

LÍVIA TAVARES COLOMBO

**METAGENÔMICA DE UM CONSORCIO MICROBIANO TERMOFÍLICO NA
DETECÇÃO DE GENES DE ENZIMAS CELULOLÍTICAS E EXPRESSÃO
HETERÓLOGA EM *Kluyveromyces marxianus* CCT 7735 (UFV-3)**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Microbiologia Agrícola, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2015

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

C718m
2015 Colombo, Lívia Tavares, 1985-
Metagenômica de um consórcio microbiano termofílico na
detecção de genes de enzimas celulolíticas e expressão
heteróloga em *Kluyveromyces marxianus* CCT 7735 (UFV-3) /
Lívia Tavares Colombo. – Viçosa, MG, 2015.
xii, 102f. : il. (algumas color.) ; 29 cm.

Orientador: Flávia Maria Lopes Passos.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Genética molecular. 2. Genômica. 3. Enzimas
celulolíticas. 4. *Kluyveromyces marxianus*. I. Universidade
Federal de Viçosa. Departamento de Microbiologia. Programa de
Pós-graduação em Microbiologia Agrícola. II. Título.

CDD 22. ed. 572.86

LÍVIA TAVARES COLOMBO

**METAGENÔMICA DE UM CONSORCIO MICROBIANO TERMOFÍLICO NA
DETECÇÃO DE GENES DE ENZIMAS CELULOLÍTICAS E EXPRESSÃO
HETERÓLOGA EM *Kluyveromyces marxianus* CCT 7735 (UFV-3)**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Microbiologia Agrícola, para obtenção do título de *Doctor Scientiae*.

APROVADA: 13 de março de 2015.

Denise Mara Soares Bazzolli

Cynthia Canêdo da Silva

Monique Renon Eller

Júlio César Câmara Rosa

Flávia Maria Lopes Passos
(Orientadora)

Ao meu Salvador Jesus Cristo

Ao Marcelo

À minha família

Dedico.

BIOGRAFIA

Livia Tavares Colombo, filha de Robson de Souza Colombo e Eunice Tavares da Silva Colombo, nasceu em 13 de fevereiro de 1985 na cidade de Manhuaçu, Estado de Minas Gerais.

Em outubro de 2004 ingressou na Universidade Federal do Espírito Santo (UFES), graduando-se em Licenciatura e Bacharelado em Ciências Biológicas em janeiro de 2009.

Em março de 2009 ingressou no curso de mestrado do Programa de Pós-graduação em Microbiologia Agrícola na UFV, com término em fevereiro de 2011.

Em março de 2011 iniciou o curso de doutorado neste mesmo programa, submetendo-se à defesa de tese em 13 de março de 2015.

AGRADECIMENTOS

A Deus, por ter me permitido passar cada momento desses quatro anos, firme, com amor e alegria no coração. Essa foi a força de que precisei para enfrentar os obstáculos que apareceram. Creio em um Deus, pra quem tudo é possível.

Agradeço à FAPEMIG e CAPES pela bolsa de estudo concedida.

À Universidade Federal de Viçosa, ao Departamento de Microbiologia, e ao Programa de Pós-graduação em Microbiologia Agrícola, pela oportunidade de crescer profissionalmente em uma universidade e curso de excelência.

Ao Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), onde se realizou este trabalho.

À professora Flávia Maria Lopes Passos, pelos ensinamentos, pelo carinho em todos os momentos, pela amizade e por me fazer crescer como profissional.

Aos professores Wendel Batista da Silveira, Marisa Vieira de Queiroz, Juraci Alves de Oliveira, Denise Mara Soares Bazzolli, Galvão Antônio do Nascimento, Poliane Alfenas, Maria Cristina Dantas Vanetti e Hilário Cuquetto Mantovani, pela amizade, incentivo e contribuição para minha formação científica.

Às professoras Cynthia Canedo da Silva e Monique Renon Eller por terem sido sempre tão solícitas em todos os momentos que precisei, pelas dicas e por dividirem comigo a empolgação a cada etapa concluída. E ao Dr. Júlio César Câmara Rosa, amigo e inteligente pesquisador, por ter se prontificado a participar da banca.

Aos funcionários do DMB, Danilo (Danilim), Hugo, Sr. Paulo, Letícia, Sandra, Evandro e Emília, pelo carinho, amizade e prontidão em ajudar. Sou muito feliz e grata por conhecer vocês.

À secretária da Pós-graduação Nilcéia, por toda atenção, prontidão em ajudar e resolver tudo, pelo carinho e simpatia, risadas compartilhadas e por fazer do telefone 2553 inesquecível para mim (“Miiicrobiologia Agrícola, Nilcéa”).

Ao meu amado, esposo, amigo, companheiro e pai do meu futuro filho, Marcelo, pelo incentivo, conselhos, pela dedicação em todos os momentos não só comigo, mas com o meu trabalho! Pela paciência e prontidão em ir me ver todos os finais de semana em Viçosa e por dividir comigo mais essa conquista.

À minha mãe Eunice, pelas orações, pela torcida, e por ter me incentivado em todos os momentos. Por ter me ensinado os caminhos do Senhor. Ao meu pai, Robson, pelo incentivo e orgulho por ter uma filha tão estudada. Aos meus irmãos Victor, Priscila e Gina, cunhados Sara, Juninho e Denise, e minhas amadas sobrinhas Catarina e Milena, pelo imenso amor, por acreditarem e torcerem por mim.

Aos amigos do Laboratório de Fisiologia: Mari, Josi, Deisy, Fernando (Fefê), Hugo, Éder, Robson, Lílian, Raquel, Breno, Maria Fernanda, Juan, Amanda, Fábio, Nívea e Juliana, por fazerem com que todos os dias de trabalho fossem sempre com sorriso no rosto. Vocês foram essenciais e sou grata a cada um pela colaboração e pela ótima convivência! Em especial agradeço à Deisy, minha querida estagiária, pela ajuda, disposição e prontidão, e por me fazer crescer ao vê-la crescer profissionalmente.

Ao meu grupo de convivência (GC) da Igreja Presbiteriana de Viçosa, por permitirem que minha estadia em Viçosa não resultasse apenas em crescimento profissional, mas espiritual também! Estudar a Palavra, orar com vocês, e criar amizades tão sinceras são detalhes da minha vida em Viçosa que jamais esquecerei.

Aos amigos dos Laboratórios de Microbiologia Industrial, Genética de Microrganismos, Café, Alimentos, Anaeróbios, Petróleo e Micorrizas, o meu muito obrigada pela ajuda, atenção, por compartilharem o conhecimento comigo e fazer meus dias no trabalho mais felizes.

Aos meus amigos Caio, Monalessa, Mari, Josi, Danilo (Danilim), Natan, Lygia, Thamy, Fernando, Ciro, Izadora, Carol, Tati, Gina, Meiry, Sara, Sabeline, Fernando, Sr. Lucas e D. Ivone, por me apoiarem, por sofrerem junto comigo, por se alegrarem comigo, por me incentivarem a não desistir, e pelo simples fato de saber que somos verdadeiros amigos.

A todos que, direta ou indiretamente, participaram desse trabalho e de minha formação.

ÍNDICE

RESUMO.....	ix
ABSTRACT	xi
INTRODUÇÃO GERAL.....	1
CAPÍTULO 1.....	9
Applying functional metagenomics to search for novel lignocellulosic enzymes in a microbial consortium derived from a thermophilic composting phase of sugarcane bagasse and cow manure.....	9
ABSTRACT.....	10
INTRODUCTION.....	10
MATERIALS AND METHODS.....	12
Total DNA Extraction and Metagenomic Fosmid Library Construction.....	12
Functional Screening and Enzymatic assays	13
Sequence and Phylogenetic Analysis of Fosmid DNA.....	14
Nucleotide Sequence Accession Numbers.....	15
RESULTS	17
Metagenomic Library Screening for Clones Expressing Lignocellulolytic Activity and General Characteristics.....	17
Fosmid sequence, phylogenetic and functional analyses	17
Contig analysis of clones associated with carbohydrate hydrolysis	19
DISCUSSION.....	27
ACKNOWLEDGMENTS.....	33
REFERENCES	34
CAPÍTULO 2.....	40
Construção de um sistema de expressão para <i>Kluyveromyces marxianus</i> CCT 7735 (UFV-3): Obtenção de linhagem hospedeira <i>ura3⁻</i> e de vetor integrativo para clonagem e transformação.....	40
RESUMO	41
INTRODUÇÃO.....	41

MATERIAL E MÉTODOS.....	44
Micro-organismos, condições de cultivo e manutenção.....	44
Construção dos cassetes para inativação do gene <i>URA3</i>	45
Confirmação das linhagens de <i>K. marxianus ura3⁻</i>	48
Análise da estabilidade mitótica das linhagens recombinantes	48
Construção do vetor de expressão pKmLPG	48
RESULTADOS E DISCUSSÃO.....	50
Deleção do gene <i>URA3</i> da levedura <i>K. marxianus</i> CCT 7735 (UFV-3) e isolamento de mutantes <i>ura3⁻</i>	51
Construção do vetor de expressão pKmLPG	57
CONCLUSÕES E PERSPECTIVAS.....	60
REFERÊNCIAS BIBLIOGRÁFICAS	62
CAPÍTULO 3.....	65
Expressão heteróloga de genes de β -glicosidases de fontes microbianas termofílicas na levedura <i>Kluyveromyces marxianus</i> CCT 7735 (UFV-3)	65
RESUMO	66
INTRODUÇÃO	67
MATERIAL E MÉTODOS.....	68
Análise <i>in silico</i> dos genes de β -glicosidase e predição das proteínas.....	68
Linhagens, meios e condições de crescimento	69
Construção dos cassetes de expressão	69
Transformação da levedura.....	70
Expressão e atividade de β -glicosidase	71
Análise Estatística	72
RESULTADOS E DISCUSSÃO.....	72
Amplificação dos genes de β -glicosidase e análise da estrutura das proteínas	72

Transformação de <i>K. marxianus</i> CCT 7735 (UFV-3) com os cassetes de expressão pKMCL/P3Gli, pKMCL/P8Gli e pKMCL/P9Gli, e seleção de linhagens recombinantes	78
Análise da expressão dos genes P3Gli, P8Gli e P9Gli pela atividade de β -glicosidase	82
CONCLUSÕES	92
REFERÊNCIAS BIBLIOGRÁFICAS	94
CONCLUSÕES GERAIS	99
PERSPECTIVAS PARA TRABALHOS FUTUROS	102

RESUMO

COLOMBO, Livia Tavares de, D.Sc., Universidade Federal de Viçosa, março de 2015. **Metagenômica de um consórcio microbiano termofílico na detecção de genes de enzimas celulolíticas e expressão heteróloga em *Kluyveromyces marxianus* CCT 7735 (UFV-3)**. Orientadora: Flávia Maria Lopes Passos. Coorientadores: Marisa Vieira de Queiroz e Wendel Batista da Silveira.

Neste trabalho foi construída uma biblioteca metagenômica a partir de um consórcio microbiano termofílico com aproximadamente 135.000 clones, abrigando cerca de 3,5 Gpb de DNA metagenômico. Essa biblioteca foi construída com o objetivo de identificar genes que codificam enzimas lignocelulolíticas para serem expressos na levedura *Kluyveromyces marxianus* CCT 7735 (UFV-3). De 6.720 clones analisados (5 % do total), foram obtidos 182 *hits* positivos para atividades de celulases, xilanases e β -glicosidases, com seleção e sequenciamento de 30 fosmídeos que deram origem à identificação de 34 ORFs codificando para enzimas glicosil hidrolases, relacionadas com a hidrólise de celulose e hemicelulose. Paralelo à construção da biblioteca, foi obtido uma linhagem de *K. marxianus* CCT 7735 (UFV-3) mutante (*ura3⁻*). Pela primeira vez nesta levedura a deleção deste gene foi realizada utilizando-se a técnica *split-marker*, com seleção dos mutantes primeiramente em placas contendo geneticina e, posteriormente, associados à seleção em placas contendo ácido 5-fluorótico (5-FOA). Os resultados apresentados mostram que esta técnica de deleção pode ser eficientemente empregada para *K. marxianus* CCT 7735 (UFV-3), mesmo levando-se em conta a diploidia desta linhagem. A partir da obtenção do mutante *ura3⁻*, deu-se a construção do vetor pKmLPG para ser usado como sistema de expressão para essa levedura, uma vez que a marca de seleção contida neste vetor é o gene *URA3*. A característica diferencial e promissora deste sistema de expressão é o uso da sequência do promotor do gene que codifica a enzima endo-polygalacturonase endógena de *K. marxianus* CCT 7735 (UFV-3). A confirmação do vetor pKmLPG será realizada por sequenciamento. Sem essa confirmação, utilizou-se então o vetor de expressão pKMCL para clonagem de três genes de β -glicosidase isolados a partir da biblioteca fosmidial na levedura *K. marxianus* CCT 7735 (UFV-3). O vetor pKMCL é derivado do vetor de expressão pKLAC2 de *K. lactis* e utiliza a marca de seleção que confere resistência ao antibiótico geneticina. Por realização de teste enzimático dos transformantes em meio sólido, observou-se a presença de atividade extracelular de β -glicosidase endógena de *K. marxianus* CCT 7735 (UFV-3). Com o genoma desta levedura anotado, foi possível identificar o gene, e

juntamente com os demais genes de β -glicosidase (P3Gli, P8Gli e P9Gli), foram preditas as estruturas tri-dimensionais dessas proteínas. A confirmação dos transformantes para os genes P3Gli, P8Gli e P9Gli foi realizada por PCR, e atividade de β -glicosidase foi detectada tanto no sobrenadante extracelular quanto na região do periplasma em oito das nove cepas transformantes analisadas. Os resultados de expressão heteróloga de β -glicosidases em *K. marxianus* CCT 7735 (UFV-3) mostram que esta linhagem pode ser usada como hospedeira na expressão de diferentes genes lignocelulolíticos, com expectativa de uso das linhagens recombinantes capazes de secretar celulasas visando a produção de etanol celulósico.

ABSTRACT

COLOMBO, Livia Tavares de, D.Sc., Universidade Federal de Viçosa, March, 2015. **Thermophilic microbial consortium metagenomic for detection of genes coding cellulolytic enzymes and heterologous expression in *Kluyveromyces marxianus* CCT 7735 (UFV-3)**. Adviser: Flávia Maria Lopes Passos. Co-advisers: Marisa Vieira de Queiroz and Wendel Batista da Silveira.

A metagenomic library harboring around 3,5 Gpb was constructed from a thermophilic microbial consortium. The aim of this library was to identify genes coding lignocellulolytic enzymes that could be expressed in the yeast *Kluyveromyces marxianus* CCT 7735 (UFV-3). In total, 6,720 clones (5% of all clones) were analysed and 182 positive hits were found for cellulases, xylanases and β -glucosidases. The sequencing of 30 selected fosmids resulted in the identification of 34 putative open reading frames (ORFs) coding glycoside hydrolases involved with cellulose and hemicellulose degradation. Simultaneously, a defective mutant strain of *K. marxianus* CCT 7735 (UFV-3) for gene *ura3* was obtained. The gene was deleted by using, for the first time in this yeast, the split-marker technique that consisted in a first selection of mutants in solid medium with the antibiotic geneticin followed by a second selection in the presence of 5-Fluoroorotic Acid. The results presented here demonstrated the efficiency of the split-marker technique in *K. marxianus* CCT 7735 (UFV-3), even though this is a diploid strain. Once the *ura3*⁻ was obtained, the vector pKmLPG was constructed using the *URA3* as a marker gene in order to have an expression system for the *ura3*⁻ mutant strain. The differential and promising feature of this vector lies in the presence of the promoter sequence of the gene coding an endo-polygalacturonase enzyme from *K. marxianus* CCT 7735 (UFV-3). The confirmation of the correct vector construction will be performed by sequencing. In order to express genes isolated from the metagenomic library in the yeast *K. marxianus* CCT 7735 (UFV-3) three genes coding β -glucosidase were cloned onto the vector pKMCL. This vector is derived from pKLAC2 of *Kluyveromyces lactis* and it uses the resistance gene for the antibiotic geneticin as a marker. Extracellular activity of endogenous β -glucosidase from *K. marxianus* CCT 7735 (UFV-3) was detected by enzymatic assay in solid medium during the analyses of transformants, and based in the annotated genome of this yeast, was possible to identify the β -glucosidase gene. The three dimensional structure of the endogenous protein as well from those codified by the β -glucosidase genes (P3Gli, P8Gli and P9Gli) were predicted. The confirmation of transformants for the cloned genes P3Gli, P8Gli and P9Gli was done

by PCR, and the β -glucosidase activity was detected in both extracellular supernatant and periplasm region in eight of nine transformants analysed. Results of heterologous expression of β -glucosidase in *K. marxianus* CCT 7735 (UFV-3) demonstrated that this strain can be used as a host for expression of various lignocellulolytic genes, with the expectation to use recombinant yeast strains capable to secret celulasas targeting the production of cellulosic etanol.

INTRODUÇÃO GERAL

O etanol tem recebido grande atenção no cenário nacional e internacional por ser alternativa aos combustíveis fósseis que são poluidores e não renováveis. Os biocombustíveis denominados de “primeira geração” produzidos comercialmente, como o álcool combustível, são alvos de severas críticas no cenário internacional por competir com a produção de alimentos (Naik et al., 2010). Pesquisas vêm sendo realizadas para viabilizar a produção de biocombustíveis de “segunda geração”, que utilizam como matéria-prima a biomassa lignocelulósica, um recurso abundante e de baixo custo. O uso de material lignocelulósico para produção de biocombustíveis pode aumentar a oferta de etanol sem aumentar a área plantada (Naik et al., 2010; Sims et al., 2010), característica crucial para o Brasil consolidar sua posição de destaque no mercado de biocombustíveis. O país já é o maior produtor de etanol proveniente de cana de açúcar, e a integração da produção de açúcar e etanol na mesma planta pode auxiliar a trazer maior competitividade ao etanol brasileiro.

O Estado de Minas Gerais é atualmente o terceiro maior produtor de etanol no país (Unica, 2013) e, portanto, grandes quantidades de resíduos vegetais, dentre os quais o bagaço da cana-de-açúcar, são gerados durante o processo de produção do etanol de primeira geração. Estima-se que se 90% dos açúcares fermentescíveis encontrados no bagaço pudessem ser recuperados e fermentados, a produção de etanol no Brasil aumentaria 50% (Camargo, 2005). Contudo, até o momento a produção desse tipo de combustível não é economicamente viável e muitos obstáculos precisam ser sobrepostos para que sua relação custo-benefício seja favorável.

A partir de cada tonelada de cana são produzidos aproximadamente 140 kg de bagaço seco. Um dos grandes desafios da produção de etanol celulósico é a disponibilização dos açúcares para a fermentação, uma vez que a biomassa lignocelulósica que forma o bagaço é constituída por moléculas de celulose, hemicelulose, e lignina, que se encontram organizadas em uma estrutura complexa com elevada resistência estrutural (Rubin, 2008) (Figura 1). A conversão dessa biomassa em etanol envolve três etapas, sendo a primeira delas o pré-tratamento físico, químico, ou biológico do substrato, que é essencial para expor a celulose e hemicelulose à hidrólise catalisada por celulases e hemicelulases (segunda etapa). Na terceira etapa, glicose e xilose, os principais açúcares liberados após a hidrólise enzimática, são convertidos em etanol por leveduras (Camassola & Dillon, 2009; Jeffries & Kurtzman, 1994; Matsushika et al., 2009; Sims et al., 2010). Dentre os requisitos necessários para viabilizar economicamente a produção do etanol celulósico em escala industrial destaca-se a obtenção de enzimas hidrolíticas de

baixo custo (Chen, 2011), pois a conversão enzimática da celulose e hemicelulose em açúcares fermentecíveis não gera subprodutos inibidores da fermentação (Oliva et al., 2006).

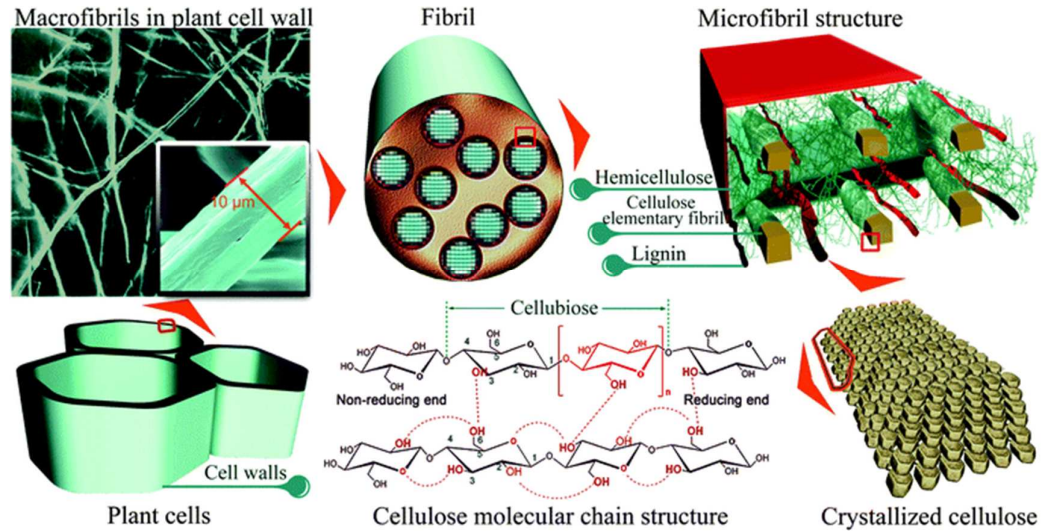


Figura 1. Estrutura da biomassa lignocelulósica. Fonte: Lu et al., 2012.

Embora existam métodos de pré-tratamento da biomassa para facilitar o processo de sacarificação, o alto custo, a baixa atividade específica dos coquetéis enzimáticos disponíveis e atualmente usados, e o número limitado de empresas produzindo enzimas para esse fim, enfraquecem a produção mundial e aumentam os custos de produção de bioetanol (Del Pozo et al., 2012; Kovacs et al., 2009).

Souza (2012), em trabalho desenvolvido no Departamento de Microbiologia da UFV, selecionou um inóculo microbiano termofílico com atividades de celulasas e xilanasas obtido a partir de comunidades microbianas provenientes de ambientes onde a matéria lignocelulósica é decomposta na natureza, tais como esterco e bagaço de cana-de-açúcar em decomposição. Culturas mistas são úteis pois podem simular condições relevantes encontradas na natureza porque uma parcela significativa dos micro-organismos ali presentes ainda não são cultiváveis em meios de cultura. Além disso, a partir dessas comunidades podem ser selecionados genes que codificam proteínas de interesse industrial por metagenômica.

Metagenômica é um tipo de estratégia de clonagem do metagenoma (DNA total de micro-organismos presentes em uma determinada amostra) que permite acessar e analisar o genoma de comunidades microbianas de uma amostra ambiental independente de cultivo (Handelsman et al., 1998; Schloss & Handelsman, 2003). Desde o ano 2000, inúmeros trabalhos científicos publicados

confirmam o quanto a metagenômica revolucionou a microbiologia ao abrir caminho para estudos e avaliações de comunidades microbianas presentes em ecossistemas complexos. Mas a atração ligada à exploração de comunidades microbianas complexas não está relacionada apenas à compreensão da filogenia, ecologia e perfis metabólicos dessas comunidades. Genomas de espécies não-cultiváveis constituem também um reservatório inexplorado de genes envolvidos na produção de novas biomoléculas, e as pesquisas demonstram o potencial da metagenômica em prospectar novos genes e biocatalisadores (Uchiyama & Miyazaki, 2009).

Por metagenômica funcional (pelo emprego de triagem funcional de bibliotecas metagenômicas) é possível analisar e identificar clones que expressam uma função desejada, como uma atividade enzimática de interesse, seguida pela caracterização dos clones positivos por sequenciamento e análises bioquímicas (Schloss & Handelsman, 2003). A principal vantagem da estratégia de triagem funcional é não necessitar que os genes de interesse sejam reconhecidos por análise de sequência, com um DNA consenso, possibilitando o potencial de identificar novas classes de genes para funções novas e/ou funções já conhecidas (Handelsman, 2004). As análises funcionais têm identificado novos genes e produtos gênicos, como por exemplo, novas enzimas celulolíticas, a exemplo de novas β -glicosidases (Del Pozo et al., 2012), endo-glucanase e mananase (Duan et al., 2009), e xilose isomerase (Parachin & Gorwa-Grauslund, 2011). No entanto, o maior desafio para a aplicação industrial desses novos genes de interesse é a obtenção de um sistema biológico capaz de sintetizar e secretar essas enzimas de forma eficiente e abundante. Neste contexto, as leveduras têm propriedades industriais que justificam sua utilização como hospedeiras para genes heterólogos e síntese de proteínas recombinantes. Essas características incluem o fato de a levedura ser eucarioto unicelular de rápido crescimento, fácil manipulação genética, e ser capaz de realizar modificações pós-traducionais, traços importantes para expressão de genes eucarióticos.

A levedura *Kluyveromyces marxianus* desponta no cenário biotecnológico por ser uma levedura capaz de fermentar em temperaturas elevadas quando cultivada em altas concentrações de substrato e baixas concentrações de oxigênio (Silveira et al., 2005; Fonseca et al., 2008), o que favorece seu uso em processos industriais para produção de bioetanol sem requerer gastos de energia com resfriamentos e aeração. Além de crescimento em elevadas temperaturas, esta levedura é também capaz de utilizar uma variedade de fontes de carbono, apresenta alto potencial para secreção de proteínas, e menor tempo de geração

conhecido para um organismo eucarioto (Fonseca et al., 2008; Groeneveld et al., 2009; Lane & Morrissey, 2010). *K. marxianus* está entre as espécies de leveduras consideradas para produção de proteínas recombinantes e tem se mostrado uma hospedeira atrativa para o uso como sistema de expressão, com estudos cada vez mais direcionados para conhecer melhor os mecanismos de transformação para esta levedura. Mas, apesar de suas propriedades, ainda são poucos os trabalhos utilizando-a como hospedeira na síntese de biomoléculas de interesse industrial, e muito conhecimento a ser descoberto acerca de sua fisiologia e manipulação genética.

Desde 2005 a linhagem *K. marxianus* CCT 7735 (UFV-3) (Silveira et al., 2005) tem sido utilizada no Laboratório de Fisiologia de Micro-organismos do Departamento de Microbiologia da Universidade Federal de Viçosa em estudos fisiológicos e de produção de etanol (Diniz et al., 2012; Santos et al., 2012; Diniz et al., 2014). Tendo em vista o interesse em aplicações biotecnológicas, o genoma desta linhagem foi recentemente sequenciado e anotado (Silveira et al., 2014), o que contribui para um maior conhecimento e compreensão da fisiologia e metabolismo desta levedura. Esta linhagem também tem se mostrado útil para a expressão de proteínas heterólogas, como proteínas estruturais do vírus da dengue (Bragança et al., 2014) e lacase de actinobactéria (estudo em andamento).

No presente trabalho, nós construímos uma biblioteca metagenômica a partir do consórcio microbiano termofílico selecionado por Souza (2012), com o objetivo de isolar novas enzimas lignocelulolíticas envolvidas na sacarificação de biomassa lignocelulósica para serem expressas na levedura *K. marxianus* UFV-3.

Esta tese de doutorado está dividida em capítulos, onde se apresentam as diferentes etapas do trabalho na forma de artigo científico contendo resumo, uma breve introdução, material e métodos, e resultados e discussão. No primeiro capítulo está descrito a construção, triagem e análise da biblioteca metagenômica. No segundo capítulo estão descritos a obtenção do mutante de *K. marxianus* CCT 7735 (UFV-3) *ura3⁻* e a construção do vetor de expressão pKmLPG. E o terceiro capítulo contempla os resultados obtidos pela expressão de três genes selecionados a partir da biblioteca metagenômica na levedura *K. marxianus* CCT 7735 (UFV-3).

REFERÊNCIAS BIBLIOGRÁFICAS

1. Bragança, C.R.S., L.T. Colombo, A.S. Roberti, M.C.T. Alvim, S.A. Cardoso, K.C.P. Reis, S.O. de Paula, W.B. Silveira & F.M.L. Passos. 2014. Construction of recombinant *Kluyveromyces marxianus* UFV-3 to express dengue virus type 1 nonstructural protein 1 (NS1). *Applied Microbiology and Biotechnology*. 99:1191-203.
2. Camargo, P.D. 2005. Força verde: um novo campo para a indústria química. *Revista Brasileira de Engenharia Química*. 18-21.
3. Camassola, M. & A.J.P. Dillon. 2009. Biological pretreatment of sugar cane bagasse for the production of cellulases and xylanases by *Penicillium echinulatum*. *Industrial Crops and Products*. 29:642-647.
4. Chen, Y. 2011. Development and application of co-culture for ethanol production by co-fermentation of glucose and xylose: a systematic review. *Journal of Industrial Microbiology and Biotechnology*. 38:581-597.
5. Del Pozo, M.V., Fernández-Arrojo L., Gil-Martínez J., Montesinos A., Chernikova T.N., Nechitaylo T.Y., Waliszek A., Tortajada M., Rojas A., Huws S.A., Golyshina O.V., Newbold C.J., Polaina J., Ferre M., Golyshin P.N. 2012. Microbial β -glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. *Biotechnol Biofuels* 5:1-13.
6. Diniz, R.H.S., M.Q.R.B. Rodrigues, L.G. Fietto, F.M.L. Passos & W.B. Silveira. 2014. Optimizing and validating the production of ethanol from cheese whey permeate by *Kluyveromyces marxianus* UFV-3. *Biocatalysis and Agricultural Biotechnology*. 3:111-117.
7. Diniz, R.H.S., W.B. Silveira, L.G. Fietto & F.M.L. Passos. 2012. The high fermentative metabolism of *Kluyveromyces marxianus* UFV-3 relies on the increased expression of key lactose metabolic enzymes. *Antonie van Leeuwenhoek (Gedruckt)*. 101:541-550.
8. Duan, C.-J., L. Xian, G.-C. Zhao, Y. Feng, H. Pang, X.-L. Bai, J.-L. Tang, Q.-S. Ma & J.-X. Feng. 2009. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *Journal of Applied Microbiology*. 107:245-256.
9. Fonseca, G.G., E. Heinzle, C. Wittmann, A.K. Gombert. 2008. The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Applied Microbiology and Biotechnology*. 79:339-54.

10. Groeneveld, P., A.H. Stouthamer & H.V. Westerhoff. 2009. Super life--how and why "cell selection" leads to the fastest-growing eukaryote. *The FEBS journal*. 276:254-70.
11. Handelsman, J., M.R. Rondon, S.F. Brady, J. Clardy & R.M. Goodman. 1998. *Chemical Biology*. 5:R245-9.
12. Handelsman, J. 2004. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*. 68:669-685.
13. Jeffries, T.W. & C.P.Kurtsman. 1994. Strain selection, taxonomy and genetic of xylose fermenting yeasts. *Enzyme and Microbiol Technology*. 6:922-932.
14. Kovacs, K., S. Macrelli, G. Szakacs & G. Zacchi. 2009. Enzymatic hydrolysis of steampretreated lignocellulosic materials with *Trichoderma atroviride* enzymes produced in-house. *Biotechnology for Biofuels*. 2:1-14.
15. Lane, M.M. & J.P. Morrissey. 2010. *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. *Fungal Biology Reviews*. 24:17-26.
16. Lu, Y., Yun L., Qingfeng S., Dongjiang Y., Xilin S., Xiangdong Y., Guangshan Z., Yixing L., Huijun Z. & Jian L. 2012. Fabrication of mesoporous lignocellulose aerogels from wood via cyclic liquid nitrogen freezing-thawing in ionic liquid solution. *Journal of Materials Chemistry*. 22:13548-13557.
17. Macedo, I.C. 2007. The current situation and prospects for ethanol. *Estudos Avançados*. 21:157-165.
18. Matsushika, A., H. Inoue, T. Kodaki, T. & S. Sawayama. 2009. Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: current state and perspectives. *Applied Microbiology and Biotechnology*. 84:37-53.
19. Naik, S.N., V.V. Goud, P.K. Rout & A.K. Dalai. 2010. Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews*. 14:578-597.
20. Oliva, J.M., M.J. Negro, F. Sáez, I. Ballesteros, P. Manzanares, A. González & M. Ballesteros. 2006. Effects of acetic acid, furfural and catechol combinations on ethanol fermentation of *Kluyveromyces marxianus*. *Process Biochemistry*. 41:1223-1228.
21. Parachin, N.S. & M.F. Gorwa-Grauslund. 2011. Isolation of xylose isomerases by sequence- and function-based screening from a soil meTagenomic library. *Biotechnol Biofuels*. 4:1-10.
22. Rubin, E.M. 2008. Genomics of cellulosic biofuels. *Nature Reviews*. 454:841-845.

23. Santos, V.C., C.R.S. Bragança, F.J.V. Passos & F.M.L. Passos. 2012. Kinetics of growth and ethanol formation from a mix of glucose/xylose substrate by *Kluyveromyces marxianus* UFV-3. *Antonie van Leeuwenhoek (Gedrukt)*. 103:153-161.
24. Schloss, P.D. & J. Handelsman. 2003. Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology*. 14:303-310.
25. Silveira, W.B., Diniz, R.H.S., Cerdan, M.E., Gonzalez-Siso, M.I., et al. 2014. Genomic Sequence of the Yeast *Kluyveromyces marxianus* CCT 7735 (UFV-3), a Highly Lactose-Fermenting Yeast Isolated from the Brazilian Dairy Industry. *Genome Announcements*. 2:e01136-14-e01136-14.
26. Silveira, W.B., F.J.V. Passos, H.C. Mantovani & F.M.L. Passos. 2005. Ethanol production from cheese whey permeate by *Kluyveromyces marxianus* UFV-3: a flux analysis of oxido-reductive metabolism as a function of lactose concentration and oxygen levels. *Enzyme and Microbial Technology*. 36:930-6.
27. Sims, R.E.H., W. Mabee, J.N. Saddler & M. Taylor. 2010. An overview of second generation biofuel technologies. *Bioresource Technology*. 101:1570-1580.
28. Souza, R.A. Obtenção de Inoculante e de Coquetel Enzimático Lignocelulolítico a partir de Comunidades Microbianas Termofílicas. 2012. 56f. Dissertação (mestrado). Universidade Federal de Viçosa, Viçosa, MG, 2012.
29. Uchiyama, T. & K. Miyazaki. 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* 20:616-622.
30. UNICA. União da Indústria de Cana-de-açúcar. Disponível em: <<http://www.unica.com.br>>. Acesso em: 13 mar/2015.

