ foi feita e constatou-se que a proteína CBH II está, aproximadamente, 53% pura.



Gel de poliacrilamida 12%. P: passado na coluna Histrap; L: lavado na coluna Histrap; 50: eluído com 50mM de imidazole; 100: eluído com 100mM de imidazole; 500: eluído com 500mM de imidazole; M: marcador PageRuler plus. Seta vermelha: CBH II 45kDa.

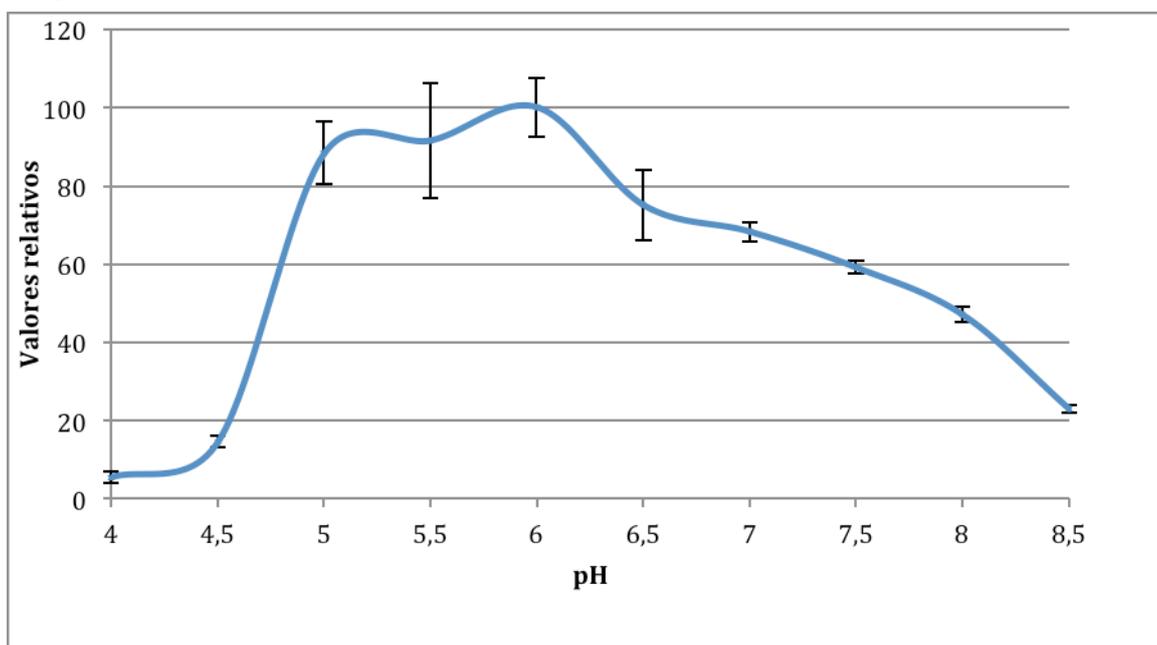
Figura 18: Gel com o resultado da purificação da CBH II

4.3.5 Caracterização de temperatura e pH ótimos da proteína

Foram feitos os ensaios para determinar a temperatura e o pH ótimos da proteína CBH II lisada e parcialmente purificada, usando o substrato pNP-C (Figuras 19 e 20).

Os valores de pH ótimos e de temperatura ótimos da proteína CBH II lisada não puderam ser mostrados pois os valores de desvio padrão ficaram muito altos e por isso os resultados não foram conclusivos.

Figura 19: Gráfico do pH ótimo da CBH II parcialmente purificada.



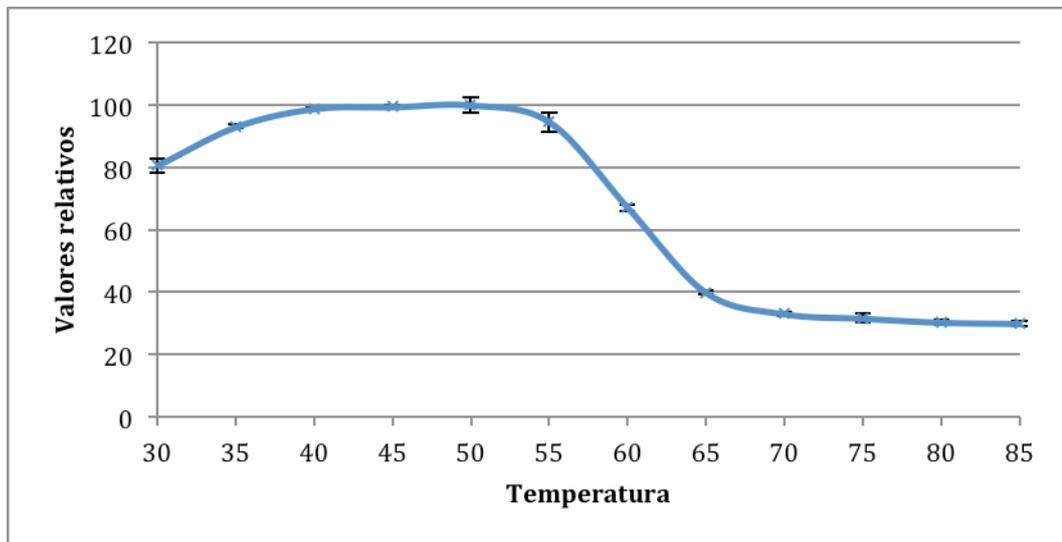


Figura 20: Gráfico da temperatura ótima da CBH II parcialmente purificada.

5 DISCUSSÃO

5.1 Resultado da triagem do clones

As tabelas 10 e 11 mostram que nenhum dos clones dos insertos, obtidos a partir da biblioteca metagenômica de rúmen de caprinos, perdeu o seu fenótipo de degradação. É possível, também, verificar que alguns clones tiveram uma diferença na intensidade da coloração. Uma possível explicação seria as diferentes condições em que os experimentos foram realizados. Tendo em vista que os experimentos não ocorreram simultaneamente e, sim, em dias diferentes, não foi possível reproduzir exatamente as mesmas condições.

A tabela 12 compara os resultados da placa de bagaço-de-cana suplementado com arabinose e os resultados da placa de bagaço-de-cana. Pode ser verificado que a maioria dos clones triados, da biblioteca de rúmen de caprino, possuem um promotor próprio não dependendo assim do promotor AraC, induzido por arabinose, encontrado no vetor pCF430. Isso pode ser interessante, do ponto de vista biotecnológico, e abre a possibilidade de novas pesquisas que façam uma comparação com os promotores existentes atualmente.

5.2 Resultado das análises do sequenciamento dos clones

A metagenômica busca aprimorar o conhecimento sobre genes de microrganismos que não podem ser cultivados em laboratórios (YEH et al., 2013). Desta forma, é de se esperar que muitas das ORFs dos clones metagenômicos, que foram totalmente sequenciados, resultaram em um e-value alto. Provavelmente, essas ORFs codificam proteínas que ainda não foram caracterizadas e assim ainda não estão disponíveis em bancos de dados. Mesmo assim, foi possível identificar algumas candidatas a serem responsáveis pelo fenótipo de degradação de celulose.

Após análises das ORFs nos bancos de dados GeneBank e Pfam, verificou-se que a ORF 274 aa do clone C2-16 apresenta resultados positivos em ambos. No Pfam, a ORF tem um domínio de glicosil hidrolase, já no GeneBank, há uma cobertura com uma proteína glicosil hidrolase. As

endoglucanases fazem parte do grupo das glicosil hidrolases e por isso essa ORF é uma forte candidata a ter sua sequência clonada e expressa separadamente (BHAT, 2000).

Uma outra candidata é a ORF 119 aa do clone A1-7 que, de acordo com a análise no GeneBank, apresenta uma proteína de celulosoma com um e-value de 0.002. Celulosomas são um complexo de enzimas celulolíticas que atuam em conjunto presentes em microrganismos anaeróbicos (BAYER et al., 1998).

5.3 Resultado da análise do sequenciamento de gene *endo 03*

No caso do gene *endo 03*, após o sequenciamento, foi confirmado que o clone possui um gene de β -manosidase (sequência pode ser vista nos anexos). B-manosidases são glicosil hidrolases que hidrolisam dímeros de ligação manosídicas β -1,4 resultantes da degradação de manana, glucomanana e galactomanana por exo-endo-manosidases. Embora esses substratos façam parte da hemicelulose (PALUAN, 2011), o gene foi triado no fenótipo para degradação de celulose e possui atividade de uma endoglucanase. É comum encontrar enzimas que possuem dupla função, como o possível caso do *endo 03*.

5.4 Resultado da amplificação dos genes

Um ponto crítico do procedimento de amplificação de genes, na etapa de expressão das proteínas celulolíticas, é o aparecimento de bandas que não correspondem ao fragmento de tamanho específico para o gene. Nas figuras 13 e 15 é possível ver que, não somente o *endo 03* (Figura 13) e os genes *cbh I* e *cbh II* (Figura 15) foram amplificados, mas existem outras bandas, bem fortes no caso do gene *endo 03*, que foram amplificadas também. Levando em conta alguns dos fatores que podem influenciar em uma reação de PCR (CARVALHO; SILVA, 2010), cuidados especiais foram tomados, em várias tentativas, para se otimizar essas amplificações e, assim, obter-se somente a banda correspondente ao gene de interesse. Foram modificadas a temperatura de anelamento, a quantidade de DNA molde, a enzima polimerase usada e a quantidade de $MgCl_2$. Melhorias parciais foram obtidas, embora não tenha sido possível eliminar por completo o aparecimento de outras bandas no gel resultante.

5.5 Resultado das clonagens com o gene *endo 03*

As maiores dificuldades, contudo, foram observadas nas clonagens feitas com o gene *endo 03*. Foram testados dois vetores comerciais diferentes (pGEM-T Easy e pJET) que deveriam facilitar o processo, mas os resultados obtidos não foram satisfatórios. O vetor de clonagem pGEM-T Easy (Promega) foi testado primeiramente, porém a clonagem não foi confirmada. Utilizou-se, então, o vetor pJET para a clonagem do gene no vetor de expressão em *E. coli* pET21a, porém também não foi possível. Essas dificuldades de se trabalhar com o gene *endo 03* talvez possam ser explicadas por se tratar de um gene novo. Embora o gene *endo 03* apresente grandes similaridades com um gene de β -manosidase de acidobactérias, não há relatos, na literatura científica, de clonagens bem sucedidas e, certamente, o gene apresenta suas particularidades.

5.6 Resultado da expressão da proteína ENDO 03

Peptídeos sinais são sequências N-terminais de 15 a 60 aminoácidos que encaminham as proteínas para o lugar em que atuam e, após esse encaminhamento, podem ser removidas (HEGDE; BERNSTEIN, 2006). O vetor de pPICZ α A possui, na sua estrutura, uma sequência de peptídeo sinal chama fator α . Essa sequência é responsável por secretar a proteína heteróloga produzida.

A proteína ENDO 03, após ser induzida, mostrou possuir atividade no seu sedimentado lisado, porém o esperado era que essa atividade aparecesse no sobrenadante. O gene *endo 03* foi clonado no vetor pPICZ α A, que possui uma sequência de peptídeo sinal e, portanto, deveria encaminhar a proteína produzida para ser secretada para fora das células.

Uma das possibilidades, para que isso não tenha ocorrido, é que a sequência fator α não esteja atuando, impedindo que a proteína seja secretada. A outra possibilidade é que as proteínas estejam sendo secretadas, porém, por uma questão de afinidade, não sejam muito solúveis no meio de cultura. Assim, na centrifugação, ao invés das proteínas se dissolverem no sobrenadante, estariam se depositando no fundo junto com a porção do sedimentado (KAY et al., 2008)

5.7 Resultados das análises dos genes *cbh I* e *cbh II*

Foi visto que os genes de *cbh I* e *cbh II*, possivelmente, são uma Celobiosidase e uma Celodextrinase, respectivamente. Os alinhamentos,

verificados com o programa Blast, mostraram identidades de 93% para a *cbh I* e de 91% para a *cbh II*. Tanto a celobiosidase quanto a celodextrinase foram descobertas em uma triagem de uma biblioteca de rúmen de bovinos (DUAN et al., 2009).

