Universidade de São Paulo

Faculdade de Medicina de Ribeirão Preto

Programa de Pós-Graduação em Bioquímica

Departamento de Bioquímica e Imunologia

Lílian dos Santos Castro

Análise global da expressão gênica durante a formação de celulases pelo fungo *Trichoderma reesei*

Ribeirão Preto – São Paulo

2015

LÍLIAN DOS SANTOS CASTRO

Análise global da expressão gênica durante a formação de celulases pelo fungo *Trichoderma reesei*

Tese apresentada ao programa de Pós-graduação em Bioquímica da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Doutora em Ciências.

Área de Concentração: Bioquímica

Orientador: Prof. Dr. Roberto do Nascimento Silva

Coorientadora: Dra. Gabriela Felix Persinoti

Ribeirão Preto – São Paulo

2015

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catalogação da Publicação
Serviço de Documentação
Faculdade de Medicina de Ribeirão Preto - USP

Ficha Catalográfica

Castro, Lílian dos Santos Castro

Análise global da expressão gênica durante a formação de celulases pelo fungo *Trichoderma reesei*/Lílian dos Santos Castro; Orientador: Profº. Dr. Roberto do Nascimento Silva; Coorientadora: Dra. Gabriela Felix Persinoti. Ribeirão Preto – São Paulo, 2015.

p.290: il. ; 30 cm

Tese (Doutorado) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2015. Área de concentração: Bioquímica.

1. *Trichoderma reesei*; RNA-seq; Fator de transcrição XYR1; Expressão gênica; Enzimas.

Nome: CASTRO, Lílian dos Santos

Título: Análise global da expressão gênica durante a formação de celulases pelo fungo *Trichoderma reesei*.

Tese apresentada ao programa de Pós-graduação em Bioquímica da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Doutora em Ciências.

Área de Concentração: Bioquímica

Aprovado em: ___/___/___

Banca Examinadora

Dr. Roberto do Nascimento Silva — Instituição: FMRP-USP

Julgamento: _____ Assinatura: _____

Dra. Maria de Lourdes Teixeira de Moraes Polizeli — Instituição: FFCLRP-USP

Julgamento: _____ Assinatura: _____

Dra. Tie Koide — Instituição: FMRP-USP

Julgamento: _____ Assinatura: _____

Dr. Cirano José Ulhoa — Instituição: UFG

Julgamento: _____ Assinatura: _____

Dra. Maria Celia Bertolini — Instituição: IQ-UNESP-Araraquara

Julgamento: _____ Assinatura: _____

Ribeirão Preto – São Paulo

2015

Apoio e Suporte Financeiro

Este trabalho foi realizado com o apoio financeiro das seguintes entidades e instituições:

1) Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (Processo 2010/15683-8).

2) Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES.

3) Fundação de Apoio ao Ensino, Pesquisa e Assistência – FAEPA (FMRP/USP).

4) Faculdade de Medicina de Ribeirão Preto - FMRP/USP

Normalização Adotada

Esta tese está de acordo com:

Associação Brasileira de Normas Técnicas (ABNT): NBR 6023, NBR 6024, NBR 6027 e NBR 6028.

Universidade de São Paulo. Sistema Integrado de Bibliotecas da USP. Diretrizes para apresentação de dissertações e teses da USP: documento eletrônico e impresso Parte I (ABNT)/Sistema Integrado de Bibliotecas da USP; Vânia Martins Bueno de Oliveira Funaro, coordenadora... [et al.] . - - 2. ed. rev. ampl. - - São Paulo: Sistema Integrado de Bibliotecas da USP, 2009. 102 p. - - (Cadernos de Estudos ; 9).

“No que diz respeito ao desempenho, ao compromisso, ao esforço, à dedicação, não existe meio termo. Ou você faz bem-feito ou não faz.”

(Ayrton Senna)

Ao meu esposo, **Sidon Oliveira Cardoso**,

Por sua existência,

Por estar sempre ao meu lado,

Pelo companheirismo, respeito e incentivo,

Pela paciência e sabedoria para transmitir seus conhecimentos e experiências de vida,

Sempre presente em minha vida,

Com amor e carinho,

Dedico-lhe este trabalho.

AGRADECIMENTOS

Primeiramente, Deus, por sua benevolência, inteligência suprema, fonte de sabedoria, amor, compreensão, por me conceder a vida e a oportunidade da existência, de modo que esta me traz transformação e evolução.

Ao meu marido, Sidon Oliveira Cardoso, por torcer tanto por mim, por rir e chorar comigo e por não desistir de acreditar em mim. Por estar sempre ao meu lado. Pelo nosso amor ter resistido a distância. Pela compreensão e paciência. Agradeço a você por ter feito de mim uma pessoa melhor, mais paciente e tolerante. Você é na Minha Vida meu Porto Seguro, meu Destino, meu Passado, Presente e Futuro! Eu amo muito VOCÊ! Meu eterno amor...

Aos meus pais, Alcione da Silva Castro e Maria Dalva dos Santos Castro, e aos meus irmãos, Fernando dos Santos Castro e Letícia dos Santos Castro. Família esta que amo muito e que só tenho a agradecer por me incentivarem na fase de adquirir conhecimentos dando-me força para este título que estou conquistando para meu crescimento profissional. Crescimento este que me fará analisar a cada dia com novos olhos o significado da minha pessoa diante do mundo.

Um agradecimento especial e sincero ao Prof. Dr. Roberto do Nascimento Silva, meu Orientador, pela orientação, dedicação à minha formação científica e profissional, pela paciência que teve nos momentos mais difíceis, pelos seus conhecimentos repassados durante todo o desenvolvimento do projeto e pela amizade. Agradeço-o ainda por me mostrar que não existem limites para se obter conhecimentos.

À minha Coorientadora, Dra. Gabriela Felix Persinoti, obrigada pela confiança, amizade, dedicação, incentivo, discussões científicas, oportunidades e pelo seu exemplo de profissionalismo. A sua orientação foi fundamental para o desenvolvimento das análises de Bioinformática deste trabalho.

À Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP/USP) e ao Departamento de Bioquímica e Imunologia que forneceram suporte e estrutura para a realização deste trabalho.

Ao Laboratório de Biotecnologia Molecular (LBM), do qual fiz parte de sua inauguração e me tornei a primeira aluna de Pós-Graduação *stricto sensu* (Doutorado). Local de fundamental importância para o desenvolvimento do projeto.

Aos colegas e amigos do Laboratório LBM: Amanda Cristina Campos Antoniêto Eriston Vieira Gomes, Karoline Maria Vieira Nogueira, Mariana do Nascimento Costa, Renato Graciano de Paula e Wellington Ramos Pedersoli. Obrigada pela amizade, pelo carinho e convivência. A vocês, Wellington e Renato, ofereço a letra da música "amizade sincera" – autoria de Renato Teixeira.

À técnica do Laboratório LBM, Zuleica Aparecida Soares de Moraes, por todo o auxílio técnico, no preparo de materiais, nos experimentos, preparo de vidrarias e soluções, extrações de RNA, e, acima de tudo, pela amizade e carinho. Minha primeira companheira e amiga do Laboratório.

Ao Laboratório de Bioquímica e Biologia Molecular de Micro-organismos e do Laboratório de Genética e Biologia Molecular de Fungos (Grupo GBM USP - Genética e Biologia Molecular), principalmente aos professores Dr. Antônio Rossi e Dra. Nilce Maria Martinez Rossi pela valiosa colaboração, assim como por ter disponibilizado os seus laboratórios e equipamentos para o desenvolvimento das análises de Bioinformática. Agradeço também pelas amizades adquiridas durante a convivência nesse laboratório, principalmente: Aline Cruz, Glauce Trevisan, Elza Lang, Gabriela Persinoti, Hemelin Ludmila, Larissa Gomes, Nalu Peres, Pablo Sanches, Rodrigo Silva e a técnica Vanderci Oliveira “Cuca”. Deixo meu muito obrigada da “Agregada lá de cima”!

Aos meus amigos Aline Helena da Silva Cruz e Rodrigo da Silva Santos pela amizade conquistada no Laboratório “vizinho”. Quero agradecer a vocês por tudo. Em especial, por estarem ao meu lado, sempre. Saibam que eu também quero fazer por vocês o que for possível. Eu, hoje, agradeço a Deus por ter conhecido vocês. Com vocês aprendi que não podemos ter medo de lutar para ser feliz, devemos vencer nossos obstáculos, pois Deus sempre está do nosso lado. Disponham da minha amizade sincera. Meu eterno agradecimento!

Aos docentes e funcionários do Departamento de Bioquímica e Imunologia. Principalmente ao Profº. Dr. Eduardo Brandt de Oliveira por me proporcionar o conhecimento

não apenas teórico, mas a manifestação do caráter e afetividade da educação no processo de formação profissional, não somente por ensinar, mas por saber como fazer o processo do aprender. Você é um exemplo de como ser um professor.

À secretária Ivone pela atenção e dedicação.

Ao Pós-doutorando do Laboratório de Biologia de Sistemas de Microrganismo (FMRP), Rafael Silva Rocha, e atual professor da FMRP, pela colaboração, dedicação e ajuda nas análises deste projeto. Sua experiência científica é fantástica, e é um exemplo a ser seguido. A este mesmo Laboratório destaco um agradecimento especial à Lívia Soares Zaramela, pela sua amizade, carinho e conhecimentos compartilhados. Admiro-a muito!

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo financiamento concedido ao projeto.

À CAPES, Coordenação de Aperfeiçoamento de Pessoal de nível Superior, pela bolsa de doutorado concedida.

Aos membros da banca, agradeço a disposição para avaliar e discutir o presente trabalho.

Obrigada a todos por participarem desta minha etapa, pois direta, ou indiretamente, fizeram-me crescer, tanto pessoalmente como profissionalmente. Enfim, quero demonstrar o meu agradecimento, a todos aqueles que, de um modo ou de outro, tornaram possível a realização desta presente tese.

A todos, o meu sincero e profundo **Muito Obrigada!**

RESUMO

CASTRO, L. S. **Análise global da expressão gênica durante a formação de celulases pelo fungo *Trichoderma reesei*** 2015. 290f. Tese (Doutorado) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto-SP, 2015.

O fungo filamentoso *Trichoderma reesei* (*Hypocrea jecorina*) é o principal produtor de enzimas celulolíticas e hemicelulolíticas utilizadas para a produção de biocombustíveis a partir de biomassa lignocelulósica. Para que o custo de produção de etanol de segunda geração (2G), em escala industrial, seja competitivo, uma mistura eficiente de enzimas hidrolíticas é necessária para degradar a biomassa vegetal em açúcares fermentescíveis. Dada a crescente demanda pelo desenvolvimento de processos que reduzam os custos de enzimas de interesse biotecnológico, o presente trabalho se propôs a analisar o perfil global do transcriptoma de *T. reesei* em presença de celulose, soforose e glicose como fontes de carbono, a fim de prover um melhor entendimento de como essas enzimas são produzidas por esse microrganismo. Para isso, foi utilizada a abordagem RNA-seq e duas linhagens de *T. reesei*: QM9414 (parental) e *Δxyr1* (mutante). No Capítulo I, foram anlisados 22 genes que codificam celulases e xilanases, com o intuito de verificar a regulação da expressão gênica pelo fator de transcrição XYR1 em diferentes fontes de carbono. Foi observada uma dependência da fonte de carbono e redução da expressão destes genes que codificam celulases e xilanases quando se comparou a linhagem mutante com a parental, cultivados na presença de celulose e soforose. Na presença de glicose, a maior parte dos genes analisados apresentou aumento de expressão no mutante *Δxyr1*. Por meio de análises *in silico* foi identificado o elemento regulatório *cis* para o fator de transcrição XYR1 usando quatro conjuntos de dados de genes dependente-condição envolvendo os experimentos RNA-seq e qPCR-TR. Foram identificados dois motivos com o consenso de ligação proposto para o regulador XYR1 (nomeado PWM_{XYR1}). Ao usar esses motivos identificados, foi analisada a presença e disposição dos putativos elementos regulatórios *cis* nesses conjuntos de dados, sendo que sítios com intervalos curtos foram mais associados a promotores dependentes de XYR1 do que sítios simples. Além disso, a abordagem utilizada permitiu a identificação de sítios de ligação XYR1, na região promotora dos genes *cel7a* e *xyn1* e mapear a potencial sequência alvo do regulador na região promotora do gene *cel6a*. Adicionalmente, sete outros promotores (genes *cel7b*, *cel61a*, *cel61b*, *cel3c*, *cel3d*, *xyn3* e *swo*) apresentaram sítios de ligação putativos de XYR1. Usando a arquitetura do regulador PWM_{XYR1}, foram identificados potenciais genes alvos de regulação direta por XYR1 no genoma de *T. reesei*. O mapeamento do referido sítio foi realizado, também, nos genes diferencialmente expressos na linhagem mutante *Δxyr1*, destacando que a regulação indireta desempenha um papel fundamental nas vias de sinalização. Estes resultados fornecem novas informações sobre os mecanismos de sinalização mediados por XYR1 na regulação de promotores de celulases. No Capítulo II, foi analisado o transcriptoma global da linhagem parental (QM9414) em diferentes fontes de carbono: celulose; soforose e glicose. Aplicando limites rigorosos de corte, 2060 genes foram identificados como diferencialmente expressos em pelo menos uma das fontes de carbono analisadas. Clusterização hierárquica desses genes diferencialmente expressos identificaram-se três possíveis regulons, representando 123 genes controlados por celulose, 154 genes controlados por soforose e 402 genes controlados por glicose. A análise da rede regulatória demonstrou a inter-relação existente entre as condições, permitindo a identicação de 75 genes específicos do crescimento em soforose e 107 de celulose. Esses resultados revelaram novos genes envolvidos na degradação da celulose, tais como: proteínas acessórias, transportadores,

fatores de transcrição e *CAZymes* que respondem especificamente a presença de celulose ou soforose. No Capítulo III, analisou-se o perfil transcricional da linhagem mutante *Δxyr1* comparando com a linhagem parental (QM9414), na presença das três fontes de carbono. As análises de expressão gênica revelaram que 2185 genes foram diferencialmente expressos na condição celulose, 2124 genes na condição soforose e 46 genes em glicose. Esses genes foram utilizados para analisar a inter-relação existente entre as condições, com destaque para grande número de genes exclusivos nas condições de indução (celulose e soforose). Por meio da clusterização hierárquica foram identificados 6 possíveis regulons, sendo 3 deles compostos por genes up-regulados e 3 regulons por genes down-regulados. Na ausência do regulador XYR1, a expressão de genes *CAZys*, fatores de transcrição, transportadores, entre outros, foram afetadas de modo dependente da fonte de carbono. Essas análises permitiram verificar que o fator de transcrição XYR1 é fundamental no processo de indução de enzimas de interesse biotecnológico, além de participar de outros processos bioquímicos/moleculares. Desse modo, todos os resultados obtidos fornecem informações sobre os eventos moleculares envolvidos na regulação da expressão gênica durante a adaptação de *T. reesei* frente a diferentes condições ambientais como: diferentes fontes de carbono e necessidades nutricionais. Contribuindo, assim, para uma melhor caracterização desse fungo filamentoso em relação à produção de enzimas e o desenvolvimento de mutantes metabólicos para utilização industrial.

Palavras-chave: *Trichoderma reesei*; RNA-seq; Fator de transcrição XYR1; Expressão gênica; Enzimas.

ABSTRACT

CASTRO, L. S. **Global analysis of gene expression during the formation of cellulases by *Trichoderma reesei*** 2015. 290f. Thesis (PhD) - Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto-SP, 2015.

The filamentous fungus *Trichoderma reesei* (*Hypocrea jecorina*) is a major producer of cellulolytic enzymes used for biofuels production from lignocellulosic biomass. In order to achieve a low cost second generation ethanol production in industrial scale, an efficient mix of hydrolytic enzymes are required for the degradation of plant biomass into fermentable sugars. Given the growing demand for development of processes in order to reduce the cost of enzymes production, the present work proposes a large-scale transcriptomic analysis of *T. reesei* in presence of cellulose, sophorose, and glucose as carbon sources, in order to provide insights about the mechanisms of enzymes production by this microorganism. For this purpose, RNA-seq approach was employed, as well, as two *T. reesei* strains: QM9414 (parental) and Δ*xyr1* (mutant). In Chapter I, 22 genes encoding cellulases and xylanases were analyzed regarding the regulation through the transcription factor XYR1. This regulation seems to occur in a carbon source dependence manner, and the expression of cellulases and xylanases was diminished in the mutant strain compared to the parental strain in the presence of cellulose and sophorose. In glucose, most of the analyzed genes showed higher expression in the mutant compared to the parental strain. In silico analysis identified a cis regulatory element for XYR1 using four datasets of condition-dependent genes based on RNA-seq and qPCR experiments. Two binding motifs have been identified with the proposed consensus for the XYR1 (named PWMXYR1). Using these motifs, the presence and arrangement of putative cis regulatory elements was analyzed, and the results indicate that the sites with short intervals were stronger associated with XYR1 dependent promoters than single sites, even with high scores. Moreover, this approach allowed the identification of XYR1 binding sites in the promoter region of *cel7a* and *xyn1*, as well as, map the potential target sequence in the promoter of *cel6a* gene. In addition, seven other promoters from genes *cel7b*, *cel61a*, *cel61b*, *cel3c*, *cel3d*, *xyn3* and *swo* presented XYR1 putative binding sites. Using the cis-regulatory architecture of PWM_{XYR1}, potential targets for direct regulation by XYR1were identified in a genome wide manner. The putative binding sites were also mapped in the differentially expressed genes identified in the mutant strain Δ*xyr1*, suggesting that indirect regulation plays a key role in signaling pathways. Taken together, the data provided here provides important information regarding signaling mechanisms mediated by XYR1 in the regulation of cellulases promoters. In Chapter II, the transcriptome of the parental strain (QM9414) was analyzed in three carbon sources, cellulose, sophorose and glucose. By applying a stringent cut-off threshold, 2,060 genes were identified as differentially expressed in at least one of the carbon sources analyzed. Hierarchical clustering of differentially expressed genes identified three possible regulons, representing 123 genes controlled by cellulose, 154 genes controlled by sophorose and 402 genes controlled by glucose. The analysis of the regulatory network demonstrated the interrelation between the conditions, allowing the identification of 75 genes specifically expressed in soforose and 107 in cellulose. These results revealed new players involved in cellulose degradation, such as accessory proteins, transporters, transcription factors and CAZymes, that specifically respond to the presence of either cellulose or sophorose. In Chapter III, a genome wide comparative transcriptional analysis were performed between the mutant Δ*xyr1* strain and the parental strain (QM9414) using the three carbon sources. Gene expression analysis revealed 2,185 genes differentially expressed in

cellulose, 2124 genes in sophorose and 46 genes in glucose. These differentially expressed genes were used to analyze the relationship between the conditions, revealing a greater number of exclusive genes in theinducing conditions (cellulose and sophorose). The hierarchical clustering analysis revealed 6 possible regulons, being three composed by up-regulated genes and 3 regulons composed by down-regulated genes. In the absence of the regulator XYR1, the expression of CAZy genes, transcription factors among other genes were affected in a carbon-dependent manner. These analyses showed that the transcription factor XYR1 has an essential role in induction of enzymes of biotechnological interest and participates in other biochemical/molecular processes. Taken together these results provide important information about the molecular events, involved in gene expression regulation during *T. reesei* adaptation to different environments such as: carbon sources and/or nutritional needs, thus contributing to a better understanding of this filamentous fungus regarding the the production of enzymes and the development of mutants for industrial applications.

Keywords: *Trichoderma reesei*; RNA-seq; XYR1 transcription factor; Gene expression; Enzymes.

LISTA DE FIGURAS

FIGURA 1 – Estrutura biomassa lignocelulósica. 37
FIGURA 2 – Representações da estrutura da celulose. 39
FIGURA 3 – Representação esquemática da estrutura da hemicelulose. 40
FIGURA 4 – Estrutura da celulose antes e depois do pré-tratamento (hidrólise ácida). 44
FIGURA 5 – Alterações estruturais do complexo celulose-hemicelulose-lignina determinadas pelo pré-tratamento. 45
FIGURA 6 – Descrição geral do processo para a produção de bioetanol de primeira e segunda geração. 47
FIGURA 7 – Usinas autorizadas pela ANP por região. 48
FIGURA 8 – Primeira usina de etanol celulósico do Brasil. 49
FIGURA 9 – Palha de cana-de-açúcar sendo estocada para produção de etanol de segunda geração (GranBio). 50
FIGURA 10 – Fotos da Usina de Crescentino em operação na Itália. 51
FIGURA 11 – Via metabólica simplificada de conversão de xilose em etanol. 52
FIGURA 12 – Etapas do processo de produção do etanol celulósico. 53
FIGURA 13 – Ciclo de vida assexuada de *T. reesei*. 55
FIGURA 14 – Genealogia dos mutantes de *T. reesei* QM6a. 57
FIGURA 15 – Representação esquemática da morfologia e as vias propostas a síntese e secreção de proteínas em *T. reesei*. 60
FIGURA 16 – Representação esquemática da hidrólise e transglicosilação (reações catalisadas por β-glicosidases). 61
FIGURA 17 – Um esquema simplificado da visão atual sobre a degradação enzimática da celulose. 62
FIGURA 18 – Visão esquemática da degradação de um sistema hemicelulolítico. 63
FIGURA 19 – Modelo de indução de celulases em *T. reesei*. 65
FIGURA 20 – Representação esquemática dos fatores de transcrição que afetam expressão de celulases e hemicelulases em *T. reesei*. 68
FIGURA 21 – Esquema das etapas experimentais realizadas neste estudo. 74
FIGURA 22 – Representação esquemática da abordagem usada para a descoberta do motivo XYR1 em *T. reesei*. 81
FIGURA 23 – Esquema da estratégia utilizada para obtenção dos RNAs para sequenciamento. 84
FIGURA 24 – Sequenciamento simplificado para o *Illumina Hiseq™ 2000*. 84
FIGURA 25 – Etapas para preparo das bibliotecas de cDNAs. 86
FIGURA 26 – Geração de *cluster paired end*. 88
FIGURA 27 – Sequenciamento plataforma *Illumina HiSeq™ 2000*. 89
FIGURA 28 – Sequenciamento das duas extremidades do fragmento de cDNA (*Read* 1 e *Read* 2). 90
FIGURA 29 – Análises dos dados de RNA-seq por ferramentas bioinformáticas. 92
FIGURA 30 – Método de contagem das *reads*. 94
FIGURA 31 – Alinhamentos e contagens das *reads*. 95

FIGURA 32 – *Heatmap* da expressão de genes celulolíticos e xilanolíticos de *T. reesei* (linhagens: parental QM9414 e mutante Δ*xyr1*). 107
FIGURA 33 – Diagrama de *Venn* de genes celulolíticos e xilanolíticos diferencialmente expressos em *T. reesei* (Δ*xyr1*/QM9414). 108
FIGURA 34 – Representação esquemática de putativos sítios de ligação de XYR1 (5'-GGCWWW-3') na região 5'-*upstream* de genes celulolíticos e xilanolíticos. 109
FIGURA 35 – Motivos de ligação de XYR1 identificados em celulases. 111
FIGURA 36 – Distribuição do Score do sítio de ligação de XYR1 em diferentes conjuntos de dados (região promotora)..... 112
FIGURA 37 – Sítio de ligação de XYR1 em diferentes conjuntos de dados (região promotora). 113
FIGURA 38 – Identificação de sítios de ligação XYR1 nos promotores alvos. 114
FIGURA 39 – Definindo as categorias dos genes potencialmente regulados diretamente por XYR1. 115
FIGURA 40 – Réplicas biológicas das amostras celulose, soforose e glicose para a linhagem parental utilizadas nas análises de RNA-seq. 122
FIGURA 41 – Análise de Componentes Principais (PCA) das amostras analisadas na linhagem parental. 123
FIGURA 42 – *Box plot* das amostras utilizadas nas análises bioinformáticas dos dados do RNA-seq na linhagem parental. 124
FIGURA 43 – Comparação dos perfis de expressão gênica da linhagem parental QM9414 cultivado em celulose, soforose e glicose como fontes de carbono. 124
FIGURA 44 – Gráficos *volcano plot* dos genes diferencialmente expressos up- e down-regulados nas condições estudadas. 125
FIGURA 45 – Diagramas de Venn representando o número de genes diferencialmente expressos na linhagem parental QM9414. 126
FIGURA 46 – Perfil de expressão gênica de *T. reesei* QM9414 durante crescimento na presença de celulose, soforose e glicose como fontes de carbono. 127
FIGURA 47 – Expressão de genes *CAZy* obtidos por RNA-seq. 129
FIGURA 48 – Rede de regulação Gênica (GRN) de genes diferencialmente expressos em *T. reesei*. 140
FIGURA 49 – Análise de enriquecimento de classes funcionais de acordo com o *Gene Ontology* (GO) de genes up-regulados nas condições celulose e soforose em *T. reesei*. 141
FIGURA 50 – Correlação entre RNA-seq e PCR quantitiva em Tempo Real (qPCR)..... 142
FIGURA 51 – Gráficos representando a correlação de Pearson entre as réplicas biológicas de cada condição para a linhagem mutante Δ*xyr1*. 148
FIGURA 52 – Análise de Componentes Principais (PCA) das amostras da linhagem mutante em relação à linhagem parental. 149
FIGURA 53 – *Box plot* das amostras utilizadas nas análises bioinformáticas dos dados do RNA-seq na linhagem mutante..... 150
FIGURA 54 – Visualização da região genômica correspondente ao gene *xyr1* de *T. reesei* utilizando dados do RNA-seq. 150

FIGURA 55 – Comparação do perfil de expressão gênica da linhagem mutante Δ*xyr1* comparado com a linhagem selvagem QM9414 crescidos em celulose, soforose e glicose como fontes de carbono obtidos por RNA-seq. 152
FIGURA 56 – Gráficos *volcano plot* dos genes diferencialmente expressos up- e down-regulados nas condições estudadas (Δ*xyr1*/QM9414). 153
FIGURA 57 – Diagramas de *Venn* representando o número de genes diferencialmente expressos na linhagem mutante Δ*xyr1* comparado com a linhagem parental QM9414. 153
FIGURA 58 – Perfil de expressão de genes de *T. reesei* durante crescimento na presença de celulose, soforose e glicose como fontes de carbono na linhagem mutante Δ*xyr1* comparado com a linhagem parental QM9414 e categorização funcional baseado no *Gene Ontology*. .. 154
FIGURA 59 – Rede regulatória de 1379 genes diferencialmente expressos na linhagem mutante Δ*xyr1* comparado com a linhagem selvagem QM9414 em resposta a diferentes fontes de carbono. 155
FIGURA 60 – Categorização funcional de acordo com o GO de genes diferencialmente expressos na linhagem mutante (Δ*xyr1*) comparado com a linhagem parental (QM9414). .. 158
FIGURA 61 – Genes *CAZy* e expressão das análises por RNA-seq em Δ*xyr1*/QM9414 159
FIGURA 62 – Árvore filogenética contendo 63 genes que codificam transportadores identificados em Δ*xyr1*/QM9414 nas três fontes de carbono estudadas. 164
FIGURA 63 – Rede regulatória de genes regulados diretamente pelo regulador positivo XYR1. 165
FIGURA 64 – Esquema envolvendo a regulação direta de XYR1 em *Trichoderma reesei*. . 166
FIGURA 65 – Correlação entre RNA-seq e PCR quantitiva em Tempo Real (qPCR) para a linhagem mutante (Δ*xyr1*) versus a linhagem parental (QM9414). 167

LISTA DE TABELAS

Tabela 1 – Composição química de biomassas lignocelulósicas. 38
Tabela 2 – Composição do bagaço e palha da cana-de-açúcar..... 43
Tabela 3 – Teores de carboidratos no bagaço de cana-de-açúcar..... 43
Tabela 4 – Efeitos de diferentes pré-tratamentos nos materiais lignocelulósicos. 46
Tabela 5 – Enzimas acessórias fúngicas para a clivagem dos resíduos derivados de hemicelulose. 63
Tabela 6 – Indutibilidade de celulases e hemicelulases em *T. reesei*..... 66
Tabela 7 – Preparo do meio *Mendels-Andreotti* (1L). 76
Tabela 8 – *Primers* utilizados na análise de expressão gênica de 22 genes por qPCR-RT..... 80
Tabela 9 – Comparações realizadas entre as amostras utilizadas no RNA-seq..... 96
Tabela 10 – Transportadores de diferentes espécies utilizados na construção da árvore filogenética. 101
Tabela 11 – *Primers* utilizados na validação de genes diferencialmente expressos (QM9414). 103
Tabela 12 – *Primers* utilizados na validação de genes diferencialmente expressos (Δ*xyr1*). 104
Tabela 13 – *Reads* obtidas por RNA-seq (*Illumina HiSeq^{TM} 2000*) para a linhagem parental QM9414..... 121
Tabela 14 – Top 10 genes diferencialmente expressos em celulose, glicose e soforose..... 128
Tabela 15 – Genes *CAZy* up-regulados em soforose. 131
Tabela 16 – Genes *CAZy* up-regulados em celulose. 132
Tabela 17 – Genes *CAZy* up-regulados em glicose. 133
Tabela 18 – Principais fatores transcricionais induzidos em celulose, soforose e glicose. 135
Tabela 19 – Expressão de fatores transcricionais envolvidos na regulação dos genes de celulases e hemicelulases..... 136
Tabela 20 – Transportadores up-regulados em presence de celulose, soforose e glicose. 138
Tabela 21 – MFS permeases reguladas nas condições celulose e soforose..... 139
Tabela 22 – *Reads* obtidas por RNA-seq (*Illumina HiSeq^{TM} 2000*) para a linhagem Δ*xyr1*.. 147
Tabela 23 – Top 15 genes down-regulados em celulose, soforose e glicose na ausência do regulador XYR1. 156
Tabela 24 – Principais fatores transcricionais modulados na ausência do regulador XYR1. 161
Tabela 25 – Ausência do ativador transcricional Zn2Cys6 (XYR1) altera a expressão dos principais fatores de transcrição envolvidos na regulação de genes de celulases e xilanases. 162

LISTA DE SIGLAS

A – Adenina

AAs – Atividades auxiliares

abf1 – Arabinofuranosidase 1

abf2 – Arabinofuranosidase 2

abf3 – Arabinofuranosidase 3

ABFI, ABFII e ABFIII – Arabinofuranosidases

ACE1 – Repressor de celulase 1

ACE2 – Ativador de celulase 2

ACE3 – Ativador de celulase 3

Ácido cítrico – Ácido 2-hidroxi-1,2,3-propanotricarboxílico (Fórmula molecular $C_6H_8O_7$)

Act – Actina

AESI – Acetil esterase I

α – Alfa

a-AF – Alfa-L-arabinofuranosidase

agl1 – Alfa-galactosidase 1

agl2 – Alfa-galactosidase 2

AGLI, AGLII e AGLIII – α-galactosidases

~ – Aproximadamente

atm – Atmosfera (refere-se a uma unidade de pressão)

AXEI e AXEII – Acetilxilanas esterases

β – Beta

BAM – Binary SAM

bgl1/cel3a – β-glicosidase

bgl2/cel1a – β-glicosidase 2

BglR – Regulador de β-glicosidase

BGs – β-glicosidases

BXLI – β-xilosidase I

bxl1: β-xilosidase 1

BWT – *Burrows-Wheeler Transform*

C – Carbono

^oC – Grau Celsius

$CaCl_2.2H_2O$ – Cloreto de cálcio di-hidratado

CAZy – Carbohydrate-Active enZYmes

CEs – Carboidratos Esterases

CBHs – Celobiohidrolases

CBHI – Celobiohidrolases I

CBHII – Celobiohidrolases II

CBM – Módulo de ligação a carboidratos

CDH – Celobiose desidrogenase

Cel – Celulose

cel5b – Endoglucanase

5'-UTR – Região 5' não traduzida (*Untranslated region*)

CIPII – Glucuronoil esterase II

CRE1 – Repressão catabólica por carbono

CO_2 – Dióxido de carbono

$CoCl_2.2H_2O$ – Cloreto de cobalto di-hidratado

2G – Etanol de segunda geração

DEPC – Dietilpirocarbonato

ddNTP – Trifosfato de didesoxirribonucleosídeo

dNTP – Trifosfato de desoxirribonucleosídeo

EGs – Endoglucanases

egl1/cel7b – Endoglucanase 1

egl2/cel5a – Endoglucanase 2

egl3/cel12a – Endoglucanase 3

egl5/cel45a: Endoglucanase 5

ER – Retículo endoplasmático

$FeSO_4.7H_2O$ – Sulfato de ferro II heptahidratado

FPKM – Fragmentos por Kilobase por Milhão de Reads Mapeadas

GB – Gigabase

GH – Glicosil Hidrolase

GFF – General Feature File

g/L – Grama por litro

Gli – Glicose

GLRI – α-glucuronidase

glr1 – α-glucuronidase 1

GO – *Gene Ontology*

GRN – Rede Regulatória Gênica

GTs – Glicosil Transferases

H – Hidrogênio

H_2SO_4 – Ácido sulfúrico

$H3K9^{me3}$ – Histona H3 trimetilada na lisina 9 – metilação

HFB1 – Hidrofobina 1

ID – Número relacionado à proteína

IGV– *Institute* o *Integrative Genomics Viewer*

JGI – *Joint Genome Institute* (*United States Department of Energy*)

KH_2PO_4 – Fosfato monobásico de potássio

Kpb – Quilo pares de bases

L – Litro

LA – Acelerador linear

LAE1 – Proteína metiltransferase

> - Maior

< - Menor

≥ – Maior igual

≤ – Menor igual

M – Molar

m^3 – Metros cúbicos

MANI – Mananase I

MEME – *Multiple Expectation Maximisation for Motif Elicitation*

µg – Micrograma

$MgCl_2$ – Cloreto de magnésio

µm – Micrômetro

mg/L – Miligrama por litro

$MgSO_4.7H_2O$ – Sulfato de magnésio heptahidratado

mL – Mililitros

mM – Milimolar

$MnSO_4.H_2O$ – Sulfato de Manganês Monohidratado

min – Minuto

n – Número

NaCl – Cloreto de sódio

$Na_2HPO_4.2H_2O$ – Fosfato de sódio bibásico di-hidratado

NCBI – *National Center for Biotechnology Information*

$(NH_4)_2SO_4$ – Sulfato de amônio

nM – Nanomolar

NTG – Nitrosoguanidina

O – Oxigênio

ORFs – A seqüência de DNA de um gene é uma fase de leitura aberta (ORF, do inglês *open reading frame*)

p – Posição para (p) é 1-4 em química orgânica para dar nome aos aromáticos que contém apenas um anel aromático

pb – Pares de base

PCA – Análise de Componentes Principais

PCR – Reação em cadeia da polimerase

pH – Potencial hidrogeniônico

PLs – Polissacarídeos Liases

PMOs – Polissacarídeos monooxigenase

p/v – Peso por volume

% – Porcentagem

PWM – *Position Weight Matrixes*

PWM_{XYR1} – Sítio XYR1 analisado

qPCR-RT – PCR em Tempo Real quantitativa

q.s.p – Quantidade suficiente para

R$ - Real

Rd1 SP – *Read 1 Sequence Primer*

Rd2 SP – Read 2 Sequence Primer

RNAm – RNA mensageiro

RNA-Seq – Sequenciamento de RNA

rpm – Rotações por minuto

SAM – *Sequence Alignment/Mapping*

Sof – Soforose

SSH – Secure Shell Client

SV – Vesículas secretoras

SWOI – Solenina

T – Timina

U$ – Dólar

UV – Radiação ultravioleta

v/v – volume por volume

XAE – Ativador de xilanase

xynI – Xilanase 1

xynII – Xilanase 2

xyn3 – Xilanase 3

xynIV– Xilanase 4

xynV – Xilanase 5

XYR1 – Regulador de xilanase 1 (principal fator transcrional em *T. reesei*)

$ZnSO_4.7H_2O$ – Sulfato de zinco heptahidratado

SUMÁRIO

1 INTRODUÇÃO 31

2 FUNDAMENTAÇÃO TEÓRICA 37

2.1 Características estruturais da biomassa lignocelulósica **37**

2.2 Produção de etanol de segunda geração (2G) **44**

2.2.1 Etanol 2G: uma realidade no Brasil **48**

2.3 *Trichoderma reesei* **54**

2.3.1 Gênero, taxonomia, morfologia e mutações **54**

2.3.2 *CAZy* **57**

2.3.3 O processo de hidrólise enzimática: ação sinérgica dos sistemas celulolítico e hemicelulolítico de *T. reesei* **58**

2.3.4 Regulação da expressão de genes de celulases e hemicelulases **64**

3 OBJETIVOS 72

3.1 Objetivo geral **72**

3.2 Objetivos específicos **72**

4 MATERIAIS E MÉTODOS 74

4.1 Delineamento experimental **74**

4.2 Linhagens fúngicas **75**

4.3 Condições de crescimento **75**

4.3.1 Meio MEX **75**

4.3.2 Meio *Mendels-Andreotti* **75**

4.4 Extração de RNA e verificação de sua integridade **77**

4.5 Remoção do ácido desoxirribonucléico e síntese da primeira fita de cDNA **77**

4.6 Análises de transcritos por PCR quantitativo (qPCR-RT) e análises estatísticas **78**

4.6.1 *Primers* utilizados nesta análise por PCR em Tempo Real (qPCR-TR) **79**

4.7 Análise *in silico* de putativos sítios de ligação para XYR1 em regiões promotoras dos genes celulolíticos e xilanolíticos **80**

4.7.1 Descoberta do suposto motivo de XYR1 em *T. reesei* **81**

4.7.1.1 Suposto motivo (*de novo*) de XYR1 ..... **82**

4.7.1.2 Análise do elemento regulatório cis de XYR1 no genoma de *T. reesei* **82**

4.8 Sequenciamento de alto desempenho do transcriptoma (RNA-Seq) **83**

4.8.1 Preparo das amostras de RNAs totais para realização do sequenciamento **83**

4.8.2 Sequenciamento utilizando plataforma *Illumina Hiseq^{TM} 2000* **84**

4.8.2.1 Preparo das bibliotecas de cDNAs **85**

4.8.2.2 Geração de *clusters* **87**

4.8.2.3 Sequenciamento *Paired end* **88**

4.9 Bioinformática **91**

4.9.1 Alinhamento das *reads* no genoma de referência **93**

4.9.2 R e *Bioconductor* **93**

4.9.2.1 Sumarização: *Reads Counts* (Tabelas) **94**

4.9.2.1.1 *rtracklayer* **95**

4.9.2.1.2 *GenomicRanges* **95**

4.9.2.2 Pacote DESeq **96**

4.9.2.2.1 Normalização **96**

4.9.2.2.2 Análise de Expressão diferencial **97**

4.9.3 Anotação Funcional **97**

4.9.4 Redes regulatórias **98**

4.10 Outras ferramentas bioinformáticas utilizadas neste estudo **98**

4.10.1 *Box plot* **98**

4.10.2 *Volcano plot* **99**

4.10.3 Diagramas de *Venn* **99**

4.10.4 Construção de *heatmaps* no estudo de regulons específicos de cada condição **99**

4.10.5 Categorização funcional **99**

4.10.6 Genes *Carbohydrate active enZymes* (*CAZy*) **100**

4.10.7 Análise Filogenética **100**

4.11 Validação das análises transcricionais **101**

4.11.1 Análise da expressão gênica por PCR em Tempo Real (qPCR-TR) **101**

4.11.1.1 *Primers* utilizados na etapa de validação **102**

5 CAPÍTULO I **106**

5.1 RESULTADOS **106**

5.1.1 Análises da influência de fontes de carbono na regulação da expressão gênica de genes celulolíticos e xilanolíticos por XYR1 ..... **106**

5.1.2 Análises *in silico* de putativos sítios de ligação para XYR1 em regiões promotoras dos 22 genes celulolíticos e xilanolíticos **108**

5.1.3 Determinando a arquitetura do elemento regulatório cis para XYR1 ..... **110**

5.1.4 Inspeção de promotores de celulases por sítio de ligação de XYR1 ..... **113**

5.1.5 Identificação de potenciais genes alvos de XYR1 ..... **115**

5.2 DISCUSSÕES **116**

6 CAPÍTULO II **121**

6.1 RESULTADOS **121**

6.1.1 Perfil global da expressão gênica de *T. reesei* cultivado em diferentes fontes de carbono **121**

6.1.2 *CAZYome* **129**

6.1.3 Fatores transcricionais **134**

6.1.4 Transportadores **136**

6.1.5 Decifrando a rede regulatória de *T. reesei* em resposta a celulose/soforose **139**

6.1.6 Análises de PCR quantitativa em Tempo Real (qPCR-TR) **141**

6.2 DISCUSSÕES 142

7 CAPÍTULO III 147

7.1 RESULTADOS 147

7.1.1 Perfil de expressão gênica da linhagem mutante (*Δxyr1*) comparado à linhagem parental (QM9414) em diferentes fontes de carbono **147**

7.1.1.1 Sequenciamento e alinhamento **147**

7.1.1.2 Genes Diferencialmente Expressos **151**

7.1.1.3 Expressão de genes *Carbohydrate Active enZymes* (*CAZy*) identificados em *T. reesei Δxyr1*/QM9414 ..... **159**

7.1.1.4 Expressão de genes que codificam fatores de transcrição identificados em *T. reesei Δxyr1*/QM9414 ..... **160**

7.1.1.5 Expressão de transportadores identificados em *T. reesei Δxyr1*/QM9414 ..... **162**

7.1.1.6 Genes regulados diretamente pelo fator transcricional XYR1 ..... **165**

7.1.1.7 Análises de PCR quantitativa em Tempo Real (qPCR-TR) **166**

7.1.2 DISCUSSÕES 167

8 CONCLUSÕES 176

REFERÊNCIAS 179

CD ANEXO 205

Arquivo 1: Tabelas 1.1 a 1.3 205

Tabela 1.1 – Fatores Transcricionais up-regulados na condição de crescimento celulose **205**

Tabela 1.2 – Fatores Transcricionais up-regulados na condição de crescimento soforose **205**

Tabela 1.3 – Fatores Transcricionais up-regulados na condição de crescimento glicose **205**

Arquivo 2: Tabela 2 – 9129 genes analisados por RNA-seq em *T. reesei* QM9414 205

Arquivo 3: Tabela 3 - 9129 genes analisados por RNA-seq em *T. reesei Δxyr1* comparado ao *T. reesei* QM9414 205

Arquivo 4: Tabela 4 – Predição de todo o genoma de sítios de ligação de XYR1 em *T. reesei* 205

Arquivo 5: Tabela 5 – Predição de sítios de ligação XYR1 em genes down-regulados induzidos por celulose no mutante *Δxyr1* 205

Arquivo 6: Tabela 6 – Predição de sítios de ligação XYR1 em genes up-regulados induzidos por celulose no mutante *Δxyr1* 205

Arquivo 7: Tabela 7 – Predição de sítios de ligação XYR1 em genes down-regulados induzidos por soforose no mutante *Δxyr1* 205

Arquivo 8: Tabela 8 – Predição de sítios de ligação XYR1 em genes up-regulados induzidos por soforose no mutante *Δxyr1* 205

Arquivo 9: Tabelas 9.1 a 9.3 205

Tabela 9.1. Genes do regulon de Celulose **205**

Tabela 9.2. Genes do regulon de Glicose **205**

Tabela 9.3. Genes do regulon de Soforose **205**

Arquivo 10: Tabela 10 – Genes diferencialmente expressos em celulose e soforose. **205**

Arquivo 11: Tabela 11 – Comparação dos níveis de expressão de genes avaliados por RNA-seq e qPCR. **205**

Arquivo 12: Tabelas 12.1 a 12.6 205

Tabela 12.1 – Genes regulon glicose (Down-regulados em *Δxyr1*/QM9414) **205**

Tabela 12.2 – Genes regulon soforose (Down-regulados em *Δxyr1*/QM9414) **205**

Tabela 12.3 – Genes regulon celulose (Up-regulados em *Δxyr1*/QM9414)..... **205**

Tabela 12.4 – Genes regulon soforose (Up-regulados em *Δxyr1*/QM9414) **205**

Tabela 12.5 – Genes regulon celulose (Down-regulados em *Δxyr1*/QM9414)..... **205**

Tabela 12.6 – Genes regulon glicose (Up-regulados em *Δxyr1*/QM9414) **205**

Arquivo 13: Tabelas 13.1 a 13.10 206

Tabela 13.1 – Rede regulatória: genes up-regulados em Celulose (*Δxyr1*/QM9414)..... **206**

Tabela 13.2 – Rede regulatória: genes down-regulados em Celulose (*Δxyr1*/QM9414) **206**

Tabela 13.3 – Rede regulatória: genes up-regulados em Soforose (*Δxyr1*/QM9414)..... **206**

Tabela 13.4 – Rede regulatória: genes down-regulados em Soforose (*Δxyr1*/QM9414)..... **206**

Tabela 13.5 – Rede regulatória: genes up-regulados em Glicose (*Δxyr1*/QM9414) **206**

Tabela 13.6 – Rede regulatória: genes down-regulados em Glicose (*Δxyr1*/QM9414) **206**

Tabela 13.7 – Rede regulatória: Celulose versus Soforose (*Δxyr1*/QM9414) **206**

Tabela 13.8 – Rede regulatória: Soforose versus Glicose (*Δxyr1*/QM9414) **206**

Tabela 13.9 – Rede regulatória: Celulose versus Glicose (*Δxyr1*/QM9414) **206**

Tabela 13.10 – Rede regulatória: todas as condições (*Δxyr1*/QM9414) **206**

Arquivo 14: Tabelas 14.1 a 14.3 ..... **206**

Tabela 14.1 – Genes CAZy up- ou down-reguladas em celulose (*Δxyr1*/QM9414) **206**

Tabela 14.2 – Genes CAZy up- ou down-reguladas em soforose (*Δxyr1*/QM9414) **206**

Tabela 14.3 – Genes CAZy up- ou down-reguladas em glicose (*Δxyr1*/QM9414) **206**

Arquivo 15: Tabela 15 – Transportadores up- e down-regulados em presença de celulose, soforose e glicose (*Δxyr1*/QM9414) **206**

Arquivo 16: Tabela 16 – Comparação dos níveis de expressão de genes testados por RNA-seq e qPCR-RT em Δ*xyr1*/QM9414 ..... **206**

Arquivo 17: Tabela 17.1 a 17.3 ..... **206**

Tabela 17.1 – GHs específicas das linhagens QM9414 e Δ*xyr1* **206**

Tabela 17.2 – Transportadores específicos das linhagens QM9414 e Δ*xyr1* **206**

Tabela 17.3 – Fatores específicos das linhagens QM9414 e Δ*xyr1* **206**

PARTICIPAÇÕES EM CONGRESSOS **208**

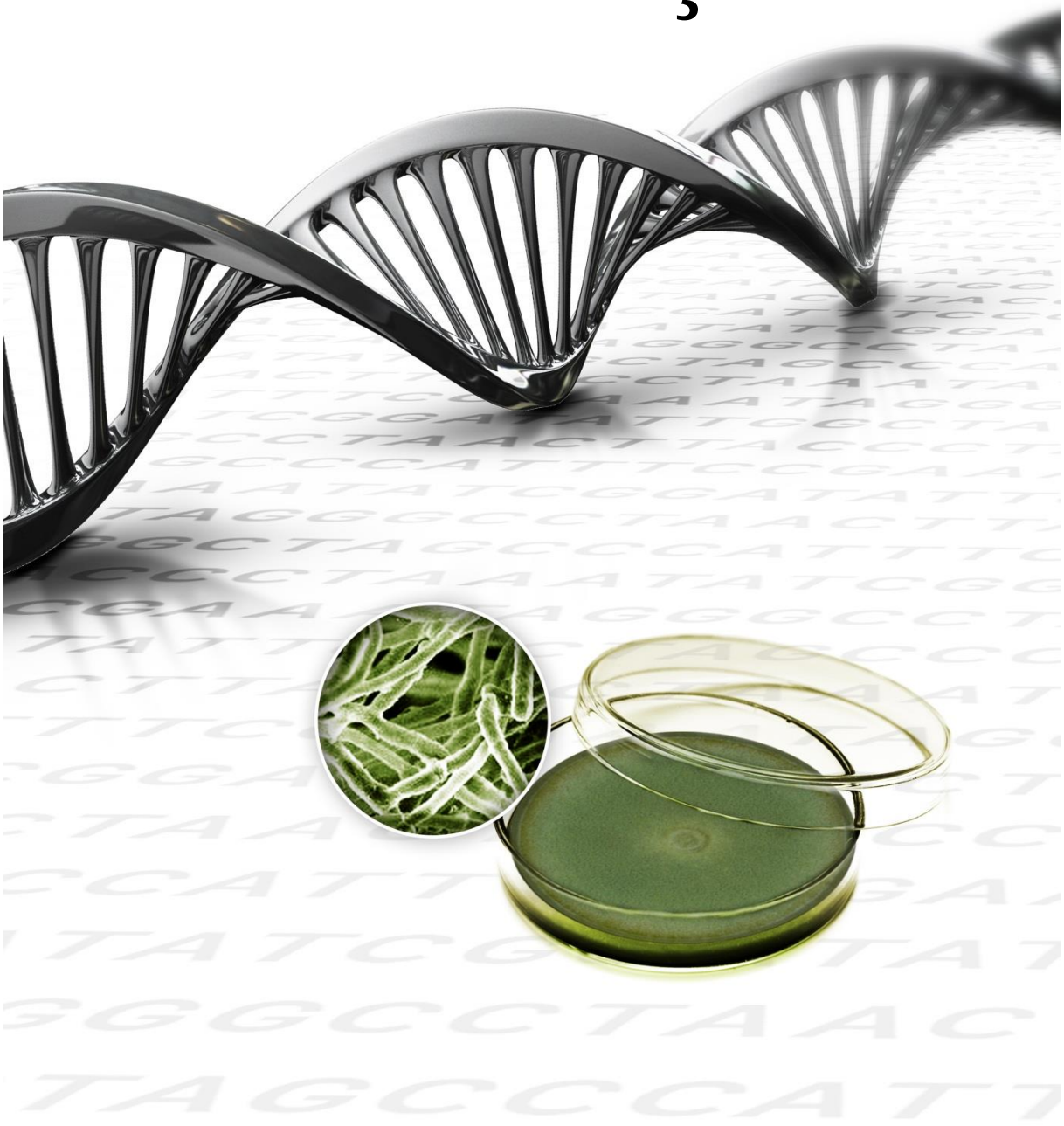
Resumos publicados em anais de congressos nacionais **208**

Resumos publicados em anais de congressos internacionais **208**

PARTICIPAÇÕES EM CURSOS **211**

ANEXOS DE PUBLICAÇÕES **212**

INTRODUÇÃO



1 INTRODUÇÃO

Impulsionada pelo aumento da industrialização e da população, a procura global por energia e matérias-primas está em constante crescimento. Uma vez que as principais fontes mundiais de energia e produtos químicos têm origem nos combustíveis fósseis, este crescimento levanta importantes questões em nível ambiental, econômico e social (PEDRO, 2013). Petróleo, gás natural e seus derivados representam 55% do consumo mundial de energia. São esses combustíveis que permitem a existência dos meios de transporte rápidos e eficientes que temos hoje, bem como boa parte das atividades industriais, porém suas reservas são finitas (BNDES E CGEE, 2008). A diminuição da participação de fontes fósseis na matriz energética é uma das medidas que se faz necessária para reduzir a emissão de gases do efeito estufa e a consequente redução do aquecimento global. As mudanças climáticas mundiais devido às emissões de gases do efeito estufa são conhecidas e trazem consigo consequências indesejáveis ao meio ambiente, causando transtornos à sociedade e até colocando em risco a existência de vida na Terra (SANTOS et al., 2013).

O desenvolvimento de fontes renováveis de energia é um tema cada vez mais explorado na agenda global. O melhoramento da qualidade de vida e o crescimento populacional exige uma solução para os problemas que se enfrenta hoje e que poderão se agravar no futuro, como a poluição ambiental, a carência de alimentos e a de energia. Com a escassez das reservas de petróleo e a iminência das mudanças climáticas, as energias renováveis surgem como uma importante alternativa ao futuro do planeta para suprir as necessidades energéticas da atual e das futuras gerações (DRABER, 2013). É necessário encontrar substitutos para esses combustíveis. Nada mais racional do que produzi-los com base em matéria orgânica renovável (biomassa), da qual, no passado distante, os combustíveis fósseis foram produzidos pela natureza (SILVA, 2010). Assim, com o esgotamento dos recursos petroquímicos não renováveis e um sentimento ambiental cada vez mais forte, fundamentado pelas drásticas mudanças climáticas, os biocombustíveis surgem como uma alternativa eficaz (LIU et al., 2013).

Biocombustíveis são derivados de biomassa renovável que podem substituir, parcial ou totalmente, combustíveis derivados de petróleo e gás natural em motores a combustão ou em outro tipo de geração de energia. Os dois principais biocombustíveis líquidos usados no Brasil são o etanol (bioetanol) extraído de cana-de-açúcar e, em escala crescente, o biodiesel, que é produzido a partir de óleos vegetais ou de gorduras animais e adicionado ao diesel de

petróleo em proporções variáveis. Cerca de 45% da energia e 18% dos combustíveis consumidos no Brasil já são renováveis. No resto do mundo, 86% da energia vêm de fontes energéticas não renováveis. Pioneiro mundial no uso de biocombustíveis, o Brasil alcançou uma posição almejada por muitos países que buscam fontes renováveis de energia como alternativas estratégicas ao petróleo (ANP, 2012).

O Brasil vem se destacando com a produção de bioetanol a partir da cana-de-açúcar, também conhecido como etanol de primeira geração. O etanol trata-se de uma fonte de energia natural, limpa, renovável e sustentável e que pode ser produzido a partir de várias matérias-primas além da cana-de-açúcar, como: milho, trigo e beterraba (AZEVEDO et al., 2012). Dentre as diversas fontes vegetais, a cana-de-açúcar é a que tem oferecido maiores vantagens energéticas e econômicas. O Brasil produz apenas etanol a partir de cana-de-açúcar (ANP, 2012), e é o segundo maior produtor de etanol no mundo, atrás apenas dos Estados Unidos. Porém, a produtividade de álcool de cana-de-açúcar do país é a maior do mundo, e os EUA são os maiores produtores mundiais de álcool de milho. Contudo, o biocombustível brasileiro leva vantagem devido à produtividade da cana-de-açúcar, isto é, em uma mesma área a cana-de-açúcar brasileira produz mais etanol do que o milho (GRUBISICH, 2011).

Esse combustível, com potencial energético consideravelmente superior ao do milho ganhou destaque no Brasil na década de 1970, diante da crise energética mundial que levou o país a criar o Programa Nacional do Álcool – Proálcool, a fim de reduzir a dependência brasileira em relação ao petróleo. A ação do programa justifica o Brasil ser o segundo maior produtor mundial de etanol (VIEIRA et al., 2007).

De acordo com novaCana (2014a), durante a produção do etanol de primeira geração são obtidos dois tipos de etanol: anidro (sem água) e o hidratado. O etanol anidro é misturado à gasolina para baratear o combustível, aumentar sua octanagem e reduzir a emissão de poluentes. O Brasil atualmente utiliza a mistura na proporção de 27% (AMATO; MATOSO, 2015). Mais de 40 países, como Estados Unidos, Canadá, Paraguai e China também utilizam essa mistura, porém em proporções que costumam ser de 5% ou 10%. O etanol anidro ainda é utilizado na fabricação de tintas, vernizes, solventes, bebidas destiladas, entre outros produtos. Já o etanol hidratado é utilizado como combustível somente no Brasil, desde o fim da década de 70. Esse uso torna-se viável economicamente apenas no país, algo possível graças a incentivos governamentais com o programa Proálcool, 1970, e, de 1975. Além de combustível, o etanol hidratado também está presente em cosméticos, produtos de limpeza,

antissépticos, vinho, cerveja e outros líquidos, em graduações alcoólicas que variam de produto a produto.

Com o etanol, o Brasil é o pioneiro no mundo na utilização em larga escala de biocombustível renovável como energia veicular. Com o início da produção de veículos, a etanol, criou-se no país uma nova e extensa cadeia econômica, da produção do etanol na agroindústria canavieira à distribuição e utilização em larga escala do combustível diretamente nos motores de ciclo Otto (etanol hidratado) e para adição de até 27% (AMATO; MATOSO, 2015) na gasolina consumida no país (etanol anidro). Em 2003, com desenvolvimentos tecnológicos próprios, foram lançados no Brasil os veículos *flex fuel*, que podem consumir indistintamente, ou ao mesmo tempo, etanol e gasolina, em qualquer proporção (CNI, 2012). Os automóveis que saem das fábricas brasileiras já são produzidos com a tecnologia *flex fuel* e espera-se que, até o ano de 2022, um terço dos 30 milhões de automóveis brasileiros utilizará essa tecnologia (LORA; ANDRADE, 2009), justificando, assim, a necessidade do aumento da capacidade brasileira de produção de bioetanol de primeira geração.

Segundo a CNI - Confederação Nacional da Indústria (2012), a viabilização técnica e econômica do etanol combustível consolidou essa nova e importante cadeia econômica no país. Os efeitos tecnológicos, econômicos e sociais dessa atividade são intensos, interiorizando o desenvolvimento com investimentos, produção, geração de empregos, renda, consumo e qualidade de vida das regiões produtoras do combustível renovável, bem como ainda movimentando a cadeia automotiva propriamente dita, com a produção dos veículos *flex fuel*. Ao lado desses benefícios socioeconômicos, estão os ganhos ambientais da produção e do consumo do etanol, com a redução das emissões de CO_2 na atmosfera. O balanço ambiental do etanol é positivo, considerando-se que suas emissões de CO_2, durante o consumo, são compensadas pelo cultivo de cana-de-açúcar para a produção do combustível.

A produção mundial de etanol é alta, mas devido a essa série de aumento no consumo deste combustível, torna-se muito importante aumentar ainda mais a produção. A busca por novas fontes energéticas fez a biomassa ganhar espaço, sendo utilizada diretamente para cogeração de energia na forma de queima como também na produção de novos combustíveis, como o etanol celulósico, ou etanol lignocelulósico, ou, ainda, etanol de segunda geração (2G), (ALVES; MACRI, 2013). No entanto, muito ainda precisa ser feito para aprimorar as cadeias de produção de etanol, visando sempre ao aumento de produção e a minimização de impactos negativos ao meio ambiente. Diante disso, surge a possibilidade de

aproveitamento de resíduos agrícolas, como uma alternativa promissora que alia aumento de produção e diminuição da degradação ambiental (SILVA, 2012a).

Atualmente, a pesquisa de etanol lignocelulósico está em constante expansão. Enquanto a produção de etanol de primeira geração teve um bom desempenho, o processo de produção de etanol de segunda geração, a partir de materiais lignocelulósicos, ainda está à procura de viabilidade econômica. Além disso, as matérias-primas utilizadas para a produção do etanol de primeira geração são também, geralmente, fontes de alimentos importantes, e, futuramente, poderá ser questionado quanto à segurança alimentar da população, e suas disponibilidades não serão suficientes para cobrir o total necessário para a produção do bioetanol em demanda mundial e, ao mesmo tempo, a alimentação (MARYANA et al., 2014). Assim, matérias-primas lignocelulósicas não competindo com a produção de alimentos, podem fornecer benefícios ambientais e econômicos e se tornar uma importante estratégia para a produção de biocombustíveis (VIIKARI et al., 2012).

Materiais lignocelulósicos, como o bagaço ou a palha de cana-de-açúcar são compostos principalmente de três componentes: celulose, hemicelulose e lignina. O processo de conversão de biomassa lignocelulósica a etanol consiste, basicamente, em quatro etapas: pré-tratamento, a hidrólise enzimática, fermentação e destilação (ENSINAS et al., 2013). O pré-tratamento físico-químico é etapa essencial para quebrar a estrutura do material lignocelulósico, a fim de aumentar a acessibilidade dos polímeros de celulose e hemicelulose a enzimas celulolíticas. Esse grupo de enzimas hidrolisa, os polissacarídeos em monossacarídeos (glicose ou xilose) que podem ser, subsequentemente, convertidos a etanol por fermentação. O processo de produção de etanol lignocelulósico tem sido uma área amplamente pesquisada, a fim de compreender os gargalos que existem em cada uma das etapas do processo, como: 1) matéria-prima para obtenção do etanol (as características e disponibilidade de biomassa que determina maior ou menor sucesso); 2) técnicas para degradar a parede celular da biomassa (eleger a melhor opção pode ser uma tarefa difícil) e 3) enzimas eficientes para obtenção de monossacarídeos constituintes na matéria-prima (o custo e a regularidade em escala industrial são fundamentais) (KOPPRAM et al., 2014).