CAPÍTULO 1

APPLYING FUNCTIONAL METAGENOMICS TO SEARCH FOR NOVEL LIGNOCELLULOSIC ENZYMES IN A MICROBIAL CONSORTIUM DERIVED FROM A THERMOPHILIC COMPOSTING PHASE OF SUGARCANE BAGASSE AND COW MANURE

ABSTRACT

Discovering new hydrolytic enzymes from environments where lignocellulosic biomass is naturally decomposed is an ideal strategy to overcome one of the most significant barriers in the saccharification process to efficient second-generation ethanol production: the high cost of enzymatic cocktails. Aiming to identify novel carbohydrate-depleting enzymes, we applied functional metagenomic analysis to a laboratory subculture of a microbial community obtained from two natural environments: a mix of sugarcane bagasse and cow manure in the thermophilic composting phase. A fosmid library was constructed containing 135,000 clones, and 182 clones were selected regarding the ability to hydrolyze carbohydrate among almost 7,000 clones tested. The DNA of 30 fosmids was sequenced, generating 67 contigs whose genes related to biomass degradation were distributed in 12 contigs encoding 34 putative carbohydrate-active enzymes belonging to 17 glycosyl hydrolase (GH) families. One third of the putative proteins belong to the GH3 family, which includes β -glucosidase enzymes known to be important in the cellulose-deconstruction process but present with low activity in commercial enzyme preparations. Phylogenetic analysis of the amino acid sequences of seven selected proteins, including three β -glucosidases, showed very low identity with protein sequences deposited in databases. Our findings highlight microbial consortium obtained from a mixture of decomposing biomass residues, such as sugar cane bagasse and cow manure, as a rich resource of novel enzymes potentially useful in biotechnology for saccharification of lignocellulosic substrate.

INTRODUCTION

Sugar cane bagasse is one of the most abundant agro-industry residues in Brazil and its high carbohydrate content could be used as a feedstock for second-generation ethanol production in biorefineries. It is estimated that, if nearly 90% of the fermentable sugars in this feedstock could be recovered and fermented, the Brazilian ethanol production would increase by 50% (1,2). However, the use of sugarcane bagasse or other sources of lignocellulosic biomass is still not economically viable for efficient ethanol fuel production. Among the four stages of ethanol production from lignocellulosic waste - pretreatment, enzymatic hydrolysis, fermentation, and distillation (3,4) - the step involving enzymatic hydrolysis to release soluble sugars (5) remains cost prohibitive because of the large amount of

enzymes necessary and the low activity of the enzymatic cocktails currently available (6).

The identification of new hydrolytic enzymes or genetic improvement of existing ones could help lower the costs of enzyme production for biomass degradation, and there is considerable efforts now being directed toward improving cellulosic ethanol production by these and other means (7-12).

Metagenomics is a promising approach for functional analysis of the microbial genomes present in natural environments, especially the genomes of uncultured microbial species that represent an unexplored reservoir of new biomolecules (13). Several studies have shown how metagenomic libraries provide an almost limitless pool of new genes encoding biotechnology products of interest (14-18). Among the diverse array of non-cultivable microorganisms there likely exist novel cellulases and hemicellulases with better catalytic performance, exhibiting higher affinity and activity toward cellulosic substrates and having desirable traits for industrial operational conditions such as high activity at temperatures of 45-55 °C. The metagenomic strategy is important for mining new genes from untapped feedstock sources aiming to increase the final cellulose and hemicellulose hydrolysis yield.

The successful discovery of biomass-degrading genes from various environmental samples, including cow rumen (19), compost soils (10) and termite gut (20), encouraged us to investigate complex and unexploited microbial communities obtained by mixing different samples enriched for microorganisms capable of degrading cellulose. In addition, the screening of microbial genomes in their natural environments can accelerate the discovery of new genes. Thus, here we evaluated the possibility of isolating genes of carbohydrate catabolic enzymes from a metagenomic library constructed from a microbial consortium derived from naturally degrading sugarcane bagasse associated with cow manure in a thermophilic composting phase. Functional screening was employed and led to the identification and phylogenetic analysis of seven novel putative protein sequences, among which we found β -glucosidase and α -xylosidase that are essential to enzymatic conversion of lignocellulosic biomass to fermentable sugars for bioethanol production.

MATERIALS AND METHODS

Total DNA Extraction and Metagenomic Fosmid Library Construction

Sample used in this work as source of metagenomic DNA was obtained by Souza (21). Briefly, samples of decomposing sugarcane bagasse and cow manure were usually inoculated in flasks containing cellulose-peptone solution with sugarcane bagasse at a ratio of 1 % (w/v), pH 7.7, a strip of Whatman filter paper n. 1 as an indicator of cellulase activity and incubated at 55 °C under static aerobic conditions (21). When degradation of the filter paper strip was detected, 1.0 mL of suspension was transferred into the same fresh medium. This strategy favored presence of lignocellulosic activity.

Total DNA from a microbial consortium was extracted using a protocol described by Stevenson and Weimer (22). The cells were lysed using beads and SDS (20% w/v), and DNA was isolated by purification using phenol/chloroform extraction followed by alcohol precipitation. The pellet of DNA was resuspended in TE (10 mM Tris/HCl, 1 mM EDTA, pH 8.0), treated with RNase at 37 °C for two hours and stored at -20 °C.

Cloning of metagenomic DNA into the vector Fosmid pCC2FOS (Epicentre®, Madison, WI, USA) was performed according to the manufacturer's instructions. Briefly, the metagenomic DNA was size selected by preparative pulsed field (Pulsed-field CHEF DRIII System – Bio-Rad, Hercules, CA, USA) gel electrophoresis at an angle of 120°, 6 V.cm⁻¹, 0.5 s – 0.5 s switch time, 5 h at 14 °C. DNA fragments of approximately 40 kb were excised from the gel and their ends repaired using the End-Repair Enzyme Mix before ligation into the vector. CopyControl fosmids containing the inserts were packaged with MaxPlax Lambda Packaging Extract and used to infect the *Escherichia coli* EPI300-T1® plating strain. Infected cells were spread on Luria-Bertani (LB) plates [1% (w/v) bactotryptone; 0.5% (w/v) yeast extract; 0.5% (w/v) NaCl] supplemented with 12.5 µg.mL⁻¹ chloramphenicol and incubated at 37 °C overnight to select for the CopyControl Fosmid clones. Clones were transferred to 96-well plates containing LB medium and stored after growth at -80 °C in the presence of glycerol [20% (v/v)].

Estimation of the size of the metagenomic library was carried out with ten randomly selected clones. The fosmid DNA from each clone was extracted using a Wizard Plus DNA Purification kit (Promega, Madison, WI, USA), and digested using 10 U *NotI* restriction enzyme (Promega, Madison, WI, USA) at 37 °C overnight. The restriction fragments of the fosmid clones were separated by preparative pulsed field

gel electrophoresis at an angle of 120° ; 6 V.cm^{-1} , 0.5 s – 0.5 s switch time, 10.5 h at 14°C .

Functional Screening and Enzymatic assays

Clones from the metagenomic library were pre-cultured in 96-well microplates containing $150 \mu\text{L}$ of liquid LB medium supplemented with chloramphenicol ($12.5 \mu\text{g.mL}^{-1}$) and incubated at 37°C on a rotary shaker (200 rpm) for 16 h. After growth, the clones were plated on different media for lignocellulose degradation activity screening using a replicator. All functional screening in solid medium was carried out in triplicate.

Screening for cellulase activity was performed using the method described by Kasana et al. (23) on agar CMC medium [0.2% (w/P) NaNO_3 , 0.1% (w/v) K_2HPO_4 , 0.05% (w/v) MgSO_4 , 0.05% (w/v) KCl, 0.2% (w/v) carboxymethylcellulose sodium, 0.02% (w/v) peptone and 1.7% (w/v) agar]. Three microliters of culture were transferred to agar CMC plates using a replicator. Following a 48-h incubation, Gram's iodine solution [6.6% (w/v) KCl and 3.33% (w/v) iodine) was spread on the plate for three to five minutes to visualize the results. Formation of a clear zone around the colony indicated the presence of cellulolytic enzymes.

Screening for xylanases was carried out according to the method described by Teather and Wood (24), with modifications. Cells were inoculated on plates containing 1.5% agar (w/v) and 0.1% xylan (w/v), followed by incubation at 37°C for 48 h, after which the plates were stained with a solution of Congo Red [0.1% (w/v)] for 20 min and rinsed with 1 M NaCl for 20 min. Positive clones were identified by a clear zone around colonies.

Screening for β -glucosidases was carried out according to two methods. In the first, described by Eberhart et al. (25), clones were cultured on LB agar plates containing 0.2% esculin (w/v), 0.05% ferric ammonium citrate (w/v), and $12.5 \mu\text{g.mL}^{-1}$ chloramphenicol at 37°C overnight. Those colonies forming clear black halos were selected as positive colonies. In the second method, described by Del Pozo et al. (6), clones were cultured on LB agar plates containing $12.5 \mu\text{g.mL}^{-1}$ chloramphenicol at 37°C overnight, after which the plates were covered with an agar buffer substrate solution [40 mL of 50 mM $\text{C}_2\text{H}_3\text{NaO}_2$, pH 5.6, 0.4% (w/v) agarose and 5 mg.mL^{-1} of pNP β G as substrates]. Positive clones were identified by formation of a yellow color.

Enzymatic activities from positive clones identified in the functional screening were confirmed by enzymatic assays for CMCase, β -glucosidase and xylanase. CMCase activity was analyzed by the DNS (3,5-dinitrosalicylic acid) method that measures sugars released during enzymatic hydrolysis of cellulose, as described by

Miller (26), with the reaction mix adapted from Ghose (27). Xylanase activity was also spectrophotometrically measured by the DNS method using a protocol from Bailey et al. (28). In both cases, one unit of enzyme activity was defined as the amount of enzyme necessary to release one μmol of reduced sugar per gram of protein per minute under assay conditions.

β -glucosidase activity was determined according to Chen et al. (29). One unit of β -glucosidase activity was defined as the amount of enzyme required to hydrolyze 1 μmol of substrate per minute, under assay conditions, per milligram of protein. Total protein concentration was determined according to Bradford (30).

Thirty positive clones were selected for sequencing based on functional screening and enzymatic assays, and their fosmid DNA was digested with 10 units of *Bgl*III restriction enzyme (Promega, Madison, WI, USA) at 37 °C overnight for restriction profile analysis to confirm that each fosmid was unique. Fosmid DNAs were pooled in equal amounts before sequencing using a HiSeq2000 (Illumina) with the 100-bp paired-end protocol at MacroGen Inc. (<http://dna.macrogen.com>).

Sequence and Phylogenetic Analysis of Fosmid DNA

A sequence quality check was performed using FastQC software (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Reads were filtered before assembly such that for a pair of Paired-Ended reads each read should have >90% of bases with base quality \geq Q20. Reads matching the fosmid vector sequence were removed and the remaining reads were assembled into contigs using the SOAPdenovo2 method (<http://soap.genomics.org.cn/soapdenovo.html>).

Contig sequences were screened for Open Reading Frames (ORFs) in the PRODIGAL program (Prokaryotic Dynamic Programming Genefinding Algorithm) (<http://prodigal.ornl.gov>) (31). The predicted genes were submitted to the MG-RAST metagenomic online server (32) for both taxonomic and functional annotation. Taxonomic annotation was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (33). Functional annotation was performed using the subsystems technology, in which genes are classified in a hierarchical structure in which all genes required for a specific function are grouped into subsystems (34). All annotation was done using an E-value cutoff of 1E-5 and a minimum identity cutoff of 60%.

In order to identify carbohydrate-active enzymes, the amino acid sequences were submitted to the CAZymes Analysis Toolkit (CAT) (<http://mothra.ornl.gov/cgi-bin/cat/cat.cgi>) (35) for annotation. Enzymes were searched using the sequence similarity-based annotation against the entire non-redundant sequences of the CAZy

database, with the optional Pfam-based annotation that uses association rules inferred considering the CAZy database using the association rule-learning algorithm. Then, 14 pairs of primers were designed to amplify cellulase and hemicellulase genes using the ORF sequences identified by the PRODIGAL program after enzymes selection done by CAZy database. To amplify those glycoside hydrolase genes, pools of fosmid DNA were used as template, followed by individual PCRs to identify which clones were harboring which genes. The PCR conditions were as follows: 95 °C for 120 s, followed by 30 cycles of 95 °C for 60 s, 61 °C for 45 s and 72 °C for 85 s, with a final annealing at 72 °C for 5 min. The PCR was performed using GoTaq[®] DNA Polymerase (Promega). The pairs of primers and the corresponding encoded enzyme for each gene are listed in Table 1.

In this dataset, we choose two contigs (both > 30 kb) containing carbohydrate-active enzymes for further analysis. Translated ORF sequences were searched against the Non-redundant Protein Sequences (nr) and the Clusters of Orthologous Groups databases at NCBI to predict their functions.

To study the evolutionary relationships among the putative proteins identified in Contigs 31 and 61, the amino acid sequences of the identified ORFs encoding carbohydrate-active enzymes were compared to sequences from the Non-redundant Protein Sequences (nr) database at NCBI using the BLASTP algorithm (36). Sequences with high identity, as identified by the search, were imported into MEGA 6 software (37), aligned using ClustalW (38), and calculation of the phylogenetic trees was based on these sequence alignments using the Neighbor-joining method. To check the robustness of the resulting tree and the statistical significance levels of the interior nodes, bootstrap analysis with 1000 replicates was carried out and values above 70% were reported.

Nucleotide Sequence Accession Numbers

Sequences can be accessed at <https://copy.com/vAUcdc7Vr9UXbiKS>. The sequences of contigs 31 and 61 have been deposited in GenBank under the accession number KP867044 and KP867045, respectively.

Table 1. Primers used in this study

Primer	Sequence (5'-3')	Corresponding encoded enzyme
P1-F	CGCAGATCTATGGACAAAAAGAAAC	Contig 31_21
P1-R	CCCGGCGGCCGCGCAGTTGATACTATT	(α -L-fucosidase)
P2-F	CGCAGATCTATGGTGCTTGAAGTCAG	Contig 31_24
P2-R	CGTTGCGGCCGCGAGAATACGCTCTCA	α -xylosidase
P3-F	CGCAGATCTATGGAACAAAAAGAAC	Contig 31_25
P3-R	CCCGGCGGCCGCTTTATTTATTCCAC	β -glucosidase
P4-F	CGCAGATCTATGAAAAAGCAAA	Contig 32_1
P4-R	CCTTGCGGCCGCTTATGCCTCAACTTT	α -N-arabinofuranosidase
P5-F	CGCAGATCTATGTTCAAACCTGCGT	Contig 42_5
P5-R	CCTTGCGGCCGCGGACATCAAAAGACA	glycosyl hydrolase
P6-F	CGCAGATCTATGAATGAGTATATGAGAG	Contig 54_1
P6-R	CCTTGCGGCCGCTCTTATTAATATTCC	β -xylosidase
P7-F	CGCAGATCTATGATAACCACTGAGG	Contig 61_21
P7-R	CCTTGCGGCCGCGAAAAATCGATTGCTA	glucan 1,3- β -glucosidase
P8-F	CGCAGATCTATGGATGGGCAAGAA	Contig 61_33
P8-R	CCTTGCGGCCGCTTAATCCGTTTGGT	β -glucosidase
P9-F	CGCAGATCTATGTTTCCTCTAGGT	Contig 63_9
P9-R	CCTTGCGGCCGCGATTAAAAACCGTTG	β -glucosidase
P10-F	CGCAGATCTATGACGCTCAGGGAG	Contig 65_6
P10-R	CCCAGCGGCCGCGACCTTATTCATCCTA	β -glucosidase
P11-F	CCCAGATCTATGGCGGTAGATATCA	Contig 66_9
P11-R	CCCAGCGGCCGCAAACCCATTTATTC	β -glucosidase
P12-F	CGCAGATCTATGACGCTCAGGGAG	Contig 67_8
P12-R	CCTTGCGGCCGCTTATCATATCATCC	β -glucosidase
P13-F	CGCAGATCTATGAACGGTAAAAATG	Contig 67_13
P13-R	CCAAGCGGCCGCGTGGTTTATTCTTCC	glycosidase
P14-F	CGCAGATCTATGACGGGAAAAATG	Contig 67_26
P14-R	CCTTGCGGCCGCTCAGCATAAATTACC	β -glucosidase

F: Forward primer; R: Reverse primer

RESULTS

Metagenomic Library Screening for Clones Expressing Lignocellulolytic Activity and General Characteristics

To identify novel genes coding for cellulolytic enzymes from new sources, a fosmid metagenomic library was constructed with total DNA isolated from a microbial consortium originated from sugar cane bagasse and cow manure in a thermophilic composting phase (21). A total of 135,000 clones was obtained from the metagenomic library. Ten randomly chosen fosmid clones were analyzed by restriction enzyme digestion to determine the average size of the DNA inserts; this was calculated to be approximately 26 kb, with the full library harboring about 3.5 Gbp of metagenomic DNA.

About 5% of the clones from the library were screened for lignocellulolytic activity on solid medium. Of the 6,720 clones screened, 159 and 9 clones showed a clear zone around the colonies to indicate CMCase and xylanase activity, respectively, and 14 clones showed a black zone around the colonies to indicate β -glucosidase activity. These results represent a hit rate of approximately 42, 746 and 480 tested clones for CMCase, xylanase and β -glucosidase activity, respectively. Clones positive for laccase activity were not found in this screening.

In choosing fosmids for sequencing, the functional screening test on solid plates was repeated to confirm the positive clones, followed by enzymatic assay for CMCase, xylanase and β -glucosidase, with the highest clone reaching 204.57, 1689.45 and 1.34 U.mg⁻¹ specific activity, respectively. Fosmid DNA from selected positive clones of 9 xylanases, 12 β -glucosidases (two of the original 14 showed very low activity and were discarded) and 9 cellulases (chosen for their stronger halo formation and CMCase activity) were extracted and digested with the same enzyme to confirm that all inserted DNA fragments were different (so as to avoid sequencing identical DNA fragments).

Fosmid sequence, phylogenetic and functional analyses

Thirty fosmids were selected for sequencing based on their positive results for CMCase, xylanase and β -glucosidase activity as indicated above. A total of 9,256,230 high-quality reads were generated in this study, of which 2,996,474 matched fosmid vector sequence and were discarded from further analysis. The remaining 6,259,756 reads (average length of 101 bp) were assembled by using the SOAPdenovo2 method.

Sixty-seven contigs (average length of 16,743 bp) were generated after assembly, with the longest contig being 63,351 bp and the shortest 259 bp. The N50 (which means that half of all bases reside in contigs of this size or longer) was 33,374 bp. The average GC content of the reads was 59.33%.

Sequences of the 67 contigs were screened for ORFs using the PRODIGAL program and annotated by the MG-RAST program. Based on the KEGG protein database used by MG-RAST, of the 1,100 ORFs submitted to MG-RAST, 98.3% of the annotated ORFs showed identity to proteins found in Bacteria, 0.39% to Eukarya, 0.13% to Archaea and 1.18% was not assigned to any microbial group. Within the Bacteria domain, the most abundant phylum was Firmicutes (55.2%), with Clostridia being the most abundant class (90.8%) and comprising mostly the order Clostridiales (93.88%); Proteobacteria was the next most abundant phylum (39.47%), with Gammaproteobacteria (68.92%) being the most abundant class, comprising mostly the order Chromatiales (52.45%) (Figure 1).

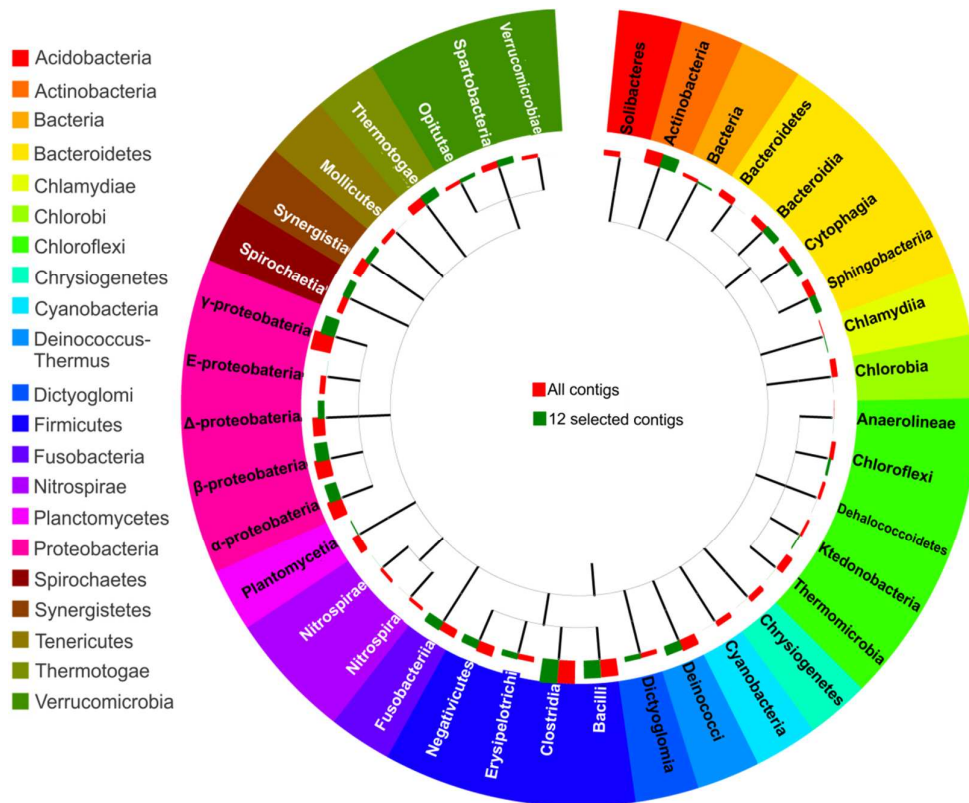


Figure 1. ORF distribution according with their taxonomic classification using the KEGG database in the MG-RAST program (32). The tree was constructed at the order level and it is colored at the phylum level. The stacked bars indicate the abundance of ORFs assigned to each order in the analysis

The amino acid sequences conceptually translated from the 67 contigs were used to predict a function based on homology by the MG-RAST platform using the SEED subsystem (39). Seven hundred nineteen ORFs were assigned to 25 functional groups (subsystems). Clustering-based subsystems and carbohydrate metabolism had the largest quantity of annotated reads assigned, representing 15.43% and 12.38% of the total of ORFs, respectively (Figure 2). Genes associated with Miscellaneous (10.15%), Amino Acid and Derivatives (8.48%), and Cofactors, Vitamins and Prosthetic Groups (7.78%), and Membrane Transport (5.70%) were the next most prevalent functional groups in the assigned annotated reads (Figure 2). Those groups accounted for more than 50% of the hits. Based on the MG-RAST subsystem classification, most of the annotated ORFs (10.98% of all ORFs) fell into the Clustering-based subsystem, and were related to an uncharacterized second level which includes the category Sugar Utilization in Thermotogales (3.47% of all ORFs) containing such genes as β -glucosidase, α -galactosidase, endo-1,4- β -xylanase, xylose transporter and others. The Carbohydrate subsystem, the second largest category, was also dominated by the uncharacterized category at the second level (containing 18.38% of the ORFs), which includes the category Sugar Utilization in Thermotogales mentioned above. The next subcategories in the Carbohydrate subsystem are enzymes involved in Central Carbohydrate Metabolism (13.97% of ORFs within the category), Di- and Oligosaccharides (13.24%), CO₂ Fixation (12.50%) and Monosaccharides Utilization (10.29%).

Contig analysis of clones associated with carbohydrate hydrolysis

To search for genes related to hydrolysis of carbohydrate, analysis of the amino acid sequences of ORFs from all 67 contigs was performed using the Carbohydrate-Active enZYme Database (CAZy). This analysis identified 188 gene modules across 45 families of glycosyl hydrolases, glycosyltransferases and carbohydrate esterases, with 74 genes in 26 families of glycosyl hydrolases. Twelve contigs containing 34 genes coding for carbohydrate-degrading enzymes, which are interesting in terms of the saccharification step of cellulose- and hemicellulose-deconstruction for bioethanol production, were detected and manually annotated by BLASTp. Table 2 summarizes the information about the glycoside hydrolase families detected. The most similar protein and the microbial class for each ORF in those 12 contigs were assigned.

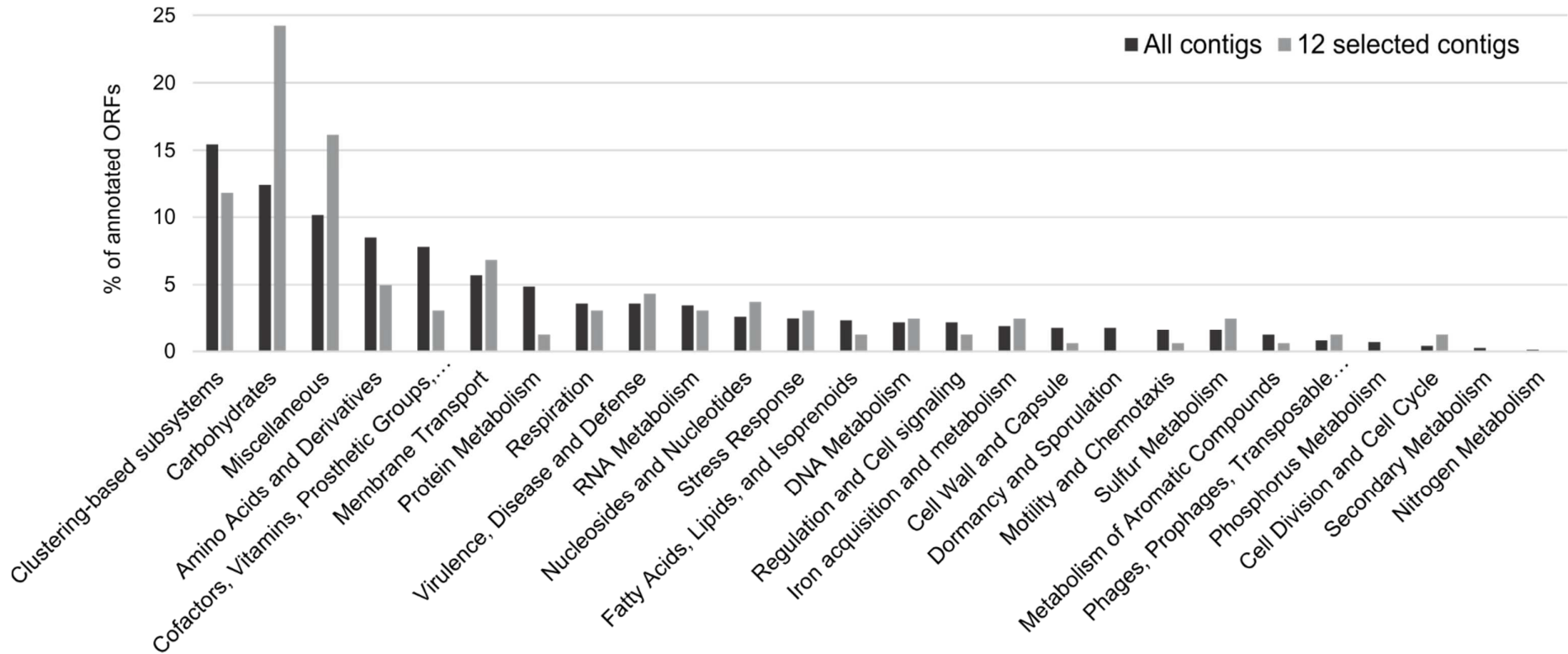


Figure 2. Relative distribution (in percentage of annotated ORFs) of the metabolic subsystems (using SEDD subsystems in the MG-RAST program) detected in the fosmid library. In black, ORFs detected in all 67 contigs. In gray, ORFs detected in the 12 selected contigs containing carbohydrate-degrading enzymes.

Table 2. Annotation of protein sequences related to glycosyl hydrolases. Annotation was performed using the CAZymes Analysis Toolkit (CAT) (35).

Gene ID ^a	Length (aa)	CAZy Family	Most similar protein	Identity (%)	Organism (GenBank Accession Number)	Class
ORF 31_15	794	GH95	α -L-fucosidase	41	<i>Paenibacillus mucilaginosus</i> KNP414 (AEI43267.1)	Bacilli
ORF 31_21	452	GH29	α -L-fucosidase	48	<i>Asticcacaulis excentricus</i> CB 48 CS 48 (ADU14844.1)	α -proteobacteria
ORF 31_24	792	GH31	α -xylosidase	54	<i>Clostridium cellulolyticum</i> H10 ATCC 35319 (ACL75366.1)	Clostridia
ORF 31_25	741	GH3	β -glucosidase	53	<i>Roseburia intestinalis</i> XB6B4 (CBL12457.1)	Clostridia
ORF 31_27	339	GH105	Rhamnogalacturonyl hydrolase	76	<i>Paenibacillus polymyxa</i> M1 (CCC83780.1)	Bacilli
ORF 31_28	733	GH36	α -galactosidase	86	<i>Clostridium stercorarium</i> DSM 8532 (AGI38523.1)	Clostridia
ORF 32_1	504	GH51	α -N-arabinofuranosidase	70	<i>Geobacillus</i> sp. GHH01 (AGE22472.1)	Bacilli
ORF 32_4	454	GH4	α -galactosidase	49	<i>Clostridium pasteurianum</i> BC1 (AGK98481.1)	Clostridia
ORF 41_5	760	GH3	β -glucosidase	51	<i>Niastella koreensis</i> GR20-10 (AEV97206.1)	Sphingobacteria
ORF 41_10	1018	GH3	β -glucosidase	49	<i>Bacillus</i> sp. GL1 (BAA36161.1)	Bacilli
ORF 42_5	304	GH130	glycosyl hydrolase	60	<i>Thermoanaerobacterium thermosaccharolyticum</i> (ADL69414.1)	Clostridia
ORF 42_9	1077	GH3	β -glucosidase	41	<i>Clostridium saccharoperbutylacetonicum</i> N1-4 (AGF58529.1)	Clostridia
ORF 54_1	735	GH3	β -xylosidase	69	<i>Mahella australiensis</i> 50-1 BOM (AEE97207.1)	Clostridia
ORF 54_2	1354	GH43	β -xylosidase	62	<i>Caldicellulosiruptor saccharolyticus</i> DSM 8903 (ABP67988.1)	Clostridia
ORF 54_3	316	GH10	endo-1,4- β -xylanase	87	<i>Clostridium stercorarium</i> DSM 8532 (AGI38766.1)	Clostridia
ORF 59_8	624	GH20	N-acetyl- β -hexosaminidase	78	<i>Symbiobacterium thermophilum</i> IAM 14863 (BAD39289.1)	Clostridia
ORF 59_9	388	GH3	N-acetylglucosaminidase	47	<i>Halobacillus halophilus</i> DSM 2266 (CCG44816.1)	Bacilli
ORF 59_14	538	GH3	β -glucosidase	73	<i>Symbiobacterium thermophilum</i> IAM 14863 (BAD39285.1)	Clostridia
ORF 61_21	422	GH5	glucan 1,3- β -glucosidase	39	<i>Azoarcus</i> sp. KH32C (BAL26553.1)	Y-proteobacteria
ORF 61_33	296	GH3	β -glucosidase	43	<i>Carnobacterium maltaromaticum</i> LMA28 (CCO12977.2)	Bacilli
ORF 63_7	335	GH16	endo-1,3(4)- β -glucanase	100	<i>Clostridium thermocellum</i> DSM 1313 (ADU75066.1)	Clostridia
ORF 63_9	472	GH1	β -glucosidase	100	<i>Clostridium thermocellum</i> ATCC 27405 (ABN51453.1)	Clostridia
ORF 64_17	482	GH13	α -amylase	66	<i>Aromatoleum aromaticum</i> EbN1 (CAI10105.1)	β -proteobacteria
ORF 64_18	377	GH13	α -amylase	50	<i>Nitrosococcus halophilus</i> Nc4 (ADE15576.1)	Y -proteobacteria
ORF 65_6	450	GH1	β -glucosidase	58	<i>Halobacteroides halobius</i> DSM 5150 (AGB40269.1)	Clostridia
ORF 65_20	788	GH3	xylan 1,4- β -xylosidase	73	<i>Caldanaerobius polysaccharolyticus</i> KMCJ (AFM44649.1)	Clostridia
ORF 66_9	756	GH3	β -glucosidase	99	<i>Clostridium thermocellum</i> ATCC 27405 (ABN52488.1)	Clostridia
ORF 66_25	682	GH43	arabinoxylan arabinofurano hydrolase	99	<i>Clostridium thermocellum</i> DSM 1313 (ADU74055.1)	Clostridia
ORF 67_8	749	GH3	β -glucosidase	82	<i>Clostridium stercorarium</i> DSM 8532 (AGI38411.1)	Clostridia
ORF 67_10	309	GH130	Glycosidase	40	<i>Ilyobacter polytropus</i> DSM 2926 (AGI38411.1)	Fusobacteria
ORF 67_13	330	GH130	Glycosidase	68	<i>Paenibacillus mucilaginosus</i> KNP414 (AEI40120.1)	Bacilli
ORF 67_21	304	GH20	xylose isomerase	82	<i>Clostridium stercorarium</i> DSM 8532 (AGI38409.1)	Clostridia
ORF 67_25	848	GH78	α -L-rhamnosidase	39	<i>Paenibacillus</i> sp. Y412MC10 (ACX64747.1)	Bacilli
ORF 67_26	752	GH3	β -glucosidase	79	<i>Clostridium stercorarium</i> DSM 8532 (AGI38411.1)	Clostridia

The ORFs detected in those 12 contigs were predominantly affiliated to organisms from the Firmicutes (76.38% of ORFs in the 12 contigs) and Proteobacteria (12.56%) phyla. At the hierarchical class level, Clostridia (87.5%) and Gamma-proteobacteria (44%) were the most frequent classes found within the Firmicutes and Proteobacteria phyla, respectively. ORFs in the 12 selected contigs represent 36.7% of all ORFs identified in Firmicutes considering the 67 contigs, demonstrating an enrichment of genes affiliated to this phylum in these contigs.