5.8 Resultados da expressão da proteína CBH II

O pH ótimo da CBH II parcialmente purificada ficou na faixa de 5,0 a 6,0. Não foi possível definir um valor com menor grau de incerteza para o pH ótimo pois o desvio padrão ficou muito alto. Mesmo repetindo o teste diversas vezes, não foi possível chegar a valores mais baixos no desvio padrão. Possivelmente, isso se deve, em parte, pelo fato de terem sido utilizados 10 tampões diferentes aumentando muito o número de variáveis envolvidas que poderiam afetar a leitura. Em testes de pH ótimo, como os que foram realizados, basta uma pequena diferença para gerar dados muito diferentes entre as triplicatas e, assim, aumentar o desvio padrão.

Em termos comparativos, a temperatura ótima para a celodextrinase encontrada em rúmen de bovinos é de 45°C e o pH ótimo de 6.0. A temperatura ótima da CBH II parcialmente purificada objeto desta pesquisa foi de 50°C e o pH ótimo ficou na faixa de 5.0 a 6.0 (DUAN et al., 2009). Esse é um resultado promissor para seu possível aproveitamento industrial. Os processos de despolimerização, na indústria, acontecem entre uma faixa de pH de 4.5 – 5.0 e temperaturas de 40 a 50°C (HORN et al 2012). A CBH II parcialmente purificada apresentou temperatura e uma faixa de pH ótimos que se encaixam nessas especificações, mostrando assim que ela pode ser uma candidata para ser usada no processo de despolimerização em nível industrial.

A simples leitura da temperatura ótima (50°C), obtida para a CBH II parcialmente purificada, não traduz corretamente seu potencial de uso. Como pode ser observado no gráfico da figura 20, a proteína é bastante estável em uma faixa bem ampla de temperatura o que pode ser uma vantagem. A indústria, geralmente, utiliza um coquetel enzimático composto por endoglucanases, celobiohidrolases e β -glicosidases (SINGHANIA et al., 2010). Ter uma enzima com atividade em uma ampla faixa de temperatura pode ser uma vantagem. Isto garantiria a atividade do coquetel enzimático mesmo quando outras enzimas não estivessem mais ativas devido à variação de temperatura.

Ainda em relação ao aproveitamento das proteínas objeto desta pesquisa (CBH II e ENDO 03) em um coquetel enzimático, há outro fator positivo nesse sentido: a complementaridade. No grupo de pesquisa do Laboratório de Ciências Genômicas e Biotecnologia da Universidade Pontifícia Católica de Brasília, a pesquisadora Jessica Bergmann fez a purificação e a caracterização de uma β -glicosidase (BERGMANN, 2013). Juntando-se a esse trabalho, a purificação e a caracterização de uma celobiohidrolase (CBH II) e de uma endoglucanase (ENDO 03), os resultados das duas pesquisas disponibilizam as enzimas básicas necessárias para a degradação da celulose em glicose, representando um passo importante na fabricação de bioetanol de 2ª geração. Isso reforça mais uma vez a importância dos estudos desenvolvidos com as proteínas citadas.

6 CONCLUSÕES

A pesquisa realizou significativos avanços na busca de celulases que possam ser aproveitadas comercialmente na degradação de celulose para a fabricação de biocombustíveis de 2ª geração. Trabalhando em três frentes distintas, a pesquisa realizou a triagem de clones em bibliotecas metagenômicas; e clonou, expressou duas enzimas celulolíticas: uma endoglucanase (ENDO 03) e uma celubiohidrolase (CBH II). Em relação à enzima CBH II, foi possível avançar até a etapa de purificação (parcialmente) e caracterização, com resultados promissores para seu aproveitamento industrial.

Nas pesquisas com a biblioteca metagenômica, foram triados dez clones da biblioteca de rúmen de caprinos para degradação de celulose e seus fenótipos foram confirmados em 4 substratos diferentes, sendo eles: CMC LV, CMC HV, Avicel e Bagaço de cana pré- tratado.

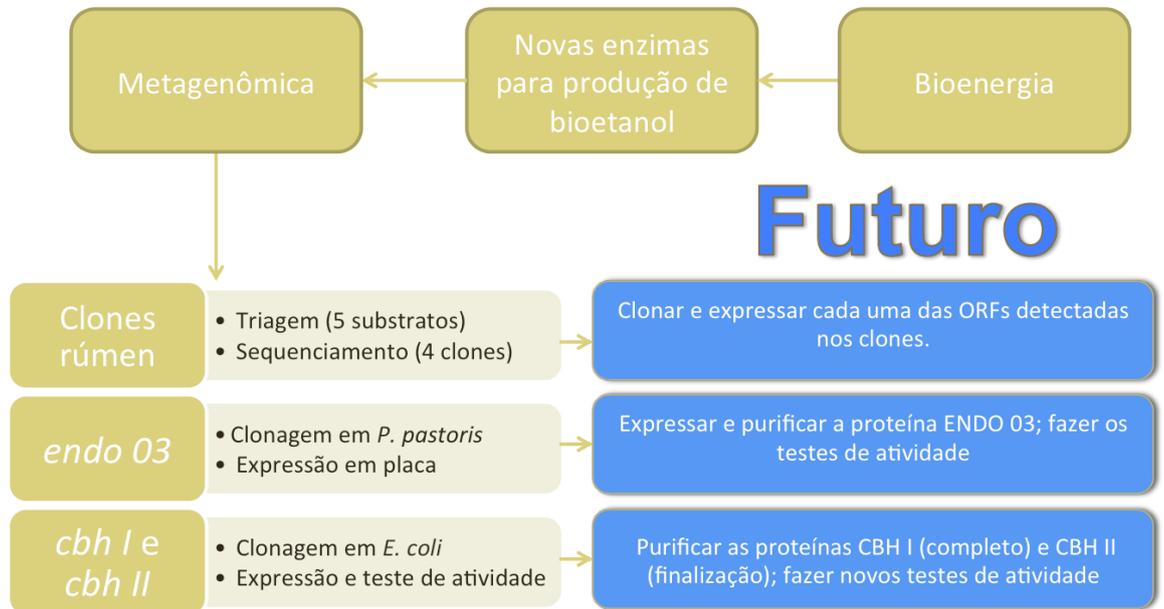
Desses 10 clones, 4 deles foram totalmente sequenciados e suas possíveis ORFs analisadas nos principais bancos de dados de sequenciamento de material genético (NCBI e Pfam), com a identificação de duas fortes ORFs (119 aa do clone A1-7 e 274 aa do clone C2-16) candidatas à clonagem e expressão.

No desenvolvimento de enzimas celulolíticas comercialmente viáveis, foram clonados, com sucesso, os genes *cbh II* e *cbh I* no vetor de expressão, em *E. coli*, pET21a e do gene *endo 03* no vetor de expressão, em *P. pastoris*, pPICZαA. A proteína CBH II foi expressa e parcialmente purificada, e testes de pH e temperatura ótimos foram feitos com a proteína parcialmente purificada indicando resultados promissores para seu aproveitamento comercial em um coquetel enzimático.

As pesquisas comprovam a utilidade da técnica metagenômica na identificação de genes potencialmente interessantes na obtenção de novas enzimas celulolíticas, bem como a capacidade dos grupos de pesquisa dos laboratórios envolvidos (Laboratório de Ciências Genômicas e Biotecnologia da Universidade Católica de Brasília, Laboratório de Genética e Biotecnologia da Embrapa Agroenergia e o Laboratório de Enzimologia da Universidade de

Brasília) em utilizar as técnicas biotecnológicas para a produção e o estudo destas enzimas visando a obtenção do biocombustível de 2ª geração.

Figura 21: Quadro resumo do trabalho de pesquisa realizado e a indicação das próximas etapas na busca de celulases para a fabricação de biocombustíveis de 2ª geração



7 PERSPECTIVAS

Assim como o trabalho de pesquisa se deu em três frentes distintas, uma na identificação de possíveis genes de interesse para a produção de enzimas celulolíticas, e duas outras na produção e caracterização de enzimas com potencial comercial, os avanços devem contemplar, no mínimo, esses mesmos objetivos. Espera-se explorar o grande potencial das bibliotecas metagenômicas que foram construídas (solo da Amazônia e rúmen de caprinos) para se identificar novas enzimas responsáveis por produzir o fenótipo de degradação de celulose. A partir das triagens feitas, as próximas pesquisas deverão clonar e expressar cada uma das ORFs detectadas nos clones sequenciados (119 aa do clone A1-7 e 274 aa do clone C2-16). Deverão, ainda, dar prosseguimento ao trabalho de sequenciamento e análise dos clones na possível identificação de outras ORFs candidatas.

No tocante à produção de enzimas e sua caracterização, há diversos passos a serem feitos para compor um coquetel enzimático comercialmente viável composto por uma β -glicosidade, uma celobiohidrolase e uma endoglucanase. A exemplo do trabalho feito com a proteína BGL 17 (β -glicosidade), espera-se conseguir purificar a proteína CBH II (celobiohidrolase) de maneira total e com isso poder fazer mais testes para verificar o Kcat, Vmáx e a estabilidade da CBH II.

Espera-se, também, expressar e purificar a proteína ENDO 03 (celobiohidrolase), em *P. pastoris*, e fazer os testes para verificar a temperatura e o pH da proteína, assim como o seu Kcat, Vmáx e a estabilidade da mesma.

Espera-se expressar e purificar a proteína CBH I (celobiohidrolase), em *E. coli*, e fazer os testes para verificar a temperatura e o pH da proteína, assim como o seu Kcat, Vmáx e a estabilidade da mesma.

Espera-se expressar e purificar as proteínas CBH II e CBH I em *P. pastoris*, fazendo o uso de primers específicos que já foram construídos e fazer os testes para verificar a temperatura e o pH da proteína, assim como o seu Kcat, Vmáx e a estabilidade da mesma para depois fazer uma comparação da expressão dessas proteínas em *E. coli* e em *P. pastoris*.

8 REFERÊNCIAS BIBLIOGRÁFICAS

AGGER, J.; VIKSØ-NIELSEN, A.; MEYER, A. S. Enzymatic xylose release from pretreated corn bran arabinoxylan: differential effects of deacetylation and deferuloylation on insoluble and soluble substrate fractions. **Journal of agricultural and food chemistry**, v. 58, n. 10, p. 6141–6148, 26 maio 2010.

BAYER, E. A. et al. Cellulose, cellulases and cellulosomes. **Current opinion in structural biology**, v. 8, n. 5, p. 548–557, out. 1998.

BERGMANN, J. C. **Utilização de estratégias metagenômicas para aplicações biotecnológicas no setor de biocombustíveis**. Universidade Católica de Brasília. 2013.

BHAT, A. et al. Isolation of cold-active, acidic endocellulase from Ladakh soil by functional metagenomics. **Extremophiles**, v. 17, n. 2, p. 229–239, 1 mar. 2013.

BHAT, M. K. Cellulases and related enzymes in biotechnology. **Biotechnology advances**, v. 18, n. 5, p. 355–383, ago. 2000.

CARVALHO, M. C. DA C. G. DE; SILVA, D. C. G. DA. Next generation DNA sequencing and its applications in plant genomics. **Ciência Rural**, v. 40, n. 3, p. 735–744, mar. 2010.

CGEE. **Bioetanol de Cana-de-açúcar. Energia para o desenvolvimento sustentável**, 2008.

CHANG, L. et al. Characterization of a bifunctional xylanase/endoglucanase from yak rumen microorganisms. **Applied Microbiology and Biotechnology**, v. 90, n. 6, p. 1933–1942, 2011.

CHEN, F.; DIXON, R. A. Lignin modification improves fermentable sugar yields for biofuel production. **Nature biotechnology**, v. 25, n. 7, p. 759–761, jul. 2007.

CREGG, J. M. et al. Recombinant protein expression in *Pichia pastoris*. **Molecular biotechnology**, v. 16, n. 1, p. 23–52, set. 2000.