A conversão dos componentes celulósicos em açúcares fermentáveis é, no entanto, um grande desafio tecnológico e econômico na produção de biocombustíveis. Especialmente, a hidrólise enzimática que ainda constitui um fator de custo importante. Sendo que este custo é dependente da eficiência, rendimento e os custos do pré-tratamento, a ação sinérgica de

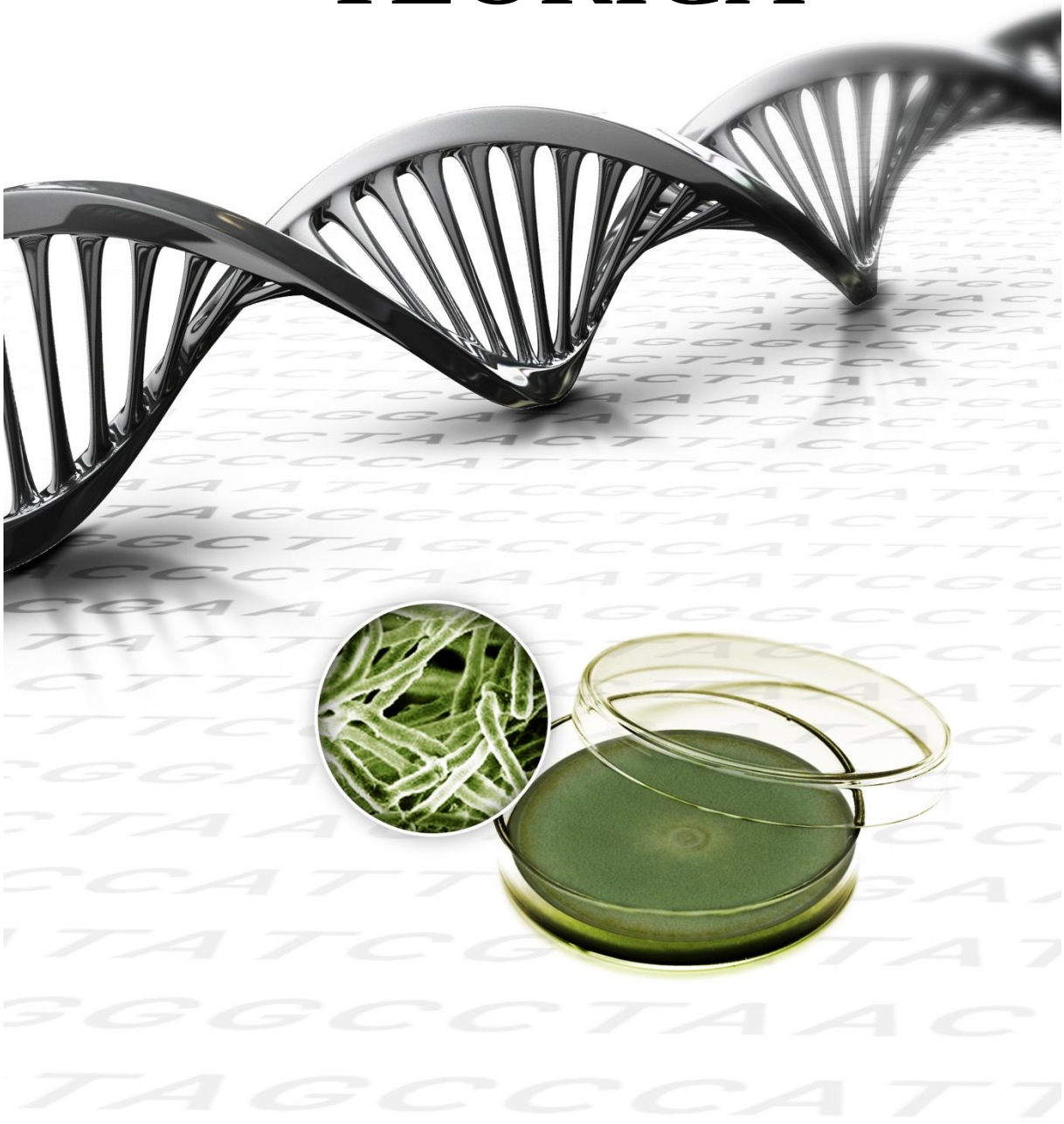
celulases e enzimas acessórias, bem como a quantidade necessária de enzimas adicionadas externamente (VIIKARI et al., 2012).

A escolha do fungo filamentoso *Trichoderma reesei* para o desenvolvimento desta tese foi pelo fato de ser um dos organismos mais amplamente investigados para a produção de um conjunto completo de enzimas celulolíticas para a degradação da biomassa. Destacando, da mesma forma, o seu potencial para desenvolver mutantes capazes de produzir enzimas eficientes para utilização em escala industrial. Além disso, essas enzimas poderão ser utilizadas na obtenção de coquetéis enzimáticos[1], uma combinação de enzimas diferentes, para transformar material vegetal em açúcares que tenham valor para a indústria de biocombustíveis. Se enzimas podem produzir açúcares de forma mais eficiente, isso significa que o custo dos coquetéis, que é um dos principais fatores de custo do processo de conversão de biomassa em biocombustível, ficará mais baixo. Dessa forma, o estudo do genoma/transcriptoma deste fungo poderá ser utilizado para ajudar a resolver um dos principais gargalos existentes na produção de etanol de segunda geração.

O trabalho apresentado encontra-se estruturado em: 1 Introdução; no tópico 2, a Fundamentação Teórica desse estudo, destacando-se: a) composição do material lignocelulósico da cana-de-açúcar, por se tratar da principal fonte de matéria-prima na produção de etanol de primeira geração no Brasil; b) bioetanol de 1ª e de 2ª geração destacando as suas diferenças de produção; c) usinas de etanol e açúcares existentes no Brasil; d) realidade do etanol 2G no Brasil, apresentando as unidades industriais de produção de etanol de 2ª geração, mostrando a sua localização e capacidade de produção; e) enzimas existentes no mercado e f) características do fungo *T. reesei*. No tópico 3, destacam-se os objetivos da presente tese. No tópico 4, têm-se as metodologias utilizadas no desenvolvimento desta tese. No tópico 5 (Capítulo I), os resultados referentes à expressão de 22 enzimas conhecidas na literatura, tanto na linhagem parental *T. reesei* QM9414, como na linhagem mutante *T. reesei Δxyr1* (PCR em Tempo Real), bem como análises de motivos de XYR1. No tópico 6 (Capítulo II), os resultados e discussões obtidos para a linhagem parental de *T. reesei* QM9414 (Análises RNA-seq). No tópico 7 (Capítulo III), há os resultados e discussões obtidos para a linhagem mutante de *T. reesei Δxyr1* comparado à linhagem parental QM9414 (Análises RNA-seq). Como término, no tópico 8, a enunciação das principais conclusões resultantes do presente estudo.

[1] É um conjunto de proteínas que reconhecem especificamente regiões de determinadas moléculas e, na presença de água, as quebram (clivam) em moléculas menores (GRANBIO, 2014).

FUNDAMENTAÇÃO TEÓRICA



2 FUNDAMENTAÇÃO TEÓRICA

2.1 Características estruturais da biomassa lignocelulósica

A biomassa lignocelulósica constitui a maior fonte de carboidratos naturais do mundo (SANTOS et al., 2012). É composta por três componentes principais (Figura 1): celulose (40-50%); hemicelulose (20-40%); e lignina (20-30%). Além de constituir em menores proporções: proteínas, lipídeos, pectina, açúcares solúveis e minerais (QUIROZ-CASTAÑEDA; FOLCH-MALLOL, 2013). A composição da biomassa depende: das espécies de plantas, do tecido específico e do estágio de crescimento. A Tabela 1 resume a composição lignocelulósica encontrados nas fontes mais comuns de biomassa.

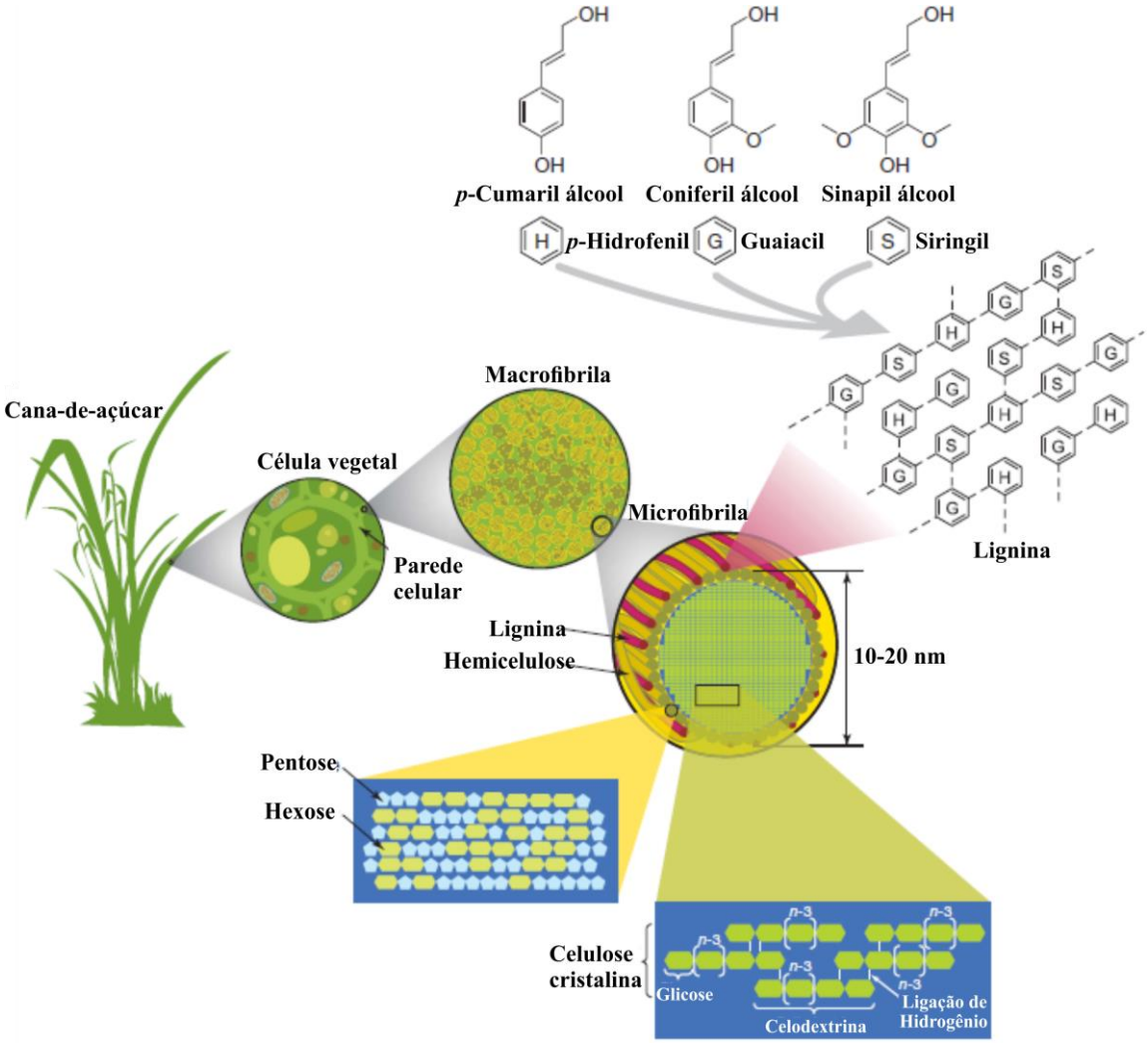


FIGURA 1 – Estrutura biomassa lignocelulósica.
Fonte: Adaptado de (RUBIN, 2008).

Tabela 1 – Composição química de biomassas lignocelulósicas.

Biomassa Lignocelulósica	% Celulose	% Hemicelulose	% Lignina
Palha de cana	40-44	30-32	22-25
Bagaço de cana	32-48	19-24	23-32
Madeira dura	43-47	25-35	16-24
Madeira mole	40-44	25-29	25-31
Talo de milho	35	25	35
Espiga de milho	45	35	15
Algodão	95	2	0,3
Palha de trigo	30	50	15
Sisal	73,1	14,2	11
Palha de arroz	43,3	26,4	16,3
Forragem de milho	38-40	28	7-21
Fibra de coco	36-43	0,15-0,25	41-45
Fibra de bananeira	60-65	6-8	5-10
Palha de cevada	31-45	27-38	14-19

Fonte: (SANTOS et al., 2012).

Celulose, o polissacarídeo estrutural mais abundante nas paredes celulares, é caracterizado por um polímero linear, constituído por repetidas unidades de celobiose[2] (Figuras 2A e 2B) (GUO et al., 2012), ou também, pode ser brevemente descrita como um polímero linear de alta massa molar, constituído exclusivamente de unidades de anidroglicopiranose (Figura 2B). Essas unidades são unidas por ligações glicosídicas β-(1→4), e por isso a unidade anidroglicopiranose também é denominada de β-D-anidroglicopiranose (EYLEY; THIELEMANS, 2014). Durante a biossíntese da celulose, as cadeias glicosídicas formadas interagem umas com as outras por pontes de hidrogênio e forças de *van der Waals*, cristalizando-se e formando redes insolúveis, também denominadas regiões cristalinas (Figura 2C). As cadeias individuais de celulose são orientadas paralelamente com a extremidade redutora e a extremidade não redutora em lados opostos (Figura 2B). Essa subestrutura resulta em microfibrilas que contêm cadeias organizadas em zonas altamente cristalizadas, bem como regiões amorfas (Figura 2C), onde as cadeias não são organizadas paralelamente (CHAMI KHAZRAJI; ROBERT, 2013).

A estrutura da celulose pode ser classificada em três níveis organizacionais. O primeiro é definido pela sequência de resíduos β-D-glicopiranosídicos unidos por ligações

[2] Celobiose é formada pela união de duas moléculas de glicose (ERIKSEN et al., 2013).

covalentes, formando o homopolímero de anidroglicopiranose com ligações β-D(1→4) glicosídicas, de fórmula geral $(C_6H_{10}O_5)_n$ [n = 10000 a 15000, sendo dependente da fonte constituída da celulose (MOON et al., 2011)]. O segundo nível descreve a conformação molecular, isto é, a organização espacial das unidades repetitivas, e é caracterizado pelas distâncias das ligações e respectivos ângulos e pelas ligações de hidrogênio intramoleculares. O terceiro nível define a associação das moléculas, formando agregados com uma determinada estrutura cristalina. Esses agregados conferem elevada resistência à tensão, tornando a celulose insolúvel em água e em um grande número de outros solventes, o que explica a sua resistência à degradação microbiana. As microfibrilas formadas por cadeias de celulose, por sua vez, estão envolvidas por hemicelulose e lignina compondo uma matriz que fornece rigidez e resistência à parede celular dos vegetais lignificados (Figura 1) (SANTOS et al., 2012).

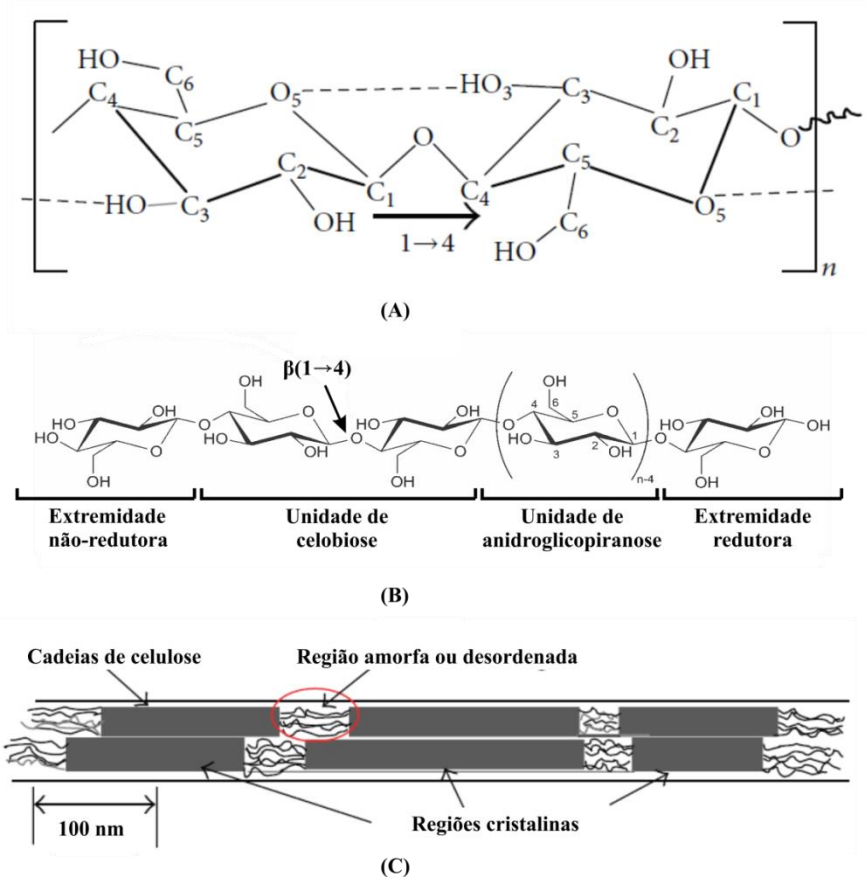


FIGURA 2 – Representações da estrutura da celulose.
(A) unidade de repetição de uma simples cadeia de celulose, mostrando a direcionalidade da ligação β-(1→4) e ligações de hidrogênio intracadeias (linhas tracejadas). (B) cadeia de celulose mostrando as unidades de celobiose e anidroglicopiranose juntamente com a numeração dos átomos, a ligação glicosídica e ambas as extremidades redutora e não redutora do polímero. (C) regiões cristalinas e amorfas.
Fonte: Adaptado de (CHAMI KHAZRAJI; ROBERT, 2013) e (EYLEY; THIELEMANS, 2014).

Hemicelulose, o segundo componente mais abundante da biomassa lignocelulósica, são polímeros heterogêneos de pentoses (incluindo xilose e arabinose), hexoses (principalmente manose), ácidos urônicos, grupos acetila e ácido 4-O-metil-glucurônico (PAULY et al., 2013). Sua estrutura apresenta ramificações e cadeias laterais que interagem facilmente com a celulose, dando flexibilidade ao agregado (SANTOS et al., 2012). A hemicelulose interage por ligações de hidrogênio com as microfibras de celulose, atua como um elo químico entre a celulose e a lignina, fornecendo a espinha dorsal da parede celular (ALMEIDA, 2012).

A hemicelulose, portanto, é um polissacarídeo que difere da celulose por apresentar cadeias moleculares mais curtas, podendo conter grupos pendentes. A existência de grupos pendentes na cadeia polissacarídica principal faz com que a hemicelulose não apresente regiões cristalinas. Alguns açúcares típicos que dão origem à cadeia principal são: glicose, manose e xilose. Enquanto que arabinose, galactose, ácido 4-*O*-metil-glucurônico e grupos acetil são comumente encontrados como grupos pendentes (PATIL et al., 2012). A Figura 3 representa um exemplo de hemicelulose com cadeia principal de xilose (também denominada β-D-Xilopiranose) e os grupos pendentes, ácido 4-*O*-metil-glucurônico, grupos acetil (acetato), α-L-arabinofuranose e ácido ferúlico. Em gramíneas, como a cana-de-açúcar, a principal hemicelulose é a 4-*O*-metil-glucuranoarabinoxilana (JACKSON et al., 2009).

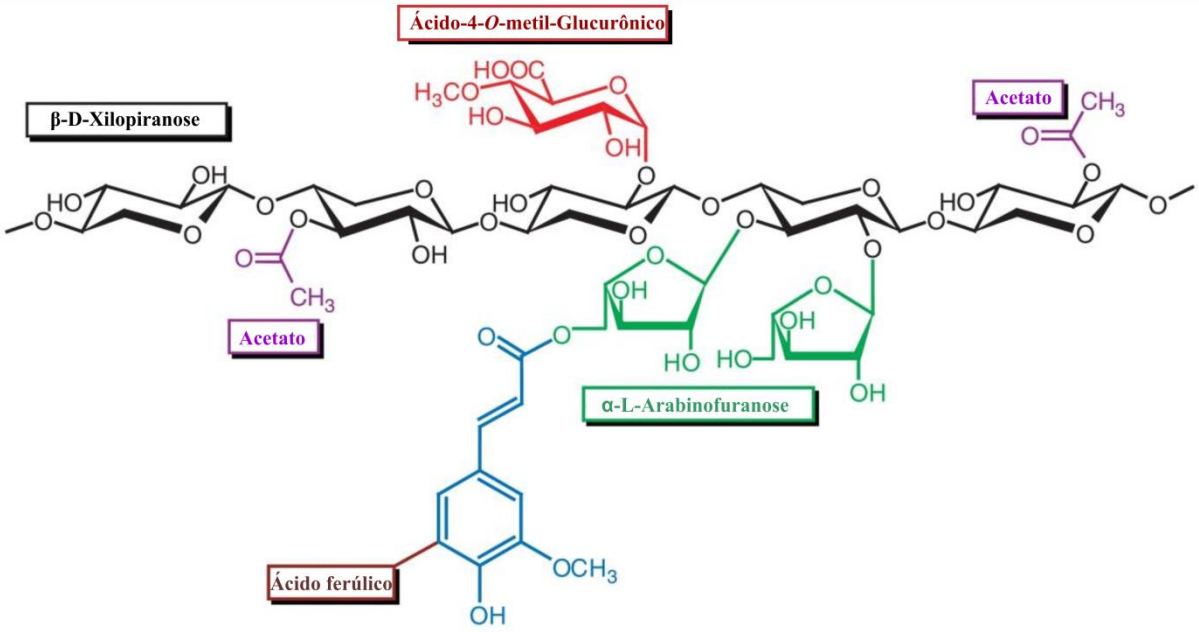


FIGURA 3 – Representação esquemática da estrutura da hemicelulose.
Fonte: Adaptado de (DODD; CANN, 2010).

Geralmente as hemiceluloses são classificadas de acordo com o resíduo de açúcar principal, como, por exemplo, xilanas, mananas e glucanas. Diferentes subclasses de hemicelulose podem ser encontradas, dependendo da espécie vegetal, do estágio de desenvolvimento, do tipo de tecido. Dentre elas, pode-se citar: glucuronoxilanas, arabinoxilanas, mananas lineares, glicomananas, galactomananas, galactoglicomananas, β-glucanas e xiloglucanas (OGATA, 2013).

A lignina, depois da celulose, é o segundo polímero mais abundante na natureza. É um heteropolímero amorfo, formado pela ligação covalente entre monolignóis (álcoois fenilpropanílicos formados na via dos fenilpropanóides). Os monolignóis se polimerizam em reações promovidas por radicais livres produzidos por oxidases. Os principais monolignóis são os alcoóis coniferílico, sinapílico e p-cumárilico que, após a polimerização, são denominados resíduos guaiacil (G), siringil (S) e p-hidroxifenil (H) [Figura 1] (MOREIRA et al., 2011).

A estrutura da lignina não é homogênea, possui regiões amorfas e estruturas globulares. A composição e a organização dos constituintes da lignina variam de uma espécie para outra, dependendo da matriz de celulose-hemicelulose (LIMA, 2013). A lignina possui uma estrutura tridimensional altamente complexa e alta massa molar. Mesmo presente em quantidades menores do que a fração polissacarídica, a lignina confere limitação suficiente para dificultar, ou mesmo impedir, completamente, a atuação microbiana sobre os materiais lignocelulósicos. O efeito protetor conferido pela lignina está baseado em sua hidrofobicidade e na sua distribuição topoquímica no interior da parede celular, impregnando os polissacarídeos e minimizando assim a porosidade da parede (MASARIN et al., 2011). Possui estrutura bem mais complexa que a dos outros dois componentes e pode afetar o processo de sacarificação da biomassa, sendo que sua remoção facilita a hidrólise da celulose (CHANG; HOLTZAPPLE, 2000).

Assim sendo, a lignina é um subproduto do processo de hidrólise que poderá ser utilizado como fonte de energia térmica, isto é, como não oferece açúcar, a lignina não é alvo do processo e pode ser queimada para gerar eletricidade (SCHARR, 2013). Já a hemicelulose e a celulose são policarboidratos que podem ser convertidos em seus monômeros, pentoses e hexoses respectivamente, através da adição de água à molécula. Essa reação é conhecida como hidrólise e, para que a mesma ocorra, é necessária a presença de um catalisador. Por várias décadas, a hidrólise eficiente de material lignocelulósico, a fermentação dos açúcares formados, as pentoses e as hexoses para produzir etanol comercialmente, têm sido um grande

desafio técnico. As rotas mais conhecidas são a hidrólise catalisada por um ácido, denominada de hidrólise ácida, e a outra, por uma enzima, a hidrólise enzimática (WU et al., 2014). Processos catalisados por enzimas são objeto da maior parte dos estudos efetuados atualmente em nível mundial; em princípio, por oferecerem maior conversão e um grande potencial de redução de custos a médio/longo prazo. Há várias opções de processos em estudo hoje.

A enorme importância do desenvolvimento de processos de hidrólise, que sejam economicamente viáveis, está relacionada com a grande disponibilidade de material celulósico na maioria das regiões do mundo e no fato dos açúcares resultantes (hexoses e pentoses) serem matérias-primas adequadas, não só para produção de etanol, mas de uma grande variedade de outros produtos químicos (ELLIOTT, 2004; PHILBROOK et al., 2013).

O setor sucroalcooleiro apresenta o maior potencial para implantação comercial da produção de etanol, a partir de hidrolisado de matéria lignocelulósica. Como o setor sucroalcooleiro no Brasil já produz etanol hidratado e anidro carburante, e já possui as utilidades e a infraestrutura para produção e estocagem do produto final, podendo esta estrutura ser partilhada com as novas unidades de hidrólise. As matérias-primas disponíveis para hidrólise são: o bagaço da cana e os resíduos da colheita, os quais estão disponíveis em grande volume e próximos das instalações industriais, o que resulta em um baixo custo de oportunidade (YADAV et al., 2014; MARJOTTA-MAISTRO; GUILHOTO, 2014).

O etanol, a partir dos açúcares extraíveis da cana, gera grandes excedentes de bagaço, com potencial para que este excedente seja transformado em etanol, o que irá permitir aumentar significativamente a oferta deste combustível, sem exigir um aumento proporcional das áreas de plantio. Nessa nova condição, o aproveitamento da cana será integral (DIAS et al., 2014; UNICA, 2013). O bagaço da cana-de-açúcar é a fração de biomassa resultante após os procedimentos de limpeza, preparo (redução através de jogos de facas rotativas niveladoras e desfibramento através de jogos de martelos oscilantes) e extração do caldo de cana (através de ternos de moagem ou de difusores) (RAMIRO, 2014). Não é uma biomassa homogênea, apresenta variações em sua composição, assim como na sua estrutura morfológica, em função dos procedimentos de corte no campo e no processamento industrial (REZENDE et al., 2011; CANILHA et al., 2012). As Tabelas 2 e 3 reproduzem a composição característica e o teor de carboidratos no bagaço de cana-de-açúcar.

Tabela 2 – Composição do bagaço e palha da cana-de-açúcar.

Composição %(Base seca)	Bagaço	Fibra	Medula	Palha
Celulose	46,6	47,7	41,2	45,1
Pentosanos	25,2	25,0	26,0	25,6
Lignina	20,7	19,5	21,7	14,1
Organosolúveis	2-3	-	-	3,5
Aquosolúveis	2-3	-	-	-
Cinzas	2-3	-	-	8
Umidade (do produto)	48-52	-	-	9,7

Fonte: (NOVACANA, 2014b).

Tabela 3 – Teores de carboidratos no bagaço de cana-de-açúcar.

Polissacarídeos	%(I)	%(II)
Glicose (hexose)	42,9	60,8
Xilose (pentose)	23,1	32,8
Arabinose (hexose)	3,3	4,7
Galactose (hexose)	1,2	1,7
Total	70,5	100

I – No bagaço moído, extraído, livre de umidade e não corrigido de cinzas.
II –Em relação ao total de carboidratos.
Fonte: (NOVACANA, 2014b).

Nesse contexto, o bagaço de cana-de-açúcar apresenta-se como um dos materiais lignocelulósicos com maior potencial para a obtenção de diversos produtos de interesse comercial. No que tange à produção de etanol, são realizadas reações químicas e/ou biológicas, aumentando a produtividade das indústrias. Estimativas estabelecem que a produção de etanol no Brasil pode ser duplicada, sem a necessidade de aumentar áreas de cultivo de cana-de-açúcar (AGUIAR et al., 2013). Entende-se, pois, que, na composição dos materiais lignocelulósicos, é possível discutir sua viabilidade para produção de bioetanol e também formas de recolhimento do material, facilitando e barateando os custos, tornando o processo em si mais viável economicamente (ALEIXO, 2012).

2.2 Produção de etanol de segunda geração (2G)

O etanol, também denominado álcool etílico (C_2H_5OH ou C_2H_6O) é produzido desde os tempos antigos pela fermentação dos açúcares encontrados em produtos vegetais. Ainda hoje, grande parte do etanol industrial é obtido pelo mesmo processo, embora também possa ser produzido a partir de eteno, derivado do petróleo (BASTOS, 2007).

Como verificado anteriormente, o material lignocelulósico apresenta uma estrutura complexa e compacta, tornando-se necessária uma etapa de pré-tratamento, tanto a fim de remover a estrutura amorfa da celulose (Figura 4) quanto a fim de remover a lignina, expondo as moléculas de celulose e hemicelulose à ação enzimática (Figura 5) para a produção do etanol de segunda geração.

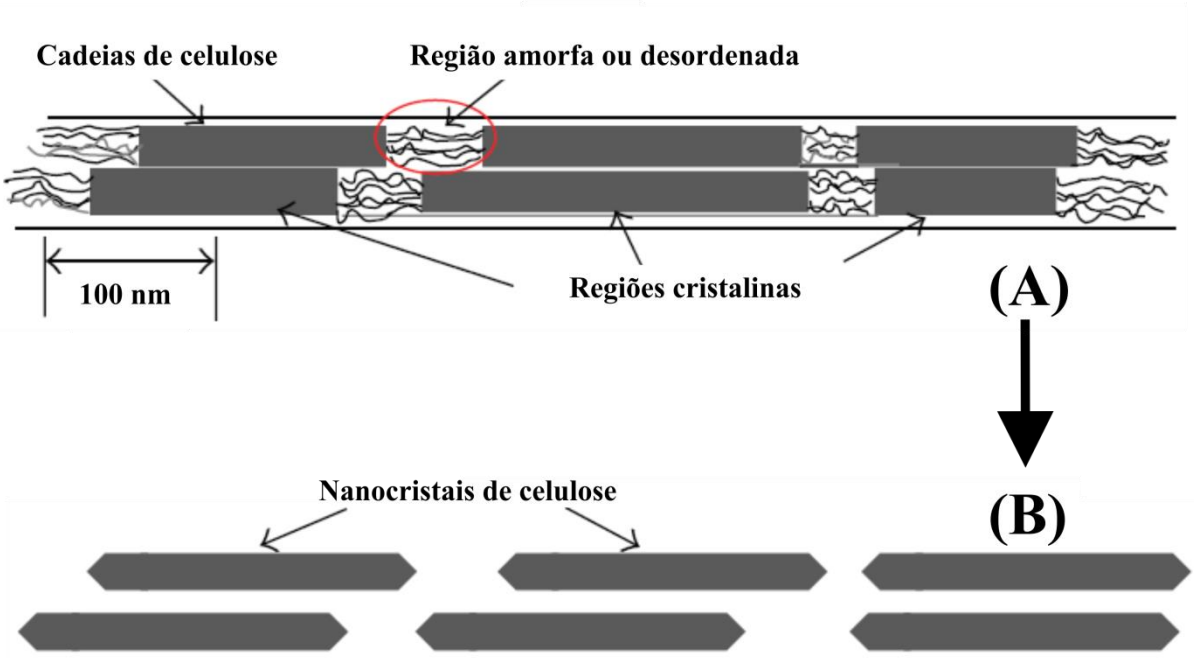


FIGURA 4 – Estrutura da celulose antes e depois do pré-tratamento (hidrólise ácida).
Fonte: (CHAMI KHAZRAJI; ROBERT, 2013).

O pré-tratamento da biomassa é necessário por conta das fortes ligações entre os compostos que formam a biomassa, seu objetivo principal é desorganizar a estrutura do material, aumentando sua porosidade e permitindo que as enzimas ou os ácidos (dependendo do tipo de hidrólise) atuem melhor (FERREIRA, 2012). Para que o pré-tratamento seja satisfatório deve atender as seguintes finalidades: melhorar a formação de açúcares ou a capacidade de futura formação de açúcares pela hidrólise; evitar a degradação ou perda de

carboidratos; evitar a formação de coprodutos que inibam a hidrólise e ter baixo custo (SANTOS et al., 2012).

Atualmente, há uma grande quantidade de processos de pré-tratamento disponíveis, podendo ser: físicos; químicos (utilizam componentes químicos ácidos e bases); biológicos (utilização de fungos que mudam a composição química da matéria); fracionamento por solvente; além da combinação de alguns (Tabela 4), sendo catalisado por temperaturas e pressões elevadas (RABONATO, 2013). Os mais conhecidos são: o emprego de álcali; hidrólise ácida; explosão a vapor; água quente; fluido supercrítico; amônia líquida; e hidróxido de sódio. Todos utilizados para desmanchar a estrutura da biomassa lignocelulósica, mas qualquer que seja o tipo deve apresentar o efeito que é mostrado na Figura 5 de maneira simplificada.

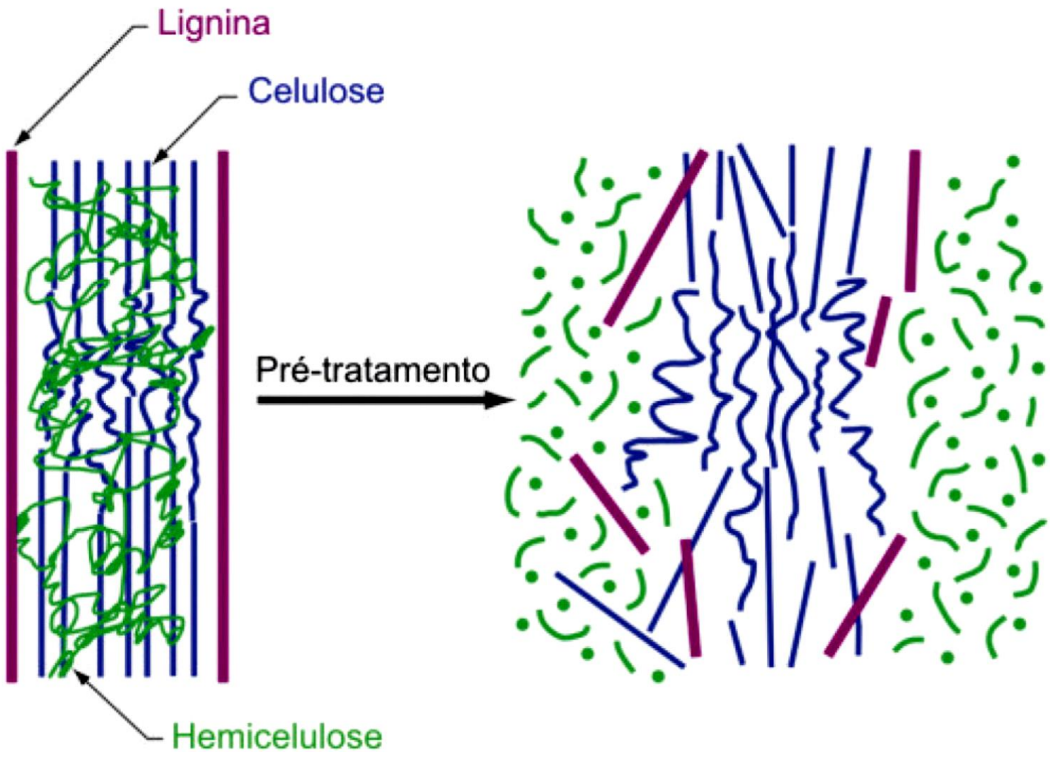


FIGURA 5 – Alterações estruturais do complexo celulose-hemicelulose-lignina determinadas pelo pré-tratamento.
Fonte: (SANTOS et al., 2012).

Na Tabela 4, são apresentados, de forma simplificada, os diversos tipos de pré-tratamentos e as respectivas mudanças causadas à biomassa lignocelulósica (SANTOS et al., 2012).

Tabela 4 – Efeitos de diferentes pré-tratamentos nos materiais lignocelulósicos.

Pré-tratamento		Características composicionais			Vantagens	Desvantagens
		Celulose	Hemicelulose	Lignina		
Físico	Moinho de bolas	Intensiva diminuição do grau de cristalinidade	Não remove	Não remove	Redução de cristalinidade	Alto consumo de energia
Químico	Ácido diluído	Pouca despolimerização	80-100% de remoção	Pouca remoção, mas ocorre mudança na estrutura	Condições médias, alta produção de xilose	Difícil recuperação do ácido, corrosivo e relativamente custoso
	Hidróxido de sódio	Inchação significativa	Considerável solubilidade	Considerável solubilidade, >50%	Remoção efetiva de ésteres	Reagente caro, recuperação cristalina
	ARP*	Menor que 5% de despolimerização	~50% de solubilidade	~70% de solubilização	Efetiva deslignificação	Recuperação alcalina, relativamente caro
	Hidróxido de cácio	Pouca despolimerização	Significativa solubilização	Solubilização parcial (~40%)	Efetiva remoção de lignina e acetil, baixo custo	Menor efetividade devido a pouca solubilidade da cal
	Ozonólise	Não foi observada despolimerização	Pequena solubilização	Solubilidade acima de 70%	Efetiva deslignificação em condições suaves	Caro, necessidade de mais ozônio
	Organosolv	Considerável inchação	Significativo, quase completa	Significativo, pode ser quase completa	Alta produção de xilose, efetiva deslignificação	Recuperação de solvente cara
Biológico	Biológico	20-30% de despolimerização	Acima de 80% de solubilização	~40% de deslignificação	Baixo requerimento de energia, efetiva deslignificação	Perda de celulose, baixa taxa de hidrólise
Combinado	Explosão a vapor	Pouca despolimerização	80-100% de remoção	Pouca remoção, mas ocorre mudança na estrutura	Energia eficiente, nenhum custo de reciclagem	Degradação da xilana como produto inibitório
	AFEX**	Diminuição do grau de cristalinidade	Acima de 60% de solubilidade	10-20% de solubilização	Menor perda de xilanas, não formação de inibidores	Recuperação de amônia, não é efetivo para alta concentração de lignina

Fonte: Adaptado de (SANTOS et al., 2012).
*ARP = Ammonia Recycle Percolation ou Percolação com reciclo de amônia;
**AFEX = Ammonia Fiber Explosion ou Explosão da Fibra com Amônia.

A principal diferença entre o etanol de primeira e segunda geração é a matéria-prima. Enquanto o primeiro é produzido a partir do caldo de cana, o segundo pode ser feito a partir da celulose da planta, presente - por exemplos – bagaço da cana-de-açúcar ou sabugo, folhas e talos de milho (Figura 6) (ROSTAGNO et al., 2014). Em relação às propriedades físico-químicas, o etanol 2G é igual ao etanol tradicional, o que muda é o seu processo de fabricação.

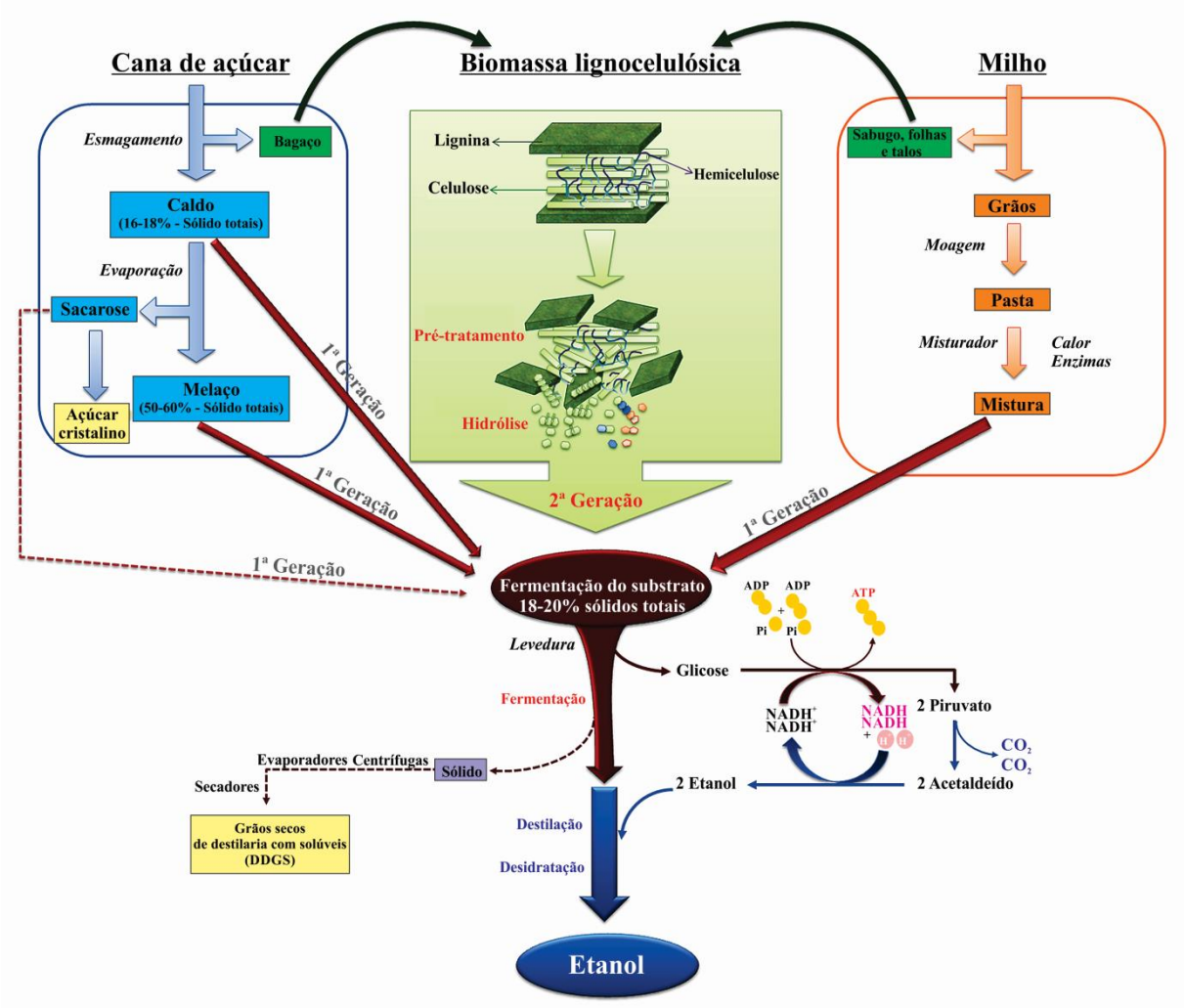


FIGURA 6 – Descrição geral do processo para a produção de bioetanol de primeira e segunda geração.
Fonte: Adaptado de (ROSTAGNO et al., 2014).

Portanto, o processo de fabricação do etanol a partir do bagaço pode ser dividido nas seguintes etapas: Preparo do bagaço (pré-tratamento), hidrólise enzimática (enzimas rompem as fibras da celulose e/ou hemicelulose, liberando açúcares mais simples de serem fermentados), fermentação (leveduras convertem a glicose e/ou xilose em etanol e CO_2 como

subprodutos metabólicos) e destilação[3] (AGUIAR et al., 2013; DRABER, 2013; ROSTAGNO et al., 2014). Normalmente, hidrólises enzimáticas possuem um rendimento de açúcar menor que 20%, enquanto que, se uma etapa de pré-tratamento for utilizada antes da hidrólise enzimática, o rendimento pode alcançar até > 90% (LUCY et al., 2010).

2.2.1 Etanol 2G: uma realidade no Brasil

Segundo a ANP – Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (2014) o Brasil possui no total 408 usinas produtoras de etanol e açúcar, sendo que dessas 382 usinas possuem autorização pela referida agência (Figura 7). Essas usinas possuem as seguintes capacidades totais para a produção de etanol de primeira geração: 3.690 (Norte), 30.716 (Nordeste), 181.731 (Sudeste), 74.793 (Centro-Oeste) e 18.494 (Sul) [todos os valores são representados em m^3 por dia].

FIGURA 7 – Usinas autorizadas pela ANP por região.
Fonte: (ANP, 2014).

[3] Operação capaz de separar dois ou mais compostos presentes em uma mistura líquida que possuam distintas temperaturas de ebulição, como, por exemplo, o etanol e a água. Nesse caso, o etanol evapora primeiro que a água, ao atingir 78°C, e separa-se do líquido. O vapor é resfriado e recuperado por condensação, obtendo-se o etanol em estado líquido com alto grau de pureza (GRANBIO, 2014).

Para a produção de etanol 2G, essas usinas fornecerão matérias-primas, como o bagaço e palha da cana-de-açúcar (considerados descartes), para que possa existir um sistema totalmente integrado com a produção de etanol de primeira geração já existente nas usinas, favorendo, a produção de etanol 2G sem que haja um aumento da área de cultivo, proporcionando, assim, uma maior eficiência da matriz energética, reduzindo os desperdícios e aumentando a produção sem exigir mais recursos (SOUZA, 2013a).

Entre 2014 e 2015, entrarão em operação as duas primeiras fábricas brasileiras de etanol 2G, utilizando a celulose como matéria-prima (bagaço e palha de cana-de-açúcar). A primeira que entrou em operação foi uma unidade da GranBio [primeira indústria de etanol celulósico da América Latina] (Figura 8), em São Miguel dos Campos, em Alagoas, com capacidade para 82 milhões de litros anuais, utilizando a tecnologia Proesa® de *Beta Renewables* (RENEWABLES, 2014) e conta também com o apoio do BNDES, e de uma série de empresas, como a *Chemtex*, DSM, *Novozymes* e Grupo Carlos Lyra (SIQUEIRA, 2014). No momento, a unidade, orçada em R$ 350 milhões, está em fase de testes, e teve a previsão de entrar em operação no final de 2014 (ZAPAROLLI, 2014). Na GranBio, os planos são de expandir a produção de etanol de segunda geração para um bilhão de litros por ano até 2020, por meio de associações com usinas convencionais para o aproveitamento do bagaço e da palha (Figura 9), e também com o desenvolvimento de uma nova variedade de cana-de-açúcar, com maior potencial energético, batizada pela companhia de cana-energia (Cana Vertix[4]), que tem seu primeiro plantio programado para 2015 (SIFAEG, 2014).

FIGURA 8 – Primeira usina de etanol celulósico do Brasil.
Fonte: (LANE, 2014).

[4] É um conceito utilizado para designar variedades de cana-de-açúcar que possuem uma elevada produção de fibra e baixo teor de açúcar, gerando muito mais biomassa por hectare. Além disso, essas plantas crescem mais rapidamente que as variedades convencionais de cana e são mais resistentes a pragas e ao estresse hídrico (GRANBIO, 2014).

FIGURA 9 – Palha de cana-de-açúcar sendo estocada para produção de etanol de segunda geração (GranBio). Fonte: (LANE, 2014).

A segunda fábrica a entrar em operação, está orçada em R$ 230 milhões, e pertence a Raízen e está sendo construída em Piracicaba, no interior paulista, ao lado da Usina Costa Pinto, que pertence ao grupo. A nova unidade terá capacidade para produzir 40 milhões de litros por ano. Evandro Curtolo da Cruz, gerente de novas tecnologias da Raízen, informa que a companhia tem planos de erguer mais sete fábricas de etanol celulósico até 2024, todas próximas de usinas convencionais já existentes para facilitar a logística da palha e do bagaço (ZAPAROLLI, 2014).

Com a tecnologia aplicada, o Brasil será o maior produtor de bionegia do mundo, através da biomassa. Cerca de 450 milhões é a quantidade de hectares de terras agricultáveis subutilizadas no Brasil, quantidade muito superior a de qualquer outro país (GRANBIO, 2014). No entanto, a Itália inaugurou, em abril de 2014, a primeira usina no mundo (Usina de Crescentino – Figura 10) com capacidade para produzir comercialmente o chamado etanol celulósico ou etanol 2G, a partir de Julho, utilizando resíduos da agricultura como as matéria-primas palhas de arroz e de trigo e uma espécie de cana gigante conhecida como cana-do-reino (SINDAG, 2014). A estrutura tem capacidade para produzir 75 milhões de litros de etanol celulósico, consome, em média, 270 mil toneladas de matéria-prima por ano. A unidade italiana, até então era a maior do planeta, mas existe uma "corrida global" pelo etanol 2G que ganhará novos contornos com a construção de uma nova planta na China, que será a maior do mundo e contará com a parceria da *Novozymes* e a *Beta Renewables*. Com investimentos de U$325 milhões (equivalente a R$ 732 milhões), a nova usina ficará pronta

até o final de 2016 na cidade de Fuyang e processará entre 970 mil e 1,3 milhão de toneladas de biomassa por ano (matéria-prima a ser definida). Esta fábrica de etanol 2G terá capacidade para processar quatro vezes mais, aproximadamente, 235 milhões de litros de etanol celulósico por ano de matéria-prima, que a usina italiana e número quase três vezes superior à capacidade projetada para a usina da GranBio, em Alagoas/Brasil (SIQUEIRA, 2014).

FIGURA 10 – Fotos da Usina de Crescentino em operação na Itália.
(A) Visão geral noturna da usina; (B) Matéria-prima (resíduos de biomassa) sendo estocada; (C e D) Matéria-prima sendo processada; (E) Etanol 2G pronto para ser comercializado.
Fonte: (SIQUEIRA, 2014).

Cellic® CTec3 deve ser a enzima adotada nos projetos brasileiros de etanol celulósico já anunciados pela GranBio e a Raízen (SCHARR, 2013). Cellic® CTec3 é um coquetel enzimático produzido pela empresa baseada em biotecnologia (com sede na Dinamarca, conhecida por *Novozymes*) que apresenta um melhor desempenho e custos competitivos para a

produção de biocombustíveis a partir da biomassa (NOVOZYMES, 2014). Esse coquetel é composto de enzimas como: celulases (incluindo GH61 e β-glicosidases), bem como uma nova gama de atividades de hemicelulases, que são responsáveis em converter as fibras de celulose e hemicelulose em açúcares fermentacíveis, como glicose e xilose (NOVOZYMES, 2014; SCHARR, 2013). Não é divulgado pela empresa de qual ou quais micro-organismos são obtidos essas enzimas.

A levedura tradicionalmente usada na fermentação alcoólica da sacarose na primeira geração não é eficiente para processar outros açúcares contidos na cana, como a xilose – monossacarídeo do tipo aldopentose – um dos principais constituintes da biomassa. E é este açúcar que está no alvo da GranBio. Trata-se da "linhagem RN1016 de Levedura *Saccharomyces cerevisiae*", que será importante para fermentar açúcares de 5 carbonos, como a xilose e a arabinose, além da tradicional glicose de 6 carbonos (SCHARR, 2014). De acordo com o Relatório de Biossegurança da linhagem RN1016 de Levedura *S. cerevisiae* para produção de etanol, produzido pela BioCelere Agroindustrial Ltda., (PEREIRA et al., 2013, p. 32) determina que:

> A Levedura *S. cerevisiae* cepa RN1016 expressa o gene XylA, codificador da enzima xilose isomerase. Este gene é oriundo do micro-organismo não patogênico *Piromyces sp*., e catalisa a conversão de D-xilose em D-xilulose. Além da inserção do gene XylA, outros genes endógenos da levedura, que participam do metabolismo de açúcar, tiveram sua expressão aumentada com a finalidade de tornar mais eficiente o processo fermentativo de carboidratos, especialmente de 5 carbonos. Este aumento de expressão se deu através da utilização de promotores constitutivos da própria levedura. Estes genes são: XKS1 (codificador de uma xiluloquinase), TAL1 (codificador de uma transaldolase), TKL1 (codificador de uma transcetolase), RPE1 (codificador de uma D-ribulose-5-fosfato epimerase) e RKI1 (codificador de uma ribose-5-fosfato ceto-isomerase). A linhagem RN1016 ainda apresenta a deleção do gene GRE3 (codificador de uma aldose redutase), visto que esse gene é responsável pela produção de xilitol a partir da xilose, diminuindo o rendimento de etanol. A Figura 11 demonstra de maneira simplificada a conversão de xilose em etanol.

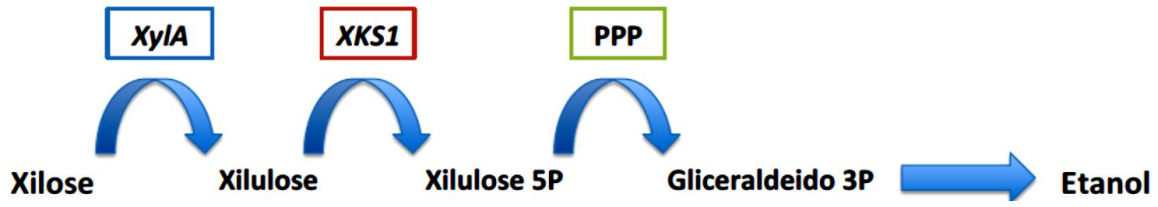


FIGURA 11 – Via metabólica simplificada de conversão de xilose em etanol.
XylA: xilose isomerase; XKS1: xiloquinase; PPP: via das pentose fosfato (*Pentose Phosphate Pathway*), representado pelos genes TAL1, TKL1, RPE1 e RKl1.
Fonte: (PEREIRA et al., 2013).

A Figura 12 resume todo o processo para transformar resíduos agrícolas, como palha e bagaço de cana-de-açúcar, em combustível (etanol de segunda geração) pela GranBio.

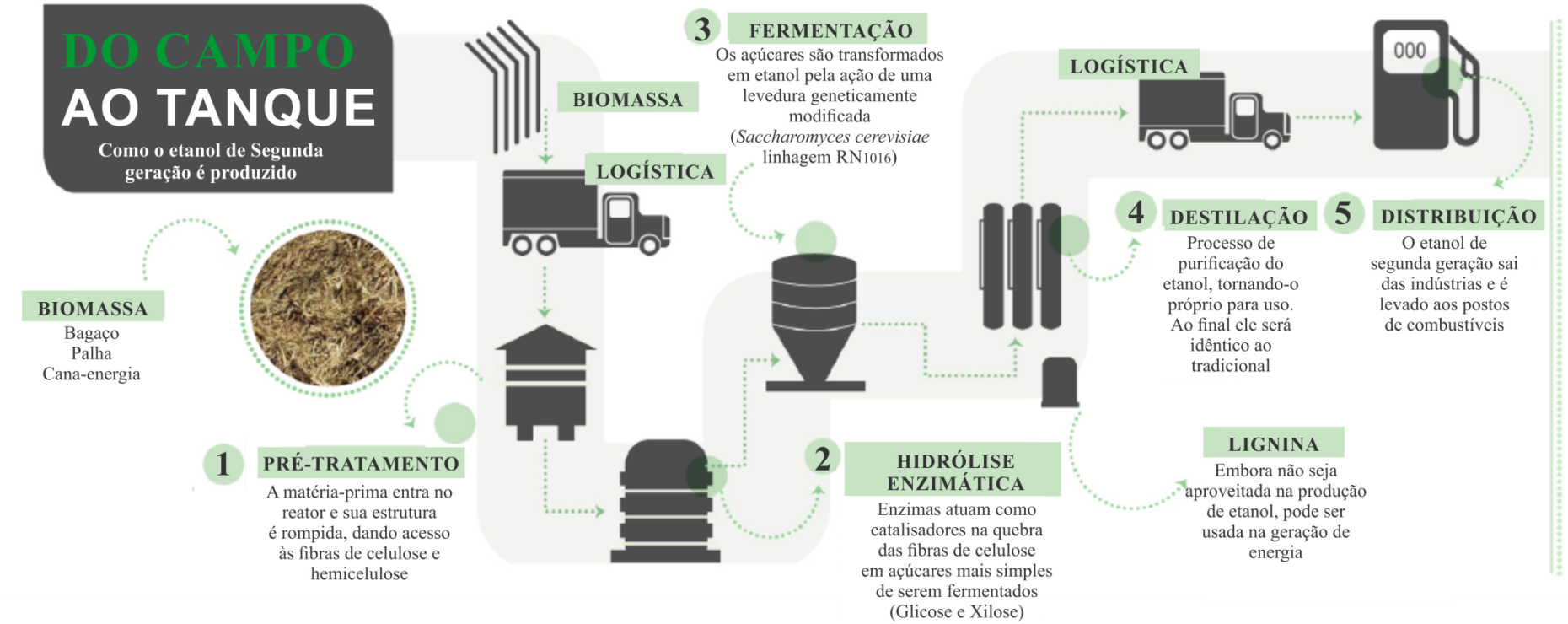


FIGURA 12 – Etapas do processo de produção do etanol celulósico.
Fonte: Adaptado de (PEREIRA et al., 2013).

Segundo Scharr (2013), outras empresas são destaques em fornecer enzimas para a produção de etanol celulósico, como: *Codexis*/EUA (A empresa produz um pacote de enzimas com atividade de celulase chamado "*CodeXyme*" e está desenvolvendo o produto para atender às exigências técnicas de processos de etanol celulósico da Shell); *Dyadic*/EUA (A empresa desenvolveu o que chamou de tecnologia de plataforma "C1" – esta é uma cepa do fungo chamado *Chrysosporium lucknowense*, originário da Rússia e cresce em solo alcalino, que pode crescer em condições extremas); *Genencor*/EUA (A empresa desenvolveu uma enzima chamada "*Accellerase 1000*", com várias atividades enzimáticas provenientes de uma mistura de diferentes tipos de enzimas, principalmente exoglucanase, endoglucasnase, hemicelulase e β-glicosidase; depois desenvolveu "*Accellerase 1500*", uma versão melhorada da enzima, a partir de uma cepa geneticamente modificada de *Trichoderma reesei* (DUPONT, 2013)).

A tecnologia da *Accellerase*® será utilizada pela usina de etanol 2G da DuPont em Nevada/EUA, no estado de Iowa, que começou a ser construída em 2012 e teve sua previsão de iniciar em escala comercial no ano de 2014. O potencial estimado da estrutura é de 113 milhões de litros por ano. A palha de milho será fornecida por, aproximadamente, 500 agricultores, que estão localizados dentro de um raio de 48 km em torno da usina. O Estado de Iowa é o maior produtor de milho dos Estados Unidos. Outras três usinas dedicadas à produção de etanol celulósico também teve previsão de iniciar as suas atividades em 2014 (UNICA, 2014). Portanto, o ano de 2014 foi um marco tanto para o Brasil como para os Estados Unidos para iniciar as operações das primeiras usinas destinadas à produção de etanol celulósico, ou de segunda geração em escala comercial.

2.3 *Trichoderma reesei*

2.3.1 Gênero, taxonomia, morfologia e mutações

Segundo Samuels (2006), *Trichoderma* foi proposto pela primeira vez como um gênero por *Persoon* em 1794, com base em materiais coletados na Alemanha, incluindo quatro espécies na descrição (sem a percepção de que apenas os anamorfos estavam sendo descritos). Somente em 1969 que foi realizada a primeira revisão taxonômica e reclassificada por Rifai. A partir daí, um maior número de espécies foi agregado ao gênero, chegando à

atualidade com cerca de 83 taxons (espécies, formas e variedades), incluindo *Trichoderma* e *Hypocrea*.

Hypocrea é o gênero que constitui uma das duas fases reprodutivas que *Trichoderma* possui, denominada de teleomórfica. Nessa fase o gênero *Hypocrea* é classificado como ascomiceto da ordem *Hypocreales*, no qual predomina uma etapa sexual. A segunda fase, denominada anamórfica, prevalece uma etapa assexual (Figura 13), que parece ser independente da fase teleomórfica, seja em nível de indivíduo ou de população (HARMAN et al., 2004). Portanto, as duas formas de uma mesma espécie acabam recebendo nomes diferentes. Caso típico é o que ocorre com *T. reesei*, considerado como a forma anamórfica do fungo *Hypocrea jecorina* (teleomorfo), ambos representando a mesma espécie. As espécies do gênero *Trichoderma* são micro-organismos presentes no solo, em materiais vegetais e substratos em decomposição, e, em geral, são organismos dominantes na microflora de solos de uma grande variedade de "*habitats*", atribuindo-se tal fato à sua diversa capacidade metabólica e à sua natureza competitiva (SAMUELS, 1996).

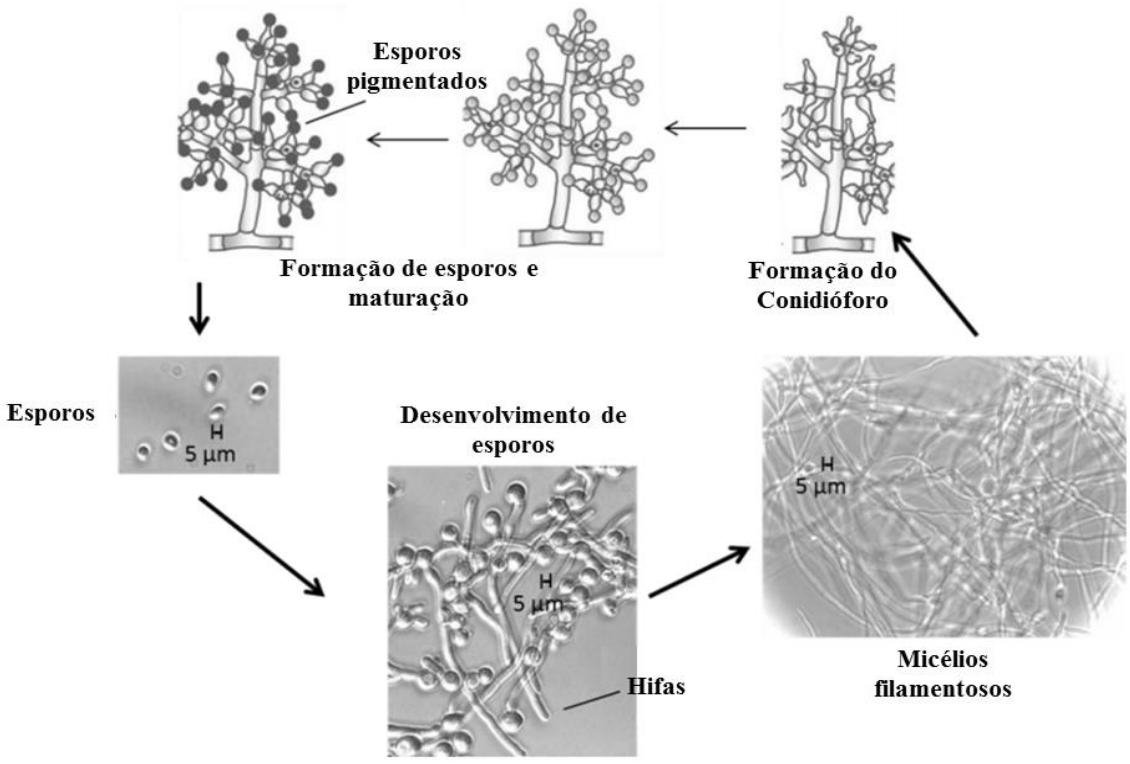


FIGURA 13 – Ciclo de vida assexuada de *T. reesei*.
Fonte: Adaptado de (RIES, 2013a).

Segundo a classificação taxonômica no banco de dados "*National Center for Biotechnology Information* – NCBI (http://www.ncbi.nlm.nih.gov/taxonomy/), *Trichoderma reesei* tem a seguinte classificação: Taxonomy ID: 51453; Super-reino: *Eukaryota*; Reino: *Fungi*; Filo: *Ascomycota*; Ordem: *Hypocreales*; Família: *Hypocreaceae*; Gênero: *Hypocrea*; Espécie: *Hypocrea jecorina* – teleomorfo e *Trichoderma reesei* – anamorfo.