A closer look at the annotation of the ORFs from the 12-selected contigs by using MG-RAST (Figure 2) showed that their subsystem abundances were the same in those 67 contigs, but most of the ORFs (24.22% of all ORFs in the 12 contigs) were assigned within the Carbohydrate subsystem (Figure 2). Those ORFs represent 43.8% of all ORFs in the 67 contigs identified within this subsystem, in addition to representing 94.4% and 78.57% of the ORFs assigned in the categories di- and oligosaccharides and monosaccharides, respectively. Furthermore, the 12 selected contigs harbor 72% of all ORFs assigned to the category Sugar Utilization in Thermotogales.

To confirm the correct assembly of the data, we designed 14 pairs of primers (Table 1) targeting the 34 genes encoding carbohydrate-degrading enzymes. Only two pairs of primers (P4 and P6) failed to amplify the corresponding putative α -N-arabinofuranosidase and β -xylosidase gene, respectively. Because PCR amplification using the other 12 pairs of primers was successful, we detected which fosmid contained each gene (Figure S1) (because the 30 fosmid DNAs had been pooled before sequencing). Although the sequenced fosmids were selected by their positive result for CMCase (9 fosmids), xylanase (9 fosmids), and β -glucosidase (12 fosmids) activity, PCR amplification was detected only in clones selected by their β -glucosidase activity. Besides these genes, one α -L-fucosidase, one glycosyl hydrolase from the GH130 family, and one α -xylosidase gene were amplified.

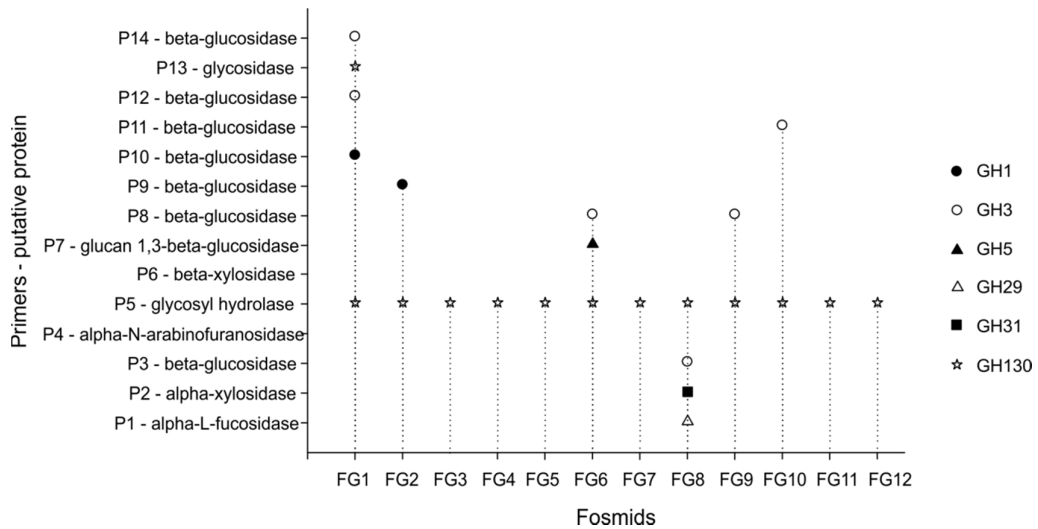


Figure S1. Presence of genes coding for carbohydrate-active enzymes in different fosmid clones from the metagenomic library. The presence of each gene was investigated by PCR using the primers listed in Table 1. Proteins of the glycosyl hydrolase (GH) family are identified.

We picked two contigs (Contig 31 and Contig 61) for further analysis based on their length and the presence of genes coding for putative glycoside hydrolases but with low identity to similar proteins in databases. The assembled sequences of each one were 42,007 bp (Contig 31) and 30,498 bp (Contig 61) (Figure 3). Of the 31 ORFs in Contig 31, 10 showed $\leq 60\%$ identity with any known gene, whereas 7 showed $\geq 80\%$ identity. For the 33 ORFs in Contig 61, 17 showed $\leq 60\%$ identity with any known gene and only 5 showed $\geq 80\%$ identity with proteins in the database. All of the ORFs of Contig 31 and Contig 61 were assigned to the Firmicutes phylum. From the 14 pairs of primers used to confirm the assembly data, three pairs were designed for ORFs in Contig 31 (ORF 31_21; 31_24; 31_25) and two pairs for those in Contig 61 (ORF 61_21; 61_23) (Tables 1 and 2). All five of these predicted proteins show low identity to proteins in the database. For example, the protein assigned to ORF 31_25 shows 53% identity with a β -glucosidase of *Roseburia intestinalis*, and the protein assigned to ORF 61_21 shows 39% identity with a glucan 1,3- β -glucosidase of *Azoarcus* sp. (Table 2).

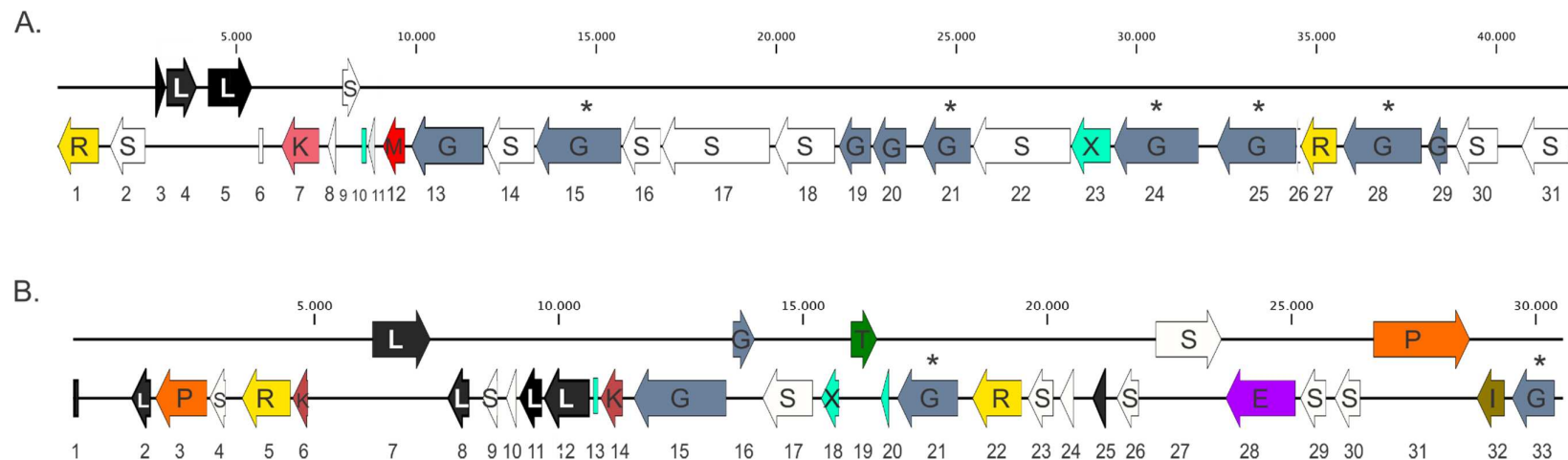


Figure 3. Schematic map of ORFs identified in Contig 31 (A) and Contig 61 (B). The ORFs are colored and labeled according to the COG functional categories as E (amino acid transport and metabolism), G (carbohydrate transport and metabolism), I (lipid metabolism), K (transcription), L (replication, recombination and repair), M (cell wall/membrane/envelope biogenesis), O (post-translational modification, protein turnover, chaperone functions), P (inorganic ion transport and metabolism), R (general function prediction only), S (function unknown), and T (signal transduction mechanisms). ORFs labeled with an X had no match in the protein or nucleotide databases. Each ORF was assigned a number shown below each map and the asterisks indicate those ORFs coding carbohydrate-degrading enzymes identified by the CAZy toolkit (32).

Contig 31 contains 31 ORFs, 4 positive- and 27 negative-stranded, whereas Contig 61 contains 33 ORFs, 5 positive- and 28 negative-stranded. All the ORFs from these two contigs were also searched against the Non-redundant protein Sequences (nr) and the COG databases to predict functional categories (Figure 3). The ORFs were classified into functional categories, with 9 (29%) and 4 (12%) classified into the G category (carbohydrate transport and metabolism), and 12 (38.7%) and 10 (30%) classified into the S category (function unknown) for Contig 31 and Contig 61, respectively (Figure 3). In Contig 61, seven (21%) ORFs were classified in the L category (replication, recombination and repair). The rest of the categories were represented by ≤ 3 ORFs each in the two contigs.

Phylogenetic analyses were done for those enzymes in Contigs 31 and 61 identified as carbohydrate-active enzymes by the CAZy tool kit (Table 2). A comparison of the β -glucosidases of these sequences with those previously deposited in the GenBank database revealed the three enzymes to be distributed in different groups supported by high bootstrap values, and two of those enzymes (Contig 31_25 and Contig 61_33) form a separate branch from other sequences from bacteria belonging to the Firmicutes phylum (Figure 4a), whereas the putative β -glucosidase encoded by ORF 61_21 clusters with β -glucosidase sequences derived from bacteria of the genera *Azoarcus* (Betaproteobacteria class), *Haloplasma* (unclassified Bacteria) and *Carnobacterium* (Bacilli class) (Figure 4a).

In the phylogenetic analysis using amino acid sequences from the putative α -L-fucosidases identified in Contig 31, the two sequences clustered with sequences of α -L-fucosidases derived from bacteria of the Clostridia, Bacilli, and Alpha- and Beta-proteobacteria classes (Figure 4b), which belong to the Firmicutes and Proteobacteria phyla. Although they are phylogenetically related, their low identity with known proteins in databases suggests they are novel proteins.

Phylogenetic reconstruction using the amino acid sequences of the ORFs encoding the putative α -xylosidase and α -galactosidase demonstrated that they are closely related to proteins derived from bacteria of the *Clostridium* genus, but form distinct branches (Figure 4c and 4d).

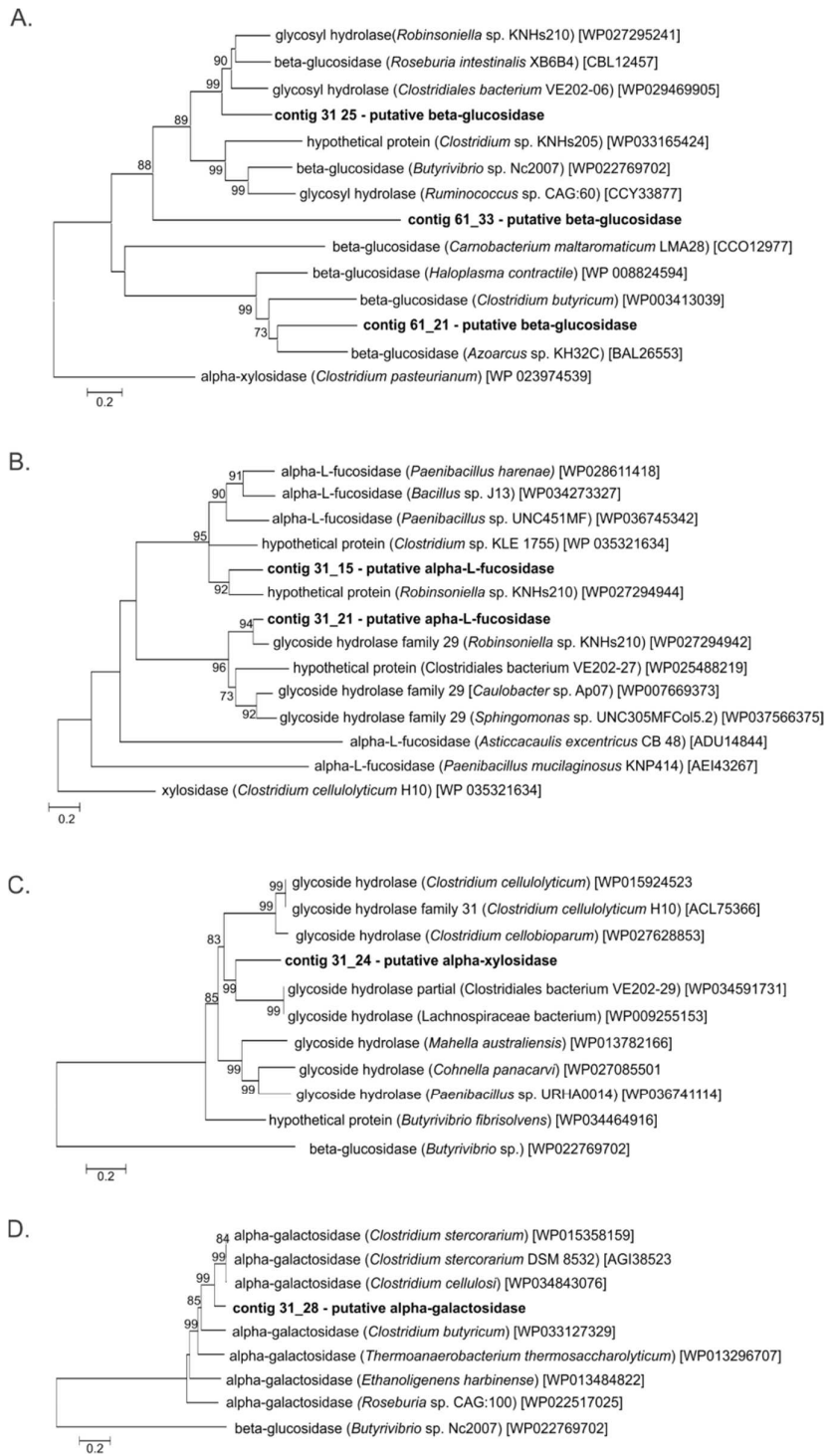


Figure 4. Neighbor-joining phylogenetic trees based on deduced amino acid sequences of seven carbohydrate-active enzymes identified by the CAZy toolkit in Contigs 31 and 61. (a) β -glucosidase; (b) α -L-fucosidase; (c) α -xylosidase; (d) α -galactosidase. Bootstrap values are shown for nodes with over 70% support. The sequences found in this study are presented in bold script. The accession numbers for each sequence extracted from the GenBank database are shown after each sequence name.

DISCUSSION

To construct a metagenomic library to search for new lignocellulosic enzymes, we used DNA fragments representing a mixed microbial population obtained from degrading lignocellulosic residue (sugar cane bagasse associated with cow manure in a thermophilic composting phase). This microbial population was reproduced in the laboratory by inoculating the mixture in medium containing cellulose as the main carbon source (21). From the library, 34 protein sequences related to glycosyl hydrolase families were identified, with 16 sharing $\leq 60\%$ identity with the most similar protein sequences deposited in databases. Phylogenetic reconstruction performed with protein sequences of seven of the putative enzymes revealed low relatedness to known sequences, suggesting they may be novel proteins.

We chose a function-based metagenomic approach to search for novel lignocellulosic enzymes. The approach had already led to the discovery of novel β -galactosidases/ α -arabinopyranosidases (40), cellulases (11,41), xylanases (42), xylose isomerases (43), and β -glucosidases (6,44), supporting this strategy as a powerful activity-based screening tool to identify entirely new classes of gene sequences encoding new or known functions (43,45). Several sources of potential cellulose- and hemicellulose-depleting microbial communities are being used for metagenomic analysis. We reproduced a microbial consortium by mixing two natural environmental sources: sugarcane bagasse and cow manure (21). Consistent with the origin of the consortium (rich in lignocellulosic materials) and the conditions of its reproduction in the laboratory, we obtained a high hit rate in our screening by favouring the presence of lignocellulosic activity. It is a common practice for enhancing the desired functions in a microbial community to induce the growth of specific microorganisms by applying pre-enrichment methods to the sample to produce an increased screening hit rate (46). For example, Grant et al. (47) used metagenomic DNA of cultures grown in medium containing carboxymethylcellulose as the only carbon source, and they observed that the number of glycosyl hydrolases detected was about four times greater than the number identified in metagenomic libraries obtained using DNA taken directly from environmental samples. Usually, the rates of positive hits using functional metagenomic approaches in search of novel cellulases and hemicellulases are low. For example, from a fosmid metagenomic library constructed with DNA samples isolated from soils from a wetland, the number of positive hits for β -glucosidase activity was just 5 from 14,000 clones screened (48). Feng et al (49) found 11 positive clones for

cellulase activity in 32,500 clones screened from a cosmid metagenomic library originated from rabbit cecum contents. In our metagenomic library, we had hit rates of 42 (one positive every 42 analyzed) (159 positives in 6,720 clones), 746 (9 in 6,720) and 480 (14 in 6,720) tested clones for CMCase, xylanase, and β -glucosidase activity, respectively.

From the 182 *E. coli* positive clones selected by cellulase, β -glucosidase and xylanase activity-based screening from the fosmid metagenomic library in this study, 30 fosmid DNAs were sequenced, yielding 67 contigs after assembly. By analyzing 1,100 ORFs generated from those contigs, we found that Firmicutes (55.2%) and Proteobacteria (39.47%) were the most abundant phyla within the Bacteria domain (representing 98.3% of the annotated ORFs) (Figure 1). These results are consistent with literature data, in that the Firmicutes was previously reported to be abundant in the thermophilic composting phase (50), in feces (51), and in the bovine rumen (52), whereas in sugarcane bagasse Proteobacteria were found to be especially abundant, followed by Firmicutes (53). Even analyzing only 30 fosmid sequences from our metagenomic library of 135,000 clones, we observed the same pattern of phylum distribution and the presence of a variety of taxa in our results.

Concerning the functional roles, the majority of the ORFs were classified in Clustering-based subsystems and Carbohydrate metabolism (Figure 2), both including the category Sugar Utilization in Thermotogales, which comprises genes encoding putative β -glucosidase, α -galactosidase, and endo-1,4- β -xylanase enzymes. Genes characterized in the Clustering-based subsystems category usually cluster together in genomic regions and they are functionally coupled (54). In addition, the order Thermotogales is represented by anaerobic, thermophilic and hyperthermophilic microorganisms (55). It is known that in anaerobic microorganisms there is an entity called the cellulosome which is an extracellular enzyme complex consisting of a scaffold and enzymes that are capable of degrading plant cell walls, whereas in aerobic bacteria several individual cellulases are secreted and act synergistically to hydrolyze plant cell walls (56). Some cellulases, hemicellulases and other carbohydrate-active enzymes work in concert to facilitate the degradation of carbohydrates. In some anaerobic microorganisms, there are large gene clusters as well as unlinked genes that encode enzymes of the cellulosomal complex (56). The predominance of ORFs assigned to anaerobic microorganisms in this study might result from the presence of carbohydrate-depleting enzyme genes colocalized in genomic regions that may encode cellulosomal enzymes. For example, in Contig 31 (Figure 3a) there are nine linked genes related to carbohydrate transport and metabolism, six of which belong to

glycoside hydrolase families (Table 2), and an additional 12 genes of unknown function.

The same profile of predominant phyla and functional groups was observed in the 12 contigs selected for the presence of carbohydrate-active enzymes. The main difference concerned the Carbohydrate functional group, which was more prevalent than the Clustered-based subsystems (Figure 2). Thirty-four ORFs (present in those 12 contigs) encode putative proteins related to glycosyl hydrolases distributed in 17 GH families, especially the GH3 (35%) and GH130 (9%) families. Sixteen protein sequences share $\leq 60\%$ identity with the most similar protein sequence deposited in databases, with six of them (ORFs: 31_15; 31_21; 31_24; 31_25; and ORFs: 61_21; 61_33) (Table 1) likely derived from representatives of the Firmicutes. Among the detected enzymes there are five β -glucosidases from the GH3 family, which is an important enzyme class needed to hydrolyze cellulose but is usually present with low activity in commercial cocktails of cellulases (6). One of the putative β -glucosidases found in this study shares only 43% identity (at the protein level) with a similar enzyme of *Carnobacterium maltaromaticum*, a bacterium of the Firmicutes phylum and Bacilli class, and another β -glucosidase shared only 41% identity with a similar enzyme of *Clostridium saccharoperbutylacetonicum*, from class Clostridia within Firmicutes (Table 2). Studies of diverse novel cellulases have shown that these enzymes have evolved independently and are unrelated in sequence and structure (57), characterizing them as a large class of nonhomologous isofunctional enzymes (58). The wide diversity of domain architectures of cellulases, even within the same protein family, as well as differences of sequence and structure can complicate the identification of novel enzymes (57), but supports the use of functional metagenomic approaches for discovering improved (for biotechnological purposes) and new hydrolase enzymes. Ferrer et al (59) isolated and characterized a novel multifunctional enzyme from the GH43 family and suggested that the great diversity of polymeric substrates imposed on a complex microbial community may drive the evolution of this enzyme category.

Most of the ORFs present in Contig 31 and Contig 61 are encoded in the same strand (Figure 3), indicating the existence of possible clusters of genes coding for enzymes involved in polysaccharide degradation. *Clostridium cellulovorans*, *C. acetobutylicum* and *C. cellulolyticum* contain unlinked genes that encode cellulosomal enzymes, but contain large gene clusters as well with related organization (60,61,62). More than 60% of the ORFs in Contig 31 and almost 50% in Contig 61 are related to carbohydrate transport and metabolism, and 'unknown function' categories. All the ORFs in those two contigs that could not be assigned to

any function did, however, show homology with hypothetical proteins of different microorganisms (Tables S1 and S2). Furthermore, from the 12 pairs of primers designed to confirm the correct assembly, three of the ORFs correspond to genes from Contig 31 (ORFs 31_21; 31_24; 31_25) and two to genes from Contig 61. The three genes of Contig 31 were found in the same fosmid DNA (FG8) (Table 2, Figure S1), and the two genes of Contig 61 were found in the same fosmid DNA (FG6), except that a gene related to ORF 61_33 is also present in fosmid FG9. These results might indicate that Contigs 31 and 61 represent a genomic region from single microorganisms from the microbial consortium. We also found genes coding for putative transposases in these contigs (3 genes in each contig) and, considering the large sizes of the two contigs (42 Mbp for Contig 31 and 30.5 Mbp for Contig 61) and information in the literature, it is reasonable to hypothesize that these contigs contain sequences that could have been horizontally transferred among bacteria present in the consortium used as the gene source. Accordingly, *Clostridium cellulovorans* contains a transposase gene at the 3' end of a cluster of genes involved in cellulose degradation, indicating that lateral gene transfer might have occurred (60).

Comparison of all amino acid sequences from the proteins of Contigs 31 and 61 with those previously deposited in the GenBank database confirmed there are multiple differences among the newly discovered proteins and known sequences. Moreover, the enzymes from the same contig clustered separately and with proteins from different bacteria. Finally, although some enzymes, such as the putative α -galactosidase encoded by ORF 31_28, have higher identity with enzymes from *Clostridium*, the comparison of the entire sequences of Contigs 31 and 61 against the Complete Genome NCBI database revealed that only 3% of sequences from Contig 31 matched sequences in the genome of *Clostridium thermocellum* and 10% the genome of *Clostridium stercoarium*; and 5% of sequences of Contig 61 had 80% identity with the genome of *Clostridium stercoarium*. Taken together, the phylogenetic analyses and the low identity values (Tables S1 and S2) indicate that the enzymes identified in this work are distantly related to known proteins and could represent new enzymes for biotechnological purposes such as cellulose degradation.

Table S1. Comparison of amino acid sequences of ORFs in Contig 31 against the -redundant protein database (nr) at NCBI.

Contig gene	Coverage (%)	Identity (%)	Closest genus at NCBI	Corresponding protein
1	71	86	<i>Clostridium</i>	HD superfamily phosphohydrolases
2	99	99	<i>Clostridium</i>	Hypothetical protein
3	95	99	<i>Clostridium thermocellum</i>	Transposase
4	99	100	<i>Clostridium</i>	Putative transposase
5	99	93	<i>Clostridium clarifavum</i>	Transposase
6	98	98	<i>Clostridium stercorarium</i>	Hypothetical protein
7	33	42	<i>Paenibacillus</i>	AraC family transcriptional regulator
8	81	31	<i>Clostridium intestinalis</i>	Hypothetical protein
9	90	76	<i>Ruminoclostridium</i>	Hypothetical protein
10	-	-	no hit	no hits
11	90	41	<i>Paenibacillus</i>	AraC family transcriptional regulator
12	93.2	67	<i>Bacillus</i>	N-acetylmuramoyl-L-alanine amidase
13	98	55	<i>Robinsoniella</i>	bgaB; beta-galactosidase [EC:3.2.1.23]
14	99	79	<i>Clostridium cellulosi</i>	Transport protein
15	97	53	<i>Robinsoniella</i>	AXY8; alpha-L-fucosidase 2 [EC:3.2.1.51]
16	98	53	Clostridiales	Hypothetical protein
17	99	46	<i>Paenibacillus</i>	Hypothetical protein
18	88	60	<i>Robinsoniella</i>	Hypothetical protein
19	95	67	<i>Robinsoniella</i>	lplC; putative aldouronate transport system permease protein
20	94	74	<i>Robinsoniella</i>	lplB; putative aldouronate transport system permease protein
21	99	68	<i>Robinsoniella</i>	FUCA; alpha-L-fucosidase [EC:3.2.1.51]
22	97	65	<i>Robinsoniella</i>	Hypothetical protein
23	99	66	Clostridiales	Trimethylamine corrinoid protein
24	96	62	Lachnospiraceae bacterium	yicL; alpha-D-xyloside xylohydrolase [EC:3.2.1.177]
25	99	54	Clostridiales bacterium	bglX; beta-glucosidase [EC:3.2.1.21]
26			no hit	no hits
27	98	78	<i>Paenibacillus</i>	yteR; unsaturated rhamnogalacturonyl hydrolase [EC:3.2.1.172]
28	99	86	<i>Clostridium</i>	E3.2.1.22B; alpha-galactosidase [EC:3.2.1.22]
29	99	77	<i>Clostridium</i>	PGAM; 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase [EC:5.4.2.11]
30	98	47	<i>Caudiprolobacter</i>	Hypothetical protein
31	97	51	<i>Treponema</i>	Multidrug transporter MATE

Table S2. Comparison of amino acid sequences of ORFs in Contig 61 against the Non-redundant protein database (nr) at NCBI.

Contig gene	Coverage (%)	Identity (%)	Closest genus at NCBI	Corresponding protein
1	71	48	<i>Thermoanaerobacter</i>	Integrase
2	98	92	<i>Bacillus</i>	ARSC2; arsenate reductase [EC:1.20.4.1]
3	97	80	<i>Clostridium</i>	TC.ACR3; arsenite transporter, ACR3 family
4	99	86	<i>Clostridium</i>	Redox-active disulfide protein 2
5	99	86	<i>Clostridium</i>	Predicted permease
6	94	79	<i>Clostridium</i>	arsR; ArsR family transcriptional regulator
7	99	86	<i>Clostridium stercorarium</i>	Transposase
8	64	73	<i>Enterococcus</i>	Recombinase
9	90	74	<i>Clostridium</i>	Hypothetical protein
10	98	58	<i>Clostridium</i>	Hypothetical protein
11	66	47	<i>Turcibacter</i>	Transposase
12	95	57	<i>Bacillus</i>	Transposase
13	66	62	<i>Bacillus</i>	no hits
14	89	37	<i>Sporolactobacillus</i>	Transcriptional Regulator
15	89	62	<i>Haloplasma</i>	PTS-EI.PTSI; phosphotransferase system, enzyme I, PtsI [EC:2.7.3.9]
16	92	54	<i>Paenibacillus</i>	PTS beta-glucoside transporter subunit
17	98	57	<i>Vlrginibacillus</i>	Hypothetical protein
18				No hits
19	99	68	<i>Bacillus</i>	Thioredoxin reductase
20				no hits
21	86	44	<i>Haloplasma</i>	Beta-glucosidase
22	99	54	<i>Acholeplasma brassicae]</i>	Predicted membrane protein, putative toxin regulator
23	91	45	<i>Listeria</i>	Hypothetical protein
24	69	61	<i>Bacillus</i>	Hypothetical protein
25	97	71	<i>Haloplasma</i>	hupB; DNA-binding protein HU-beta
26	91	44	<i>Bacillus</i>	Hypothetical protein
27	97	60	<i>Alkaliphilus</i>	Multidrug transporter MatE
28	99	56	<i>Haloplasma</i>	PepD; dipeptidase D [EC:3.4.13.-]
29	95	62	<i>Bacillus</i>	Hypothetical protein
30	94	60	<i>Flavobacterium</i>	Hypotetical protein
31	94	51	<i>Haloplasma</i>	Sulfatase
32	92	27	<i>Rhodopirelulla</i>	pgsA; CDP-diacylglycerol--glycerol-3-Phosphate 3-phosphatidyltransferase [EC:2.7.8.5]
33	85	32	<i>Clostridium</i>	Glycoside hydrolase

In conclusion, screening of the metagenomic library based on function revealed 182 positive clones with gene products able to hydrolyze polysaccharide. The sequencing results of 30 positive fosmids proved the feasibility of finding new genes for this purpose by using functional metagenomics applied to a complex microbial community obtained from decomposing sugarcane bagasse and cow manure. This result was confirmed by a more detailed analysis of 12 selected contigs containing seven carbohydrate-active enzymes sharing low identity with protein sequences in databases. Thus, such reproducible mixed cultures could serve as a reservoir of enzymes for future applications in biomass-degradation for biofuels production, including cloning the genes into expression systems to obtain hydrolytic enzymes secreted on an industrial scale with no cellular lysis, or for use in yeast transformants themselves for simultaneous saccharification and fermentation.

ACKNOWLEDGMENTS

This work was supported by CAPES, CNPq and FAPEMIG Brazilian funding agencies.

REFERENCES

1. **Camargo PD.** 2005. Força verde: um novo campo para a indústria química. *Revista Brasileira de Engenharia Química* 18-21.
2. **Rabelo SC, Fonseca NA, Andrade RR, Maciel Filho RR, Costa AC.** 2011. Ethanol production from enzymatic hydrolysis of sugarcane bagasse pretreated with lime and alkaline hydrogen peroxide. *Biomass Bioenerg.* **35**:2600-2607.
3. **Balat M, Balat H.** 2009. Recent trends in global production and utilization of bioethanol fuel. *Appl Energy* **86**:2273–2282.
4. **Parawira W, Tekere M.** 2011. Biotechnological strategies to overcome inhibitors in lignocellulose hydrolysates for ethanol production: review. *Crit Rev Biotechnol* **31**:20-31.
5. **Sun Y, Cheng J.** 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresource Technol* **83**:1–11.
6. **Del Pozo MV, Fernández-Arrojo L, Gil-Martínez J, Montesinos A, Chernikova TN, Nechitaylo TY, Waliszek A, Tortajada M, Rojas A, Huws SA, Golyshina OV, Newbold CJ, Polaina J, Ferre M, Golyshin PN.** 2012. Microbial β -glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. *Biotechnol Biofuels* **5**:1-13.
7. **Gnansounou E, Dauriat A.** 2010 Techno-economic analysis of lignocellulosic ethanol: A review. *Bioresource Technol* **101**:4980–4991.
8. **Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD.** 2007. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science.* **315**:804–807.
9. **Li LL, Mccorkle SR, Monchy S, Taghavi S, van der Lelie D.** 2009. Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. *Biotechnol Biofuels* **2**:10.
10. **Pang H, Zhang P, Duan CJ, Mo VX, Tang JL, Feng JX.** 2009. Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. *Curr Microbiol* **58**:404-408.
11. **Duan CJ, Xian L, Zhao GC, Feng Y, Pang H, Bai XL, Tang JL, Ma QS, Feng JX.** Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. 2009. *J. Appl. Microbiol.* **107**:245-256.
12. **Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VGH.** 2012. Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels.* **5**:45.

13. **Uchiyama T, Miyazaki K.** 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* **20**:616-622.
14. **Lee SW, Won K, Lim HK, Kim JC, Choi GJ, Cho KY.** 2004. Screening for novel lipolytic enzymes from uncultured soil microorganisms. *Appl Microbiol Biotechnol* **65**:720-726.
15. **Duan CJ, Feng JX.** 2010. Mining metagenomes for novel cellulase genes. *Biotechnol Lett* **32**:1765-1775.
16. **Yun J, Kang S, Park S, Yoon H, Kim MJ, Heu S, Ryu S.** 2004. Characterization of a Novel Amylolytic Enzyme Encoded by a Gene from a Soil-Derived Metagenomic Library. *Appl Environ Microbiol* **70**:7229-7235.
17. **Hjort K, Bergstrom M, Adesina MF, Jansson JK, Smalla K, Sjöling S.** 2010. Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen-suppressive soil. *FEMS Microbiol Ecol* **71**:197-207.
18. **Kim YJ, Choi GS, Kim SB, Yoon GS, Kim YS, Ryu YW.** 2006. Screening and characterization of a novel esterase from a metagenomic library. *Protein Express Purif* **45**:315-323.
19. **Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringle SG, Visel A, Woyke T, Wang Z, Rubin EM.** 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**:463–467.
20. **Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, et al.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*. **450**:560–565.
21. **Souza, R.A.** Obtenção de Inoculante e de Coquetel Enzimático Lignocelulolítico a partir de Comunidades Microbianas Termofílicas. 2012. 56f. Dissertação (mestrado). Universidade Federal de Viçosa, Viçosa, MG, 2012.
22. **Stevenson DM, Weimer PJ.** 2007. Dominance of *Prevotella* and low abundance of phylogenetic analysis of the microbial community in industrial sugarcane bagasse feedstock piles. *Biosci Biotechnol classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR.* *Appl Microbiol Biotechnol* **75**:165–174.
23. **Kasana RC, Salwan R, Dhar H, Dutt S, Gulati A.** 2008. A rapid and easy method for the detection of microbial cellulases on agar plates using Gram's iodine. *Curr Microbiol* **57**:503-507.