CRUEGER, A.; and CRUEGER, W. Glucose transforming enzymes. p. 177-226. In FOGARTY, W. M. and KELLY, C. T. (ed.), **Microbial enzymes and biotechnology**, 2nd ed. Elsevier Science, London, United Kingdom. 1990.

CUNHA, I. S. et al. Bacteria and Archaea community structure in the rumen microbiome of goats (*Capra hircus*) from the semiarid region of Brazil. **Anaerobe**, v. 17, n. 3, p. 118–124, jun. 2011.

DESHPANDE, M. V.; ERIKSSON, K. E.; PETTERSSON, L. G. An assay for selective determination of exo-1,4,-beta-glucanases in a mixture of cellulolytic enzymes. **Analytical biochemistry**, v. 138, n. 2, p. 481–487, 1 maio 1984.

DUAN, C.-J. et al. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. **Journal of Applied Microbiology**, v. 107, n. 1, p. 245–256, 2009.

FENG, Y. et al. Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. **Applied microbiology and biotechnology**, v. 75, n. 2, p. 319–328, maio. 2007.

FERRER, M. et al. Functional metagenomics unveils a multifunctional glycosyl hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen. **PLoS ONE**, v. 7, n. 6, 2012.

GENG, A. et al. Expression and characterization of a novel metagenome-derived cellulase Exo2b and its application to improve cellulase activity in *Trichoderma reesei*. **Applied microbiology and biotechnology**, v. 96, n. 4, p. 951–962, nov. 2012.

GONG, X. et al. Cloning and identification of novel hydrolase genes from a dairy cow rumen metagenomic library and characterization of a cellulase gene. **BMC research notes**, v. 5, p. 566, 2012.

GUO, Y. et al. Cloning and heterologous expression of glucose oxidase gene from *Aspergillus niger* Z-25 in *Pichia pastoris*. **Applied**

biochemistry and biotechnology, v. 162, n. 2, p. 498–509, set. 2010.

HANDELSMAN, J. et al. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. **Chemistry & biology**, v. 5, n. 10, p. R245–249, out. 1998.

HARTLEY, J. L. Cloning technologies for protein expression and purification. **Current opinion in biotechnology**, v. 17, n. 4, p. 359–366, ago. 2006.

HEGDE, R. S.; BERNSTEIN, H. D. The surprising complexity of signal sequences. **Trends in biochemical sciences**, v. 31, n. 10, p. 563–571, out. 2006.

HENRISSAT, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. **The Biochemical journal**, v. 280 (Pt 2), p. 309–316, 1 dez. 1991.

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

JIANG, C. et al. Characterization of a novel β -glucosidase-like activity from a soil metagenome. **The Journal of Microbiology**, v. 47, n. 5, p. 542–548, 1 out. 2009.

JIANG, C. et al. Identification of a metagenome-derived β -glucosidase from bioreactor contents. **Journal of Molecular Catalysis B: Enzymatic**, v. 63, n. 1–2, p. 11–16, abr. 2010.

KIM, S.-J. et al. Characterization of a gene encoding cellulase from uncultured soil bacteria. **FEMS Microbiology Letters**, v. 282, n. 1, p. 44–51, 2008.

KUMAR, R.; SINGH, S.; SINGH, O. V. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. **Journal of industrial microbiology & biotechnology**, v. 35, n. 5, p. 377–391, maio. 2008.

KWON, E. J. et al. Construction of a metagenomic library from compost and screening of cellulase- and xylanase-positive clones. **Journal of Applied Biological Chemistry**, v. 53, n. 6, p. 702–708, 2010.

KAY, R. et al. Enrichment of low molecular weight serum proteins using acetonitrile precipitation for mass spectrometry based proteomic analysis. **Rapid communications in mass spectrometry: RCM**, v. 22, n. 20, p. 3255–3260, out. 2008.

LI, G. et al. Molecular cloning and characterization of a novel β -glucosidase with high hydrolyzing ability for soybean isoflavone glycosides and glucose-tolerance from soil metagenomic library. **Bioresource technology**, v. 123, p. 15–22, nov. 2012.

LIU, J. et al. Cloning and functional characterization of a novel endo- β -1,4- glucanase gene from a soil-derived metagenomic library. **Applied Microbiology and Biotechnology**, v. 89, n. 4, p. 1083–1092, 2011.

LYND, L. R. et al. Microbial cellulose utilization: fundamentals and biotechnology. **Microbiology and molecular biology reviews: MMBR**, v. 66, n. 3, p. 506–577, table of contents, set. 2002.

MACHADO, C. M. M. et al. Microrganismos na Produção de biocombustíveis líquidos. **Editora Técnica. 1a edição.** 2013.

MBA MEDIE, F. et al. Genome analyses highlight the different biological roles of cellulases. **Nature reviews. Microbiology**, v. 10, n. 3, p. 227–234, mar. 2012.

MICHAEL, R.GREEN and JOSEPH SAMBROOK. Molecular Cloning: A Laboratory Manual. 4th Edition. **Cold Spring Harbor Laboratory Press. 2012.**

MIR, B. A. et al. Recombinant hyperthermophilic enzyme expression in plants: a novel approach for lignocellulose digestion. **Trends in biotechnology**, v. 32, n. 5, p. 281–289, maio 2014

NACKE, H. et al. Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. **Biotechnology Letters**, v. 34, n. 4, p. 663–675, abr. 2012.

NEWMAN, J. R. et al. "Broad - host - range expression vectors that carry the L-arabinose-inducible *Escherichia coli* araBAD promoter and the araC regulator." **Gene**, 227: 197-203, 1999.

NGUYEN, N. H. et al. Identification and characterization of a cellulase-encoding gene from the buffalo rumen metagenomic library. **Bioscience, biotechnology, and biochemistry**, v. 76, n. 6, p. 1075–1084, 2012.

NIMCHUA, T. et al. Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes. **Journal of microbiology and biotechnology**, v. 22, n. 4, p. 462–469, abr. 2012.

Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB)

PACHECO, T.F.. Produção de Etanol: primeira ou segunda geração?. Acessado em 20/04/2013.

PAULAN, S.F. **Construção de Biblioteca Metagenômica de DNA Microbiano do Solo Amazônico e Prospecção de Enzimas de Interesse Biotecnológico**. Universidade Católica de Brasília. 2011.

PHITSUWAN, P. et al. Present and potential applications of cellulases in agriculture, biotechnology, and bioenergy. **Folia microbiologica**, v. 58, n. 2, p. 163–176, mar. 2013.

QIN, H. et al. The expression of neuroglobin in astrocytoma. **Brain tumor pathology**, v. 29, n. 1, p. 10–16, jan. 2012.

RASHAMUSE, K. J. et al. Characterisation of two bifunctional cellulase-xylanase enzymes isolated from a bovine rumen

metagenome library. **Current microbiology**, v. 66, n. 2, p. 145–151, fev. 2013.

SINGHANIA, R. R. et al. Advancement and comparative profiles in the production technologies using solid-state and submerged fermentation for microbial cellulases. **Enzyme and Microbial Technology**, v. 46, n. 7, p. 541–549, 7 jun. 2010.

VOGET, S. et al. Prospecting for Novel Biocatalysts in a Soil Metagenome. **Applied and Environmental Microbiology**, v. 69, n. 10, p. 6235–6242, out. 2003.

WANG, W. et al. The porcine gut microbial metagenomic library for mining novel cellulases established from growing pigs fed cellulose-supplemented high-fat diets. **Journal of animal science**, v. 90 Suppl 4, p. 400–402, dez. 2012.

WOOLF, D. et al. Biofuels from pyrolysis in perspective: trade-offs between energy yields and soil-carbon additions. **Environmental science & technology**, 30 abr. 2014.

XING, M.-N.; ZHANG, X.-Z.; HUANG, H. Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. **Biotechnology advances**, v. 30, n. 4, p. 920–929, ago. 2012.

XIAO, Z.; STORMS, R.; TSANG, A. Microplate-based carboxymethylcellulose assay for endoglucanase activity. **Analytical Biochemistry**, v. 342, n. 1, p. 176–178, 1 jul. 2005.

YEH, Y.-F. et al. A metagenomic approach for the identification and cloning of an endoglucanase from rice straw compost. **Gene**, v. 519, n. 2, p. 360–366, 1 maio 2013.

9 ANEXOS

Sequência do gene ENDO 03:

ATGTTTCGCGAGTGGCGCGTGGATGTGAAGAATGAGCTCAAACCCGGCGCGAACCAAC
TTCGCGTGGTCTTCCCTTACCCATCAAAGCGGCTTCGGAAATCGCCGCCAACGATCCCTG
GCGTGAGCGCACACATGTCGAGGCCAAGACCTACATCCGCAAAGCCGCTTACGAATACGGA
TGGGACTGGGGTCCGCGCTTCGTCACCAGCGGCATCTGGCGTCCCGCCCGCCTCGAAACC
TGGAACGACGCGCGCATTAGCAACATCCAGATCCGTCAGCTCGACGTGAACAGCGCCGTCG
CGCATCTGCTTGCAGAAGTCGAAGTCACATCCTCCGACGGAGTAATGCGAATGTCTCGGT
CGCGTACGGCGTCACCGGCAAACCCGCCGACGCATCCCGTAGTATTCAACTCCAGCCCGG
CTTAAACCACGTCACCATACCCATCGATATCGCAGGTCCCTCCTTGTGGTATCCCGCCGGAT
ACGGCTCCCAACCGATCTACAAATTCAGGTGCAACTCAAATCGGCAAGCTCATCGAAGAC
GAAGCCGTTGCCAAAACCGGCCCTTCGATCGGTGGTTCTGCGTCGCGAGCCCGATCCGTGG
GGACGCTCCTTCGAGTTCGTGGTCAACGGCATCCCGGTCTTCGGCAAAGGCGCCGACGTG
ATCCCGTTCGATAGCTTTCCCGCACGCGTTACCCTCGAGCAATACCGCCAGGTCCTCGAATC
CGTGGTCGACGCCAACATGAACATGATTCGCCACTGGGGTGGCGGCTATTACGAAACCGAC
GAGTTCTACAACATGTGCGACGAGCTCGGCATCATGATCTGGCAGGACTTCATGTTCCGGCAA
CGACTGGCAGCCCGGCACCTACGAATGGAACTCAACGTCGCCGCCGAAGCCGAGCAGCA
GCTGAAACGCCTGCGCAACCATCCAGCATCGTAATCTGGTGTGGCAACAACGAAACTGAG
GTCGCACTCGCGTGGCAGCACCAGCGACAAGCTGCCCGCCGACGTCCGCTTCCAGATGTGG
CAGGACTATGTGTCGACTTTCAGCGGAATCCTCGCCGCGGCGGTCAGCCGCTATGCACCTG
AAACGCCTTATTGGCCAGCTCACCCAGTGGGACTATGAGACTGTCAACGATAAGTACCAG
TCCGGCGACATGCACGATTGGACCGTCTGGCACGGCCGTGTTCCCTTACCCGACTACGAGA
AGCACAACCCGCGCTTCATGACCGAGTACGGCTTCCAGTCGTTTCCGGAGATGCGCACAGT
CGACAGCTTACGATTCCCGAAGACCGTCTCAATATCTTACGCCCGTCATGCTGGCGCATC
AGAAGAACACCGCGGGCAACGCCATCATCCGCGACTATATGCTGCGCGACTATCCGCAACC
CAAAGACTTTGCTTCGTTCTTTACGCCAGCCAGGTCTTGCAAGCCGAAGGCATCAAATCG
GAGCCGAGCATCTGCGCCGCAATCGACCGCGCACCATGGGTTCCATCTACTGGCAGCTCAA
CGACTGCTGGCCTGTCGCCTCCTGGTCGAGCCTCGACTACTACGGACGCTGGAAGGCCCTT
CATTACTACGCCCGCCGCTTCTACGCGCCAATTCTCGTCTCGCCCCATCAGGAAGAAGGCA
ACGTGGCCGTCTACATCGTCTCCGACAAGACTTCGCCACAGTTGCTACGTTGCGCGTGGC
CATCCTTAGCTTCGACGGCAATCTGCTCAGCGACCAGTCGCAGCCGGTTCAAATCACGCCG
CTTCCAGCAAGGTATATCTCACGATTCCGATGCTCGACATCAGCAATCTGCCCAATTGCGA
TCTCGCCAAGATTGCCTCGTGGCCGACGCATT**TAA**