Morfologicamente, *T. reesei* apresenta hifas septadas, ramificadas (5-10 μm de diâmetro), polinucleadas, de núcleo haplóide, que formam colônias brancas de rápido crescimento, formando almofadas verdes ou amarelas de filamentos de esporulação. Os filamentos férteis, conidióforos, produzem fileiras laterais de pequenas fiálides. Os mitoesporos (conídios de cor verde e de 3 a 5 μm de diâmetro) são produzidos sucessivamente no extremo da fiálide e agrupados em pequenas massas (SCHUSTER; SCHMOLL, 2010).

Juntamente com *Neurospora*, *Penicillium* e *Aspergillu*s, o gênero *Trichoderma* é um dos mais estudados entre os fungos filamentosos devido ao seu grande potencial de aplicação do ponto de vista industrial e biotecnológico (SAMUELS, 1996). Com destaque para *T. reesei*, um fungo filamentoso, mesófilo dotado com um enorme repertório de enzimas hidrolíticas (celulases e hemicelulases) relacionadas com a desconstrução da biomassa lignocelulósica de plantas que são de grande importância para os processos biotecnológicos, principalmente na indústria de papel ou produção de combustível de segunda geração (TOMME et al., 1988; STERNBERG; MANDELS, 1979; SCHUSTER; SCHMOLL, 2010; MAURYA et al., 2012; MINTY et al., 2013).

Uma particularidade importante de *T. reesei* é que todas as linhagens utilizadas na produção e pesquisa industrial são derivadas de um único isolado natural QM6a (o tipo selvagem sequenciado por Martinez et al., (2008), compreendendo 9129 genes com, aproximadamente, 34 Mbp), que foi originalmente isolado nas Ilhas Salomão durante Segunda Guerra Mundial (VITIKAINEN et al., 2010). Todas as mutações com potencial impacto sobre a produção de celulases foram analisadas com o objetivo de aumentar a produção. Assim, duas séries foram obtidas a partir da linhagem QM6a, Figura 14, (KUBICEK, 2013, p.137):

> 1) Série *Rutgers* – na Universidade *Rutgers* (*New Jersey*/EUA), foram desenvolvidas as linhagens RUT-M7, RUT-NG14 e RUT-C30. Esta última linhagem, é mais bem caracterizada, sendo referência entre os grandes produtores de celulases é a mais utilizada desta série, e que passou por três mutações (uma truncagem do gene que codifica o repressor catabólico de carbono CRE1 (ILMÉN et al., 1996); uma mutação de "*frameshift*" no gene *gls2* da subunidade alfa envolvido na glicosilação

de proteína que aumenta a secreção da proteína (GEYSENS et al., 2005); e uma deleção de 85-kb, que eliminou 29 genes, incluindo transportadores, fatores de transcrição e enzimas metabólicas primárias (SEIDL et al., 2008) foram descritas anteriormente para ocorrer na linhagem RUT-C30.
2) Série *Natick*: Laboratório *Natick* (*Massachusetts*/EUA), foram desenvolvidas as linhagens QM9123 e QM9414 por acelerador linear (LA), sendo que foram identificados mais de quarenta mutações de vários tipos, e a linhagem QM9414 produz quatro vezes mais enzimas do que a cepa isolada QM6a (VITIKAINEN et al., 2010; KUBICEK, 2013). Todas as mutações em QM9414 podem ser verificadas no trabalho de Vitikainen e colaboradores (2010).

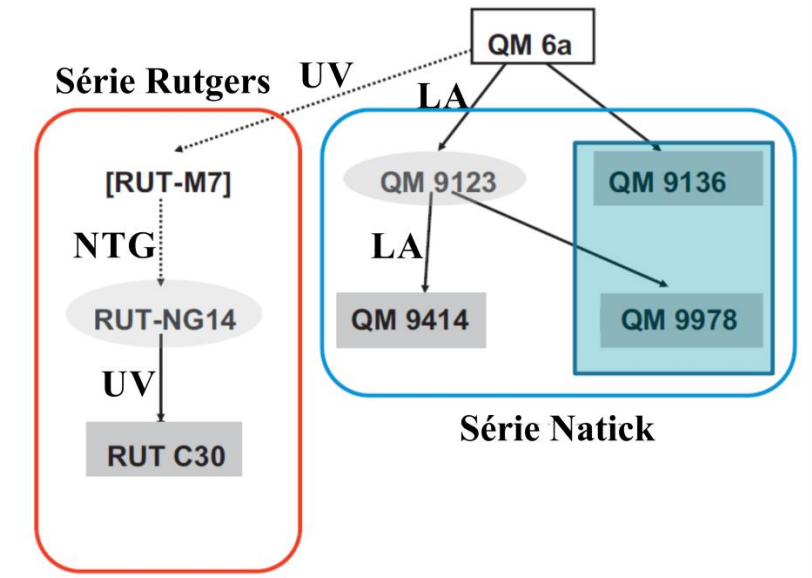


FIGURA 14 – Genealogia dos mutantes de *T. reesei* QM6a.
UV: radiação ultravioleta; LA: acelerador linear e NTG: nitrosoguanidina. Os mutantes negativos para produção de celulases estão localizados dentro da caixa azul (QM 9136 e QM 9978).
Fonte: Adaptado de (KUBICEK, 2013; VITIKAINEN et al., 2010).

2.3.2 *CAZy*

O fungo filamentoso *T. reesei* produz enzimas capazes de clivar ligações existentes em carboidratos. Estas enzimas são conhecidas como Glicosil Hidrolases (GHs), Glicosil Transferases (GTs), Polissacarídeos Liases (PLs), Carboidratos Esterases (CEs), Módulo de ligação a carboidratos (CBM) e Polissacarídeos monooxigenase (PMOs – e, recentemente, consideradas como uma nova classe de enzima, "Atividades auxiliares (AAs)" – (LOMBARD et al., 2014), que são classificadas e descritas no banco de dados *Carbohydrate-Active enZYmes Database* (*CAZy* - www.*CAZy*.org) (HÄKKINEN et al., 2012). Esse banco de dados descreve as famílias de proteínas estruturalmente relacionadas à atividade catalítica ou de ligação a carboidratos que sejam capazes de degradar, modificar ou criar ligações glicosídicas. Utiliza também a sequência de aminoácidos para correlacionar os mecanismos

de ação e enovelamentos proteicos em famílias distintas. As famílias proteicas do *CAZy* são normalmente designadas através de um formulário simples, incluindo a classe ou categoria do módulo e um número que reflete a ordem da criação familiar dentro do mesmo grupo (CANTAREL et al., 2009). O genoma de *T. reesei* codifica caracterizadas enzimas como: 10 celulases e 16 hemicelulases que demonstrou ser importante para aplicações industriais (HÄKKINEN et al., 2014), além de apresentar uma quantidade de genes que codificam *CAZy* (FOREMAN et al., 2003; MARTINEZ et al., 2008). De acordo com uma anotação de atualização, realizada em 2012, o genoma de *T. reesei* codifica 201 genes glicosil hidrolases, 22 genes carboidratos esterases e 5 genes polissacarídeos liases, dos quais, pelo menos, 66 são genes conhecidos ou previstos para codificar atividades celulolíticas e hemicelulolíticas (HÄKKINEN et al., 2012).

2.3.3 O processo de hidrólise enzimática: ação sinérgica dos sistemas celulolítico e hemicelulolítico de *T. reesei*

Apesar de *T. reesei* ser o mais proeminente degradador lignocelulósico entre o gênero *Trichoderma*, esta espécie tem um número reduzido de enzimas celulolíticas quando comparado com outros fungos lignocelulósicos (MARTINEZ et al., 2008). Essa competência é atribuída para *T. reesei*, pois possui sistemas eficientes para o transporte de nutrientes e a indução/secreção de celulases. Subsequentemente, o estudo do sistema celulolítico em *T. reesei* é de interesse substancial para a indústria biotecnológica (CASTRO et al., 2014a).

O sistema celulolítico deste fungo consiste em pelo menos três tipos de enzimas: exoglucanases (CBHs – celobiohidrolases, EC 3.2.1.91), endoglucanases (EGs – EC 3.2.1.4) e β-glicosidases (BGs – EC 3.2.1.21) que ocorrem em várias isoformas (SALOHEIMO et al., 1997). *T. reesei* produz, pelo menos, duas celobiohidrolases (CEL7A e CEL6A), cinco endoglucanases (CEL7B, CEL5A, CEL5B, CEL12A, CEL45A), duas β-glicosidases caracterizadas (CEL3A e CEL1A), e um adicional de cinco preditas β-glicosidases (CEL3B, CEL3D, CEL1B, CEL3C, CEL3E) (HÄKKINEN et al., 2012). Além dessas enzimas clássicas, outras proteínas/enzimas recentemente identificadas em *T. reesei* estão envolvidas na degradação da celulose, tais como: a solenina (SWOI - que são proteínas com ação de relaxamento e extensão da parede celular em plantas facilitando a ação de outras enzimas, isto é, perturba a estrutura de celulose cristalina, sem aparente liberação de açúcares

(SALOHEIMO et al., 2002); expansinas/família 45 endoglucanase-*like* (EEL1, EEL2 e EEL3) que apresentam semelhanças ao gene *swo1* (MARTINEZ et al., 2008; VERBEKE et al., 2009); e polissacarideos monooxigenases (PMOs – também conhecidos como glicosil hidrolases da família 61 (GH61)), foram recentemente re-anotados como AA9 (*Auxiliary Family Activity* 9, www.*CAZy*.org), que ajudam na degradação lignocelulósica por um mecanismo oxidante (HÄKKINEN et al., 2012; DIMAROGONA et al., 2012; HORN et al., 2012; BEY et al., 2013; LEVASSEUR et al., 2013).

Hemicelulases produzidas por *T. reesei* incluem cinco xilanases (EC 3.2.1.8 – XYNI, XYNII, XYNIII, XYNIV e XYNV) (TÖRRÖNEN et al., 1992; TENKANEN et al., 1992; XU et al., 1998; METZ et al., 2011), uma mananase (MANI) (STÅLBRAND et al., 1995), duas acetilxilanas esterases (EC 3.1.1.6 – AXEI e AXEII [predita]) (MARGOLLES-CLARK et al., 1996a; FOREMAN et al., 2003), uma α-glucuronidase (EC 3.2.1 – GLRI) (MARGOLLES-CLARK et al., 1996b), uma caracterizada (ABFI) e duas preditas (ABFII e ABFIII) arabinofuranosidases (EC 3.2.1.55) (MARGOLLES-CLARK et al., 1996c; FOREMAN et al., 2003; HERPOËL-GIMBERT et al., 2008), três α-galactosidases (AGLI, AGLII e AGLIII) (MARGOLLES-CLARK et al., 1996d), bem como uma β-xilosidase (EC 3.2.1.37 – BXLI) (MARGOLLES-CLARK et al., 1996c) que digere oligossacarídeos derivados de xilana. Uma acetil esterase (AESI) que remove o grupo acetil da hemicelulose também foi identificada (LI et al., 2008). Sugere-se também que a glucuronoil esterase (CIPII) participa da degradação da biomassa lignocelulósica, clivando ligações éster entre lignina e hemicelulose, facilitando a remoção da lignina (FOREMAN et al., 2003; POKKULURI et al., 2011).

Nevalainen e Peterson (2014) esquematizaram como essas proteínas/enzimas são produzidas em *T. reesei,* e, que estas são secretadas pela ponta crescente das hifas (Figura 15). Neste sentido, a degradação da biomassa lignocelulósica por fungos ocorre de forma extracelular e é atribuída à ação das enzimas retratadas anteriormente que atuam sinergicamente. Os componentes da biomassa lignocelulósica devem ser inicialmente despolimerizados até compostos menores, tornando-se suscetíveis ao transporte através da parede celular e ao metabolismo intracelular do fungo (MARCO, 2014). Em *T. reesei*, as celobiohidrolases I (CBHI), correspondem a 60% do total de proteínas secretadas, enquanto celobiohidrolases II (CBHII) e endoglucanases são responsáveis, respectivamente, por 20 a 10% (TIWARI et al., 2013) e as BGs tipicamente tornam-se menos do que 1% do número total de enzimas secretadas (MARGEOT et al., 2009).

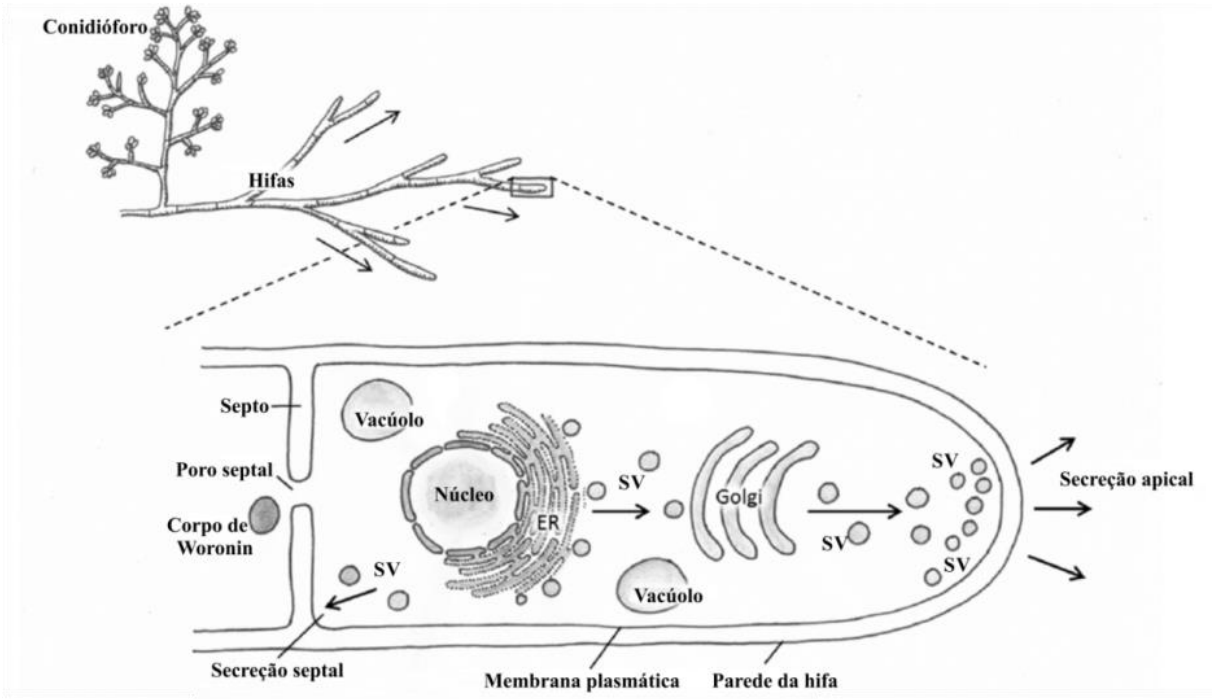


FIGURA 15 – Representação esquemática da morfologia e as vias propostas a síntese e secreção de proteínas em *T. reesei*.
Proteínas são sintetizadas no retículo endoplasmático (ER), em seguida, viajam em vesículas secretoras (SV) para o *Golgi* e sofrem modificações pós-transcricionais. Vesículas secretoras, transportam as proteínas modificadas para a ponta das hifas para secreção apical, ou possivelmente para os septos como uma via alternativa para secreção.
Fonte: Adaptado de (NEVALAINEN; PETERSON, 2014).

A ação sinérgica de diferentes tipos de celulases é um essencial fenômeno na hidrólise de celulose. Dois tipos de sinergismo têm sido identificados, uma endo-exo entre EGs e CBHs (JALAK et al., 2012) e outra exo-exo entre duas CBHs, uma atuando na extremidade redutora e outra atuando na extremidade não redutora (WANG et al., 2013). Assim, endoglucanases clivam internamente as cadeias de celulose, liberando unidades a serem degradadas por CBHs e/ou BGs. As CBHs clivam em unidades de celobiose atuando nas extremidades da cadeia de celulose, isto é, dependendo do tipo da enzima esta pode atuar uma na extremidade redutora e outra na extremidade não redutora (ARO et al., 2005). Finalmente, as BGs hidrolisam as unidades de celobiose em glicose, uma fonte de carbono monomérica prontamente metabolizável para o fungo (BÉGUIN, 1990). Essa mesma enzima pode converter celobiose em soforose, fenômeno denominado de transglicosilação (EL-GOGARY et al., 1989). A ligação covalente intermediária formada nas reações catalisadas por BGs é comum para iniciar duas rotas divergentes: hidrólise e transglicosilação. O ataque de água neste complexo intermediário caracteriza a hidrólise, ao passo que o ataque por um

segundo substrato (chamado aceitador neste caso) define o percurso para transglicosilação (Figura 16) (TEUGJAS; VÄLJAMÄE, 2013; FRUTUOSO; MARANA, 2013).

Essas três classes de enzimas precisam agir sinergicamente e sequencialmente em ordem para degradar completamente a matriz de celulose. Porém, participam também as proteínas expansinas (*swo1*) que, rapidamente, induz o afrouxamento da parede celular da planta, enfraquecendo as interações não covalentes que ajudam a manter a sua integridade. Enzimas da família GH61 agem em sinergia com outras enzimas celulolíticas. Desde então, diversos estudos (HORN et al., 2012; DIMAROGONA et al., 2012; HÄKKINEN et al., 2012) têm demonstrado que GH61 são enzimas oxidativas (PMOs) que agem em sinergia com celobiose desidrogenase (CDH). Esta última é um potencial doador de elétrons para as monooxigenases (PMOs) que traz um novo modelo de degradação da celulose. CDH podem utilizar uma molécula de celobiose como receptor e, consequentemente, oxida-a. Os elétrons liberados podem ser utilizados por GH61/PMOs para despolimerizar a celulose cristalina através de eliminação redutiva. Após ação das GH61, destaca-se a ação de EGs e PMOs que clivam internamente as cadeias de celulose liberando novas extremidades da cadeia que tornam alvos para celobiohidrolases (VANHOLME et al., 2013). Uma visão esquemática da degradação da celulose é representada na Figura 17.

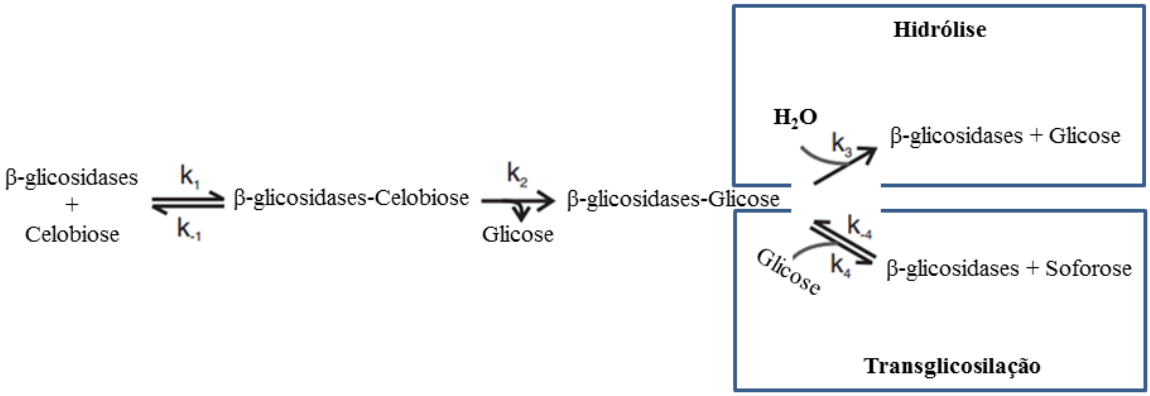


FIGURA 16 – Representação esquemática da hidrólise e transglicosilação (reações catalisadas por β-glicosidases).
Celobiose (ligação glicosídica β(1→4); Soforose (ligação glicosídica β(1→2).
Fonte: Adaptado de (TEUGJAS; VÄLJAMÄE, 2013; FRUTUOSO; MARANA, 2013).

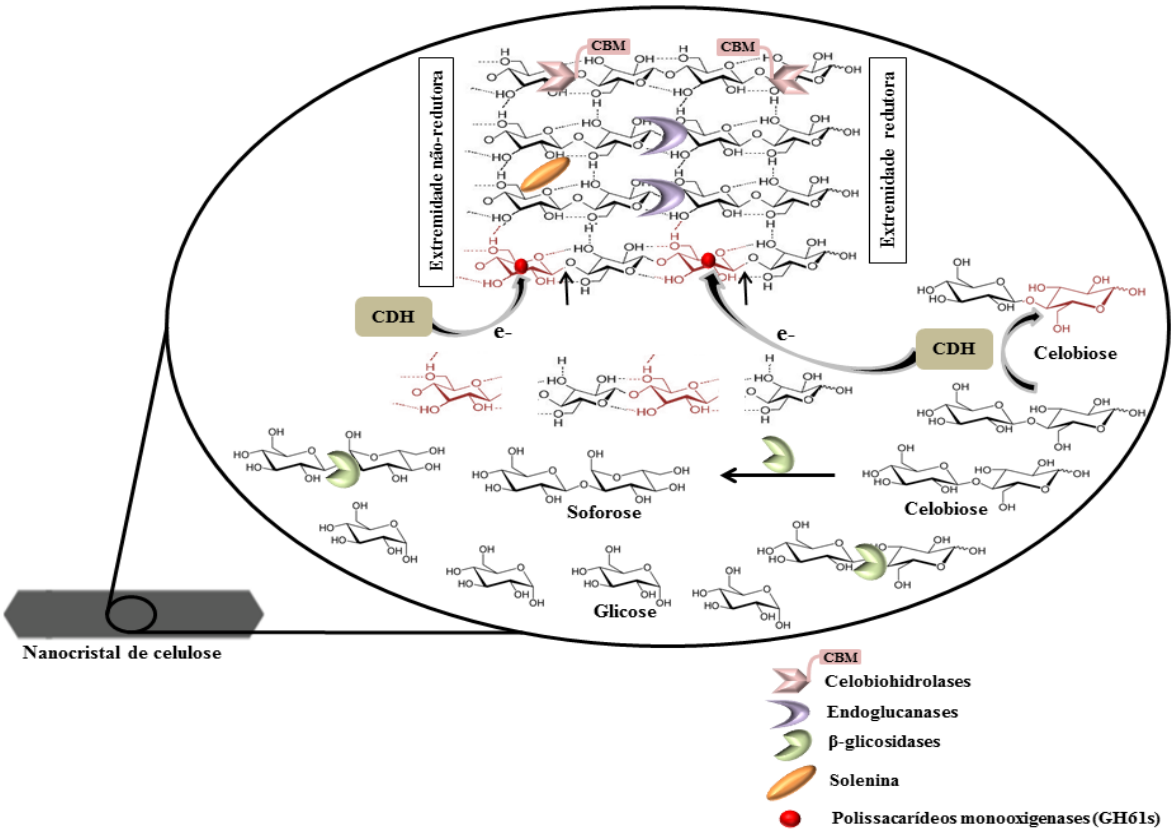


FIGURA 17 – Um esquema simplificado da visão atual sobre a degradação enzimática da celulose.
Estruturas em vermelho identificam as glicoses que foram oxidadas. A possível consequência da ação GH61 é ilustrada, onde novos pontos de ataque para celobiohidrolases estão indicadas por setas. A maioria das enzimas é acoplada aos Módulos de Ligação a Carboidrato (CBMs), que não têm atividade catalítica, mas assegura a ligação do domínio catalítico no polissacarídeo. Adicionalmente, uma rota para a formação de soforose por transglicosilação realizada pela β-glicosidase é apresentada.
Fonte: Adaptado de (HORN et al., 2012; VANHOLME et al., 2013).

Como descrito anteriormente, a hemicelulose é uma matriz de polissacarídeo complexo composto por diferentes grupos pendentes de sua cadeia principal, cadeia esta que pode ser de três tipos: xilana, glucana e manana. Para a complexidade de hemicelulose, a hidrólise/degradação da mesma requer uma ação combinada de endo-enzimas que clivam a parte interna da cadeia principal, exo-enzimas liberam açúcares monoméricos, e enzimas acessórias (Tabela 5) clivam polímeros das cadeias laterais ou oligossacarídeos associados, levando a liberação de vários monossacarídeos ou dissacarídeos dependendo do tipo de hemicelulose (SOUZA, 2013b). Como a hemicelulose pode variar em sua composição de acordo com a fonte lignocelulósica, a Figura 18 representa apenas a hemicelulose que possui, em sua cadeia principal, a D-xilose (xilana), que será também o alvo para a produção de etanol no Brasil, e, portanto, as setas representam cada enzima ativa para um substrato determinado.

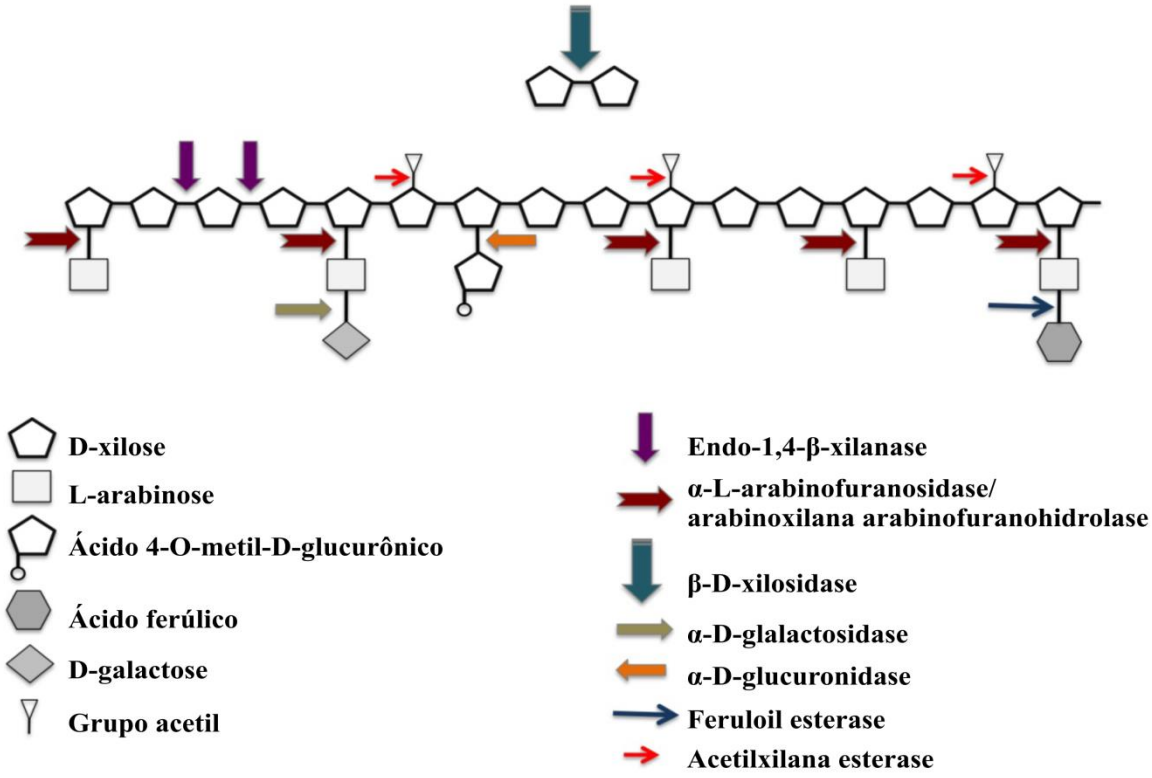


FIGURA 18 – Visão esquemática da degradação de um sistema hemicelulolítico.
Fonte: Adaptado de (SOUZA, 2013).

Tabela 5 – Enzimas acessórias fúngicas para a clivagem dos resíduos derivados de hemicelulose.

Polímero de Hemicelulose	Resíduo liberado	Enzima
Xiloglicana/xilana	L-arabinose	α-arabinofuranosidases
		arabinoxilana arabinofuranohidrolases
Xiloglicana	D-xilose	α-xilosidases
Xiloglicana	L-fucose	α-fucosidases
Xilana/galactomananas	D-galactose	α-galactosidases
Xilana	Ácido D-glucurônico	α-glucuronidases
Xilana	Grupo acetil	acetil xilana esterases
Xilana	ácido p-coumárico	p-coumaroil esterases
Xilana	Ácido ferúlico	feruloil esterases

Fonte: Adaptado de (SOUZA, 2013b).

2.3.4 Regulação da expressão de genes de celulases e hemicelulases

Desde a década de 1960, Mary Mandels e Elwyn T. Reese levantaram o seguinte paradigma sobre como ocorre a indução da formação de celulases pelo fungo *T. reesei*: "Celulases são enzimas adaptativas, mas seu substrato natural, celulose, é insolúvel. Então como ocorre a indução? (MANDELS et al., 1962). Muitos estudos têm sido realizados numa tentativa de descobrir o indutor natural da formação de celulases (GRITZALI; BROWN, 1979; KUBICEK et al., 1988; SEIBOTH et al., 1992; KUBICEK et al., 1993; ILMÉN et al., 1997; SEIBOTH et al., 1997; FOREMAN et al., 2003). No entanto, Suto e Tomita (2001, p307) propuseram um modelo de indução em *T. reesei* (Figura 19):

> Assim que esporos dos conídios entram em contato com a celulose, a celobiohidrolase presente na superfície do esporo degrada a celulose produzindo celo-oligossacarídeos. Estes são então hidrolisados até glicose e transglicosilados pela ação de uma β-glicosidase, também constitutiva, formando assim a soforose (1). Segue-se então a germinação dos esporos dos conídios. A glicose como fonte de carbono e a soforose como um indutor entra então na célula (2). As celulases são induzidas por soforose e secretadas para o meio extracelular (3). As celulases então degradam a celulose, promovendo um aumento nas quantidades de celo-oligossacarídeos e glicose (4). A glicose é assimilada. Celo-oligossacarídeos são hidrolisados a glicose e transglicosilados para formar soforose novamente (5) reiniciando o passo (3). Este ciclo continua até que a celulose esteja esgotada do meio (Figura 19).

Zhou e colaboradores (2012) demonstraram que uma percepção adequada de celulose, no exterior da célula, transforma em um sinal intracelular e garante a produção rápida de celulases e que a participação de β-glicosidases (*cel3a* – extracelular; *cel1a* e *cel1b* – intracelulares) são de fundamental importância para uma rápida e eficiente indução de genes de celulases, isto é, através da indução na presença de celulose, celobiose é liberada por ação sinérgica de exoglucanases e endoglucanases, e, em seguida, é transportada intracelularmente para iniciar a indução transcricional ou será hidrolisada por β-glicosidases intra- ou intercelulares em glicose. Qualquer que seja o modelo que, de fato, seja responsável pelo ataque inicial da molécula de celulose, o produto final deve ser um indutor de baixa massa molecular que é capturado pelas células do fungo. A identidade desse composto ainda não é conhecida, porém diversos oligossacarídeos podem funcionar como indutores de celulases em experimentos laboratoriais (SILVA, 2008).

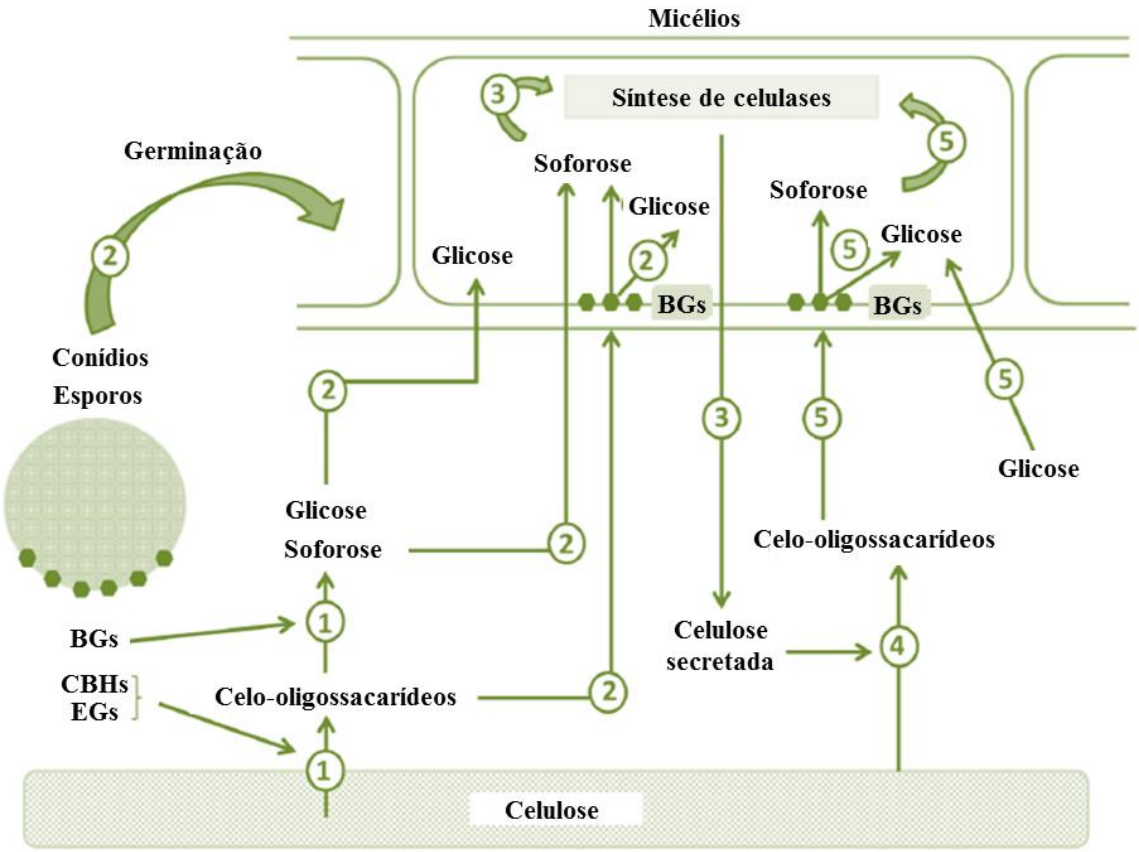


FIGURA 19 – Modelo de indução de celulases em *T. reesei*.
Os números de 1 a 5 correspondem as etapas já explicadas no texto.
BGs: β-glicosidases; CBHs: celobiohidrolases e EGs: endoglucanases.
Fonte: (SUTO; TOMITA, 2001).

Consequentemente, a produção de enzimas é coordenadamente regulada e induzida na presença de substratos, isto é, existem fontes de carbono diferentes para induzir expressão de genes que codificam as celulases e hemicelulases (Tabela 6), como celulose e outros oligossacarídeos (soforose, 1,5-lactona-δ-celobiono[5], D-xilose, xilobiose, galactose e lactose) (SCHMOLL; KUBICEK, 2003; ARO et al., 2005; KARAFFA et al., 2006; MORIKAWA et al., 1995; STRICKER et al., 2006; STRICKER et al., 2007). Contudo, os mecanismos de indução da expressão de genes de celulases e hemicelulases envolvem a ativação da expressão de genes pelas respectivas hidrólises e/ou produtos de transglicosilação de celulose e/ou xilana, como exemplo, a soforose em *Aspergillus terreus* e *T.reesei* (HRMOVÁ et al., 1991; MANDELS et al., 1962). Soforose é um forte indutor de celulases e é considerado um possível indutor natural, isto é, como relatado anteriormente, a soforose é formado por *T.*

[5] Análogo de celobiose (SZAKMARY et al., 1991).

reesei durante a hidrólise da celulose por uma reação de transglicosilação (EL-GOGARY et al., 1989).

Tabela 6 – Indutibilidade de celulases e hemicelulases em *T. reesei*.

Substratos / Enzimas	Glicerol	Glicose	Sorbitol	Celulose	Celobiose	Xilose	Soforose	Lactose	Xilobiose	Galactose	Xilana	Arabinose	Arabinitol	Sorbose	Referências
bgl1					+	+	+		+						(ARO et al., 2005)
bgl2						+	+		+						(STRICKER et al., 2006)
bxl1						+	+		+						(STRICKER et al., 2006)
cbh1	-		-	+	+	+	+	+			+				(CARLE-URIOSTE et al., 1997; ILMÉN et al., 1997)
cbh2	-		-	+	+	+	+		+						(ILMÉN et al., 1997)
cel5b	+	+					+	+							(FOREMAN et al., 2003)
egl1	-		-	+			+								(CARLE-URIOSTE et al., 1997; ILMÉN et al., 1997)
egl2	-		-	+			+								(ILMÉN et al., 1997)
egl3							+							+	(NOGAWA et al., 2001)
egl5	-		-	+			+								(ILMÉN et al., 1997)
xynI		x			+	+	+		+		+			+	(NOGAWA et al., 2001)
xynII		x			+	+	+		+		+			+	(NOGAWA et al., 2001)
xyn3						-	+							+	(NOGAWA et al., 2001; XU et al., 2000)
a-AF														+	(NOGAWA et al., 2001)
abf1		x										+	+		(AKEL et al., 2009)
abf2		x										+	+		(AKEL et al., 2009)
abf3		x										+	+		(AKEL et al., 2009)
agl1									+	+					(ARO et al., 2005)
agl2									+	+					(ARO et al., 2005)
glr1									+						(ARO et al., 2005)
axe1										+					(ARO et al., 2005)

Repressor (x); indutor (+); nenhuma ação (-). *bgl1/cel3a*: β-glicosidase 1; *bgl2/cel1a*: β-glicosidase 2; *bxl1*: β-xilosidase 1; *cbh1/cel7A*: celobiohidrolases 1; *cbh2/cel6A*: celobiohidrolases 2; *cel5b*: endoglucanase; *egl1/cel7b*: endoglucanase 1; *egl2/cel5a*: endoglucanase 2; *egl3/cel12a*: endoglucanase 3; *egl5/cel45a*: endoglucanase 5; *xynI*: xilanase 1; *xynII*: xilanase 2; *xyn3*: xilanase 3; *a-AF*: alfa-L-arabinofuranosidase; *abf1*: arabinofuranosidase 1; *abf2*: arabinofuranosidase 2; *abf3*: arabinofuranosidase 3; *agl1*: alfa-galactosidase 1; *agl2*: alfa-galactosidase 2; *glr1*: α-glucuronidase 1 e *axe1*: Acetilxilana esterase 1.
Fonte: Adaptado de (AMORE et al., 2013).

Em adição à indução de genes por mono e oligossacarídeos, que codificam enzimas hidrolíticas envolvidas em desconstrução da parede celular da planta em fungos filamentosos pode ser reprimida durante o crescimento na presença de fontes de carbono facilmente metabolizáveis, tais como: a glicose e frutose (CHAMBERGO et al., 2002). Na presença de

glicose, transporte de indutores, tais como soforose nos micélios é inibida (ZEILINGER et al., 2003). A disponibilidade de distintas fontes de carbono representa mais uma variável ambiental na qual os micro-organismos devem se adaptar. Um dos mecanismos regulatórios nesta adaptação é a repressão por carbono. Nesse mecanismo, a presença de fontes de carbono, rapidamente metabolizáveis, promove a repressão da síntese de enzimas relacionadas ao catabolismo de fontes alternativas de carbono, como a celulose. Segundo Ruijter e Visser (1997), sob o ponto de vista da fisiologia celular, a existência de tal mecanismo regulatório é vantajosa na medida em que a fonte de carbono energeticamente mais favorável é consumida, não havendo gasto desnecessário de energia na síntese de enzimas que degradam substratos complexos. Esses mesmos autores ressaltam que esse mecanismo regulatório pode ser desencadeado por diversas fontes de carbono, mas como a glicose é uma das fontes mais repressora, o termo "repressão por glicose" também é utilizado, embora se refira a um caso específico da repressão por carbono ou pode-se referir "repressão catabólica por carbono", com a sigla CCR na língua inglesa.

Assim, a presença de celulose, xilana e misturas de polímeros de planta presentes no meio de cultura do fungo provoca a produção de atividades celulolíticas e xilanolíticas por *T. reesei*, como relatado em estudos anteriores (BISARIA; MISHRA, 1989; MACH; ZEILINGER, 2003; OLSSON et al., 2003; JUHÁSZ et al., 2005; STRICKER et al., 2008; BISCHOF et al., 2013). Portanto, a regulação da expressão gênica de celulases e hemicelulases ocorre em nível transcricional de maneira coordenada e é dependente da presença do indutor (KUBICEK, 2013). Ao longo das últimas décadas, esforços contínuos foram realizados na identificação de genes envolvidos na regulação transcricional da expressão do gene da celulases e hemicelulases. São conhecidos cinco reguladores positivos (XYR1 (STRICKER et al., 2006), ACE2 (ARO et al., 2001), complexo HAP2/3/5 (ZEILINGER et al., 2001), BglR (NITTA et al., 2012) e ACE3 (HÄKKINEN et al., 2014) e dois repressores (ACE1 (SALOHEIMO, 2000; ARO et al., 2003) e o repressor catabólico por carbono CRE1 (ILMÉN et al., 1996)), todos envolvidos na expressão de genes de interesse biotecnológico em *T. reesei* (Figura 20) (KUBICEK et al., 2009; PORTNOY et al., 2011a; NITTA et al., 2012; HÄKKINEN et al., 2014).

Os fungos filamentosos *Aspergillus niger* e *T. reesei* foram objetos de muitos estudos que investigaram os mecanismos de regulação da transcrição de genes de celulases e hemicelulases (STRICKER, A. et al., 2008). O regulador transcricional XlnR, que foi

inicialmente identificado em *A. niger*, controla a transcrição de cerca de 20-30 genes que codificam as celulases e hemicelulases. O ortólogo *xyr1* tem uma função similar a XlnR, embora, em alguns pontos, os mecanismos parecem ser diferentes (RAUSCHER et al., 2006). O principal regulador positivo na expressão dos genes de celulases e hemicelulases em *T. reesei* é representado por XYR1 (Regulador de xilanase 1) (STRICKER et al., 2006; STRICKER et al., 2008), que possui um domínio bi-nuclear de zinco de ligação ao DNA podendo ligar-se a vários motivos 5'-GGC(A//T)$_3$-3' (STRICKER et al., 2006). É responsável por ativar a transcrição de genes que codificam hidrolases, tais como: *xyn1*, *xyn2*, *bxl1*, *cbh1*, *cbh2*, *egl1* e *bgl1*. E a deleção de *xyr1* elimina a indução de celulases em presença de celulose e soforose como fontes de carbono para o fungo e prejudica a indução de genes de hemicelulases envolvidos na degradação de xilana e arabinana (STRICKER et al., 2008; AKEL et al., 2009), provando, assim, o seu papel essencial no processo de indução.

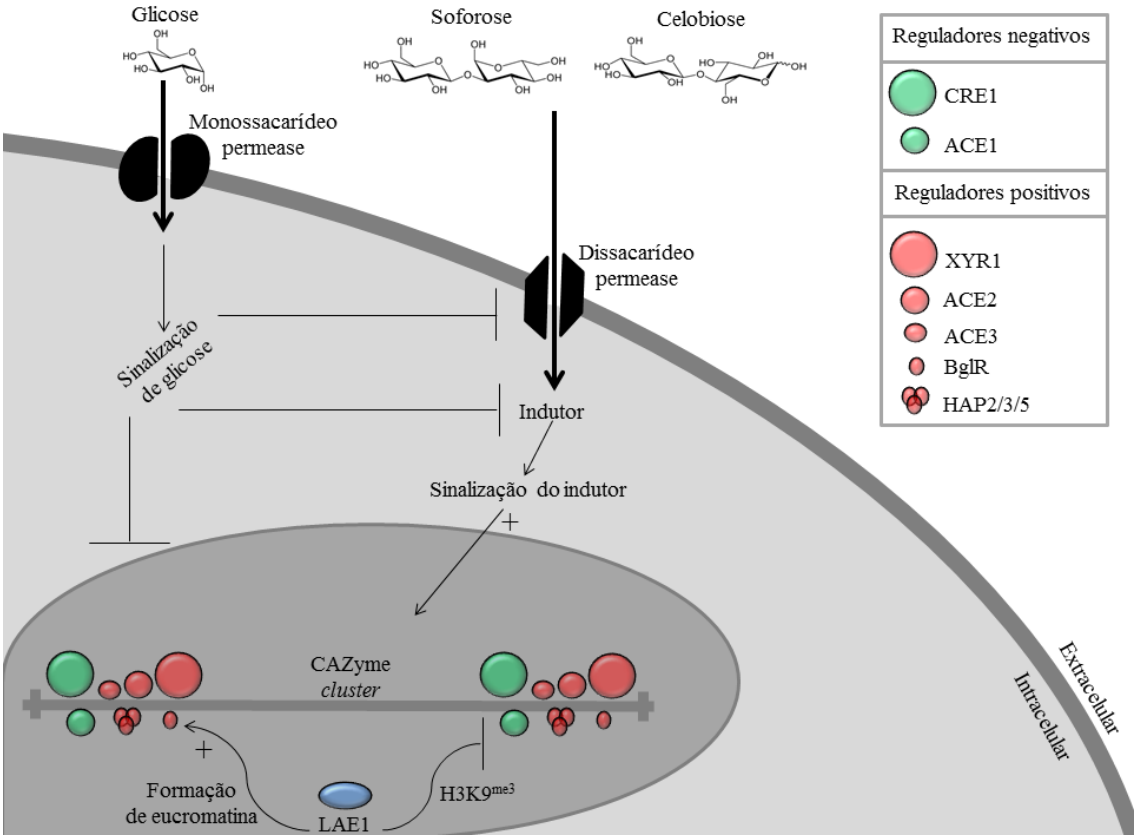


FIGURA 20 – Representação esquemática dos fatores de transcrição que afetam expressão de celulases e hemicelulases em *T. reesei*.
LAE1 (proteína metiltransferase); H3K9^{me3} (Histona H3 trimetilada na lisina 9 – metilação).
Fonte: Adaptado de (SEIBOTH et al., 2011); BglR (NITTA et al., 2012); ACE3 (HÄKKINEN et al ., 2014); LAE1 (SEIBOTH et al., 2012; KARIMI-AGHCHEH et al., 2013; FEKETE et al., 2014).

É sabido, ainda, que a transcrição de *xyr1* não é induzida durante o crescimento em celulose (MACH-AIGNER et al., 2008). A maioria dos ativadores de transcrição eucarióticos estão presentes nas células apenas em pequenas quantidades necessárias para iniciar a expressão do gene, e, em muitos casos, eles são ainda induzidos pelas condições para as quais são necessários e depois são degradados, uma vez que não são mais imperativos (TANSEY, 2001). Pelo contrário, a expressão de *xyr1* é regulada por CRE1 (dependente de CCR) (ANTONIÊTO et al., 2014) e pela repressão pelo fator de transcrição ACE1 (MACH-AIGNER et al., 2008; STRICKER et al., 2007).

A repressão na presença de glicose na produção de genes de celulases e hemicelulases é regulada por *cre1*, e sítios de ligação para este fator é encontrado nas regiões promotoras dos genes *cbh1*, *cbh2*, *eg1* e *eg2*, indicando que a regulação de tais genes ocorre por CCR. A remoção de CCR foi uma das primeiras tentativas da engenharia de produção em *T. reesei*, como exemplo o mutante RUT-C30 que foi obtido a partir do *T. reesei* QM6a que contém o gene *cre1* truncado, e é, portanto, capaz de produzir celulases em presença de glicose (PETERSON; NEVALAINEN, 2012). Esse fator, consequentemente, reprime tanto diretamente como indiretamente genes envolvidos na degradação de polissacárido, assim como genes envolvidos na utilização de aminoácidos e de etanol como potencial fontes de carbono (RUIJTER; VISSER, 1997; FELENBOK et al., 2001; PORTNOY et al., 2011b; ANTONIÊTO et al., 2014). CRE1 está envolvido na regulação do principal fator de transcrição XYR1 reprimindo a transcrição do gene *xyr1* na presença de altas concentrações de D-xilose, por conseguinte, também influencia indiretamente transcrições de genes de hemicelulases (MACH-AIGNER et al., 2010). Além disso, CRE1 também desempenha um papel na regulação do crescimento de fungos em substratos sólidos e a formação de hifas aéreas, conídios e colônias fúngicas (NAKARI-SETÄLÄ et al., 2009). Além de estar envolvido na captação de indutor e na sua formação, na repressão direta de celulases e na expressão de genes ativadores de transcrição (Figura 20).

O ativador transcricional ACE2 regula também a expressão de genes de celulases e hemicelulases em *T. reesei* e tem sido mostrado ocorrer apenas em *Trichoderma* spp. Foi verificado que é um fator que se liga e ativa a região promotora da principal celulase (*cbh1*) e a deleção do gene *ace2* leva a uma redução na indução da expressão de celulases na presença de celulose, enquanto que, em presença de soforose, não tem afetado a indução de celulases (ARO et al., 2001; STRICKER et al., 2008). Hap2/3/5 é um outro ativador transcricional que

se liga a região promotora do gene *cbh2* (ZEILINGER et al., 2001). Evidencias também foram apresentadas que esse complexo se liga no elemento ativador de xilanase (XAE) (ZEILINGER et al., 1996; ZEILINGER et al., 2001) e, por conseguinte, está envolvido na regulação da *xyn2* (WÜRLEITNER et al., 2003). O seu papel na regulação da expressão de genes de celulases e hemicelulases pode ser um potencializador geral para aumentar a acessibilidade de outros fatores, tais como XYR1 nas regiões promotoras de celulases (KUBICEK et al., 2009). Ao contrário dos três ativadores, ACE1 é um repressor, e a deleção deste gene resulta em um aumento na expressão de todas os principais genes de celulases e dois genes de xilanases em culturas contendo os indutores celulose e soforose, indicando que ACE1 atua como um repressor na expressão de celulases e xilanases (ARO et al., 2003) e de *xyr1* durante o crescimento em D-xilose (MACH-AIGNER et al., 2008).

Dois novos fatores foram identificados em *T. reesei*, o primeiro em 2012, conhecido como BglR (proteína ID: 52368, regulador de β-glicosidase) que foi identificado como um novo fator de transcrição que regula a expressão de genes específicos que codificam β-glicosidases. Uma das funções de BglR é regular positivamente os genes específicos β-glicosidase (com a exceção de *bgl1*, que aparentemente, está sob o controle direto de XYR1) (NITTA et al., 2012). O segundo fator foi identificado por Häkkinen e colaboradores em 2014, nomeado como ACE3 (proteína ID: 77513, ativador da expressão de celulase 3). No estudo realizado por esses autores, ficou claro que a deleção desse gene aboliu a produção e a expressão de vários genes de celulases estudados. Essa deleção também reduziu significativamente a atividade de xilanases e a expressão de genes de enzimas que degradam a xilana.

Seiboth e coautores (2012) demonstraram que uma proteína ortóloga conhecida como LAE1 em *T. reesei* também controla a expressão de genes envolvidos na degradação lignocelulósica através de mecanismos que ainda não foram totalmente elucidados. Experimentos confirmaram que LAE1 modula a expressão de genes de celulases e que foi dependente do principal regulador positivo XYR1, mas também a expressão de *xyr1* foi dependente de LAE1. Os autores ainda afirmaram que essa proteína também foi fundamental na conidiação de *T. reesei*. Em adição a expressão de grupos de *CAZymes* são também reguladas em nível de cromatina por LAE1, com potencial alvo na formação de eucromatina e na trimetilação da H3K9 (necessário para a formação de heterocromatina em fungos) (Figura 20) (SEIBOTH et al., 2011). Porém, LAE1 não afeta a expressão de *CAZymes* por modular

diretamente a trimetilação da H3K9, sugerindo que o efeito sobre a expressão de genes é um processo indireto (SEIBOTH et al., 2012).

Com esses dados, existe uma regulação na expressão genética que define os genes que serão ativados/reprimidos em um determinado momento. Qual/quais o/os mecanismo(s) envolvido(s) na produção de enzimas de interesse biotecnológico/industrial quando o fungo filamentoso *T. reesei* crescer em fontes de carbono envolvendo a biomassa lignocelulósica, isto é, em presença do polissacarídeo mais complexo (celulose) e outras fontes de carbono, relacionados como produtos da degradação da biomassa (soforose e glicose)? Compreender como o fungo se comporta em presença dessas fontes é entender como o fungo produz as enzimas, e, desta forma, será uma etapa fundamental para o desenvolvimento de linhagens mutantes para, futuramente, comercializar coquetéis enzimáticos destinados ao processamento da biomassa oriunda da cana-de-açúcar para a produção de etanol 2G.

3 OBJETIVOS

3.1 Objetivo geral

Analisar o perfil global de expressão gênica de *Trichoderma reesei* em condições de indução (celulose, soforose) e repressão (glicose) da produção de celulases nas linhagens parental QM9414 e mutante *Δxyr1* para entender como as enzimas de interesse biotecnológico são produzidas por esse micro-organismo.

3.2 Objetivos específicos

3.2.1 Construção de bibliotecas de cDNA nas condições indutoras (celulose e soforose) e na condição repressora (glicose);

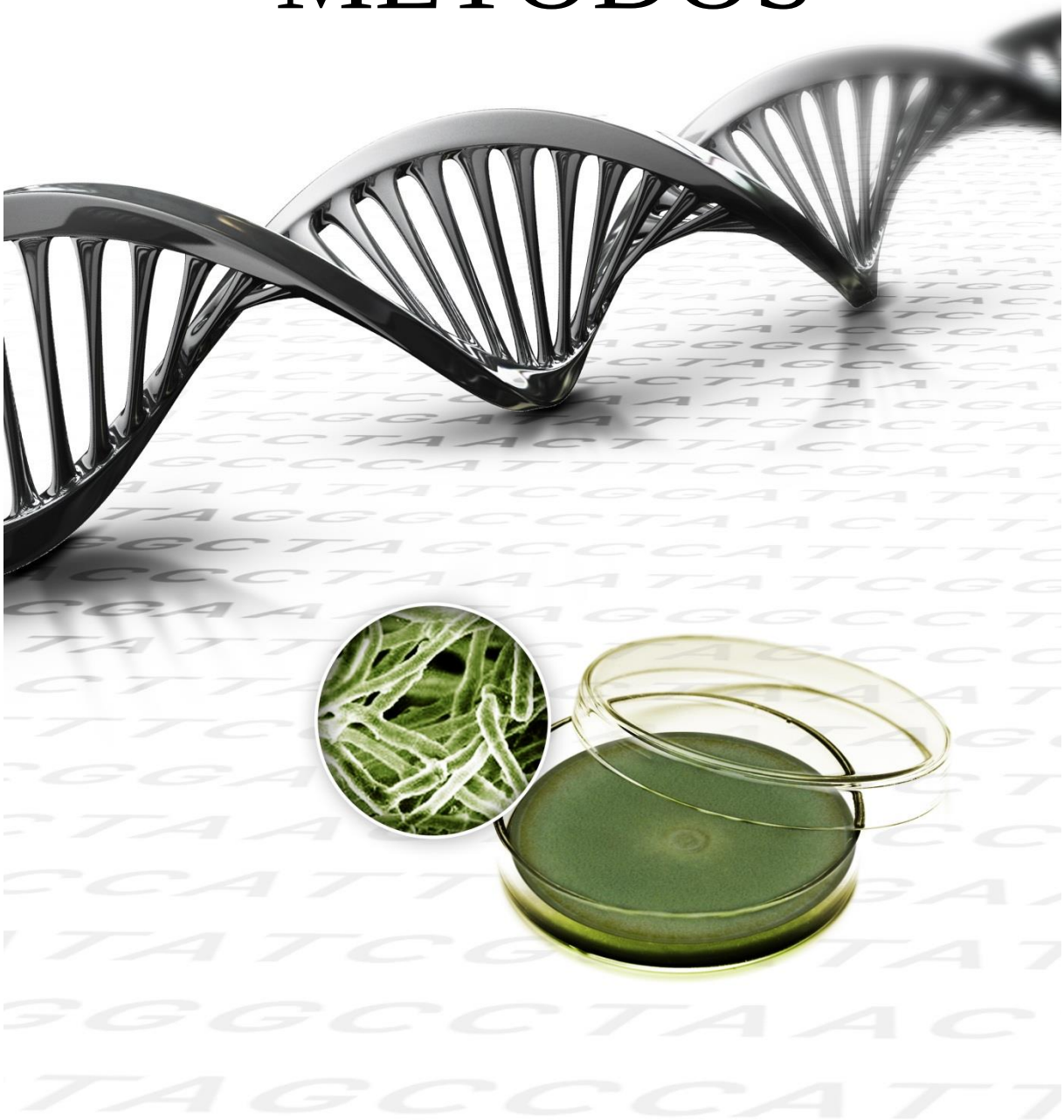
3.2.2 Sequenciamento das bibliotecas de cDNA utilizando a Plataforma *Illumina HiSeq*^{TM} *2000*;

3.2.3 Análise global da regulação da expressão gênica de *T. reesei* quando cultivado na presença dos indutores e do repressor por meio de ferramentas de Bioinformática;

3.2.4 Analisar os mecanismos de formação de celulases pelo fungo *T. reesei* nas diferentes condições estudadas;

3.2.5 Validação de genes diferencialmente expressos por PCR em tempo real (qPCR-TR).

MATERIAIS E MÉTODOS



4 MATERIAIS E MÉTODOS

4.1 Delineamento experimental

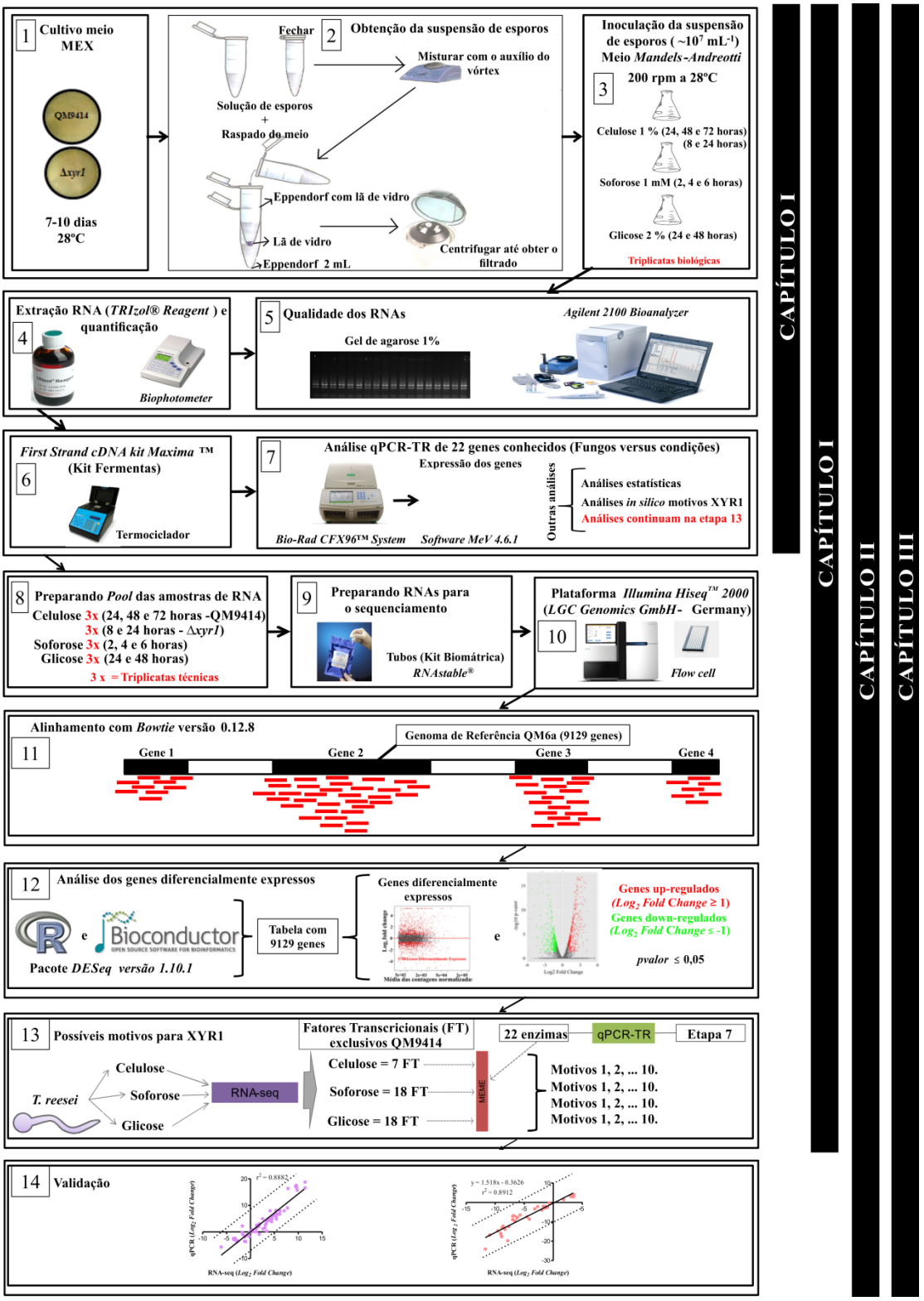


FIGURA 21 – Esquema das etapas experimentais realizadas neste estudo.

Na Figura 21 é ilustrado o esquema geral das etapas experimentais utilizadas para o desenvolvimento desta tese.

4.2 Linhagens fúngicas

Foram utilizadas duas linhagens de *Trichoderma reesei*: QM9414 (ATCC 26921 – linhagem parental) e a linhagem mutante *Δxyr1* (STRICKER et al., 2006), obtidas junto ao"*Molecular Biotechnology Laboratory, Institute*, TU Vienna, Austria".

4.3 Condições de crescimento

4.3.1 Meio MEX

O meio MEX foi utilizado para a manutenção das linhagens fúngicas no Laboratório de Biotecnologia Molecular (LBM – FMRP/USP) e foi preparado com extrato de malte 3% (p/v) e ágar 2% (p/v) diluídos com água destilada e, em seguida, autoclavado a 1 atm por 20 minutos. As linhagens QM9414 e *Δxyr1* foram cultivadas a 28^{o}C por 7-10 dias para completa esporulação.

4.3.2 Meio *Mendels-Andreotti*

O meio *Mendels-Andreotti* foi preparado conforme as instruções presentes na Tabela 7.

Tabela 7 – Preparo do meio *Mendels-Andreotti* (1L).