24. **Teather RM, Wood PJ.** 1982. Use of Congo red polysaccharide interactions complex formation between Congo red and polysaccharide in detection and assay of polysaccharide hydrolases. *Methods Enzymol* **160**:59–74.
25. **Eberhart B, Cross DF, Chase LR.** 1964. Beta-glucosidase system of *Neurospora crass.* I. beta-glucosidase and cellulose activities of mutant and wild-type strains. *J Bacteriol* **87**:761–770.
26. **Miller GL.** 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem* **31**:426-428.
27. **Ghose TK.** 1987. Measurement of cellulase activity. *Pure Appl Che* **59**:257–268.
28. **Bailey MJ, Biely P, Poutanen K.** 1992. Interlaboratory testing of methods for assay of xylanase activity. *J Biotechnol* **23**:257-270.
29. **Chen H, Li X, Ljundahl LG.** 1994. Isolation and properties of an extracellular beta-glucosidase from the polycentric rumen Fungus *Orpinomyces* sp. strain PC-2. *Appl Environ Microbiol* **60**:64-70.
30. **Bradford MM.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal Biochem* **72**:248–254.
31. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**:119.
32. **Meyer F, Paarmann D, Souza MD, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA.** 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**:386.
33. **Kanehisa M, Goto S.** 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**:27–30.
34. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*. **9**:75.
35. **Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B.** 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**:D233–238

36. **Altschul S, Gish W, Miller W, Myers E, Lipman D.** 1990. Basic local alignment search tool. *J Mol Biol* **215**:403-410.
37. **Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S.** 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* **30**:2725-2729.
38. **Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H, Valentin F, Wallace I, Wilm A, Lopez R, Thompson J, Gibson T, Higgins D.** 2001. ClustalW and Clustal X version 2.0. *Bioinformatics* **23**:2947-2948.
39. **Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, Crecy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Ruckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V..** 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**:5691–5702.
40. **Beloqui A, Nechitaylo TY, López-Cortés N, Ghazi A, Guazzaroni ME, Polaina J, Strittmatter AW, Reva O, Waliczek A, Yakimov MM, Golyshina OV, Ferrer M, Golyshin PN.** 2010. Diversity of glycosyl hydrolases from cellulose depleting communities enriched from casts of two earthworm species. *Appl Environ Microbiol* **76**:5934–5946.
41. **Kim SJ, Chang-Muk L, Bo-Ram H, Min-Yong K, Yun-Soo Y, Sang-Hong Y, Bon-Sung K, Hong-Ki J.** Characterization of a gene encoding cellulase from uncultured soil bacteria. 2008. *FEMS Microbiol Lett* **282**:44-51.
42. **Hu Y, Guimin Z, Aiyang L, Jing C, Lixin M.** 2008. Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach. *Appl Microbial Biotechnol* **80**:823-830.
43. **Parachin NS, Gorwa-Grauslund MF.** 2011. Isolation of xylose isomerases by sequence- and function-based screening from a soil metagenomic library. *Biotechnol Biofuels*. **4**:1-10.
44. **Biver S, Stroobants A, Portetelle D, Vandenberg M.** 2014. Two promising alkaline β -glucosidases isolated by functional metagenomics from agricultural soil, including one showing high tolerance towards harsh detergents, oxidants and glucose. *J. Ind Microbiol Biot* **41**:479-488.
45. **Handelsman J.** 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol R.* **68**:669-685.

46. **Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R, Wittwer P.** 2005. Metagenomic gene discovery: past, present and future. *Trends in Biotechnol* **23**:321-329.
47. **Grant S, Sorokin DY, Grant WD, Jones BE, Heaphy S.** 2004. A phylogenetic analysis of Wadi el Natrun soda lake cellulase enrichment cultures and identification of cellulase genes from these cultures. *Extremophiles*. **8**:421-429.
48. **Kim SJ, Lee CM, Kim MY, Yeo YS, Yoon SH, Kang HC, Koo BS.** 2007. Screening and characterization of an enzyme with beta-glucosidase activity from environmental DNA. *J Microbiol Biotechnol* **17**:905–912.
49. **Feng Y, Duan CJ, Pang H, Mo XC, Wu CF, Yt Y, Hu YL, Wi J, Tang JL, Feng JX.** 2009. Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. *Appl Microbiol Biotechnol* **75**:319-328.
50. **Li R, Li L, Huang R, Sun Y, Mei X, Shen B, Shen Q.** 2014. Variations of culturable thermophilic microbe numbers and bacterial communities during the thermophilic phase of composting. *World J Microb Biot* **30**:1737-1746.
51. **de Oliveira MNV, Jewell KA, Freitas FS, Benjamin LA, Tótola MR, Borges AC, Moraes CA, Suen G.** 2013. Characterizing the microbiota across the gastrointestinal tract of a Brazilian Nelore steer. *Vet Microbiol* **164**:307-314.
52. **Jami E, Mizrahi I.** 2012. Composition and Similarity of Bovine Rumen Microbiota across Individual Animals. *PLoS ONE*. **7**:e33306.
53. **Rattanachomsri U, Kanokratana P, Eurwilaichitr L, Igarashi Y, Champreda V.** 2011. Culture-independent *Biochem* **75**:232-239.
54. **Lu HP, Wang YB, Huang SW, Lin CY, Wu M, Hsieh CH, Yu HT.** 2012. Metagenomic analysis reveals a functional signature for biomass degradation by cecal microbiota in the leaf-eating flying squirrel (*Petaurista alborufus lena*). *BMC Genomics*. **13**:466.
55. **Huber R, Karlo S.** 1992. The Order Thermotogales, p 3809-3815. In: Balows A, Truper H, Dworkin M, Harder W, Schleifer KH (eds), *The Prokaryotes*, 2nd ed. Springer, New York.
56. **Doi R, Kosugi A.** 2004. Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nature* **2**:541-551.
57. **Sukharnikov LO, Cantwell BJ, Podar M, Zhulin IB.** 2011. Cellulases. ambiguous nonhomologous enzymes in a genomic perspective. *Trends Biotechnol* **29**:473–479.

58. **Omelchenko MV, Galperin MY, Wolf YI, Koonin EV.** 2010. Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. *Biol Direct* **5**:31.
59. **Ferrer M, Ghazi A, Beloqui A, Vieites JM, López-Cortéz N, Marín-Navarro J, Necgutaylo TY, Guazzaroni ME, Polaina J, Waliczek A, Chernikova TN, Reva ON, Golyshina OV, Golyshin PN.** 2012. Functional metagenomics unveils a multifunctional glycosyl hydrolase from the Family 43 catalysing the breakdown of plant polymers in the calf rumen. *PLoS ONE*. **7**:e38134.
60. **Tamaru Y, Karita S, Ibrahim A, Chan H, Doi RH.** 2000. A large gene cluster for the *Clostridium cellulovorans* cellulosome. *J Bacteriol* **182**:5906–5910.
61. **Nolling J, Breton G, Omelchenko MV, Makarova KS, Zeng Q, Gibson R, Lee HM, Dubois J, Qiu D, Hitti J, W YI, Tatusov RL, Sabathe F, Doucette-Stamm L, Soucaille P, Daly MJ, Bennett GN, Koonin EV, Smith DR.** 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J Bacteriol* **183**:4823–4838.
62. **Belaich A, Parsieglia G, Gal L, Villard C, Haser R, Belaich JP.** 2002. Cel9M, a new family 9 cellulase of the *Clostridium cellulolyticum* cellulosome. *J Bacteriol* **184**:1378–1384.

CAPÍTULO 2

CONSTRUÇÃO DE UM SISTEMA DE EXPRESSÃO PARA *Kluyveromyces marxianus* CCT 7735 (UFV-3): OBTENÇÃO DE LINHAGEM HOSPEDEIRA *ura3⁻* E DE VETOR INTEGRATIVO PARA CLONAGEM E TRANSFORMAÇÃO

RESUMO

Kluyveromyces marxianus é uma linhagem de levedura cujas características industriais, como alta conversão em massa celular (alta velocidade de crescimento) e termotolerância, seriam mais bem aproveitadas tendo disponíveis construções genéticas específicas para esta levedura que permitissem clonagem, eficiente transformação, expressão e secreção de diversas proteínas recombinantes de interesse biotecnológico. Neste trabalho, construímos uma linhagem de *K. marxianus* CCT 7735 (UFV-3) mutante auxotrófico deficiente para o gene *URA3*, para ser utilizada como hospedeira para expressão de proteínas heterólogas utilizando sistemas de expressão cuja marca de seleção seja o gene *URA3*. A deleção do gene *URA3* foi realizada utilizando a técnica *split-marker*. Dois fragmentos foram obtidos por fusão por PCR, cada um contendo uma sequência flanqueadora da região codificante do gene *URA3* (extremidades 5' e 3') e uma parte do gene de resistência a geneticina (KanMX) truncado. Após recombinação *in vivo* dos dois fragmentos gerados, o cassete de deleção resultante continha o gene KanMX flanqueado pelas extremidades da região codificante do gene *URA3* e foi utilizado para inativação deste mesmo gene por recombinação homóloga no genoma da *K. marxianus* CCT 7735 (UFV-3). A eficiência de integração com os fragmentos resultantes do *split-marker* foi de 100 %. Em paralelo à obtenção do mutante foi também construído um vetor de expressão integrativo denominado pKmLPG para produção de proteínas recombinantes em *K. marxianus* CCT 7735 (UFV-3). Para isso utilizou-se o promotor da enzima endo-poligalacturonase amplificado a partir do DNA de *K. marxianus* CCT 7735 (UFV-3), o peptídeo sinal α -Mating Factor de *Kluyveromyces lactis* bem como o terminador da transcrição do gene *LAC4* desta levedura, além do gene *URA3* destinado a promover integração do vetor no genoma da levedura hospedeira (no locus *URA3*) e servir como marca de seleção.

INTRODUÇÃO

A escolha de uma célula hospedeira bem como de um vetor de transformação (epissomal ou integrativo) com um promotor apropriado e uma marca de seleção eficiente são fatores determinantes para o sucesso de um processo biotecnológico envolvendo a produção de proteínas recombinantes, especialmente quando se objetiva aumentar a concentração e a pureza da proteína de interesse.

A levedura *Kluyveromyces marxianus* é uma candidata promissora em processos biotecnológicos, inclusive para expressão de proteínas recombinantes, devido a elevada taxa de conversão de substratos em massa celular, alto potencial para secreção de proteínas, termotolerância, que dispensa gastos de energia para resfriar dornas de fermentação. Além disso, esta levedura utiliza uma variedade de carboidratos como pentoses (xilose, arabinose), e dissacarídeos (lactose, melibiose), não assimiláveis por *Saccharomyces cerevisiae* (Fonseca et al., 2008; Lane & Morrissey, 2010). A levedura *K. marxianus* CCT 7735 (UFV-3) possui todas essas propriedades, entretanto, nenhum sistema de expressão foi até então construído para permitir explorar melhor sua capacidade como hospedeira para produção de proteínas heterólogas.

Tanto vetores de expressão episomais (que replicam extracromossalmente) ou integrativos (integram no genoma hospedeiro) são comumente usados em leveduras (van Ooyen et al., 2006). Por serem sistemas mais estáveis que os vetores episomais (Romanos et al., 1992), várias estratégias estão sendo desenvolvidas para aumentar o número de cópias de vetores integrativos nesses organismos. Uma alternativa descrita por Pecota et al. (2007) é o emprego de um cassete de seleção que permite a integração do gene de interesse em múltiplas cópias ao genoma de *K. marxianus*. Neste caso, o marcador de seleção auxotrófico *URA3* é flanqueado por duas repetições do gene *hisG* de *Salmonella*, formando o cassete *URA3 blaster* (UB).

A escolha do promotor para o desenvolvimento de um sistema de expressão é fundamental para o sucesso do processo e afeta consideravelmente os rendimentos do produto, uma vez que o reconhecimento e o tipo de regulação desse promotor no organismo hospedeiro são determinantes para sua atividade (Çelik & Çalik, 2012). A utilização de promotores transcricionais fortes é necessária para que ocorra um alto nível de expressão de proteínas recombinantes. A disponibilidade desses promotores para dirigir a expressão de genes heterólogos em vetores de expressão de leveduras facilita a produção de grandes quantidades da proteína alvo de forma mais viável que a maioria dos outros sistemas eucarióticos (Yin et al., 2007). Em estudos desenvolvidos no Laboratório de Fisiologia de Micro-organismos por nosso grupo de pesquisa, verificou-se que o cultivo da linhagem *K. marxianus* CCT 7735 (UFV-3) em meio Yeast Nitrogen Base (YNB) com lactose como única fonte de carbono em cultura contínua com velocidade de $0,1 \text{ h}^{-1}$, favoreceu a secreção da enzima endo-poligalacturonase

(Endo-PG, E.C.3.2.1.67) (artigo em preparação). O uso do promotor induzível deste gene num sistema de expressão para leveduras *K. marxianus* pode favorecer altos níveis de expressão do gene de interesse nas condições fisiológicas de cultivo pré-determinadas.

A marca de seleção é outra característica necessária para o bom funcionamento de um vetor, pois permite selecionar as células transformantes. Os marcadores auxotróficos são utilizados para complementação de mutações auxotróficas na linhagem hospedeira, sendo a maior parte desses genes envolvida na biossíntese de aminoácidos e nucleotídeos (Porro et al., 2005). Através da marcação auxotrófica, ao incluir um gene que complementa um ou mais genes mutados na linhagem auxotrófica no vetor de expressão, os clones recombinantes podem ser facilmente detectados em meio mínimo (Nasser et al., 2003). Os marcadores de seleção mais usados em leveduras são *LEU2* (β -isopropilmalato desidrogenase), *HIS3* (imidazol glicerol fosfato desidrogenase), *TRP1* (triptofano sintetase), *LYS2* (α -aminoadipato redutase) e *URA3* (orotidina 5'-fosfato descarboxilase) (Romanos et al., 1992; Stearns et al., 1990). *URA3* é particularmente versátil, pois apresenta método para contra-seleção do marcador, uma vez que a seleção positiva é por complementação de uma auxotrofia, e a seleção negativa é pela habilidade de crescer em meio contendo o composto que inibe o crescimento de células selvagens (Stearns et al., 1990). Este é o marcador de seleção mais utilizado para *K. marxianus*. No entanto, a obtenção de linhagens que não expressem o gene *URA3* ainda apresenta dificuldades. Comumente a obtenção deste tipo de mutantes na levedura *K. marxianus* ocorre na presença de ácido 5-fluorótico (5-FOA), uma substância tóxica que previne o crescimento de células selvagens mas possibilita o crescimento de células mutantes para o gene *URA3* (Boeke et al., 1984), ou ainda por meio de construções de cassetes contendo o gene *URA3* mutado inativo utilizados para inativação do gene no locus *URA3* por recombinação homóloga. A deleção gênica em fungos filamentosos tem sido conduzida de forma eficaz por uma técnica denominada *split-marker*, baseada em fusão por PCR, que permite uma montagem rápida das construções para deleção de genes. Mas em leveduras esta técnica tem sido explorada em *S. cerevisiae*, algumas espécies de *Candida*, e mais recentemente em *K. lactis* (Fairhead et al., 1996; Lee et al., 2006; Colombo et al., 2014).

No presente trabalho, objetivou-se a obtenção de mutantes *ura3⁻* da linhagem *K. marxianus* CCT 7735 (UFV-3), utilizando a técnica *split-marker*, para ser empregada como hospedeira para sistema de expressão cuja marca de seleção seja o gene *URA3*. Além disso, propusemos também a montagem de um vetor de expressão integrativo para *K. marxianus* CCT 7735 (UFV-3) isolando o promotor da Endo-PG desta levedura, e ligando o mesmo ao peptídeo sinal e ao terminador da transcrição do gene *LAC4* de *K. lactis*, e o gene *URA3* como marcador de seleção utilizando o cassete *URA3 blaster*.

MATERIAL E MÉTODOS

Micro-organismos, condições de cultivo e manutenção

A levedura *Kluyveromyces marxianus* CCT 7735 (UFV-3) (Silveira et al., 2005) foi utilizada neste trabalho para obtenção da linhagem mutante para o gene *URA3*. A linhagem foi mantida em meio YPD sólido [1 % (p/v) extrato de levedura, 2 % (p/v) peptona, 2 % (p/v) glicose, 2 % (p/v) ágar]. Meio YPD acrescido de 200 µg.mL⁻¹ de geneticina (G418 - SIGMA) foi utilizado para seleção de transformantes *K. marxianus* CCT 7735 (UFV-3) *ura3⁻*.

O meio S-SPO [0.25 % (p/v) extrato de levedura, 1.5 % (p/v) acetato de potássio, 0.05 % (p/v) dextrose, 0.2 % (p/v) *drop-out mix*] foi usado para induzir a esporulação da levedura crescendo sob agitação de 180 rpm a 37 °C durante 7 dias.

O meio 5-FOA [0.7 % (p/v) Yeast Nitrogen Base, 0.1 % (p/v) 5-fluorotic acid, 0.005 % (p/v) uracila, 2 % (p/v) glicose, 2 % (p/v) ágar] foi usado para repicar as colônias transformantes após seleção em meio YPD acrescido de 200 µg.mL⁻¹ de geneticina.

Células de *Escherichia coli* DH5αTM, utilizadas para manutenção e amplificação de DNA plasmidial, foram cultivadas em meio Luria-Bertani (LB), suplementado com 50 µg.mL⁻¹ de ampicilina quando apropriado, e incubadas a 37 °C sob agitação.

Para estocagem em longo prazo, as linhagens de micro-organismos foram mantidas a -80 °C em 20 % (v/v) glicerol.

Construção dos cassetes para inativação do gene *URA3*

Para construção de linhagem *K. marxianus ura3⁻* utilizou-se a técnica *split-marker* (Catlett *et al.*, 2002), em que se utilizam dois cassetes, cada um codificando uma metade diferente de um gene de resistência, e uma extremidade do gene de interesse ao qual se deseja inativar. Um cassete foi construído por fusão por Reação em Cadeia da Polimerase (PCR) da região 5' que flanqueia o gene *URA3* ao gene de resistência KanMX truncado na sua extremidade 5', e o outro cassete foi construído por fusão da extremidade 3' do gene de resistência KanMX truncado à região 3' que flanqueia o gene *URA3* (Figura 1). A região codificadora do gene *URA3* (Bergkamp *et al.*, 1993) (*GenBank* nº Z21934.1) foi utilizada para desenhar os oligonucleotídeos KmURA3FP1-F e KmURA3RP1-R para amplificar a parte referente à extremidade 5' do gene, e os oligonucleotídeos KmURA3FP2-F e KmURA3RP2-R para amplificar a parte referente a extremidade 3' do gene. Os fragmentos correspondentes a cada extremidade foram amplificados separadamente a partir do DNA genômico de *K. marxianus*. O gene KanMX, o qual confere resistência a geneticina, foi amplificado a partir do vetor pUG6 (Guldener *et al.*, 1996) com oligonucleotídeos PUGKam1-F e PUG6Kam-NLC37-R, para amplificar a extremidade 5', e PUG6KamNLC38-F e PUG6Kam2-R para amplificar a extremidade 3'. Nesse primeiro passo foram realizadas duas PCRs para obtenção de dois fragmentos diferentes para construção de cada cassete. Os parâmetros de PCR usados para amplificar cada fragmento foram: desnaturação inicial de 98 °C por 2 min; 30 ciclos de 98 °C por 10 s, 64 °C por 45 s, e 72 °C por 25 s; e um passo de extensão final de 72 °C por 5 min. Num segundo passo, os fragmentos gerados no primeiro passo foram unidos por fusão por PCR usando os pares de oligonucleotídeos KmURA3FP1-F / PUG6KamNLC37-R para gerar o fragmento de fusão I, e o par PUG6KamNLC38-F / KmURA3RP2-R para gerar o fragmento de fusão II (Figura 1). Os parâmetros de PCR usados nesse segundo passo foram temperatura de anelamento e extensão de: 59 °C por 45 s e 72 °C por 42 s para o fragmento de fusão I; e 62 °C por 45 s e 72 °C por 54 s para o fragmento de fusão II. Todos os oligonucleotídeos utilizados neste estudo estão descritos na Tabela 1. A Phusion Taq DNA polymerase usada para as reações de PCR foi comprada da empresa New England BioLabs.

Os dois fragmentos de fusão dividem uma região central de homologia do gene KanMX que se sobrepõe. Esses fragmentos foram usados para transformar a linhagem

K. marxianus CCT 7735 (UFV-3), para que, *in vivo*, ocorresse a reconstrução completa do gene de resistência através da recombinação dos cassetes truncados. Dessa forma, o cassette de deleção completo contendo a sequência inteira do gene KanMX flanqueada pelas extremidades 5' e 3' da região codificante do gene *URA3* recombinasse com o genoma da levedura no locus *URA3*, novamente por recombinação homóloga, para inativação deste gene pela integração do cassette (Figura 1).

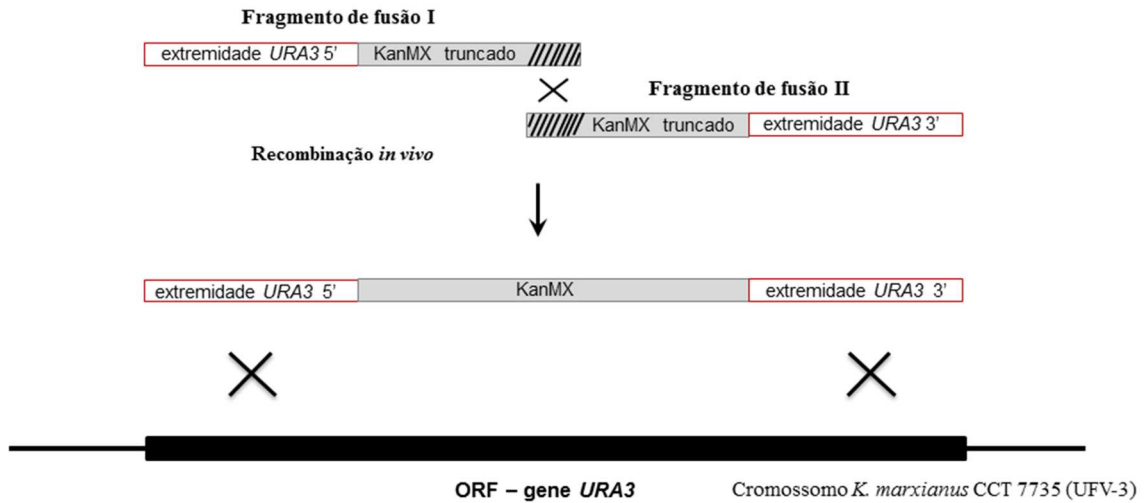


Figura 1. Técnica *split-marker* usada para inativação do gene *URA3* de *K. marxianus* CCT 7735 (UFV-3). Os fragmentos de fusão I e II amplificados separadamente foram usados para transformar *K. marxianus* onde ocorreu a recombinação entre os dois cassetes e reconstituição do gene KanMX afim de que o cassette completo recombinasse no locus do gene *URA3*. ORF: Open Reading Frame. Adaptado de Colombo et al., 2014.

A transformação de *K. marxianus* CCT 7735 (UFV-3) com os fragmentos de fusão I e II obtidos por PCR foi realizada de acordo Kooistra et al. (2004) e Sanchez et al (1993), com adaptações. As células foram inoculadas em cinco mL de meio YED [1 % (p/v) extrato de levedura, 1 % (p/v) glicose] a 37 °C sob agitação de 200 rpm por aproximadamente 14h. As culturas foram então inoculadas em 50 mL de meio YED para $DO_{600}=0,1$ por mL, e incubadas sob as mesmas condições do inóculo inicial até atingirem DO_{600} de aproximadamente 0,8. Em seguida as células foram centrifugadas por cinco minutos a 6000 rpm e lavadas com 20 mL de água destilada estéril.

Tabela 1. Oligonucleotídeos utilizados neste trabalho

Oligonucleotídeo	Sequência (5' - 3')
KmURA3FP1 – F	CTAAGAGCTTCTTCCCATTTCAG
KmURA3RP1 - R	AAGGGTTGTGCGACCTGCAGCCGAACATCAAGAGAAGCAC
KmURA3FP2 – F	GATATCAGATCCACTAGTGGCCGGTGTATACCGTATCGCCGA
KmURA3RP2 – R	GTGACGCAATGCCATTACTACTC
ExtraURA3 – F	GGTCTTCTGAGCTCATTATACCTC
PUGKam1 – F	GCTGCAGGTCGACAACCCTT
PUG6KamNLC37 – R	TCGTGATTGCGCCTGAGCGA
PUG6KamNLC38 – F	AGGATACAGTTCTCACATCACATCCGA
PUG6Kam2 – R	GCGGCCGCATAGGCCACTA
BamHI-KmPrPG – F	CGCGGATCCCGCGGTAGTATGGATGGTAGCTAGTAG
5GCG-KmPrPG – R	GGGGGCCCGCGGGGTTTTTTTTTTTTTTGGTTTCTGAGCTTAC
5CGC-CassetpKLAC – F	CCCCCGGGGGCCCCCAAATGAAATTCTCTACTATA
BamHI-CassetpKLAC – R	CGCGGATCCCGCCCCTATGCTCCAACGTATACAACATCG
BamHIKmPG2 – F	CACAGGATCCGGTAGTATGGATGGTAG
ComplementLAC1PG – R	GAGAATTTTCATCGCGGCTTTTTTTTTTTTTTTGGTTTCTGAGC
ComplemPGLAC12 – F	GCCGCGATGAAATTCTCTACTATA
SacII CassetLACIN – R	CCAACGTATACAACATCGCCGCGGTGTGGC

F: *forward*; R: *reverse*

O sedimento celular foi ressuscitado em 20 mL de tampão de pré-tratamento (YED, 25 mM DTT e 20 mM HEPES, pH 8.0) e incubado por 30 minutos a 30 °C sob agitação de 100 rpm. Após a incubação, as células foram centrifugadas e lavadas com 10 mL de água destilada. O sedimento celular foi ressuscitado em 0,3 mL de tampão de eletroporação (10 mM Tris-HCl, pH 7.5, 270 mM sacarose, e 1 mM MgCl₂). Foi retirada uma alíquota de 60 µL da suspensão de células e adicionado dois µg de cada fragmento de fusão I e II e 50 µg de DNA de esperma de Salmon (Invitrogen). Após incubação no gelo por 15 minutos, a mistura foi transferida para uma cubeta de eletroporação (2 mm). A eletroporação foi realizada no aparelho BIORAD *Gene Pulser* a 1 kV, 25 µF e 400 Ohm, e imediatamente após o pulso foi adicionado um mL de meio YED gelado. As células foram transferidas para tubos de ensaio estéreis e incubadas a 37 °C por duas horas sob agitação de 200 rpm. Após centrifugação, o sedimento celular foi ressuscitado em um mL água estéril e 0,3 mL plaqueadas em meio YPD sólido contendo 200 µg.mL⁻¹ de geneticina e as placas foram incubadas por três dias a 37 °C. As colônias transformantes foram repicadas em meio 5-FOA.

Confirmação das linhagens de *K. marxianus ura3⁻*

A confirmação da linhagem *K. marxianus ura3⁻* foi realizada por PCR com oligonucleotídeo desenhado para anelar fora do gene alvo (ExtraURA3-F) (Tabela 1) e confirmar a integração sítio-específica. Os parâmetros de PCR utilizados foram: 98 °C por 30 s para desnaturação inicial; 35 ciclos de 98 °C por 10 s, 61 °C por 45 s e 72 °C por 1 min e 20 s; 72 °C por 5 min para extensão final. Os produtos de PCR foram separados em gel de agarose 1 % (p/v) em tampão TBE (0,8 g.L⁻¹ Tris Base, 5,5 g.L⁻¹ ácido bórico e 4 mL EDTA 0,5 M pH 8,0) 1X a 80 Volts.

Análise da estabilidade mitótica das linhagens recombinantes

As células mutantes foram transferidas do meio YPD suplementado com 200 µg.mL⁻¹ de geneticina e foram repicadas três vezes em meio YPD não seletivo para avaliar a estabilidade dos transformantes. Após cada repicagem, as placas foram incubadas a 37 °C por 72 horas. Ao final das repicagens em meio não seletivo, os clones foram novamente cultivados em meio YPD suplementado com 200 µg.mL⁻¹ de geneticina e submetidos às mesmas condições de incubação.

Construção do vetor de expressão pKmLPG

Mapas físicos dos plasmídeos usados neste trabalho são mostrados na Figura 2. O plasmídeo pINGOXi (Rocha et al., 2010) foi utilizado como base para construção do vetor pKmLPG. Neste vetor, as sequências do promotor e da região terminadora utilizadas são as do gene *INU1* de *K. marxianus*, que dirigem a expressão do gene que codifica uma glicose oxidase (GOX). Essas regiões foram retiradas por hidrólise, primeiramente com a endonuclease de restrição *Bam*HI, e posteriormente, com *Bam*HI e *Sac*II. Nos dois casos, o fragmento resultante de aproximadamente 6.8 kb (denominado fragmento IN), contendo o gene de resistência à ampicilina (AmpR), origem de replicação em *E. coli*, e o gene *URA3* de *Saccharomyces cerevisiae* flanqueado por 2 sequências *hisG* (*direct repeat*), foi purificado usando o kit comercial de purificação de DNA a partir do gel Wizard SV Gel and PCR Clean-Up System da Promega (Wisconsin, EUA) para subsequente ligação do cassete linear PGp- α MF-MCS-TT_{LAC4} que foi construído.

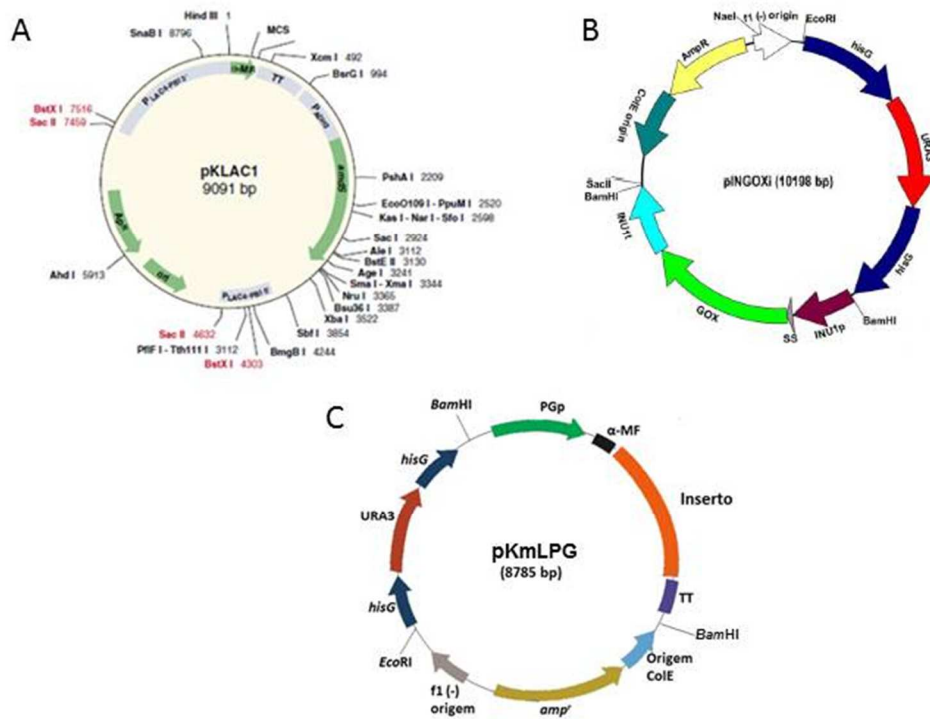


Figura 2. Mapas dos plasmídeos usados para construção do vetor de expressão pKmLPG: pKLAC1 (A), pINGOXi (B), e pKmLPG (C). pKmLPG foi criado pela ligação do fragmento denominado IN (liberado pela digestão de pINGOXi com *Bam*HI) com o cassete completo linear construído por fusão por PCR da região α -MF/MCS/TT_{LAC4} de pKLAC1 com a sequência do promotor do gene da endo-poligalacturonase (PGp) de *K. marxianus* CCT 7735 (UFV-3).