Sequência do gene de CBH II:

ATGAGAAAGAATGTTTTAGATTAGCCGTTGTGCGAACGGCACTGTTGTTTTCTCTGGGCTG
TATGGCGCAGTCAGCAAAAAGTGAACAACCTCCGTATCAAGCGCGGAACAAATGTGAGCCACT
GGCTGTCACAGTCGGAACAGCGTGGTGAGGCACGCCCTCCACATTCAGGAGGACGATTT
CAACCGTCTGGAGGATCTGGGCTTCGACTTCGTACGTATCCCCATCGACGAGGTGCAGTTC
TGGGACGAGAAGGGCAACAAGCTGCCTGAGGCTTGGAATCTGCTGAGCAACGCCCTTGACT
GGAGCCGCAAGCACAACTGCGTGCCATCGTGGATCTGCATATCATCCGCTCGCACTATTT
CAACGCTGTCAATGAAGAAGATCAGGCTGCAAACACCCTCTTCACCTCAGAGAAGGCACAG
GAGGACCTCATCAACCTCTGGCGCCAACCTCTCAGATTTCTGAAGGGTCGCAGTTGCGACT
GGGTGGCTTATGAGTTTATGAATGAGCCGGTGGCTGACGAGCATGAGCAGTGAACCAGCT
GGTAGCAAAGGTACACAAGGCGTTGCGCAAGTTGGAGCCTCAGCGTACACTCGTCATCGGT
TCAAACCGTTGGCAGGGTCATGAGACCATGAAGTATCTGAAGGTGCCCGAGGGCGACAAGA
ACATCGTGCTCTCTTTCCACTACTACAACCCGATGATTCTGACACACTATGGTGCCTGGTGG
TCGCCAATGTGCGCAGCCTACAAAGGCAAGGTTCACTATCCCGGCATATTGGTTTTCAAAGGA
GGACTATGATGCCGCACCCGACGCCATCAAGCCCGAGCTGAAGCCCTATACCGAGGAAGTA
TGGGACATCAACAAGATCCGTGAGCAGTTCAGGGATGCCATCGAGGCAGCGAAGAAATACG
ACCTACAGCTGTTCTGCGGTGAGTGGGGTGTGTATGAGCCCGTGGATCGTGAAGTGGCCTA
CAACTGGTATCGTGACGTGCTGACCGTATTCGACGAGTTCAACATTGCCTGGACCACTTGGT
GCTACGATGCCGACTTCGGCTTCTGGGACCAGCAGCGCCACAGTTTCAAGGACTATCCTCT
GGTAGAGCTCCTGATGTCAGGCAAGCCTCTGGAGAAGAAAT**TAA**

Sequência do gene de CBH I:

ATGAGTAAAAAAGTCATTACTTTGGCCGTCATCATTGCGGCACTGTTTGTAACCATGT
CATGCCAGAACAAGAGTCAGAGTCAGAGTCAGAATCAGGACCAGGCAGGACAAGTGNATAA
CTTCCGCATCAAGCGCGGCACTAATATCAGCCACTGGCTCTCGCAGTCGGAGCAGCGCGG
GGAAGCCCGTCGCCTCCATATCCAGGAGGACGACTTCGCCCGTCTGGAGGAACTGGGCTT
CGACTTCGTGCGCATCCCATCGACGAGGTACAGTTCTGGGACGAGCAGGGCAATAAACTG
CCCAGGGCGTGGGACCTGCTGAACAACGCCCTCGACTGGGCCCGCCAGCACAACTCCGT
GCCATCGTCGATCTCCACATCATCCGTTCCCACTACTTCAATGCCGTGAACGAGGCARACCA
GGCCGCCAACACCCTCTTCACCTCGGAGGAGGCACAGGAGGGACTCATCAACCTCTGGCG
CCAGTTGTCGGAATTCCTGAAGGACCACAGCAACGACTGGGTAGCCTACRAGTTCATGAAC
GAGCCCGTGGCCGAGGAGCACGAGCAGTGAACCAGTTGATTGCCAAGGTGCACAAGGCA
CTGCGCGAACTGGAGCCCCAGCGCACACTGGTCATCGGCTCGAATCTGTGGCAGGGTCAC
CAGACCATCAAGTACCTGAAGGTGCCCGAGGGCGACAAGAACATCATCTCTCCTTCCACTA
CTACAACCCGATGCTTGTGACCCATTACGGCGCCTGGTGGTCACCCCTGTGTGCAGCCTAC
AAGGGTAAGGTGAACTATCCCGGCGTGCTCGTGTGCAAGGAGGACTACAACGACGCCCC
GACGCCATCAAACCTGAACTGAAGCCCTATACCGAGGAGGTATGGAACATCGACAAGATCC
GTGAACAGTTCAAGGATGCCATCGAGGCCGCCAAGAGTACGACCTGCAACTGTTCTGCGGC
GAGTGGGGTGTCTATGAGCCCG**TGA**

Sequência do clone A1-7:

CTGCAGAACACCTTTAACTGTCATTCCGGACGTCAATTCCCAGATATTCCTTCATTTT
CGCGTTCCAGATCTGCGGCAGATCTTTGACTTCGATCTCCCGGTTGATCAGTCCCTTTTCGA
TTTCATATCTCAGCATAATGTGCAGGCAGTAGGTCAGCTCATCCGCTTCTGTGCGGATCAGG
GATGGCTGTGCGATATTGACCGCTTTATAAACTCGTACGGAACAACATCTTTAACTGTTCC
GGGAACAATTCTTTTCATCTTTGGCAGAATAAAATCACAAAAGCCGCGGGAACGTCCGATGAT
GTTTTCAAAAAGCGGGAAATGCTTTTCATGAATACTCGAGGAAGCACCGGAAGCCAATGCAG
TAAAGCGCAGATCATCACCGGTGTGCAGCTCATAGAGCGCGTGTCTCTTCATGCATCAC
GGAATAAAAGGAGCTCATGACCTCATTGAGATGATAATGGGTGCGTACGGACATCATCTT
TATTGAACTCCAGCGTAAACGGATGTTCCGGTCTCCGTGATGGCGCAGTGATTGCGGTCAATT
GTCAGGACGTCCATCAGATAATCGGAAAGTTCGCGCTGTTTTTCGATTGGCGCCTCCTGATG
GAGGAAATAATCTCTCGGCTGCCGGCCCCGTTCTGTGATCTGATGAACCAGCGGGACGATC
GTCTTTTTGATTTCCGCAAAGAAGTTATCCAGGAATTCTTCCGTCAGACCGTATTCATAATCA
TTGAGCAGCACCTCGTAAGGCTTCTTTTTGTGTTCCGGATCGCGGTATTCGATAAAGCGCAG
CTGGGAGGCAACGATTTTTTCAAGATACGGAGCAAAGGATGCAAATCATTCTCCTGTTTTG
ACTTCACCCAGACATTCTGAGCTTCCGAGACAAGCACATCATAGGCCACATACTCATCCTGC
GGGATGGCTTTCATGTGATCATAGTCCCGCTTCAGGAATTCAACCCTACGGGCATCCCGGG
CAGAAAGTTCTGACTTCCGTTTCAGAAAGTTCGCGGAGCAGTTCGCCGGTTTTTTCCGAAGCG
AATGCAGTGTACTGTTCCCTGCTCAGGTATTCCATCACCGCGCTGCGGCCGATCGTCTGTGT
CATGCGGAGCGGCCGTGTTCCGCTCCAGCTCCACCAGGGCGGAAAGATGATGAAACACGT
AAAGCGACCGTTCAGTTCGTCTAATTGCTTCACTAATTCTTCTGTGTTTTGCATTTATGATTC
TCCTTATATCGATCAGAGATTTTTGCTGCAGACTGTTCTCGATGGGAAAATGGTCCGTTCTGT
CTATGGGGACTGTCCATAAAGGGACAGTCCCCAAAAAATTTTATTCCAGACGGAAATTAT
AGATCCCGTCCACGATTGGGTAAACCCTGCCGCAGTCCGGGCAGCTCATCGCTTCTTTGGA
TTCGATCAGATGTATGCTGCCGCAGGCGGGACATTTGAAGACCGGGGCAGTCCGATCATAA
TGGCTGCCTGTGATCTGTGTGCATTTTCAGGAACACACTGGGGCTGAGTTTGACACCGGAAA
GAGCATCCTGCAGGACATTTTCCACCTCAGCATCCAGTCCAGATTTTTTGGATTGTTCTTCA
AGGCACCGATACGCAGGTTGGAAACGGAAAGCGCTTTTTCAATGGTAAAGCCGTGTGCTTC
CAGTGTGTTGCTTGATCGTTTTCCGGATGGTTATCGTAGTTTCAGTTTGACGAACTCCACCTGCTC
TTTTGTGAAGGGATTCCATTTCTGCTTCCCGGTTCATGTATCGAAGGATCGCCTTGATATTGCG
CTTGTTGGCAAATTCAGAATAAAGGTGGCATCCTTCTCCAGCACCCGCTGTATCTCCTCCA
TTGCCGGATCAATCTCAGAAAGATGATGGATCACCCGGATCATGGTTGCACCGTCAAAGACA
TTATCATTATAGGGAAGGTGATATACATCAGCCGCGGCATATACCAGCCGTTCCGATTCCGCC
CAGACGCTGTTTAGCCTGCTGCATCTGCGTCAGGGCATAATCGAAAAGATAGATCTTCTCAT
AGCCGCTATAGCGCGGAGTGTGCGTCCGCGCACCCGCACCGATTTCCAGCATGCGCTTCC
CGGATTTCCGGCAGCAGTTTTTTAAAGCGATATCTTCTGACCGGTCTCATAAGCACGATTG
GCATCATCCCAGAATGCTTTCTGGTAATCCGATCCGTCATAAACACAAAAATCCGGGGTTTT
GTCTTCTGCCATGCTGCCTCCGCTTTGAATTAAAAAAATACGTCCGGACAGGTTTGCCGGCG
GCGCCAGTAGATCTGCCTTATTGCTGCTTCCCTCTCGGACCTGACAAGGTTTCGACAGTCCCC
GCCGCGCCGGTCCCGCCCGAACGCACTCTAATCATACCATATTTTTCCGGAAATTACAAAAC
AGAGAGAATGTTTATGTTATTGTTCCGTGTAATCTTTTATGATGCGGTATAAGGAGAAAGAAGGC