1 - Solução mineral (500 mL) – 50%:
2,8 g/L $(NH_4)_2SO_4$
4,0 g/L KH_2PO_4
0,6 g/L $MgSO_4.7H_2O$
0,8 g/L $CaCl_2.2H_2O$
2 - 0,1 M tampão Citrato-fosfato pH 5,0 (480 mL) – 48%:
0,2 M $Na_2HPO_4.2H_2O$
0,2 M ácido cítrico
3 - 50X Solução elementos-traços[6] (20 mL) – 2%:
250 mg/L $FeSO_4.7H_2O$
85 mg/L $MnSO_4.H_2O$
70 mg/L $ZnSO_4.7H_2O$
100 mg/L $CoCl_2.2H_2O$
Ajustar o pH para 2.0 com H_2SO_4
As soluções 1, 2 e 3 foram preparadas separadamente de acordo com a proporção de meio desejada. Em outro frasco misturou-se as soluções 1, 2 e 3 e acrescentaram-se 1 g/L de peptona bacteriológica e 0,3 g/L de ureia para depois acrescentar 200 mL deste meio nos fracos *Erlenmeyer* 1000 mL contendo celulose 1%, e em seguida foram autoclavados. Nos cultivos em presença de glicose apenas 180 mL do meio *Mandels-Andreotti* foram preparados e acrescentados nos fracos *Erlenmeyer* 1000 mL e autoclavados e, somente depois, foram adicionados 20 mL de uma solução de glicose 20% (correspondendo a 2% do meio) esterilizada em membrana de filtro de 20 μm. Para o cultivo em presença de soforose utilizou-se um frasco *Erlenmeyer* 1000 mL com glicerol 1% e autoclavou-se o meio. O fungo foi previamente crescido por 24 horas nesta fonte de carbono neutra e depois deste tempo o micélio foi filtrado e lavado com meio *Mandels-Andreotti* sem o acréscimo da peptona para então ser transferido para um *Erlenmeyer* de 125 mL contendo 20 mL do meio *Mandels-Andreotti* sem peptona autoclavado e contendo 1 mM de soforose adicionado posteriormente. Estes procedimentos de cultivos na presença dessas três fontes de carbono foram realizados em triplicatas biológicas.

Fonte: (SCHMOLL et al., 2009).

Após o preparo desse meio, as placas de QM9414 e *Δxyr1* cultivadas em meio MEX foram utilizadas para obtenção da suspensão de esporos. A solução de esporos foi preparada com 0,8% de NaCl e *Tween* 80 0,05% diluídos em água ultra pura e a solução foi esterilizada utilizando um filtro (membrana de 0,22 μm) e mantida a 4ºC.

[6] O termo elementos-traços têm sido usado para definir metais catiônicos e oxiânicos presentes em baixas concentrações (usualmente < 0,1%) no meio ambiente e seres vivos. Esse termo tem sido utilizado como substituto de "metais pesados". Alguns elementos-traços são considerados essenciais do ponto de vista biológico, enquanto outros não são, e muitos são essenciais ao metabolismo fúngico (DIAS, 2013).

Aproximadamente, 10^7 mL^{-1} de esporos das linhagens QM9144 e *Δxyr1* previamente cultivadas em meio MEX foram inoculados em de meio *Mendels-Andreotti* com as respectivas fontes de carbono e, em seguida, foram incubados em *shaker* orbital em 200 rpm e a 28ºC por 24, 48 e 72 horas (celulose linhagem parental QM9414), 8 e 24 horas (celulose linhagem mutante *Δxyr1*), 24 e 48 horas (glicose) e 2, 4 e 6 horas (soforose). As amostras de massa micelial das culturas QM9414 e *Δxyr1* foram coletadas para serem extraídas o RNA total nos respectivos tempos de cultivos. Essas massas miceliais foram coletadas por filtração em *miracloth* e rapidamente embrulhadas em papel alumínio devidamente nomeadas e congeladas em nitrogênio líquido. Posteriormente, armazenadas no freezer -80ºC.

4.4 Extração de RNA e verificação de sua integridade

Os RNAs totais foram extraídos dos micélios de cada amostra, usando o kit de extração de *RNA TRIzol*® (*Invitrogen Life Technologies, Carlsbad, CA, USA*), de acordo com as instruções do fabricante. As quantificações foram determinadas no *BioPhotometer Eppendorf* e o grau de pureza foi determinado pela relação de absorbância 260/280. A integridade destes RNAs foi verificada tanto com o *Aligent 2100 Bioanalyzer* (*Aligent Technologies, Waldbroon, Germany*) como pela utilização do gel de eletroforese em agarose 1%.

Para a realização das quantificações dos RNAs, utilizou-se uma diluição 30 vezes e a quantificação do RNA foi realizada no *Biophotometer Eppendorf* para determinar a razão 260/280, que indica o grau de pureza da amostra (valores próximo de 2,0 para ácidos nucléicos). Para cada tubo de RNA extraído a quantificação foi realizada em duplicatas.

4.5 Remoção do ácido desoxirribonucléico e síntese da primeira fita de cDNA

Após a obtenção do RNA total, para cada tempo de cultivo, o mesmo foi tratado com DNAse-I livre de RNAse (*Thermo Scientific*) para remover todas possíveis contaminações com DNA genômico. O seguinte protocolo foi utilizado:

RNA	1µg
10X tampão com $MgCl_2$	1µL
DNase-I, livre de RNase	1µL
Água DEPC[7]	q.s.p. 10µL
Volume total	10µL

Incubou a 37ºC por 30 minutos utilizando. Adicionou 1µL de EDTA 50mM e incubou a 65ºC por 10 minutos..

A primeira fita de cDNA foi amplificada, sendo utilizada a enzima transcriptase reversa (*Thermo Scientific*) a partir do RNA tratado com DNase-I, seguindo as etapas:

5X Reaction Mix	4µL
Maxima®Enzyme Mix	2µL
RNA da aula prática 11	11µL
Água DEPC	q.s.p. 20µL
Volume Total	20µL

Misturou suavemente e incubou por 10 minutos a 25ºC seguido por 15 minutos a 50ºC. Terminou a reação por aquecimento a 85ºC por 5 minutos. O produto da síntese da primeira fita de cDNA foi utilizado diretamente para Reação em Cadeia da Polimerase em Tempo Real (qPCR-TR) ou foi estocado a -20ºC por uma semana ou estocado por um longo período a -80ºC. Congelamento e descongelamento de cDNA foram evitados.

4.6 Análises de transcritos por PCR quantitativo (qPCR-RT) e análises estatísticas

Para verificar o perfil de expressão gênica nas diversas condições de cultivo, isto é, celulose 24, 48 e 72 horas na linhagem parental QM9414, celulose 8 e 24 horas na linhagem mutante *Δxyr1*, soforose 2, 4 e 6 horas e glicose 24 e 48 horas para as duas linhagens, primeiramente os cDNAs foram diluídos 1/50 vezes e utilizados para PCR em Tempo Real (qPCR-TR) no *CFX96™ System Bio-Rad*, utilizando para detecção de sinal *SsoFast™*

[7] É essencial que toda água usada diretamente ou nos tampões seja tratada com dietilpirocarbonato (DEPC) para inativar as RNases (HIRATA et al., 2006). A água DEPC foi preparada a 0,1% (v/v) e incubada a 37°C por 18-20 horas e autoclavada por 30 min, a 1 atm de pressão e 120°C.

EvaGreen® *Supermix* (*Bio-Rad*) de acordo com as instruções do fabricante. Cada reação, realizada em um volume final de 20μL, continha 10 μL de *EvaGreen* e 500 nM de cada primer (*forward* e *reverse*), cDNA e água livre de RNAse. As seguintes reações de amplificações foram utilizadas: 95ºC por 10 minutos seguidos por 40 ciclos (95ºC por 10 segundos, 60ºC por 30 segundos) e por uma curva de dissociação (curva de *Melting*) de 60ºC a 95ºC com incremento de 0,5ºC por 10 segundos (para verificar a formação de dímeros de *primers* e amplificações não específicas). O gene β-actina foi utilizado como referência interna para normalizar a quantidade de RNA total presente em cada reação (VERBEKE et al., 2009). As expressões gênicas foram calculadas de acordo com o método $2^{-\Delta\Delta CT}$ (LIVAK; SCHMITTGEN, 2001). A linhagem QM9414 cultivada em glicerol (condição não indutora) por 24 horas foi utilizada como amostra referência (VERBEKE et al., 2009). Esse experimento foi realizado em triplicata. Os resultados foram analisados utilizando o *software MeV v.4.6.1* e *heatmaps* sendo construídos para avaliar as diferentes expressões entre as linhagens e em cada condição de análise.

Para as análises estatísticas, foram utilizados os métodos *one-way* ANOVA (*nonparametric*) seguido pelo teste de *Bonferroni* (*comparing all pairs of columns*) utilizando o *software Prism v. 5.0*. A linhagem parental QM9414 foi utilizada como amostra padrão para comparação com o mutante *Δxyr1* em todas as condições analisadas, isto é, cultivos do mutante de 2, 4 e 6 horas (soforose), 24 e 48 (glicose) foram comparadas com a linhagem parental, com exceção, o cultivo do mutante *Δxyr1* em 24 horas (celulose) que foi comparado com a linhagem parental 48 horas (celulose) pelo fato dos tempos de cultivos serem diferentes.

4.6.1 *Primers* utilizados nesta análise por PCR em Tempo Real (qPCR-TR)

Os *primers*, utilizados para as análises de expressão gênica por PCR em Tempo Real, foram desenhados de acordo com o programa *Design Software*, disponível em http://www.idtdna.com/site, utilizado com a sequência FASTA de 44 transcritos obtidos em http://genome.jgi-psf.org/Trire2/Trire2.home.html. O gene que codifica actina (*act*) foi usado como controle endógeno para as análises de acordo com Steiger et al., (2010). As sequências dos *primers* (*Forward* e *Reverse*) podem ser verificadas na Tabela 8.

Tabela 8 – *Primers* utilizados na análise de expressão gênica de 22 genes por qPCR-RT.

Proteína ID	Enzima	Sequência 5'→3' (*Foward*)	Sequência 5'→3' (*Reverse*)
72567	*cel6a*	ACAAGAATGCATCGTCTCCG	TGTTCCACCCGTTGTAGTTG
44504	*actina*	TGAGAGCGGTGGTATCCACG	GGTACCACCAGACATGACAATGTT
76672	*cel3a*	CTGTACATCACCTACCCATC	TAGCTGAGATCTCGTCGTC
123989	*cel7a*	CCGAGCTTGGTAGTTACTCTG	GGTAGCCTTCTTGACTGAGT
123992	*swo*	CCAAACTATACGAGTAGCC	GAGTGAATGTCTTGATGG
120749	*cel1a*	TTTGCCTGGTCGCTCATG	AATCAGCTCGTCAAACAGCG
73643	*cel61a*	GCGCCACTGTTCCTGGAG	ACCGCTGCCACCACACTG
49976	*cel45a*	CAGCGACGTCTACATTGG	TGGTCCAGAATGCACTCG
122081	*cel7b*	CCCTCAACACTAGCCACCAG	AGGTCTTGGAGGTGTCAACG
22197	*cel1b*	CCATCTACATCACCGAGAACG	TCCAAGTGCGAGTCAAAGTAG
121735	*cel3b*	CCAGGATAACTTCAACGAGGG	ATGTGGAGGTTGGAGAACTTG
82227	*cel3c*	GCGTACAATGGCATCAATGG	ATCCCAACCCCATTCCTTTC
120312	*cel5a*	GCCACTACTATCACCACTTCG	GTACAGCCAAAGTCAAAACCC
82616	*cel5b*	AAAGTACCGTCCAACACCC	TGGTAACAGCATTGGGAGTG
123232	*cel12a*	AGATTGCCATTCCCCAGAAG	CCCGAGTACGTGACATGATTC
120961	*cel61b*	ACTATGTCTTCCGCCATGAAC	CACGCACTGAGGATAGTTCTG
46816	*cel3d*	AACCCAGCATATCTCAACTACG	CTTGAAGGTAGCGTAAGACAGG
76227	*cel3e*	ATGTCTGGAAGTGAGGTTGC	TCGTGAGTCCAAAGTGAACAG
49081	*cel74a*	GCCTTGTATCTGACCTATTCCG	TGATGTCTTTCCAAGTTCCCC
74223	*xyn1*	GGCCAAATTATCGTCAACTGTC	TCTGTCTTTTGGGCTTGGAG
123818	*xyn2*	TGTCAACGAGCCTTCCATC	TCTGCACAGTAACAGTTCCG
120229	*xyn3*	AAGTCATCCGCACCCATG	GTTCAAAACTCACCCAAGCAC
111849	*xyn4*	TGTCAGCAATTCGGGTCTTC	GACAGGTTGGCAAAATGGTG

4.7 Análise *in silico* de putativos sítios de ligação para XYR1 em regiões promotoras dos genes celulolíticos e xilanolíticos

A sequência do genoma de *T. reesei* (MARTINEZ et al., 2008) foi obtida junto ao *Joint Genome Institute T. reesei database* (http://genome.jgi-psf.org/Trire2/Trire2.home.html) e foi utilizada para a obtenção da região *5'-upstream* (1,5-kpb) para cada gene. A ocorrência dos motivos 5'-GGCWWW-3' do fator transcricional XYR1 foram determinados em ambas as fitas de 22 genes, analisados neste estudo, bem como para os 9117 genes no genoma de *T. reesei*, usando um *Perl script*. A frequência de cada motivo foi calculada pelo número total de pares de base das sequências *upstream* analisadas pelo número de motivos, como descrito por Furukawa et al., (2009). Potenciais motivos de ligação foram identificados na região promotora (1,5 kpb), sendo utilizado o algoritmo *Scope version 2.1.0* (http://genie.dartmouth.edu/scope/) (CARLSON et al., 2007).

4.7.1 Descoberta do suposto motivo de XYR1 em *T. reesei*

A ocorrência dos motivos 5'-GGCWWW-3' foram identificados por Furukawa e colaboradores (2009), sendo que oito combinações diferentes podem ser identificadas (W representa tanto uma adenina (A) como uma timina (T)). Pergunta-se: qual o verdadeiro sítio de ligação do fator transcricional XYR1? Portanto, análises do elemento regulador em cis XYR1 foram realizadas com quatro grupos de genes corregulados obtidos de diferentes experimentos (qPCR-RT e RNA-seq). Três grupos representaram genes que codificam fatores transcricionais (FTs) e que foram identificados por análises de dados de RNA-seq (CASTRO et al., 2014b) de *T. reesei* QM9414, crescidos em celulose (7 fatores), soforose (18 fatores) ou glicose (18 fatores) como fontes de carbono e foram especificamente up-regulados. Esses genes identificados podem ser verificados no CD Anexo (Arquivo 1, Tabelas 1.1 a 1.3). O quarto grupo é representado por 22 genes que codificam celulases cujos promotores são regulados pelo fator transcricional XYR1 (CASTRO et al., 2014a). Para cada um dos 65 genes analisados, sequências FASTA (1,5 kpb), antecedentes ao códon ATG (região *5'-upstream*), foram obtidas através do genoma de *T. reesei* com um *Perl script*. Para os quatro grupos, as sequências FASTA foram utilizadas para identificar o motivo de XYR1 como descrito a seguir. A abordagem geral desse experimento pode ser verificada na Figura 22.

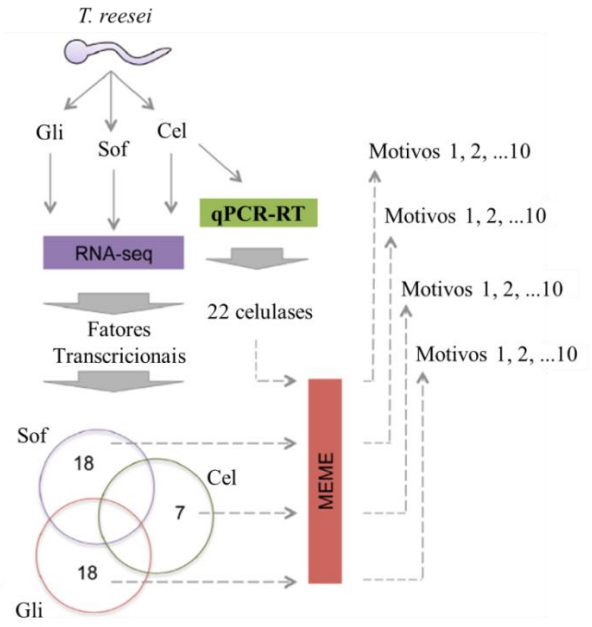


FIGURA 22 – Representação esquemática da abordagem usada para a descoberta do motivo XYR1 em *T. reesei*. Cel = celulose; Sof = soforose e Gli = glicose. MEME = ferramenta utilizada para análises dos motivos.

4.7.1.1 Suposto motivo (*de novo*) de XYR1

Para identificar o elemento regulatório cis de XYR1 (d*e novo*) nos quatro grupos já retratados anteriormente, análises das sequências das regiões promotoras foram realizadas utilizando o algoritmo MEME – *Multiple Expectation Maximisation for Motif Elicitation* (Figura 22) (BAILEY et al., 2006). Para as análises no MEME, alguns parâmetros foram selecionados para a pesquisa de motivos: sequências curtas (6 a 10 nucleotídeos) que deveria ocorrer zero ou uma vez por sequência na fita de DNA ou na sua fita complementar, seguido de, no máximo, 10 motivos diferentes obtidos pelo programa. Como resultados, motivos que foram similares aos motivos de XYR1 (*T. reesei*) já conhecidos na literatura (FURUKAWA et al., 2009) foram selecionados. Os motivos repetidos para o consenso do XYR1 foram fundidos em apenas um único motivo denominado de PWM (*Position Weight Matrixes*) de acordo com Schneider e Stephens (1990). Portanto, para XYR1, utilizou-se um PWM representando os 8 primeiros nucleotídeos e foi nomeado PWM $_{XYR1}$ que foi utilizado para análises subsequentes.

4.7.1.2 Análise do elemento regulatório cis de XYR1 no genoma de *T. reesei*

O PMW$_{XYR1}$ descoberto pelo MEME foi utilizado para analisar a região promotora de todos os genes do genoma de *T. reesei* (MARTINEZ et al., 2008). Para essa análise, regiões promotoras (1,5 kpb de ~9000 genes) foram analisadas para identificar o melhor motivo para o PWM em cada promotor. Em seguida, o mesmo conjunto de dados foi reanalisado para identificar vários motivos por promotor com um valor acima de um específico *threshold*, que foi definido como 6,2 para o PWM$_{XYR1}$. O resultado identificou sítios que foram analisados para identificar o elemento regulatório cis adjacente, localizado dentro de sequências curtas (menor que 30 pb), com arquitetura similar a sítios funcionais relatados anteriormente para XYR1 (FURUKAWA et al., 2009; LING et al., 2009; RAUSCHER et al., 2006; ARO et al., 2001; MURRAY et al., 2003 e ZEILINGER et al., 1998). Além disso, essas mesmas etapas, isto é, a identificação do melhor sítio, mapeamento de vários sítios e a procura de elementos adjacentes, foram aplicadas nas regiões promotoras dos 22 genes de celulases usando o PWM$_{XYR1}$. Para finalizar essas análises, genes diferencialmente expressos nas condições indutoras celulose e soforose em *Δxyr1* (dados

RNA-seq, Capítulo III) foram utilizados para determinar genes potencialmente diretamente regulados por XYR1 utilizando o referido motivo de análise.

4.8 Sequenciamento de alto desempenho do transcriptoma (RNA-Seq)

A metodologia de RNA-Seq (sequenciamento de RNA) emprega tecnologias de sequenciamento maciço e de alto rendimento recentemente desenvolvidas para mapear e quantificar os transcritos, sendo considerado um método revolucionário, pois possui alta sensibilidade e pode ser usado para caracterizar o transcriptoma completo de um organismo (SILVA, 2012b; VARUZZA, 2013).

4.8.1 Preparo das amostras de RNAs totais para realização do sequenciamento

As amostras de RNA totais das replicatas biológicas foram agrupadas (*Pools* de 1 a 3) referentes aos tempos de cultivo para cada linhagem, isto é, celulose (24, 48 e 72 horas – QM9414), celulose (8 e 28 horas – *Δxyr1*), soforose (2, 4 e 6 horas) e glicose (24 e 48 horas), totalizando 18 bibliotecas (Figura 23). Para a realização do sequenciamento, utilizaram-se 20µg de RNAs totais dessas amostras que foram liofilizadas e estocadas usando o *RNAstable*® *tube kit* (*Biomatrica, San Diego, CA*/*USA*). O sequenciamento foi realizado pela empresa *LGC Genomics GmbH* (*Berlin/Germany*) que utilizou a plataforma *Illumina Hiseq*^{TM} *2000* para determinar o perfil transcricional de *T. reesei* em resposta à presença de três fontes de carbono. A empresa realizou o sequenciamento de cada biblioteca em duplicatas. A seguir, as etapas detalhadas para o desenvolvimento do sequenciamento.

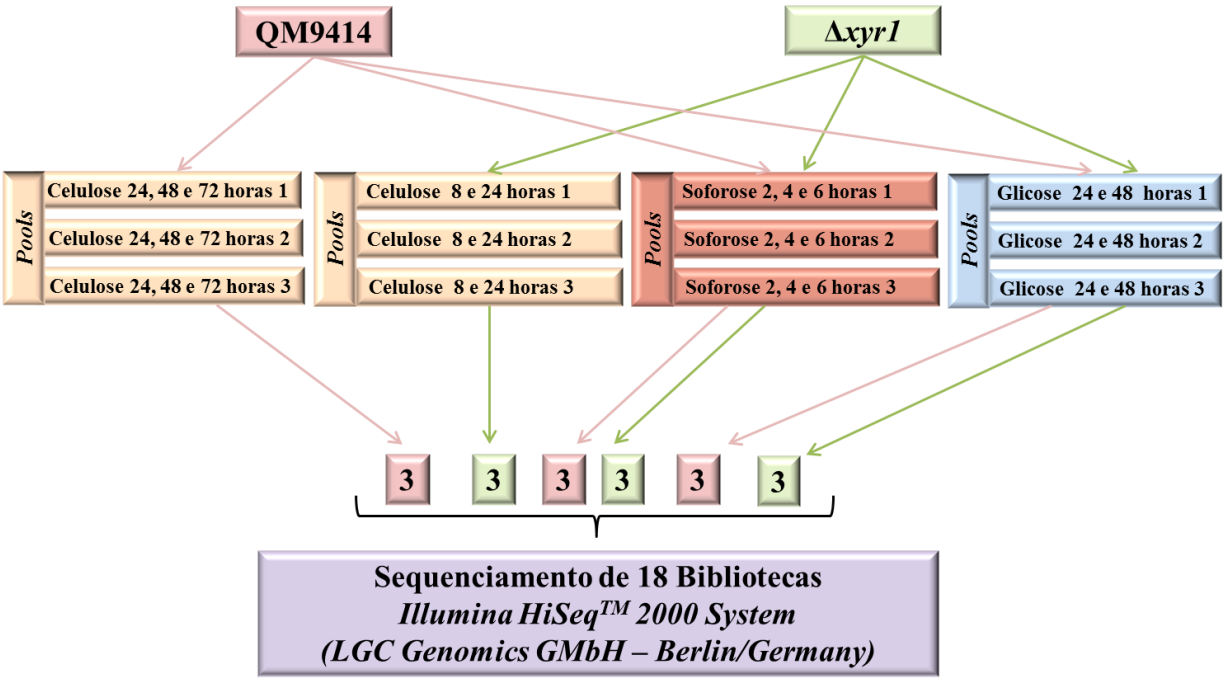


FIGURA 23 – Esquema da estratégia utilizada para obtenção dos RNAs para sequenciamento.

4.8.2 Sequenciamento utilizando plataforma *Illumina Hiseq™ 2000*

O processo de sequenciamento no *Illumina Hiseq™ 2000* consiste em três etapas: preparo das bibliotecas com *TruSeq™ RNA*, geração de clusters (no cBot) e sequenciamento (Figura 24). O software CASAVA versão 1.8.2 foi utilizado para conversão dos dados gerados no sequenciamento para o formato de sequências *fastq*.

FIGURA 24 – Sequenciamento simplificado para o *Illumina Hiseq™ 2000*.
Fonte: Adaptado de (ILLUMINA, 2011a).

4.8.2.1 Preparo das bibliotecas de cDNAs

Primeiramente, as amostras de RNAs totais referidas anteriormente foram utilizadas para o preparo das bibliotecas de cDNAs de acordo com o protocolo disponibilizado pela Illumina®, sumarizado a seguir. As bibliotecas foram preparadas com *TruSeq™ RNA Sample Preparation Kits* (ILLUMINA, 2011b), envolvendo cinco etapas principais: (1) Seleção de RNAs mensageiros (RNAm) (Figura 25A) pela cauda poli-A na região terminal 3', usando sondas de oligonucleotídeos de timina (poli-T). Após o enriquecimento da amostra em RNAm, estes foram fragmentados por nebulização[8] e a primeira fita de cDNA complementar aos RNAm foi sintetizada com uso de *primers* hexâmeros randômicos N6 e enzima transcriptase reversa (HEAD et al., 2014). (2) Logo, a segunda cadeia foi gerada para criar a cadeia dupla de cDNA (Figura 25B). (3) As extremidades das cadeias duplas de cDNA foram reparadas e as extremidades 5' foram fosforiladas (Figura 25C). (4) Às extremidades 3' foram adicionadas uma base adenina/A (Figura 25D). Na última etapa (5), ligaram-se os adaptadores às duas extremidades dos fragmentos de cDNA (Figura 25E-F). As bibliotecas foram verificadas quanto ao seu tamanho, por concentração no *NanoDrop* e através do *Bioanalyzer 2100* (*Agilent*).

[8] Nebulização: é um método mecânico submetido a condições de pressão e tempo que promovam quebras aleatórias gerando fragmentos de tamanhos variados (SECCON, 2012).

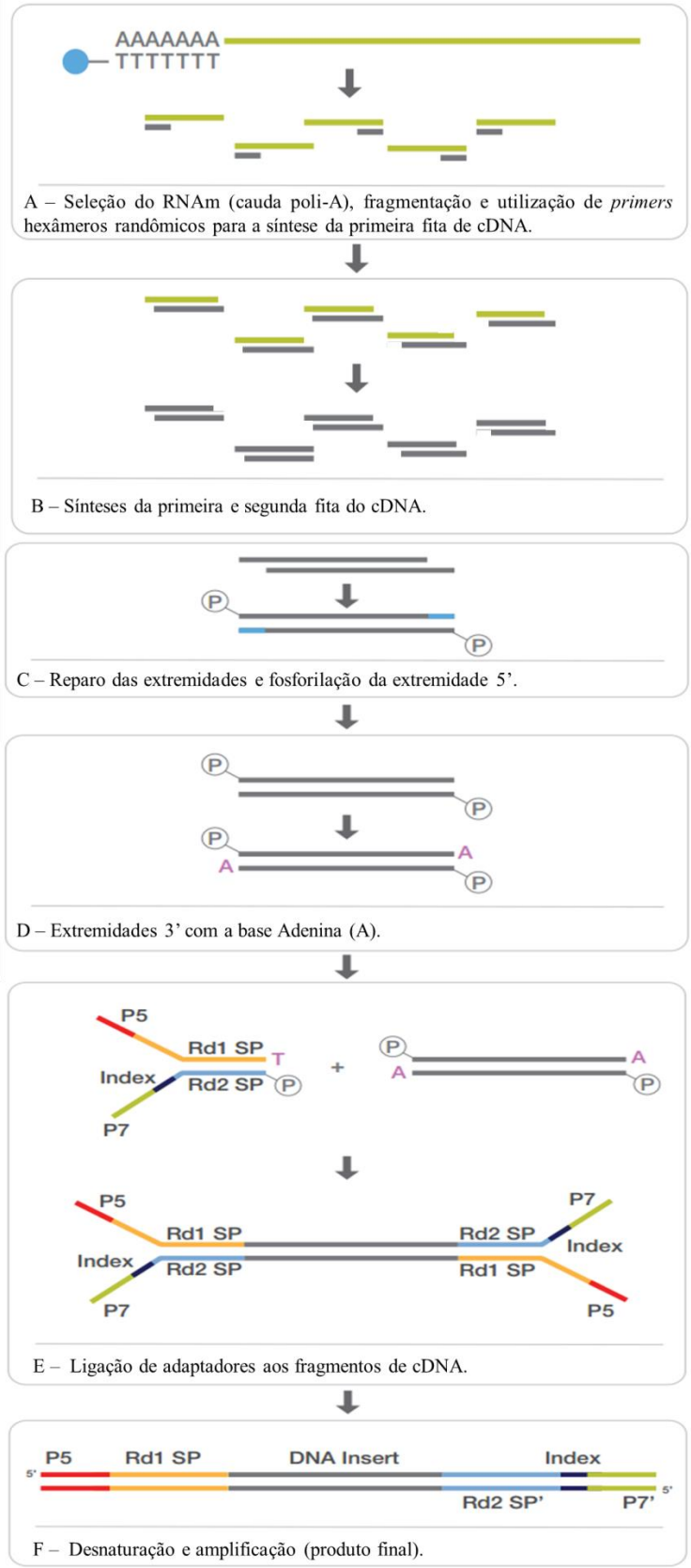


FIGURA 25 – Etapas para preparo das bibliotecas de cDNAs.
Fonte: Adaptado de (ILLUMINA, 2011b).

4.8.2.2 Geração de *clusters*

Para a realização do sequenciamento, os *templates* fita simples são primeiramente amplificados para formar clusters clonais na *flow cell* de sequenciamento. Essa possui sua superfície revestida com oligonucleotídeos denominados P5 e P7, que são complementares aos adaptadores ligados à extremidades dos fragmentos durante a construção da biblioteca.

Os fragmentos com adaptadores (P5 e P7) ligados às extremidades foram hibridizados à *flow cell* pela extremidade 5', deixando a extremidade 3' livre para servir como iniciador na reação de sequenciamento. No primeiro ciclo de amplificação, nucleotídeos não marcados foram fornecidos para a síntese da segunda fita do fragmento imobilizado no suporte. A alta densidade de adaptadores no suporte facilitou a hibridização do adaptador livre dos fragmentos imobilizados a sua sequência complementar fixa durante o ciclo de hibridização. Após esse ciclo, o fragmento formou uma estrutura em "ponte" na superfície da *flow cell* e a extensão ocorreu, formando a fita complementar também em "ponte". No ciclo de desnaturação, as fitas foram separadas e linearizadas. Esses ciclos foram repetidos 35 vezes e, assim, cerca de mil cópias geradas de cada fragmento nessa PCR de fase sólida permaneceram próximas umas das outras, formando um *cluster* de sequenciamento (CARVALHO; SILVA, 2010). Uma vez formados os *clusters*, as sequências complementares à original foram removidas para que cada *cluster* tivesse apenas uma sequência. Na Figura 26, é illustrada como ocorre a formação dos *cluster paired end* na *flow cell*.

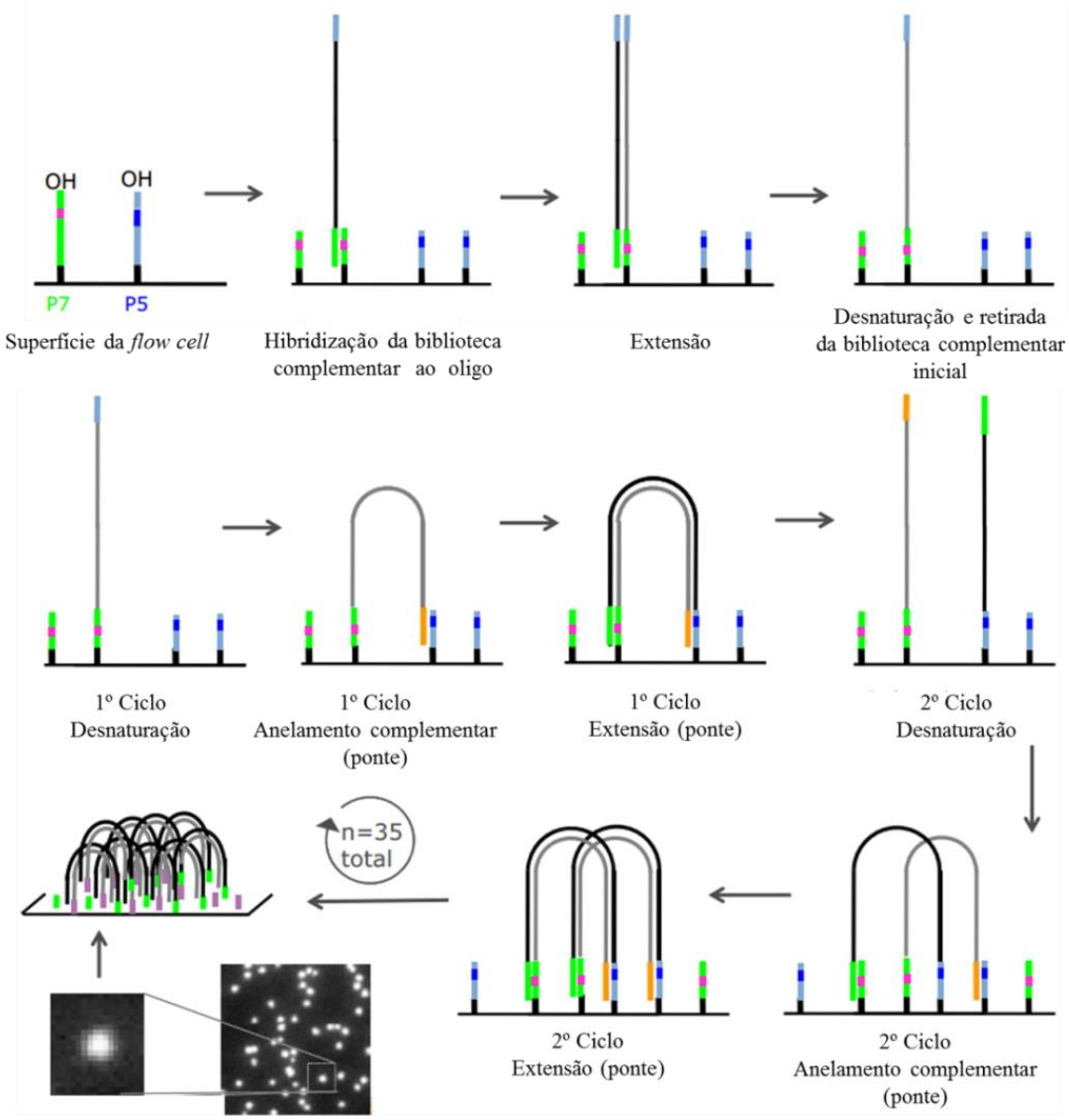


FIGURA 26 – Geração de *cluster paired end*.
Fonte: Adaptado de (FISHMAN, 2010).

4.8.2.3 Sequenciamento *Paired end*

No equipamento *HiSeq*^{TM} *2000*, milhões de *clusters* são sequenciados simultaneamente. Durante o sequenciamento, os fragmentos de cDNA foram sequenciados base a base em paralelo, usando quatro nucleotídeos marcados com fluorescência. As quatro bases competem umas com as outras para se ligarem ao alvo, esta competição natural garante a alta precisão (Figura 27A). Depois de cada ciclo, os fluorocromos foram excitados por um laser, e imagens foram capturadas por câmeras (Figura 27B). Esse fluorocromo foi depois retirado para a ligação de uma próxima base na região de sequenciamento (Figura 27B)

(KIRCHER; KELSO, 2010). Após a incorporação de cada nucleotídeo no fragmento em síntese, ocorreu uma etapa de lavagem para remoção dos reagentes excedentes e remoção do terminal 3' bloqueado (fita fixa à superfície da *flow cell*) e do fluoróforo do nucleotídeo incorporado no ciclo anterior (Figura 27C) para dar procedência a reação de sequenciamento (Figura 27A-C) (CARVALHO; SILVA, 2010). O processo se repetiu por 100 ciclos referentes ao sequenciamento da *Read 1*. Para o sequenciamento do *Read 2* (*paired end*), a fita recém sequenciada, foi desnaturada ao final do sequenciamento do *Read 1* e sua extremidade 3' foi desbloqueada para a resíntese da fita molde, ligação do *primer* de sequenciamento da *Read 2* e, finalmente, sequenciamento da *Read 2* (Figura 28).

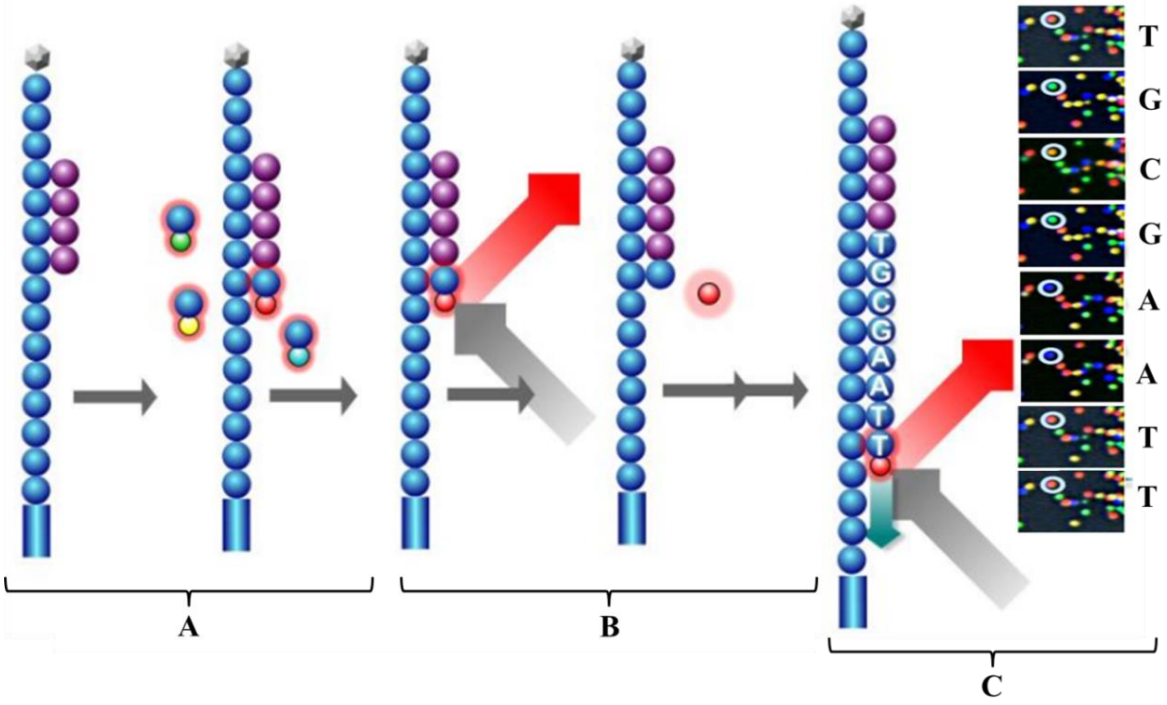


FIGURA 27 – Sequenciamento plataforma *Illumina HiSeq™ 2000.*
As letras de A-C foram identificadas no texto.
Fonte: Adaptado de (BROAD-INSTITUTE, 2014).

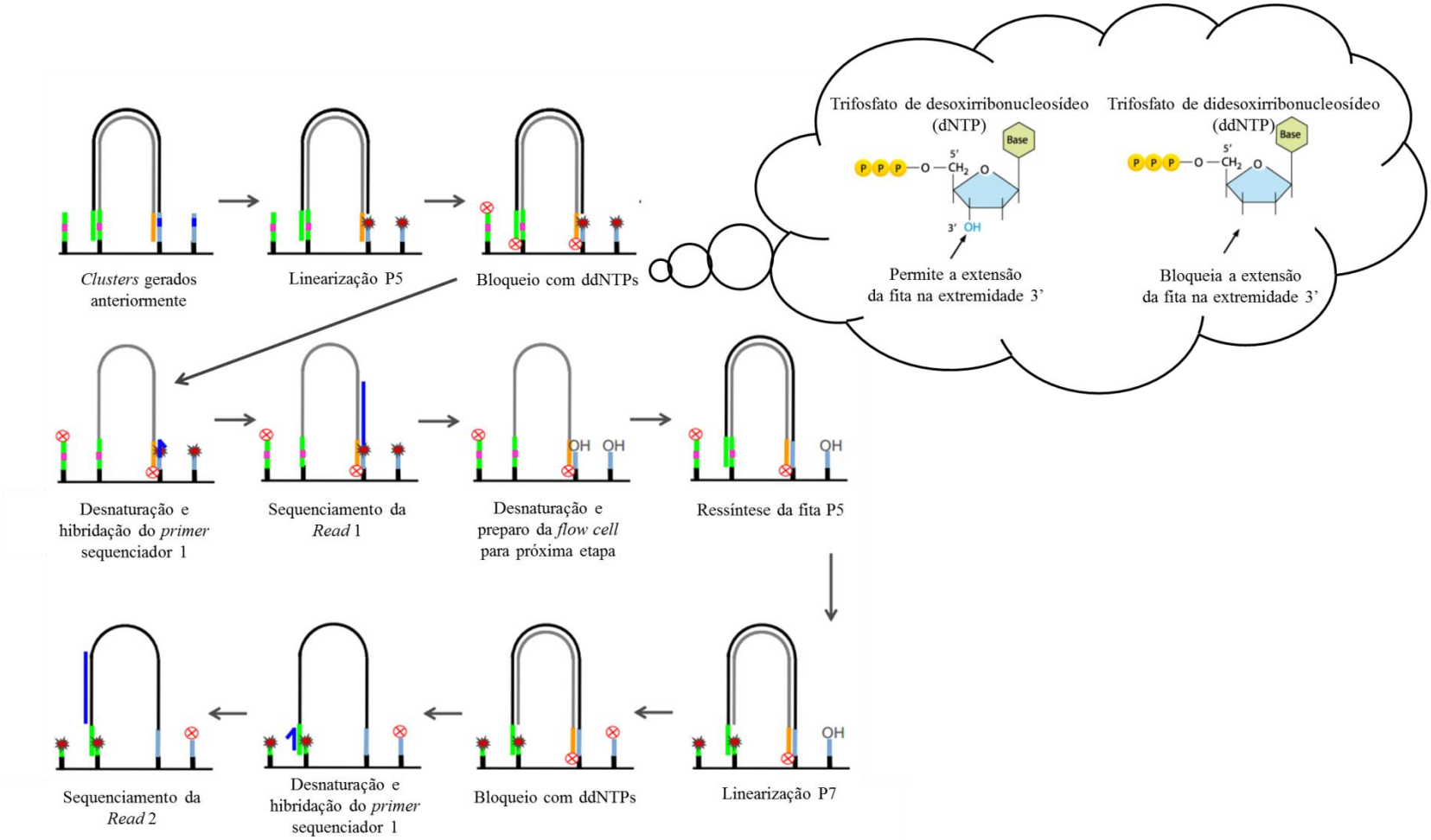


FIGURA 28 – Sequenciamento das duas extremidades do fragmento de cDNA (*Read 1* e *Read 2*).
As etapas de *clusters* gerados e linearização não estão representados no esquema na posição vertical e sim em ponte para facilitar o entendimento do esquema.
Fonte: Adaptado de (BROAD-INSTITUTE, 2014) e (PEGOS; SALAZAR, 2014).

4.9 Bioinformática

Diversas ferramentas de bioinformática foram recentemente desenvolvidas para analisar a imensa quantidade de informação gerada pela técnica de RNA-seq. Essas ferramentas diferem entre si quanto à normalização e às técnicas estatísticas aplicadas com impacto nos resultados finais (ALMEIDA, 2013). A empresa *LGC Genomics* realizou a primeira etapa de análises após o sequenciamento: controle de qualidade das *reads*. Utilizou-se a ferramenta *FastQC* (ANDREWS, 2010) para avaliação da qualidade das *reads* e, em seguida, procedeu à correção, remoção e corte das *reads* que não respeitaram os valores mínimos de qualidade estabelecidos (PABINGER et al., 2014). Alguns parâmetros utilizados foram: o tamanho das sequências; a qualidade de cada base e das *reads*; número de sequências repetidas; e a distribuição das bases. A ferramenta utilizada permitiu fazer relatórios detalhados aos vários parâmetros de qualidade. Após a realização desta etapa, iniciaram-se as análises com as ferramentas bioinformáticas que estão sumarizadas na Figura 29.

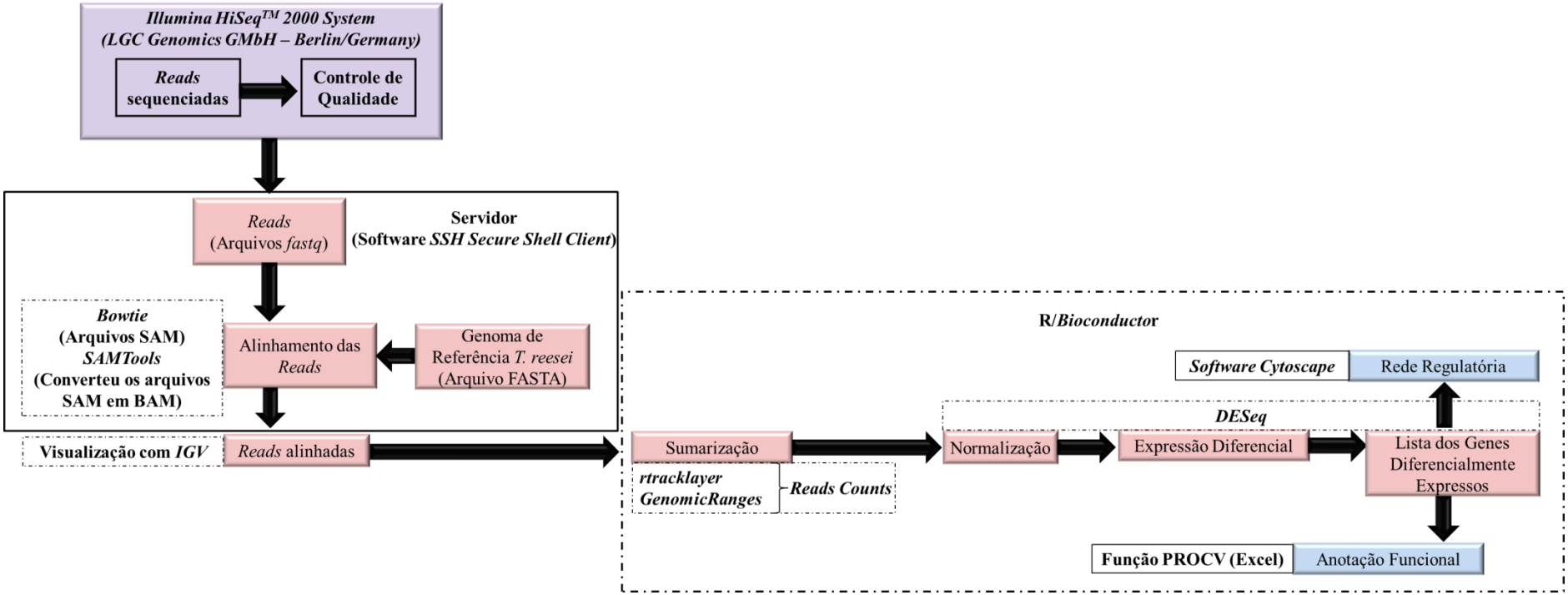


FIGURA 29 – Análises dos dados de RNA-seq por ferramentas bioinformáticas.
Os retângulos com linhas pontilhadas destacam as ferramentas bioinformáticas utilizadas nesta etapa.
Os retângulos com linhas sólidas são softwares que não fazem parte do pacote R/*Bioconductor*.

4.9.1 Alinhamento das *reads* no genoma de referência

A etapa do alinhamento foi a fase mais crítica na análise de dados, uma vez que foi deste resultado, que todas as outras operações se processaram. O alinhamento das *reads* no genoma de referência de *T. reesei*, disponível no JGI (http://genome.jgi-psf.org/Trire2/Trire2.home.html), foi realizado usando o algoritmo *Bowtie* (específico para o alinhamento de sequências curtas) (LANGMEAD, 2010). Esse algoritmo utilizou a transformação de *Burrows-Wheeler*[9] para compressão dos dados (LI; DURBIN, 2009; LANGMEAD, 2013). Com o *output* foram gerados arquivos no formato SAM. Através do algortimo *SAMtools* (LI et al., 2009), o arquivo SAM foi transformado em uma versão binária BAM, que se encontra numa forma compacta e indexada, sendo, por isso, mais eficiente no que diz respeito ao acesso dos dados e ao tamanho do arquivo. Após o alinhamento, os arquivos BAM foram utilizados para visualização dos alinhamentos das *reads* no genoma de referência de *T. reesei*, utilizando-se uma ferramenta elaborada pelo *Broad Institute* o *Integrative Genomics Viewer* (IGV) que é uma ferramenta de alto desempenho para exploração interativa de grandes conjuntos de dados genômicos. Suporta uma ampla variedade de tipos de dados, incluindo dados obtidos de sequenciamento de nova geração e anotações genômicas (THORVALDSDÓTTIR et al., 2013). Utilizando o IGV, foi possível visualizar a região genômica correspondente ao fator de transcrição *xyr1* (Proteína ID 122208) que foi de fundamental importância para confirmar a deleção deste fator no mutante utilizado neste estudo. A etapa de visualização permitiu fazer um rastreio de possíveis problemas no alinhamento, e também possibilitou a visualização de zonas específicas do genoma, verificar a qualidade das bases e a cobertura em regiões específicas do genoma.

4.9.2 R e *Bioconductor*

R (VENABLES; SMITH, 2014) é um sistema de código livre voltado para análises estatísticas e confecção de gráficos amplamente utilizados no ambiente científico, este sisrema provê um rico ambiente interativo, manipulável por meio de inserção de comandos e uma linguagem de programação, que permite a criação de *scripts* e programas mais

[9] também conhecido pela sigla em inglês BWT (*Burrows-Wheeler Transform*), é um processamento estatístico de um bloco de dados que aumenta a redundância espacial, facilitando a aplicação de técnicas de compressão de dados (LANGMEAD, 2013).

complexos. Tem suas funcionalidades extensíveis por meio de um sistema de pacotes, no qual usuários podem criar métodos de análises ou procedimentos computacionais para fins específicos. Para análises biológicas, além de pacotes inclusos na versão base, existe um projeto chamado *Bioconductor* (MORGAN, 2013; MORGAN, 2014) que agrega inúmeros pacotes adicionais ao R. Tais pacotes, que estendem às funcionalidades dos pacotes originais, são específicos para análises de dados biológicos que, em sua maioria, abrangem o campo da Bioinformática. Os softwares, listados a seguir, foram amplamente utilizados nas análises conduzidas neste trabalho.

4.9.2.1 Sumarização: *Reads Counts* (Tabelas)

A partir do arquivo resultante do alinhamento, o próximo passo foi resumir as *reads* numa unidade com algum significado biológico. A abordagem utilizada foi contar o número de *reads* mapeadas nos vários éxons de um gene, isto é, contagem das *reads* alinhadas aos éxons das diferentes isoformas[10], utilizando o método união de éxons (Figura 30) (GARBER et al., 2011).

FIGURA 30 – Método de contagem das *reads*.
Fonte: Adaptado de (GARBER et al., 2011).

Para obtenção das tabelas (*Reads Counts*) contendo a quantidade de *reads* correspondentes a cada gene do genoma de *T. reesei*, foram utilizados os pacotes disponíveis no *Bioconductor*: *rtracklayer* (LAWRENCE et al., 2014) e *GenomicRanges* (ABOYOUN et al., 2014).

[10] Diferentes formas de uma proteína que podem ser produzidos a partir de genes diferentes ou a partir do mesmo gene por *splicing* alternativo (JOAQUIM; EL-HANI, 2010).

4.9.2.1.1 *rtracklayer*

O pacote *rtraclayer* é uma interface (ou camada) entre R e o *Bioconductor* que permitiu manipular e visualizar anotações genômicas em vários formatos, como o formato *GFF* (*General Feature File*). As características do *rtracklayer* podem ser divididas em duas categorias: 1) a importação/exportação de dados; e 2) o controle e a consulta de sessões de navegadores do genoma externo (LAWRENCE et al., 2014). O pacote *rtracklayer* foi utilizado para importar o arquivo de anotação do genoma de *T. ressei* (arquivo *GFF*) para o R. Essa etapa permitiu que as informações sobre as coordenadas genômicas dos 9129 genes descritos em *T. ressei* fossem confrontadas com os dados do alinhamento em arquivos BAM.

4.9.2.1.2 *GenomicRanges*

O pacote *GenomicRanges* foi utilizado para a representação e manipulação de intervalos genômicos. Esse pacote possui uma função "*summarizeOverlaps*", que calcula as sobreposição entre as *reads* e as características genômicas, utiliza o método de união de éxons citado anteriormente. A característica genômica pode ser uma parte de uma região genômica, tais como: um gene, transcrito, exon, etc. Cada *read* foi contada uma única vez. Esse pacote, juntamente com *rtraclayer*, foi utilizado para realizar as contagens, isto é, quantas vezes as *reads* se alinharam em uma parte do genoma correspondendo a um gene, criando tabelas como resultados. Exemplos de alinhamentos e contagem das *reads* podem ser verificadas na Figura 31.

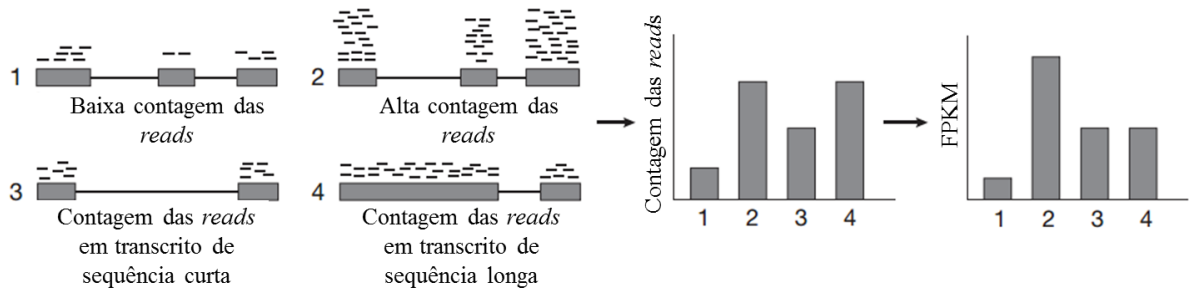


FIGURA 31 – Alinhamentos e contagens das *reads*.
Fonte: Adaptado de (GARBER et al., 2011).

Os valores de FPKM (Fragmentos por Kilobase por Milhão de Reads Mapeadas) foram obtidos com a utilização de um algoritmo *cuffdiff* que quantificou o nível de expressão

(abundância dos transcritos) dos 9129 genes de *T. reesei* normalizando os dados pelo tamanho do gene e pelo total de leituras mapeadas (TRAPNELL et al., 2010).

4.9.2.2 Pacote DESeq

O pacote DESeq foi utilizado para realizar análises de expressão gênica diferencial. A análise de expressão diferencial utilizou o número de *reads* reportadas para cada gene em cada amostra para realizar os testes estatísticos, baseados na distribuição binomial negativa, que apontam se um gene é diferencialmente expresso em uma condição em relação à outra (ANDERS; HUBER, 2010). Esse pacote foi utilizado para: normalizar os dados; estimar a variância; obter genes diferencialmente expressos; gráficos de correlação das amostras; *heatmap* dos 30 genes mais expressos; *heatmap* da distância das amostras; e análise de componentes principais (PCA). Na Tabela 9, estão representados os tipos de comparações utilizadas neste estudo, compreendendo as duas linhagens (parental QM9414 e mutante *Δxyr1*) e as fontes de carbono para a obtenção da lista de genes diferencialmente expressos.

Tabela 9 – Comparações realizadas entre as amostras utilizadas no RNA-seq.

	Condições		
QM9414	Celulose/Glicose	Soforose/Celulose	Soforose/Glicose
***Δxyr1*/QM9414**	Celulose	Soforose	Glicose

4.9.2.2.1 Normalização

A normalização dos dados é um passo de elevada importância, que visa garantir que as estimativas de expressão gênica são comparáveis entre genes/isoformas e comparáveis entre bibliotecas e amostras diferentes. A forma de normalização utilizada foi a *median log deviation* disponível no pacote DESeq, que calcula a mediana da razão para cada gene, e, o número de *reads* pela sua média geométrica em todas as amostras, por meio da função *estimateSizeFactors* (ANDERS; HUBER, 2010). A razão pela qual foi necessária esse tipo de normalização foi o fato da profundidade do sequenciamento variar entre as várias bibliotecas, o que iria influenciar o número de *reads* alinhadas a um dado éxon, não representando, no entanto uma expressão fidedigna (OSHLACK et al., 2010).

4.9.2.2.2 Análise de Expressão diferencial

A análise da expressão diferencial foi realizada utilizando o pacote DESeq, o qual baseia-se na distribuição binomial negativa para verificar a variação biológica entre as amostras. Dados como *Fold Chang*, Log_2 *Fold Change* e *p-valor* foram obtidos para cada gene em suas respectivas condições de análises. Para a obtenção da lista de genes diferencialmente expressos, foram aplicados limiares para selecionar genes estatisticamente significativos e diferencialmente expressos. Genes diferencialmente expressos são aqueles que apresentaram *p-valor ajustado* $\leq 0{,}05$. Também nessa lista, constam a identificação de genes diferencialmente expressos up-regulados ou down-regulados, de tal maneira que estabeleceu-se um "limite", em que, para ser considerado up-regulado, tenha que apresentar um valor maior que 2 de expressão e down-regulado um valor menor que -2 de expressão. Como os resultados de *Fold Change* estão em Log_2, assim, tirando a função logarítmica temos: $Log_2 2 = 1$. Portanto, os valores de Log_2 *Fold Change* ≥ 1 serão up-regulados e Log_2 *Fold Change* ≤ -1 serão down-regulados, caso, apresentem *p-valor* $\leq 0{,}05$. As tabelas contendo todos os genes que compreendem o genoma de *T. reesei* (9129 genes) tanto na linhagem parental (QM9414) como na linhagem *Δxyr1*, comparado ao parental podem ser verificadas no CD Anexo (Arquivos 2 e 3).

4.9.3 Anotação Funcional

A anotação funcional foi uma etapa que se pretendeu atribuir um significado biológico aos genes. Esse processo caracterizou-se pelo agrupamento dos genes que partilharam propriedades biológicas entre si, sendo, posteriormente, feita uma análise de modo que se verificou as categorias presentes no conjunto de genes diferencialmente expressos. Além das informações disponíveis no arquivo de anotação genômica GFF de *T. reesei*, contendo 9129 genes, buscou-se complementar essas informações em tabelas obtidas no sítio http://genome.jgi-psf.org/Trire2/Trire2.download.html, buscando para cada gene a sua respectiva proteína ID, e, outras anotações como: descrição de cada proteína ID; descrição do GO; categoria principal do GO; KOGs; domínios; vias de sinalização; e, por fim, as informações obtidas no sequenciamento (Log_2 *Fold Change* e *p-valor*) para cada gene e cada condição analisada tanto no parental como no mutante comparado ao parental.

4.9.4 Redes regulatórias

Para a construção das redes regulatórias, os mesmos genes obtidos na lista de genes diferencialmente expressos ($p \leq 0,05$) e up- e down-regulados, tanto na linhagem parental (QM9414) como na linhagem mutante comparado ao parental (*Δxyr1*/QM9414) nas condições experimentais analisadas, tabelas foram elaboradas com as seguintes informações:

- QM9414: Condições (qmCelulose/Glicose, qmSoforose/Celulose e qmSoforose/Glicose); Tipos de interação (up- ou down-regulados) e o gene alvo (proteínas ID de cada gene);
- *Δxyr1*/QM9414: Condições (xyqmCelulose, xyqmSoforose e xyqmGlicose); Tipos de interação (up- ou down-regulados) e o gene alvo (proteínas ID de cada gene);

As redes regulatórias foram geradas usando o *software Cytoscape versão 3.0.1* (SHANNON et al., 2003).

4.10 Outras ferramentas bioinformáticas utilizadas neste estudo

4.10.1 *Box plot*

Esta foi uma ferramenta utilizada para construir diagramas/gráficos de caixas ou *Box plot*. É uma ferramenta que se usa para localizar e analisar a variação de uma variável dentre diferentes grupos de dados, em que o eixo vertical representou a variável analisada (dados da contagem das *reads* dados brutos; e, após a normalização dos dados) e o eixo horizontal as condições de interesse. Esses gráficos foram construídos utilizando o programa R (KONIS, 2014).

4.10.2 *Volcano plot*

O *volcano plot* são gráficos entre a significância (*-Log$_{10}$ p-valor*) versus o *Log$_2$ Fold Change* plotados no eixo y e x, respectivamente, e foram construídos utilizando-se o pacote estatístico R. Na construção desses gráficos, foram aplicados os limiares para selecionar somente os genes estatisticamente significativos e diferencialmente expressos up- e down-regulados (*Log$_2$ Fold Change* ≥ 1 or ≤ -1 com *p-valor* $\leq 0,05$).

4.10.3 Diagramas de *Venn*

Diagramas de *Venn* foram construídos com os dados obtidos nas análises de expressão diferencial para verificar a distribuição dos genes diferencialmente expressos nas condições analisadas. Para a linhagem QM9414, foram construídos diagramas de *Venn* relacionando dois conjuntos (celulose/glicose, soforose/celulose e soforose/glicose) e para o fungo *Δxyr1*/QM9414, foram construídos diagramas relacionando três conjuntos (celulose, soforose e glicose). Para a obtenção desses diagramas foram utilizados o programa R e o pacote *limma* (SMYTH et al., 2014).

4.10.4 Construção de *heatmaps* no estudo de regulons específicos de cada condição

Clusters[11] dos genes up- e down-regulados e diferencialmente expressos foram obtidos utilizando o *software Mev v.4.6.1* para identificar os regulons de celulose, soforose e glicose. O método utilizado para a geração dos *clusters* foi *average linkage*, com correlação *uncentered* como métrica de similaridade. As cores azul e amarelo, nas figuras dos *heatmaps*, representam genes down- e up-regulados, respectivamente.

4.10.5 Categorização funcional

A categorização funcional foi realizada de acordo com o *Gene Ontology* (BLAKE; HARRIS, 2008) e as categorias super-representadas em cada uma das condições analisadas foram determinadas a partir da análise de enriquecimento (*p-valor* $\leq 0,05$) implementada pelo

[11] Permutando as linhas e as colunas de uma matriz para colocar valores semelhantes próximas umas das outras de acordo com o agrupamento.

algoritmo *BayGO* (VÊNCIO et al., 2006). Com os resultados obtidos na análise de enriquecimento, gráficos foram construídos utilizando o *software GraphPad Prism v 5.1*.

4.10.6 Genes *Carbohydrate active enZymes* (*CAZy*)

Baseado na reanotação de genes *CAZy* de *T. reesei* de acordo com HÄKKINEN et al., 2012, esses genes foram analisados com os dados obtidos do RNA-seq e, com estas informações, gráficos foram construídos utilizando *software GraphPad Prism v 5.1* contendo valores obtidos do FPKM de cada gene versus condições.