Para a construção do cassete linear completo PGp- α MF-MCS-TT_{LAC4} utilizou-se a técnica de fusão por PCR conforme descrito por Cha-aim et al. (2009), na qual dois ou mais fragmentos de DNA são fusionados via PCR usando trechos complementares simples, ou de sequências ricas em GC, como por exemplo, G₅C₅G₅. A sequência do promotor do gene da endo-poligalacturonase de *K. marxianus* CCT 7735 (UFV-3) (PGp), utilizada para construção do cassete completo linear, foi amplificada com oligonucleotídeos BamHI-KmPrPGF / 5GCG-KmPrPGR desenhados a partir da sequência genômica de *K. marxianus* CCT 7735 (UFV-3). Já a sequência codificadora do peptídeo sinal α -Mating Factor (α MF) de *K. lactis*, o sítio múltiplo de clonagem (MCS) e o terminador transcricional do gene *LAC4* de *K. lactis* (TT_{LAC4}) foram

amplificados a partir do DNA plasmidial do vetor pKLAC1 (Colussi & Taron, 2005) (Figura 2) com oligonucleotídeos 5CGC-CassetpKLACF / BamHI-CassetpKLACR. Esses dois fragmentos amplificados separadamente foram unidos por fusão por PCR. Os oligonucleotídeos 5GCG-KmPrPGR (*reverse* usado na amplificação do promotor PGp) e 5CGC-CassetpKLAC1F (*forward* usado na amplificação do α -MF do fragmento do pKLAC1) adicionam a cauda G₅C₅G₅ e C₅G₅C₅ em cada um dos fragmentos, respectivamente, unidos por fusão por PCR. Enquanto os oligonucleotídeos das extremidades (*forward* usado para amplificação do promotor PGp, e *reverse* usado para amplificação do terminador da transcrição TT_{LAC4}) adicionam sequências do sítio da endonuclease de restrição *Bam*HI (Tabela 1). O cassete linear completo resultante da fusão por PCR foi purificado, clonado no vetor pGEM®-T Easy (Promega, Madison, WI, USA), e utilizado para transformar células de *E. coli* DH5- α . O vetor pGEM contendo o cassete linear completo foi extraído e o cassete linear completo liberado pela reação de digestão com a enzima *Bam*HI, e ligado ao fragmento IN obtido da clivagem do vetor pINGOXi com a mesma enzima, para dar origem ao plasmídeo pKmLPG (Figura 2). As reações de ligação foram realizadas usando a enzima T4 DNA Ligase (Promega, Wisconsin, EUA). DNAs plasmidiais foram extraídos usando o kit Wizard Plus SV Minipreps DNA Purification System (Promega, Madison, WI, USA). Todas as manipulações de DNA foram realizadas de acordo com Sambrook et al. (1989).

O plasmídeo pKmLPG foi propagado em células de *E. coli*, e após extração, submetido à reação de clivagem com as endonucleases de restrição *Bam*HI, *Sac*II *Stu*I e *Not*I para confirmação da clonagem. Posteriormente, os fragmentos foram separados por eletroforese em gel de agarose 0,8 % (p/v), em tampão TBE 1X a 80 Volts. O vetor pKmLPG foi enviado para sequenciamento pela empresa Macrogen Inc., (Korea), utilizando a plataforma Illumina HiSeq2000, para confirmação da sequência inteira do vetor.

RESULTADOS E DISCUSSÃO

No presente trabalho objetivou-se a construção de um vetor de expressão integrativo para a levedura *K. marxianus* CCT 7735 (UFV-3) para expressão de proteínas recombinantes. Para isso, foi necessário primeiramente preparar a levedura

hospedeira, isto é, obter mutantes de *K. marxianus* CCT 7735 (UFV-3) com o gene *URA3* inativo, uma vez que o marcador de seleção do sistema de expressão a ser construído funcionaria por complementação sobre a auxotrofia em uracila.

Deleção do gene *URA3* da levedura *K. marxianus* CCT 7735 (UFV-3) e isolamento de mutantes *ura3*⁻

A inativação do gene *URA3* foi realizada utilizando-se a técnica *split-marker*. A partir do DNA genômico da linhagem CCT 7735 (UFV-3) foram amplificados um fragmento de 451 pares de bases (pb) da extremidade 5' e um fragmento de 616 pb da extremidade 3' do gene *URA3*. O gene de resistência a geneticina (KanMX) foi amplificado em duas partes, dando origem a dois fragmentos contendo o gene KanMX truncado. Em um destes fragmentos foi amplificada a região 5' contendo 914 pb do gene truncado, e no outro, a região 3' contendo 1.186 pb, e em ambos há uma região central de homologia que permite posterior recombinação entre eles. Por fusão por PCR, um fragmento de fusão de 1.365 pb (fragmento de fusão I) foi obtido utilizando os produtos da amplificação da extremidade 5' do gene *URA3* (451 pb) com a 5' do gene de resistência truncado (914 pb). Um segundo fragmento de fusão de 1.802 pb (fragmento de fusão II) foi obtido utilizando os produtos da amplificação da extremidade 3' (616 pb) do gene *URA3* com o 3' do gene KanMX truncado (1.186 pb). (Figura 3)

Os fragmentos de fusão I e II gerados foram os utilizados para transformar a linhagem *K. marxianus* CCT 7735 (UFV-3) e deletar e/ou inativar o gene *URA3* desta linhagem. O interesse em deletar/inativar o gene *URA3* é justificado pelo seu comum uso como marca de seleção em plasmídeos para leveduras. Comumente a mutação deste gene é obtida pelo uso do meio 5-FOA (ácido 5-fluorótico), que em concentrações estabelecidas, previne o crescimento de células de levedura selvagens e permite o crescimento de células que tenham uma mutação aleatória no gene *URA3* (Boeke et al., 1984). 5-FOA é um análogo de pirimidina, e células selvagens convertem o 5-FOA a 5-fluoro-orotidina monofosfato. Em seguida, é formado o composto denominado fluorodeoxiuridina que é tóxico para a célula por ser potente inibidor da timidilato sintase, enzima que converte deoxiuridina monofosfato (dUMP) à timidina monofosfato (dTMP), precursor do trifosfato de timidina essencial para a síntese de DNA. Portanto, células *ura3*⁻ podem ser selecionadas (seleção negativa) por sua

resistência a 5-FOA ao crescer no meio contendo essa substância tóxica para célula (Nasser et al., 2003).

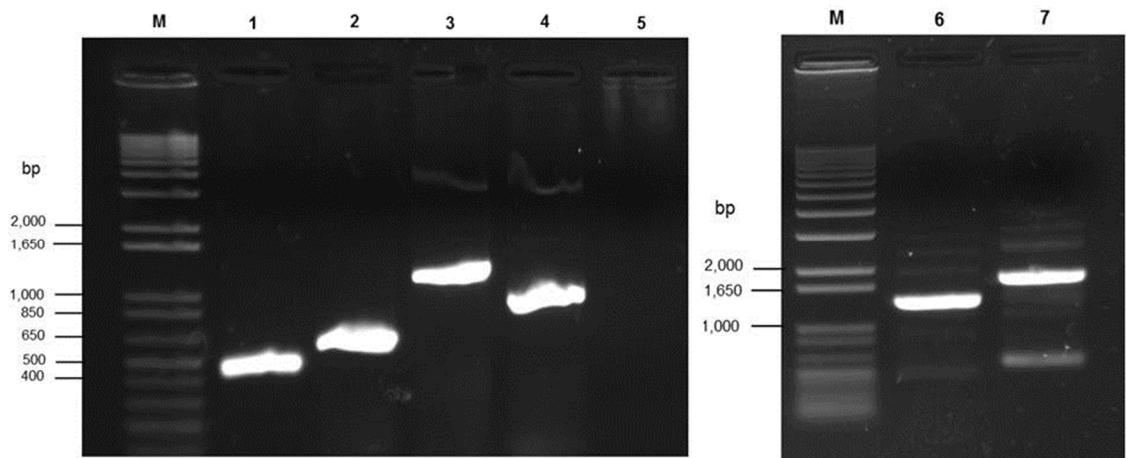


Figura 3. Perfil dos fragmentos de DNA obtidos pela técnica *split-marker* utilizando PCR. **M**, marcador 1 kb DNA *ladder plus* (Invitrogen). **1** e **2**, produtos da amplificação das extremidades 5' (451 pb) e 3' (616 pb), respectivamente, flanqueadoras do gene *URA3*, a partir do DNA genômico de *K. marxianus*. **3** e **4**, produtos da amplificação do gene de resistência a geneticina (G418) a partir do pUG6, contendo cada fragmento uma região central de sobreposição (1.186 e 914 pb, respectivamente). **5**, controle negativo sem DNA **6** e **7**, fragmentos resultantes da fusão por PCR dos fragmentos **1** e **4** (1.365 pb), e **2** e **3** (1.802 pb), respectivamente.

Neste trabalho, não houve sucesso na obtenção de mutantes *ura3⁻* por essa estratégia, mesmo submetendo as células à radiação ultra violeta por 4, 6 e 8 minutos com posterior plaqueamento em meio 5-FOA. Pecota et al. (2007) descreveram outro método para obtenção de linhagens auxotróficas pela inativação do gene *URA3* usando um cassete contendo um fragmento homólogo modificado de *URA3* e, portanto, inativo. A célula de levedura era transformada com esse único cassete *URA3Δ* inteiro e a integração no *locus* do gene ocorria por recombinação homóloga. No entanto, a desvantagem desta técnica é o cassete utilizado ser longo e dificultar o processo de transformação, diminuindo assim as chances de se obter um mutante. Em decorrência desses resultados e do sucesso obtido pela mutação do gene *KU80* de *K. lactis* pela técnica *split-marker* por nossa equipe de trabalho em estudo anterior

(Colombo et al., 2014), optamos pela obtenção de mutantes *ura3⁻* da linhagem de *K. marxianus* CCT 7735 (UFV-3) por *split-marker*. Inicialmente, após a primeira transformação da linhagem CCT 7735 (UFV-3) com os dois fragmentos de fusão descritos acima, foram obtidos apenas dois transformantes.

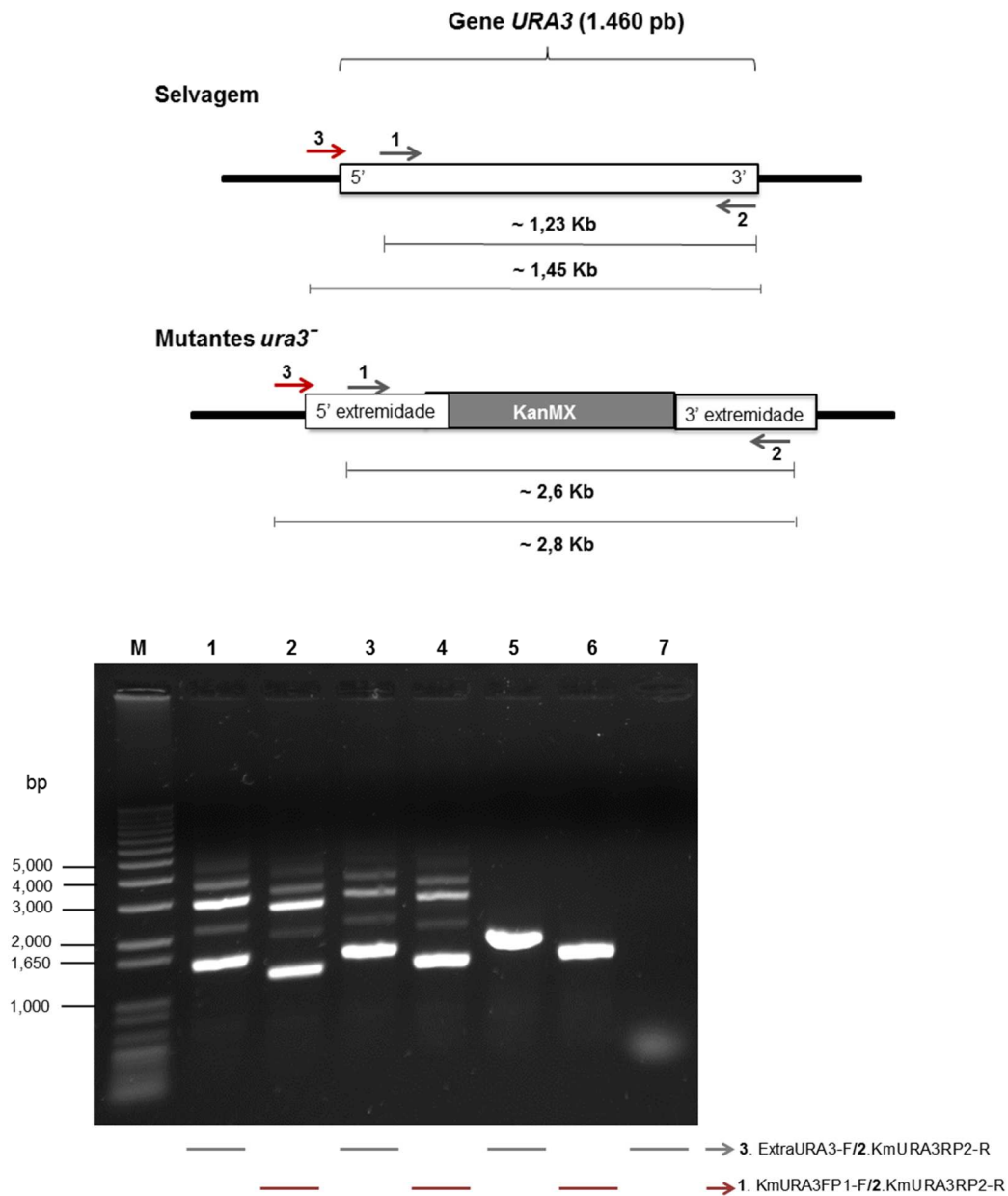


Figura 4. Produtos da amplificação do gene *URA3* do DNA genômico de *K. marxianus* utilizando os oligonucleotídeos ExtraURA3-F (externo à região codificante) + KmURA3RP2-R, e KmURA3FP1-F + KmURA3RP2-R, usados no *split-marker*.

Fragmentos resultantes da amplificação pelo primeiro par de primers são indicados pelas barras cinzas, e os resultantes da amplificação pelo segundo par de primers são indicados pelas barras vermelhas. **M**, marcador 1 kb DNA *ladder plus* (Invitrogen). **1** e **2**, linhagem transformante 1. **3** e **4**, linhagem transformante 2. **5** e **6**, controle *K. marxianus* CCT 7735 (UFV-3) selvagem. **7**, controle negativo (sem DNA). Esquema ilustrativo adaptado de Colombo et al., 2014.

Para ambos foi possível identificar a integração do cassete em apenas um dos alelos do gene *URA3* (Figura 4). Isso pode ser observado pela presença dos produtos de amplificação de mesmo tamanho dos controles, realizados com o DNA da linhagem de *K. marxianus* CCT 7735 (UFV-3) selvagem.

Dentre as características que dividem *K. marxianus* e *K. lactis* em espécies diferentes está o sistema de compatibilidade de *mating*, pois as linhagens de *K. lactis* são principalmente heterotáticas (Belloch et al., 2002; Butler et al., 2004), ou seja, o estado haplóide é predominante durante o crescimento vegetativo, enquanto quase todas as linhagens de *K. marxianus* são homotáticas (Iborra, 1993), apresentando durante o crescimento vegetativo o estado diplóide, ficando o estado haplóide restrito apenas para os esporos. Diante disso, nós induzimos a levedura *K. marxianus* UFV-3 mutante para um alelo, obtida na primeira transformação, à esporulação em meio específico, e usamos as células cultivadas por sete dias no processo de transformação por eletroporação. Como a levedura já apresentava um cassete integrado em um alelo, as células eram capazes de crescer em meio seletivo contendo geneticina. Nossa estratégia para selecionar os mutantes foi repicar as colônias transformantes em meio 5-FOA, pois apenas transformantes contendo os dois alelos mutados poderiam crescer neste meio. De 88 transformantes obtidos após a transformação, apenas 6 cresceram após repicagem no meio contendo 5-FOA (Figura 5), e destes, a eficiência de integração gene-específica foi de 100 %, ou seja, todos tiveram o gene *URA3* mutado.

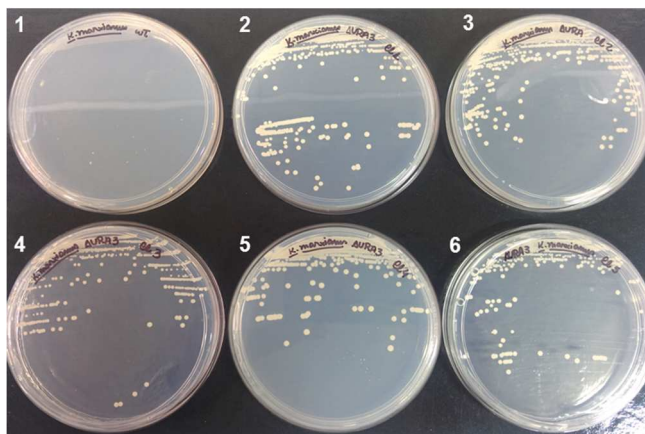


Figura 5. Crescimento celular de transformantes de *K. marxianus* UFV-3 em meio 5-FOA para confirmação da mutação nos dois alelos do gene *URA3*. 1, Controle negativo: Linhagem de *K. marxianus* selvagem. 2 a 7, transformantes colônias 1 a 6.

A Figura 6 mostra a confirmação por PCR das 6 mutantes *ura3⁻* para os dois alelos. O tamanho do gene *URA3* na linhagem parental é de 1.460 pb. O oligonucleotídeo ExtraURA3-F (*forward*) anela externo à região codificante, e amplifica na linhagem selvagem um fragmento de aproximadamente 1,45 kb (Figura 4 e 5). A construção de deleção contendo o gene KanMX e as extremidades da região codificante do gene *URA3* apresenta um tamanho de 2,6 kb, resultando na amplificação de um fragmento de cerca de 2,8 kb ao se utilizar, para os transformantes o mesmo oligonucleotídeo externo (Figura 4 e 6). Esse resultado confirmou, portanto, a integração gene-específica e a inativação do gene *URA3*.

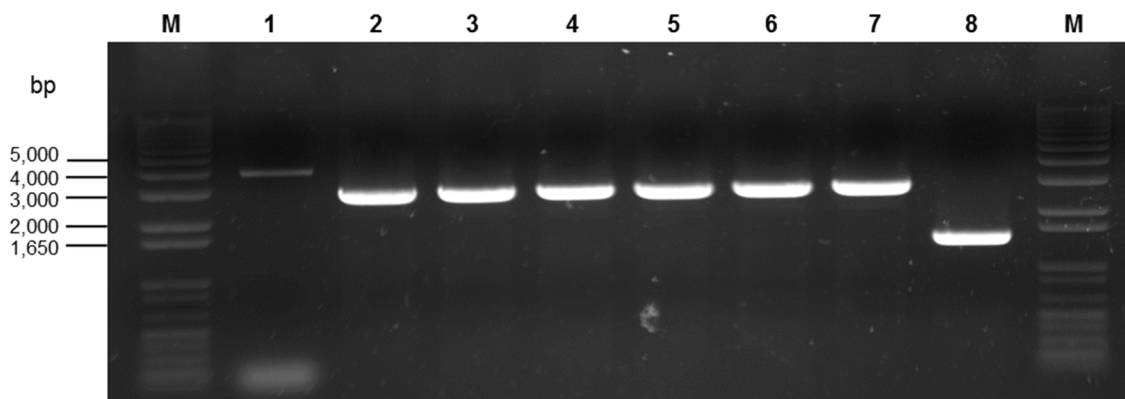


Figura 6. Confirmação de mutantes *ura3⁻* de *K. marxianus* UFV-3 pela amplificação do gene a partir do DNA genômico dos transformantes com oligonucleotídeos ExtraURA3-F e KmURA3RP2-R. **M**, marcador 1 kb DNA ladder *plus* (Invitrogen). **1**, Controle negativo (sem DNA). **2 a 7**, linhagens transformantes *ura3⁻*. **8**, linhagem selvagem.

Neste trabalho foi possível determinar que a técnica *split-marker* constitui um sistema de deleção gênica eficaz para a levedura *K. marxianus* UFV-3 quando são utilizadas na construção dos cassetes de deleção sequências do gene de interesse acima de 400 e 600 pb flanqueando o gene marcador de resistência. Esse resultado é importante, pois *K. marxianus*, assim como *K. lactis*, apresenta baixa eficiência para efetuar recombinação homóloga. Existe uma diferença entre os diversos organismos quanto à eficiência de integração de um fragmento de DNA no gene alvo (Kooistra et al., 2004). Isso acontece porque alguns organismos apresentam a via de junção de extremidades não-homólogas (*non-homologous end joining*, NHEJ) (que resulta na integração aleatória do DNA exógeno) mais ativa que a via de recombinação homóloga (Kegel et al., 2006). Foi demonstrado em uma linhagem selvagem de *K. lactis* que a eficiência de integração no gene alvo varia de 0 % com 50 pb a 88 % com 600 pb, enquanto a eficiência de integração no gene alvo em *S. cerevisiae* revela um mecanismo de recombinação homóloga com quase 100 % de integração mesmo quando o fragmento de DNA introduzido na célula apresenta apenas 40 pb de homologia em cada extremidade com o gene alvo (Kooistra et al., 2004). Em relação à *K. marxianus*, Ribeiro et al. (2007) mostraram uma eficiência de inativação do gene *LAC4* de apenas 34 % por recombinação homóloga utilizando um sistema Cre-*LoxP* contendo 723 e 713 bases flanqueando o gene marcador de seleção para a linhagem

CBS 6556. Neste estudo foi possível identificar em todos os transformantes inativação do gene *URA3* por recombinação homóloga utilizando 451 e 616 pb flanqueando o gene de resistência KanMX, provando esta técnica como ótima estratégia para inativação ou deleção gênica nesta levedura.

Construção do vetor de expressão pKmLPG

A montagem do vetor de expressão pKmLPG iniciou-se pela digestão do vetor pINGOXi com *Bam*HI para liberar o fragmento IN, de aproximadamente 6.8 kb, que contém o gene de resistência à ampicilina (AmpR), origem de replicação em *E. coli*, e o gene *URA3* de *S. cerevisiae* flanqueado por 2 sequências *hisG*. O objetivo ao se utilizar esse fragmento IN é o princípio de integração que teríamos com o nosso vetor: a recombinação do gene *URA3* presente no plasmídeo com o do genoma de *K. marxianus*, e a presença à montante e à jusante do gene *URA3* de sequências *hisG*, que possibilitariam a integração de várias cópias *in tandem* (Alani et al., 1987). Desta forma, o vetor iria conferir à levedura *K. marxianus* UFV-3 auxotrófica em uracila o fenótipo *ura+*.

A escolha do promotor para o vetor pKmLPG foi determinada em decorrência de estudos anteriores desenvolvidos no Laboratório de Fisiologia do Departamento de Microbiologia da UFV pelos quais verificou-se que o cultivo de *K. marxianus* UFV-3 em YNB contendo 2 % (p/v) de lactose em cultura contínua com taxa de diluição de 0,1 h⁻¹ favoreceu a secreção da enzima endo-poligalacturonase (Endo-PG, E.C.3.2.1.67). Uma vez que o genoma da *K. marxianus* UFV-3 foi recentemente sequenciado e anotado (Silveira et al., 2014), foi possível identificar a sequência do gene para Endo-PG (localizados na fita negativa - orientação 3'→5' - do cromossomo I), incluindo o promotor, e desenhar oligonucleotídeos para amplificação da sequência do promotor deste gene. Os oligonucleotídeos desenhados amplificam um fragmento de aproximadamente 1.035 pb (Figura 7A). Essa região apresenta, além da sequência TATA BOX padrão, 50 sítios putativos de ligação a fatores de transcrição. O fragmento do promotor amplificado foi fusionado por PCR ao fragmento de DNA de aproximadamente 1.000 pb obtido pela amplificação das sequências α MF-MCS-TT_{LAC4} do plasmídeo pKLAC1 (Figura 7A) para formar o cassete linear completo PGp- α MF-MCS-TT_{LAC4} contendo cerca de 2.040 pb (Figura 7B).

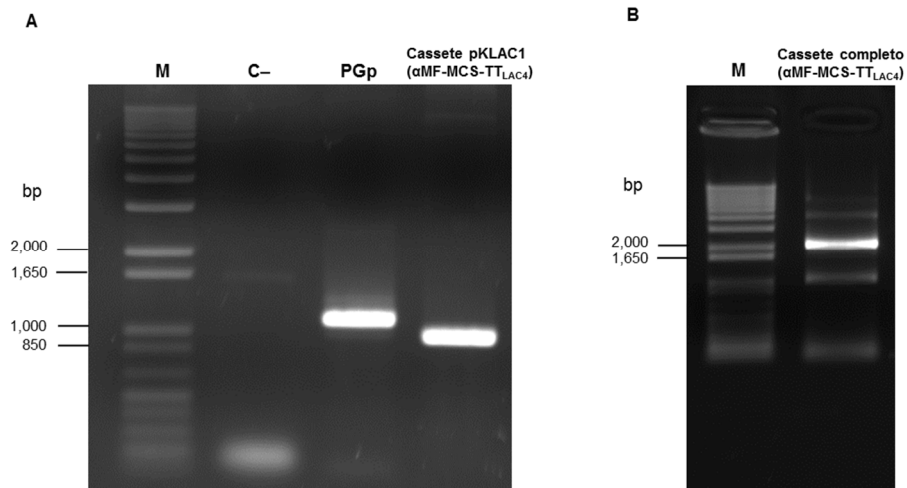


Figura 7. Amplificação da sequência de DNA referente ao promotor do gene da Endo-PG e cassete α MF-MCS- TT_{LAC4} do pKLAC1 (A), e cassete completo PGP- α MF-MCS- TT_{LAC4} (B) obtido pela fusão por PCR dos dois primeiros fragmentos. **M**, marcador 1 kb DNA *ladder plus* (Invitrogen).

O cassete completo purificado a partir do gel foi ligado ao vetor pGEM-T Easy e posteriormente digerido com a enzima *Bam*HI, purificado e utilizado para ser ligado ao fragmento IN descrito anteriormente desfosforilado, dando origem ao vetor pKmLPG. Após transformação das células de *E. coli*, 10 clones foram analisados (de 30 colônias transformantes) para confirmação da correta construção do vetor. Esses clones tiveram o DNA plasmidial extraído e hidrolisado com *Bam*HI. Contudo, os fragmentos liberados após reação de hidrólise não apresentavam o tamanho esperado do fragmento de cerca de 2 kb referente ao cassete PGP- α MF-MCS- TT_{LAC4} que deveria ser liberado (Figura 8). O fragmento IN correspondia ao tamanho correto, mas o fragmento do cassete era maior do que o esperado para confirmação da correta construção. Os demais clones foram testados e apresentaram o mesmo resultado.

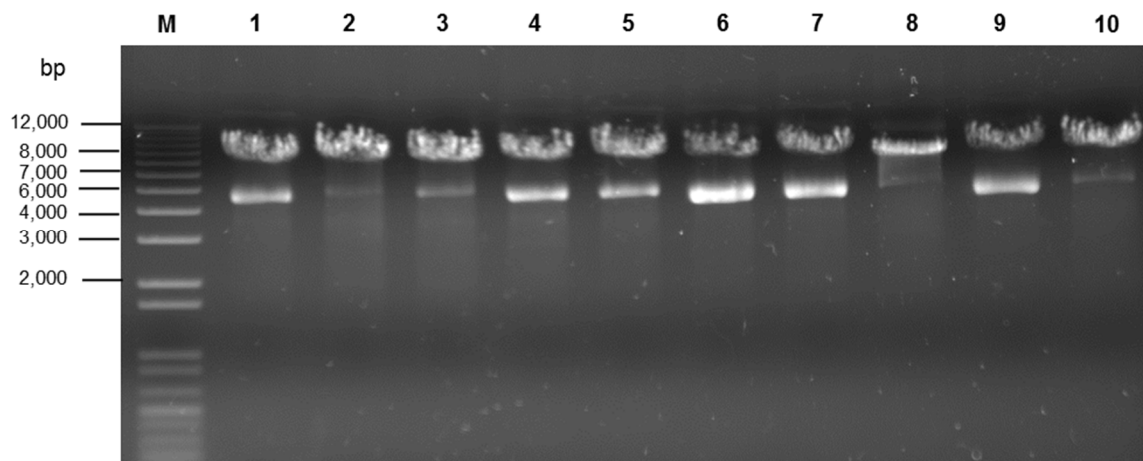


Figura 8. Hidrólise do vetor pKmLPG com a endonuclease de restrição *Bam*HI para confirmação da correta construção do vetor. **1 a 10**, DNA plasmidial extraído de 10 transformantes de *E. coli*. **M**, marcador 1 kb DNA *ladder plus* (Invitrogen).

Os experimentos foram repetidos inúmeras vezes, utilizando diferentes proporções de fragmento IN desfosforilado:cassete completo para a reação de ligação. A transformação sempre ocorria corretamente, mas a checagem da construção resultava em fragmentos de tamanhos inesperados. Como o fragmento IN e o cassete completo apresentavam sítios apenas para a enzima *Bam*HI, o experimento foi refeito utilizando duas enzimas no processo: *Sac*II e *Bam*HI. Além disso, ao invés de utilizar a sequência C₅G₅C₅ como estratégia de região de homologia para a técnica de fusão por PCR do promotor PGp à sequência α MF-MCS-TT_{LAC4} do pKLAC1, foram utilizadas sequências dos próprios fragmentos como homologia nos novos oligonucleotídeos desenhados. Novamente os tamanhos dos fragmentos resultantes da análise de restrição com as enzimas *Sac*II e *Bam*HI resultaram em fragmentos diferentes do esperado. Um fator determinante para a construção de um vetor é o conhecimento das sequências utilizadas. O vetor pINGOXi (Rocha et al., 2010) utilizado como base neste trabalho foi gentilmente cedido pelo Departamento de Biología Celular e Molecular da Faculdade de Ciências da Universidade da Caruña na Espanha, mas não apresenta sequência depositada em bancos de dados. Nem mesmo o vetor pNADFL11 (Pecota et al., 2007), do qual foi derivado o vetor pINGOXi, não apresenta sequência disponível. A ausência dessa informação foi prejudicial à construção do vetor pKmLPG, uma vez que somente a sequência do cassete completo linear PGp- α MF-MCS-TT_{LAC4}

era conhecida. Além disso, de acordo com dados do vetor pINGOXi (Rocha et al., 2010), o fragmento IN, de acordo com o plasmídeo pNADFL11 deveria conter aproximadamente 6.800 pb após excisão do promotor *INU1*, gene *GOX* e *INU1tt* (que totalizam aproximadamente 3.398 pb) com *Bam*HI ou *Bam*HI/*Sac*II. No entanto, a hidrólise de pINGOXi com essas enzimas resulta em um fragmento de cerca de 8 kb, e outro de quase 3 kb, o que dificulta ainda mais a compreensão desta sequência (Figura 9).

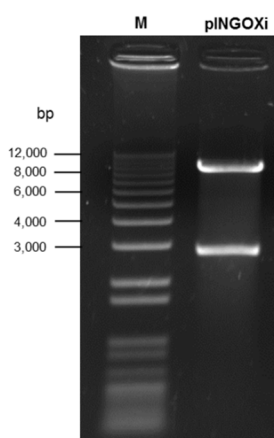


Figura 9. Hidrólise do vetor de expressão pINGOXi com a endonuclease de restrição *Bam*HI.

Na tentativa de conhecer melhor a sequência do plasmídeo pINGOXi, realizamos um perfil de restrição do vetor pINGOXi com as seguintes endonucleases: *Xho*I, *Bgl*II, *Hae*III, *Not*I, *Kpn*I e *Stu*I. Por esta análise foi possível perceber que as enzimas *Not*I e *Stu*I não clivam o vetor, e que *Bgl*II não cliva o fragmento IN, e essa informação é importante uma vez que no sítio múltiplo de clonagem do nosso vetor há sítios para essas enzimas. O vetor pKmLPG foi então enviado para sequenciamento para confirmar se este está apto para ser usado como sistema de expressão na levedura *K. marxianus* CCT 7735 (UFV-3).

CONCLUSÕES E PERSPECTIVAS

Neste trabalho foi obtida uma linhagem de *K. marxianus* CCT 7735 (UFV-3) mutante para o gene *URA3* utilizando a técnica *split-marker*. A inativação do gene

URA3, o qual codifica para a síntese de uma enzima essencial presente na via de biossíntese da uracila, ocorreu em duas tentativas, pois em decorrência da diploidia da linhagem de *K. marxianus* utilizada, na primeira tentativa apenas um alelo do gene foi mutado. Contudo, a técnica, bem como a quantidade de bases de homologia ao gene *URA3* usadas para a recombinação homóloga dos cassetes no genoma da levedura foi realizada com sucesso, indicando que assim como para *K. lactis*, a estratégia *split-marker* constitui um sistema de deleção gênica capaz de funcionar na levedura *K. marxianus* quando se utiliza na construção dos cassetes de deleção sequências do gene de interesse acima de 400 pb flanqueando o gene marcador de resistência. A linhagem hospedeira *K. marxianus* CCT 7735 (UFV-3) *ura3⁻* está pronta para ser usada em estudos cuja seleção será feita por complementação da referida mutação auxotrófica.