CGCTCAGCATGTCCGCGTAGGCTTTTCGCAGAACTTCAGCTTTGCCTTCTGAAAATTCTTCT
GTCAGCTCACCCGGGGTAATGGACAGAACAAGAGGCTGTGTACCTTGCCTCTCCCCGGCAA
GAGAACGGGATATCACCTGCATCAGTGTCCCTGCCAGTCAAATTGATCGATGCCAGCCA
GTTCCCTCCAACCCGGACGGAGGAGTAGATACACCAAGCATCTGAATATTGAGAGAAACAA
GTTTTTGCAGCAAAGCTTCAGAAATTGCCTCATCCGAAAGGAACACGGTATAGATAAAACTCT
TTTGCCTCATCGAATTGCGCGCTCCAGCTGTCCGGTGAAGATGAACCCGGGAGAGAAAC
AACCAGCGGGAAACTGACATAAGGAGCCAGAGAGGAGCGGCAATTCCCGCACATGTAATGC
GCCCCGTTCAAAAAGGCTTCTTCCGCGTGTGAACCCACACCGGGGAGTCATTAGGAATCA
TCCCGATAGAACGCCAGTCATAAGCATTCTCAGTCAGTGCCAGTCCGGCCAGAAATGGCAG
AAATGCTTCATCATAATGAACGGTCCAGGCATTGTCGTAAGCGCTTCCGGGAGAAGCGATCA
CTATAAACGGGTTTGCAGAAAGCGGTCTGTGAGGGTAGAAAGAGCTGTTGGTTCTTTGGCA
AAAATCACAAAATCCGTATCATTTCGATAGCATCTTCACCGGAAATCGTCATACAGTCATAC
CCGGCACAGATCGTATCCAATGCCTTTGTAATCAATTCCGGTACACCGGGAACCTTATCCGA
TGGAATAAATGCGATTTTTTCCCGCTGCGGGAGCGGCGTCCGGCTTTCAGTTGGTTCAGGG
ATATCTTCAACGACCGGTTCCGGTGTGTTGTAGGCGTAACGGCCACACGCTGCGGCATGT
TGCAGCCCGCAGCCATCAAACAGTAAAAATCCCGTAAAAACAACCTGAAGCTTTTTTCATC
CGATTGATTCAGCCTTATCCTCCGGATCATTTCAGATTCCGGTTCATTGGAACCAAAGACAA
TTTTCGTCAATATTCATATAATTCCGGCGGAAATAATGATGCAGCCCATAGGACAGACGCTCG
ACCATTTCTCCATACGGTCTTTTTTGTATTTGAAGAAAGAATCGATTTCCAACCTGGGCAATC
AGAGAATCAATGCGCTCATCACTTTCACTCACAGCCGCGTAATCCGGTACAGGCAGGTCTGCGT
TTTTGTTAGGAAGGCTTTCATCGTCCGAATATCTGACTTTTCGATGAGCCATTCCGTCCGCC
GACAATCAGGCAGCCCTCATTATAACGCTGTTCCAGTTCATCCAATGCGGTTCATCCATGGCC
CGATTTCCGCAGTATGAAGAAACGGAACCTGATTGACGACAACCGTATCCCCGACAAAATA
ACTTTCCGCGCCGGGATTTCCACCCAGGTGGATCCATAAGAAGGCCCCGGATGGGACTGCA
G

Sequência do clone C2-16:

CTGCAGGTTGTGCTCGCGGCGCAGCATGAACTCCACGTCGTGGAACGCCTGCTGAGCCAG
CAGCGTCAGGGCCTCGCGGCTCACCTTGATGATGTCTTGCCTCAAACCTGGCCGACGCTG
ATGCCGTCCTTGCTCAAAGTCTGTACTCGGTCTTGTCTTCGCCAGCTGAAACATCGGGGC
GTACTTGAATGTCGTTGTTGCCATATTCTTGCTTGTTTTTTTGTTTTTTCGGCTACAAAATTAC
AACATTATTTATAAATAACCATCAATGCGCACATTTTTTTCAAGAAAGAAAGGCTTAAAAAAGT
TTTCATTTTATTTGCATAATTGCAAAGAACCATATACCTTTGCGACGGTGATGCTGCCAACTTA
CTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCC
CTCCTGTTTCAGCTACTGACGGGGTGGTGCCTAACGGCAAAGCACCGCCGGACATCAGCG
CTATCTCTGCTCTCACTGCCGTAACATGGCAACTGCAGTTCACTTACACCGCTTCTCAACC
CGGTACNCACCAGAAAATCATTGATATGGCCATGAATGGCGTTGGATGCCGGGCAACAGCC
CGCATTATGGGCGTTGGCCTCAACACGATTTTACGTCACTTAAAAAACTCAGGCCGCAGTCG
GTAACCTCGCGCATAACAGCCGGGCAGTGACGTCATCGTCTGCGCGGAAATGGACGAACAGT
GGGGCTATGTCGGGGCTAAATCGCGCCAGCGCTGGCTGTTTTACGCGTATGACAGTCTCCG
GAAGACGGTTGTTGCGCACGTATTCGGTGAACGCACTATGGCGACGCTGGGGCGTCTTATG

AGCCTGCTGTCACCCTTTGACGTGGTATATGGATGACGGATGGCTGGCCGCTGTATGAAT
CCC GCCTGAAGGGAAAGCTGCACGTAATCAGCAAGCGATATACGCAGCGAATTGAGCGGCA
TAACCTGAATCTGAGGCAGCACCTGGCACGGCTGGGACGGAAGTCGCTGTCGTTCTCAAAA
TCGGTGGAGCTGCATGACAAAGTCATCGGGCATTATCTGAACATAAAACACTATCAATAAGTT
GGAGTCATTACCCTTTGCGACCATTATAACTTAATAAACCTTTTAAAACATTCAACCATGAAA
AAGATTTATCTGGCACTGATGTGTATGGCAAGCCTTTTCGCTGATGACGGCTTGTGGCGACAA
GAAATCGGACAAGGGCGACGCCGACGAGAAAGAGAAGACGGAGGAGGTGGCCGACAACGA
CGAGCAGAACGGCGAAGAGCAGGAGCTTGAGGAGCAGCCCACCTTCTGATGAGAGCGAAGT
GTGGGGCAATCCCTATGAGGCTGANNTCCAGGACTTGGCTGCCCTCTACGCCAGCGGGCAG
TTCAAGCCCGCACCGGTTGTCATCTTCGAGGACACGCTGACGGAGGCCACTGAGGAAGCAG
GCGAGCTGCCCTCAAAGTGGGACATCAAGGAAGGCAGCGCCGAGATAGGGCAGGCTGGCG
GCCACCGTTTTATCAAGATGCTGGGCGGCACCACCGTGCTCACCCCGCAGGTTGACAGCAA
GGACTTCTGCCCTCGAAGTACACTGCCGAGTTTGAGTTCATGTTCCGGCAGGGACACTTGG
TATCGCGTCTTCTTCTACGATGCCGAGGAGAATGAGGTCCGGCAACTATGTGATGTGGTATTA
TGCGGCTGAATGGAACCTTCTGCAAGGCCGACGACGAGTGGATTCATGGCAACAAGGACGAG
TTTGACAAACTGCTCAACCGCGACGGCTGGAACCACTTTGCAGCCAGCTTTGACAACGGTAA
CATGAAGCTCTTTGTCAACGGCAAGCGCATTGCTAACATTCCCAACATCAAGCAGGCCGCCT
ACTTCACTATCAGCGGCGATAATGCTGACGGCGAGAGCCACTACATCCGCAGCATCCGCGT
CTGCAAGTAACAAGTTCAAGGAGTTGAGGAAACCTACTTTACTTCTCAACTCCTCATCTCCT
CAGTCAGCACTGCTTTATTCTTTTTTCATCTGCCATTTGTACAAAACTGTAACCAGGCTGGT
TACTGACTGTAACCAGCCTGGTTACAGTCAGTAACTCGCTTGGTTACAAAATGACACATTAC
CACAAGACACTATAGATTGGTTGGCACAAAAAAACCGCCGCTATGACTGAACTTAGCGGGC
GTTATTGAAAGTGTAGTCAGGGACGATTACTTAAGCACAACTTTACACCGTTATGGATGTAC
AGGCCCTTGCGGGTGGGCTGGCGGTGCGAGCTTCTTGCCGTGACAGTGTACCAGCCGCGC
TCATTGTCAACTGTGTCTCCAATTGAAATGATGCCCGTGGTGACGTGCCCGCCGAAGTCGAG
CACGAACTGGCGGGTGCTCTTCGCAATGGCTTCGGCAAAACGGAAGTAGGCGCGGCACGA
CTTCAGCGTACGGTTAATAGCCGTATGCTTCAGCGTGTGCTGCGGTGCTGAGTAGATGT
CGTCGTCCGCTGCGGTGATGGGGAAGGCGTCGTAATAGCCGATGAAGTTCACCTGACCGC
CGGCCTGGGAGACGGTGCCTCAGCCTCGGTGGTGTGACGACAGTGACACCCTGGAACA
CTGGCTCCACGATGTTTTCGCCGCTCTCCCACTTGATGATGTAGGGCAGCCGGCCTTCAG
CTCCGTACGGGCTCCGTCTCGCTGTGCGGGTTGCCGAAGGTCAGGCTGACATTTGTGCCG
TTCACCGAGGCGCCCGTCAGTGTCTTGGCCGTGGCACCCCTCGAGGAGTGAGCCTTCAAGC
GTCACGTGCAAGGGCAGGCACAGCGTGTCCACTCCTTGCCTTAAAGAGCGTGCGGCCGC
TCAGCGTCACGTCATAGGGCTTGGCGGTGGCGGCAGCCTCGTTGATGGCCGTCTCGTTGG
CAGTGCCGTGTTGCTCAGCTCCAGCGTGCCGGGAGTAACGATGTTGAAGAACAGCTGGTT
CTTCTTTTTCAGCACATTGTTATTCTGTGCTCTGTGAAGCGGGAGCGACTGTGCGGTGCTAGT
CGGTCTTGCTGCTGTGGGCATAGAGCGCCTGGTTGGATTCTGTGCTGTACACGCGGCAGGC
CATGTGGGGCTGTGCAGAGCTGGATGCGGTGTTGTGCTCGATAATCACCGCAAGGTTGGCC
GTGCCGTGTAAGCGAAGGGCTTGTGAGCGTGACGAAAGTCCAGTCGTCAGCCACCATCG
TCACTTCGCCGGAGAAGGCCTTGTGCTCGGCTGTGCGAGTAATCCAGTCGGTCTTGCCGCT

GAACACCACCGACGTGATGCTTCCCCGTTGGTCGTACCAACTTGAAGAACTCGCACTCTCAA
ACGGGCCGACACTCATGCCGCCACGAGCCTCCTTCTGGTTGTCGTAGTAGAAGGTCAGCAC
GGTGTGTCGTTGCTAAGCACGGCATAACGGCTCCTGAGCACTTGCTTTGCTGCCGATAAGC
AGCAGGGCGGTGAGCTGGAATAATTTTTTTAGGGTTAATCGTTTCTTCATTGATTTTTGAACT
TTGTTTTACATTAATTATTTTCAGTCCAAAGCCTCGTTATAATAGACAGAAAAAGAAGGGACGT
GGACTGATTACTTCGTCTACGTTTCTCGAGGTATCGCCAAACACCTAACACACTATATCCGA
GTAAGGCCACGCGGCCAGCGTGAGCATCCTTTGCTTTTACTCTTGGTAGTTCTTGAAATTT
TGCGCAATTTGTAGAACAAGATTCAAGCGGACGCTTCTATCCTATATGTCTTTTATGCGCTGT
TGTCTATGTCATGGGCAATTGCGGTTTATCGGGTTCAACATTATGTTCTATTTTTTTTTCTAAT
AATTCTACGGCTTCAATGAATGCCTTGGCTTTCGGCAGGAGAAAGGCTACGCTACTGGCATC
ACGGAATGCATAAGCCTCATAGTCACTGGCATGACGCTGGGTTGAATAAATCTGTCAGAGTC
GCTCCGTGCTCTATGCTCAGCAAGCCCTTACGGATAAAGTTCATCGATATCATGGTCTTTACA
CCATTGTGGGTAGAGGTGTCTATTCCGTTGGCTGTAAGAAGGGCACGTGCTGCATTAACA
TGCATAATAAAGGCGGGTACTGCTCCGTTATAATATCCACCGTCACGAAGATACTCCGCCT
CTTTTATCGACTCGTGCGCCCGTTCAAAGTAATAGGCCACAAGGTCGTTACGCTGTTCTTGG
GTCAAACCTTCTCGTTCATAGCAGTACACCTTCGCGTTTTACATTTTGATAGAAAGGAGTGA
CTATACGAGGATTCTCCATTGATGCCGAGGCATAACGAGAGTGCTGATAACGACACCAGAT
TGTATTTCTATTTTATATAATACTGGCAATAGTATGCTCCTGTTCTGGAGTCACAGTTTTCT
CTGGCAGTAGAATAAGCAAGTCGATATCACTGTCAGGGCGCGCATCGCCGCGAGCCTCAGA
GCCGTAGACGATGGCTTGAGCCTTGGGCAGCATTGCATGTGTCTGCTCCCTAATCTGTTCAA
CAACATTCGTACGTCTCATATCCTTTGCTAATCATTCTACGCTTGCAAAGATACGCATTATTC
CGAACTGCAAAGGATTTCCATGAAAAAGTCTTTGCAGGGCGGTGTGGAGTGGAAGTCCGG
ACGGGGCTATGACTGGCAACGGACGGGGCAGGATCCCCCTCAGGACGGAAACCGCCTGTT
TCCCAAATTTGTCAGCATTATAGCCTGTCGATTTCTTCAATAGTGAGTCCAGTGTACTTAGA
GATAAGTTCTGCGGACATGCCATCATTTTTTATGCGGAGTGCGGTGTCAGTCTTTTCTGCTG
CCCGGCCTTTGGCCTCGGCGGTGCGGATGGCACCGTCGTTGTCGTAGTAGATCTTCTCGCT
CTCCCAGTACTTGGCGTACTCCTCCTTCGGCAGATTGCCTATCTCAGCCACCTCGAAGA
TGCCGAAGATGCGCTCCTGCAGGGCCGACGACGCTCCAACAGGCGCGGAAGGTTCTTCA
GCGCATAACAGCCACTTGTGCGCCATCGTCTCAACTCGTCTCCGTCTTGTGGAACCTCGG
CATCTCAAATAGATATAGGTCAGCTTGTGCTAAAAGACTTCTTCTTGTGACGTCCATCAG
CTGCACCGTGTGGTGGTAGTACTCCTTGTGCTTTTGTCTTCAACACGAAGTCAAGGA
TGCCGATGGTATAGAGAGCCTTCAGGCCGAAGCGCCACTGCCCTTCTCGCCCTGCTGACG
AATGGGGAACGACGAGTAGTAGACCGAGCGGTCTTGAACCACTGCTGGTCCGGCCTTCTGC
ATCTCGATGATGATCTTCTCGCCCTGCTCGTTCTCACAATAAACGTGGAACACCGAGCGGCG
GTCGTCTTCTGTGTCGCCAACTGCTCCGTTGGCAGCAGCGTCAGCGACTTGATGACCTCC
TTACCCTGCAG