4.10.7 Análise Filogenética

As sequências das proteínas de *T. reesei* e de outras espécies utilizadas na construção da árvore filogenética foram obtidas a partir do *JGI Genome Portal* e outros bancos de dados *online* (Tabela 10), respectivamente. O alinhamento múltiplo de sequência foi realizado pelo *ClustalW* e a árvore do tipo *neighbour-joining* foram criadas pelo programa Mega 4 (TAMURA et al., 2007) com 1000 *bootstraps*. Uma lipase secretada (ID 57204) de *T. reesei* foi utilizada como *outgroup*.

Tabela 10 – Transportadores de diferentes espécies utilizados na construção da árvore filogenética.

Organismo	Portal de acesso	Identificação	Descrição
Saccharomyces cerevisiae	Saccharomyces Genome Database	YHR094C	Transportador de hexose
Saccharomyces cerevisiae	Saccharomyces Genome Database	YHR092C	Transportador de hexose
Saccharomyces cerevisiae	Saccharomyces Genome Database	YDR343C	Transportador de hexose
Neurospora crassa	*Neurospora crassa* Database	NCU00821	Transportador de açúcar
Neurospora crassa	*Neurospora crassa* Database	NCU04963	Transportador de glicose
Neurospora crassa	*Neurospora crassa* Database	NCU08114	Transporte de celodextrina
Neurospora crassa	*Neurospora crassa* Database	NCU10021	Transportador de glicose
Aspergillus nidulans	AspGD	AN3199	Transportador de açúcar MFS
Aspergillus nidulans	AspGD	AN6831	Transportador de açúcar putativo
Aspergillus oryzae	AspGD	AO090103000130	Maltose permease
Aspergillus oryzae	AspGD	AO090038000233	Maltose permease
Aspergillus niger	GenBank	XP_001397059.2	Transportador ABC
Escherichia coli	GenBank	YP006127731.1	Transportador de açúcar
Metarhizium anisopliae	GenBank	GQ167043.1	Transportador de oligossacarídeos
Metarhizium anisopliae	GenBank	EFY97396.1	Transportador MFS
Metarhizium anisopliae	GenBank	EFY94560.1	Transportador ABC
Talaromyces marneffei	GenBank	XP002152585.1	Maltose permease
Talaromyces stipitatus	GenBank	XP002484658.1	Transportador de açúcar
Candida albicans	Candida Genome Database	19.13383	Sensor de glicose
Ogataea angusta	GenBank	ACA58225.1	Sensor de glicose
Beauveria bassiana	GenBank	EJP62811.1	Canal de potássio
Togninia mínima	GenBank	EOO00826.1	Transportador ABC
Colletotrichum gloeosporioides	GenBank	EQB50276.1	Transportador ABC
Fusarium fujikuroi	GenBank	CCT62487.1	Transportador de cátion

4.11 Validação das análises transcricionais

4.11.1 Análise da expressão gênica por PCR em Tempo Real (qPCR-TR)

Foi utilizado a mesma diluição e a mesma detecção de sinal retratada no item 4.6. O que diferenciou, em relação a outra análise por PCR-TR, foi que cada reação foi realizada em um volume final de 10μL, continha 7,5μL de *Master Mix* (*EvaGreen* 5μL, *primer forward* 0,3μL, *primer reverse* 0,3μL e 1,9μL de água DEPC) e 2,5μL de cDNA. As reações de amplificações foram as mesmas utilizadas no item 4.6. Transcrições de β-actina e *sar1* (pequena GTPase SAR/ARF) foram utilizadas como referências internas para normalizar a

quantidade de RNA total presente em cada reação (controles endógenos) de acordo com (STEIGER et al., 2010) e com o que está especificado no item 4.11.1.1. As expressões gênicas foram calculadas de acordo com o método $2^{-\Delta\Delta CT}$ (LIVAK; SCHMITTGEN, 2001) usando QM9414, crescido em presença de glicose para a validação da técnica de RNA-seq de *T. reesei* QM9414 e de QM9414, crescidos em presença de celulose e soforose para a validação de *T. reesei Δxyr1* comparado ao parental. Estas amostras de QM9414 citadas foram consideradas como amostras referências. Os resultados obtidos foram convertidos em Log_2. Os gráficos finais foram obtidos a partir dos valores de Log_2, usando o *software GraphPad Prism v 5.1*. Esse experimento foi realizado em triplicata.

4.11.1.1 *Primers* utilizados na etapa de validação

Os *primers* utilizados nesta etapa de validação foram desenhados conforme o item 4.6.1. Genes que codificam actina (*act*) e uma pequena *GTPase* SAR/ARF (*sar1*) foram utilizados como controle endógenos para as análises da linhagem parental (QM9414) e apenas o gene actina foi utilizado como controle endógeno para a linhagem mutante comparado a parental (*Δxyr1*/QM9414) de acordo com Steiger et al., (2010). Cerca de 40 *primers*, incluindo genes up- e down-regulados foram desenhados, sendo 20 genes para a linhagem parental e 20 genes para o mutante (Tabelas 11 e 12).

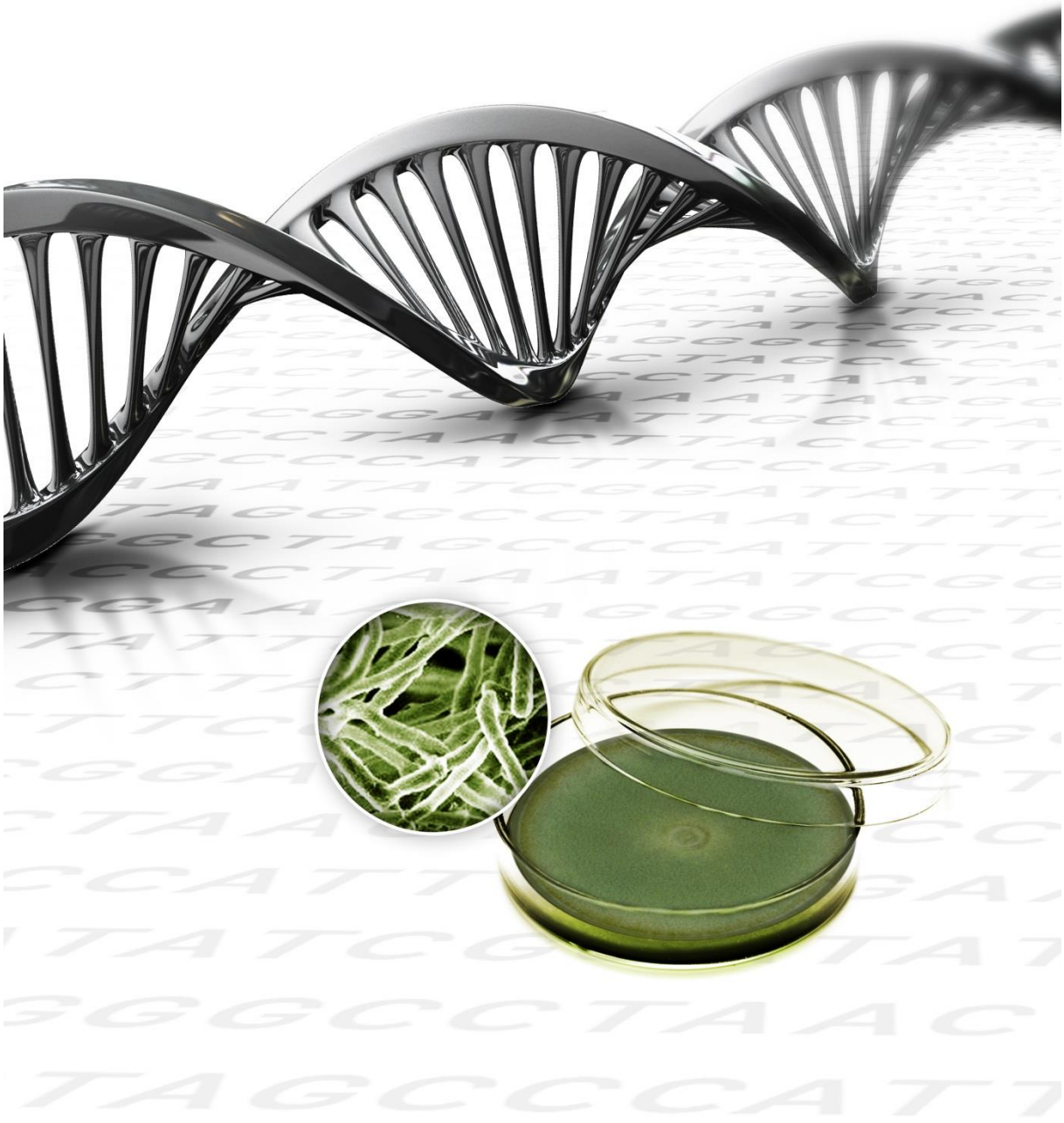
Tabela 11 – *Primers* utilizados na validação de genes diferencialmente expressos (QM9414).

Proteína ID	Sequência 5'→3' (*Foward*)	Sequência 5'→3' (*Reverse*)
58475	ACTTCCAAGGTCACAGGTTG	AGTCCGTTCAACATGACCTG
120975	GTCGTTGTGAGAACTGCAATG	ACCCAGAAATGAAAGCCTCC
21876	GTCGGGTCGTTGACTATAACAC	ATTGCTCTTCACCGTCTTCC
104251	CATTTTGGGCTGGGACTCTAG	TTCAAAGTACTGCTCAGGGC
110267	TGGAATACAATCAGGCGGTG	CGGTTGGAATTCTGCTTTGG
76359	GGGCTAGAGAATTGGAAGACG	ATCTCGCCTCCAAATCCAAG
122792	TCTCGTCAACAGGCAATCAG	CAAGCCGTCCAGTTTTACATG
106314	GTGGTCGTGACTGAGATTGTC	AATTGCCTCTCCCTTGGTG
123673	AGTCATTGATGGAGCTTCTGG	CACTTTGCGTTCTTGTTCTGTC
109945	CTGAAGCATGGATACGAGGTC	AGTAGATCCAAACCAAGGCAAG
72379	ACTCTGAATGCAAGTCTGGAC	TGGTCCTTTTCGCAGTATCC
82227	GCGTACAATGGCATCAATGG	ATCCCAACCCCATTCCTTTC
123232	AGATTGCCATTCCCCAGAAG	CCCGAGTACGTGACATGATTC
120961	ACTATGTCTTCCGCCATGAAC	CACGCACTGAGGATAGTTCTG
72567	ACAAGAATGCATCGTCTCCG	TGTTCCACCCGTTGTAGTTG
123989	CCGAGCTTGGTAGTTACTCTG	GGTAGCCTTCTTGACTGAGT
123992	CCAAACTATACGAGTAGCC	GAGTGAATGTCTTGATGG
73643	GCGCCACTGTTCCTGGAG	ACCGCTGCCACCACACTG
22197	CCATCTACATCACCGAGAACG	TCCAAGTGCGAGTCAAAGTAG
121735	CCAGGATAACTTCAACGAGGG	ATGTGGAGGTTGGAGAACTTG
44504	TGGATCGTCAACTGGTTCTACGA	GCATGTGTAGCAACGTGGTCTTT
61470	TGAGAGCGGTGGTATCCACG	GGTACCACCAGACATGACAATGTT

Tabela 12 – *Primers* utilizados na validação de genes diferencialmente expressos (*Δxyr1*).

Proteína ID	Sequência 5'→3' (*Foward*)	Sequência 5'→3' (*Reverse*)
76359	GGGCTAGAGAATTGGAAGACG	ATCTCGCCTCCAAATCCAAG
62165	GTGTTTGCTGTGAAAGATGGG	TCATATTCCAAGTCCCACTGTG
4146	GACGATTGAAGAGATGAATGCG	AATTATGATCCCACTCGGCAG
123673	AGTCATTGATGGAGCTTCTGG	CACTTTGCGTTCTTGTTCTGTC
64710	GTGTGATTTGCGATGCTTCTG	TCTAGGGAGGTTCTGAGGAAG
109945	CTGAAGCATGGATACGAGGTC	AGTAGATCCAAACCAAGGCAAG
123989	CCGAGCTTGGTAGTTACTCTG	GGTAGCCTTCTTGACTGAGT
72567	ACAAGAATGCATCGTCTCCG	TGTTCCACCCGTTGTAGTTG
122081	CCCTCAACACTAGCCACCAG	AGGTCTTGGAGGTGTCAACG
120312	GCCACTACTATCACCACTTCG	GTACAGCCAAAGTCAAAACCC
123232	AGATTGCCATTCCCCAGAAG	CCCGAGTACGTGACATGATTC
73643	GCGCCACTGTTCCTGGAG	ACCGCTGCCACCACACTG
49081	GCCTTGTATCTGACCTATTCCG	TGATGTCTTTCCAAGTTCCCC
120961	ACTATGTCTTCCGCCATGAAC	CACGCACTGAGGATAGTTCTG
76672	CTGTACATCACCTACCCATC	TAGCTGAGATCTCGTCGTC
120749	TTTGCCTGGTCGCTCATG	AATCAGCTCGTCAAACAGCG
82227	GCGTACAATGGCATCAATGG	ATCCCAACCCCATTCCTTTC
76227	ATGTCTGGAAGTGAGGTTGC	TCGTGAGTCCAAAGTGAACAG
123992	CCAAACTATACGAGTAGCC	GAGTGAATGTCTTGATGG
111849	TGTCAGCAATTCGGGTCTTC	GACAGGTTGGCAAAATGGTG
44504	TGAGAGCGGTGGTATCCACG	GGTACCACCAGACATGACAATGTT

CAPÍTULO I



5 CAPÍTULO I

5.1 RESULTADOS

5.1.1 Análises da influência de fontes de carbono na regulação da expressão gênica de genes celulolíticos e xilanolíticos por XYR1

Para avaliar a influência do fator de transcrição XYR1 na regulação da expressão de genes celulolíticos e xilanolíticos em *T. reesei*, as linhagens parental (QM9414) e mutante (*Δxyr1*) foram cultivadas na presença de celulose, soforose e glicose. A expressão de 22 genes que codificam 18 celulases e 4 xilanases (Tabela 8) foi avaliada por qPCR-RT. Em presença de celulose, a expressão de genes de celulases e xilanases no mutante *Δxyr1* (exceto para *xyn3* e *cel3e*) foram geralmente reduzidas. A clusterização hierárquica de genes revelou a formação de dois *clusters* principais (Figura 32A) em presença de celulose, apresentando regulação sinérgica de exoglucanases/endoglucanases/β-glicosidases, por diferentes grupos de genes, incluindo também xilanases. Interessantemente, *xyn3*, *cel3b* e *xyn4* não apresentaram perfil de expressão semelhante a outros genes, sugerindo uma diferente forma de regulação (Figura 32A). Em presença de soforose três *clusters*, foram formados (Figura 32B). O primeiro contém os genes *cel3c*, *cel1b* e *xyn3*; o segundo contém *cel45a*, *cel7a*, *cel61a* e *xyn1*; o terceiro *cluster* contém a maioria dos genes analisados. O gene *cel3b* foi o único que não foi agrupado com nenhum outro gene (Figura 32B). Em presença de glicose, uma condição repressora, a maioria dos genes (13 genes) apresentou um aumento de expressão quando se comparou a linhagem mutante *Δxyr1* com a linhagem parental QM9414 (Figura 32C). Sabe-se, porém, que as fontes de carbono celulose e soforose são condições indutoras de celulases e xilanases, mas a diferença de clusterização em cada condição demonstra que a regulação da expressão dos genes analisados via XYR1 é dependente da fonte de carbono.

Para complementar o perfil transcricional de QM9414 e *Δxyr1*, Diagramas de *Venn* foram construídos com os genes modulados que exibiram valores de expressão ($Log_2 \geq 2$ ou $Log_2 \leq -2$) em *Δxyr1*/QM9414 crescidos em celulose, soforose e glicose (Figura 33A-B).

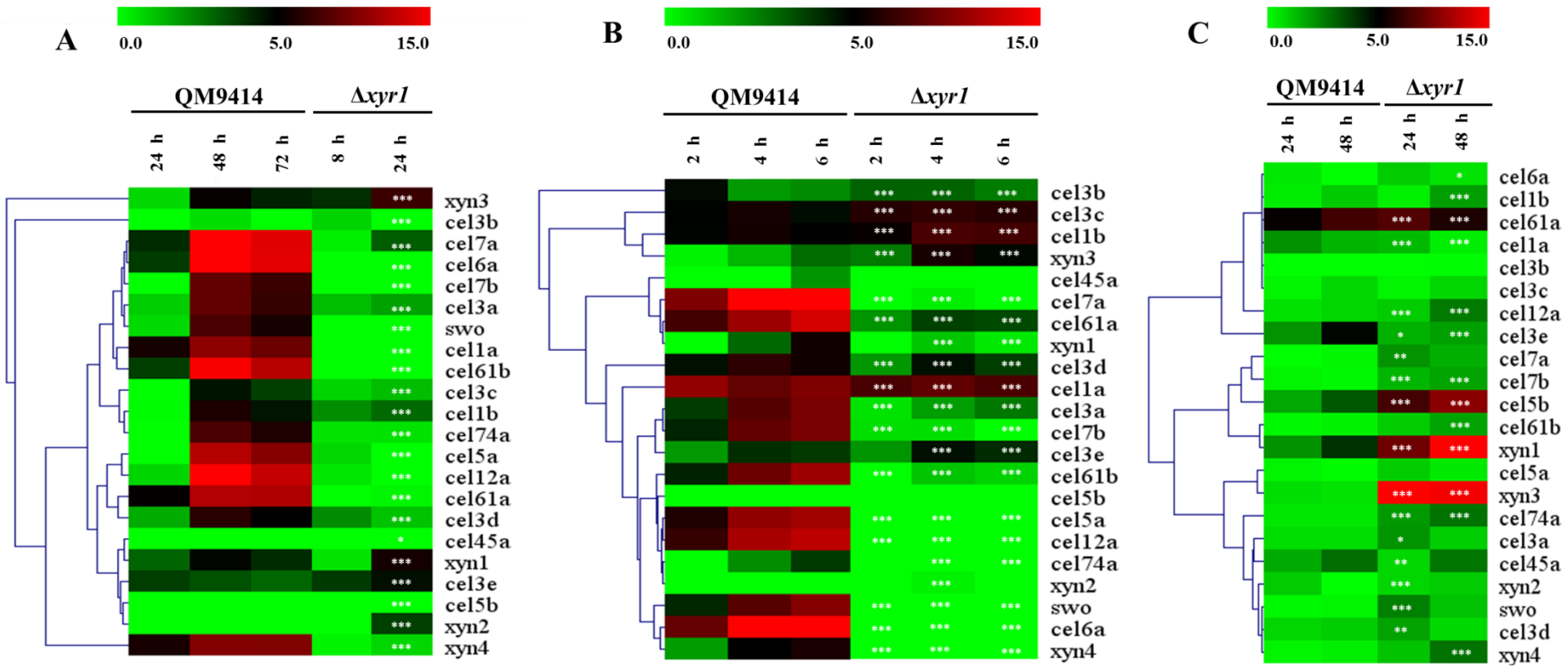


FIGURA 32 – *Heatmap* da expressão de genes celulolíticos e xilanolíticos de *T. reesei* (linhagens: parental QM9414 e mutante *Δxyr1*).
(A) Cultivo em presença de celulose; (B) em presença de soforose e (C) em glicose. A escala de expressão está representada em Log_2. *p < 0,05; **p < 0,01 e ***p < 0,001.

A comparação entre *Δxyr1*/QM9414 (Figura 33A) revelou que *xyn3* e *cel5* tiveram *Fold Change* ≥ 2 nas três condições analisadas, enquanto que nenhum gene apresentou expressão aumentada nas duas condições indutoras (celulose e soforose). Três genes tiveram aumento de expressão em celulose (*cel45a*, *cel3e* e *xyn2*) e soforose (*cel1b*, *cel3b* e *cel3c*). Curiosamente, 10 genes (*cel7a*, *cel6a*, *cel3a*, *cel7b*, *cel5a*, *cel12a*, *cel74a*, *xyn1*, *xyn4* e *swo*), a maioria deles codificando celobiohidrolases e endoglucanases, tiveram um *Fold Change* ≥ 2 de expressão em presença de glicose, sugerindo uma forte relação entre XYR1 e CCR. Logo, em *Δxyr1*/QM9414, a maioria dos genes reprimidos, 10 no total, foram identificados na presença de celulose/soforose. Os genes *cel1b*, *cel3b* e *cel3c*, que codificam β-glicosidases, foram reprimidos em celulose. Além disso, *xyn1* e *cel3e* foram reprimidos na presença de soforose e glicose, respectivamente (Figura 33B).

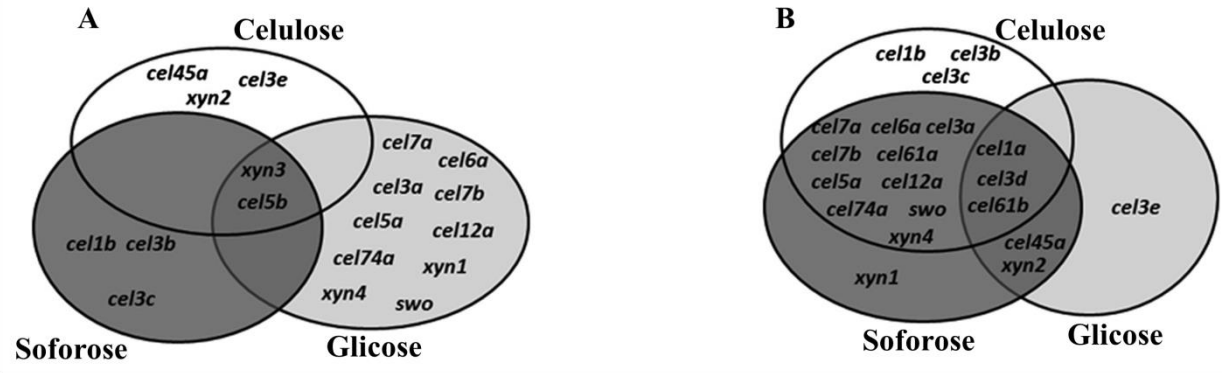


FIGURA 33 – Diagrama de *Venn* de genes celulolíticos e xilanolíticos diferencialmente expressos em *T. reesei* (*Δxyr1*/QM9414).
(A) Genes que apresentaram significativo aumento e (B) diminuição no nível de expressão em *Δxyr1*/QM9414. Para cada condição foi considerado um aumento de expressão independente do tempo de cultivo e Log2 *Fold Change* ≥ 2 e ≤ -2.

5.1.2 Análises *in silico* de putativos sítios de ligação para XYR1 em regiões promotoras dos 22 genes celulolíticos e xilanolíticos

Para estabelecer uma possível relação entre a expressão de genes celulolíticos e xilanolíticos e o número de putativos sítios de ligação de XYR1, foi verificada a presença de motivos de ligação na região *5'-upstream* (1,5 kpb) dos 22 genes celulolíticos e xilanolíticos analisados. As diversas variações da 5'-GGCWWW-3', do motivo de ligação de XYR1, foram utilizadas para comparar a frequência da presença deste motivo na região promotora dos 22 genes regulados por XYR1, e, com os demais genes do genoma de *T. reesei*. Os

motivos 5'-GGCTAA-3', 5'-GGCTAT-3' e 5'-GGCAAA-3' foram encontrados a 5,3-; 2,3-; e 1,6-vezes em maior frequência em genes regulados por XYR1 quando comparados com outras ORFs. Desta maneira, foram identificados 78 locais de ligação simples e 16 locais dispostos como uma sequência de repetição invertida na região promotora dos 22 genes analisados (Figura 34).

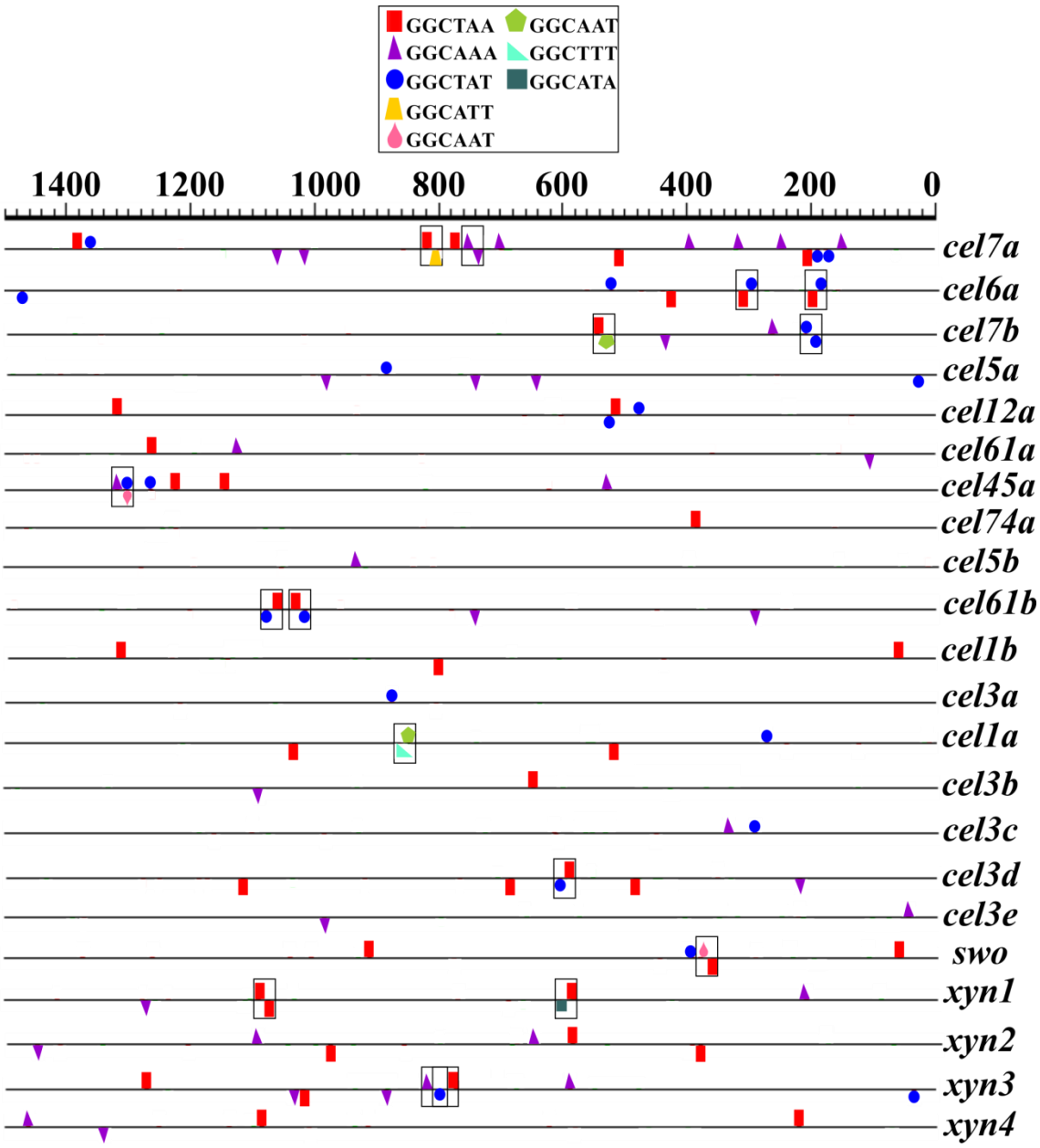


FIGURA 34 – Representação esquemática de putativos sítios de ligação de XYR1 (5'-GGCWWW-3') na região 5'-*upstream* de genes celulolíticos e xilanolíticos.
A posição dos motivos é relativo ao códon ATG (início da tradução). Os símbolos acima e abaixo da linha indicam motivos de ligação no sentido *sense* e *antisense* da fita de DNA. Repetições invertidas de sítios de XYR1 são destacados nos retângulos e sítios únicos são identificados como na legenda da figura. W = A/T, de acordo com normas da IUPAC.

Os genes que apresentaram maior frequência de putativos sítios de ligação de XYR1 foram *cel7a*, *xyn3*, *cel6a* e *cel45a* (18, 9 e 7 sítios, respectivamente, Figura 34). Entre eles, *cel7a*, *cel6a* apresentaram aumento de expressão em QM9414, quando cultivados em celulose (Figura 32A) e soforose (Figura 32B) e tiveram expressão reduzida em presença de glicose (Figura 32C), enquanto *cel45a* e *xyn3* apresentaram redução na expressão em presença de celulose (Figura 32A) e soforose (Figura 32B), e *cel45a* apresentou aumento de expressão em presença de glicose (Figura 32C). Resultados opostos foram encontrados para o mutante *Δxyr1* quando os mesmos genes foram analisados (Figura 32A-C). Entre os genes que apresentaram baixa frequência de motivos de ligação de XYR1 estão *cel74a* e *cel3a*, os quais apresentaram um comportamento de expressão semelhante ao gene *cel7a*. Para cada condição analisada, os genes foram agrupados em diferentes *clusters*, sugerindo uma dependência entre XYR1 e a fonte de carbono na regulação da expressão destes genes celulolíticos e xilanolíticos (Figura 32A-C). Curiosamente, no mutante *Δxyr1*, os genes *cel1b* e *cel3e* também apresentaram alta expressão em celulose, soforose, e glicose em comparação com QM9414 (Figura 32A-C), sugerindo a cooperação entre XYR1 e CRE1 na regulação destes genes. Os resultados das análises *in silico* não forneceram nenhuma evidência de uma correlação direta entre a expressão de genes celulolíticos e xilanolíticos e o número de putativos sítios de ligação de XYR1.

5.1.3 Determinando a arquitetura do elemento regulatório cis para XYR1

Com os resultados obtidos, e utilizando o software MEME, juntamente com os dados de RNA-seq de *T. reesei* crescido em presença de celulose e soforose (Capítulo II) não foi encontrado nenhum motivo de ligação de XYR1 que fosse semelhante a sequências consenso conhecidas. Porém, ao utilizar os dados dos 22 genes de celulases dependentes de XYR1 (CASTRO et al., 2014a), nestes genes foi possível identificar dois motivos denominados Celulases – Motivo 1 e Celulases – Motivo 2 (Figura 35) apresentando uma região altamente conservada (GGC) seguida por uma região rica em A e T (FURUKAWA et al., 2009). Cada um desses dois motivos foi encontrado na região promotora dos 22 genes analisados, e divergiram principalmente nas bases que foram conservadas na extremidade 3' (Figura 35A). Em seguida, as sequência desses motivos (Celulases – Motivos 1 e 2) foram fundidas para

gerar o consenso que, potencialmente, representa o sítio de ligação para o fator de transcrição XYR1 em *T. reesei* (Figura 35B).

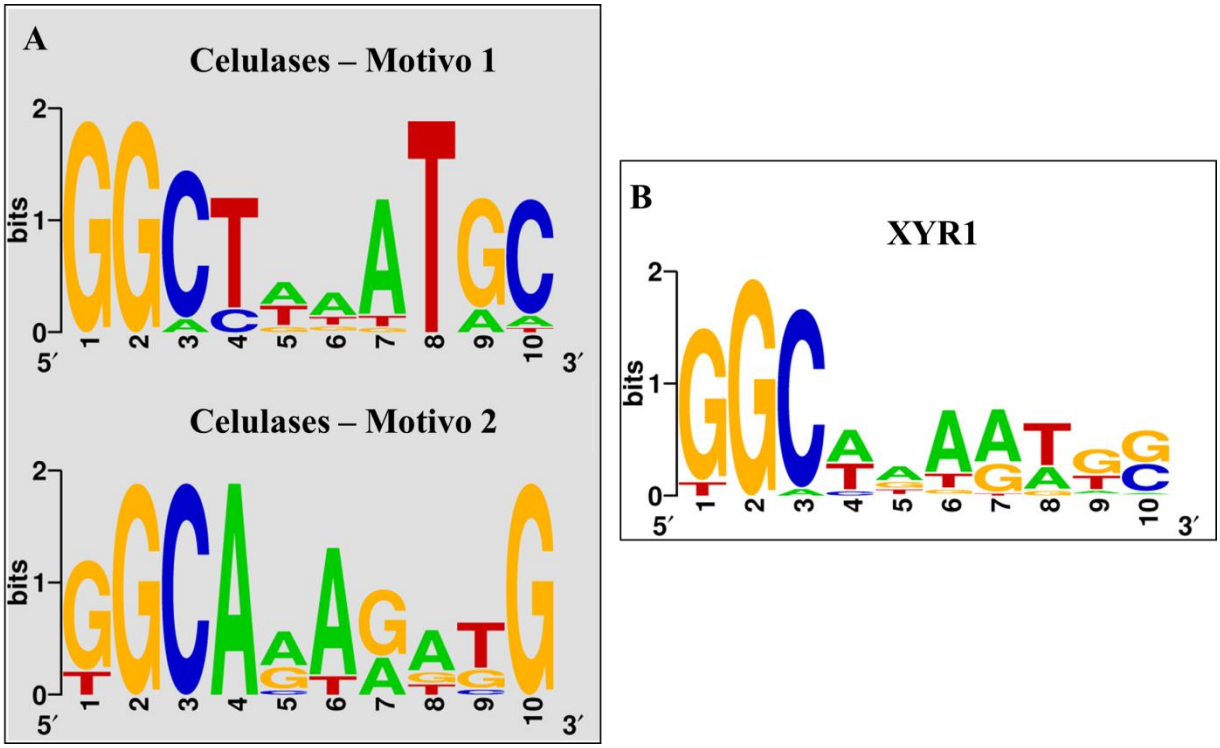


FIGURA 35 – Motivos de ligação de XYR1 identificados em celulases.
(A) Dois motivos identificados em 22 celulases (região promotora) que possui semelhança ao consenso de XYR1 conhecido na literatura (5'-GGCWWW-3'). (B) Combinação do Motivo 1 e Motivo 2 para criar o motivo consenso de XYR1 utilizado nesta pesquisa para localizar os sítios de ligação em *T. reesei* (denominado PWM_{XYR1}).

O motivo identificado na Figura 35B apresenta a primeira representação de alta resolução do sítio de ligação XYR1 reconhecido em *T. reesei*, e é uma ferramenta valiosa para investigar regulons neste organismo. Desta maneira, depois de definir o suposto sítio de ligação para o fator de transcrição XYR1, este foi utilizado para decifrar como este regulador reconhece seus promotores alvos. Utilizando a pontuação obtida por meio do *Position Weight Matrixes* (PWM), que representa o consenso de ligação de XYR1 de acordo com Hallikas et al., (2006), e acessando a pontuação de locais identificados na escala do genoma e sobre os conjuntos de dados específicos (ou seja, o grupo de 22 celulases), identificou-se apenas o melhor *hit* por promotor de acordo com o PWM.

Como apresentado na Figura 36A, o suposto motivo para XYR1, quando analisado no genoma, seguiu uma distribuição normal, com os *scores* variando de, 6,2 a 7,3 e um pico próximo de 6,7. No entanto, quando a região promotora dos 22 genes de celulases foram

analisadas, verificou-se uma distribuição muito fechada com dois picos, um perto de 6,7 e o outro cerca de 7 (Figura 36B). Na verdade, essa análise revela que quase metade dos promotores de celulases são dotados de um sítio com *score* superior a 7, enquanto uma parte, consideravelmente menor de todo genoma, apresenta sítios acima deste *score*.

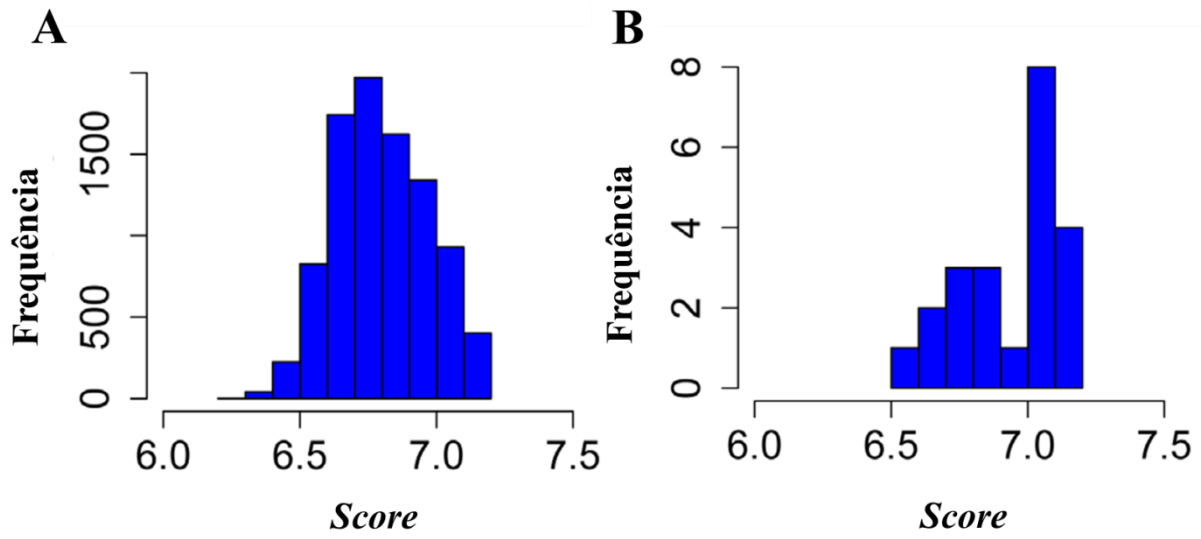


FIGURA 36 – Distribuição do *Score* do sítio de ligação de XYR1 em diferentes conjuntos de dados (região promotora).
(A) No genoma de *T. reesei*. (B) Na região promotora de 22 celulases.

Desse modo, os resultados, apresentados na Figura 36, revelaram que *score* eram muito abundantes no nível genômico, portanto, foi investigado sítios de ligação dispostos em arquiteturas específicas propostas anteriormente (AMORE et al., 2013; FURUKAWA et al., 2009; LING et al., 2009; RAUSCHER et al., 2006; SHIDA et al., 2008), isto é, sítios de ligação de XYR1 foram focados em repetições invertidas (para dentro) e evertidas (para fora) localizados a uma distância entre 8 a 30 nucleotídeos de um sítio a outro (Figura 37A). Em seguida, procurou-se sítios que cumpriram este requisito e utilizou-se diferentes *thresholds* em termo de *score* para o sítio de XYR1 (variando de 6,1 a 6,6). Como apresentado na Figura 37B, e como esperado, o aumento no valor do *threshold* gerou a identificação de menos sítios por promotores em ambos os grupos celulases e genoma. Porém, quando utilizado um valor alto de *threshold*, foi observado um enriquecimento de sítios nos promotores de celulases em comparação ao genoma de *T. reesei* (Figura 37B). Com destaque, a utilização desse valor alto de *threshold* foi para o gene *cel7a*, que codifica um dos mais abundantes genes de celulases produzidos por *T. reesei* (AMORE et al., 2013).

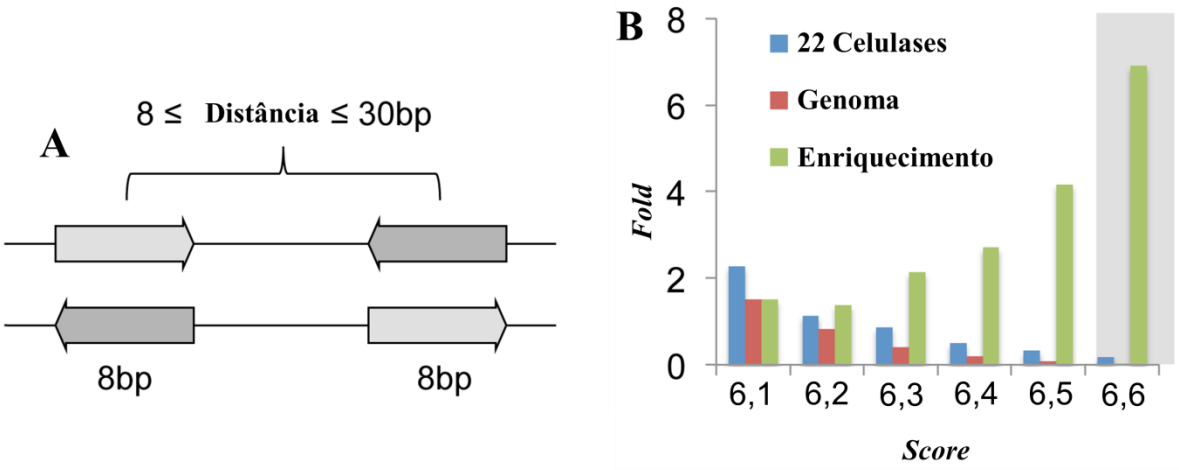


FIGURA 37 – Sítio de ligação de XYR1 em diferentes conjuntos de dados (região promotora).
(A) Em XYR1, ambos sítios invertidos e evertidos foram considerados e somente sítios dentro da distância de 8 a 30 pb. (B) Representação dos sítios de ligação repetidos nos promotores de celulases e dos genes do genoma de *T. reesei*. O eixo y do gráfico (*Fold*) representa o número de sítios identificados relativos ao número de promotores dos conjuntos de dados. O grupo enriquecimento representa a taxa entre sítios por promotores dos promotores de celulases e o correspondente valor do genoma. A área sombreada em cinza destaca a região de maior enriquecimento.

5.1.4 Inspeção de promotores de celulases por sítio de ligação de XYR1

Uma vez definida a arquitetura do elemento regulatório cis de XYR1, procurou-se a presença deste elemento nos promotores de genes que codificam celulases. Como representado na Figura 38, utilizando um valor de *threshold* de 6,4 para o sítio de ligação de XYR1, pôde-se verificar que, nesta figura, foi identificado este sítio em 10 promotores de celulases, sendo que a maioria destes promotores apresentaram sítios simples, e apenas o promotor do gene *cel7b* apresentou dois sítios de ligação, para o fator XYR1, que apresentam poucos espaços entre os dois sítios e foi localizado a 600 pb na região *upstream* antes do códon de iniciação ATG (Figura 38A). Além disso, a maioria dos promotores identificados (9 de 10) apresentaram o putativo sítio de ligação para XYR1 localizado a menos de 1 kpb do códon ATG. Outro achado interessante foi que três promotores (*cel7a*, *cel6a* e *cel7b*), considerados enzimas eficientes produzidas por *T. reesei*, apresentaram putativos sítios duplos de ligação XYR1 que foi detectado usando a abordagem desta pesquisa.

A comparação do sítio de ligação XYR1, encontrado *in silico* com aqueles caracterizados anteriormente (*in vivo*) em alguns promotores de celulase, mostrou um notável nível de concordância, isto é, no promotor do gene *xyn1* (RAUSCHER et al., 2006), no promotor do gene *cel7a* (RIES et al., 2014) e no promotor do gene *cel6a* (ZEILINGER et al.,

1998). No caso do promotor *xyn1*, um GGCTAA-box formado por duas repetições invertidas do elemento GGCWWW e localizado na posição –410 foi encontrado por ser requerido para interação *in vivo* e *in vitro* (RAUSCHER et al., 2006). No caso do promotor *cel7a*, dois potenciais sítios de ligação têm sido propostos como funcionais, na posição –320 e –733, mas nenhuma evidência direta de sua função é conhecida (RIES et al., 2014). As análises *in silico* do promotor *cel7a* revelou um sítio de repetição direta com alto *score*, que inclui na sua região 3', o sítio –733 anteriormente relatado. Finalmente, o promotor do gene *cel6a*, representa o exemplo de maior interesse. Para exemplificar, a Figura 38B sumariza a arquitetura putativa do promotor do gene *cel6a*, incluindo o potencial sítio de ligação XYR1 (denominado de UAS_{XYR1}) e os elementos regulatórios cis caracterizados anteriormente CAE[12] e TATA-box (ZEILINGER et al., 2003; ZEILINGER et al., 1998).

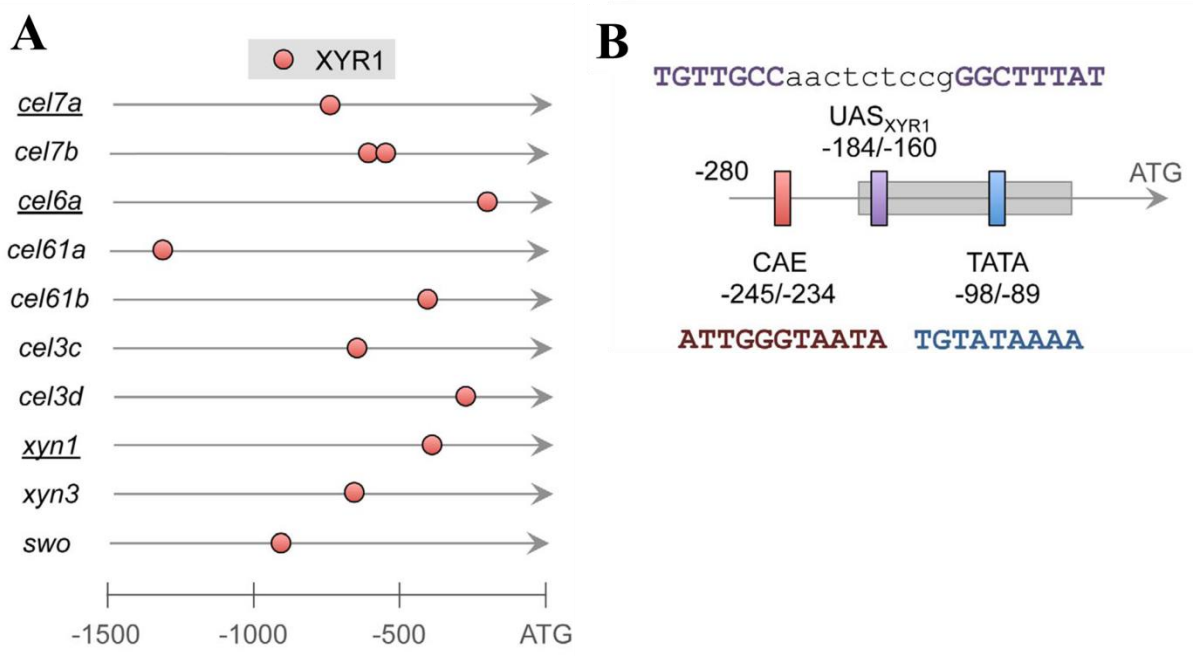


FIGURA 38 – Identificação de sítios de ligação XYR1 nos promotores alvos.
(A) Representação dos sítios de ligação XYR1 encontrados na região promotora de 10 genes que codificam celulases. (B) Região promotora do gene *cel6a*, apresentando o CAE (barra vertical em vermelho), o TATA-box (barra vertical em azul), a região que se liga ao nucleosoma 1 (retângulo horizontal em cinza) e o sítio de ligação para XYR1, rotulado como UAS_{XYR1}. A sequência de cada elemento regulatório é apresentado em (ZEILINGER et al., 2003). As posições apresentadas do sítio de ligação de XYR1 são relativos ao códon de iniciação ATG do gene *cel6a*.

[12] CAE = Elemento de ativação por *cbh2*/*cel6a* (ZEILINGER et al., 2003).

5.1.5 Identificação de potenciais genes alvos de XYR1

Para esta análise, uma lista de 9115 genes de *T. reesei* anotados com suas regiões promotoras, foram analisados com os mesmos critérios definidos na Figura 37. Utilizando um *threshold* rigoroso de 6,6 para o sítio de ligação XYR1, foram identificados 233 genes potencialmente regulados por este fator (CD Anexo, Arquivo 4) – Predição de todo o genoma de sítios de ligação de XYR1 em *T. reesei*). A próxima etapa obtida foi a categorização funcional (KOG) de todos os genes identificados (Figura 39). Como apresentado na Figura 39, as categorias funcionais em destaque foram: *metabolism carbohydrate and amino acid*; *chromatin structure and dynamics*; *RNA processing and modification*; *translation* e entre outros.

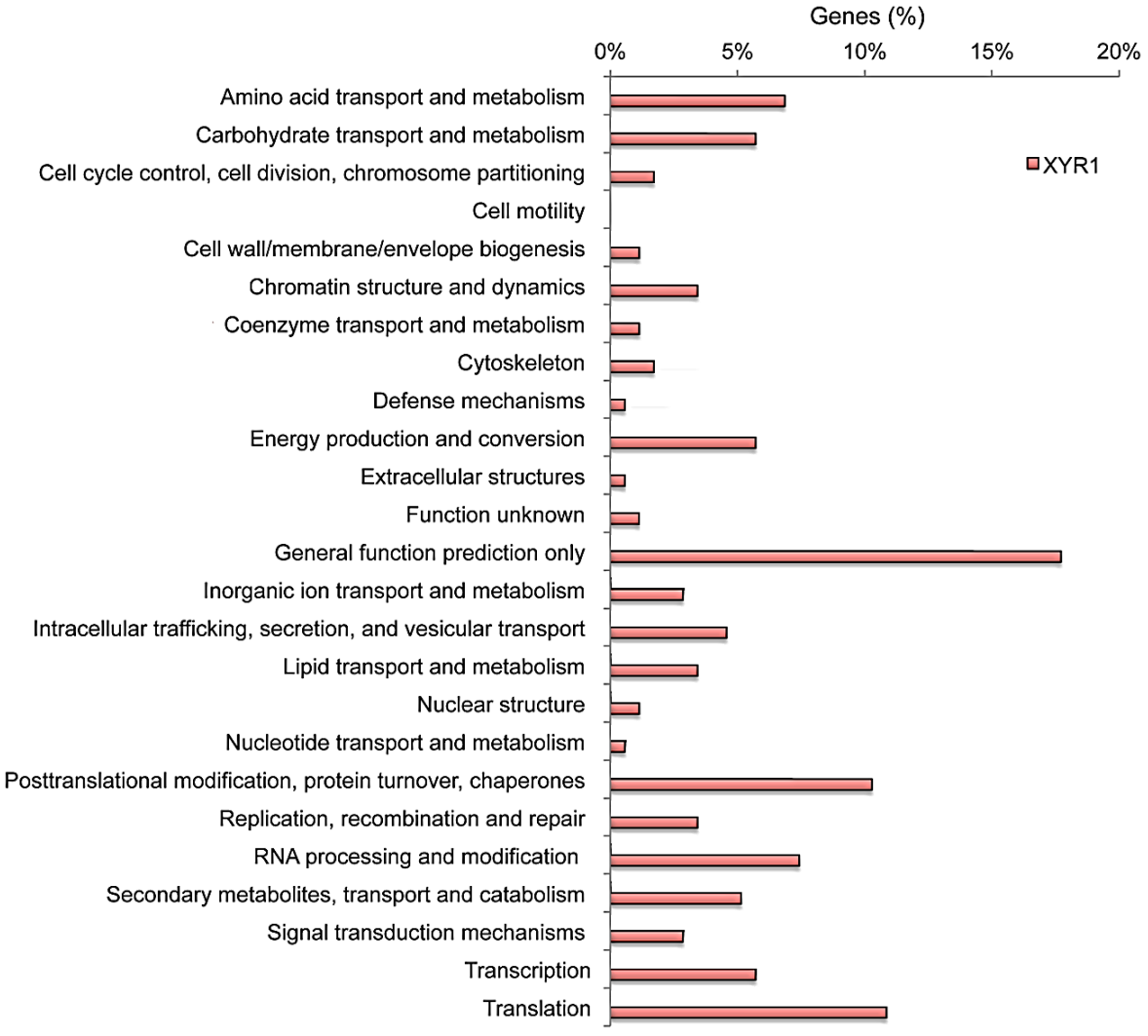


FIGURA 39 – Definindo as categorias dos genes potencialmente regulados diretamente por XYR1.

Desse modo, uma melhor compreensão da abrangência funcional dessas proteínas requer a integração de dados de expressão *in vivo*. Para isso, sítios de ligação XYR1 foram identificados nas regiões promotoras dos genes up- e down-regulados na linhagem mutante *T. reesei* (*Δxyr1*) comparado à linhagem parental crescidos em presença de celulose e soforose (Capitulo III). Usando esses conjuntos de dados, foi encontrado que 15,8% dos genes diferencialmente regulados apresentaram putativos sítios de ligação XYR1. As listas dos genes diferencialmente expressos utilizadas nesta análise com seus respectivos valores de expressões determinados por RNA-seq podem ser verificadas no CD Anexo (Arquivos 5 a 8).

5.2 DISCUSSÕES

Devido a sua habilidade para produzir uma grande quantidade de celulases e xilanases, *T. reesei* tem sido extensivamente estudado para utilização na produção de etanol celulósico (PORTNOY et al., 2011a). Neste estudo, foi demonstrado que os genes de celulases e xilanases respondem diferentemente, dependendo da fonte de carbono disponível e também do tipo de regulação, neste caso pelo fator transcricional XYR1.

Os resultados relacionados ao mutante *Δxyr1*, em geral, apresentaram uma redução na expressão de genes de celulases e xilanases quando comparado à linhagem parental QM9414 cultivados em presença de celulose e soforose (Figura 32A-B). Estes achados corroboram com os dados de Furukawa et al., (2009), que também observaram uma significante redução na expressão de genes de celulases e xilanases quando utilizou a linhagem mutante PC-3-7 em presença de L-sorbose ou soforose. Ao contrário do que foi observado nas condições indutoras, aproximadamente 60% dos genes analisados tinham aumento de expressão no mutante *Δxyr1* quando comparado ao parental, em presença de glicose nos tempos de 24 e 48 horas (Figura 32C). Além do fator de transcrição XYR1, o fator CRE1 e outros dois fatores (ACE1 e ACE2) são conhecidos por regular a expressão de genes de celulases e xilanases, por reprimi-los ou ativá-los (RAUSCHER et al., 2006). Uma explicação plausível para esta observação é que pode ser uma expressão constitutiva parcial de ACE2, uma vez que, o *xyr1* é sujeito a CCR, *Ace2* não é, reforçando a explicação para estes resultados obtidos (PORTNOY et al., 2011b).

Além das celulases clássicas, novos tipos de celulases tem sido recentemente descritas em *T. reesei*, incluindo solenina (SWO) (SALOHEIMO et al., 2002). SWO é

positivamente regulada tanto em celulose como em soforose e negativamente regulada em presença de glicose (VERBEKE et al., 2009). Os resultados encontrados sugerem que este gene *swo* pode ser alvo de regulação por XYR1, independente da fonte de carbono.

Sabe-se que hidrolases atuam sinergicamente na degradação de substratos poliméricos, tais como: celulose e xilana. Foi relatado, em estudos anteriores, que *cel3a* e *bxl1* são regulados por XYR1 na presença de soforose, mas *cel1a* não é regulado (STRICKER et al., 2008). Neste estudo, os genes *cel3a* e *cel1a* foram diferencialmente regulados e essa regulação foi dependente não somente de XYR1, mas também da fonte de carbono. No entanto, β-glicosidases requerem um fator de transcrição específico (BglR) para esta expressão (NITTA et al., 2012).

Genes que codificam xilanases também parecem ser dependentes do regulador XYR1 (Figura 33). Herold e outros pesquisadores (2013) verificaram que a maioria das xilanases podem ser induzidas em presença de D-xilose (com exceção da *xyn3*) e que, em presença de L-arabinose, induzem as xilanases *xyn1*, *xyn4* e *xyn5*. Seja qual for a maneira da regulação da expressão de celulases e xilanases, os dados obtidos sugerem que essa regulação pode ocorrer direta ou indiretamente por um fator ou um conjunto de fatores de transcrição e que esse processo é dependente da fonte de carbono.

As análises referentes ao suposto sítio de ligação de XYR1 ajudou a decifrar como esse regulador reconhece os seus promotores alvos. Embora vários trabalhos têm tentado compreender esse processo em escala global, um modelo unificado para interação proteína-DNA para esse regulador ainda não está disponível (AMORE et al., 2013; GLASS et al., 2013a). Por exemplo, alguns trabalhos relatam que os sítios de ligação funcionais de XYR1 tem sido caracterizados por repetições invertidas (RAUSCHER et al., 2006; FURUKAWA et al., 2008). Como alternativa, Furukawa e colegas (2009) sugeriram que os promotores dos genes regulados por XYR1 possuem uma maior quantidade de sequências referentes a sequência consenso GGCWWW quando comparado com o restante do genoma. Portanto, o motivo gerado neste estudo representou uma alta resolução na descrição do sítio de ligação para o regulador XYR1.

Os resultados apresentados na Figura 37 mostraram que a procura por sítios duplos de ligação permitiu uma melhor definição dos alvos de XYR1. Vale a pena ressaltar que a disposição dos locais de ligação pareceu ser mais importante do que o PWM, uma vez que os *thresholds*, próximos dos valores médios na distribuição do genoma, ainda forneceram um alto enriquecimento nos promotores alvo (Figura 37). Todos os dados apresentados sugerem

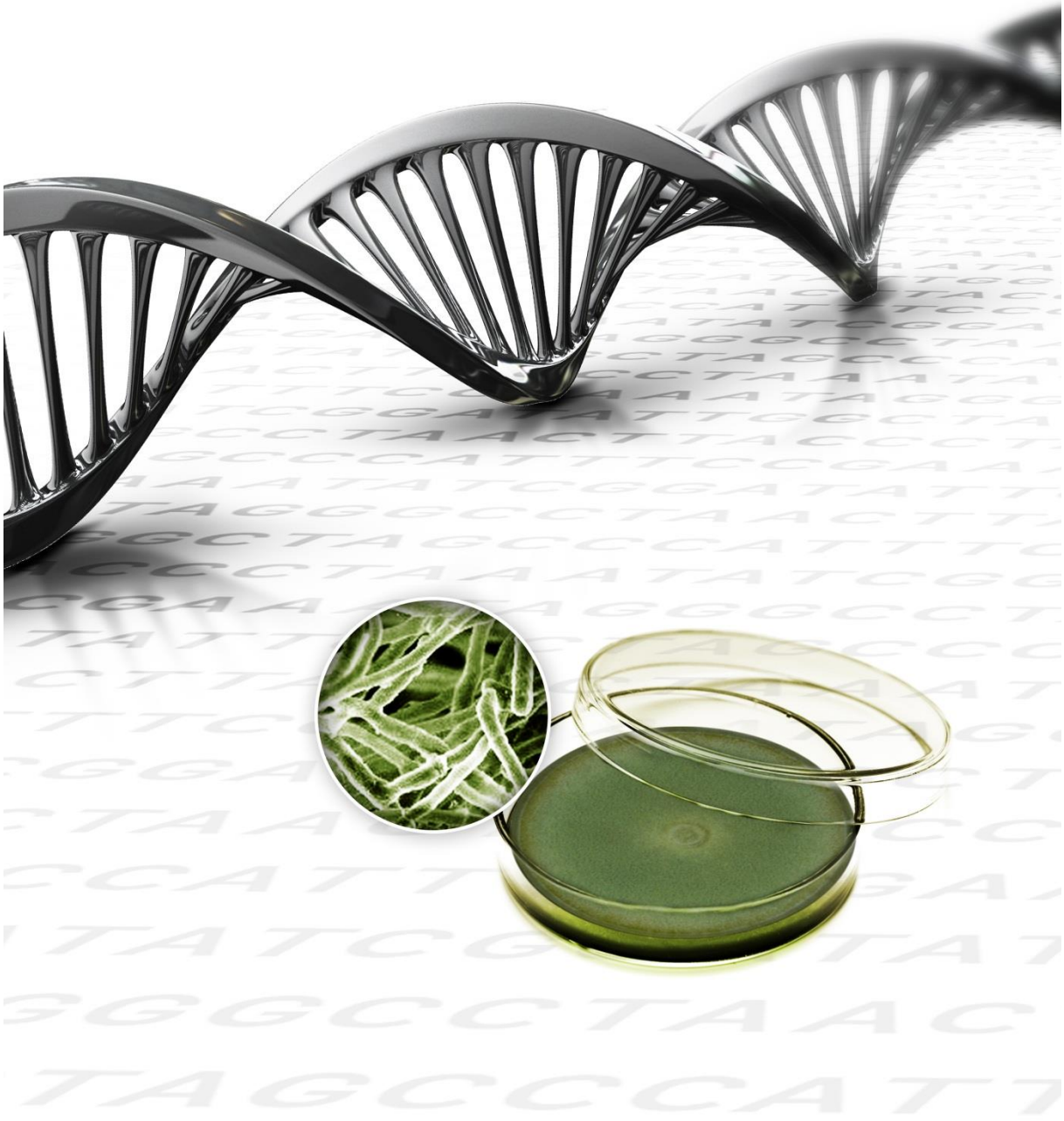
que os sítios duplos de ligação são mais relevantes para o reconhecimento dos promotores-alvo por XYR1 do que os sítios individuais.

É conhecido que os genes que codificam celulases são controlados em nível transcricional por indução na presença de substratos tais como celulose e soforose e a repressão mediada por uma fonte de carbono facilmente metabolizável como a glicose. Considerando a participação dos reguladores XYR1 e CRE1 tem sido caracterizada pela indução e repressão nos promotores de celulases, respectivamente, continua a questão se os promotores são regulados diretamente por estes reguladores ou se existe mecanismos indiretos envolvendo outros fatores de transcrição desconhecidos (AMORE et al., 2013; (GLASS et al., 2013a).

Evidências *in vivo* e *in vitro* para interações direta entre XYR1 e CRE1 foram relatadas para os promotores dos genes *cel7a* (LING et al., 2009; STRICKER et al., 2006), *xyn1* (HEROLD et al., 2013; STRICKER et al., 2006; RAUSCHER et al., 2006) e *cel6a* (ZEILINGER et al., 2003; GORSCHE et al., 2014; STANGL et al., 1993), sendo este último destacado como exemplo nas análises. Uma vez que, análises anteriores revelaram a existência da região CAE entre as posições –234 e –245 que é essencial para a indução deste promotor em resposta à presença de celulose e soforose (ZEILINGER et al., 1998), e esta região foi encontrada sendo regulada pelo complexo HAP2/3/5 e por uma proteína ainda não caracterizada (ZEILINGER et al., 2003). Esses autores ainda estudaram a ocupação nucleossoma do promotor *cel6a* e descobriram que o CAE está localizado numa região livre e que as proteínas de ligação a este elemento controlam a montagem de um nucleossoma (denominado nucleossoma 1) abrangendo uma região de –192 a –49 que inclui o TATA-box. Os mesmos autores ainda propuseram um modelo de indução em que a remoção do nucleossoma 1 é necessária para permitir a interação de proteínas com o TATA-box para seguir a indução do promotor *cel6a* em resposta à celulose e soforose. No entanto, a sequência alvo do regulador XYR1 ainda não tinha sido identificada, embora os dados disponíveis sugerem que a ligação ocorra em algum lugar *downstream* da região CAE (ZEILINGER et al., 2003; ZEILINGER et al., 1998). A busca por sítios de ligação XYR1, usado neste estudo, permitiram a identificação de uma repetição evertida no promotor *cel6a* na posição –160 a –184, concordando perfeitamente com a informação atualmente disponível sobre a regulamentação deste promotor. Vale a pena ressaltar que os locais de ligação evertidas são reconhecidos por outras proteínas *zinc finger* (MACPHERSON et al., 2006), mas não tem XYR1 associado a esses elementos até o momento.