Além disso, foi construído um sistema de expressão denominado pKmLPG para ser usado em *K. marxianus* CCT 7735 (UFV-3) utilizando o promotor do gene da Endo-PG, que em condições estabelecidas de cultivo para esta levedura mostrou forte indução, observada pela síntese e secreção da enzima Endo-PG. Uma característica importante do vetor é que possui o gene *URA3* como marcador de seleção, para que possa ser utilizado na linhagem mutante auxotrófica *K. marxianus* CCT 7735 (UFV-3) *ura3⁻* como hospedeira para expressão de genes heterólogos. A presença de um promotor fortemente induzido e de uma marca de seleção auxotrófica *ura3⁻* no sistema de expressão construído são determinantes para o estabelecimento deste sistema de expressão para clonagem de diferentes genes heterólogos na levedura *K. marxianus* CCT 7735 (UFV-3) *ura3⁻*. O vetor pKmLPG está sendo sequenciado para conhecimento da sequência referente ao fragmento IN provindo do vetor pINGOXi, bem como para confirmação de sua correta construção, a qual não foi possível ser confirmada por hidrólise com endonucleases de restrição.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Alani, E., L. Cao, et al. 1987. A method for gene disruption that allows repeated use of *URA3* selection in the construction of multiply disrupted yeast strains. *Genetics*. 116:541-5.
2. Belloch, C., T. Fernández-Espinar, A. Querol, M.D. García, E. Barrio. 2002. An analysis of inter- and intraspecific genetic variabilities in the *Kluyveromyces marxianus* group of yeast species for the reconsideration of the *K. lactis* taxon. *Yeast*. 19:257-268.
3. Bergkamp, R.J., Geerse, R.H., Verbakel, J.M., Planta, R.J. 1993. Cloning and sequencing of the *URA3* gene of *Kluyveromyces marxianus* CBS 6556. *Yeast*. 9:677-681.
4. Boeke, J.D., F. LaCroute, G.R. Fink. 1984. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Molecular and General Genetics*. 197:345-346.
5. Butler, G., C. Kenny, A. Fagan, C. Kurischko, C. Gaillardin, K.H. Wolfe. 2004. Evolution of the *MAT* locus and its Ho endonuclease in yeast species. *PNAS*. 101:1632-1637.
6. Catlett, N.L., Lee, B.N., Yoder, O.C., Turgeon, B. G. 2003. Split-marker recombination for efficient targeted deletion of fungal genes. *Fungal Genetics Newsletter*. 50:9–11.
7. Cha-aim, K., Fukunaga, T., Hoshida, H. & Akada, R. 2009. Realible fusion PCR mediated by GC-rich overlap sequences. *Applied and Microbial Biotechnology*. 67:364-369.
8. Colombo, L.T., Rosa, J.C.R, Bragança, C.R.S., Ignacchiti, R.P., Alvim, M.C.T., Silveira, W.B., de Queiroz, M.V., Bazzolli, D.M.S., Passos, F.M.L. 2014. Construction of a *Kluyveromyces lactis ku80⁻* Host Strain for Recombinant Protein Production: Extracellular Secretion of Pectin Lyase and a Streptavidin–Pectin Lyase Chimera. *Molecular Microbiology*. 56:319-28.
9. Colussi, P.A. & Taron, C.H. 2005. *Kluyveromyces lactis LAC4* Promoter Variants That Lack Function in Bacteria but Retain Full Function in *K. lactis*. *Applied and Environmental Microbiology*, 71, 7092-7098.
10. Çelik, E. & P. Çalık. 2012. Production of recombinant proteins by yeast cells. *Biotechnology Advances*. 30:1108-1118.

11. Fairhead, C., B. Llorente, F. Denis, M. Soler & B. Dujon. 1996. New Vectors for Combinatorial Deletions in Yeast Chromosomes and for Gap-repair Cloning using "Split-marker Recombination". *Yeast*. 12:1439-1457.
12. Fonseca, G.G., E. Heinzle, C. Wittmann, A.K. Gombert. 2008. The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Applied Microbiology and Biotechnology*. 79:339-54.
13. Guldener, U., Heck, S., Fielder, T., Beinhauer, J., Hegemann, J. H. 1996. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Research*. 24:2519–2524.
14. Iborra, F. 1993. High efficiency transformation of *Kluyveromyces marxianus* by a replicative plasmid. *Current Genetics*. 24:181-183.
15. Kegel, A., P. Martinez, S.D. Carter, S.U. Åström. 2006. Genome wide distribution of illegitimate recombination events in *Kluyveromyces lactis*. *Nucleic Acids Research*. 34:1633-1645.
16. Kooistra, R., Hooykaas, P. J. J., Steensma, H. Y. 2004. Efficient gene targeting in *Kluyveromyces lactis*. *Yeast*. 21:781–792.
17. Lane, M.M. & J.P. Morrissey. 2010. *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. *Fungal Biology Reviews*. 24:17-26.
18. Lee, T.H., M. Kim & J. Seo. 2006. Development of Reusable Split URA3-Marked Knockout Vectors for *Saccharomyces cerevisiae*. *Journal Microbiology Biotechnology*. 16:979-982.
19. Nassser, M.W., V. Pooja, M.Z. Abdin, S.K. Jain. 2003. Evaluation of Yeast as an Expression System. *Indian Journal of Biotechnology*. 2:477-493.
20. Pecota, D.C., V. Rajgarhia, et al. 2007. Sequential gene integration for the engineering of *Kluyveromyces marxianus*. *Journal of Biotechnology*. 127:408-16.
21. Porro, D., M. Sauer, P. Branduardi, D. Mattanovich. 2005. Recombinant Protein Production in Yeasts. *Molecular Biotechnology*. 31:245-259.
22. Ribeiro, O., A.K. Gombert, et al. 2007. Application of the Cre-loxP system for multiple gene disruption in the yeast *Kluyveromyces marxianus*. *Journal of Biotechnology*. 131:20-6.
23. Rocha, S.N., Abrahão-Neto, J., Cerdán, M.E., González-Sico, M.I., Gombert, A.K. 2010. Heterologous expression of glucose oxidase in the yeast *Kluyveromyces marxianus*. *Microbial Cell Factories*. 9:1-12.

24. Romanos, M.A., C.A. Scorer, J.F. Clare. 1992. Foreign gene expression in yeast: a review. *Yeast*. 8:423–488.
25. Sambrook, J., Fritsch, E.F., Maniatis, T. 1989. *Molecular cloning: a laboratory manual* (2nd ed.). Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
26. Sanchez M., Iglesias, F.J., Santamaria, C., Dominguezl, A. 1993. Transformation of *Kluyveromyces lactis* by electroporation. *Applied and Environmental Microbiology*. 59:2087–2092.
27. Silveira, W.B., Diniz, R.H.S., Cerdan, M.E., Gonzalez-Siso, M.I., et al. 2014. Genomic Sequence of the Yeast *Kluyveromyces marxianus* CCT 7735 (UFV-3), a Highly Lactose-Fermenting Yeast Isolated from the Brazilian Dairy Industry. *Genome Announcements*. 2:e01136-14-e01136-14.
28. Silveira, W.B., F.J.V. Passos, H.C. Mantovani & F.M.L. Passos. 2005. Ethanol production from cheese whey permeate by *Kluyveromyces marxianus* UFV-3: a flux analysis of oxido-reductive metabolism as a function of lactose concentration and oxygen levels. *Enzyme and Microbial Technology*. 36:930-6.
29. Stearns, T., H. Ma, D. Botstein. 1990. Manipulating Yeast Genome Using Plasmid Vectors. *Methods in Enzymology*. 185:280-297.
30. van Ooyen, A.J.J., P. Dekker, et al. 2006. Heterologous protein production in the yeast *Kluyveromyces lactis*. *FEMS Yeast Research*. 6:381-392.
31. Yin, J., G. Li, X. Ren, G. Herrler. 2007. Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *Journal of Biotechnology*. 127:335-347.

CAPÍTULO 3

EXPRESSÃO HETERÓLOGA DE GENES DE B-GLICOSIDASES DE FONTES MICROBIANAS TERMOFÍLICAS NA LEVEDURA *Kluyveromyces marxianus* CCT 7735 (UFV-3)

RESUMO

β -glicosidases são enzimas que realizam a hidrólise de biomassa celulósica. Neste trabalho, três genes que codificam para a enzima β -glicosidase isolados de uma biblioteca metagenômica obtida a partir de um consórcio microbiano termofílico foram clonados e expressos na levedura *K. marxianus* CCT 7735 (UFV-3). Um dos genes (P3Gli) clonados codifica para uma β -glicosidase pertencente à família 1 de glicosil hidrolases (GH1) com 100 % de identidade com a enzima de *Clostridium thermocellum*, enquanto os outros dois genes (P8Gli e P9Gli) pertencem à família GH3 e dividem apenas 53 e 43 % de identidade com enzimas similares presentes nos bancos de dados, possivelmente representando novas β -glicosidases. Células transformantes *K. marxianus* UFV3/P3Gli, UFV3/P8Gli e UFV3/P9Gli foram selecionadas em meio contendo o antibiótico geneticina, e a integração gene-específica dos cassetes de expressão lineares contendo os genes de recombinantes de β -glicosidase no genoma da levedura foi confirmada por PCR pela obtenção do fragmento de DNA de 2,3 kb utilizando oligonucleotídeos iniciadores específicos. Um fragmento maior de aproximadamente 2,8 kb obtido para todos os clones analisados confirma também a integração de mais de uma cópia do cassete de expressão. Durante a realização de um teste enzimático dos transformantes em meio sólido observou-se a presença de atividade extracelular de β -glicosidase na levedura controle *K. marxianus* CCT 7735 (UFV-3) selvagem. Com o genoma desta levedura anotado, foi possível identificar o gene, e juntamente com os demais genes P3Gli, P8Gli e P9Gli, foram preditas as estruturas tridimensionais dessas proteínas. A análise da expressão dos genes recombinantes foi realizada por atividade enzimática de β -glicosidase a temperatura de 50 °C na tentativa de reduzir a interferência da enzima endógena. Atividade enzimática foi detectada tanto no sobrenadante das culturas quanto no espaço periplasmático. Apenas um dos clones analisados (UFV3/P9Gli-2) não apresentou diferença significativa em relação à cepa controle nas análises de atividade do sobrenadante das culturas. Já para a atividade presente no periplasma, apenas os clones UFV3/P8Gli-3 e UFV3/P8Gli-1 apresentaram diferença significativa em relação à cepa controle. O valor de maior atividade enzimática alcançado no sobrenadante das culturas foi 17 vezes maior que a atividade da cepa controle CCT 7735 (UFV-3), enquanto para as amostras ensaiadas do periplasma o maior valor de atividade apresentado foi 2,5 vezes maior que a cepa controle CCT 7735 (UFV-3). Os resultados obtidos neste trabalho indicam

que *K. marxianus* CCT 7735 (UFV-3) é capaz de expressar heterologicamente as enzimas β -glicosidases.

INTRODUÇÃO

β -glicosidases (EC 3.2.1.21) catalisam a hidrólise de ligações β -1,4-glicosídicas com liberação de glicose a partir de extremidades não-redutoras (Li et al., 2013). Essas enzimas em sinergia com endoglucanases e exoglucanases (celobiohidrolases), que realizam a hidrólise das cadeias de celulose em celobiose, convertem esse substrato em glicose (Rubin, 2008; Chir et al., 2011; Li et al., 2013). Entre as celulases, β -glicosidase é essencial para o passo final de sacarificação da celulose, pois reduz a inibição por *feedback* de endoglucanases e celobiohidrolases através da hidrólise de celobiose e celo-oligossacarídeos (Bhatia et al., 2002), sendo assim uma enzima que pode limitar o rendimento do produto final. Coquetéis enzimáticos comerciais são compostos por essas três enzimas. No entanto, as celulases comerciais geralmente exibem baixa atividade de β -glicosidases (Del Pozo et al., 2012). Por ser uma enzima importante na degradação de biomassa lignocelulósica, novas fontes bem como novas estruturas físico-químicas de β -glicosidases que confirmam termoestabilidade, maior especificidade de ligação ao substrato, e com propriedades catalíticas que proporcionem maior velocidade de hidrólise são particularmente atrativas para aplicações biotecnológicas.

A produção viável de biocombustível celulósico não depende apenas de um bom sistema de enzimas celulolíticas, depende também de um sistema eficiente para produção e disponibilidade dessas enzimas por um baixo custo (Chang et al., 2013). Nesse contexto a clonagem do gene que codifica a enzima de interesse num sistema de expressão eficiente continua sendo uma expectativa da biotecnologia moderna. Um sistema de expressão eficiente depende de uma célula hospedeira que não só receba bem o gene mas que o reconheça e o processe até sua proteína funcional. Por outro lado, na fabricação do etanol de segunda geração o desafio é encontrar um micro-organismo capaz de hidrolisar a lignocelulose e fermentar os diferentes açúcares presentes no hidrolisado. Alguns esforços têm sido conduzidos para o desenvolvimento de cepas de leveduras capazes de sintetizar celulases e hemicelulases recombinantes (Raynal & Guerinéau, 1984; Hong et al., 2007; Chang et al., 2013), pois apesar de algumas leveduras serem consideradas boas produtoras de etanol a partir de

açúcares, falta em seus genomas genes para enzimas celulolíticas necessários para eficiente hidrólise da biomassa lignocelulósica (Chang et al., 2013). Dentro dessa perspectiva, leveduras pertencentes ao táxon *Kluyveromyces* apresentam, além das vantagens de um organismo eucariótico como processamento pós-traducional, são organismos unicelulares de crescimento rápido e apresentar ampla diversidade metabólica. A levedura *Kluyveromyces marxianus* CCT 7735 (UFV-3), isolada por nosso grupo de pesquisa (Silveira et al., 2005), é capaz de fermentar em altas temperaturas quando cultivada em altas concentrações de açúcar e baixas tensões de oxigênio, e além de vantagens comuns à espécie, como termotolerância e elevada taxa de crescimento (Fonseca et al., 2008), provou ser capaz de expressar e secretar proteínas heterólogas (Bragança et al., 2014).

No presente estudo, três genes que codificam β -glicosidases isolados de uma biblioteca metagenômica obtida de uma cultura mista reproduzida em laboratório a partir de ambientes naturais (bagaço de cana-de-açúcar e esterco em fase termofílica de decomposição) foram clonados e expressos na levedura *K. marxianus* CCT 7735 (UFV-3), e os transformantes foram analisados quanto à secreção e atividade de β -glicosidase.

MATERIAL E MÉTODOS

Análise *in silico* dos genes de β -glicosidase e predição das proteínas

Dois genes de β -glicosidase que apresentam baixa identidade de aminoácidos (53 e 43 %) e um que apresenta 100 % de similaridade com sequências depositadas nos bancos de dados foram selecionados após análise do sequenciamento de 30 fosmídeos de uma biblioteca metagenômica construída a partir de um consórcio microbiano termofílico obtido por nosso grupo de pesquisa (Capítulo 1; Souza, 2012). O programa ExPASy Bioinformatics Resource Portal (Gasteiger et al., 2005) (http://web.expasy.org/compute_pi/) e Phyre2 server (Kelley and Sternberg, 2009) (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) foram usados para estimar a massa molecular das proteínas e para predizer as estruturas, respectivamente. Similaridade de sequência foi investigada usando o programa Blastp fornecido pelo National Center for Biotechnology Information (NCBI;

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Altschul et al., 1990), e alinhamentos múltiplos foram feitos usando usando o programa ClustalX 2.1 (Larkin et al., 2007)

Linhagens, meios e condições de crescimento

A linhagem de levedura *K. marxianus* CCT 7735 (UFV-3) (Silveira et al., 2005) usada neste trabalho foi mantida a 4 °C em meio YPD sólido [1 % (p/v) extrato de levedura, 2% (p/v) peptona, 2 % (p/v) dextrose e 2 % (p/v) ágar]. Os transformantes foram selecionados em meio YPD sólido suplementado com 200 µg.mL⁻¹ de geneticina (G418, Sigma®). As linhagens selvagem e transformantes foram rotineiramente pré-cultivadas em YPD líquido *overnight* a 37 °C sob agitação. Indução da expressão do gene da β-glicosidase foi realizada em meio YNB líquido (“Yeast Nitrogen Base”, Sigma-Aldrich, St. Louis, MO, USA) com 2 % (p/v) de lactose como única fonte de carbono. Estoque das culturas foram preparados em meio YP com 20 % (v/v) de glicerol e estocadas a -80 °C.

Escherichia coli DH5α foi usada para manutenção e amplificação dos DNA's plasmidiais. Células de *E. coli* foram cultivadas em meio “Luria-Bertani” (LB) suplementado com 50 µg.mL⁻¹ de ampicilina e incubadas a 37 °C sob agitação.

Construção dos cassetes de expressão

Três diferentes genes de β-glicosidasas foram amplificados por PCR a partir de DNA's fosmidiais da biblioteca metagenômica obtida por nosso grupo de pesquisa (Capítulo 1; Souza, 2012). Os pares de primers P3Gli-F e R, P8Gli-F e R, e P9Gli-F e R foram desenhados usando a sequência resultante da montagem do sequenciamento de 30 fosmídeos desta biblioteca (Capítulo 1). Os parâmetros de amplificação foram: 95 °C por 2 min; 35 ciclos de 95 °C por 1 min, 60 °C por 45 s, e 72 °C por 2 min; e 72 °C por 5 min. Os fragmentos correspondentes a cada um dos três genes foram ligados ao vetor pGEM®-T Easy (Promega, Madison, WI, USA) para formar pGEM/P3Gli, pGEM/P8Gli e pGEM/P9Gli, e utilizados para transformar células de *E. coli* DH5α. Após extração dos plasmídeos, estes foram digeridos com *Bam*HI e clonados no vetor pKMCL (Bragança et al., 2014), previamente digerido com a mesma enzima, para dar origem aos vetores pKMCL-P3Gli, pKMCL-P8Gli e pKMCL-P9Gli.

DNA's plasmidiais foram extraídos usando o kit Wizard Plus SV Minipreps DNA Purification System (Promega, Madison, WI, USA). Todas as manipulações de DNA foram realizadas de acordo com Sambrook et al. (1989).

O primers PKmLac-F e PKmLac-R (Tabela 1) foram usados para obter os cassetes de expressão na forma linearizada dos plasmídeos pKMCL-P3Gli, pKMCL-P8Gli e pKMCL-P9Gli. Os cassetes de expressão linearizados foram usados para transformar a linhagem de levedura *K. marxianus* CCT 7735 (UFV-3).

A confirmação dos transformantes após recombinação homóloga no locus do promotor *LAC4* foi realizada por PCR usando primers Integration Primer 1 e Integration Primer 2 (Manual de Instruções, New England Biolabs, 2009) (Tabela 1).

Transformação da levedura

A transformação de *K. marxianus* CCT 7735 (UFV-3) com os cassetes de expressão linearizados obtidos por PCR foi realizada de acordo com Kooistra *et al.* (2004) e Sanchez et al (1993), com adaptações. As células foram inoculadas em cinco mL de meio YED [1 % (p/v) extrato de levedura, 1 % (p/v) glicose] a 37 °C sob agitação de 200 rpm *overnight*. As culturas foram então inoculadas em 50 mL de meio YED para $DO_{600}=0,1$ por mL, e incubadas sob as mesmas condições do inóculo inicial até atingirem DO_{600} de aproximadamente 0,8. Em seguida as células foram centrifugadas por cinco minutos a 6.000 rpm e lavadas com 20 mL de água destilada estéril. O sedimento celular foi ressuspensionado em 20 mL de tampão de pré-tratamento (YED, 25 mM DTT e 20 mM HEPES, pH 8.0), e incubado por 30 minutos a 30 °C sob agitação de 100 rpm. Após o tempo de incubação, as células foram centrifugadas e lavadas com 10 mL de água destilada. O sedimento celular foi então ressuspensionado em 0,3 mL de tampão de eletroporação (10 mM Tris-HCl, pH 7.5, 270 mM sacarose, e 1 mM $MgCl_2$). Foi retirada uma alíquota de 60 μ L da suspensão de células e adicionado dois μ g do cassete de expressão linearizado e 50 μ g de DNA de esperma de Salmon (Invitrogen). Após incubação no gelo por 15 minutos, a mistura foi transferida para uma cubeta de eletroporação (2 mm). A eletroporação foi realizada no aparelho BIORAD Gene Pulser a 1 kV, 25 μ F e 400 Ohm, e imediatamente após o pulso foi adicionado um mL de meio YED gelado. As células foram transferidas para tubos de ensaio estéreis e incubadas a 37 °C por duas horas sob agitação de 200 rpm. Após esse período as células foram centrifugadas e o sedimento celular foi

ressuspendido em um mL de água estéril e 0,3 mL plaqueadas em meio YPD sólido contendo 200 $\mu\text{g}\cdot\text{mL}^{-1}$ de geneticina e incubadas por 72 h a 37 °C.

Expressão e atividade de β -glicosidase

Todos os transformantes contendo os cassetes de expressão pKMCL-P3Gli, pKMCL-P8Gli ou pKMCL-P9Gli integrados no locus *LAC4* foram cultivados em 50 mL de meio YNB (Sigma-Aldrich St. Louis, MO, USA) suplementado com 3 % (p/v) de lactose a 37 °C com agitação por 72 h. As células foram centrifugadas duas vezes e os sobrenadantes das culturas (denominado Sobr_1) foram ensaiados para atividade de β -glicosidase. Após centrifugação o sedimento celular foi utilizado para liberar a fração de enzimas associadas à célula (ligadas à parede celular/retidas no espaço periplasmático) por meio de tratamento com reagentes sulfidril, de acordo com protocolo descrito por Rouwenhorst et al. (1988). As amostras resultantes deste tratamento foram denominadas de Sobr_2, e também foram ensaiadas para a atividade de β -glicosidase.

Atividade de β -glicosidase foi determinada espectrofotometricamente pelo monitoramento do aumento na absorbância a 420 nm (A_{420}) como descrito por Chen et al. (1994) e Rouwenhorst et al., (1988), com adaptações. 0,250 mL do sobrenadante da cultura de cada linhagem expressando o gene de β -glicosidase foi adicionado a 0,75 mL de 12 mM de *p*-NP β G (*p*-nitrophenyl- β -D-glucopyranoside) em tampão fosfato de potássio 50 $\text{mmol}\cdot\text{L}^{-1}$ (pH 7.0). Préincubações foram realizadas a 40 °C por 60 minutos. Imediatamente a absorbância foi medida a 420 nm. Uma unidade de atividade de β -glucosidase foi definida como a quantidade de enzima requerida para a hidrólise de 1 μmol de substrato por minuto, sob as condições de ensaio, por miligrama de proteína.

O experimento foi realizado em triplicata e repetido duas vezes. A atividade específica ($\text{U}\cdot\text{mg}^{-1}$) foi calculada da atividade de β -glicosidase ($\text{U}\cdot\text{mL}^{-1}$) normalizada pela concentração de proteína total da cultura ($\text{mg}\cdot\text{mL}^{-1}$). A concentração de proteína total foi determinada de acordo com Bradford (1976).

O conteúdo de proteína total do sobrenadante das culturas foi coletado durante a fase de indução e precipitado com ácido tricloroacético e acetona e, posteriormente, ressuspendido em tampão de amostras [125 mM Tris-HCl, pH 6,8, 10 % beta-mercaptoetanol, 10 % (p/v) SDS, 10 % (v/v) glicerol e 0,05 (p/v) azul de bromofenol]

para SDS-PAGE. As amostras foram aquecidas a 100 °C por 5 min, e as proteínas foram separadas por SDS-PAGE. A eletroforese em gel de poliacrilamida 12 % (v/v) foi realizada conforme descrito por Sambrook et al., (1989).

Análise Estatística

Todos os ensaios foram conduzidos em triplicata e repetidos duas vezes. Análises estatísticas para avaliar as diferenças na atividade enzimática foram realizadas usando o *software* estatístico Minitab (Minitab, 2006). Subsequente à ANOVA, comparações significativas foram feitas usando teste de Turkey a um valor de significância de $p < 0.05$.

RESULTADOS E DISCUSSÃO

Amplificação dos genes de β -glicosidase e análise da estrutura das proteínas

A partir da montagem de 67 contigs originados do sequenciamento de 30 fosmídeos de uma biblioteca metagenômica construída a partir de um consórcio microbiano termofílico desenvolvido por nosso grupo de pesquisa (Capítulo 1; Souza, 2012), foram identificadas 34 ORFs (Open Reading Frame) presentes em 12 dos contigs que codificam proteínas relacionadas ao grupo de glicosil hidrolases (GH) (Capítulo 1). Foram selecionadas duas β -glicosidasas, as quais foram denominadas P3Gli, P8Gli, pertencentes à família GH3, e uma, denominada P9Gli pertencente à família GH1, para serem clonadas e expressas na levedura *K. marxianus* CCT 7735 (UFV-3). Estas enzimas foram escolhidas primeiramente, dada a importância de seu papel na hidrólise de celulose. Além disso, as proteínas P3Gli e P8Gli dividem apenas 53 e 43 % de identidade (a nível de proteína) com enzimas similares de *Roseburia intestinalis* (Classe Clostridia) e *Carnobacterium maltaromaticum* (Classe Bacilli), respectivamente, ambos micro-organismos anaeróbios pertencentes ao Filo Firmicutes. Os baixos valores de identidade somados à comparação dos contigs 31 e 61, de onde foram selecionadas estas enzimas, contra os dados de base de Genoma Completo do NCBI indicam que estas enzimas são distantemente relacionadas a proteínas conhecidas e podem representar novas enzimas (Capítulo 1). Já a enzima P9Gli, que apresentou 100 % de identidade com um β -glicosidase de *Clostridium thermocellum* ATCC 27405 (Wilson et al., 2014), foi escolhida por representar uma proteína

conhecida, apresentando desta forma uma, maior chance de encontrar atividade e poder ser utilizada como um controle, além de poder ser expresso em levedura. Apesar desta enzima ainda não ter sido caracterizada, apenas identificada e anotada, espécies celulolíticas do gênero *Clostridia* são conhecidas pela produção e secreção de grandes complexos celulolíticos chamados celulosomo, e produção e secreção de enzimas livres, em ambas as situações, ocorre uma eficiente degradação de celulose e polissacarídeos relacionados da parede celular de planta (Mingardon et al., 2005), o que representa grandes chances de P9Gli ser uma enzima ativa.

No curso do estudo, ao realizar um teste de atividade de β -glicosidase em meio sólido contendo esculina e citrato amoníaco de ferro III com as leveduras transformantes, surpreendentemente identificamos que a linhagem de *K. marxianus* CCT 7735 (UFV-3) selvagem, usada como controle, também apresentou teste positivo (halo escuro ao redor da colônia) (Figura 1), indicando que esta linhagem é capaz não somente de sintetizar, mas também de secretar para o meio extracelular a enzima β -glicosidase. Esse resultado é inesperado, pois a literatura sobre expressão heteróloga de β -glicosidases em *K. marxianus* (Ball et al., 1999; Hong et al., 2007; Yanase et al., 2010; Chen et al., 2012), é mais abundante que sobre a enzima nativa nesta levedura. Raynal & Guerneau (1984) isolaram o gene da β -glicosidase de *K. fragilis* ATCC 12424 (atualmente classificada como *K. marxianus* ATCC 12424 (Lachance, 1998) para ser clonado na levedura *S. cerevisiae*. O estudo mais citado ao se realizar uma busca literária por produção de β -glicosidase em *K. marxianus* é o trabalho desenvolvido por Yoshida et al. (2010), onde os autores realizaram a purificação, cristalização e análises preliminares da raio-X da β -glicosidase intracelular de *K. marxianus* NBRC 1777. Demais trabalhos a partir desta data apenas citam os estudos de Raynal & Guerneau (1984) e Yoshida et al. (2010). A falta de dados na literatura sobre a expressão dessa enzima em outras linhagens de *K. marxianus*, ou o crescimento desta levedura em celobiose retrata sua diversidade metabólica, confirmada por seu isolamento em diversos ambientes, sendo esta uma levedura altamente polimórfica (Fonseca, et al., 2008).

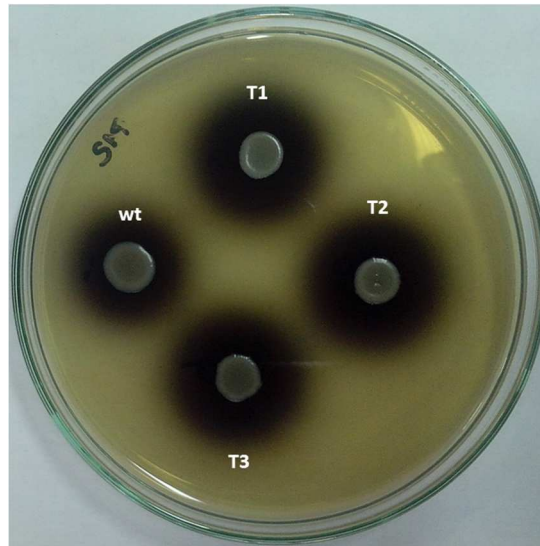
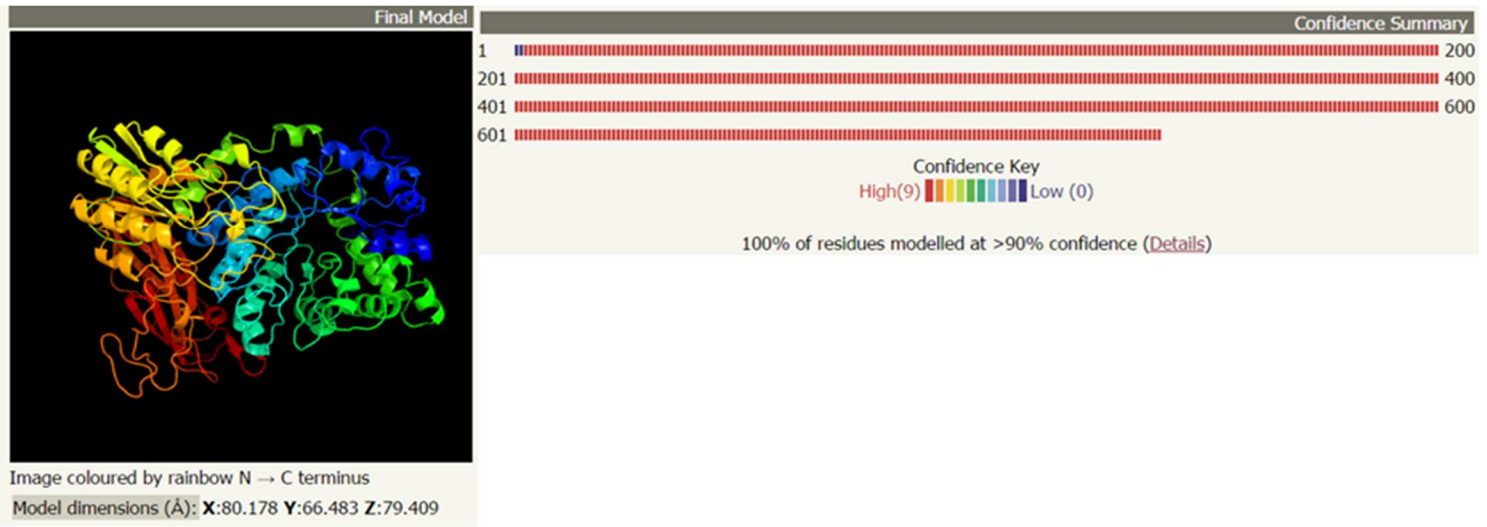


Figura 1. Análise qualitativa de β -glicosidase com linhagens de *K. marxianus* CCT 7735 (UFV-3) selvagem (wild-type –wt) e transformantes (T1 a T3) em meio sólido contendo esculina e citrato amoníaco de ferro III. Formação de halo preto ao redor das colônias indica resultado positivo para atividade de β -glicosidase.

Diante desses resultados, foi pesquisado no genoma da linhagem *K. marxianus* CCT 7735 (UFV-3), recentemente sequenciado e anotado (Silveira et al., 2014), o gene da β -glicosidase baseado na sequência do gene *K. marxianus* NBRC 1777 (*GenBank* nº ACY95404.1), para compará-lo com os demais genes de β -glicosidase utilizados neste trabalho. O gene da β -glicosidase de *K. marxianus* CCT 7735 (UFV-3) encontra-se localizado no cromossomo 3 (posições 39698 a 36161); é codificado pela fita complementar (fita negativa); possui 2.538 pb; 846 resíduos de aminoácidos; e massa molecular da proteína estimada em 93,75 kDa.

O gene referente à β -glicosidase P3Gli isolado da biblioteca metagenômica contém 2.225 pb, correspondendo a 741 resíduos de aminoácidos, e uma massa molecular estimada de 83,31 kDa; o gene de P8Gli contém 889 pb, 296 resíduos de aminoácidos e 34,34 kDa; enquanto P9Gli contém 1.418 pb, 472 resíduos de aminoácidos e peso molecular 54,08 kDa. Os três genes de β -glicosidase clonados apresentam diferenças significativas de tamanho e massa molecular em relação à β -glicosidase endógena de *K. marxianus* CCT 7735 (UFV-3). A Figura 2 mostra a predição da estrutura 3D dessas quatro proteínas utilizando o programa Phyre2. De todas as estruturas tridimensionais geradas, somente P8Gli apresentou <90 % de confiança, mas apresentou 59 % dos resíduos modelados com confiança

A**B**

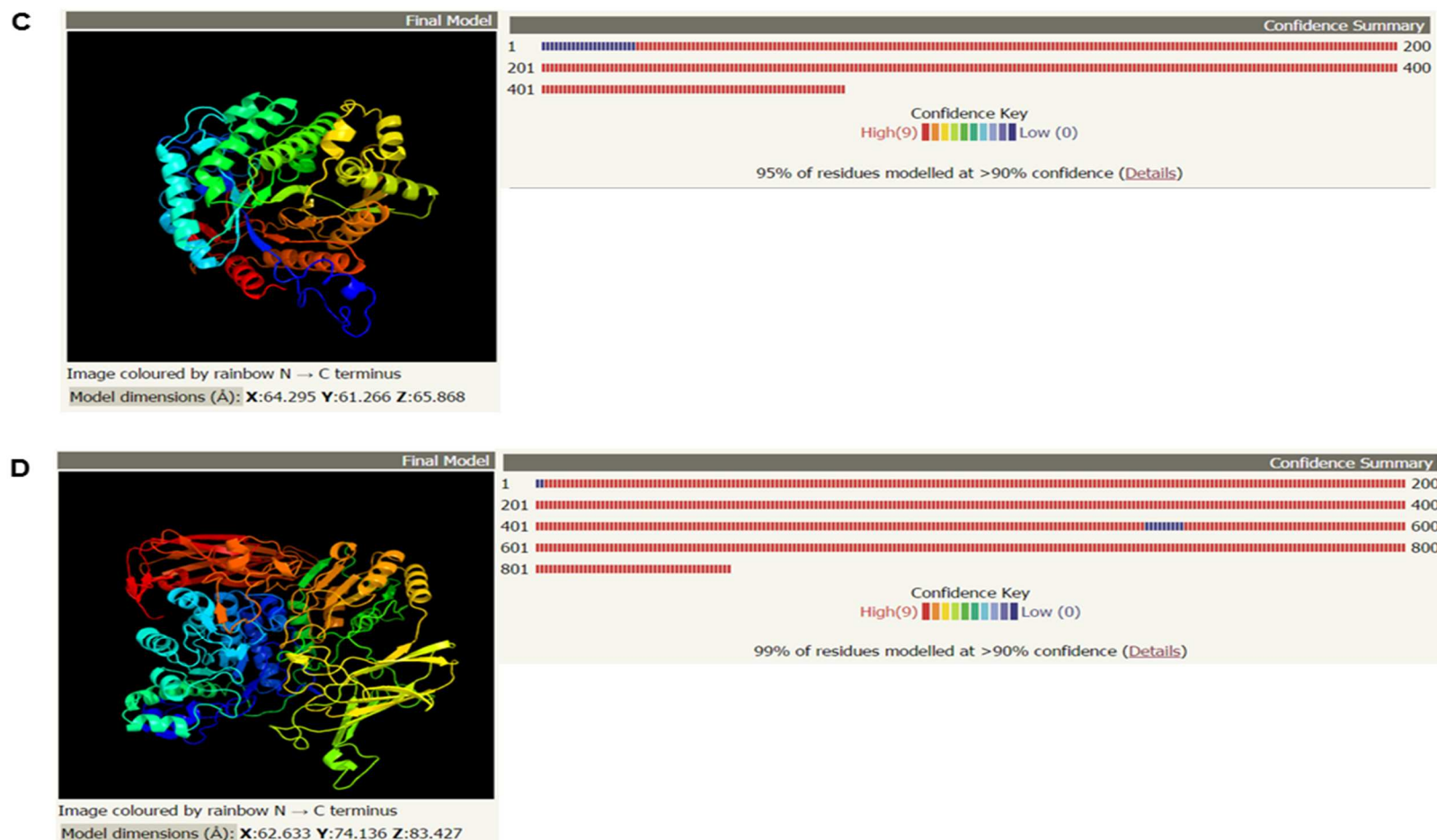


Figura 2. Predição da estrutura tri-dimensional de enzimas β -glicosidases P3Gli, P8Gli, P9Gli e β -glicosidase de *K. marxianus* CCT 7735 (UFV-3). **A**, P3Gli; **B**, P8Gli; **C**, P9Gli; **D**, β -glicosidase de *K. marxianus* CCT 7735 (UFV-3). Nível de confiança superior a 90 % indica uma maior confiança de que a proteína adote aquele enovelamento mostrado e que o núcleo da proteína é modelado com alta precisão.

superior a 90 %. Confidência superior a 90 % indica uma maior confiança de que a proteína adote aquele enovelamento mostrado e que o núcleo da proteína está modelado com alta precisão.