Sequência do clone G8-127:

CTGCAGGTTTTGACCCTCAGAGAGAACATCAGTTGCCACAAGAATTCTTGTCTCACT
CTCGACTGATATATCTGCCTTGTGCTGACAGGTGAGAAATGTTCTACGATACTTGTAGGATT
CTTACTGCCACCAGTAGCACGATCTATCTGTTTGAAGCCTCGTTTACGCAGTTGGCGATAAA

TATAGTTGGCGGTGTCGCTATATTGGGTAAATATCAGCACCTTGTCATCCTTGTGTGTCTTGT
TAAGCAACTTCTCCAGACAATTGAGCTTTTTGGTCTGTCTGCGGGTTCCACTTGTGCGCAAAGC
AGAATCATCTCGATAATCTTCTCGCAGTCCTGACGCAGATGCTGTTTGAGAGTACGCTTGAA
ATAGGCAGGATTGATCCACGACACATTGTTCTTCATGGAGAGTATGTTGTAGTAATCCTTGG
CCTTCTCCATATACACCTTCATATCTGTAGGGATGTCTATCAGACCGTCACCAGCCTTGCGG
TCTTCCTCGTCATCGTCGTGCGAGAAGATATCGTTGATGTCCGCATCTTCAATCCACTCTTCA
GGCAGTTGGTTCTCGTCGCCAATAGGTAATGGCAGTTTATTCTCGATGGCATAGATAAACAC
GGCATTACGCAGGATGTGGCGATAGAGTGTGAGCAAGAATGAGAAACCACTACTGTCAATG
CGCTTGAAGAAGGTGCTCTTGCGAGAAGCCATCATGCGCTGACCTGCTCGCGATAGTTTT
CCAACAGCTGTTTGTGCTGGTCGAGGCATCTGCCGTCTTGCCCTTGTCTTCATACTGGGAG
AGACCATAGCGAGGCAGCCGCGAGCTCTTCCATCATGGCAATCATATCGGAACTATAGAGGC
GGCTGTACTGATCGCCCTCCTTCGTATCAAACCTGATGGCCTTGGGTACGCGGTCTGGGAA
ATAGGACTTATGCCCATCGCGGAACTCCAGATACTTGCGTCCGTTCTCAGGGTCGGTCTTGG
CATAGTTCTCCTTGATGAAAGTACGTGTGCGACGCACTAGGAAGAGTTTCATCAGTTCCTGC
CAGTCGTCGAGCTTGTGCTTTCGTTCAAATGCCTTGATGCTACGGATGAACACCTCGCTATG
TTTTTGCAGGAAGGCACGGTCACCACCCAACGAGCGGATGTATTCCTCAGGACGGATACCC
AGGTCGTCATCCTCGCTGATAAAGAGTTTCAGCTGGCTGGCAAGGTCGGAAAAGTCTTTATT
ATAGGGTGTGGCAGTCAGTAGCAGCACCTTGCTGTCCTGACGCTCAATGAGGGCCTTGATG
TTACGGTAGCGCACACCATTCTCGTTGCGCAGGTTGTGACTTTCATCCACGATAATCAGGCG
ATAGTAGCGGGCATTATCCACATCTATAGGCTTGGCCATCGACATGATATCTGCTTTCAAGTC
ATAGCGCCGACGGTACTTGTCCACATATCTTGAAGGTTGGCAGGACAGATAATAAGCGTGC
TGCTGGCAAAGTCATCTCATAAATCTTGGCAATTGCGCAAGCAGTAATGGTCTTACCCAGT
CCCACCACATCGCCAATCATAGCACCACTCGCTTGTGCTGTTTCAGATGGCGGGCAGCTA
TCTTGACGGCTGTCTGCTGGAAGTCGAAGAGGTCGTTTTGGAACCTTGCGGGCAACGTGAA
CTCCTTGATGCCGCTACGCGCATCGGCACTCAGATGGTAGGCTGTCTTCAGATAGATATAAT
AAGGTGGAATGACGTCCTCGCCAGCCCAACTGTTGTGCGATGGCGTCAATGAGTTCCTTGGT
GATGTGATAACAGAAGCGGTCGTTCCAGCGGTCGTCAAACCAGGCAGAGAGTTTCTGCGCA
CTATCACTATCACCAAACCTCGGCATTCAGTTCACCCTGCTTGGTCAATCCGGCATAGGTGAG
GTTGCTGCTACCCATGATGGCCTGAATGGGATTGAAGTTGTGCTCAGGGCGATGGGCCATA
TAGAGTTTGGCATGAAGTGGTTCTGCTACATAGAGTTTACGCAAACCTTGTCTTCTTTCAAC
TGGGCTGACAGGCGGCGTAGTGTCCATTGCTCTTGTGGTAGGTAAGCCCAAGAGGAGCT
GTTTCTTGAATTCACGTGCTATTTCCAACCTGCATTGTTGGGCGTAGGCAGCATCAGTAGGTT
GTTCTTTGGAATAGAGTCGTCGCACCAGTTCCTGATCTGGCTGATGCATACCAATCAGCAGG
CGACAGCAACGAACTCCTGCCTGTTGTCTCGTAAACGTAGTCGCCAGGAATCTGGTCTAT
CTGATTGACCACCAACTGCCAGCCGCGAAGATTGAAATATCCCACGCAGAAGTCCACACGC
TTCACACCCATATTGGTGATGATATCCTGTAGCCCCTGCGTGAATTTAGTCTCAATGTTGTCG
TAAATTCTCGCCATTCTGCACTTATAGTTGTTAGGAGCATAGCACTTACCTGGGCTTTTACAG
GCACCACCCACCAACCGAACAGACACGAAAAAAGTGCGAACTACACTATATCCGTGTCTGAG
GTTGTAGGAAAGACCTGTTCTGCTAGATAAGTGACATGTCCGCACTATCAAGCTGCAGATTA
TCAATTACCAGTAGACGAACTACGTGGTATAAACTTCCTACATTTACAGACACGCCGAATAAC
TGCTAATGCTGTTATTATTTACCCTCAAAAATCATACCGCCACCCTGCGCTCAGAGCCCTAAG

CCGTAGCTCGCGTCGTTTACGACGATTAGGCGATACTTGATGCAAAAGTACAATTTTTATTTG
TAAATCCGCACTTTATAATTA AAAAGATGTGTAAAAATGGTGT TTTTGAGGGCGAAAAAGACAT
GATTTGCCTTCAAACACACAAAAAGAGTAGAGTTACTTATAGTACCCCGGTCTGATATCGTG
CCTTGCTGTGGCAAAGGGATTCCAGGAGATGTGGAGCCAGCCGTTGCCCGTCAGGAGTTG
GTCGGTATACTCACATGCCTTGAGGAACGAGACCAGTCGCTGGAAGTCTTCGGATCGGCA
GGTCGGATGTGCGGCGGCCTGGCCGATGAGGTGCTGCGAGCTCTTGGCTCCACCTACCAGA
CGGTTGACGCGAGCGTTGCGGAAGCCAGAGTTGATGATGATGGGACAGCCGATAGCTTGTG
GGGCGGGTTCCAGAAGCAGATGACAGCCATAGGACATGTTACCACATCTTGTAGCGAAGG
CTTGTGTCGGGATACTTGGCAACGTTCCAGGAACTCGTCGAGTCGGAAATGAGGCGACAA
GCGAACGTCATTCGAGAGTTTGATGTCATTAAGTGTGATCATGGTAATGCCTTTTAATGA
TTCTGATATACTGATCGAGTCCGAAGATCGAGCCGGCAAGCAGGAACGACTGAGCCACATA
GTAGAGCAGCCCTCCTGCCACGTCTCTGATCTCGACCACCTGATAGCTCACC AAAGCGATG
CTGCTCACAATGAGCAGCACCGCACAGGCGTACTGAGAGATTTTGAAGCCTTTGTGTTCCAT
GGTATCCAAGATTATTCGTATGTGGCTTCTCTTACATTGCGGCACCGCCTGCCGCACCAAGG
ATAAGGGTGATGATGTAAGTACTGATCGCCTGAAGGATCGTCTTAACTGTTTTGTTTCATAATTCTT
AATGTTTTAAATGGTGAATAATAGGGTTATTTATCTCACGCATATTGCGTTTTGTTTTACCCTAC
AAAGATACGAAAAAACTGAGACTACCAAATGTTTTAGTCAGAAATTTTCGATCGCATTCTT
GCGCCGAATCATCTGCCCTTGATTCCATCAATGTGCCACGTATATTGCACTCAAAGCTTTT
TTGCGGGTGCGAAATAATGTCCCGAAATGAAAATCGGGTGTTTCGTAAGTTTGCATGAAGCAA
ATAGCAACACAAAAACATAGTAATTATGGCAAAGGAGAAAAAGAAGGGCAAGACTCTAAAGA
AAGTGAAAAAGGAATTTACTCCTAAGACACGTCTATTTGATTTGGTAAGCAAGAAATATCATA
GTGACCTATCTCTGCTTGTGGAGGAAATAAGTCCATGGCTCATTGATAGAGAGAAAATTGATT
GGAAGAATCTGAGAAGGCTTGAAGTAGACTACTCAAAGACAACCAAAGACAGTTTTCTCTTA
GTTATTACCAACTTTCTGTATAAGGTTGGACGAGGTAAAGGACTGAAATGCAAAATGGAGAC
ATTAATAAGGTATTTGGCTTCAACAGAACATAGCAATTTTGGCCTCAAGGAAAGCTCACTTAA
AGCAAAGATATACAGCATGTTGCGTTATCATGAAAGCGTAAAAAAGAATAAGAATTTGGTAT
CACATGAAACATAAGAAAACATATTGATTGTTATGGCAAAGGGGTAGGATTTGTATTGTTATC
AGAATAATTGTACATTTGCAAAAAATAGTTATAGATATGGTGGCAAATAATGAATATCCGGAA
CGCGAGCCTCGCCTCGTTGGAACCATTATCATAGAAATGCTTCATGATGGCTCCGCGCTTGC
TGTTGGTTACCGTCGATACCTAATTAATAATATATATAGAGAGTAAATGATTATGGAAACTAAG
AGATACGTTGTATCAGCCTTTGCGCCGACCGATATAACAACCAATTCCGTGTGGTGAATCTC
GTTGTTGACGTTTATCGCTATGAGGTGAGTACTAGAGCAGTGGAAACGCAGAAAATACTTC
GCGCCATGACCGACGAGAAAGCCAAAAAGCTTTTAAGGACAAGAAGTTCGACTGGGTGAC
CCCAGCAGGAACGTTGAGCTATGCCAGTGATGACAAACAAGTACTGACTGCATCTGAAGTCGTCT
GTATGGATCTGGATTATCTCTGTCTGCCATCGGAGATTGATGAGGAGAACGGTGACCCAGTA
ACGGAGCTGAAGGAGAACTGTTGAAGGATTCTACTTTGAGACACTGCTGTTATTCCGCTC
GCCTCGCGGACGCGGATTA AAATGGTGGGTGCCCGTTGAACTCACAAGTGCGACCATCGC
ACATGGTTTACTGCGATTAGGAATTATGTGATGCAGCAATATAAACTGACAAATATACAATGT
GACAAACAGGTGGTCAACGAGTCGAGAGGATGTTTCTCGGCCATGATTCTCACTGCTATCT
CAAACCCGAATTGTTTGAATACTTTTAAAATATTATCACTATGGAAAAGAAGAATTTGACCCA
ACACAGTGGGTCAATGCCGGAAGTGAAAGTGAGCTGCCAAAAAGACGGCAGCTTCTGTGT

CTAACCTGCTAGCGACAGTCTTGCTTTAGCCGACAACGAGTTGGAAAAGGCCCGCGCCGT
 ATGCAATGAACTGCTGCAG

Tabela 14 - Primers construídos para o sequenciamento.