As análises, em conjunto, sugerem que os sítios identificados neste estudo representam uma alta confiança do sítio de ligação para XYR1 em *T. reesei*. Enquanto que as análises em relação à Figura 39 fornecem pistas sobre o potenciais alvos de XYR1, deve-se notar que não fornece uma descrição completa de seus regulons desde que fatores transcricionais adicionais poderiam mediar regulação indireta nos promotores alvos (AMORE et al., 2013; GUELZIM et al., 2002). Portanto, as últimas análises, com dados dos experimentos do RNA-seq (*Δxyr1*/QM9414), fortemente sugerem que a regulação indireta desempenha um papel importante no controle de genes alvos por XYR1 em *T. reesei*.

CAPÍTULO II



6 CAPÍTULO II

6.1 RESULTADOS

6.1.1 Perfil global da expressão gênica de *T. reesei* cultivado em diferentes fontes de carbono

T. REESEI QM9414 foi cultivado em presença de celulose, soforose e glicose como descrito (ver materiais e métodos). Nove bibliotecas foram sequenciadas, usando o equipamento *Illumina Hiseq*^{TM} *2000*, gerando aproximadamente 117 milhões de *reads paired-end* de 100 pb correspondendo a 23,32 GB de dados (Tabela 13). As *reads* foram mapeadas no genoma de referência *T. reesei* QM6a disponível no JGI (*Trichoderma reesei* v2.0), usando o algoritmo *Bowtie*. Em geral, 68% das *reads* foram mapeadas no genoma de referência (Tabela 13).

Tabela 13 – *Reads* obtidas por RNA-seq (*Illumina HiSeq*^{TM} *2000*) para a linhagem parental QM9414.

Amostras	***Reads* brutas**	***Reads* de qualidade**	***Reads* Mapeadas**	**% *Reads* Mapeadas**	**Número de nucleotídeo (Gigabase/GB)**
QM9414 Cel1	28483604	24284506	16951434	69,80	3,39
QM9414 Cel2	16036616	14420917	10172333	70,54	2,03
QM9414 Cel3	19224967	17304525	11683287	67,52	2,34
QM9414 Sof1	14311488	11341066	8445229	68,83	1,69
QM9414 Sof2	17782393	33172592	11041736	68,34	2,21
QM9414 Sof3	19691061	23138683	12645886	70,86	2,53
QM9414 Gli1	13341102	12269998	7814715	68,91	1,56
QM9414 Gli2	36490949	16157814	22484949	67,78	4,50
QM9414 Gli3	25600303	17845732	15366090	66,41	3,07
Total	190962483	169935833	116605659	68,78	23,32

Foram utilizadas três réplicas biológicas de cada condição para as análises transcricionais, e foram obtidas altas correlações entre as réplicas (Figura 40), indicando que

as amostras apresentam boa confiabilidade e, assim, foram consideradas satisfatórias e utilizadas em análises subsequentes. Além disso, criou-se um modelo matemático por meio da Análise de Componentes Principais (PCA), uma vez que esta ferramenta permitiu não apenas uma visualização adequada dos dados, como também uma modelagem matemática. Nessa análise ficou muito clara a separação dos grupos quando a técnica PCA foi aplicada as amostras (Figura 41).

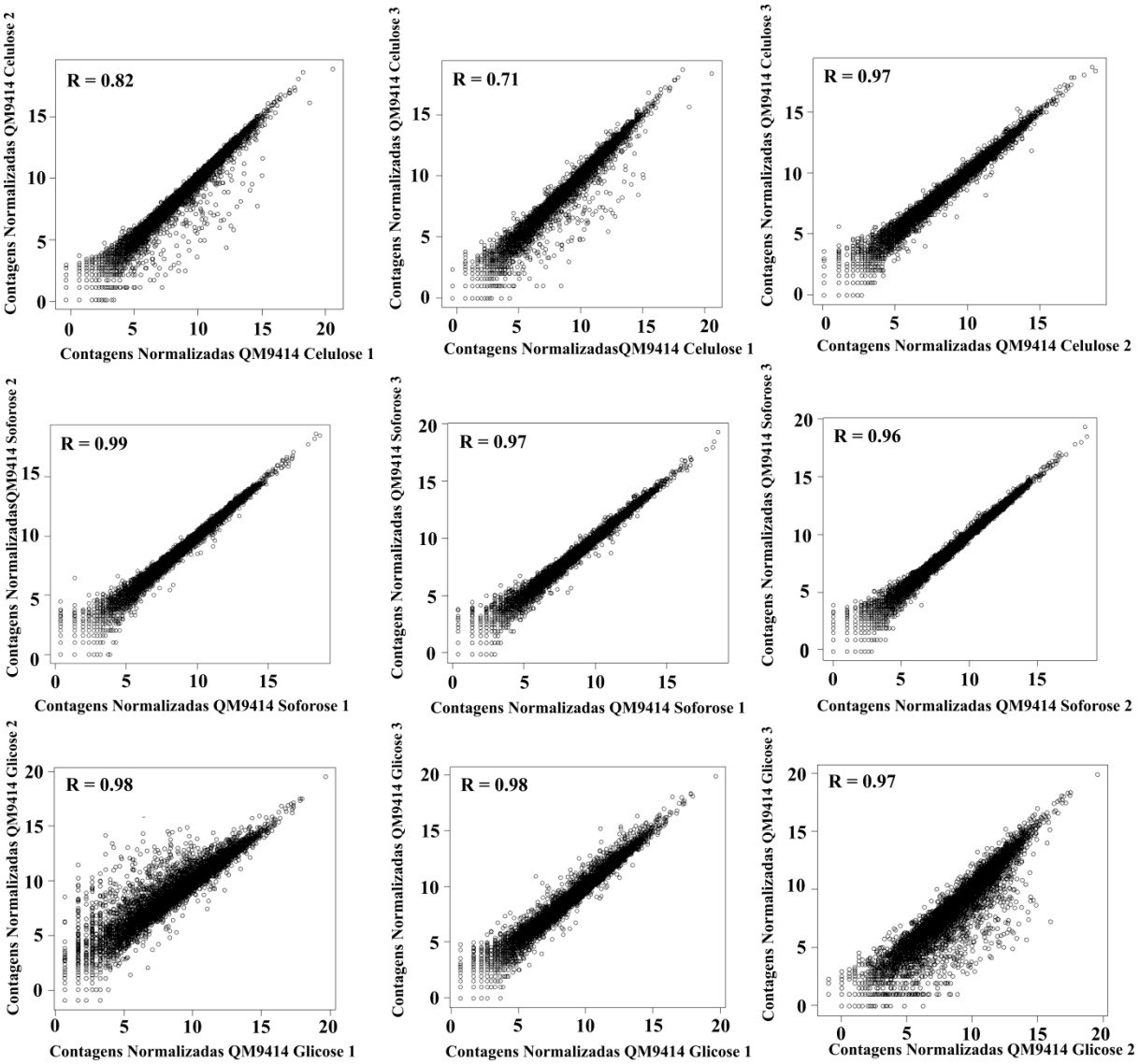


FIGURA 40 – Réplicas biológicas das amostras celulose, soforose e glicose para a linhagem parental utilizadas nas análises de RNA-seq.
Gráficos representando a Correlação de *Pearson* entre as réplicas biológicas de cada amostra.

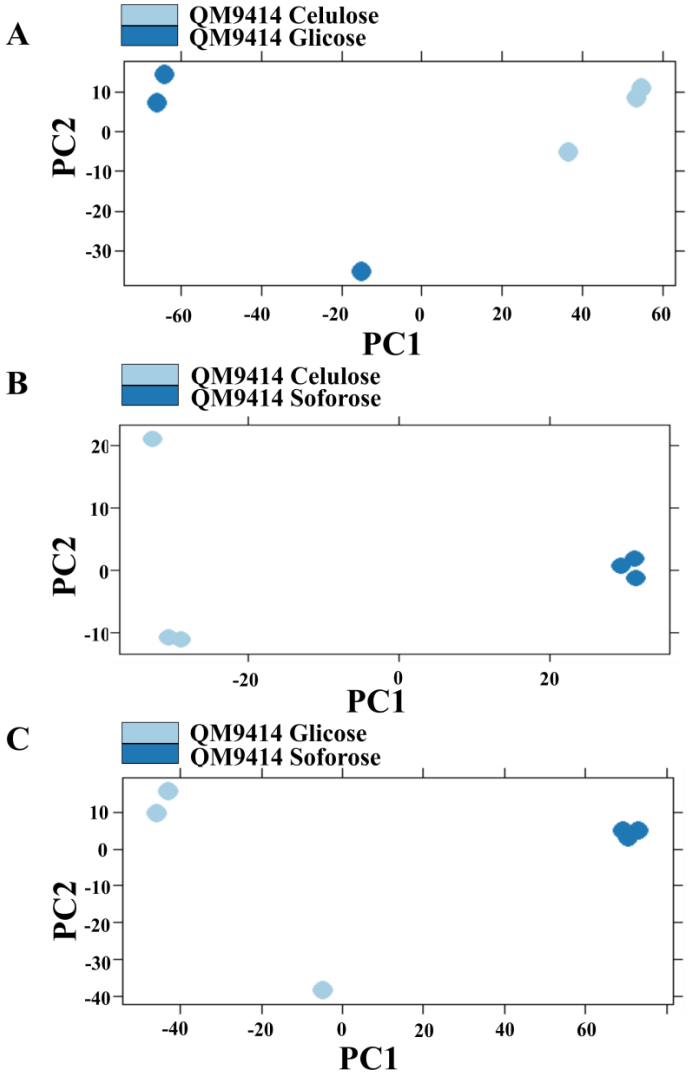


FIGURA 41 – Análise de Componentes Principais (PCA) das amostras analisadas na linhagem parental. (A) QM9414 Celulose/Glicose; (B) QM9414 Soforose/Celulose e (C) QM9414 Soforose/Glicose.

Para comprovar que entre as condições utilizadas e seus respectivos tempos de cultivos são amostras comparáveis entre si, gráficos *box plots* (dados brutos e após a etapa de normalização dos dados do RNA-seq) foram construídos e os resultados podem ser visualizados na Figura 42. Portanto, os dados normalizados foram utilizados para obtenção do perfil de expressão gênica de *T. reesei* obtidos de diferentes fontes de carbono, analisados pelo *R* e *Bioconductor* (pacote *DESeq*). Dos 9129 genes de *T. reesei*, 1788 genes foram identificados como sendo diferencialmente expressos ($p \leq 0,05$) em celulose/glicose, 2545 genes em soforose/celulose e 2481 genes em soforose/glicose (Figura 43). Com esses dados, gráficos *volcano plot* foram construídos para facilitar a visualização dos genes up e down-regulados do total de genes diferencialmente expressos (Figura 44).

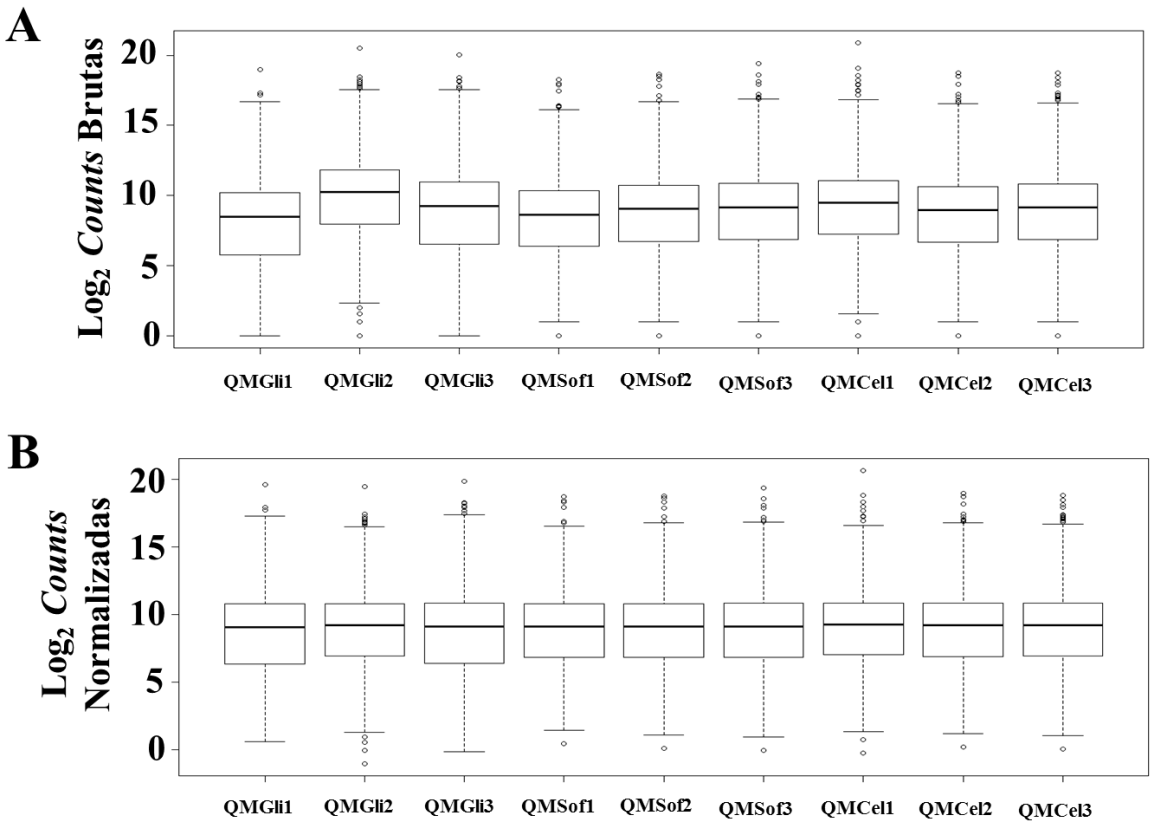


FIGURA 42 – *Box plot* das amostras utilizadas nas análises bioinformáticas dos dados do RNA-seq na linhagem parental.
(A) Dados brutos; (B) Dados normalizados.

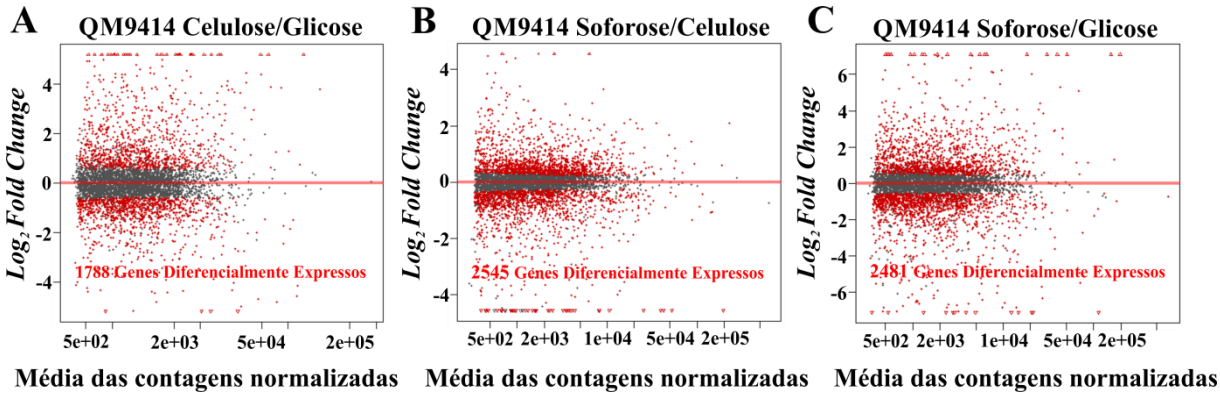


FIGURA 43 – Comparação dos perfis de expressão gênica da linhagem parental QM9414 cultivado em celulose, soforose e glicose como fontes de carbono.
(A) Celulose/Glicose; (B) Soforose/Celulose e (C) Soforose/Glicose. Genes diferencialmente expressos encontrados através do pacote DESeq estão representados em vermelho.

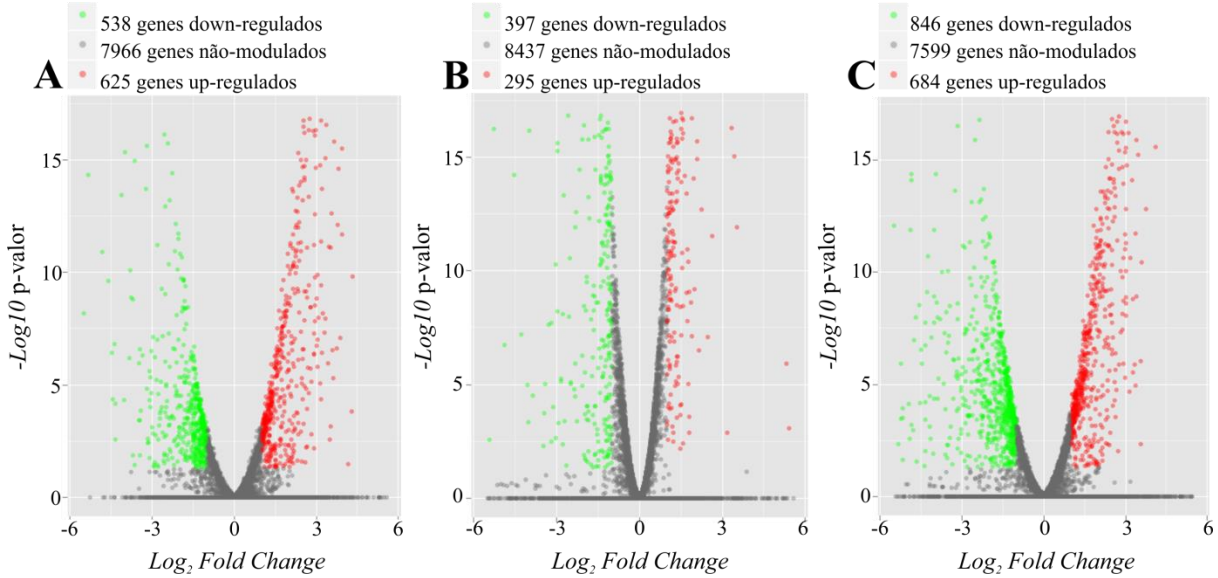


FIGURA 44 – Gráficos *volcano plot* dos genes diferencialmente expressos up- e down-regulados nas condições estudadas.
Log$_2$ *Fold Change* e seu correspondente *p-valor* (em *Log10*) de todos os genes significativos e regulados positivamente e negativamente foram utilizados para construir estes gráficos. (A) QM9414 Celulose/Glicose; (B) QM9414 Soforose/Celulose e (C) QM9414 Soforose/Glicose. Genes up-regulados (*Log*$_2$ *Fold Change* ≥ 1 e *p-valor* 0,05) estão destacados em pontos vermelhos e genes down-regulados (*Log*$_2$ *Fold Change* ≤ -1 e *p-valor* 0,05) estão em verde. Os genes não significativos estão em cinza.

A Figura 45A destaca os Diagramas de *Venn* das condições estudadas, apresentando 1886 genes que foram diferencialmente expressos em celulose e glicose, sendo 703 genes e 491 genes up- e down-regulados exclusivamente em glicose, e 254 e 102 genes sendo up- e down-regulados exclusivamente em celulose, respectivamente. Por outro lado, 1889 genes foram diferencialmente expressos em soforose e em celulose, com 321 e 405 sendo up- e down-regulados em soforose, e 262 e 97, sendo up- e down-regulados em celulose, respectivamente (Figura 45B). No entanto, em glicose e soforose 1670 genes foram modulados, com 262 e 245 genes, sendo up- e down-regulados em soforose, e 505 e 473 genes up- e down-regulados em glicose, respectivamente (Figura 45C). Curiosamente, o número de genes transcricionalmente modulados em soforose (726 genes) foi maior do que em celulose (359 genes), quando ambos foram comparados com glicose.

Aplicando o limiar estabelecido, isto é, Log$_2$ *Fold Change* ≥ 1 ou ≤ -1 e *p-valor* $\leq$ 0,05 como *threshold*, 2060 genes foram identificados como diferencialmente expressos em pelo menos uma das fontes de carbono analisadas (Celulose/Glicose; Soforose/Celulose e Soforose/Glicose). Com esses dados, ou seja, 2060 genes, construiu-se a clusterização hierárquica dos genes diferencialmente expressos em cada condição, permitindo a identificação de três possíveis regulons, representando 123 genes modulados por celulose, 154

genes modulados por soforose e 402 genes modulados por glicose, totalizando 679 genes (Figura 46A e CD Anexo (Arquivo 9, Tabelas 9.1 a 9.3)). A anotação, de acordo com o *Gene Ontology* (GO), dos 679 genes específicos nas fontes de carbono revelou que 46%, 34% e 39% dos genes dos regulons de celulose, glicose e soforose, respectivamente, foram genes que apresentaram função desconhecida. Esses resultados enfatizam o potencial de descoberta de genes envolvidos na produção de celulases em *T. reesei* durante o crescimento em condições indutoras e repressora.

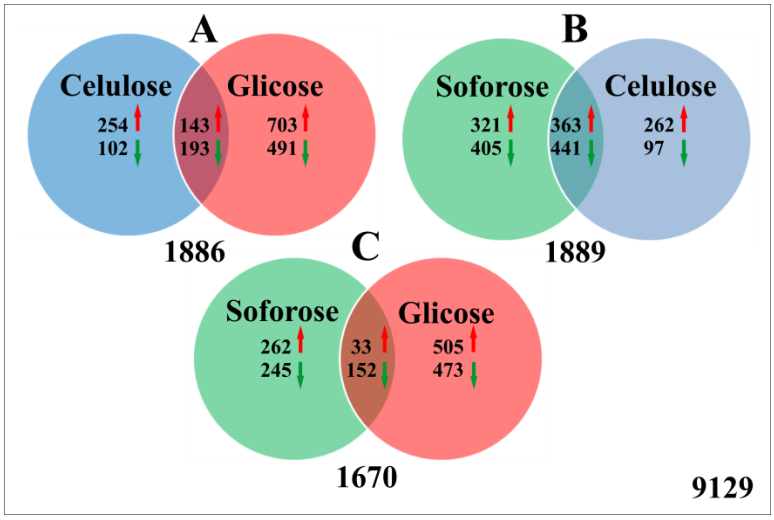


FIGURA 45 – Diagramas de *Venn* representando o número de genes diferencialmente expressos na linhagem parental QM9414.
(A) Linhagem QM9414 cultivada em celulose e glicose como fontes de carbono. (B) Linhagem QM9414 cultivada em celulose e soforose. (C) Linhagem QM9414 cultivada em soforose e glicose. As setas em vermelho indicam o número de genes up-regulados e as setas em verde o número de genes down-regulados nas condições analisadas. Os números abaixo de cada diagrama de *Venn* representam o total de genes regulados. O número no canto inferior direito representa o número de transcritos no genoma de *T. reesei*.

A fim de avaliar os regulons específicos de cada fonte de carbono apresentados na Figura 46, o top 10 genes diferencialmente expressos em celulose, glicose e soforose foram identificados na Tabela 13. O top 10 genes up-regulados em celulose incluiu as glicosil hidrolases (GH5, GH31 e GH16), a carboidrato esterase (CE5), uma oxidase, uma permease do facilitador maior da superfamília MFS e cinco proteínas com função desconhecida (Tabela 14). O top genes up-regulados em soforose incluiu um GH76, quatro oxidoredutases, duas MFS permeases e três proteínas com funções desconhecidas (Tabela 14). É interessante destacar que existem mais GHs nos top 10 genes de celulose do que em soforose. Como esperado, o top 10 genes em glicose não apresentou nenhum gene que codifica hidrolases

(Tabela 14). Esses resultados indicam uma expressão de genes específicos em resposta à fonte de carbono disponível em *T. reesei*.

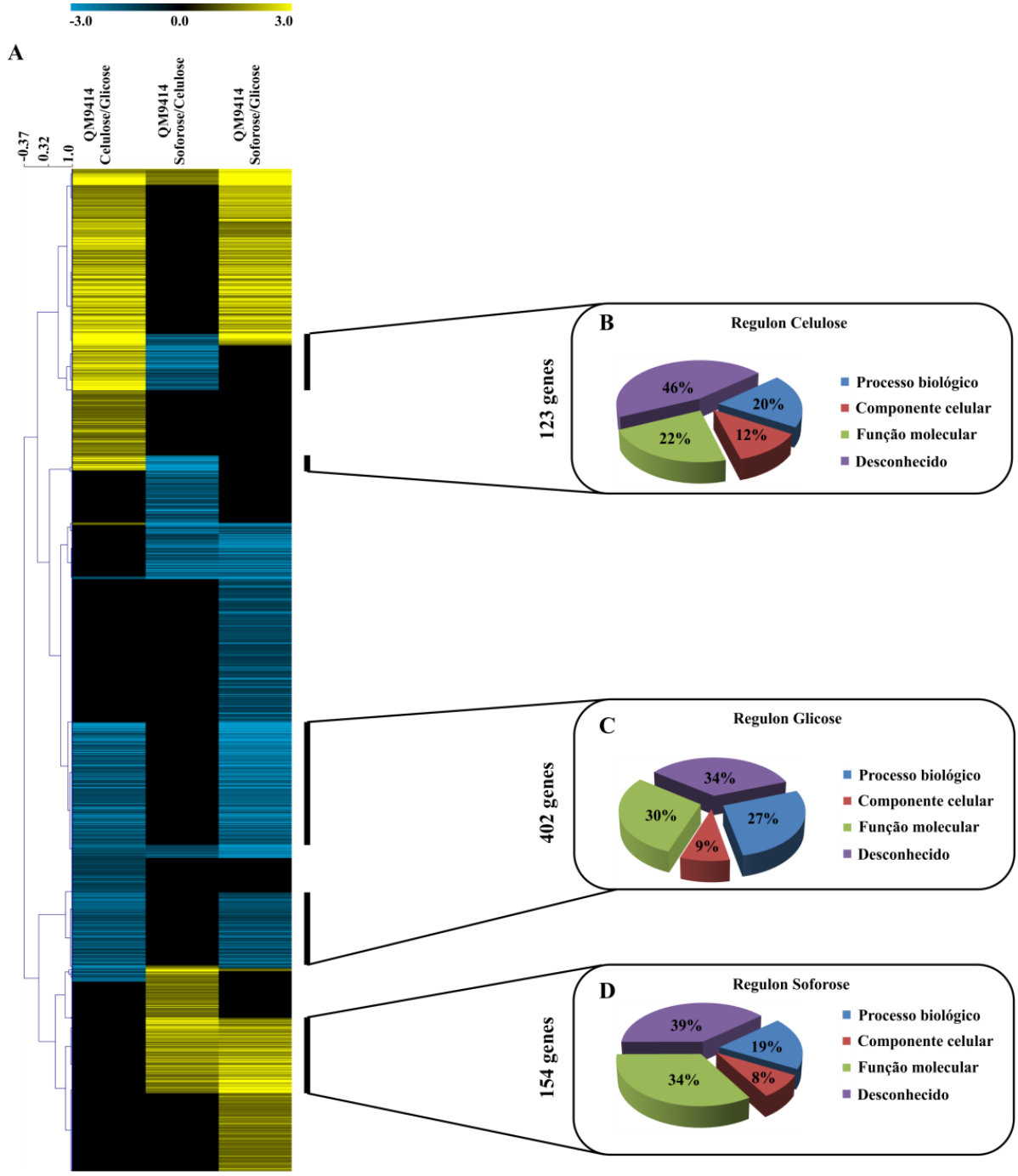


FIGURA 46 – Perfil de expressão gênica de *T. reesei* QM9414 durante crescimento na presença de celulose, soforose e glicose como fontes de carbono.
(A) Clusterização hierárquica dos genes significativos up- e down-regulados em *T. reesei* QM9414. Utilizou-se o *software Mev v.4.6.1*, com o método "*average linkage*" para gerar os *clusters* e correlação "*uncentered*". A distância Euclidiana foi usada para verificar as diferenças entre os 2060 genes e os grupos (condições) baseado na distância do centro dos grupos, $p < 0,05$. (B) Genes up-regulados em celulose. (C) Genes up-regulados em glicose. (D) Genes up-regulados em soforose. Em (B), (C) e (D) também estão representados a anotação de acordo com o *Gene Ontology*.

Tabela 14 – Top 10 genes diferencialmente expressos em celulose, glicose e soforose.

Condição	Proteína ID	Descrição	GO	FC Cel/Gli	FC Sof/Cel	FC Sof/Gli
Celulose	69957	MFS permease	Cellular component	11,001	-3,673	0
	56996	GH5 β-Mannanase MAN1	Biological process	10,849	-5,403	0
	69944	GH31 α-xylosidase/α-glucosidase	Biological process	10,5	-2,007	0
	73632	CE5 acetyl xylan esterase AXE1	Molecular function	8,06	-1,025	0
	108642	Unknown protein	Unknown	3,644	-8,621	0
	112258	Unknown protein	Unknown	4,665	-8,069	0
	124079	Multicopper oxidases	Molecular function	3,787	-7,762	0
	123236	SSCRP	Unknown	2,666	-7,713	0
	119552	Unique protein	Unknown	3,884	-7,09	0
	55886	GH16 glucan endo-1,3(4)-β-D-glucosidase	Unknown	4,086	-7,067	0
Glicose	79816	Unknown protein; secreted	Molecular function	-5,793	0	-9,354
	70520	Short chain dehydrogenase/reductase	Molecular function	-4,333	0	-9,263
	23382	Aldehyde reductase AKR7	Molecular function	-4,028	0	-9,213
	30759	Zinc containing alcohol dehydrogenase superfamily	Molecular function	-3,684	0	-8,508
	123084	Chloroperoxidase	Molecular function	-4,587	0	-8,369
	122998	Unknown protein	Unknown	-3,64	0	-7,637
	69115	Dienelactone hydrolase	Unknown	-4,827	0	-7,496
	81525	Isoflavone reductase	Molecular function	-4,43	0	-7,448
	81586	Unknown protein	Unknown	-2,903	0	-7,338
	76641	MFS permease	Cellular component	-4,454	0	-7,309
Soforose	106164	Short chain dehydrogenase/reductase	Molecular function	0	2,026	10,007
	59628	Unknown protein	Unknown	0	3,846	5,711
	48444	MFS maltose permease	Cellular component	0	2,109	5,524
	5345	FAD-containing oxidoreductase	Molecular function	0	2,579	5,443
	122087	Unknown protein	Cellular component	0	2,898	5,225
	21876	Zinc-binding oxidoreductase	Molecular function	0	5,034	3,828
	22915	Glucose oxidase	Molecular function	0	1,156	4,544
	110267	Unknown protein	Biological process	0	4,486	1,287
	60945	MFS permease	Cellular component	0	1,341	4,451
	55802	GH76 α-1,6-mannanase	Molecular function	0	1,956	4,441

GO = *Gene Ontology*; FC = Log_2 *Fold Change*; MFS = permease do facilitador maior da superfamília; GH = glicosi lhidrolase; CE = carboidrato esterase; AXE = acetil xilana esterase; SSCRP = pequena proteína secretada rica em cisteína; Cel = celulose; Gli = glicose e Sof = soforose.

6.1.2 *CAZYome*

A média do FPKM (fragmentos por kilobase de exon por milhões fragmentos mapeados) para todos os genes da família de GH foram calculados em cada condição analisada. Esses dados foram utilizados para demonstrar o potencial enzimático geral e a resposta transcricional global (Figura 47). Durante o crescimento em glicose, a transcrição das GHs, em geral, foi baixa, porém, quando analisado no crescimento na presença de celulose ou soforose, resultou numa drástica indução de várias famílias de GH, refletindo na indução transcricional de *CAZymes*, como exemplos, as duas condições indutoras resultaram em aumento da transcrição de celobiohidrolases, membros das famílias GH6 e GH7 (Figura 47).

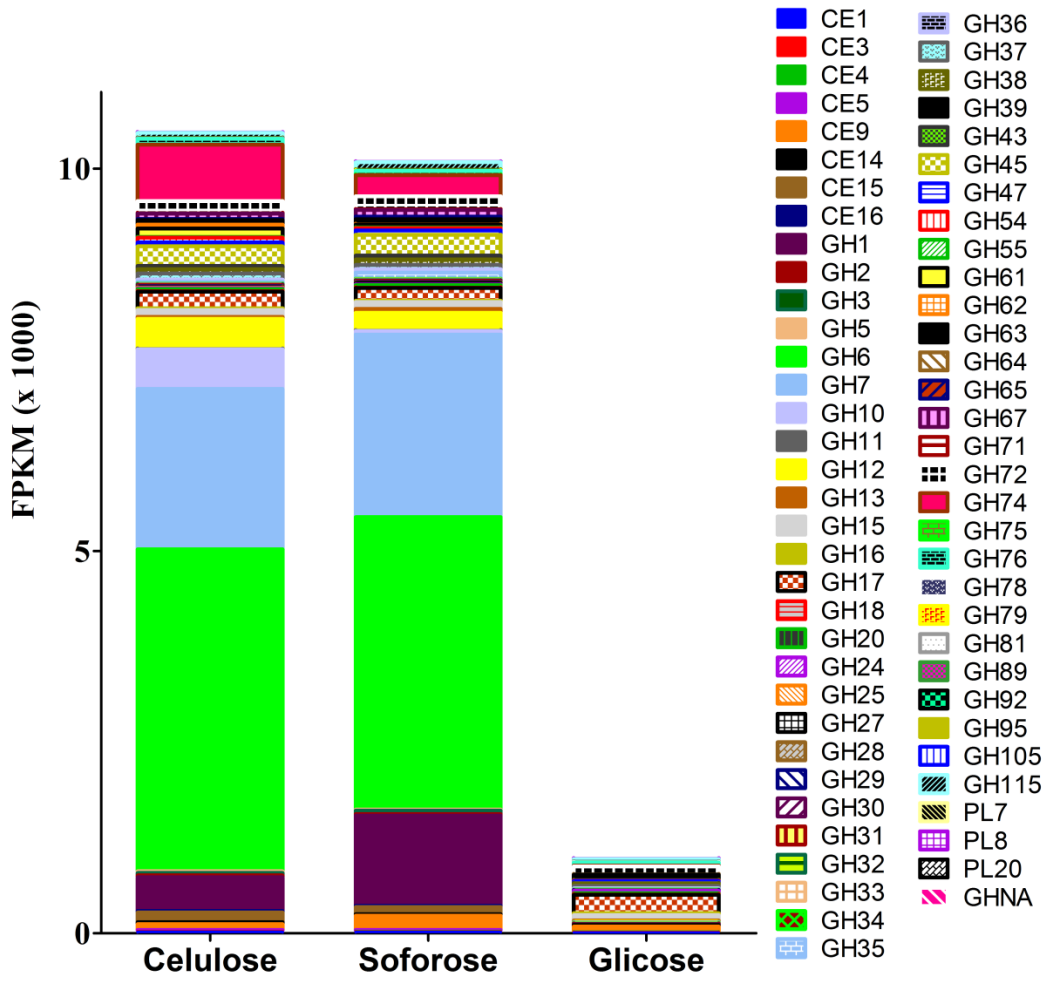


FIGURA 47 – Expressão de genes *CAZy* obtidos por RNA-seq.
Fragmentos por kilobase de exon por milhões fragmentos mapeados (FPKM) foram obtidos da média de cada família glicosil hidrolase (GH) quando cultivado em presença de celulose, soforose e glicose. A classificação *CAZy* foi realizada de acordo com a re-anotação dos genes de *CAZy* de *T. reesei* de acordo com Häkkinen e colaboradores (2012).

As Tabelas 15, 16 e 17 retratam com mais detalhes as enzimas que são up-reguladas em soforose, celulose e glicose. Vinte GHs e uma CE foram up-reguladas em resposta à presença de soforose (Tabela 15), enquanto que 23 GHs e duas CEs foram induzidas na presença de celulose (Tabela 16). Curiosamente, genes que codificam enzimas envolvidas na degradação de xilana, tais como: as xilanases (*xyn2*, *xyn3* e *xyn4*); acetil xilana esterase (*axe1*); xiloglucanase (*cel74a*); α-xilosidase (GH31); degradação de arabinoxilanas, tais como: arabinofuranosidase (ABF1 e ABF2) foram preferencialmente expressas em presença de celulose, mas não em soforose (Tabelas 15 e 16). Além disso, uma polissacarídeo monooxigenase (*cel61a*) foi up-regulada somente em celulose, e está de acordo com sua função no processo de oxidação da celulose.

Esses resultados também foram observados por Bischof e colaboradores (2013), quando os dados transcricionais obtidos pelo cultivo da palha de trigo comparado ao cultivo em presença de lactose. Por outro lado, tanto genes que codificam α- e β-glicosidades (incluindo *cel3c*, *cel3b* e *cel1b*), e uma candidata a α-amilase e α-1,6-mananase (enquanto uma β-mananase foi expressa em celulose), foram up-reguladas em soforose (Tabela 15). Interessantemente, em ambas as condições indutoras, enzimas que degradam trealose foram induzidas, indicando que o fungo pode catabolizar trealose armazenada, produzindo glicose, durante a produção de celulases.

Quando se comparou a expressão gênica entre as três condições foram observados que na presença de glicose, 17 GHs foram up-reguladas (Tabela 17). Esses genes codificam enzimas, tais como: endoglucanase (*cel5b*), β-1,4-glucanase (GH5), β-1,3-glucanosiltransferase (GH72), e uma não caracterizada GH (ID121136) e parecem não estar sujeitas a repressão catabólica.

Tabela 15 – Genes *CAZy* up-regulados em soforose.

Proteína ID	Nome	Classe	Família	Anotação	Log_2 FC Sof/Cel	*p-valor*
105956		Glicosil hidrolase	13	Candidate α-amylase	2,1286	6,00E-24
55802		Glicosil hidrolase	76	Candidate α-1,6-mannanase	1,9559	8,70E-15
82235		Glicosil hidrolase	31	Candidate α-glucosidase	1,8953	8,51E-20
82227	*cel3c*	Glicosil hidrolase	3	Candidate β-glucosidase	1,6899	2,57E-32
120749	*bgl2/cel1a*	Glicosil hidrolase	1	β-glucosidase	1,6795	1,01E-18
50215		Glicosil hidrolase	16	Candidate endo-1,3-β-D-glucosidase/1,3-glucan binding protein	1,6128	1,15E-23
59082	*chi18-2*	Glicosil hidrolase	18	Candidate chitinase	1,5828	0,003875319
80240	*bga1*	Glicosil hidrolase	35	β-galactosidase	1,5736	5,07E-31
66041	*chi18-18*	Glicosil hidrolase	18	Candidate chitinase	1,5377	7,42E-11
106575		Glicosil hidrolase	79	Candidate β-glucuronidase	1,5174	1,22E-14
123456		Glicosil hidrolase	65	Candidate α,α-trehalase	1,5006	5,69E-17
79671		Carboidrato esterase	9	Candidate N-acetyl-glucosamine-6-phosphate deacetylase	1,3821	0,00000921
108477		Glicosil hidrolase	13	Candidate α-glucosidase/oligo α-glucosidase	1,3297	9,49E-14
74198		Glicosil hidrolase	92	Candidate α-1,2-mannosidase	1,2728	5,09E-08
122511		Glicosil hidrolase	16	Candidate glucan endo-1,3(4)-β-D-glucosidase	1,2565	2,07E-11
121735	*cel3b*	Glicosil hidrolase	3	Candidate β-glucosidase	1,2107	9,89E-12
71532		Glicosil hidrolase	71	Candidate α-1 3-glucanase	1,1734	8,21E-08
22197	*cel1b*	Glicosil hidrolase	1	Candidate β-glucosidase	1,1577	2,97E-19
79669		Glicosil hidrolase	3	Candidate β-N-acetylglucosaminidase	1,048	9,44E-16
58117		Glicosil hidrolase	89	Candidate α-N-acetylglucosaminidase	1,0084	0,000390284
65380		Glicosil hidrolase	47	Candidate α-1,2-mannosidase	1,001	1,66E-13

Tabela 16 – Genes *CAZy* up-regulados em celulose.

Proteína ID	Nome	Classe	Família	Anotação	Log_2 FC Sof/Cel	*p-valor*
80833	*chi46*	Glicosil hidrolase	18	Chitinase	-1,0173	3,20E-09
73632	*axe1*	Glicosil hidrolase	5	Acetyl xylan esterase	-1,0245	2,32E-08
53731		Glicosil hidrolase	5	Candidate endo-β-1,4-glucanase	-1,0746	0,006503566
65162	*Endo T*	Glicosil hidrolase	18	Endo-N-acetyl-β-D-glucosaminidase	-1,0917	1,46E-08
123226		Glicosil hidrolase	37	Candidate α,α-trehalase	-1,0944	0,000000518
2735	*chi18-6*	Glicosil hidrolase	18	Candidate chitinase	-1,0999	0,000000414
49081	*cel74a*	Glicosil hidrolase	74	Xyloglucanase	-1,3971	6,56E-15
45717	*mds1*	Glicosil hidrolase	47	α-1,2-mannosidase	-1,5459	2,41E-09
123232	*egl3/cel12a*	Glicosil hidrolase	12	Endo-β-1,4-glucanase	-1,6085	3,92E-18
79602		Glicosil hidrolase	81	Candidate endo-1,3-β-glucanase	-1,6627	4,93E-19
44366		Glicosil hidrolase	3	Candidate esterase	-1,8139	3,39E-32
120312	*egl2/cel5a*	Glicosil hidrolase	5	Endo-β-1,4-glucanase	-1,8718	8,51E-43
123818	*xyn2*	Glicosil hidrolase	11	Endo-β-1,4-xylanase	-1,9745	3,37E-26
69944		Glicosil hidrolase	31	Candidate α-xylosidase/α-glucosidase	-2,007	3,94E-17
111849	*xyn4*	Glicosil hidrolase	30	Endo-β-1,4-xylanase	-2,0563	5,74E-46
74807		Glicosil hidrolase	76	Candidate α-1,6-mannanase	-2,7202	5,29E-89
123283	*abf1*	Glicosil hidrolase	54	α-L-arabinofuranosidase I	-2,7247	1,30E-48
122495		Glicosil hidrolase	76	Candidate α-1,6-mannanase	-3,0661	1,33E-53
120229	*xyn3*	Glicosil hidrolase	10	Endo-β-1,4-xylanase	-3,3849	6,11E-45
62166		Glicosil hidrolase	2	Candidate β-mannosidase	-3,4682	5,64E-26
120961	*cel61b*	Glicosil hidrolase	61	Candidate copper-dependent polysaccharide monooxygenase	-3,5574	1,35E-85
76210	*abf2*	Glicosil hidrolase	62	Candidate α-L-arabinofuranosidase	-3,8807	1,08E-65
56996	*man1*	Glicosil hidrolase	5	β-Mannanase	-5,4028	5,28E-84
124175		Glicosil hidrolase	64	Candidate endo-1,3-β-glucanase	-5,639	2,01E-137
55886		Glicosil hidrolase	16	Candidate glucan endo-1,3(4)-β-D-glucosidase	-7,067	3,59E-67

Tabela 17 – Genes *CAZy* up-regulados em glicose.

Proteína ID	Nome	Classe	Família	Anotação	Log_2 FC Gli/Sof/Cel	*p-valor*
121136		Glicosil hidrolase	NA	-	-5,1769	0,022304559
49274		Glicosil hidrolase	16	Candidate glucan endo-1,3(4)-β-D-glucosidase	-3,63	0,00598636
121294		Glicosil hidrolase	16	Candidate glucan endo-1,3(4)-β-D-glucosidase	-2,6584	0,005798218
50215		Glicosil hidrolase	16	Candidate endo-1,3-β-D-glucosidase/1,3-glucan binding protein	-2,5328	0,0007701
82616	*cel5b*	Glicosil hidrolase	5	Candidate membrane bound endoglucanase	-2,4871	0,000368036
82633		Glicosil hidrolase	72	Candidate β-1 3-glucanosyltransferase	-2,1732	1,15E-15
82633		Glicosil hidrolase	72	Candidate β-1 3-glucanosyltransferase	-2,0714	0,002177737
122511		Glicosil hidrolase	16	Candidate glucan endo-1,3(4)-β-D-glucosidase	-2,0364	3,25E-12
53731		Glicosil hidrolase	5	Candidate endo-β-1,4-glucanase	-2,008	0,0000499
39942		Glicosil hidrolase	17	Candidate glucan endo-1,3-β-glucosidase	-1,9124	0,030050127
66041	*chi18-18*	Glicosil hidrolase	18	Candidate chitinase	-1,5663	1,06E-19
39755		Glicosil hidrolase	16	Candidate glucan endo-1,3(4)-β-D-glucosidase	-1,3516	0,002774499
77284		Glicosil hidrolase	12	Candidate endo-β-1,4-glucanase	-1,3141	0,000101813
65333		Glicosil hidrolase	15	Candidate alpha-glycosidase (Glucoamylase and related glycosyl hydrolases)	-1,3056	0,000892332
81598	*chi18-7*	Glicosil hidrolase	18	Candidate chitinase	-1,1877	0,0000114
68064		Glicosil hidrolase	43	Candidate β-xylosidase/α-L-arabinofuranosidase	-1,1055	0,0000069
71399		Glicosil hidrolase	16	Candidate endo-1,3-β-glucanase	-1,062	0,000340308

6.1.3 Fatores transcricionais

A Tabela 18 apresenta os fatores transcricionais que foram induzidos em cada condição analisada. Nessa análise, 7 fatores foram up-regulados em celulose, 18 em soforose e outros 18 foram específicos em glicose. Focou-se então nos fatores que foram up-regulados, dependendo da fonte de carbono. O gene caracterizado pela proteína ID105269 apresentou um alto nível de expressão em presença de celulose, enquanto o ID123881 apresentou um alto nível de expressão em soforose, e 112499 foi up-regulado em glicose (Tabela 18). Genes que codificam fatores transcricionais do subgrupo Zn2Cys6, conhecido como C2H2 (um dos mais comuns tipos de fatores transcricionais encontrados em eucariotos), foram somente induzidos em presença de celulose e soforose, mas não na condição repressora, sugerindo uma resposta específica de C2H2 à presença de moléculas indutoras de celulases. Genes que codificam fatores transcricionais da família bZIP, por outro lado, apresentaram uma alta expressão em celulose (ID110152), mas também estavam presentes em soforose (ID73654) e em glicose (ID119759).

A expressão de genes que codificam fatores transcricionais já caracterizados em estar envolvidos na regulação da expressão de celulases e hemicelulases estão apresentados na Tabela 19. Entre os fatores de transcrição que atuam positivamente (XYR1, ACE2, CLR-1, CLR-2 e BglR), o gene *xyr1* apresentou o maior nível de expressão, seguido por CLR-1 e BglR (Tabela 18). Os fatores ACE2 e CLR-2 não apresentaram diferença significativa na expressão nas condições analisadas. Esses resultados reforçam a hipótese que o fator XYR1 é o principal regulador positivo na expressão de genes de celulases e hemicelulases. Por outro lado, os fatores transcricionais que atuam negativamente na transcrição de enzimas hidrolíticas, tais como ACE1 e CRE1, apresentaram um baixo nível de expressão comparado aos fatores que atuam positivamente, ou não foram transcricionalmente modulados dependendo da fonte de carbono, sugerindo que estes fatores podem atuar de maneira cooperativa ou tem um modo mais eficaz de ação. Além disso, o gene PacC (fator transcricional responsível ao pH) foi regulado dependentemente da fonte de carbono, apresentando uma alto nível de expressão na presença de celulose (Tabela 19).

Tabela 18 – Principais fatores transcricionais induzidos em celulose, soforose e glicose.

Condição	Proteína ID	Descrição	Cel/Gli	Sof/Cel	Sof/Gli	Regulação
Celulose	110152	BZIP transcriptional regulator	2,45	-0,95	ns*	↑
	120698	C2H2 transcriptional regulator	2,12	-0,54	ns	↑
	108775	Transcription factor AbaA	1,70	-1,77	ns	↑
	68254	Zn2Cys6 transcriptional regulator	1,33	-1,51	ns	↑
	69972	Zn2Cys6 transcriptional regulator	1,85	-1,07	ns	↑
	121164	Zn2Cys6 transcriptional regulator	1,75	-1,19	ns	↑
	105269	Zn2Cys6 transcriptional regulator	2,84	-1,99	ns	↑
Soforose	73654	BZIP transcriptional regulator	ns	1,25	1,96	↑
	67418	C2H2 transcription factor	ns	0,60	1,70	↑
	120224	C2H2 transcriptional regulator	ns	0,76	1,59	↑
	120428	C2H2 transcriptional regulator	ns	0,62	1,28	↑
	120908	myb transcriptional regulator	ns	1,01	0,90	↑
	80200	transcription factor (Snd1/p100)	ns	1,48	0,93	↑
	62244	Zn2Cys6 transcriptional regulator	ns	1,90	1,91	↑
	65746	Zn2Cys6 transcriptional regulator	ns	0,82	1,22	↑
	66828	Zn2Cys6 transcriptional regulator	ns	1,00	0,69	↑
	68455	Zn2Cys6 transcriptional regulator	ns	0,98	1,47	↑
	55274	Zn2Cys6 transcriptional regulator	ns	1,48	2,75	↑
	70351	Zn2Cys6 transcriptional regulator	ns	0,79	2,86	↑
	58389	Zn2Cys6 transcriptional regulator	ns	0,96	1,69	↑
	21997	Zn2Cys6 transcriptional regulator	ns	1,30	1,44	↑
	123881	Zn2Cys6 transcriptional regulator	ns	2,24	3,21	↑
	121107	Zn2Cys6 transcriptional regulator	ns	1,68	1,88	↑
	73792	Zn2Cys6 transcriptional regulator	ns	1,04	1,09	↑
	72611	Zn2Cys6 transcriptional regulator	ns	0,78	2,41	↑
Glicose	119759	BZIP transcriptional regulator	-1,41	ns	-1,60	↑
	21270	CAP20 virulence factor	-1,36	ns	-1,42	↑
	78049	Elongation factor Tu (G)	-1,33	ns	-1,34	↑
	75472	Transcriptional regulator, unknown	-0,89	ns	-1,00	↑
	74346	Translation elongation factor precursor from Aspergillus fumigatus	-1,34	ns	-1,81	↑
	57676	Translation initiation factor 3, subunit i (elf-3i)	-1,16	ns	-0,92	↑
	74252	Translation initiation factor 6 (eIF6) by homologyToThe corresponding protein in other eukaryotes	-0,76	ns	-1,03	↑
	54437	Zn2Cys6 transcriptional regulator	-1,28	ns	-2,05	↑
	66047	Zn2Cys6 transcriptional regulator	-1,21	ns	-1,72	↑
	55759	Zn2Cys6 transcriptional regulator	-1,03	ns	-0,80	↑
	57534	Zn2Cys6 transcriptional regulator	-0,87	ns	-1,14	↑
	102497	Zn2Cys6 transcriptional regulator	-3,88	ns	-3,70	↑
	102499	Zn2Cys6 transcriptional regulator	-4,39	ns	-3,34	↑
	109394	Zn2Cys6 transcriptional regulator	-3,22	ns	-4,11	↑
	104182	Zn2Cys6 transcriptional regulator	-1,67	ns	-2,48	↑
	112202	Zn2Cys6 transcriptional regulator	-2,37	ns	-2,83	↑
	112499	Zn2Cys6 transcriptional regulator	-4,40	ns	-6,82	↑
	105520	Zn2Cys6 transcriptional regulator	-0,89	ns	-1,41	↑

↑ Up-regulados; Cel = celulose; Sof = soforose e Gli = glicose.
*ns = Não significativo (p > 0,05)

Tabela 19 – Expressão de fatores transcricionais envolvidos na regulação dos genes de celulases e hemicelulases.

Proteína ID	Nome	Cel/Gli	Sof/Cel	Sof/Gli
122208	XYR1	6,062	2,087	8,131
78445	ACEII	ns*	0.069	ns
27600	CLR-1	2,263	0,516	2,766
26163	CLR2	ns	ns	ns
52368	BglR	1,492	ns	1.57
120117	ACEI	0,807	-0,674	ns
120117	CREI	0,807	-0,674	ns
124286	HAP2	ns	0,455	ns
121080	HAP3	ns	ns	ns
62979	HAP5	ns	ns	ns
76817	AreA	ns	0,458	ns
120698	PacC	2,123	-0,538	ns

Cel = celulose; Sof = soforose e Gli = glicose. * ns = não significativo

Outros fatores de transcrição com potencial papel regulatório, tais como HAP2/3/5 e AreA, não apresentaram modulação significativa na expressão gênica nas condições analisadas. Esses dados mostram que existe um sistema complexo de fatores que regulam as enzimas hidrolíticas e revelou, também, novos fatores não caracterizados influenciando o processo de produção de enzimas.

6.1.4 Transportadores

Os genes que codificam proteínas envolvidas no transporte compreendem cerca de 5% (459 genes) do genoma de *T. reesei*. Resultados mostram que, entre estes, 14 são regulados exclusivamente por celulose, 14 por soforose e 30 por glicose (Tabela 20).

As permeases MFS (*Major Facilitator Superfamily*) são as proteínas mais abundantes nas três condições analisadas. Essas proteínas permitem a absorção de nutrientes essenciais e íons, excreção dos produtos finais do metabolismo e comunicação célula/ambiente (PAO et al., 1998). O gene que codifica para MFS permease (ID69957) foi especificamente e altamente up-regulado em celulose e pode estar envolvido no transporte de dissacarídeos, devido à alta similaridade com uma putativa maltose permease do fungo patogênico humano *Talaromyces marneffei* (BOYCE; ANDRIANOPOULOS, 2013). Outra maltose permease que, codificada pelo gene ID48444, foi altamente induzida em soforose.

Por outro lado, a MFS permease gene ID76641 foi a mais expressa em glicose (Tabela 20). Esta MFS permease apresentou 85% de identidade com a vesícula sináptica transportadora SVOP e também semelhança com o transportador de glicose humano 1 (Glut1) (YAO; BAJJALIEH, 2009). Curiosamente, um gene que codifica uma potencial galactose permease (ID62380) foi especificamente expressa em glicose e uma MFS permease que codifica o gene ID 76800 foi induzido em celulose, ambos genes se assemelham com transportadores de *S. cerevisiae*, transportadores que, quando deletados, contribuíram com uma total perda de absorção de hexoses (WIECZORKE et al., 1999).

Outra família que apresentou dependência da fonte de carbono na regulação transcricional foi a transportadores ABC que foram altamente up-regulados em celulose e soforose. A família AAA (ATPases associadas com uma variedade de atividades celulares) e genes aquagliceroporinas foram altamente expressos em soforose, enquanto que genes que relacionados ao transporte de ADP/ATP foram altamente expressos em glicose (Tabela 20). Além disso, transportadores de aminoácidos, oligopeptídeos e genes relacionados ao transporte de íons foram regulados pelas três fontes de carbono, com um número maior de genes em celulose.

A fim de identificar as MFS permeases que são comuns as condições celulose e soforose, a condição glicose foi utilizada para a normalização entre ambas condições indutoras. Dentre as 85 MFS permeases, anotadas no genoma de *T. reesei*, 22 são comuns às duas condições (Tabela 21). Entre elas, a mais expressa foi: *crt1*, a qual tem sido demonstrada ser requerida em *T. reesei* para o crescimento em celulose e lactose, mas não em xilana (ZHANG et al., 2013); *hxt1*, uma glicose permease; uma MFS permease ID50894, que tem alta afinidade com transportador de glicose (RIES et al., 2013b); e uma MFS permease relacionada com a sinalização de celulose (ID79202) (IVANOVA et al., 2013). Interessantemente, uma MFS permease, descrita recentemente, *stp1*, que está envolvida no transporte de celobiose e glicose (ZHANG et al., 2013), apresentou um alto nível de expressão em soforose do que em celulose (Tabela 21), indicando uma complexa regulação na absorção de celobiose/soforose em *T. reesei*.

Tabela 20 – Transportadores up-regulados em presence de celulose, soforose e glicose.

Condição	Proteína ID	Descrição	Cel/Gli	Sof/Cel	Sof/Gli	Regulação
Celulose	123611	Amino acid permease	2,15	-1,02	*ns	↑
	81442	Amino acid transporters	1,97	-1,18	ns	↑
	49970	H+ nucleoside cotransporter	1,93	-1,26	ns	↑
	79644	Metal ion transporter SMF2	2,04	-1,33	ns	↑
	75165	MFS permease	2,30	-2,41	ns	↑
	69957	MFS permease	11,00	-3,67	ns	↑
	58561	MFS permease	1,26	-1,99	ns	↑
	80058	MFS permease	3,71	-1,71	ns	↑
	74953	MFS permease	2,20	-2,73	ns	↑
	76800	MFS permease	3,45	-2,36	ns	↑
	68122	MFS permease	2,60	-2,76	ns	↑
	79329	MFS permease	1,58	-1,07	ns	↑
	60116	MRP-type ABC transporter	3,48	-1,87	ns	↑
	76682	PDR-type ABC transporters	4,17	-5,42	ns	↑
Soforose	122124	AAA ATPase	ns	2,80	1,83	↑
	103149	AAA ATPase	ns	1,79	1,94	↑
	64710	AAA+-type ATPase	ns	3,31	3,04	↑
	62693	ABC-transporter Ste6p	ns	1,33	2,09	↑
	58584	Aquaglyceroporin	ns	1,36	2,71	↑
	81082	Aquaglyceroporin	ns	2,34	2,67	↑
	48444	MFS maltose permease	ns	2,11	5,52	↑
	69026	MFS permease	ns	1,07	4,04	↑
	69611	MFS permease	ns	1,51	3,12	↑
	60945	MFS permease	ns	1,34	4,45	↑
	77785	MFS permease	ns	1,02	1,95	↑
	28409	MFS permease	ns	2,55	2,99	↑
	66111	MRP-type ABC transporter	ns	1,91	1,12	↑
	80028	MRP-type ABC transporter	ns	1,97	3,53	↑
Glicose	104106	ADP/ATP carrier protein	-4,32	ns	-7,02	↑
	67806	Amino acid permease	-1,45	ns	-2,05	↑
	110316	Amino acid transporters	-5,18	ns	-5,47	↑
	105752	C4-dicarboxylate transporter/malic acid transport protein	-2,56	ns	-4,53	↑
	123588	electron transport protein, probably involved in cytochrome C assembly	-1,89	ns	-2,87	↑
	80879	MFS H+/oligopeptide transporter	-3,91	ns	-4,38	↑
	43701	MFS multidrug transporter	-4,17	ns	-4,77	↑
	80086	MFS peptide transporter	-1,34	ns	-1,19	↑
	66657	MFS permease	-2,39	ns	-3,19	↑
	69164	MFS permease	-1,19	ns	-1,11	↑
	59272	MFS permease	-3,19	ns	-4,68	↑
	60086	MFS permease	-1,71	ns	-3,31	↑
	21595	MFS permease	-2,56	ns	-5,24	↑
	78585	MFS permease	-1,06	ns	-5,16	↑
	76641	MFS permease	-4,45	ns	-7,31	↑
	76775	MFS permease	-1,41	ns	-1,25	↑
	107936	MFS permease	-1,82	ns	-3,67	↑
	108893	MFS permease	-1,06	ns	-2,12	↑
	26642	MFS permease	-2,62	ns	-2,69	↑
	78970	Mitochondrial (phosphate) carrier	-1,47	ns	-2,45	↑
	22251	Mitochondrial carnitine-acylcarnitine carrier protein	-1,32	ns	-2,24	↑
	78679	Mitochondrial carrier protein	-1,35	ns	-2,30	↑
	64818	Mitochondrial substrate carrier	-1,12	ns	-1,88	↑
	120556	Mitochondrial substrate carrier	-1,35	ns	-1,41	↑
	64920	Monocarboxylate transporter	-3,43	ns	-4,85	↑
	76910	Monocarboxylate transporter	-2,15	ns	-2,28	↑
	56176	MRP-type ABC transporter	-1,56	ns	-1,55	↑
	44476	MRP-type ABC transporter	-1,80	ns	-1,97	↑
	59014	PDR-type ABC transporters	-2,19	ns	-3,88	↑
	77703	PDR-type ABC transporters	-3,06	ns	-3,79	↑

↑ Up-regulados; Cel = celulose; Sof = soforose e Gl i= glicose. Valores são expressos em *Log2 Fold Change*.
*ns = Não significativo (p > 0,05)

Tabela 21 – MFS permeases reguladas nas condições celulose e soforose.

Proteína ID	Descrição	Celulose	Soforose
3405	MFS permease (*Crt1*)	7,802	9,893
22912	MFS permease (glucose permease HXT1)	7,142	6,368
50894	MFS permease	6,683	7,069
79202	MFS permease, associated with cellulose signalling	5,742	7,646
44175	MFS H+ sugar transporter	4,413	4,102
121482	MFS permease	4,364	3,887
67752	MFS permease	4,085	3,919
68812	MFS permease	3,921	3,002
104549	MFS permease	3,871	4,482
47710	MFS permease (*Stp1*)	3,629	5,065
78833	MFS permease (fucose permease)	3,34	4,967
62502	MFS permease	2,971	3,137
54632	MFS permease	2,841	2,94
80026	MFS permease	2,718	3,073
50618	MFS permease	2,439	2,932
119789	MFS permease	2,135	2,055
106330	MFS permease	1,875	3,465
67334	MFS permease	1,74	2,234
61350	MFS permease	1,665	2,995
65153	MFS permease	1,663	2,731
68813	MFS permease	1,392	1,102
64874	MFS toxin efflux pump	1,155	1,017

6.1.5 Decifrando a rede regulatória de *T. reesei* em resposta a celulose/soforose

Utilizando os 2060 genes identificados como diferencialmente expressos (Figura 46) construiu-se uma rede regulatória com o intuito de identificar genes que são modulados de maneira dependentes das fontes de carbono estudadas e, também, quais genes são comuns entre as condições (Figura 48). Foi observado uma ampla coincidência entre genes diferencialmente expressos em celulose e soforose (710 genes). Além disso, genes especificamente associados com cada condição foram identificados, como exemplificados por um grande número de genes (441 genes) cuja expressão foi especificamente modulada durante o crescimento em soforose comparada à glicose. Esses genes representam tanto genes silenciados em glicose, mas induzidos por soforose (154 genes up-regulados), ou genes que são necessários para crescimento em glicose, mas dispensável para crescimento em soforose (287 genes down-regulados). Desse modo, as análises dos genes específicos da condição celulose apresentou uma super-representação de genes up-regulados (132 genes) que estão relacionados com a expressão de genes de celulases quando comparado à glicose (201 genes no total) (Figura 48).