Análise de similaridade de sequências de aminoácidos de P3Gli, P8Gli e a β -glicosidase de *K. marxianus* CCT 7735 (UFV-3) foram realizadas por alinhamento com outras β -glicosidases pertencentes à família GH3 de outros seis micro-organismos representantes dos grupos de fungo (*Neocallimastix patriciarum* e *Aspergillus niger*), bactéria (*Ruminoclostridium thermocellum* e *Clostridium stecorarium*), arquea (*Thermofilum pendens*) e levedura (*K. marxianus* NBRC 1777) (Figura 3). O sítio catalítico de todas as proteínas submetidas ao alinhamento encontra-se localizado entre os resíduos 304 e 320.

Todas as sequências apresentam o domínio BglX (ou Glyco_hydro_3) típico da família GH3 de β -glicosidases (Li et al., 2013). As pequenas diferenças na sequência de aminoácidos podem resultar em alteração de domínios proporcionando maior ou menor especificidade de ligação ao substrato. Termoestabilidade por exemplo, é uma propriedade obtida através de muitas modificações estruturais mínimas. Comparações de proteínas termofílicas com sua homóloga mesofílica mostram que geralmente ocorre uma troca de alguns resíduos de aminoácidos (Sadeghi et al., 2006). Apesar de apresentar uma fração grande de conservação entre as nove sequências alinhadas, apenas as sequências das duas leveduras são mais conservadas ao longo da sequência, enquanto as outras diferem destas e entre si em alguns resíduos de aminoácidos, podendo resultar na diferença de atividade enzimática de cada uma.

A sequência de aminoácidos predita para β -glicosidase de *K. marxianus* CCT 7735 (UFV-3) mostrou 99% de identidade com a mesma enzima de *K. marxianus* NBRC 1777, e pelos dados gerados pelo programa Phyre2 para predição da estrutura 3D da proteína, houve 100 % de confidência de 97 % da β -glicosidase da linhagem CCT 7735 (UFV-3) adotar a mesma topologia da enzima da linhagem NBRC 1777, sugerindo serem estas proteínas homólogas.

Um dos principais interesses em se utilizar a levedura *K. marxianus* CCT 7735 (UFV-3) na degradação de biomassa celulósica para produção de bioetanol está na potencialidade de utilizá-la em processos de sacarificação e fermentação simultânea. Portanto, obter uma linhagem desta levedura capaz de produzir e secretar a enzima β -glicosidase endógena que antes se pensava ser intracelular, bem como obter

linhagens mutantes dessa levedura que também expressem novas β -glicosidasas recombinantes ainda não estudadas representa uma importante estratégia para seu uso neste tipo de aplicações biotecnológicas.

Transformação de *K. marxianus* CCT 7735 (UFV-3) com os cassetes de expressão pKMCL/P3Gli, pKMCL/P8Gli e pKMCL/P9Gli, e seleção de linhagens recombinantes

Com o sequenciamento dos 30 fosmídeos obtidos da biblioteca metagenômica, identificou-se os genes de glicosil hidrolases, selecionou-se genes que codificam β -glicosidase putativas e outras enzimas celulolíticas e hemicelulolíticas, e desenhou-se primers para amplificação desses genes (Capítulo 01). Todos os genes escolhidos não apresentam peptídeo sinal. Considerando que os 30 fosmídeos foram misturados antes do sequenciamento, realizou-se reação de PCR para cada gene usando como molde os 30 fosmídeos, conseguindo assim identificar o fosmídeo ao qual pertencia cada gene. Após amplificação dos genes P3Gli, P8Gli e P9Gli, estes foram clonados no plasmídeo pGEM®-T Easy e posteriormente, subclonados no vetor pKMCL (Bragança et al., 2014). O vetor pKMCL é derivado do vetor de expressão pKLAC2 (Colussi & Taron, 2005), originalmente construído para *K. lactis*, e que tem sido usado para expressão heteróloga de enzimas celulolíticas em *K. marxianus* (Chang et al., 2012; Chang et al., 2013). A seleção de transformantes contendo o vetor pKLAC2, é o crescimento em meio seletivo contendo acetamida como única fonte de nitrogênio, pois o vetor apresenta o gene da acetamidase (*amdS*), sob controle do promotor *ADH1*, como marcador de seleção. No entanto, a levedura selvagem *K. marxianus* CCT 7735 (UFV-3) é capaz de crescer em meio seletivo contendo acetamida como única fonte de nitrogênio, impossibilitando o uso da marca de seleção presente no pKLAC2. Assim, o vetor pKMCL, derivado do pKLAC2, contém, ao invés de acetamidase, o cassete LoxP-KanMX-LoxP, que confere resistência ao antibiótico geneticina, como gene marcador de seleção.

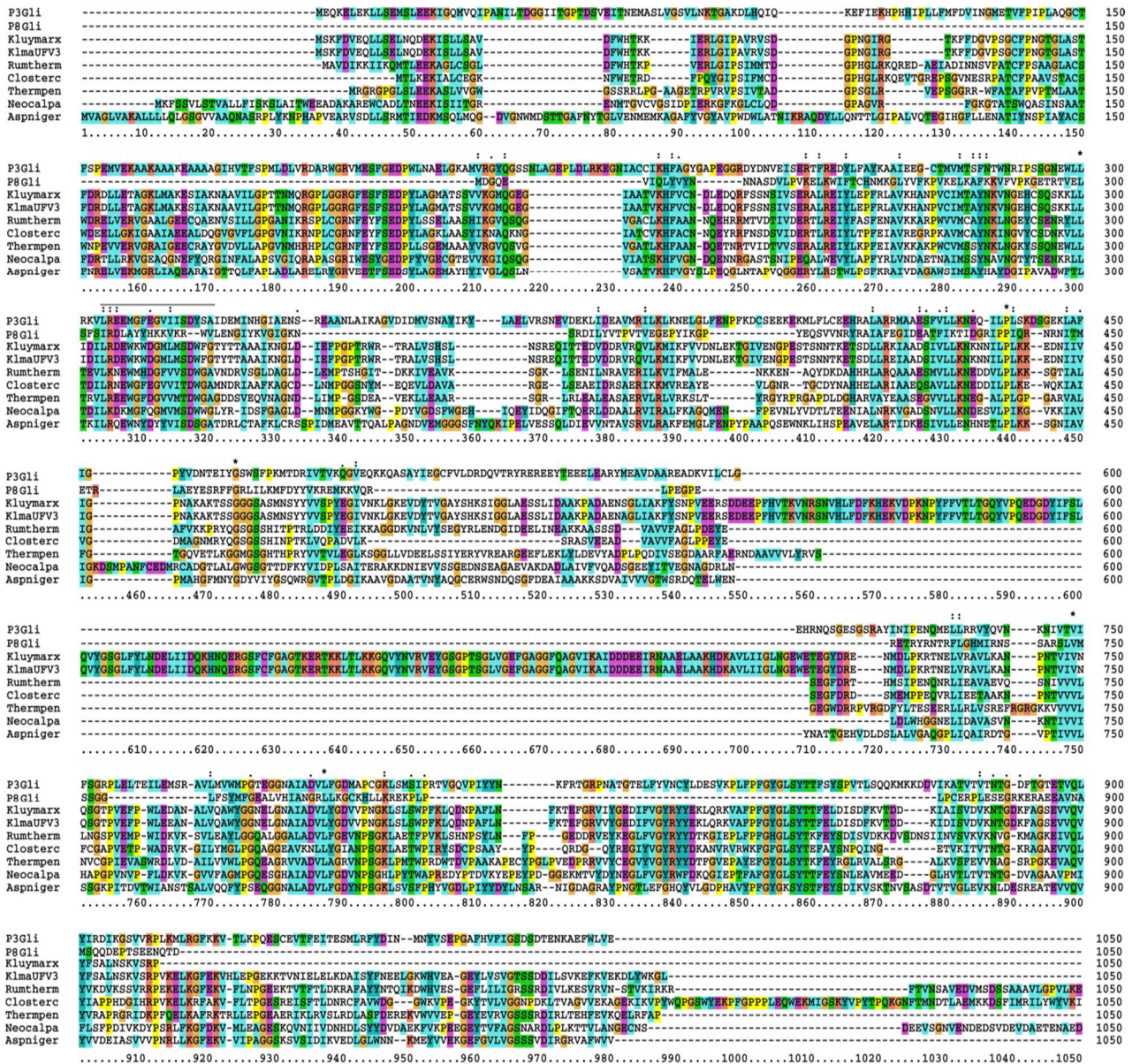


Figura 3. Alinhamento múltiplo de sequências das sequências traduzidas de P3Gli, P8Gli e β-glicosidase de *K. marxianus* CCT 7735 (UFV-3) com outras β-glicosidasas da família GH3 (ClustalX 2.1). O sítio catalítico das enzimas está indicado com um traço na parte superior. A cor dos resíduos em cada coluna foi determinada de acordo com o resíduo consenso para cada coluna. Kluymarx – *Kluyveromyces marxianus* NBRC 1777 (ACY95404.1); KlimaUF3 – *Kluyveromyces marxianus* UFV-3 (PRJNA255779); Rumtherm - *Ruminoclostridium thermocellum* (ABN52488.1); Thermpen – *Thermofilum pendens* (YP_920894.1); Neocalpa - *Neocallimastix patriciarum* (AFI47447.1); Closterc – *Clostridium stecorarium* (AGI38411.1); Aspigner – *Aspergillus niger* (XP_001398259.2). Asterisco (*) indica posições que apresentam um único resíduo completamente conservado. Dois pontos (:) indica conservação entre grupos de propriedades fortemente similares. Ponto (.) indica conservação entre grupos de propriedades fracamente similares.

Os vetores pKMCL-P3Gli, pKMCL-P8Gli e pKMCL-P9Gli deveriam ser linearizados com a endonuclease de restrição SacII antes de serem usados, separadamente, para a transformação da levedura *K. marxianus* CCT 7735 (UFV-3). No entanto, esta enzima cliva os genes P8Gli e P9Gli, impossibilitando seu uso no processo de linearização do vetor. Então, um cassete linear foi obtido de cada vetor, pKMCL-P3Gli, pKMCL-P8Gli e pKMCL-P9Gli, por amplificação *in vitro* usando o par de oligonucleotídeo PKmLac-F e PKmLac-R (Tabela 1). O fragmento amplificado referente a cada cassete contém o gene da β -glicosidase (P3Gli, P8Gli ou P9Gli), a marca de seleção com o gene de resistência à geneticina (LoxP-KanMX-LoxP), e também sequências 3'P_{LAC4} (1.578) e 5'P_{LAC4} (525), nas extremidades dos cassetes, as quais dirigem a integração do cassete no locus do promotor *LAC4* no genoma da levedura por recombinação homóloga. É necessário que ocorra esta recombinação corretamente para que haja a reconstituição do promotor e expressão do gene recombinante.

Após transformação de células de *K. marxianus* CCT 7735 (UFV-3) com cada cassete separadamente, foram obtidos 20, 21 e 19 transformantes UFV3/P3Gli, UFV3/P8Gli e UFV3/P9Gli, respectivamente. O DNA de 3 transformantes em cada caso foi extraído para análises de integração gene-específica. Um fragmento de 2,3 kb obtido da amplificação *in vitro* usando os oligonucleotídeos Primer de Integração 1-F (P1) e Primer de Integração 2-R (P2) (Figura 4A) revelou que a recombinação homóloga ocorreu dentro do promotor *LAC4* para todos os transformantes analisados. Mesmo a via de recombinação homóloga nesta levedura não ser a via preferencial de integração de DNA exógeno no seu genoma, e sim a via de junção de extremidades não homólogas (NHJE) (Abdel-Banat et al., 2010), a quantidade de bases utilizadas nas sequências de homologia ao promotor *LAC4* [3'P_{LAC4} (1.578) e 5'P_{LAC4} (525)] presente nas extremidades de cada cassete linear foram suficientes para induzir e efetivamente promover a integração homóloga dos cassetes. Além disso, todas as linhagens analisadas apresentaram mais de uma cópia de cada cassete, evidenciado pelo fragmento de 2.8 kb obtido por amplificação por PCR utilizando o par de oligonucleotídeos Primer de Integração 3-F (P3) e Primer de Integração 2-R (P2) (Figura 4B). Isso ocorre porque a inserção do cassete no locus *LAC4* por recombinação homóloga duplica a região do promotor *LAC4*, que pode então ser alvo

de uma nova recombinação homóloga por outro cassete resultando na integração de múltiplas cópias.

No vetor pKMCL derivado do pKLAC2, o gene heterólogo é expresso fusionado ao domínio de secreção α -mating fator de *K. lactis*, o qual endereça a proteína para a via secretória para ser processada pela peptidase sinal e proteases no retículo endoplasmático e Complexo de Golgi, respectivamente, resultando na secreção da proteína ativa (Colussi & Taron, 2005). A presença do peptideo sinal na proteína recombinante e a integração de mais de uma cópia do cassete de expressão no genoma da levedura hospedeira pode potencialmente aumentar os níveis de expressão e secreção da proteína recombinante (Colombo et al., 2014). Portanto, a síntese extracelular das β -glicosidases clonadas foi avaliada.

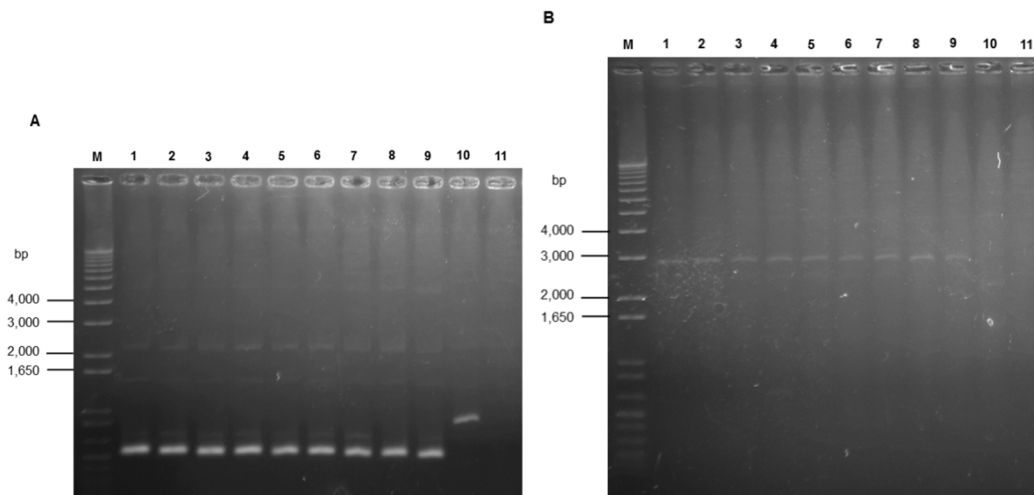


Figura 4. Integração dos cassetes de expressão. Produtos amplificados por PCR a partir do DNA genômico de transformantes UFV3/P3Gli, UFV3/P8Gli e UFV3/P9Gli. A, produto de 2,3 kb amplificado com o par de oligonucleotídeos Primer de Integração 1 (P1) + Primer de Integração 2 (P2) que indica integração gene-específico do cassete de expressão. B, produto de 2,4 kb amplificado usando o par de oligonucleotídeo Primer de Integração 3 (P3) + P2, que indica ocorrência de integração de mais de uma cópia do cassete de expressão. **M**, marcador 1 kb DNA *ladder plus* (Invitrogen). **1 a 3**, transformantes UFV3/P3Gli; **4 a 6**, transformantes UFV3/P8Gli; **7 a 9**, transformantes UFV3/P9Gli; **10**, controle negativo utilizando o DNA total da linhagem selvagem CCT 7735 (UFV-3); **11**, controle negativo sem DNA.

Análise da expressão dos genes P3Gli, P8Gli e P9Gli pela atividade de β -glicosidase

Três clones de cada cepa transformante *K. marxianus* UFV3/P3Gli, UFV3/P8Gli e UFV3/P9Gli foram investigados quanto à capacidade de expressar e secretar a enzima β -glicosidase. Ensaio enzimáticos de β -glicosidase foram realizados com o sobrenadante das culturas (denominado Sobr_1) e para a fração de enzimas ligadas à parede celular (retidas no periplasma) (denominado Sobr_2). A Figura 5 mostra os valores de atividade específica ($\text{U}\cdot\text{mg}^{-1}$ de proteína) de cada clone e cepa controle CCT 7735 (UFV-3) sobre o substrato sintético pNPG. Foi detectada atividade da enzima no sobrenadante da cepa controle CCT 7735 (UFV-3), o que já era esperado após visualização de atividade em teste enzimático realizado em meio sólido. No entanto, os ensaios foram realizados a temperatura de 50 °C, uma vez que os genes foram isolados a partir de um consórcio microbiano termofílico selecionado nesta temperatura, enquanto o teste de β -glicosidase em placa havia sido realizado a temperatura de 37 °C, temperatura ótima de crescimento de *K. marxianus* CCT 7735 (UFV-3). Essa estratégia pode ter diminuído a interferência da enzima endógena nos ensaios enzimáticos. Geralmente a hidrólise de celulose é realizada sob condições de temperatura de 45 °C. Portanto, encontrar β -glicosidases com estabilidade de temperatura mais elevada é de especial interesse na produção de bioetanol, pois além de necessitar de grandes quantidades desta enzima para a hidrólise do material celulósico, as enzimas termoestáveis são facilmente manuseadas, armazenadas e transportadas, o que diminui potencialmente o custo da hidrólise (Uchima et al., 2012).

No sobrenadante das culturas (Sobr_1) apenas o clone UFV3/P9Gli-2 não apresentou diferença significativa na atividade de β -glicosidase em relação à cepa controle (Figura 5A), e o clone UFV3/P8Gli-1 não apresentou atividade detectável. Os demais clones foram significativamente diferentes do controle. Já para a atividade de Sobr_2 relacionado ao periplasma, apenas os clones UFV3/P8Gli-3 e UFV3/P8Gli-1 (cuja atividade não foi detectada no Sobr_1) apresentaram diferenças significativas em relação à cepa controle (Figura 5B).

O valor de maior atividade enzimática alcançado no sobrenadante das culturas foi de 0,0848 $\text{U}\cdot\text{mg}^{-1}$ referente ao clone UFV3/P3Gli-3, seguido da atividade de 0,0832 $\text{U}\cdot\text{mg}^{-1}$ do clone UFV3/P8Gli-3, sendo ambas 17 vezes maior que a atividade da cepa controle CCT 7735 (UFV-3) (0,0049 $\text{U}\cdot\text{mg}^{-1}$). Quanto aos ensaios enzimáticos das

amostras do periplasma, o maior valor apresentado foi de 0,0981 U.mg⁻¹ também do clone UFV3/P8Gli-3, um valor de atividade 2,5 vezes maior que a cepa controle CCT 7735 (UFV-3), cuja atividade foi de 0,0398 U.mg⁻¹.

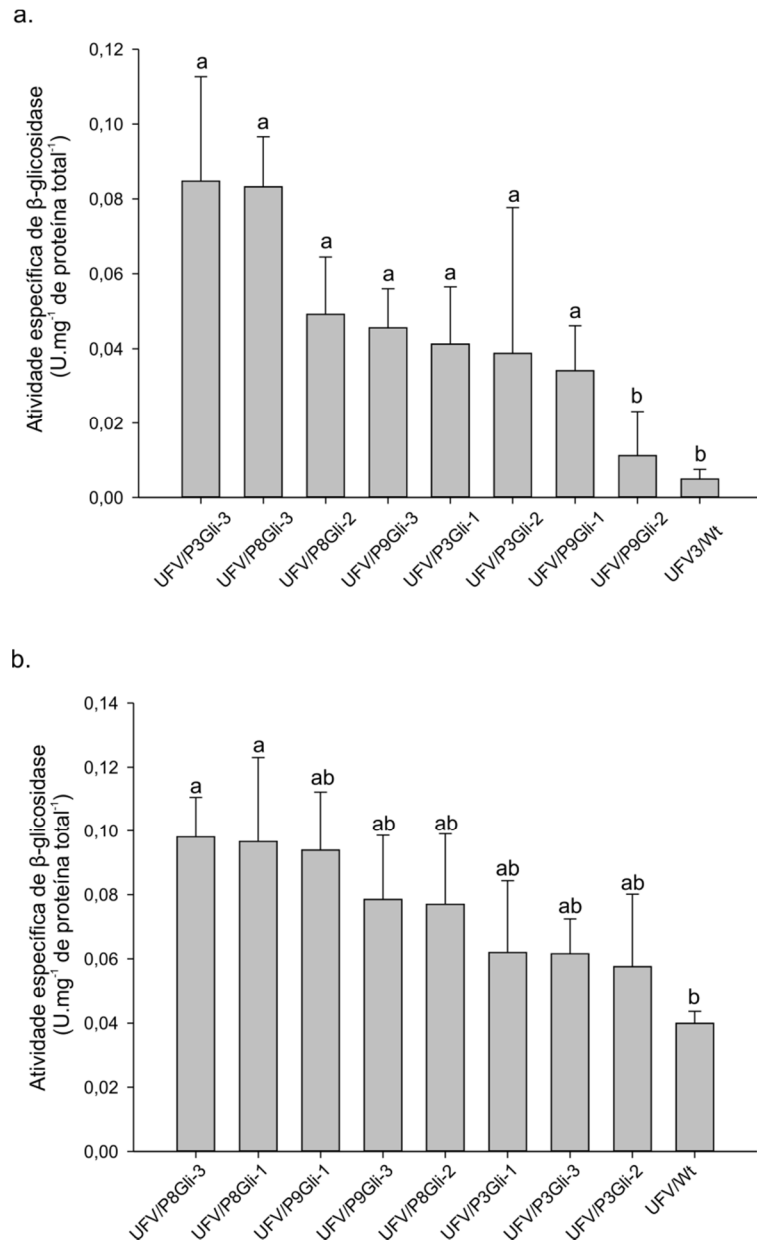


Figura 5. Atividade específica de β-glicosidase dos transformantes *K. marxianus* UFV3/P3Gli, UFV3/P8Gli, UFV3/P9Gli, e linhagem controle CCT 7735 (UFV-3) em meio YNB + 3 % (p/v) de lactose. As atividades de β-glicosidase foram calculadas: **(A)** no sobrenadante das culturas (Sobr_1) e **(B)** na fração de enzimas ligadas à parede celular (retidas no espaço periplasmático) (Sobr_2). Barras indicam desvio padrão das

médias em um mínimo de três amostras. As médias seguidas da mesma letra são estatisticamente iguais pelo teste de Tukey ao nível de 5 % de probabilidade.

Uchima et al. (2012) relataram a expressão de uma β -glicosidase de *Nasutitermes takasagoensis* em *Pichia pastoris* com a linhagem recombinante apresentando atividade de 0,037 U.mg⁻¹. Com essa expressão eles purificaram e caracterizaram a enzima. No entanto, a comparação dos resultados apresentados neste trabalho com a maioria das β -glicosidases heterólogas expressas em diferentes micro-organismos é difícil em decorrência dos ensaios enzimáticos apresentados em tais trabalhos serem realizados com o sobrenadante da cultura juntamente com o sobrenadante obtido após lise celular, ou ainda, com a enzima purificada a partir desses sobrenadantes, o que resulta em valores superiores de atividade. Um exemplo é o trabalho desenvolvido por Bergmann et al. (2014), onde se apresenta uma β -glicosidase purificada com atividade de 85 U.mg⁻¹. Para micro-organismos produtores de tais enzimas os valores também são diferentes. Por exemplo, o fungo *Trichoderma reesei*, produtor de celulases, apresenta atividade de β -glicosidase endógena de 2,2 U.mg⁻¹ (Dashtban & Qin, 2012), enquanto o fungo *Phanerochaete chrysosporium* apresenta atividade de 0,254 U.mg⁻¹ (Lyman et al., 1995). Portanto, não podemos concluir que a expressão das β -glicosidases em *K. marxianus* CCT 7735 (UFV-3) foram baixas, especialmente por se tratar de expressão heteróloga e de ensaios realizados com o sobrenadante da cultura sem haver concentração prévia. O que podemos observar no entanto, é que parte das enzimas ficaram retidas no espaço periplasmático, como pode ser observado com o clone UFV3/P8Gli-3, que apresentou atividade específica nos dois casos analisados e que por análise estatística ao nível de 5 % de probabilidade não apresentou diferença significativa de valores (dados não mostrados).

Esperava-se obter maior atividade extracelular das enzimas, pois o gene foi expresso fusionado ao domínio de secreção α -*mating factor* de *K. lactis*, o qual já foi demonstrado funcionar em *K. marxianus* (Bragança et al., 2014) utilizando o mesmo vetor usado no presente trabalho. Essa sequência sinal dirige a proteína expressa para a via secretória para ser processada por peptidase sinal e proteases presentes no retículo endoplasmático e no Complexo de Golgi, respectivamente, resultando na secreção da proteína ativa (Colussi & Taron, 2005). A atividade de β -glicosidase detectada no periplasma em todas as cepas transformantes, incluindo a cepa controle, pode indicar retenção de parte da enzima expressa, e pode estar relacionada a fatores

de glicosilação da proteína. Glicosilação é um tipo comum de modificação pós-traducional observada em organismos eucariotos, e que pode influenciar na solubilidade, dobramento, estabilidade térmica e atividade das proteínas, sendo a N-glicosilação no retículo endoplasmático o tipo de modificação bem conservada entre leveduras (Çelik & Çalik, 2012). No entanto, diferentes tipos ou graus de glicosilação podem ocorrer em uma mesma proteína em diferentes leveduras. Por exemplo, no Complexo de Golgi, as glicoproteínas podem sofrer adição de manose (hiperglicosilação) em diferentes graus dependendo da levedura - *Pichia pastoris* adiciona até 20 resíduos, enquanto *S. cerevisiae* pode adicionar de 50 a 100 (Çelik & Çalik, 2012). Rocha et al. (2010), ao expressar a enzima glicose oxidase (GOX) de *Aspergillus niger* heterologicamente em *K. marxianus* SLC33, observaram que esta levedura apresenta uma tendência de glicosilação da proteína heteróloga, mas uma menor tendência de hiperglicosilar esta proteína. Da mesma forma, Rocha et al. (2011) observaram uma menor tendência de hiperglicosilação por esta levedura durante a expressão heteróloga de uma esterase, que apresentou-se em três níveis de glicosilação: desglicosilada, glicosilada em um ponto, e glicosilada em dois pontos. Em ambos os trabalhos, as proteínas heterólogas foram ativas. Esta característica de *K. marxianus* é uma vantagem em relação à *S. cerevisiae* para expressão de proteínas heterólogas, pois a hiperglicosilação de proteínas apresentada por esta, pode reduzir o nível de secreção da proteína de interesse. Entretanto, mesmo apresentando uma menor tendência de hiperglicosilação, a glicosilação de proteínas ocorre nesta levedura, então, a manutenção da atividade da enzima dependerá da própria sequência da proteína, do correto enovelamento e formação de pontes dissulfeto. Para o exemplo citado acima, a glicosilação não afeta os parâmetros catalíticos de GOX, por exemplo, mas causa uma mudança na estrutura da proteína (Kohen et al., 1997). De acordo com cálculo realizado pelos servidores NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc/>) e NetOGlyc (<http://www.cbs.dtu.dk/services/NetOGlyc/>), a proteína P3Gli apresenta apenas dois potenciais sítios para N-glicosilação e três para O-glicosilação, já P8Gli apresenta 3 e 2, respectivamente, enquanto P9Gli contém apenas 3 sítios para N-glicosilações. A presença de atividade de β -glicosidase extracelular nos três tipos de recombinantes analisados sugere que a enzima foi corretamente enovelada e houve processamento do peptídeo sinal para secreção da enzima funcional. No entanto, apesar de serem considerados baixos

níveis de glicosilação, não sabemos como esses carboidratos interagem com a superfície da proteína β -glicosidase ou com seu sítio ativo. Não há relatos na literatura para afirmar ausência de interferência na atividade da enzima, como redução de sua atividade com a adição dessas cadeias de polissacarídeos, mas é possível que essas glicosilações estejam interferindo na secreção das proteínas para o meio extracelular.

Outro fator que deve ser levado em consideração ao se analisar a expressão de proteínas heterólogas e que pode estar interferindo com a secreção é a competição das proteínas recombinantes com as proteínas endógenas pela entrada na via de secreção (Mattanovich et al., 2004). No caso dos transformantes, cada um apresenta, além da β -glicosidase endógena, uma β -glicosidase recombinante. Sabe-se que a expressão de proteínas recombinantes em altos níveis pode limitar outros processos celulares pela competição por substratos ou interferência direta no metabolismo (Mattanovich et al., 2004). Tudo isso pode inclusive comprometer a secreção de proteínas endógenas do hospedeiro. Portanto, a competição entre a proteína recombinante e a endógena pode também estar causando retenção de ambas as proteínas no periplasma. Neste caso, as análises das β -glicosidases recombinantes foram favorecidas, pois reduziu a interferência da enzima endógena nos ensaios do sobrenadante extracelular. Entretanto, foram observadas mudanças no perfil de secreção e/ou expressão de outras proteínas endógenas. A Figura 6 mostra o fracionamento de proteínas do sobrenadante das culturas recombinantes e controle em gel de poliacrilamida (SDS-PAGE). A discussão a partir das análises do gel torna-se complicado, pois não foi possível identificar as proteínas P3Gli, P8Gli e P9Gli pelo gel, e a comparação de cada amostra com as demais e com o controle ficou de difícil visualização, mas pelo gel é possível observar a alteração no perfil de secreção dos transformantes citado acima.