A1-7	A1-7_F_p1 GTGATCTGATGAACCAGCGG A1-7_R_p1 GTTACGCCTACAAACACACCG A1-7_p2_F TGAAACACGTAAAGCGACCG A1-7_p2_R GTGTTCTTTTCGGATGAGGC A1-7_p3_R GAATCCGAACGGCTGGTATA
G2-8	G2-8_p1_R TCCACTACCAGTTCTACACGC G2-8_p2_R GAATGCCTTTGGAACGACCC G2-8_p1_F GTGAAGATGGTTGGCAAGCC G2-8_p1.2_F GGAGTGAAGATGGTTGGCAA G2-8_p4_R AAAGAATCAGGACCTTGCCC G2-8_p2_F ATCACACCGCTGAATGTCC G2-8_p5_R GACAAAGATGCTGTGATGGC G2-8_p5.2_R TGC TCT ACG ATA AAG TCC GC G2-8_p2.2_F GCA GGA ACG CTT TGA AAA GG
E10-15	E10-15_p1_F GTTATCTCGACAGCCTCTGC E10-15_p1_R GCAGACGAAAAGACTGACACC E10-15_p2_F TGGTTGTCTGGCTGATGACC E10-15_p2_R CCAAGAACAGCGTAACGACC E10-15_p3_F CATTGGCAGTGTGAAGCTG E10-15_p3_R CTTATCGGCAGCAAAGCAAG
C2-16	C2-16_p1_R

	CTTACACCGCTTCTCAACCC C2-16_p1_F TGAACGGCACAATGTCAGC C2-16_p2_F GCCGCTAAGTTCAGTCATAGC C2-16_p2_R AGCCCACTTCTGATGAGAGC
H11-35	H11-35_p1_F TTTTTGGGCTTTCTCGTCGG H11-35_p1_R TGTGGCGATAGAGTGTGAGC H11-35_p1.2_F TGGTTGTATATCGGTCGGCG H11-35_p2_R TGCTACGGATGAACACCTCG
G10-36	G10-36_p1_R TGTAGTCGTAACGCTCGAAGG G10-36_p2_R GGTATCAACCAACACGACACC G10-36_p1_F CTGCTCGAAGAATTTCCGAAGG G10-36_p2_F CGCAAGGCTGAAACTCAAAG G10-36_p3_R ACTCAAAGGAATTGACGGGG
G8-127	G8-127_p1_F CAACGAGATTTACCACACGG G8-127_p1_R TGTGGCGATAGAGTGTGAGC G8-127_p2_F CGGGACATTATTTGCACCC G8-127_p2_R CAATCATAGCACCACTCGC G8-127_p3_F GGCATGTGAGTATACCGACC G8-127_p3_R CCAACCGAACAGACACGAAA
B3-24	B3-24_p1_F TGCCAACATTGTCGATCTGC B3-24_p1_R CATTGCCTGTGTCAGCATGG B3-24_p2_F AGCATCTCCAGCAGCTATGG B3-24_p2_R CGGGTTCTTCTCAAACAGCG B3-24_p3_F

		TGGTGGGAAGGTGAACATCC
		B3-24_p2.2_R
		GTACAGGCTGTTTTTCATCGC
		B3-24_p4_F
		GTT ACC TGA CCA TCT CCA CC
		B3-24_p3_R
		AAT TTC CAG TCT CAG CGT CC
E7-23		E7-23_p1_R
		CGAGTCTGTGGGTGATTTTGC
		E7-23_p2_R
		TAACCTGCCAACCAAAAAGCG
		E7-23_p1_F
		TATTGTGCTGGCAGGAGTCC
		E7-23_p2_F
		GCCTTTGTCATGCTATGTGC
		E7-23_p2.2_R
		GACTCGCTGTATACCGTTGG
		E7-23_p3_F
		CCCATGTACCGAAGATGACC
A7-93		A7-93_p1_F
		ATC GCT CTC TGC CTT AAT GC
		A7-93_p1_R
		TCT TTA CCA ATT CTG CCC CG
		A7-93_p2_F
		GAA AGA ACT TGA CGC ATG GC
		A7-93_p2_R
		AAATCCGTTGAGAAGCTGGG
		A7-93_p3_R
		AGGAAGAAATAGACGCCTGC

Tabela 15 - Clone A1-7. Resultado do alinhamento das ORFs do clone no BLAST P.

Clone A1-7 (3.960pb)		
ORF	Tamanho aa	Alinhamento BLAST P nr
1	168 aa	hypothetical protein ANT_03110 [Anaerolinea thermophila UNI-1] 87% 8e-18 35%* beta-lactamase domain protein [Geobacter uraniireducens Rf4]64% 4e-05 29% beta-lactamase domain containing protein [Sporomusa ovata DSM 2662]46% 2e-04 33% beta-lactamase [Geobacter sp. M18] >gb ADW14610.1 beta-lactamase domain protein [Geobacter sp. M18] 48% 3e-04 32% glyoxylase-related zinc-dependent hydrolase [Geobacter metallireducens GS-15] 64% 0.002 31%
2	119 aa	hypothetical protein ANT_27790 [Anaerolinea thermophila UNI-1] 43% 3e-10 54%

		hypothetical protein CLDAP_07470 [Caldilinea aerophila DSM 14535 = NBRC 104270] 36% 1e-04 50%
		cyclopropane-fatty-acyl-phospholipid synthase [Lactobacillus oris F0423] 54% 0.85 34%
		Cyclopropane-fatty-acyl-phospholipid synthase [Lactobacillus antri DSM 16041] 54% 1.8 34%
		putative RNA-dependent DNA polymerase [Aeromonas hydrophila subsp. hydrophila ATCC 7966] 51% 1.9 28%
		cellulosome-anchoring protein, partial [Peptoniphilus lacrimalis] 32% 5.5 36%
3	305	hypothetical protein ANT_27790 [Anaerolinea thermophila UNI-1] 79% 3e-76 48%
		hypothetical protein CLDAP_07470 [Caldilinea aerophila DSM 14535 = NBRC 104270] 78% 1e-46 35%
		hypothetical protein [Streptomyces sp. BoleA5] 58% 3e-14 31%
		Transcriptional regulator, ArsR family [Thermoplasmatales archaeon A-plasma] 77% 2e-11 25%
		Phosphatidylethanolamine N-methyltransferase [uncultured bacterium] 61% 3e-11 29%
4	347	hypothetical protein [Chloroflexi bacterium oral taxon 439] 95% 3e-35 28%
		hypothetical protein ANT_29740 [Anaerolinea thermophila UNI-1] 96% 5e-25 29%
		hypothetical protein A1O1_01234 [Capronia coronata CBS 617.96] 31% 2.1 31%
		PREDICTED: uncharacterized protein LOC101883550 [Danio rerio] 39% 2.6 27%
		hypothetical protein [Fischerella sp. PCC 9339] 18% 5.6 34%
5	404	Zn-dependent carboxypeptidase [Faecalibacterium prausnitzii L2-6] 98% 2e-138 52%
		Zn-dependent carboxypeptidase [Coprococcus catus GD/7] 97% 3e-137 52%
		zn-dependent carboxypeptidase [Firmicutes bacterium CAG:110] 93% 4e-137 55%
		zn-dependent carboxypeptidase [Faecalibacterium sp. CAG:82] 97% 7e-132 51%
		zn-dependent carboxypeptidase [Firmicutes bacterium CAG:83] 98% 1e-129 49%
6	174	hypothetical protein BN977_02820 [Mycobacterium cosmeticum] 59% 4.2 29%
		hypothetical protein [Magnetospirillum magneticum AMB-1] 30% 8.9 44%
7	288	hypothetical protein BN130_805 [Cronobacter malonaticus 507] 64% 6e-05 22%
		PREDICTED: cytochrome b-c1 complex subunit 2, mitochondrial-like [Alligator sinensis] 31% 1.9 30%
		PREDICTED: cytochrome b-c1 complex subunit 2, mitochondrial-like [Alligator mississippiensis] 31% 4.2 30%
		ATPase [Microbacterium sp. 11MF] 36% 6.0 29%
		ATPase [Microbacterium paraoxydans] 36% 6.1 29%
8	163	predicted protein [Monosiga brevicollis MX1] 49% 1.2 30%
		histidine kinase [Rhodococcus sp. R1101] 32% 7.6 36%
		hypothetical protein CAEBREN_25022 [Caenorhabditis brenneri] 29% 9.0 38%
9	318	C4-dicarboxylate anaerobic carrier, DcuC family protein [Dichelobacter nodosus VCS1703A] 32% 1.3 28%
		PREDICTED: glucan endo-1,3-beta-glucosidase 8-like [Glycine

		max] 42%	4.2	28%
		PREDICTED: glucan endo-1,3-beta-glucosidase 8-like [Glycine		
		max] 42%	6.6	28%
10	127	hypothetical protein, conserved [Plasmodium vivax]	39%	
		1.9	40%	
		acetate kinase [Haemophilus influenzae PittGG]	55%	5.0
		35%		
		acetate kinase [Haemophilus aegyptius ATCC 11116]	55%	
		5.3	35%	
		hypothetical protein CGSHiR3021_05314 [Haemophilus		
		influenzae 22.4-21]	55%	5.3 35%
		Acetate kinase [Haemophilus influenzae R2866]	55%	5.5
		35%		

* valores de cobertura, e-value e similaridade.

Tabela 16 - Clone C2-16. Resultado do alinhamento das ORFs do clone no BLAST P.