A comparação entre as duas condições indutoras forneceu informações complementares por revelar novos genes diferencialmente expressos que não foram

identificados através da comparação com a glicose. Dos genes diferencialmente expressos (up- ou down-regulados) entre celulose e soforose, somente 75 e 107 genes foram designados especificamente para soforose ou celulose, respectivamente (Figura 48).

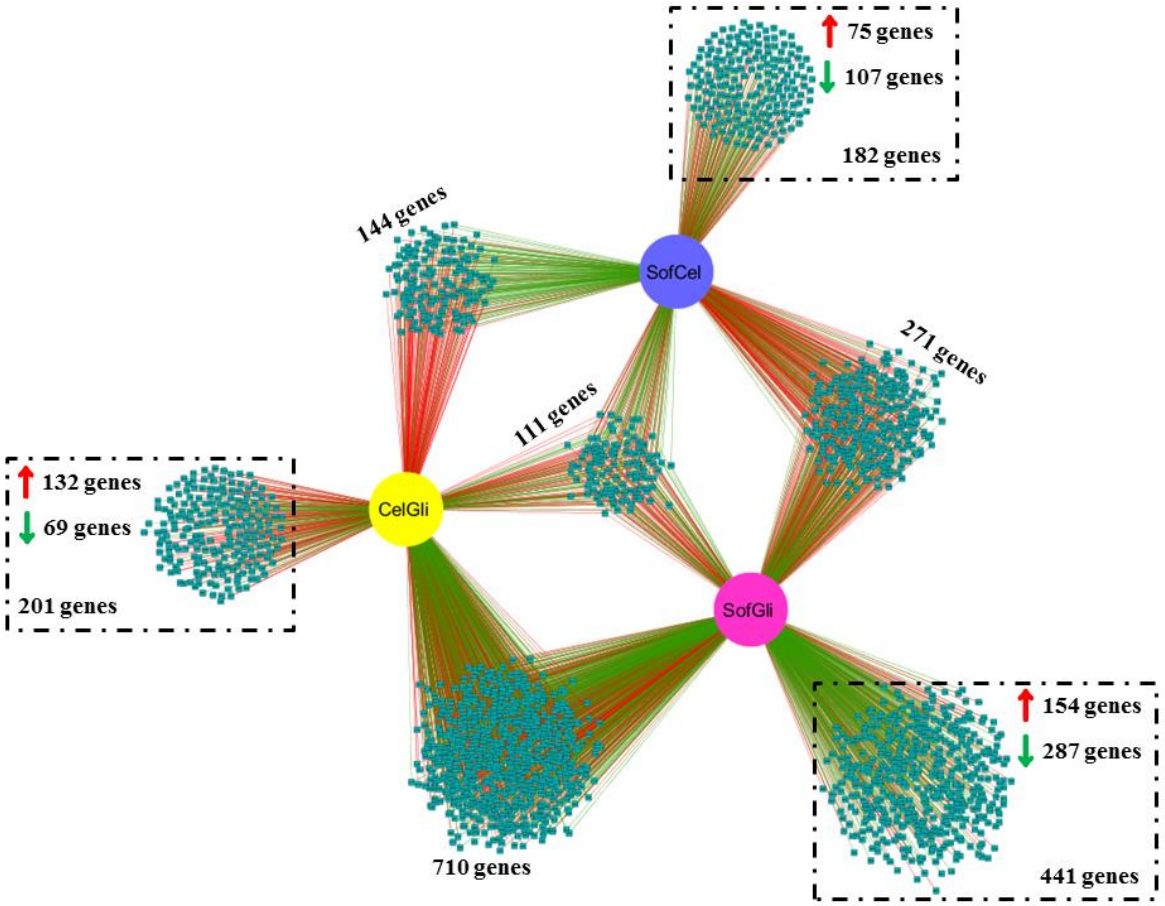


FIGURA 48 – Rede de regulação Gênica (GRN) de genes diferencialmente expressos em *T. reesei*.
Celulose versus Glicose (CelGli), Soforose versus Celulose (SofCel) e Soforose versus Glicose (SofGli). Genes são representados como "*nodes*" (quadrados azuis), e as interações são representadas como linhas, interações vermelhas (up-regulados) e interações verdes (down-regulados), que conectam "*nodes*", totalizando 3385 interações.

A maioria dos genes identificados nesta rede regulatória apresentou funções desconhecidas. É importante ressaltar que foi observado um enriquecimento de diferentes classes de genes celulose- ou soforose-específicos (Figura 49). Durante o crescimento em celulose, foram enriquecidos genes que codificam: *CAZy* (GH64, GH62, GH81, GH76 e GH54); proteínas acessórias (pequena proteína secretada rica em cisteína (SSCRP, OOC1 e Epl1); transportadores (a maioria deles relacionados ao transporte de metais e ferro); fatores transcricionais (C2H2 e Zn2Cys6); proteína *lae1*; e uma variedade de proteínas relatadas ao

transporte de elétrons (CD Anexo (Arquivo 10, Tabela 10)). Por outro lado, apenas três genes que codificam *CAZy* são específicos do crescimento em soforose. No entanto, quatro genes que codificam proteínas específicas da espécie de *Trichoderma* foram somente induzidos em soforose, sugerindo que *Trichoderma* possui um sistema de metabolismo especializado para a soforose (CD Anexo (Arquivo 10, Tabela 10)). Uma coincidência substancial entre o transcriptoma de celulose e soforose suporta a hipótese de que soforose é indutor natural na transcrição de celulases, enquanto que em celulose existe um enriquecimento específico de *CAZymes* e de proteínas acessórias, refletindo na dificuldade na desconstrução deste substrato insolúvel.

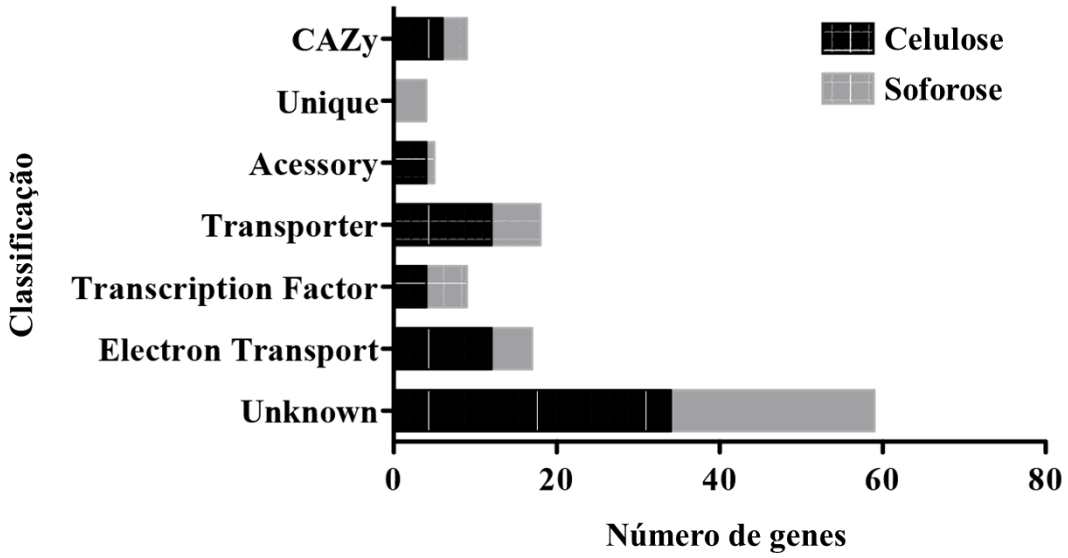


FIGURA 49 – Análise de enriquecimento de classes funcionais de acordo com o *Gene Ontology* (GO) de genes up-regulados nas condições celulose e soforose em *T. reesei*.
Categorias enriquecidas significamente ($p \leq 0,05$) estão apresentadas.

6.1.6 Análises de PCR quantitativa em Tempo Real (qPCR-TR)

Os resultados obtidos pelas análises de RNA-seq foram validados usando 20 genes diferencialmente modulados nas condições analisadas (Celulose versus Glicose, Soforose versus Celulose e Soforose versus Glicose) (CD Anexo (Arquivo 11, Tabela 11)). Nove genes codificam GHs e onze genes foram selecionados arbitrariamente. Os níveis de expressão foram obtidos por RNA-seq usando o pacote DESeq (ANDERS; HUBER, 2010) e por qPCR, utilizando o método $2^{-\Delta\Delta CT}$ (LIVAK; SCHMITTGEN, 2001) que foram comparados entre as condições. Uma alta correlação e significância estatística foram obtidas representadas pela

correlação de Pearson r^2= 0,8882; *p-valor* < 0,001, indicando a confiabilidade das análises do RNA-seq (Figura 50).

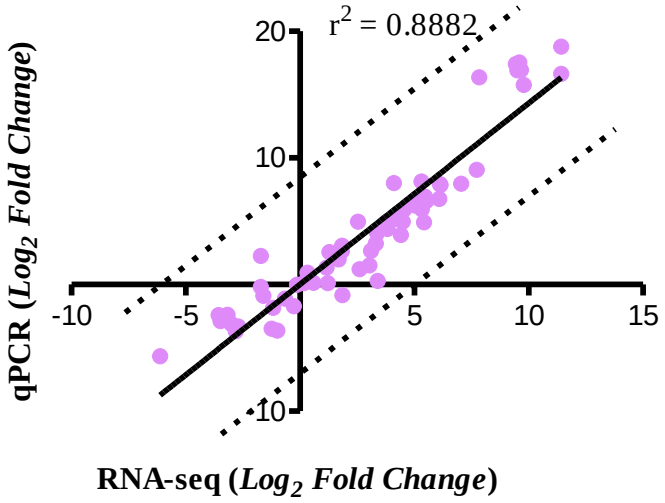


FIGURA 50 – Correlação entre RNA-seq e PCR quantitiva em Tempo Real (qPCR).

6.2 DISCUSSÕES

Em ambientes naturais, os organismos de vida livre são continuamente desafiados com condições de mudança rápida e tem um impacto considerável sobre seu estilo de vida. Técnicas de genômica e pós-genômicas revelaram que organismos de vida livre dedicam uma grande porcentagem de seus genes para detecção de sinais ambientais e a subsequente coordenação da expressão de genes em resposta a tais sinais. O fungo *T. reesei* reconhece substratos e ativa a transcrição de genes que codificam transportadores e fatores transcricionais que culminam na produção de enzimas hidrolíticas e é um assunto de especulação desde os anos 1960. Com as análises genômicas, tanto na condição de repressão (glicose) como também nas condições de de-repressão (celulose e soforose), foi possível identificar novos genes envolvidos na degradação da celulose por enzimas/proteínas produzidas por *T. reesei*. Além disso, a comparação do perfil transcricional de *T. reesei* entre a celulose e soforose, indutores naturais hipotéticos de produção de celulases, revelou uma notável semelhança nos perfis globais de expressão.

Por meio do estudo do transcriptoma de *T. reesei* QM9414, foram identificados 123 genes que foram especificamente induzidos por celulose, 154 por soforose e 402 genes por glicose (Figura 46). Dentre esse conjunto de genes, 8 genes que codificam permeases

(transportadores) foram induzidos em celulose, 6 em soforose e 11 em glicose (CD Anexo (Arquivo 9, Tabelas 9.1 a 9.3). Desses 25 transportadores, 10 apresentaram uma possível homologia com *N. crassa* e um deles, uma MFS permease (ID 76800, altamente induzida em presença de celulose), associada ao transporte de xilose em *S. cerevisiae* (DU et al., 2010; SUN et al., 2012). Além disso, um gene que codifica uma putativa galactose permease (ID62380) identificada em glicose e uma outra MFS permease (ID76800) regulada por celulose, apresentaram similaridade a transportadores envolvidos na absorção de hexoses em *S. cerevisiae* (WIECZORKE et al., 1999). Foram induzidos transportadores adicionais em ambas condições indutoras (celulose e soforose), sugerindo que soforose poderia ser o indutor natural na transcrição de genes de celulase em *T. reesei*. Apesar disso, as funções desses transportadores permanecem obscuras. O transportador (ID3405) foi recentemente identificado por estar especificamente envolvido na indução de celulases por lactose (IVANOVA et al., 2013), mas também, é descrito como envolvido no transporte de celobiose (ZHANG et al., 2013). Além disso, o mesmo transportador foi up-regulado durante o crescimentos em palha de trigo (BISCHOF et al., 2013; RIES et al., 2013b), celulose e soforose (Tabela 21). A falta de especificidade dos transportadores pode estar relacionada à estrutura fechada de celobiose/lactose/soforose ou pelo fato de que alguns transportadores podem atuar como transportadores e "sensores" de nutrientes. No entanto, serão necessários estudos mais detalhados para caracterizar esses transportadores e gerar uma melhor compreensão do sistema de transporte de indutores/repressores em *T. reesei*.

Análises da expressão gênica global por RNA-seq permitiu a construção da rede de regulação gênica (GRN) que reforçou a compreensão da interação entre diferentes genes durante a degradação e metabolismo de celulose. Estudos sobre o controle de genes catabólicos relacionados com o metabolismo de substratos simples (tais como glicose) realizados em organismos modelo têm revelado uma complexa GRN. Assim, espera-se que uma GRN ainda mais complexa controle funções catabólicas relacionadas com metabolismo de substratos complexos, como celulose (GLASS et al., 2013b). O conhecimento das vias regulatórias responsáveis por controlar a expressão gênica em *T. reesei* ainda é limitado, apesar de existir um modelo bem estabelecido de degradação de celulose (BISCHOF et al., 2013; GLASS et al., 2013b; MUKHERJEE et al., 2013). Na verdade, apesar dos fatores de transcrição XYR1 e CRE1 (FURUKAWA et al., 2009; STRAUSS et al., 1995), que regulam a indução e repressão de enzimas celulolíticas, respectivamente, e outros reguladores específicos (ACE1, ACE2, BglR) (PORTNOY et al., 2011b; NITTA et al., 2012) terem sidos

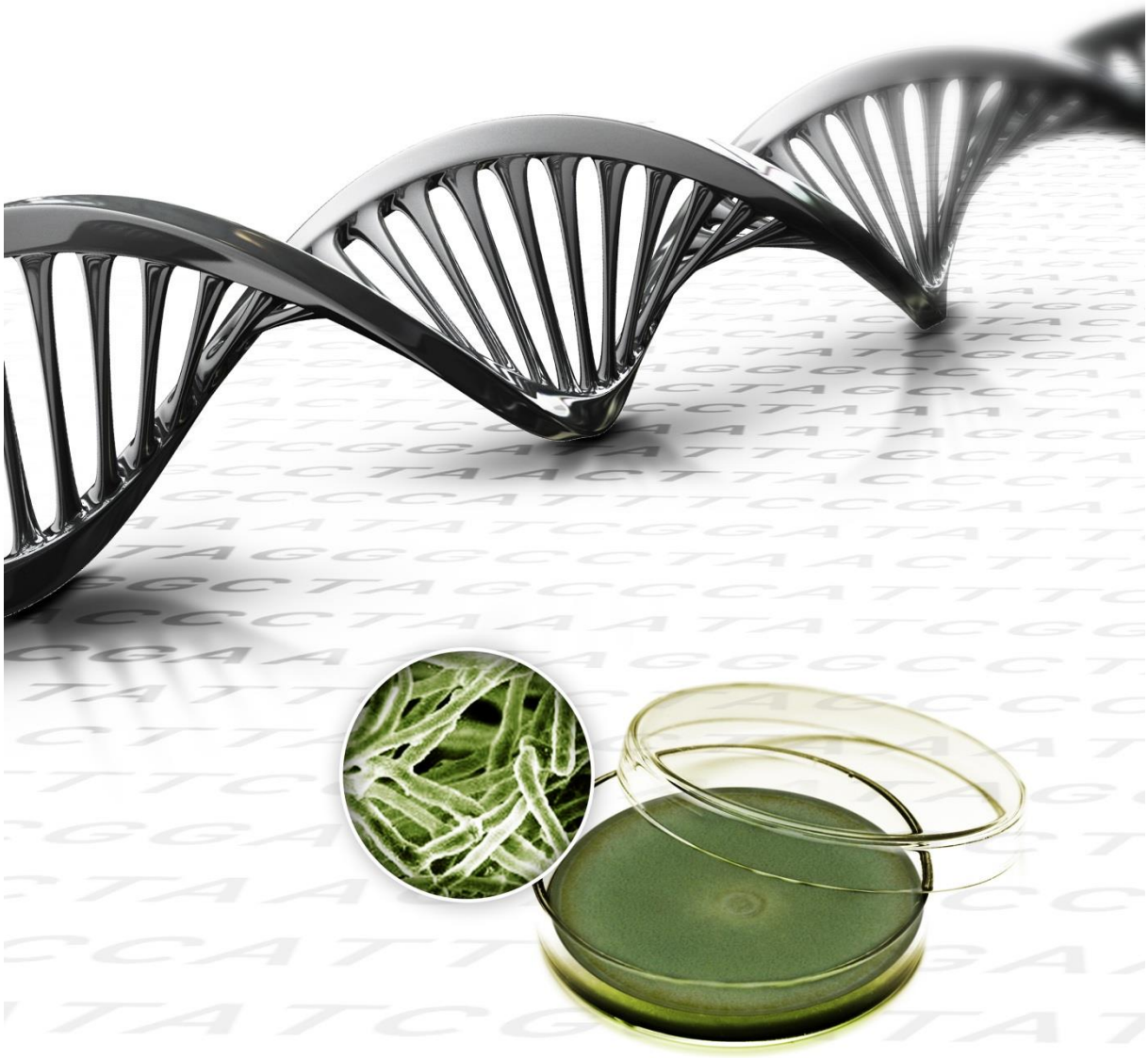
caracterizados experimentalmente, ainda não se sabe como e até que ponto a expressão dessas enzimas estão ligadas ao núcleo de GRN de *T. reesei* (GLASS et al., 2013b). Isso é importante porque a GRN em organismos de vida livre são geralmente altamente conectadas a regulação da expressão de um conjunto de genes, e é, geralmente, controlada por muitos sinais externos/internos (MARTÍNEZ-ANTONIO; COLLADO-VIDES, 2003). Os dados apresentados neste trabalho, a partir de análises genômicas em larga escala, visa preencher esta lacuna, fornecendo uma análise global de *T. reesei* cultivados em três fontes de carbono diferentes (celulose, soforose e glicose). A partir das análises apresentadas, a rede de regulação permitiu identificar de 43 fatores de transcrição induzidos em algumas condições específicas de crescimento (Tabela 18). A GRN revelou que alguns fatores são exclusivamente induzidos em resposta a celulose ou soforose (Tabela 18). A metiltransferase LAE1 foi anteriormente descrita por controlar a expressão de celulases, auxiliando fatores de transcrição auxiliares, β-glicosidases e xilanases para a degradação de celulose (SEIBOTH et al., 2012), que são proteínas comumente encontradas em resposta a indutores (celulose, lactose e palha de trigo (BISCHOF et al., 2013; RIES et al., 2013b). No entanto, a LAE1 foi preferencialmente expressa em resposta à celulose, indicando que o fungo tem uma específica sinalização para o metabolismo de celulose. Essa hipótese é corroborada pelo fato que um estudo recente apresentou que LAE1 afeta outros componentes da degradação de celulose, tais como a síntese de peptídeos não ribossomais, proteínas com repetições de anquirina, absorção de ferro, receptores PTH11 e oxidases/monoxigenases (KARIMI-AGHCHEH et al., 2013), e outros genes que foram up-regulados em presença de celulose em nossos resultados, e em presença de palha de trigo (BISCHOF et al., 2013). Outro gene up-regulado em presença de celulose (Fator de transcrição, ID120698) apresentou homologia com o fator transcricional PacC (responsivo ao pH) de *Aspergillus nidulans*. Sabe-se que este fator controla uma variedade de funções em fungos filamentosos (ROSSI et al., 2013). Alguns estudos têm demonstrado que o pH está envolvido na produção de celulases em *T. reesei* (LI et al., 2013), porém a regulação de genes de celulases por qualquer fator de transcrição em resposta ao pH, é ainda desconhecido.

O nível de expressão do gene *cre1* foi baixo, mesmo em presença de glicose. Esse resultado pode estar associado ao fato de que alguns fatores transcricionais podem atuar tanto diretamente nos genes que codificam *CAZymes* ou indiretamente por regular outros fatores transcricionais que, por sua vez, regulam genes *CAZymes*. Foram identificados alguns fatores transcricionais que podem regular indiretamente de maneira dependente da fonte de carbono

(Tabela 18). Alguns destes fatores poderiam desempenhar uma importante função em coordenar a expressão de genes down-regulados na rede gênica, tanto na associação de fatores identificados anteriormente em promotores alvos ou de forma isolada, em uma sinalização em cascata. Além disso, os fatores de transcrição identificados poderiam funcionar como pontos de verificação para integração de diferentes sinais fisiológicos e ambientais, tais como: estado metabólico da célula, luz, presença de estresse, etc. (MACH; ZEILINGER, 2003; GASCH et al., 2000). Os fatores transcricionais identificados são candidatos para estudos mais aprofundados dos mecanismos de integração de sinais nesse organismo de relevância biotecnológica. Entender como acontecem essas interações regulatórias é de fundamental importância para tentar sintetizar, por meio de engenharia biotecnológica, enzimas celulolíticas aprimoradas de *T. reesei*.

Apesar do extenso trabalho relacionado com a regulação das celulases em *T. reesei*, o verdadeiro indutor natural ainda não é totalmente estabelecido. Recentes evidências indicam celobiose e celodextrinas como fortes candidatos como indutores naturais (ZHOU et al., 2012). De fato, estudos com *N. crassa* (ZNAMEROSKI et al., 2012) e *A. niger* (GIELKENS et al., 1999) têm desacreditado soforose como indutor natural. Sabe-se que *T. reesei* possui um mecanismo diferente para a regulação da produção de celulases em resposta a soforose quando comparado a outros fungos que degradam lignocelulose (KUBICEK et al., 2011). Os dados de GRN mostraram pouca diferença na regulação da expressão gênica pelos indutores celulose e soforose, sugerindo que a soforose poderia ser um indutor natural. Mas como esta divergência entre *T. reesei* e outros fungos ocorreu? Genômica comparativa entre *T. atroviride*, *T. virens* e *T. reesei* sugerem que o ancestral de *Hypocrea*/*Trichoderma* era um micoparasita, possivelmente basidiomicetos degradadores de madeira (KUBICEK et al., 2011). *T. reesei* pode ter mantido essa característica com relação à competição de substratos, convertendo celobiose em soforose por reação de transglicosilação e, em seguida, metabolizando soforose. Essa hipótese pode ser corroborada pelo fato de que novas proteínas espécies-específicas foram up-reguladas somente em soforose e pelo fato que celobiose e soforose serem transportadas e metabolizadas em diferentes velocidades (ZHOU et al., 2012). Por esse motivo, propõe-se que, tanto celobiose e soforose atuam como coindutores na formação de celulases em *T. reesei*. Esses fatos poderiam explicar porque entre fungos que degradam lignocelulose, *T. reesei*, é o mais eficiente degradador, apesar do seu arsenal enzimático ser menor.

CAPÍTULO III



7 CAPÍTULO III

7.1 RESULTADOS

7.1.1 Perfil de expressão gênica da linhagem mutante (*Δxyr1*) comparado à linhagem parental (QM9414) em diferentes fontes de carbono

7.1.1.1 Sequenciamento e alinhamento

Foram sequenciados nove bibliotecas correspondentes à linhagem mutante *Δxyr1* cultivada em celulose, soforose e glicose. O sequenciamento resultou em cerca de 144 milhões de *reads* de 100 pb *paired-end*, correspondendo a 28.75GB (Tabela 22). Essas *reads* foram mapeadas com o genoma de referência de *T. reesei* disponíveis no JGI (http://genome.jgi-psf.org/Trire2/Trire2.home.html), usando o algoritmo *Bowtie* (LI; DURBIN, 2009), sendo que cerca de 68,32% das *reads* obtidas em cada biblioteca foram mapeadas unicamente no genoma de referência.

Tabela 22 – Reads obtidas por RNA-seq (*Illumina HiSeq^{TM} 2000*) para a linhagem *Δxyr1*.

Amostras	***Reads* brutas**	***Reads* de qualidade**	***Reads* Mapeadas**	**% *Reads* Mapeadas**	**Número de nucleotídeo (Gigabase/GB)**
Δxyr1 Cel1	10623635	9062198	6417643	70,82%	1,28
Δxyr1 Cel2	15136419	13767140	9529046	69,22%	1,91
Δxyr1 Cel3	13582541	12320492	8514772	69,11%	1,70
Δxyr1 Sof1	18340537	15481677	10277748	66,39%	2,06
Δxyr1 Sof2	21571982	19650430	13249815	67,43%	2,65
Δxyr1 Sof3	50039657	49960458	31806513	63,66%	6,36
Δxyr1 Gli1	47624261	47537831	30830313	64,85%	6,17
Δxyr1 Gli2	25791799	23575558	17543037	74,41%	3,51
*Δxyr1*Gli3	24925235	22590723	15591702	69,02%	3,12
Total	227636066	213946507	143760589	68,32%	28.75

A avaliação da qualidade dos dados, ou seja, a remoção de dados não satisfatórios foram etapas essenciais para as análises dos dados da linhagem mutante. Uma vez que, para obtenção de genes diferencialmente expressos, precisou verificar a qualidade das amostras para que, posteriormente, fossem utilizadas em testes de expressão diferencial, eliminando, assim, amostras que poderiam ser prejudiciais nas análises dos dados. Os resultados a seguir foram obtidos através do software R, utilizando-se o pacote DESeq disponível no projeto *Bioconductor*. Altas correlações de *Pearson* ($R \geq 0{,}71$) foram obtidas quando as amostras da linhagem mutante *Δxyr1* foram comparadas (Figura 51). Entretanto, quando realizado o PCA das amostras de glicose entre as linhagens mutante versus parental, observou-se que uma das amostras *Δxyr1* Gli2 não foi informativa para diferenciar as duas linhagens (Figura 52). Dessa forma, a amostra *Δxyr1* Gli2 foi removida das análises subsequentes.

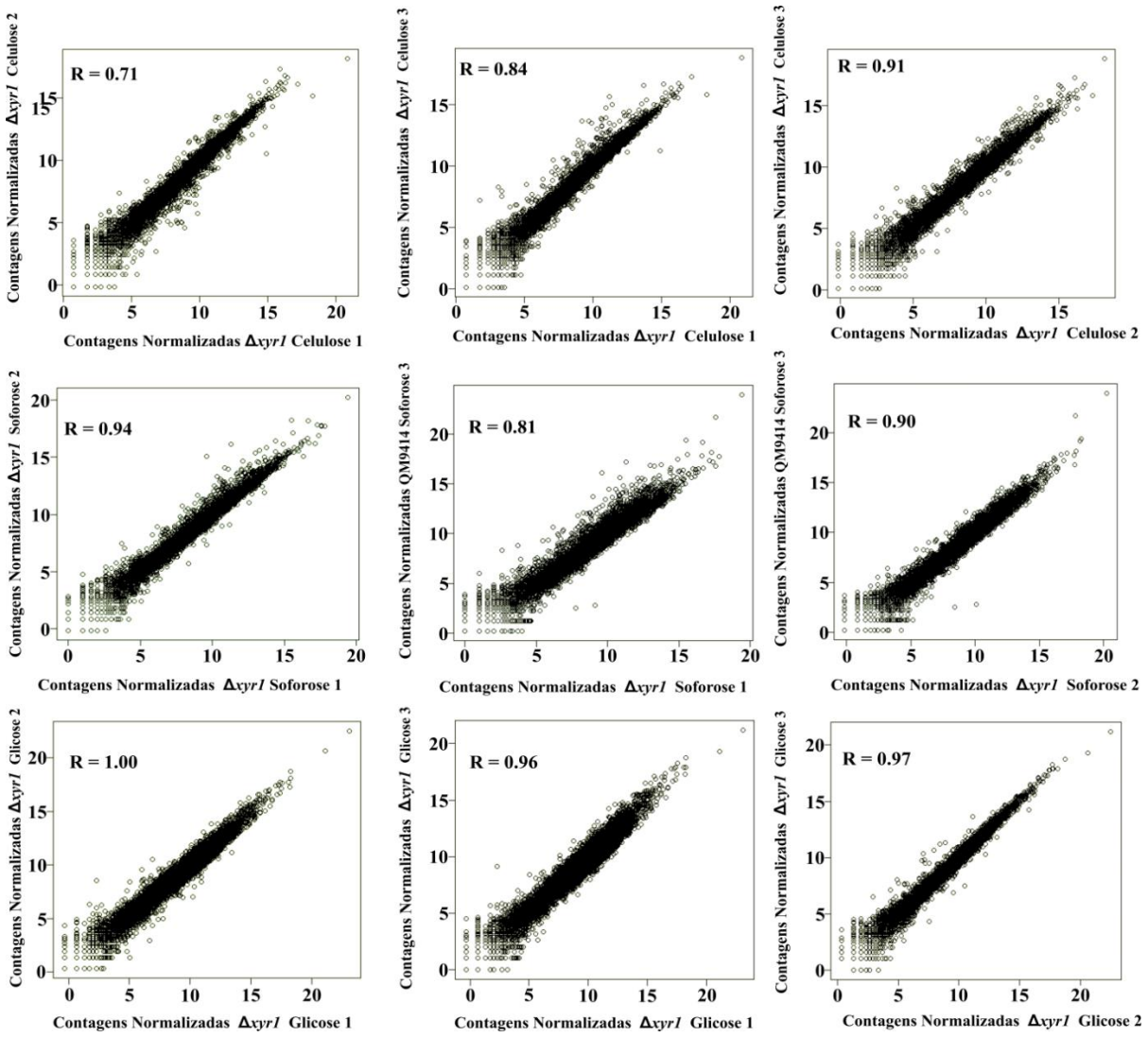


FIGURA 51 – Gráficos representando a correlação de *Pearson* entre as réplicas biológicas de cada condição para a linhagem mutante *Δxyr1*.

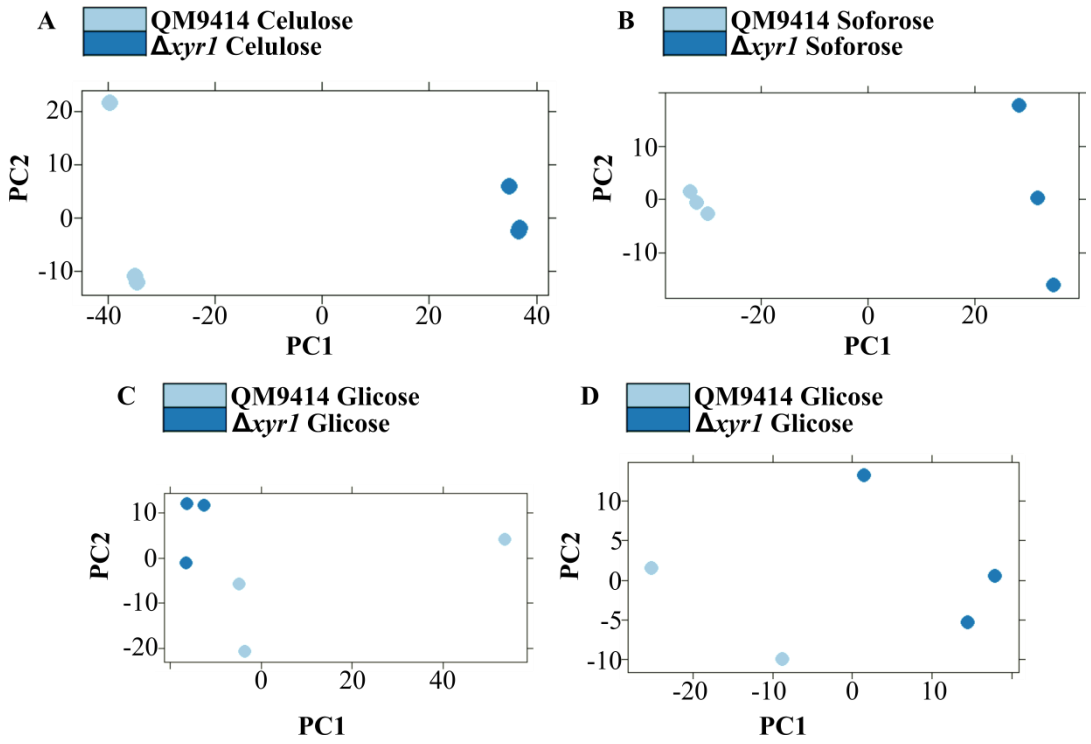


FIGURA 52 – Análise de Componentes Principais (PCA) das amostras da linhagem mutante em relação à linhagem parental.
(A) *Δxyr1*/QM9414 Celulose; (B) *Δxyr1*/QM9414 Soforose; (C) *Δxyr1*/QM9414 Glicose; (D) *Δxyr1*/QM9414 Glicose após exclusão de uma das réplicas biológicas.

Para a linhagem mutante também foi comprovado com a elaboração de gráficos *box plot* (dados brutos e após a etapa de normalização dos dados do RNA-seq) de todas as condições/amostras utilizadas e seus respectivos tempos de cultivos, demonstrando que são amostras comparáveis entre si (Figura 53). Após o alinhamento com o *software Bowtie*, os arquivos gerados em BAM foram utilizados para visualização dos alinhamentos das *reads* com o genoma de referência de *T. reesei* por meio da ferramenta IGV. Nesse caso específico, a região correspondente ao fator de transcrição XYR1 foi utilizada para verificar a presença e ausência deste fator nas linhagens parental (QM9414) e mutante (*Δxyr1*), respectivamente, confirmando, portanto, que a linhagem mutante não possui o fator de transcrição XYR1 (responsável pela produção de enzimas de interesse industrial), eliminando qualquer suspeita de contaminação durante o crescimento da linhagem mutante nas condições celulose, soforose e glicose (Figura 54). Na condição soforose, a região 5'-UTR (região 5' não traduzida – UTR = *untranslated region* – que se estende até o códon de iniciação que, apesar de ser representada na Figura 54, não permite a síntese protéica, isto é, a região que compreende o gene *xyr1* foi deletado.

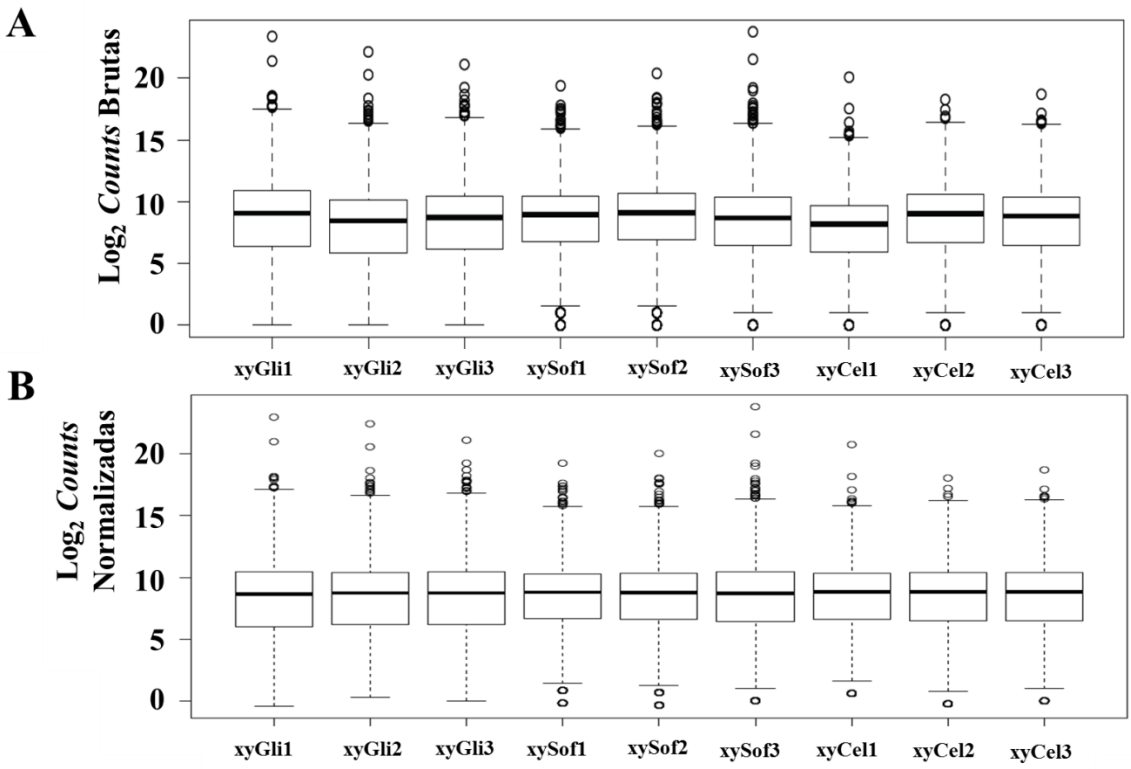


FIGURA 53 – *Box plot* das amostras utilizadas nas análises bioinformáticas dos dados do RNA-seq na linhagem mutante.
(A) Dados brutos; (B) Dados normalizados.

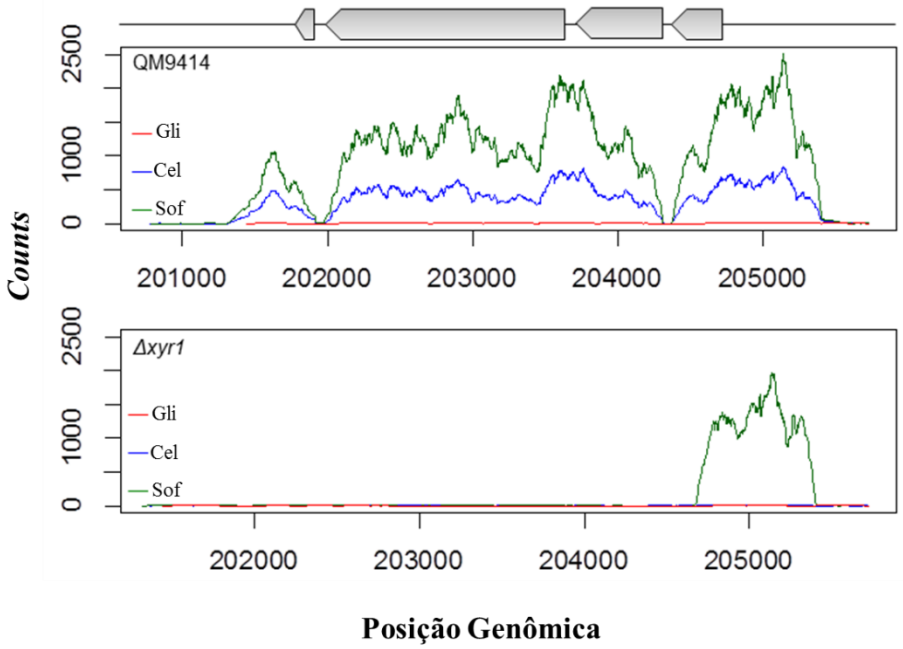


FIGURA 54 – Visualização da região genômica correspondente ao gene *xyr1* de *T. reesei* utilizando dados do RNA-seq.
Esta figura exibe contagens dos números de *reads* alinhados a região do gene *xyr1* (Proteína ID 122208, scaffold 11, posição genômica 201774 – 204723).

7.1.1.2 Genes Diferencialmente Expressos

Neste estudo, verificou-se a influência da ausência do fator de transcrição XYR1 na expressão gênica de *T. reesei*, em que, a linhagem mutante *Δxyr1* foi comparada à parental (QM9414) nas mesmas condições. Assim, dos 9129 genes que compõem o genoma de *T. reesei*, 2185 genes foram diferencialmente expressos em celulose, 2124 genes em soforose e apenas 46 genes em glicose (Figura 55). Gráficos *Volcano plot* também foram construídos para demonstrar quantos desses genes foram up- ou down-regulados, isto é, utilizou-se o limiar já aplicado anteriormente (*p-valor* $\leq 0{,}05$ e Log_2 *Fold Change* ≥ 1 or ≤ -1) nos dados gerados, identificando os genes up- e down-regulados, resultando no total 1379 genes (Figura 56). Diagramas de *Venn* dos genes up- e down-regulados em resposta às condições estudadas (Figura 57), apresentam a maioria dos genes modulados nas duas condições indutoras e poucos genes foram modulados na condição repressora. No total, foram identificados 749 genes up-regulados, sendo 333 genes específicos da condição celulose, 364 genes da condição soforose e 18 genes da condição glicose. Entre os genes down-regulados, 623 genes foram identificados, sendo 380 genes específicos da condição celulose, 221 genes da condição soforose e 8 genes da condição glicose (Figura 57).

Analisando o perfil transcricional dos 1379 genes, diferencialmente expressos tanto up- como down-regulados, construiu-se um *heatmap* para verificar a clusterização dos genes e condições (Figura 58). Dessa forma, foram identificados 6 possíveis regulons, sendo 3 deles compostos por genes up-regulados (294 genes controlados por celulose; 338 genes controlados por soforose e 15 genes controlados por glicose) e outros 3 regulons compostos por genes down-regulados (350 genes controlados por celulose; 181 genes por soforose e 8 genes por glicose), totalizando 1186 genes regulados pelas condições analisadas. Por meio da categorização funcional, baseando-se no *Gene Ontology* (GO), dos 1186 genes, foi possível destacar que a maioria dos genes tem função desconhecida e reforça a importância de outros componentes na regulação da transcrição de genes de tal maneira que, mesmo com a ausência do principal fator de transcrição envolvido na produção de enzimas de interesse biotecnológico, muitos genes sofreram regulação. A lista de todos os genes dos regulons identificados para a linhagem mutante comparado a linhagem parental pode ser verificada no CD Anexo (Arquivo 12, Tabelas 12.1 a 12.6).

Analisando-se os 1379 genes identificados como diferencialmente expressos entre as linhagens mutante e parental, pergunta-se: como esses genes se inter-relacionam? Existem

genes comuns às três condições? Quantos genes são específicos de cada condição analisada? Portanto, do mesmo modo que foi realizada para a linhagem parental QM9414, construiu-se uma rede regulatória dependente das fontes de carbono, utilizando todos estes genes modulados positivamente ou negativamente (Figura 59). Assim, 643 genes são específicos somente da condição celulose, 517 genes da condição soforose e 23 genes da condição glicose. Os genes que se inter-relacionam são: 170 genes entre as condições celulose e soforose, 14 genes entre as condições soforose e glicose, 7 genes entre as condições celulose e glicose. Apenas 2 genes são regulados de maneira comum nas três condições analisadas (Figura 59 – para mais detalhes ver CD Anexo, Arquivo 13, Tabelas 13.1 a 13.10 para identificação de todos os genes da rede regulatória).

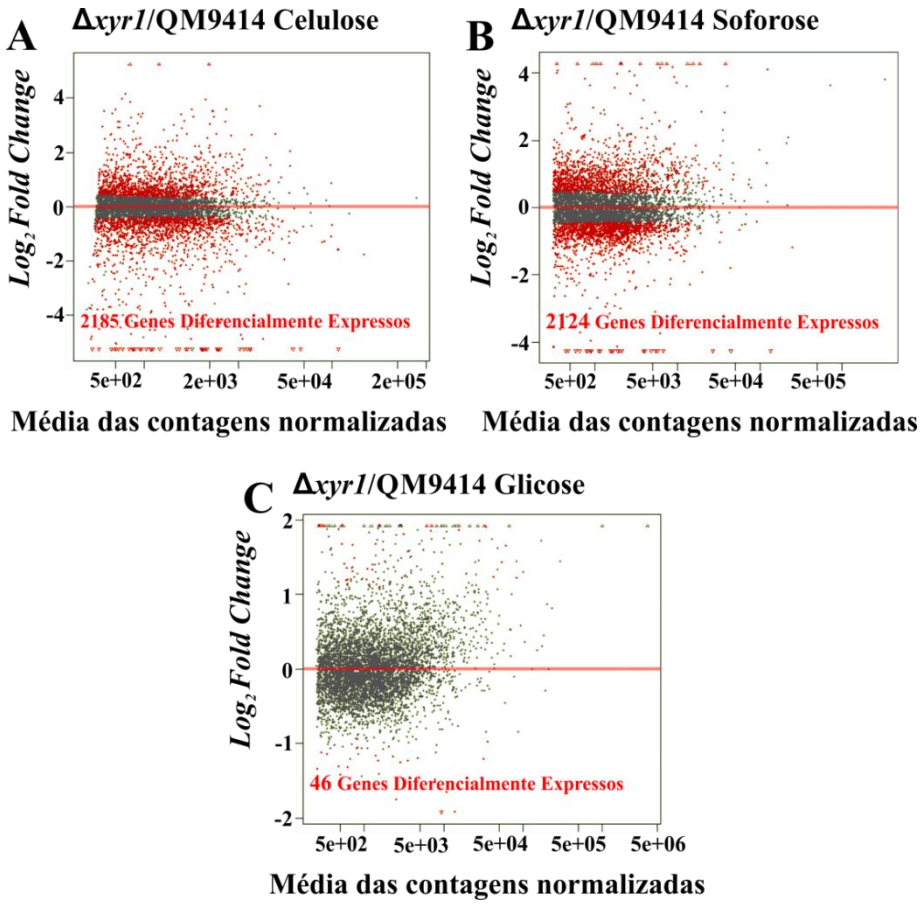


FIGURA 55 – Comparação do perfil de expressão gênica da linhagem mutante *Δxyr1* comparado com a linhagem parental QM9414 crescidos em celulose, soforose e glicose como fontes de carbono obtidos por RNA-seq.
(A) *Δxyr1*/QM9414 Celulose; (B) *Δxyr1*/QM9414 Soforose e (C) *Δxyr1*/QM9414 Glicose. Genes diferencialmente expressos obtidos pelo pacote DESeq estão plotados em vermelho.

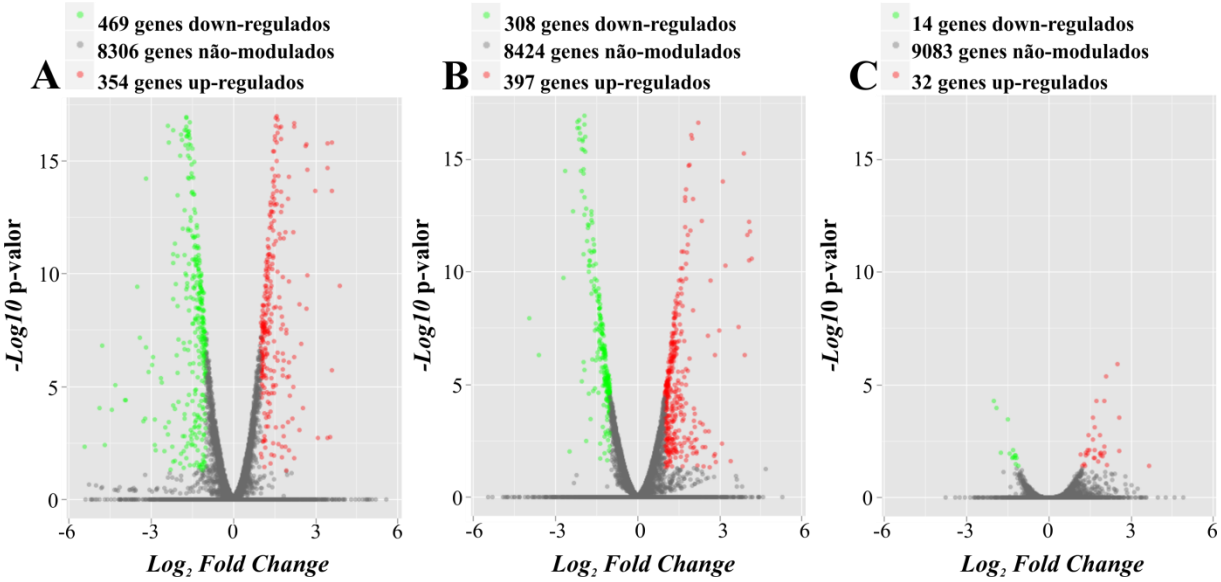


FIGURA 56 – Gráficos *volcano plot* dos genes diferencialmente expressos up- e down-regulados nas condições estudadas (*Δxyr1*/QM9414).
(A) *Δxyr1*/QM9414 Celulose; (B) *Δxyr1*/QM9414 Soforose e (C) *Δxyr1*/QM9414 Glicose. Genes up-regulados (Log_2 *Fold Change* ≥ 1 e *p-valor* 0,05) estão destacados em pontos vermelhos e genes down-regulados (Log_2 *Fold Change* ≤ -1 e *p-valor* 0,05) estão em verde.

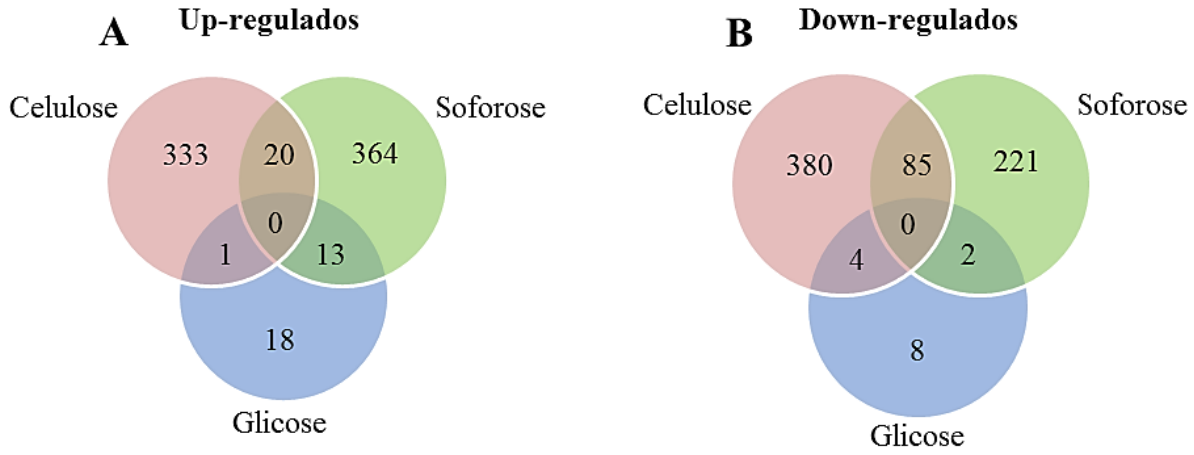


FIGURA 57 – Diagramas de *Venn* representando o número de genes diferencialmente expressos na linhagem mutante *Δxyr1* comparado com a linhagem parental QM9414.
(A) Genes up-regulados. (B) Genes down-regulados.

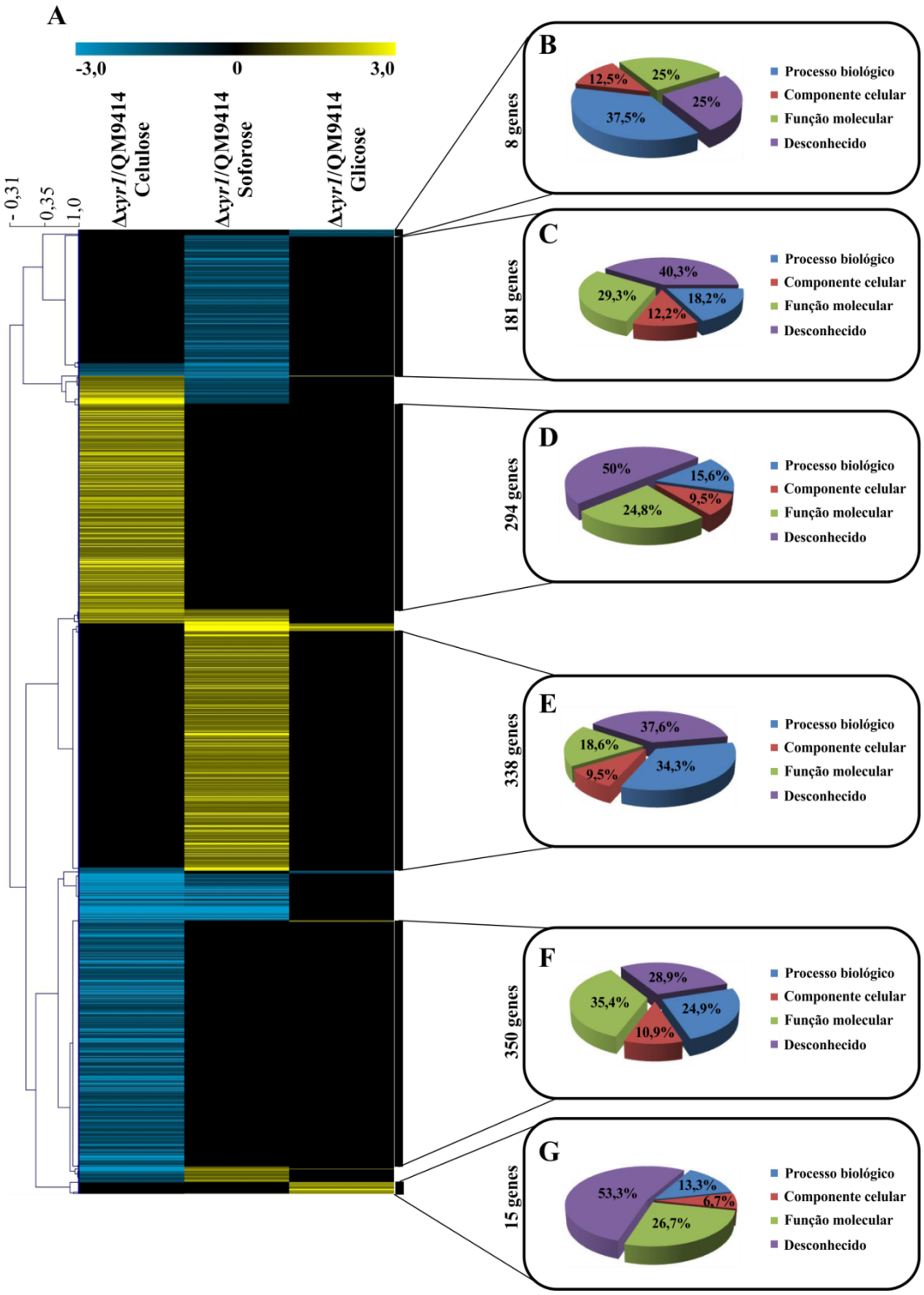


FIGURA 58 – Perfil de expressão de genes de *T. reesei* durante crescimento na presença de celulose, soforose e glicose como fontes de carbono na linhagem mutante *Δxyr1* comparado com a linhagem parental QM9414 e categorização funcional baseado no *Gene Ontology*.
(A) Clusterização hierárquica dos genes significativos up- e down-regulados em *T. reesei*. Utilizou-se o *software Mev v.4.6.1*, com o método "*average linkage*" para gerar os *clusters* e correlação "*uncentered*". A distância Euclidiana foi usada para verificar as diferenças entre os 1379 genes e os grupos (condições) baseado na distância do centro dos grupos, $p < 0,05$. (B) Genes down-regulados em glicose. (C) Genes down-regulados em soforose. (D) Genes up-regulados em celulose. (E) Genes up-regulados em soforose. (F) Genes down-regulados em celulose. (G) Genes up-regulados em glicose.

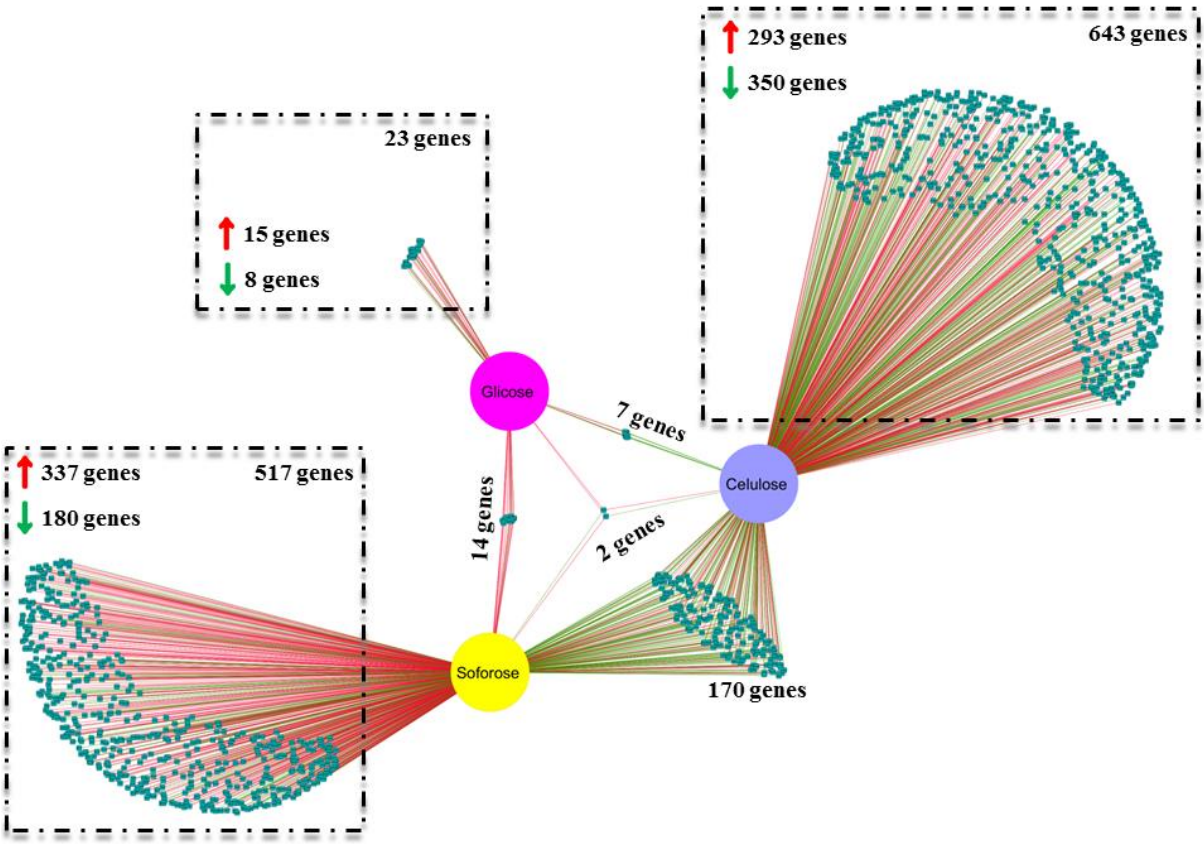


FIGURA 59 – Rede regulatória de 1379 genes diferencialmente expressos na linhagem mutante *Δxyr1* comparado com a linhagem parental QM9414 em resposta a diferentes fontes de carbono.
Genes são representados como "*nodes*" (quadrados azuis), e as interações são representadas como linhas, interações vermelhas (up-regulados) e interações verdes (down-regulados), que conectam os *nodes* com 1574 interações.

Dentre os genes diferencialmente expressos foram destacados os top 15 genes down-regulados na ausência do fator de transcrição XYR1 (Tabela 25). A ausência deste fator afetou a expressão de glicosil hidrolases (GHs) nas duas condições indutoras, sendo 8 GHs em ambas condições. As mais down-reguladas na condição celulose foram: GH12 (ID123232, -11.71); GH61 (ID120961, -10.95); GH10 (ID120229, -9.99). Em soforose: GH7 (ID123989, -8.95) e GH6 (ID72567, -8.91). Porém, na condição repressora, somente a GH16 (ID50215, -1,48) foi afetada. Duas glicosil hidrolases, GH5 (ID56996, -9.66) e GH11 (ID123818, -8.77) foram exclusivamente as top down-reguladas em celulose. Enquanto que, na condição soforose outras glicosil hidrolases foram exclusivamente identificadas: GH3 (ID121127, -7.95), GH5 (ID49976, -7.68), GH115 (ID123940, -7.56) e GH7 (ID122081, -6.59).

Em glicose, entre os top genes down-regulados estão uma HFB1[13] (ID73173, -2.00) e outros 13 genes (não sendo verificado estes genes nas outras duas condições) que estão apresentados na Tabela 23. Com esses dados, nota-se que o regulador XYR1 é essencial na expressão de enzimas importantes para a degradação da biomassa.

Tabela 23 – Top 15 genes down-regulados em celulose, soforose e glicose na ausência do regulador XYR1.