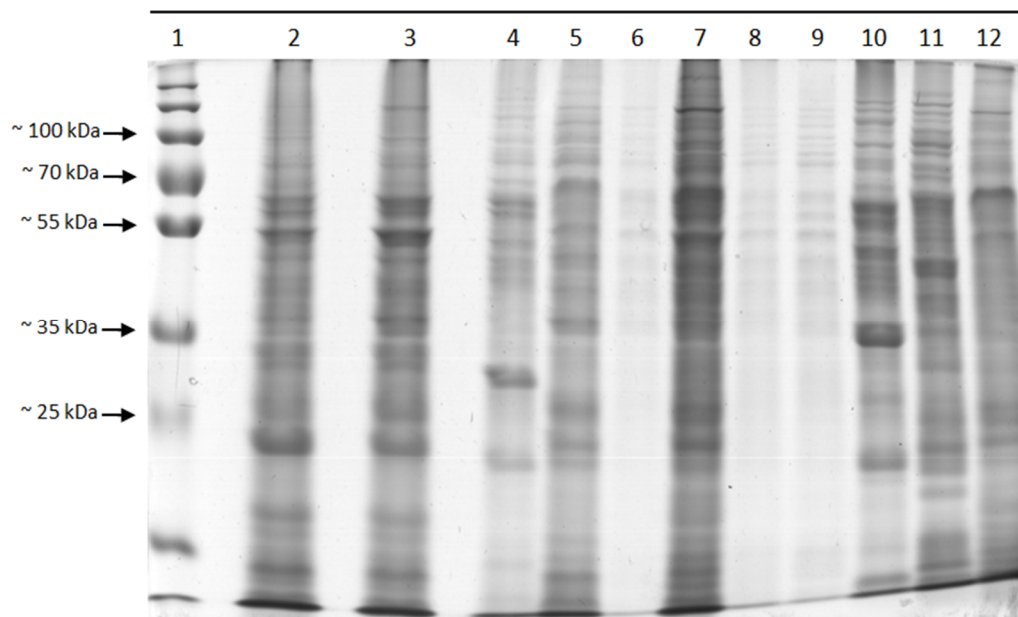


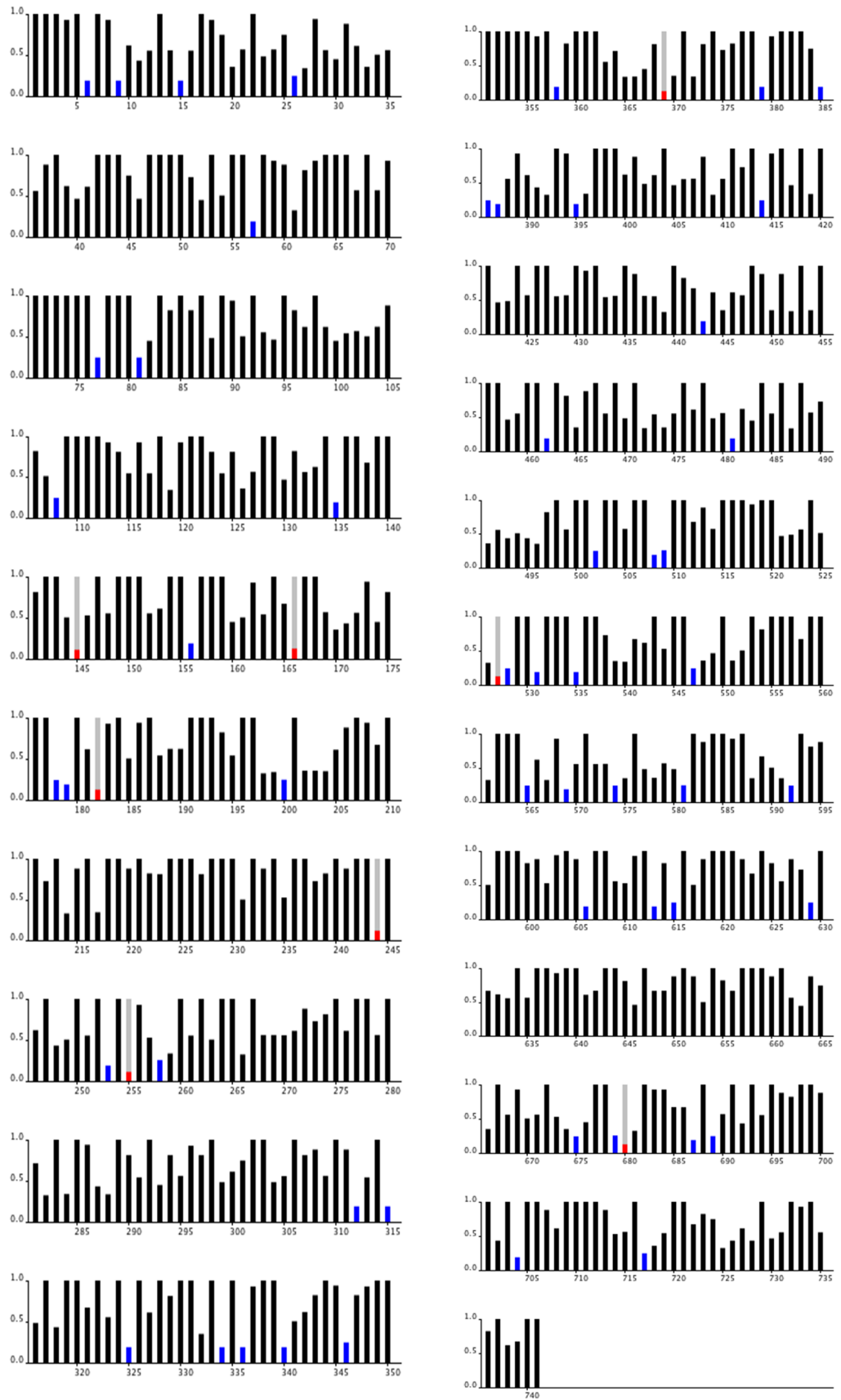
Figura 6. Fracionamento em gel de poliacrilamida (SDS-PAGE) 12 % (p/v) em condições desnaturantes para análise de β -glicosidase presentes no sobrenadante das culturas de *K. marxianus* CCT 7735 (UFV-3) recombinantes e cepa controle. Marcador de proteína (Invitrogen) (1); Controle CCT 7735 (UFV-3) (2); Controle CCT 7735 (UFV-3) (3); P9Gli-1 (4); P9Gli-2 (5); P9Gli-3 (6); P8Gli-1 (7); P8Gli-2 (8); P8Gli-3 (9); P3Gli-1 (10); P3Gli-2 (11); P3Gli-3 (12). As setas identificam a enzima β -glicosidase.

Análise *in silico* das seqüências de códons das β -glicosidasas expressas

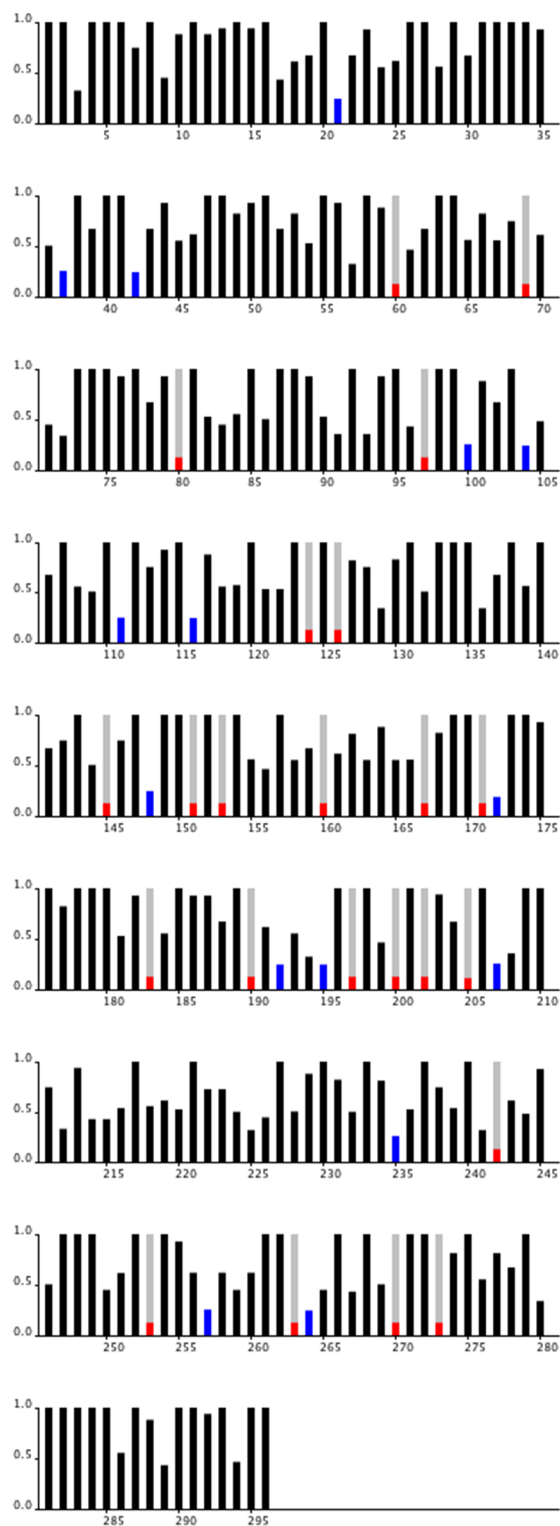
Por fim, para verificar se o nível de expressão das enzimas poderia ser alterado pela relação de códon preferencial por se tratar de enzimas providas do domínio Bacteria, foram realizadas análises de códons preferenciais usando o programa Condon Usage Enhancer (CUE) (Yu & Cheng, 2004) com adaptações feitas por membro do Laboratório de Fisiologia de Micro-organismos usando todas as seqüências CDS do genoma de *K. marxianus* CCT 7735 (UFV-3). A Figura 7 mostra o resultado das análises para cada um dos três genes de β -glicosidase. As barras azuis presentes nos gráficos indicam códons que podem ser melhorados e as barras vermelhas códons não usuais em *K. marxianus* CCT 7735 (UFV-3) que podem ser trocados para se alcançar um melhor nível de expressão da proteína. O gene de P8Gli (Figura 7B) é o que apresenta maior número de códons que devem ser melhorados para a expressão, enquanto o gene P9Gli (Figura 7C), o que apresenta o menor número de diferenças em relação aos códons preferenciais de *K. marxianus* CCT 7735 (UFV-3). O índice CAI

(*codon adaptation index*) é frequentemente usado para medir o nível de expressão de um gene através do desvio de uma dada sequência de um gene que codifica uma proteína com relação a um conjunto de genes de referência (Sharp & Li, 1987). Por análise do índice CAI para os três genes, mesmo com algumas diferenças de códons preferenciais mostrados na Figura 7, esses genes podem ser bem expressos em *K. marxianus* CCT 7735 (UFV-3), pois o valor de CAI para os três genes ficou acima de 0.6, que baseado na literatura (Friberg et al., 2004), pode ser expresso na levedura. Isso é confirmado pelo fato de um dos maiores resultados de atividade enzimática entre os clones ter sido justamente de um clone do transformante UFV3/P8Gli. Essas análises são importantes pois direcionam os trabalhos para a realização de mutações silenciosas que poderão aumentar os níveis de expressão da proteína de interesse.

A



B



C

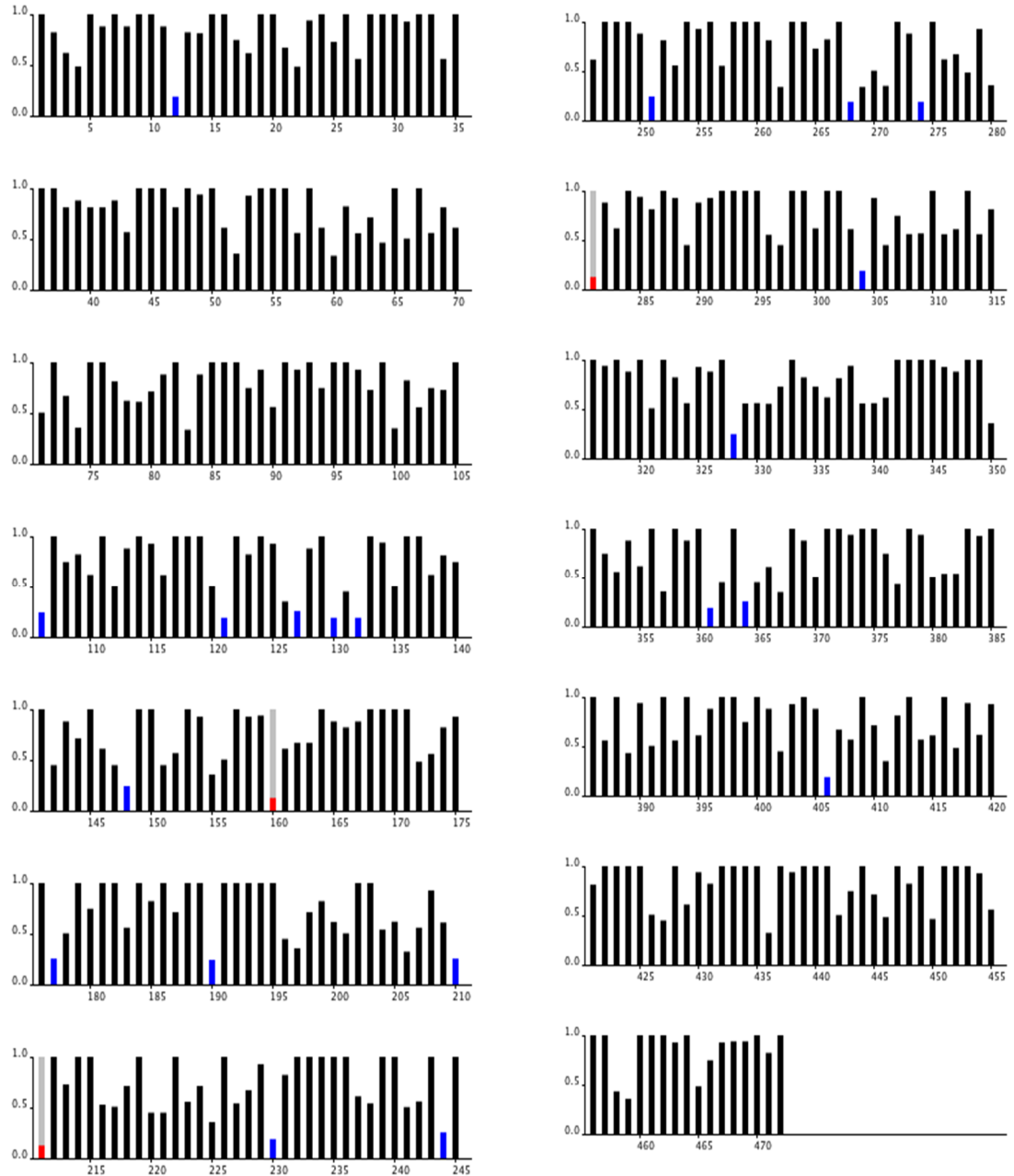


Figura 7. Análise de códons preferenciais da sequência dos genes (A) P3Gli, (B) P8Gli e (C) P9Gli em comparação com os códons preferenciais presentes no genoma de *K. marxianus* CCT 7735 (UFV-3) utilizando o programa Codon Usage Enhanced (CUE) adaptado para a levedura *K. marxianus* CCT 7735 (UFV-3). Barras vermelhas indicam códons não usuais em *K. marxianus* CCT 7735 (UFV-3) e que devem ser trocados para melhorar a expressão do gene. Barras azuis indicam códons que são presentes nesta levedura mas que podem ser trocados para melhorar a expressão.

CONCLUSÕES

O presente trabalho consistiu na construção de linhagens da levedura *K. marxianus* CCT 7735 (UFV-3) mutantes expressando genes heterólogos de β -glicosidases, visando o uso destas linhagens para estudos de hidrólise de material celulósico para produção de etanol. Para clonagem dos genes utilizou-se o sistema de expressão integrativo pKMCL, cuja expressão dos genes é dirigida pelo promotor *LAC4* após inserção da construção no genoma da levedura. Além disso, os genes neste vetor são transcritos *in frame* com a sequência do peptídeo sinal α -*mating factor* de *K. lactis*, que dirige a proteína para via de secreção. Amplificações por PCR confirmaram a integração das construções em cada linhagem transformante, e ensaios enzimáticos de β -glicosidase confirmaram a expressão e secreção das proteínas recombinantes. Os resultados apresentados mostram que as construções gênicas usando o vetor pKMCL foram novamente eficientes para a transformação da levedura.

A realização de ensaios enzimáticos a temperaturas de 50 °C, bem como a competição existente entre proteínas recombinantes e proteínas endógenas, podem ter contribuído para a redução na interferência da enzima β -glicosidase endógena de *K. marxianus* CCT 7735 (UFV-3) nos ensaios das proteínas recombinantes. A sequência de aminoácido predita da β -glicosidase de *K. marxianus* CCT 7735 (UFV-3) mostrou 99% de identidade com a mesma enzima de *K. marxianus* NBRC 1777.

A expressão e secreção das enzimas recombinantes causaram mudanças no perfil de secreção de proteínas nas cepas *K. marxianus* CCT 7735 (UFV-3) transformantes.

Apesar da presença de atividade de β -glicosidase no sobrenadante das culturas, a secreção destas enzimas pode ter sido prejudicada ou por processos de glicosilação, ou por competição com outras proteínas endógenas pela via de secreção, como pode ser observado pela presença de atividade equiparada ao do sobrenadante extracelular com ensaios realizados com proteínas retidas no espaço periplasmático das cepas transformantes.

A análise de sequências dos genes expressos em comparação com os códons preferenciais presentes em *K. marxianus* CCT 7735 (UFV-3) revelou que mesmo com algumas diferenças de códons preferenciais, os três genes podem ser expressos eficientemente nesta levedura.

Todos esses resultados representam informações importantes para aplicação desta levedura na produção de proteínas heterólogas pois proporcionam o aumento do rendimento da proteína de interesse.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Abdel-Banat, B.M.A., S. Nonklang, H. Hoshida & R. Akada. 2010. Random and targeted gene integrations through the control of non-homologous end joining in the yeast *Kluyveromyces marxianus*. *Yeast*. 27:29-39.
2. Altschul, S., W. Gish, W. Miller, E. Myers & D. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology*. 215:403-410.
3. Ball, M.M., A. Raynal, M. Guerineau & F. Iborra. 1999. Construction of efficient centromeric, multicopy and expression vectors for the yeast *Kluyveromyces marxianus* using homologous elements and the promoter of a purine-cytosine-like permease. *Journal Molecular Microbiology and Biotechnology*. 1:347-353.
4. Bergmann, J.C., O.Y.A. Costa, J.M. Gladden, S. Singer, R. Heins, P. D'haeseleer, B.A. Simmons & B.F. Quirino. 2014. Discovery of two novel β -glucosidases from an Amazon soil metagenomic library. 351:147-155.
5. Bhatia Y., S. Mishra & V.S. Bisaria. 2002. Microbial beta-glucosidases: cloning, properties, and applications. *Critical Reviews in Biotechnology*. 22:375–407.
6. Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72, 248-254.
7. Bragança, C.R.S., L.T. Colombo, A.S. Roberti, M.C.T. Alvim, S.A. Cardoso, K.C.P. Reis, S.O. de Paula, W.B. Silveira & F.M.L. Passos. 2014. Construction of recombinant *Kluyveromyces marxianus* UFV-3 to express dengue virus type 1 nonstructural protein 1 (NS1). *Applied Microbiology and Biotechnology*. 99:1191-203.
8. Çelik, E. & Çalık, P. 2012. Production of recombinant proteins by yeast cells. *Biotechnology Advances*. 30:1108-1118.
9. Chang, J.J., C.Y. Ho, F.J. Ho, T.Y. Tsai, H.M. Ke, H.T.C. Wang, H.L. Chen, M.C. Shih, C.C. Huang & W.H. Li. 2012. PGASO: A synthetic biology tool for engineering a cellulolytic yeast. *Biotechnology for Biofuels*. 5:53.
10. Chang, J.J., F.J. Ho, C.Y. Ho, Y.C. Wu, Y.H. Hou, C.C. Huang, M.C. Shih & W.H. Li. 2013. Assembling a cellulase cocktail and a cellodextrin transporter into a yeast host for CBP ethanol production. *Biotechnology for Biofuels*. 6:19.

11. Chen H, Li X, Ljundahl LG. 1994. Isolation and properties of an extracellular beta-glucosidase from the polycentric rumen Fungus *Orpinomyces* sp. strain PC-2. *Appl Environ Microbiol* 60:64-70.
12. Chen, H-L., Y-C. Chen, M-Y. J. Lu, J-J. Chang, H-T. C. Wang, H-M. Ke, T-Y. Wang, S-K. Ruan, T-Y. Wang, K-Y. Hung, H-Y. Cho, W-T. Lin, M-C. Shih & W-H. Li. 2012. A highly efficient β -glucosidase from the buffalo rumen fungus *Neocallimastix patriciarum* W5. *Biotechnology for Biofuels*. 5:24.
13. Chir, J. L., C. F. Wan, C.H. Chou & A.T. Wu. 2011. Hydrolysis of cellulose in synergistic mixtures of β -glucosidase and endo/exocellulase Cel9A from *Thermobifida fusca*. *Biotechnology Letters*. 33:777e782.
14. Colombo, L.T., Rosa, J.C.R, Bragança, C.R.S., Ignacchiti, R.P., Alvim, M.C.T., Silveira, W.B., de Queiroz, M.V., Bazzolli, D.M.S., Passos, F.M.L. 2014. Construction of a *Kluyveromyces lactis ku80⁻* Host Strain for Recombinant Protein Production: Extracellular Secretion of Pectin Lyase and a Streptavidin–Pectin Lyase Chimera. *Molecular Microbiology*. 56:319-28.
15. Colussi, P.A., & C.H. Taron. 2005. *Kluyveromyces lactis LAC4* promoter variants that lack function in bacteria but retain full function in *K. lactis*. *Applied and Environmental Microbiology*. 71:7092–7098.
16. Dashtban, M. & W. Qin. 2012. Overexpression of an exotic thermotolerant β -glucosidase in *Trichoderma reesei* and its significant increase in cellulolytic activity and saccharification of barley straw. *Microbial Cell Factories*. 11:63.
17. Del Pozo, MV, Fernández-Arrojo L, Gil-Martínez J, Montesinos A, Chernikova TN, Nechitaylo TY, Waliszek A, Tortajada M, Rojas A, Huws SA, Golyshina OV, Newbold CJ, Polaina J, Ferre M, Golyshin PN. 2012. Microbial β -glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. *Biotechnol Biofuels* 5:1-13.
18. Fonseca, G.G., E. Heinzle, C. Wittmann, A.K. Gombert. 2008. The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Applied Microbiology and Biotechnology*. 79:339-54.
19. Friberg, M., P. von Rohr & G. Gonnet. 2004. Limitations of codon adaptation index and other coding DNA-based features for prediction of protein expression in *Saccharomyces cerevisiae*. *Yeast*. 21:1083-93.

20. Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel & A. Bairoch. 2005. Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press.
21. Hong, J., Y. Wang, H. Kumagai & H. Tamaki. 2007. Construction of thermotolerant yeast expressing thermostable cellulase genes. *Journal of Biotechnology*.130:114-123.
22. Kelley, L.A. & M.J.E Sternberg. 2009. Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols* 4, 363 – 371.
23. Kohen, A., T. Jonsson & J.P. Klinman. 1997. Effects of protein glycosylation on catalysis: changes in hydrogen tunneling and enthalpy of activation in the glucose oxidase reaction. *Biochemistry*. 36:2603-11.
24. Kooistra, R., Hooykaas, P. J. J., Steensma, H. Y. 2004. Efficient gene targeting in *Kluyveromyces lactis*. *Yeast*. 21:781–792
25. Lachance, M.A. 1998. *Kluyveromyces* van der Walt emend. van der Walt. In: Kurtzman CP, Fell JW (eds) *The yeasts: a taxonomic study*, 4th edn. Elsevier, Amsterdam, pp 227–247.
26. Larkin, M. A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, R. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*. 23:2947e2948.
27. Li, D., X. Li, W. Dang, P.L. Tran, S-H. Park, B-C. Oh, W-S. Hong, J-S. Lee & K.-H. Park. 2013. Characterization and application of an acidophilic and thermostable β -glucosidase from *Thermofilum pendens*. *Journal of Bioscience and Bioengineering*. 115:490-496.
28. Lymar, E.S. & V. Renganathan. 1995. Purification and Characterization of a Cellulose-Binding β -Glucosidase from Cellulose-Degrading Cultures of *Phanerochaete chrysosporium*. *Applied and Environmental Microbiology*. 61:2976-2980.
29. Mattanovich, D., B. Gasser, H. Hohenblum & M. Sauer. 2004. Stress in recombinant protein producing yeast. *Journal of Biotechnology*. 113:121-135.
30. Mingardon, F., S. Perret, A. Bélaich, C. Tardif, J-P. Bélaich & H-P. Fierobe. 2005. Heterologous Production, Assembly, and Secretion of a Minicellulosome by *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology*. 71:1215-1222.

31. Minitab, I. 2006. MINITAB statistical software. Version: Release, 15.
32. New England Biolabs, K. lactis Protein Expression Kit, Instruction Manual. 26p. 2009.
33. Raynal, A. & M. Guerineau. 1984. Cloning and expression of the structural gene for beta-glucosidase of *Kluyveromyces fragilis* in *Escherichia coli* and *Saccharomyces cerevisiae*. *Molecular General Genetics*. 195:108-115.
34. Rocha, S.N., J. Abrahão-Neto, M.E. Cerdán, M.I. González-Sico & A.K. Gombert. 2010. Heterologous expression of glucose oxidase in the yeast *Kluyveromyces marxianus*. *Microbial Cell Factories*. 9:1-12.
35. Rocha, S.N., J. Abrahão-Neto, M.E. Cerdán, A.K. Gombert & M.I. González-Sico. 2011. Heterologous expression of a thermophilic esterase in the *Kluyveromyces* yeasts. *Applied Microbiology and Biotechnology*. 89:375-385.
36. Rouwenhorst, R.J., Visser, L.E., Van der Baan, A.A., Scheffers, W.A. & Van Dijken, J.P. 1988. Production, Distribution, and Kinetic Properties of Inulinase in Continuous Cultures of *Kluyveromyces marxianus* CBS 6556. *Applied and Environmental Microbiology*. 54:1131-1137.
37. Rubin, E.M. 2008. Genomics of cellulosic biofuels. *Nature Reviews*. 454:841-845.
38. Sadeghi, M., H. Naderi-Manesh, M. Zarrabi & B. Ranjbar. 2006. Effective factors in thermostability of thermophilic proteins. *Biophysical Chemistry*. 119:256-270.
39. Sanchez M., Iglesias, F.J., Santamaria, C., Dominguezl, A. 1993. Transformation of *Kluyveromyces lactis* by electroporation. *Applied and Environmental Microbiology*. 59:2087–2092.
40. Sambrook, J., Fritsch, E.F., Maniatis, T. 1989. *Molecular cloning: a laboratory manual* (2nd ed.). Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
41. Sharp, P.M. & W.H. Li. 1987. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*. 15:1281–1295.
42. Silveira, W.B., Diniz, R.H.S., Cerdan, M.E., Gonzalez-Siso, M.I., et al. 2014. Genomic Sequence of the Yeast *Kluyveromyces marxianus* CCT 7735 (UFV-3), a Highly Lactose-Fermenting Yeast Isolated from the Brazilian Dairy Industry. *Genome Announcements*. 2:e01136-14-e01136-14.
43. Silveira, W.B., F.J.V. Passos, H.C. Mantovani & F.M.L. Passos. 2005. Ethanol production from cheese whey permeate by *Kluyveromyces marxianus* UFV-3: a

- flux analysis of oxido-reductive metabolism as a function of lactose concentration and oxygen levels. *Enzyme and Microbial Technology*. 36:930-6.
44. Souza, R.A. Obtenção de Inoculante e de Coquetel Enzimático Lignocelulolítico a partir de Comunidades Microbianas Termofílicas. 2012. 56f. Dissertação (mestrado). Universidade Federal de Viçosa, Viçosa, MG, 2012.
 45. Uchima, C.A., G. Tokuda, H. Watanabe, K. Kitamoto & M. Arioka. 2012. Heterologous Expression in *Pichia pastoris* and Characterization of an Endogenous Thermostable and High-Glucose-Tolerant-Glucosidase from the Termite *Nasutitermes takasagoensis*. *Applied and Environmental Microbiology*. 78:4288-4293.
 46. Wilson, C.M., M. Jr. Rodriguez, C.M. Johnson, S.L. Martin, T.M. Chu, R.D. Wolfinger, L.J. Hauser, M.L. Land, D.M. Klingeman, M.H. Syed, A.J. Ragauskas, T.J. Tschaplinski, J.R. Mielenz & S.D. Brown. 2013. Global transcriptome analysis of *Clostridium thermocellum* ATCC 27405 during growth on dilute acid pretreated Populus and switchgrass. *Biotechnology for Biofuels*. 6:179.
 47. Yanase, S., T. Hasunuma, R. Yamada, T. Tanaka, C. Ogino, H. Fukuda & A. Kondo. 2010. Direct ethanol production from cellulosic materials at high temperature using the thermotolerant yeast *Kluyveromyces marxianus* displaying cellulolytic enzymes. *Applied Microbiology and Biotechnology*. 88:381-388.
 48. Yoshida, E., M. Hidaka, S. Fushinobu, T. Koyanagi, H. Minami, H. Tamaki, M. Kitaoka, T. Katayama & H. Kumagai. 2010. Purification, crystallization and preliminary x-ray analysis of b-glucosidase from *Kluyveromyces marxianus* NBRC 1777. *Biochemical Journal*. 431:39-49.
 49. Yu, N. & Y-S. Cheng. 2004. Codon Usage Enhancer (CUE): A JAVA-based Freeware for Manipulating Codon Usage. Published on The 9th Symposium on Recent Advances in Biophysics (<http://biophys.sinica.edu.tw/Symposium/9TH/>) in Institute of Biological Chemistry, Academia Sinica, Taiwan University, Taiwan.

CONCLUSÕES GERAIS

O presente trabalho teve como objetivo desenvolver uma estratégia para isolamento e clonagem de genes em linhagens hospedeiras recombinantes para serem vetores na produção de enzimas envolvidas em processos biotecnológicos da produção de bioetanol.

No primeiro capítulo desta tese, a construção de uma biblioteca metagenômica a partir de um consórcio microbiano termofílico, com aproximadamente 135.000 clones contendo cerca de 3,5 Gpb de DNA metagenômico, mostrou que a metodologia utilizada promoveu o enriquecimento da biblioteca com clones que apresentam atividades de CMCCase, xilanase e β -glicosidases. A comparação das atividades com o sequenciamento dos fosmídeos comprovou a existência de genes codificando estas enzimas, e possibilitou a geração de um banco de sequências que constitui fonte de genes a serem explorados, seja pela sua relação com atividades lignocelulolíticas ou pelo envolvimento com a síntese de outras biomoléculas com possível interesse biotecnológico.

Na segunda etapa desta tese, descrita no capítulo 2, foi obtida uma linhagem de *K. marxianus* CCT 7735 (UFV-3) mutante para o gene uracila (*ura3⁻*) utilizando-se a técnica *split-marker*, até então inédita para esta levedura e com esta finalidade. A partir da obtenção do mutante *ura3⁻*, iniciou-se a construção do vetor pKmLPG para ser usado como sistema de expressão para essa levedura, uma vez que a marca de seleção contida neste vetor é o gene *URA3*. O diferencial e, considerado por nossa equipe de trabalho, promissor deste sistema de expressão, é o uso da sequência do promotor da enzima Endo-Poligalacturonase, que se mostrou sob forte indução, identificada por excelente secreção da enzima sob condições estabelecidas em trabalho de doutorado realizado em nosso laboratório. Dada as circunstâncias de tempo para confirmação da construção do vetor pKmLPG por sequenciamento, não foi possível utilizá-lo, tão pouco a linhagem CCT 7735 (UFV-3) *ura3⁻*, para expressar os genes isolados da biblioteca metagenômica descritos no primeiro capítulo. No entanto, tanto a linhagem mutante, quanto o vetor poderão ser usados em trabalhos futuros desenvolvidos em nosso laboratório.

Na terceira etapa, foi estudada e expressão de três genes de β -glicosidase, isolados a partir da biblioteca metagenômica, na levedura *K. marxianus* CCT 7735 (UFV-3) utilizando o sistema de expressão pKMCL. Os transformantes para os genes P3Gli, P8Gli e P9Gli foram confirmados por PCR e os resultados da atividade de β -

glicosidase indicam que houve expressão diferencial das cepas transformantes em relação à cepa controle *K. marxianus* CCT 7735 (UFV-3), mas houve retenção das proteínas recombinantes no espaço periplasmático. Além disso, a expressão e secreção das proteínas heterólogas causaram mudanças no perfil de secreção de proteínas das cepas recombinantes de *K. marxianus* CCT 7735 (UFV-3). A presença de atividade no sobrenadante extracelular mesmo em meio a esses fatores indicam que a *K. marxianus* CCT 7735 (UFV-3) foi capaz de expressar e secretar as enzimas β -glicosidases recombinantes.

O conhecimento aqui apresentado sobre a fonte de novos genes presentes na biblioteca metagenômica, sobre a manipulação genética e a expressão heteróloga de β -glicosidases em *K. marxianus* CCT 7735 (UFV-3) pode ser empregado no uso desta linhagem como hospedeira na expressão de diferentes genes lignocelulolíticos em vetor de expressão desenvolvido especificamente para esta levedura, na tentativa de usar as linhagens recombinantes capazes de secretar celulasas em processos de hidrólise e fermentação dos açúcares resultantes para aplicação na produção de etanol celulósico.

PERSPECTIVAS PARA TRABALHOS FUTUROS

A partir dos resultados apresentados nesta tese de doutorado, seguem-se sugestões para estudos futuros que poderão aprofundar o tema abordado aqui e contribuir para aplicação da levedura *K. marxianus* CCT 7735 (UFV-3) na produção de etanol celulósico:

- Analisar funcionalmente mais clones da biblioteca metagenômica construída em busca de novos genes para serem isolados, analisados e clonados na levedura *K. marxianus* CCT 7735 (UFV-3). Mediante interesse por uma enzima específica a ser produzida por *K. marxianus*, produzir uma sonda utilizando a sequência do sítio catalítico típico de família glicosil hidrolase de interesse e realizar a triagem dos clones por análise baseada em sequência;
- Validar o vetor de expressão pKmLPG utilizando um gene de interesse como de celulasas ou xilanases, verificando a eficácia na secreção da proteína heteróloga; e confirmar a seleção por complementação da linhagem *ura3⁻*;
- Realizar estudos estruturais e de caracterização das β -glicosidases heterólogas, bem como da enzima endógena de *K. marxianus* CCT 7735 (UFV-3), podendo relacionar as diferenças entre elas e delas com outras β -glicosidases relatadas na literatura;
- Analisar os níveis de glicosilação das proteínas expressas em *K. marxianus* CCT 7735 (UFV-3) para compreender se estas modificações interferem na atividade e secreção da enzima;
- Realizar mutações silenciosas na sequência da β -glicosidase, ou de qualquer outra enzima de interesse isolada na biblioteca, alterando os códons raros identificados pelo programa para manipulação de *codon usage* em *K. marxianus* CCT 7735 (UFV-3) por outros comuns a essa levedura e verificar se há alteração no nível de expressão;
- Realizar cultivos em cultura contínua com taxa de diluição de $0,1 \text{ h}^{-1}$ para que seja alcançada altas taxas de expressão do gene heterólogo ao se utilizar o sistema de expressão pKmLP para clonagem gênica.