Clone C2-16 (3418 pb)		
ORF	Tamanho aa	Alinhamento BLAST P nr
1	389	hypothetical protein PBR_2752 [Prevotella bryantii B14] 56% 1e-34 36%*
		hypothetical protein HMPREF9144_1053 [Prevotella pallens ATCC 700821] 65% 7e-15 27%
		hypothetical protein HMPREF9296_1839 [Prevotella disiens FB035-09AN] 68% 2e-14 30%
		hypothetical protein HMPREF0670_00005 [Prevotella sp. oral taxon 317 str. F0108] 62% 1e-13 29% GLUG domain protein [Prevotella ruminicola 23] 19% 2e-04 39%
2	175	hypothetical protein RSal33209_3075 [Renibacterium salmoninarum ATCC 33209] 54% 1.9 28%
		hypothetical protein MYO_17400 [Synechocystis sp. PCC 6803] 53% 6.3 26%
		oxidoreductase [Arthrobacter crystallopoietes BAB-32] 46% 7.4 29%
3	109	hypothetical protein ARALYDRAFT_893970 [Arabidopsis lyrata subsp. lyrata] 57% 9.7 29%
4	236	insertion element family protein [Shigella sonnei str. Moseley] 98% 3e-170 99%
		insertion element IS1 5 protein insB [Escherichia coli O104:H4 str. 11-3677] 98% 2e-169 99%
		insertion element IS1 1/5/6 protein insB [Shigella sonnei 53G] 98% 3e-169 99%
		insA N-terminal domain protein [Escherichia coli ThroopD] 98% 2e-168 98%
		insertion element IS1 1/5/6 protein insB [Shigella dysenteriae 1617] 98% 1e-167 98%
5	274	OmpA/MotB domain protein [Bacteroides helcogenes P 36-108] 64% 3e-25 35%
		OmpA domain protein [Parabacteroides sp. CAG:2] 98% 4e-25 31%
		hypothetical protein HMPREF0999_00635 [Parabacteroides sp. D25] 98% 2e-23 30%
		membrane protein [Bacteroides sp. 2_1_7] 71% 7e-23 35%
		glycosyl hydrolase [Dysgonomonas gadei] >gb EGK02235.1]
		hypothetical protein HMPREF9455_01505 [Dysgonomonas

		gadei ATCC BAA-286] 50%	7.0	24%
6	150	hypothetical protein GLOTRDRAFT_70926 [Gloeophyllum trabeum ATCC 11539] 59%	0.90	36%
		Ndr family protein [Desulfitobacterium hafniense DP7] 37%	9.5	39%
7	122	transcriptional regulator, tetR family [Lactococcus lactis] 48%	1.4	33%
		response regulator receiver modulated metal dependent phosphohydrolase [Syntrophobacter fumaroxidans MPOB] 72%	4.4	26%
		hypothetical protein [Spirochaeta bajacaliforniensis] 54%	4.8	34%
		hypothetical protein LLT3_03595 [Lactococcus lactis subsp. cremoris TIFN3] 48%	4.9	33%
		transcriptional regulator [Lactococcus lactis subsp. lactis bv. diacetylactis str. LD61] 48%	5.0	33%

*valores de cobertura, e-value e similaridade.

Tabela 17 - Clone E10-15. Resultado do alinhamento das ORFs do clone no BLAST P.

Clone E10-15 (pb)		
ORF	Tamanho aa	Alinhamento BLAST P nr
1	196	conserved hypothetical protein [Prevotella ruminicola 23] 89% 1e-22 31%* putative uncharacterized protein [Bacteroides sp. CAG:702] 54% 1e-04 30% putative uncharacterized protein [Bacteroides cellulosilyticus CAG:158] 98% 3e-04 24% putative uncharacterized protein [Bacteroides dorei CAG:222] 73% 5e-04 28% hypothetical protein HMPREF1065_04108 [Bacteroides dorei CL03T12C01] 73% 5e-04 28%
2	231	hypothetical protein HMPREF9376_03037, partial [Enterococcus faecalis S613] 76% 1e-10 44% hypothetical protein RASY3_07585 [Ruminococcus albus SY3] 50% 0.86 27% hypothetical protein L916_07872 [Phytophthora parasitica] 66% 4.4 20% hypothetical protein L915_07944 [Phytophthora parasitica] 66% 4.4 20% hypothetical protein F444_08166 [Phytophthora parasitica P1976] 66% 4.6 20%
3	118	coiled-coil domain-containing protein 40-like [Strongylocentrotus purpuratus] 72% 0.003 38% coiled-coil domain-containing protein KIAA1407 homolog [Otolemur garnettii] 65% 0.072 30%
4	100	PREDICTED: ryanodine receptor 2-like [Ciona intestinalis] 49% 0.28 40% EH domain-binding protein 1-like protein 1 [Vicugna pacos] 47% 4.5 42% Zinc finger, C2H2-like [Penicillium roqueforti] 73% 6.6 33% hypothetical protein TCDM_01790 [Trypanosoma cruzi Dm28c] 61% 7.5 34%

		hypothetical protein TCSYLVIO_009706 [Trypanosoma cruzi] 61% 9.7 34%
5	104	nada
6	242	uncharacterized protein BN637_00624 [Prevotella sp. CAG:386] 41% 5e-07 32% hypothetical protein ThisiDRAFT_2483 [Thiorhodospira sibirica ATCC 700588] 28% 0.002 34% integrase [Kutzneria sp. 744] 24% 7.5 39% putative cytochrome oxidase assembly protein [Nocardia cyriacigeorgica GUH-2] 25% 8.2 38%
7	110	uncharacterized protein LOC101737051 [Bombyx mori] 55% 3.1 26%
8	122	Rhs element Vgr protein [Chitinophaga pinensis DSM 2588] 36% 2.5 40% hypothetical protein [Yaniella halotolerans] 58% 3.9 31% plasmid and phage replicative helicase [Sinorhizobium fredii USDA 257] 36% 5.6 39% spermatogenesis-associated protein 5 [Ciona intestinalis] 68% 8.3 29%
9	244	putative uncharacterized protein [Bacteroides sp. CAG:443] 99% 3e-98 59% hypothetical protein BSFG_02824 [Bacteroides sp. 4_3_47FAA] 99% 4e-96 56% Conserved hypothetical protein CHP01784 [Bacteroides salanitronis DSM 18170] 99% 4e-90 56% hypothetical protein BACEGG_01936 [Bacteroides eggerthii DSM 20697] 99% 1e-89 54% hypothetical protein C802_04306 [Bacteroides massiliensis dnLKV3] 97% 2e-88 54%
10	177	uncharacterized protein BN772_01365 [Bacteroides sp. CAG:754] 75% 3e-45 56% hypothetical protein HMPREF1064_00381 [Bacteroides dorei CL02T12C06] 75% 5e-44 53% putative uncharacterized protein [Bacteroides dorei CAG:222] 75% 7e-44 53% putative uncharacterized protein [Bacteroides intestinalis CAG:315] 75% 3e-36 49% uncharacterized protein BN644_00213 [Bacteroides thetaiotaomicron CAG:40] 74% 1e-32 44%
11	109	uncharacterized protein BN543_01763 [Bacteroides dorei CAG:222] 95% 4e-32 57% putative uncharacterized protein [Bacteroides intestinalis CAG:315] 95% 4e-31 53% uncharacterized protein BN772_01366 [Bacteroides sp. CAG:754]

		95% 5e-31 55% hypothetical protein HMPREF1062_02557 [Bacteroides cellulosilyticus CL02T12C19] 95% 6e-30 55% nucleotidyltransferase domain-containing protein [Bacteroides intestinalis CAG:564] 95% 3e-29 50%
12	102	oxidoreductase, partial [Pseudomonas aeruginosa PAK] 85% 1.1 30% pyochelin biosynthetic protein PchG [Pseudomonas aeruginosa BWHPA042] 85% 1.4 30% Pyochelin biosynthetic protein PchG,oxidoreductase (NAD- binding) [Pseudomonas aeruginosa 18A] 85% 2.2 30% pyochelin biosynthetic protein PchG [Pseudomonas aeruginosa BWHPA018] 85% 2.3 30% oxidoreductase, partial [Pseudomonas aeruginosa] 85% 2.4 30%
13	551	hypothetical protein HMPREF9332_01082 [Alloprevotella rava F0323] 62% 5e-82 43% surface protein [Prevotella nigrescens ATCC 33563] 61% 9e-78 48% PF03382 family protein [Prevotella intermedia 17] 61% 2e-77 49% hypothetical protein HMPREF1173_02170 [Prevotella nigrescens CC14M] 61% 2e-77 48% bacterial surface protein 26-residue [Prevotella nigrescens F0103] 61% 5e-77 48%

* valores de cobertura, e-value e similaridade.

Tabela 18 - Clone G8-127. Resultado do alinhamento das ORFs do clone no BLAST P.

Clone G8-127 (5346 pb)		
ORF	Tamanho aa	Alinhamento BLAST P nr
1	107	hypothetical protein [Flexithrix dorotheae] 37% 3.4 41%* PREDICTED: fatty acyl-CoA reductase 1-like [Chrysemys picta bellii] 48% 5.4 36% hypothetical protein EUTSA_v10003537mg [Eutrema salsugineum] 73% 9.1 27% hypothetical protein [Anaeromusa acidaminophila] 60% 9.7 34%
2	116	ABC-type transporter, integral membrane subunit [Streptomyces violaceusniger Tu 4113] 57% 2.1 36% putative dioxygenase [Roseobacter denitrificans OCh 114] 58% 3.9 35% ABC transporter permease [Streptomyces rapamycinicus NRRL 5491] 58% 4.6 36% putative ABC transporter, permease protein [Streptomyces himastatinicus ATCC 53653] 57% 6.6 36%
3	111	putative membrane protein [Prevotella ruminicola 23] 61% 3e-24 69% hypothetical protein HMPREF9019_0970 [Prevotella timonensis CRIS

		5C-B1] 46% 0.37 35% putative uncharacterized protein [Prevotella sp. CAG:1092] 49% 0.44 28% hypothetical protein HMPREF0645_2536, partial [Prevotella bergensis DSM 17361] 46% 0.62 35% unnamed protein product [Chondrus crispus] 72% 3.5 26%
4	100	PREDICTED: kelch domain-containing protein 4-like [Saccoglossus kowalevskii] 77% 7.3 27% hypothetical protein AOL_s00080g200 [Arthrotrrys oligospora ATCC 24927] 47% 8.4 40% Hypothetical protein CAP_6009 [Chondromyces apiculatus DSM 436] 46% 8.9 35% PREDICTED: kelch domain-containing protein 4-like [Nasonia vitripennis] 68% 9.4 28%
5	821	hypothetical protein HMPREF9140_00967 [Prevotella micans F0438] 98% 0.0 74% helicase C-terminal domain protein [Prevotella melaninogenica ATCC 25845] 98% 0.0 74% helicase domain protein [Prevotella pallens ATCC 700821] 98% 0.0 74% helicase C-terminal domain protein [Prevotella melaninogenica D18] 98% 0.0 74% DEAD/DEAH box helicase family protein [Bacteroides fragilis str. S24L34] 98% 0.0 68%
6	118	peptidase M15 [Prevotella stercorea CAG:629] 67% 1e-17 38% peptidase M15 [Prevotella baroniae F0067] 61% 2e-17 38% peptidase M15 [Prevotella stercorea DSM 18206] 67% 4e-17 38% peptidase M15 [Leptospira interrogans str. FPW1039] 62% 6e-16 38% peptidase M15 [Prevotella timonensis CRIS 5C-B1] 62% 7e-16 43%
7	163	hypothetical protein Moror_4160 [Moniliophthora roreri MCA 2997] 50% 2.7 31% aspartate carbamoyltransferase [Saccharophagus degradans 2-40] 95% 6.1 20%
8	263	hypothetical protein HMPREF9069_00508 [Atopobium sp. oral taxon 810 str. F0209] 25% 1.8 34% hypothetical protein HMPREF0649_01059 [Prevotella buccae D17] 29% 1.9 31% hypothetical protein PRUPE_ppa023919mg [Prunus persica] 37% 2.1 30% hypothetical protein BACUNI_01515 [Bacteroides uniformis ATCC 8492] 35% 3.7 26% hypothetical protein PRABACTJOHN_03530, partial [Parabacteroides johnsonii DSM 18315] 35% 4.8 27%
9	125	hypothetical protein A33Q_1309 [Indibacter alkaliphilus LW1] 64% 0.005 27%

hypothetical protein Celgi_2983 [[Cellvibrio] gilvus ATCC 13127]	32%	1.2	45%
polypeptide [Lygus lineolaris virus 1]	52%	2.2	36%
hypothetical protein ETSY1_34915 [Candidatus Entotheonella sp. TSY1]	36%	2.5	47%
putative exporters of the RND superfamily [Halonotius sp. J07HN4]	39%	3.0	43%
GDSL family lipase [Rhodobacteraceae bacterium PD-2]	%	8.2	32%

* valores de cobertura, e-value e similaridade.