Celulose			
Proteína ID	**Descrição**	**FC**	**GO (Categoria Principal)**
123232	GH12 endo-ß-1,4-glucanase	-11,71	molecular_function
120961	GH61 polysaccharide monooxygenase CEL61b	-10,95	#N/D
120229	GH10 endo-ß-1,4-xylanase XYN3	-9,99	biological_process
56996	GH5 β-Mannanase MAN1	-9,66	biological_process
120312	GH5 endo-β-1,4-glucanase EGL2/CEL5a	-9,54	molecular_function
69957	MFS permease	-9,1	cellular_component
73638	CIP1	-9,06	molecular_function
72567	GH6 Cellobiohydrolase CEL6A/CBH2	-9,05	molecular_function
123989	GH7 Cellobiohydrolase CBH1/CEL7a	-8,84	molecular_function
123818	GH11 endo-ß-1,4-xylanase XYN2	-8,77	biological_process
73643	GH61 polysaccharide monooxygenase CEL61a	-8,63	molecular_function
121418	lipase G-D-S-L	-8,59	molecular_function
72526	GH67 α-Glucuronidase GLR1	-8,49	molecular_function
73632	CE5 acetyl xylan esterase AXE1	-8,46	molecular_function
Soforose			
Proteína ID	**Descrição**	**FC**	**GO (Categoria Principal)**
123989	GH7 Cellobiohydrolase CBH1/CEL7a	-8,95	molecular_function
72567	GH6 Cellobiohydrolase CEL6A/CBH2	-8,91	molecular_function
73638	CIP1	-8,49	molecular_function
123232	GH12 endo-ß-1,4-glucanase	-8,29	molecular_function
121418	lipase G-D-S-L	-8,2	molecular_function
72526	GH67 α-Glucuronidase GLR1	-8,18	molecular_function
121127	GH3 β-xylosidase BXL1	-7,95	biological_process
120312	GH5 endo-β-1,4-glucanase EGL2/CEL5a	-7,84	molecular_function
49976	GH5 endo-β-1,4-glucanase EGL5/CEL45a	-7,68	molecular_function
120961	GH61 polysaccharide monooxygenase CEL61b	-7,65	#N/D
56840	unknown protein	-7,6	molecular_function
123940	GH115 methylglurunoyl esterase CIP2	-7,56	molecular_function
122081	GH7 Endo-β-1,4-glucanase EGL1/CEL7b	-6,59	molecular_function
120229	GH10 endo-ß-1,4-xylanase XYN3	-6,42	biological_process
73643	GH61 polysaccharide monooxygenase CEL61a	-6,29	molecular_function
Glicose			
Proteína ID	**Descrição**	**FC**	**GO (Categoria Principal)**
73173	HFB1	-2	biological_process
50212	GT Glycosyltransferases not yet assigned to a family	-1,91	#N/D
65522	unknown protein	-1,75	biological_process
50215	GH16 endo-1,3-β-D-glucosidase/1,3-glucan binding protein	-1,48	biological_process
76784	sorbitol dehydrogenase	-1,42	molecular_function
42053	unknown protein	-1,33	cellular_component
78738	sodium/hydrogen exchanger family protein	-1,31	#N/D
112499	Zn2Cys6 transcriptional regulator	-1,3	molecular_function
78864	GcvTGlycine cleavage system T protein	-1,24	biological_process
73937	Ornithine carbamoyltransferase OTC/ARG3	-1,23	biological_process
73631	isoamyl alcohol oxidase	-1,2	biological_process
80268	fumarate lyase	-1,19	molecular_function
75294	BCAT_beta_family	-1,17	biological_process
120370	unknown protein	-1,11	#N/D

*#N/D = Desconhecido.

[13] São um grupo de pequenas proteínas ricas em cisteína (com aproximadamente 100 aminoácidos) que são expressas apenas por fungos filamentosos. Com base em diferenças nos padrões de hidropatia e propriedades biofísicas, dividem-se em duas categorias: classe I e classe II. São geralmente encontradas na superfície exterior dos conídios e da parede das hifas, e podem estar ligadas à mediação de contato e comunicação dos fungos com o seu ambiente (HAKANPÄÄ et al., 2006).

Após determinação dos perfis transcricionais, foi realizada a categorização funcional dos genes diferencialmente expressos de acordo com o *Gene Ontology*. As categorias super-representadas em cada uma das condições analisadas, foram determinadas a partir da análise de enriquecimento implementada pelo algoritmo *BayGO* (VÊNCIO et al., 2006). Na Figura 60, verifica-se a participação de poucos genes na condição glicose, sendo a maioria das funções encontradas exclusivamente nesta condição foram: *copper ion trnasmembrane transporter activity*, *monooxygenase activity*, *ferric-chelate reductase activity*, *copper ion transport*, *nucleotide binding*, *ATPase activity coupled to transmembrane movement of substances* e *nucleoside-triphosphatase activity*, todos encontrados de maneira up-regulados. Apenas *catalytic activity* foi identificada como down-regulado em glicose. Somente a categoria *electron transport* foi comum à condição repressora (glicose) e a condição indutora (soforose). Na condição indutora celulose, a categoria *metabolic process* foi a mais afetada na ausência do fator XYR1 seguida por: *oxidoreductase activity*, *membrane*, *transport*, *integral to membrane* e *electron transport*. Outras categorias foram super-representadas tanto em celulose como em soforose: *oxidoreductase activity*, *carbohydrate metabolic process*, *hydrolase activity hydrolyzing O-glycosyl compounds*, *sugar* (*hydrogen symporter activity*) e *carbohydrate transport*.

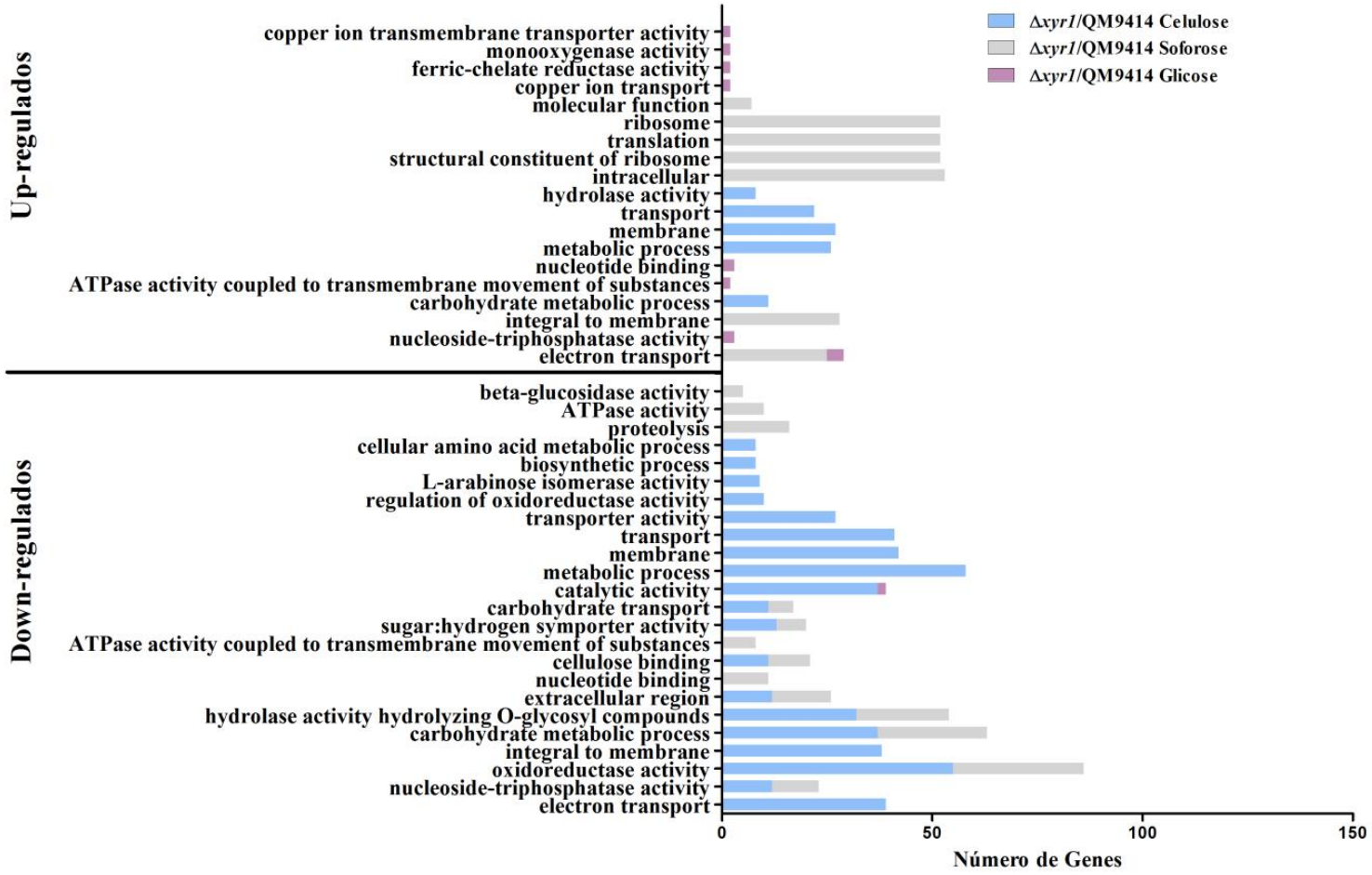


FIGURA 60 – Categorização funcional de acordo com o GO de genes diferencialmente expressos na linhagem mutante (*Δxyr1*) comparado com a linhagem parental (QM9414).
Análise de enriquecimento de genes up- e down-regulados em *Δxyr1*/QM9414 em celulose, soforose e glicose como fontes de carbono. Os termos enriquecidos do GO correspondem à função molecular, componente celular e processo biológico em *T. reesei*. Categorias significativamente enriquecidas são mostradas ($p \leq 0,05$).

7.1.1.3 Expressão de genes *Carbohydrate Active enZymes* (*CAZy*) identificados em *T. reesei Δxyr1*/QM9414

Do mesmo modo que foi observado para os genes *CAZy* na linhagem parental QM9414 (CASTRO et al., 2014b), utilizou-se a média do FPKM para todos os genes dentro de uma mesma família GH em *Δxyr1*/QM9414, em todas as condições analisadas. Assim, na Figura 61, é demonstrado o perfil dos genes *CAZy* na ausência do regulador XYR1, sendo que foi observado uma drástica redução nas três fontes de carbono quando comparado à linhagem parental, principalmente GH7 e GH6. Com dados obtidos do RNA-seq, foi demonstrado que das 228 enzimas classificadas como *CAZy* em *T. reesei* (HÄKKINEN et al., 2012), 95 foram moduladas significativamente ($p \leq 0,05$) em celulose, 71 em soforose e apenas 3 genes *CAZys* em glicose foram identificadas na linhagem mutante quando comparada à linhagem parental. Como resultados esperados, nas duas condições indutoras, a maioria das enzimas foram down-reguladas em *Δxyr1*/QM9414.

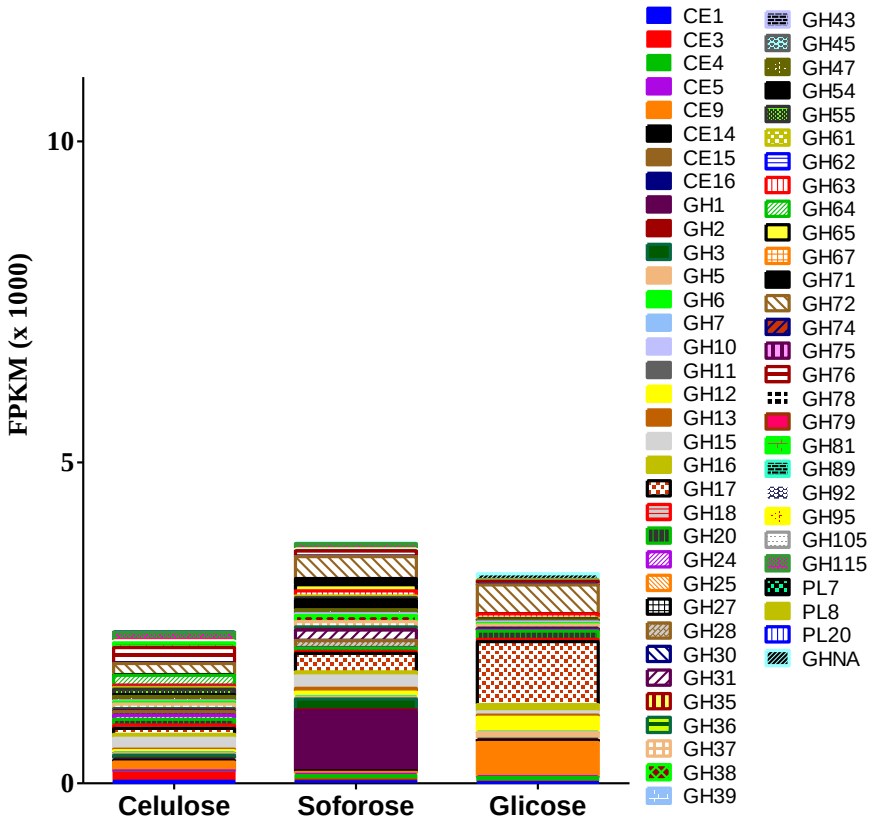


FIGURA 61 – Genes *CAZy* e expressão das análises por RNA-seq em *Δxyr1*/QM9414 .
Fragmentos por kilobase de exon por milhões fragmentos mapeados (FPKM) foram obtidos da média de cada família glicosil hidrolase (GH) quando cultivado em presença de celulose, soforose e glicose. A classificação *CAZy* foi realizada de acordo com a re-anotação dos genes de *CAZy* de *T. reesei* de acordo com Häkkinen e colaboradores (2012).

É interessante destacar que, dos genes mencionados anteriormente como modulados significativamente em cada condição, 41 genes foram especificamente identificados em celulose, 18 genes na condição soforose e 2 genes na condição repressora (estes genes estão destacados em vermelho nas Tabelas 14.1 a 14.3 no Arquivo 14 no CD em Anexo). Dos 27 genes modulados em celulose, 10 foram up- e 17 down-regulados. Em soforose, 6 genes sofreram regulação, sendo 3 genes up- e 3 genes down-regulados. Na condição repressora, apenas a regulação positiva foi identificada em 2 genes. Dentre estas *CAZy* identificadas, a α-xylosidase/α-glucanase (GH31, ID69944, -100) e β-mannanase *Man1* (GH5, ID56996, -9,66) foram as GHs que mais sofreram regulação negativa na condição celulose. A β-glucosidase *cel3d* (GH3, ID46816, -2,39), em soforose, foi a mais afetada na ausência do fator de transcrição XYR1. Outras GHs moduladas sugerem uma possível regulação não dependente do regulador XYR1, destacando em celulose duas endo-1,3-β-glucanases (GH81, ID79602, 3,60; GH64, ID124175, 3,54); em soforose uma α,α-trehalase (GH37, ID123226, 1,45), e, em glicose a GH61 (ID27554, 2,02) e uma GH com anotação não disponível (ID121136, 1,74) todas sendo alvos de uma possível regulação indireta.

7.1.1.4 Expressão de genes que codificam fatores de transcrição identificados em *T. reesei Δxyr1*/QM9414

Além da regulação da produção de enzimas envolvendo o fator de transcrição XYR1, foram identificados outros fatores, principalmente fatores transcricionais (FTs) que sofreram regulação na ausência do fator XYR1. Com as análises dos dados obtidos do RNA-seq, 31 fatores foram regulados em celulose (17 up- e 14 down-regulados); 30 regulados em soforose (14 up- e 16 down-regulados) e, em glicose 1 up-regulado e 1 down-regulado (Tabela 24), sendo 46,03% da família Zn2Cys6. O regulador "Zn2Cys6 transcription regulator *C. albicans* Fcr1" (ID122271, 3.52) apresentou o maior nível de expressão em presença do indutor celulose e o regulador "BZIP transcriptional regulator" (ID110152, - 2.40) o mais baixo nível de expressão na mesma condição indutora; quatro reguladores Zn2Cys6 (IDs 122523, 121121, 105520 e 112499; 2.97, - 1.77, 1.35 e - 1.30, respectivamente) foram identificados, sendo os dois primeiros em soforose e dois últimos em glicose. Esses dados sugerem, mais uma vez, a possível regulação direta ou indireta e a participação de outros fatores na produção de enzimas de interesse biotecnológico.

Tabela 24 – Principais fatores transcricionais modulados na ausência do regulador XYR1.

	Up-Regulados			Down-Regulados		
Condição	**Proteína ID**	**Fatores**	***Log₂ Fold Change***	**Proteína ID**	**Fatores**	***Log₂ Fold Change***
Celulose	120365	BZIP transcriptional regulator	1,34	110152	BZIP transcriptional regulator	-2,40
	67418	C2H2 transcription factor	1,48	58011	C2H2 conidiation transcription factor FlbC	-1,79
	54703	C2H2 transcriptional regulator	1,24	119826	C2H2 transcriptional regulator	-1,70
	77878	C2H2 transcriptional regulator	1,46	41617	lae1	-1,01
	23190	mRNA cleavage factor complex II protein Clp1. putative	1,10	72259	RgsA. regulator of G-protein signaling	-1,40
	122864	pre-mRNA splicing factor	1,16	54675	Transcriptional regulator PacG/VIB-1	-1,79
	52924	transcriptional activator. zinc finger. NF-X1-type	1,22	36543	transcriptional regulator STE12	-1,10
	49873	transcriptional regulator. unknown	1,07	53484	transcriptional regulator transcriptional regulator APSES type	-1,06
	122271	Zn2Cys6 transcription regulator. C. albicans Fcr1	3,52	68254	Zn2Cys6 transcriptional regulator	-1,53
	62244	Zn2Cys6 transcriptional regulator	1,93	121415	Zn2Cys6 transcriptional regulator	-1,22
	53067	Zn2Cys6 transcriptional regulator	1,46	121894	Zn2Cys6 transcriptional regulator	-1,68
	59067	Zn2Cys6 transcriptional regulator	1,09	80291	Zn2Cys6 transcriptional regulator	-1,43
	58389	Zn2Cys6 transcriptional regulator	1,11	26871	Zn2Cys6 transcriptional regulator	-1,24
	60627	Zn2Cys6 transcriptional regulator	1,03	39221	Zn2Cys6 transcriptional regulator	-1,18
	73792	Zn2Cys6 transcriptional regulator	1,18			
	111446	Zn2Cys6 transcriptional regulator	1,19			
	72611	Zn2Cys6 transcriptional regulator	1,75			
Soforose	120698	C2H2 transcriptional regulator	1,05	57322	ARO Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family)	-1,04
	55422	elongation factor 2 kinase. putative	1,17	73654	BZIP transcriptional regulator	-1,02
	120969	NADPH oxidase regulator NoxR	1,69	35552	C2H2 transcription factor	-1,40
	80709	pre-mRNA-splicing factor CWC21. putative	1,10	120428	C2H2 transcriptional regulator	-1,00
	122457	transcriptional activator Helix-turn-helix type 3	1,57	121148	C2H2 transcriptional regulator	-1,20
	74346	Translation elongation factor precursor from Aspergillus fumigatus.	1,19	36391	C2H2 transcriptional regulator	-1,13
	71559	Translation initiation factor eIF-5A gene based on homologyToThe corresponding genes in several filamentous fungi.	1,23	105643	myb transcriptional regulator	-1,11
	33359	translation initiation factor SUI1 and density-regulated protein.	1,38	5659	pre-mRNA splicing factor. putative [Cryptococcus neoformans].	-1,30
	76927	translation initiation factor SUI1.	1,09	26575	transcriptional regulator. unknown	-1,09
	122523	Zn2Cys6 transcriptional regulator	2,97	70351	Zn2Cys6 transcriptional regulator	-1,60
	122783	Zn2Cys6 transcriptional regulator	1,15	21997	Zn2Cys6 transcriptional regulator	-1,19
	80139	Zn2Cys6 transcriptional regulator	1,34	123699	Zn2Cys6 transcriptional regulator	-1,09
	103138	Zn2Cys6 transcriptional regulator	1,11	123881	Zn2Cys6 transcriptional regulator	-1,01
	105263	Zn2Cys6 transcriptional regulator	1,53	121107	Zn2Cys6 transcriptional regulator	-1,09
				121121	Zn2Cys6 transcriptional regulator	-1,77
				102497	Zn2Cys6 transcriptional regulator	-1,38
Glicose	105520	Zn2Cys6 transcriptional regulator	1,35	112499	Zn2Cys6 transcriptional regulator	-1,30

Fatores transcricionais (Tabela 25) caracterizados por estar envolvidos na expressão de genes de celulases e xilanases também foram afetados pela ausência do regulador XYR1, isto é, quando se comparou a Tabela 25 com a Tabela 19 (Capítulo II, linhagem parental QM9414), verificou-se que vários fatores como ACE2, CLR-2, ACE1, CREI, HAP2 e HAP5 não apresentaram modulação significativa ($p > 0,05$), já os fatores CLR-1, BglR, HAP3, AreA e PacC sofreram regulação positiva ou negativa (de maneira condição-dependente) na ausência do ativador transcricional Zn2Cys6 (XYR1). Na condição repressora, nenhum dos fatores de transcrição analisado apresentou regulação. Com esses resultados, reforça-se a hipótese de que XYR1 tem um papel crítico na regulação de genes de celulases e xilanases (CASTRO et al., 2014a), uma vez que a ausência deste fator afetou a expressão de fatores conhecidos e novos fatores identificados neste trabalho, sugerindo que respondem a sinais específicos e desencadeiam respostas transcricionais.

Tabela 25 – Ausência do ativador transcricional Zn2Cys6 (XYR1) altera a expressão dos principais fatores de transcrição envolvidos na regulação de genes de celulases e xilanases.

Proteína ID	Nome	xyqmCelulose	xyqmSoforose	xyqmGlicose
122208	XYR1	-6,11	-6,12	ns*
78445	ACE2	ns	ns	ns
27600	CLR-1	ns	-0,81	ns
26163	CLR-2	ns	ns	ns
52368	BglR	ns	-0,50	ns
75418	ACE1	ns	ns	ns
120117	CREI	ns	ns	ns
124286	HAP2	ns	ns	ns
121080	HAP3	ns	0,55	ns
62979	HAP5	ns	ns	ns
76817	AreA	-0,52	-0,71	ns
120698	PacC	0,88	1,05	ns

*ns = Não significativo ($p > 0,05$). Os resultados apresentados estão em *Log_2 Fold Change*.

7.1.1.5 Expressão de transportadores identificados em *T. reesei Δxyr1*/QM9414

Movimento transmenbrana de solutos é vital para o crescimento de todas as células (YAN, 2013). Sabendo-se que a utilização de fontes de carbono interferem na expressão de genes, ou seja, ao serem metabolizados intervêm na produção de enzimas e, para tal processo de metabolização é requerido transportadores para facilitar a sua utilização por parte da célula. Por meio da análise dos dados de RNA-seq, foi possível identificar transportadores diferencialmente expressos especificamente em cada condição analisada, uma vez que, de

5,29% dos genes do genoma *T. reesei* correspondentes a transportadores (483 genes no total) foram identificados, e 63 desses transportadores são específicos (up- ou down-regulados na ausência do regulador XYR1), sendo 38 genes em celulose, 22 na condição soforose e apenas 3 transportadores na condição repressora (CD Anexo, Arquivo 15, Tabela 15).

Esses 63 genes foram utilizados para a construção de uma árvore filogenética para verificar relações de ancestralidade e descendência com transportadores de outros micro-organismos, como: *Saccharomyces cerevisiae*; *Neurospora crassa*; *Aspergillus sp*; *Escherichia coli*; *Metarhizium anisopliae*; *Talaromyces sp*; *Candida albicans*; *Ogataea angusta*; *Beauveria bassiana*; *Togninia mínima*; *Colletotrichum gloeosporioides*; e *Fusarium fujikuroi* (Figura 62). Essas análises filogenéticas permitiram identificar sete grupos de transportadores, sendo: dois grupos de *ABC transportes*, e outros cinco distintos grupos de transportadores (*Sugar transportes*, *MFS permeases*, *Aminoacid transporters*, *Potassium channels* e *Phosphate transporters*).

O maior grupo identificado, dentre todos os citados, foi o *Sugar transporters* composto por 7 proteínas de *T. reesei* (IDs 47710,62380, 104072, 121482, 121850, 62502 e 76800) e outras 8 proteínas de diferentes espécies (*C. albicans* (ID19.13383), *O. angusta* (IDACA58225.1), *N. crassa* (IDs NCU04963, NCU10021 e NCU00821) e *S. cerevisiae* (IDs YHR094C, YHR092C e YDR343C)). Os transportadores identidicados, nessas outras espécies, sugerem que os transportadores de *T. reesei*, que participam deste grupo, podem estar envolvidos no transporte de glicose nas duas condições indutoras (Figura 62). Um dos dois grupos formados por *ABC trasporters* destaca a participação de uma única proteína de *T. reesei* (ID59014) expressa, na condição repressora (up-regulada) que, juntamente com a proteína de *T. reesei* (ID76682), presente em celulose (down-regulada) revelaram uma similaridade com *ABC transporters* de *C. gloeosporioides*, *T. minima* e *M. anisopliae*. Outras duas proteínas de *T. reesei* expressas na condição glicose não foram agrupadas em nenhum dos grupos identificados. O segundo grupo *ABC trans*porters destacou a participação de dois genes comumente expressos em celulose (down-regulados) e que foram agrupados com *ABC transporter* de *A. niger* (IDXP_001397059.2). Similarmente a esse grupo, a proteína de *T. reesei* (ID79169) também foi expressa somente na condição celulose (down-regulada) e apresentou semelhança ao canal de potássio de *B. bassiana*. Outros três grupos (*MFS permeases*, *Aminoacid transporters* e *Phosphate transportes*) não apresentaram similaridade com outras espécies, mas se mostraram regulados somente nas condições indutoras, sugerindo que a regulação gênica nestas duas condições podem ser similares (Figura 62).

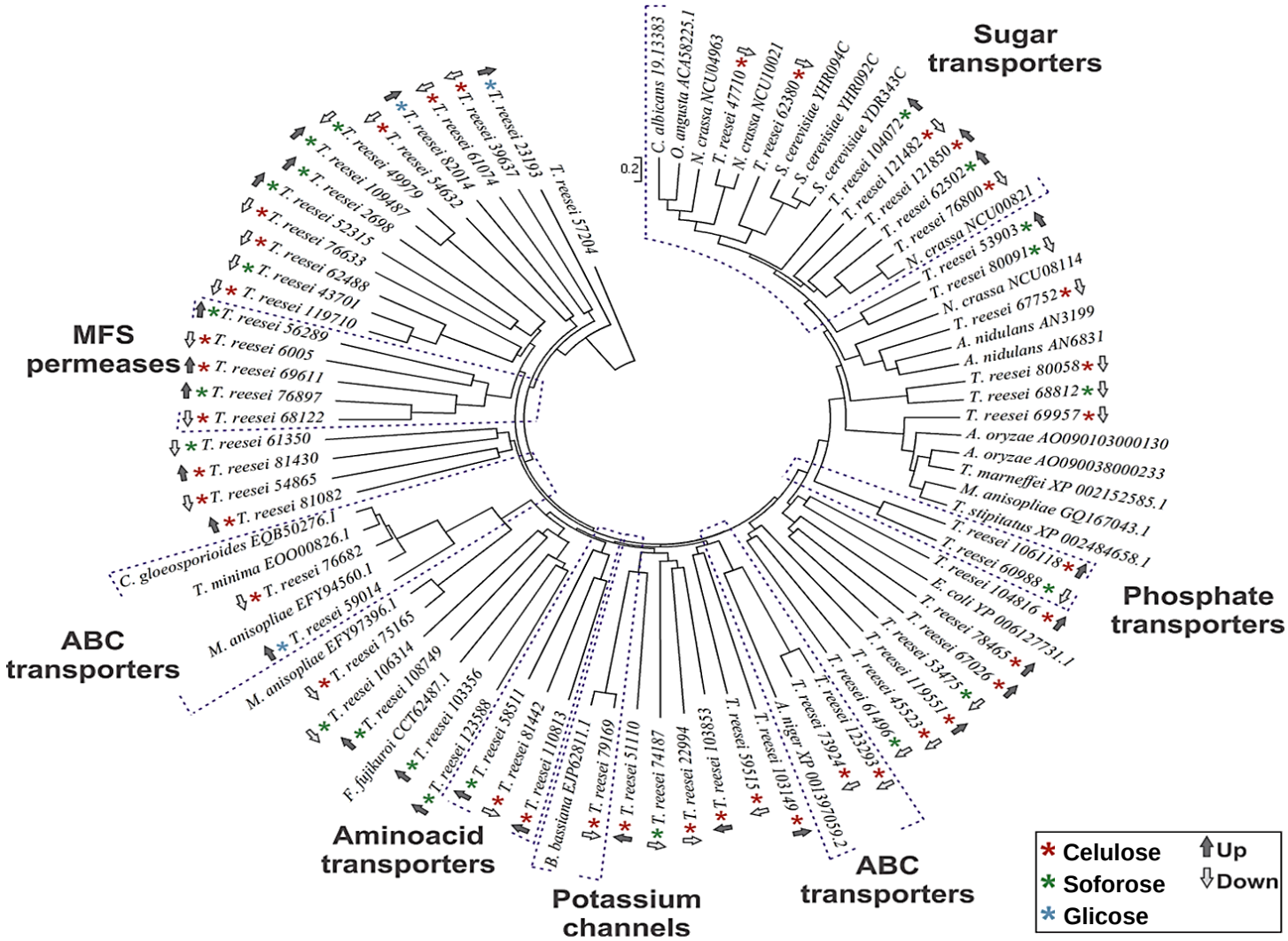


FIGURA 62 – Árvore filogenética contendo 63 genes que codificam transportadores identificados em *Δxyr1*/QM9414 nas três fontes de carbono estudadas.

7.1.1.6 Genes regulados diretamente pelo fator transcricional XYR1

Esta análise permitiu identificar genes que são possivelmente regulados diretamente pelo regulador XYR1, utilizando o $PWM_{XYR1,}$ descrito no Capítulo I, para pesquisar possíveis sítios de ligação de XYR1 nas regiões promotoras de genes diferencialmente regulados nas análises de RNA-seq. Assim sendo, partiu-se do pressuposto que os genes que possuem um ou mais sítios de ligação para o fator de transcrição XYR1 estão sob direta regulação, enquanto que a ausência de sítios de ligação sugere uma regulação indireta. As Figuras 63 e 64 enfatizam a regulação direta pelo ativador XYR1, identificado após análise, e as mesmas destacam genes que são exclusivos de regulação direta nas duas condições indutoras. Dentre os genes identificados, 96 foram modulados na condição celulose (39 up-regulados e 57 down-regulados) e 70 genes na condição soforose (47 up-regulados e 23 down-regulados), sendo que, cerca de 43,98% ainda não possuem função definida [para mais detalhes da caracterização desses genes ver CD Anexo (Arquivos 5 a 8)].

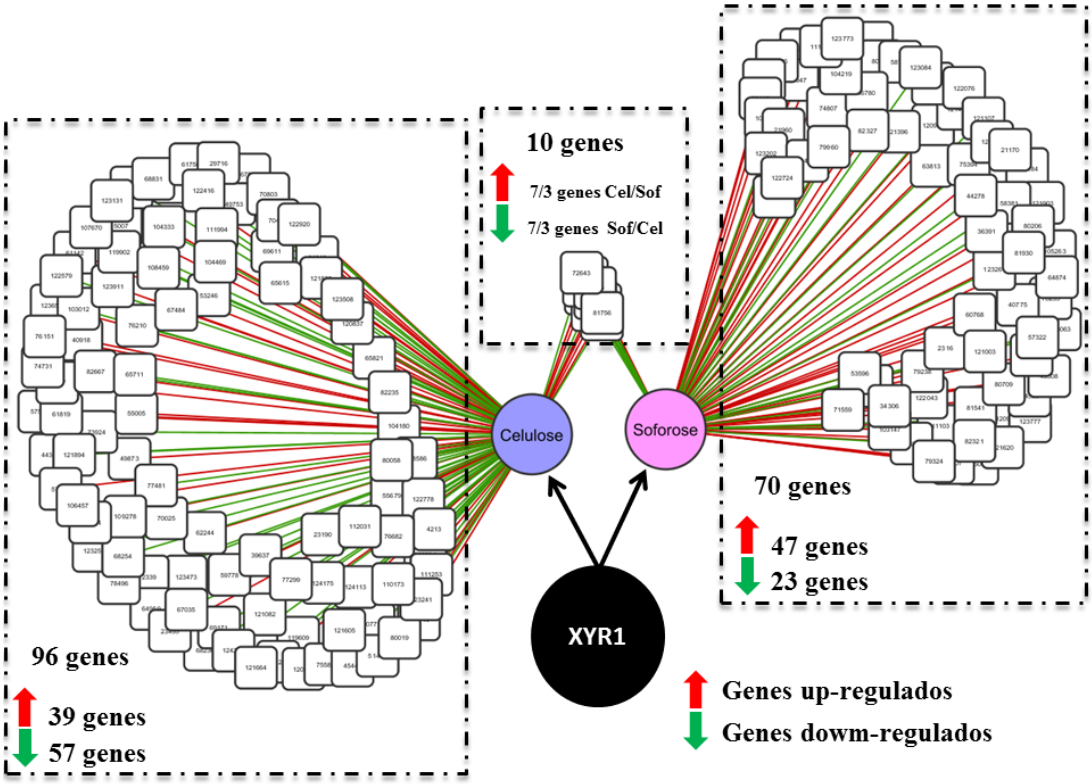


FIGURA 63 – Rede regulatória de genes regulados diretamente pelo regulador positivo XYR1.
Genes são representados como "*nodes*" (quadrados brancos), e as interações são representadas como linhas, interações vermelhas (genes up-regulados) e interações verdes (genes down-regulados), que conectam os *nodes* com 186 interações.

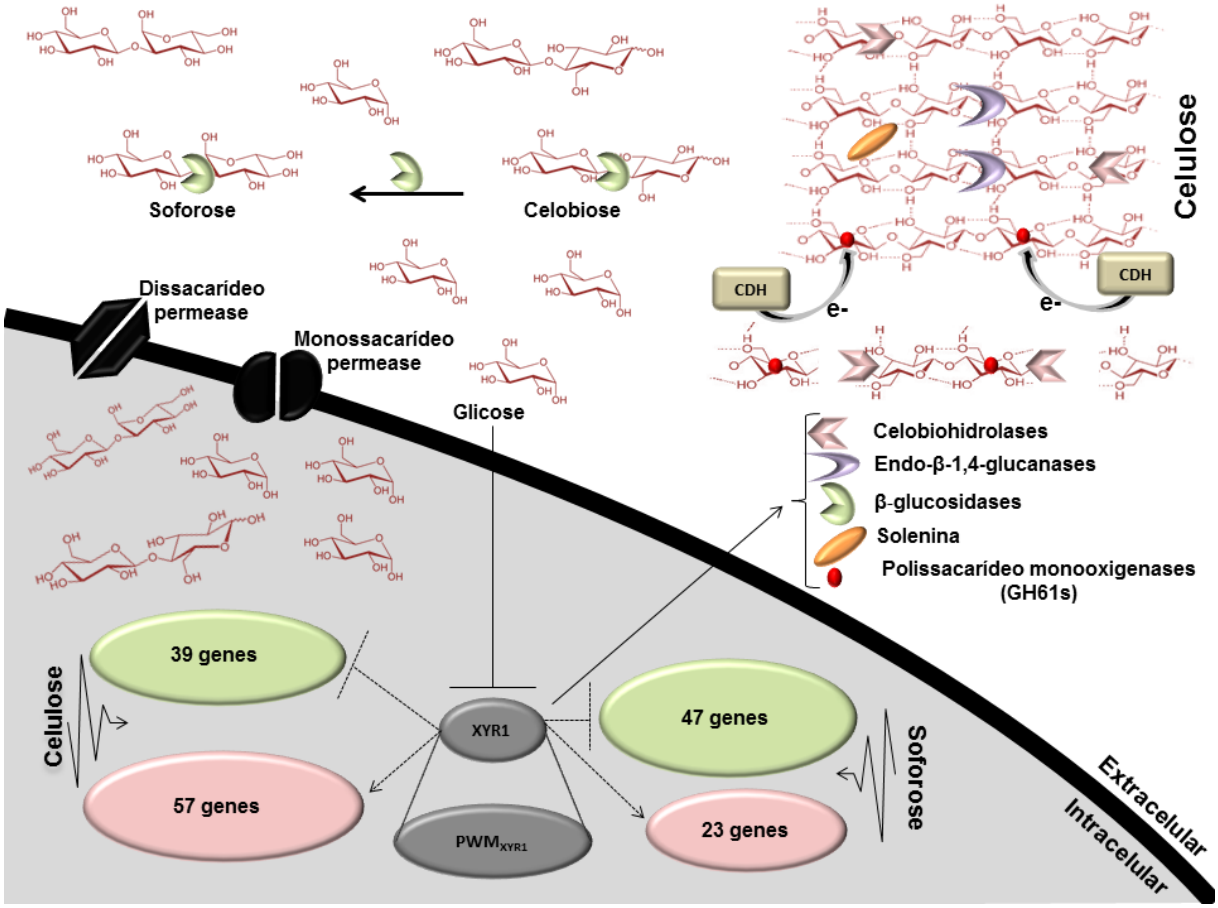


Figura 64 – Esquema envolvendo a regulação direta de XYR1 em *Trichoderma reesei*.

7.1.1.7 Análises de PCR quantitativa em Tempo Real (qPCR-TR)

Para validar os perfis de expressões de *Δxyr1*/QM9414 obtidos por análises de dados do RNA-seq, foram selecionados 20 genes diferencialmente expressos (CD Anexo, Arquivo 16, Tabela 16), nas condições indutoras, sendo 13 glicosil hidrolases (IDs: 123989, 72567, 122081, 120312, 123232, 73643, 49081, 120961, 76672, 120749, 82227, 76227 e 111849), down-reguladas no *Δxyr1*/QM9414 em presença de celulose e soforose, exceto GH3 (*cel3e*, ID76227) que foi up-regulada na condição soforose. Os resultados mostraram uma excelente correlação entre qPCR e os dados do transcriptoma do RNA-Seq (com coeficiente de correlação *Pearson* de 0,8912), o que sugere uma boa qualidade dos dados de RNA-seq (Figura 65 e CD Anexo, Arquivo 16, Tabela 16).

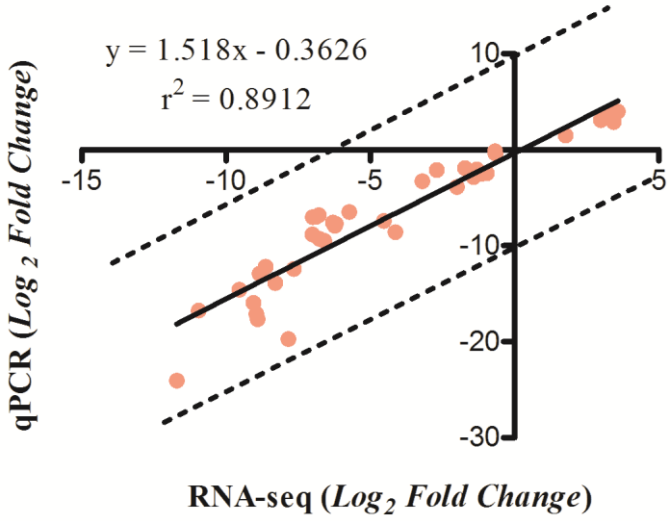


FIGURA 65 – Correlação entre RNA-seq e PCR quantitiva em Tempo Real (qPCR) para a linhagem mutante (*Δxyr1*) versus a linhagem parental (QM9414).
A linha sólida, r^2 e a equação associada foram gerados por análise de regressão linear utilizando *GraphPad Prism Version 5.0.*

7.1.2 DISCUSSÕES

Este estudo apresenta uma análise da expressão global de genes que codificam enzimas extracelulares envolvidas na degradação da celulose e xilana e o papel central do regulador XYR1 em *T. reesei*. Sabe-se que a transcrição, tradução e secreção de um conjunto de enzimas extracelulares eficientes na degradação de polímeros que constituem a biomassa é um processo que requer energia considerável. Consequentemente, *T. reesei* desenvolveu mecanismos sofisticados que assegurem que estas(es) enzimas/genes são formadas(os) apenas sob condições em que esses polímeros estão disponíveis como fonte de carbono e energia. Claramente, isto requer a presença de sistemas eficientes para detectar o tipo (ou composição) da biomassa vegetal, a fim de garantir que todas as enzimas necessárias para a degradação sinérgica da biomassa são de fato formadas (KUBICEK, 2012). Assim, o fungo filamentoso irá entrar em contato com uma mistura complexa de: celulose, hemicelulose, pectina, proteínas e outros componentes da parede celular das plantas e, em condições naturais, pode precisar requintadamente regular a produção de enzimas para uma ótima desconstrução da parede celular da planta (SUN et al., 2012). Nesse sentido, o fator de transcrição XYR1 tem sido relatado como um fator de ativação transcricional geral e essencial para o controle da expressão de genes celulolíticos e xilanolíticos, independente do agente indutor em *T. reesei* (FURUKAWA et al., 2009; CASTRO, et al., 2014a). O modo como esse ativador recebe os

sinais dos diferentes indutores e seu modo de ativação ainda permanence desconhecido (WENDLAND et al., 2013). Recentemente, Lichius e colaboradores (2014) demonstraram que a expressão de genes de celulases e xilanases requer a biossíntese de XYR1 e, simultaneamente, a sua importação para o núcleo, e, que imediatamente após o processo de indução para-se a transcrição dos genes de celulases e xilanases, de tal maneira que este processo foi acompanhado por uma degradação rápida do fator envolvido. Ainda demonstraram que nas hifas maduras, os núcleos são os locais de síntese desse principal regulador e transcrição dos genes.

Mecanismos de regulação de genes de celulases e hemicelulases têm sido estudados em fungos filamentosos, principalmente em *Aspergillus* (NOGUCHI et al., 2009; PEIJ et al., 1998a e b) e *T. reesei* (STRICKER, A. et al., 2008). O fungo filamentoso, *N. crassa*, também responde a uma variedade de moléculas indutoras que afetam a expressão de enzimas que degradam a parede celular de plantas (SUN et al., 2012). O ativador transcricional XlnR, identificado primeiramente em *A. niger* (PEIJ et al., 1998b), apresenta ortólogos nesses fungos filamentosos citados (BATTAGLIA et al., 2013; AMORE et al., 2013), porém os genes que são controlados por esse regulador são altamente espécies-específicos, resultando em diferentes enzimas produzidas por esses fungos (KLAUBAUF et al., 2014), isto é, em *Aspergillus* e *T. reesei*, XlnR/XYR1 regulam a expressão dos genes de celulases e hemicelulases, respectivamente (GIELKENS et al., 1999; STRICKER, A. et al., 2008; PEIJ et al., 1998a). Enquanto que, nas espécies *Fusarium* e *Neurospora*, os homólogos dos fatores XlnR/XYR1 são importantes para a indução de hemicelulases, não desempenhando um papel importante na expressão de genes de celulases (BRUNNER et al., 2007; SUN et al., 2012). Esses estudos sugerem que algumas espécies de fungos possuem fatores de transcrição que corregulam a expressão de genes de celulases e hemicelulases, enquanto que, em outras espécies, envolvem mecanismos regulatórios independentes (ZNAMEROSKI; GLASS, 2013). Outros fatores de transcrição, como CLR-1 e CLR-2 (identificados em *N. crassa*) são essenciais para o crescimento em presença de celulose, em que *clr-1* pode promover a expressão de transportadores de celodextrinas e genes de β-glicosidases, e também do segundo regulador *clr-2* que é responsável por ativar a produção de celulases (CORADETTI et al., 2012; XIONG et al., 2014); em *A. nidulans* e *A. oryzae*, um homólogo *clr-2*, chamado *clrB*/*manB*, respectivamente, são também essenciais para a expressão de celulases (CORADETTI et al., 2012; CORADETTI et al., 2013; OGAWA et al., 2013). Em *T. reesei*, o homólogo *clr-2* (ID26163 – recentemente descrito como uma lactose permease) é essencial

para a indução de *cbh1* e *cbh2*, enquanto que o homólogo *clr-1* (ID27600) não foi caracterizado neste fungo, sugerindo uma diferença importante nos mecanismos de ativação de *clr-2*/ID26163 em *N. crassa* e em *T. reesei* (HÄKKINEN et al., 2014). Desse modo, existem fatores essenciais para a produção de enzimas, destacando em *T. reesei* o fator de transcrição XYR1, sendo importante para expressão de genes celulolíticos e xilanolíticos, enquanto que, em *N. crassa* o fator *clr-2* é essencial apenas para a produção de celulases em presença de celulose (STRICKER et al., 2006 ; HÄKKINEN et al., 2014).

Em *T. reesei*, os resultados obtidos pelo RNA-seq no mutante *Δxyr1* permitiram verificar que o fator de transcrição XYR1 desempenha um papel crítico na ativação da transcrição de genes que codificam celulases e hemicelulases. Esse fungo pode utilizar uma variedade de fontes de carbono (CASTRO, et al., 2014b; DASHTBAN et al., 2011; CHEN et al., 2014; JUN et al., 2013; BISCHOF et al., 2013) e isto requer uma ótima adaptação ambiental para sintetizar transportdores e secretar enzimas que são induzidos por substratos ou metabólitos específicos. Para *T. reesei* se desenvolver na presença de vários constituintes da parede celular das plantas, sugere-se que exista um baixo nível constitutivo extracelular ou enzimas da biomassa vegetal que ajudem a liberar outros metabólitos (IVANOVA et al., 2013). Esses metabólitos são transportados para dentro das células e servem como indutores diretamente ou são convertidos em moléculas de sinalização necessárias para desencadear a transcrição de genes que codificam enzimas de degradação da biomassa.

Na presença das três fontes de carbono (celulose, soforose e glicose) utilizadas para o crescimento da linhagem *Δxyr1*, destaca-se as diferenças de suas complexidades, ou seja, da fonte mais complexa a menos complexa, para serem desconstruídas pelas enzimas produzidas por *T. reesei* (*Δxyr1*) e seus produtos serem metabolizados. Diferentes genes foram regulados negativamente e positivamente nas três fontes de carbono, considerando a ausência do regulador XYR1, sugerindo que a transcrição de genes, via este regulador, ocorre de maneira carbono-dependente. Em *S. cerevisiae* e outras leveduras, a expressão de genes também ocorre de maneira dependente da fonte de carbono (WEINHANDL et al., 2014). Na ausência de XYR1, cerca de 469 genes foram down-regulados em celulose, 308 genes foram down-regulados em sophorose e 14 genes em glicose. Outros 783 genes foram up-regulados na ausência do principal fator de transcrição, sendo 354, 397 e 32 genes identificados nas condições celulose, soforose e glicose, respectivamente. Em resumo, XYR1 funciona como um fator de transcrição global em *T. reesei* e afeta a expressão de genes direta e/ou

indiretamente. Esse tipo de regulação foi identificado em *N. crassa* que apresentou alvos de regulação direta e indireta (SUN et al., 2012).

Curiosamente, o remodelamento da cromatina foi um mecanismo indireto identificado por Mello-de-Sousa e colaboradores (2015), envolvendo a participação do fator de transcrição XYR1 na regulação da expressão gênica. De tal maneira que, para a regulação da expressão gênica, o remodelamento constante e ordenado da arquitetura da cromatina permite que diversos complexos proteicos com ou sem atividades enzimáticas possam acessar o DNA. O complexo enzimático responsável por promover a epigenética das histonas[14] é comumente recrutado ao DNA por fatores de transcrição. Os fatores de transcrição reconhecem sequências específicas no DNA, sinalizando à remodeladores da cromatina, correguladores e enzimas que permitem ocorrer o remodelamento da cromatina e consequente ativação ou repressão de determinado gene. Portanto, na pesquisa realizada pelos autores citados revelou que tanto as condições de cultivo (soforose e glicose) e a presença do regulador XYR1 influenciam no estado da cromatina. Em 2014, Mello-de-Sousa e outros pesquisadores destacaram também que o fator de transcrição CRE1 contribui indiretamente na abertura da cromatina, regulando o remodelamento da cromatina. Dessa forma, a identificação desse mecanismo indireto (estudo da modulação da estrutura da cromatina) juntamente com a identificação de fatores de transcrição (XYR1 e CRE1) que acarretam na sua modificação e as consequências sobre o contexto celular proteico são de suma importância para a compreensão de mecanismos de produção de enzimas degradadoras da parede celular vegetal.

Foi possível verificar com os dados obtidos por meio da abordagem de RNA-seq, a modulação diferencial de outros fatores de transcrição (Tabela 26), transportadores (CD Anexo, Arquivo 15) e de genes *CAZy* (CD Anexo, Arquivo 14, Tabelas 14.1 a 14.3) que são afetados pela ausência do principal regulador da expressão de enzimas de interesse biotecnológico nas três fontes de carbono em *T. reesei*. Esses resultados destacam que XYR1 é necessário para a expressão de muitos genes, mas sugere que este fator pode trabalhar de forma independente ou sinergicamente com outros fatores de transcrição conhecidos e desconhecidos, corroborando o que foi destacado por Bischof et al. (2013), quando utilizou palha de trigo e lactose como fontes de carbono. No CD Anexo (Arquivos 5, 6, 7, 8 e 14; Tabelas 5, 6, 7, 8 e 14.1 a 14.3) é possível verificar que o fator transcricional Zn2Cys6

[14] A epigenética das histonas consiste de modificações pós-traducionais das porções aminoterminais das caudas destas proteínas altamente básicas. Como fator determinante da regulação transcricional, estas mudanças de cargas das histonas podem afetar a ligação de proteínas remodeladoras à cromatina e/ou interferir no contato entre os nucleossomos, unidades repetitivas de DNA e um octâmero de histonas (LEITE, 2013).

(XYR1) pode ativar e modular a expressão de enzimas de degradação da parede celular vegetal e também de outros genes envolvidos em diversos processos bioquímicos/moleculares, como: *Carbohydrate transport and metabolism*; *Secondary metabolites biosynthesis, transport and catabolism*; *Energy production and conversion*; *Amino acid transport and metabolism*; *Posttranslational modification, protein turnover, chaperones*; *Signal transduction mechanisms and Transcription* (em celulose); *Translation, ribosomal structure and biogenesis*; *Intracellular trafficking, secretion, and vesicular transport and Amino acid transport and metabolism* (em soforose).

Vários genes *CAZy* necessários para a degradação de diferentes polímeros da parede celular vegetal foram fortemente reprimidos na ausência de XYR1 (Figura 61; CD Anexo, Arquivo 14, Tabelas 14.1 a 14.3) nas duas fontes de carbono indutoras, com destaque para GH12 (*egl3*/*cel12a*), GH61 (*cel61b*), GH5 (*man1* e *egl2*/*cel5a*), GH6 (*cbh2*/*cel6a*), GH7 (*cbh1*/*cel7a*) entre outras. Curiosamente, em soforose (conhecida como um potente indutor desde 1962 (MANDELS et al., 1962)) as principais exoglucanases produzidas por *T. reesei* (MIETTINEN-OINONEN, 2004), *cbh1* e *cbh2* (*Log_2 Fold Change*: -8.95 e -8.91, respectivamente, Tabela 23) foram as mais afetadas pela falta do regulador XYR1. Desse modo, a indução desses dois genes, em soforose ou nas duas outras condições analisadas, pode ser um resultado direto do aumento do nível de transcrição de *xyr1* na linhagem parental QM9414. Resultados semelhantes foram ressaltados por Derntl et al. (2013) utilizaram várias linhagens de *T. reesei*, inclusive QM6a, sendo que a formação de *cbh1*/*chb2* seguiu rigorosamente o nível de expressão de *xyr1*. Portanto, sugere-se que o nível de expressão de genes que codificam celulases ou até mesmo de outros genes (down-regulados) identificados no mutante *Δxyr1* (que são alvos diretos de XYR1) podem ser dependentes do nível de expressão de XYR1. Ao contrário, na linhagem parental QM9414 (Capítulo II), genes *CAZy* de interesse biotecnológico foram up-regulados em presença das duas condições indutoras (CASTRO et al., 2014b), porém destaca-se desta análise (Capítulo II) que 26 genes *CAZy* (sendo 16 up-regulados e 10 down-regulados) só estão presentes (frequentemente em alta abundância) nesta linhagem, enquanto que estas não foram detectadas na linhagem *Δxyr1*. Ao mesmo tempo, foram identificadas outros genes *CAZy* modulados exclusivamente na linhagem mutante (CD Anexo, Arquivo 17, Tabela 17.1). Por exemplo, GHs18 (*chitinase CHI18-2* e *chitinase CHI18-6*) somente estavam moduladas na linhagem parental, enquanto GHs18 (*chitinase CHI18-12* e *chitinase CHI18-15*) somente na linhagem mutante, o que sugere uma estratégia de compensação em *T. reesei*. Essa estratégia também foi identificada

em *Fusarium graminearum* (KLAUBAUF et al., 2014). Curiosamente, outras 9 enzimas foram diferencialmente expressas no mutante que não foram identificadas no transcriptoma da linhagem parental (CD Anexo, Arquivo 17, Tabela 17.1). O perfil de compensação na expressão de genes no mutante também foi verificado para transportadores e fatores de transcrição (CD Anexo, Arquivo 17, Tabelas 17.2 e 17.3).

Transportadores são fundamentais na via metabólica de açúcares. Como base nos dados obtidos do transcriptoma, 63 transportadores foram identificados nas três fontes de carbono estudadas, sendo, uma classe de transportadores identificadas em ambas condições indutoras. (CD Anexo, Arquivo 15, Tabela 15). Os transportadores *MSF permeases* fazem parte de uma grande família de proteínas envolvidas na importação de açúcares (por exemplo, glicose ou xilose) para o interior da célula; e são classificados como MSF transportadores. Participam de uma variedade de processos fisiológicos e são vitais para o metabolismo e homeostase energética em bactérias, fungos, protozoários, plantas e animais, podendo mediar a absorção celular de glicose e outros mono- e dissacarídeos (BÜTTNER, 2007; HENDERSON; BALDWIN, 2012; HENDERSON; MAIDEN, 1990; LI et al., 2011; ÖZCAN; JOHNSTON, 1999; WILSON-O'BRIEN et al., 2010). Com a análise filogenética (Figura 62), 7 grupos de transportadores foram identificados, sendo que o grupo *Sugars transporters* apresentou similaridade com outras espécies que estão envolvidas com transporte de glicose ou hexose nas duas condições indutoras estudadas, sugerindo que após a degradação da complexa celulose, produtos metabólitos intermediários foram produzidos e que o metabólito final (glicose ou outra hexose) foi transportado para o interior das células fúngicas para ser convertido em energia. Recentemente, uma pesquisa realizada com duas MFS transportadores, *Stp1* (ID47710) e *Crt1* (ID3405) foram identificadas como envolvidas no sensoriamento de celulose e na indução de genes de celulases em *T. resei*. (ZHANG et al., 2013). *Stp1* reprime a indução de celulases e hemicelulases em presença de celulose, enquanto que *Crt1* é requerida para indução de celulases por celulose, lactose e, possivelmente, soforose. Porém, *Crt1* não foi requerida para expressão de hemicelulases em presença de xilana (ZHANG et al., 2013; IVANOVA et al., 2013). Curiosamente, na análise dos dados do RNA-seq *Δxyr1*/QM9414 demonstrou que o transportador *Stp1* foi alvo direto do regulador XYR1, sendo down-regulado durante o crescimento em presença de celulose (CD Anexo, Arquivo 15, Tabela 15), e como já foi caracterizado por ZHANG e colaboradores (2013), a interrupção do gene *Stp1* resultou no aumento de indução de celulases em presença de celulose. Outros 24 transportadores foram alvos de regulação direta por XYR1 em celulose e

10 em soforose reforçando a ideia de que XYR1 regula outros processos celulares além da regulação de enzimas envolvidas na degradação de biomassa (CD Anexo, Arquivo 15, Tabela 15). Para uma melhor caracterização dos transportadores identificados, novos experimentos serão necessários para elucidar os complexos mecanismos envolvidos na indução de enzimas de interesse biotecnológico. Com este estudo, destaca-se a importância de transportadores para uma eficiente indução e utilização de açúcares no metabolismo em *T. reesei*.

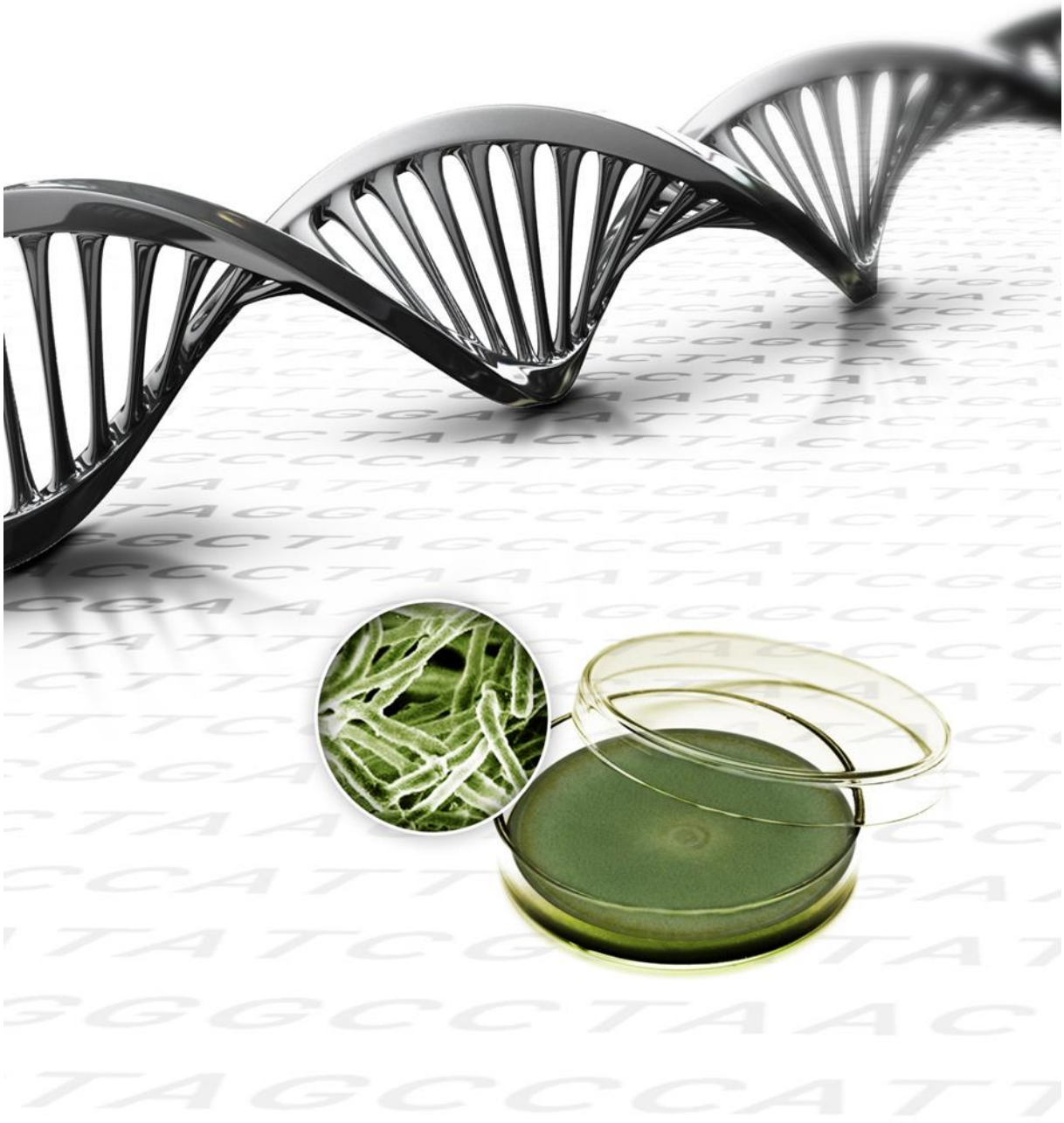
A ausência do fator de transcrição XYR1 não teve grande efeito na expressão gênica de *T. reesei* em glicose, evidenciado pelo baixo número de genes diferencialmente expressos nesta condição (Figura 55). Esses dados corroboram com informações obtidas em 2013, nas quais verificou-se que um nível muito baixo de transcrição de genes *xyn1*, *xyn2*, *cbh1*, *cbh2 e xyr1* foram formados durante o crescimento em glicose (DERNTL et al., 2013). Sabe-se que a presença de glicose, facilmente metabolizável e energeticamente favorável, resulta na repressão de vários genes, uma vez que o fungo utiliza a fonte glicose para obtenção de energia para a sua sobrevivência (CHAMBERGO et al., 2002). O que poderia explicar essa redução de expressão gênica na presença da condição repressora? Em outro estudo realizado em nosso Laboratório de Biotecnologia Molecular (FMRP-USP), envolvendo o fator de transcrição CRE1 e a técnica RNA-seq, foram identificados 552 genes diferencialmente expressos na presença de glicose no mutante Δ*cre1* quando comparado à linhagem parental QM9414, um número muito maior do que o que foi identificado nas análises do mutante *Δxyr1* (46 genes). Nesse estudo, foi possível destacar que o regulador transcricional Zn2Cys6 (XYR1) foi diretamente reprimido por CRE1 em presença de glicose (ANTONIÊTO et al., 2014). Também foi verificado por Portnoy e colaboradores (2011b) que *xyr1* foi positivamente regulado por indução e pela repressão catabólica de carbono (CCR) via CRE1, quando *T. reesei* foi cultivado em presença do indutor lactose. Consequentemente, a expressão de poucos genes no *T. reesei Δxyr1*, em presença de glicose, além de ser afetada por CCR/CRE1, foi afetada também pela ausência do principal regulador do sistema de celulases/hemicelulases em *T. reesei* (XYR1).

E o que poderia justificar o aumento de expressão de genes (up-regulados) em todas as condições analisadas na ausência do regulador XYR1? A resposta envolve todos os genes up-regulados (genes diferencialmente expressos com expressão ≥ 1 [genes *CAZy*, fatores, transportadores, regulons das condições]) que, na ausência do fator de transcrição XYR1, outro(s) fator(es) pode(m) ativar a produção de genes. Que fator ou fatores são esses? Como foi observado em *A. niger*, dois fatores XlnR (HASPER et al., 2000) e AraR (BATTAGLIA et

al., 2011), envolvidos na via catabólica de pentoses (PCP) em presença de mono- e polissacarídeos, e a deleção de XlnR, pode resultar na up-regulação do fator AraR, sendo o fator responsável pela indução de vários genes (BATTAGLIA et al., 2011). Outro(s) fator(es) de transcrição pode(m) atuar de maneira semelhante ao estudo demonstrado por Battaglia e outros pesquisadores (2011). Nesse sentido, outro fator conhecido que tem função fundamental na ativação de genes de celulases e xilanases em *T. reesei* é o ACE2 (ARO et al., 2001), mas pelos dados apresentados (Tabela 25), sugere-se que nenhum fator transcricional caracterizado neste fungo filamentoso (ARO et al., 2001; ARO et al., 2003; NITTA et al., 2012; ZEILINGER et al., 2001; HÄKKINEN et al., 2014; PORTNOY et al., 2011a) parece estar envolvido na regulação dos genes apresentados, visto que a princípio todos sofrem regulação direta de XYR1. Portanto, o principal regulador da produção de enzimas (XYR1) em todas as condições analisadas parece não agir isoladamente. Todos os genes up-regulados apresentados na Tabela 24, são suspeitos de assumir a função deste regulador para a regulação positiva de genes, principalmente uma classe de fatores: Zn2Cys6 (identificada nas três condições analisadas). Essa família é encontrada exclusivamente em fungos e está envolvida em várias funções reguladoras (ZHAO et al., 2011; COSTE et al., 2008; PARDO; OREJAS, 2014; CHANG; EHRLICH, 2013).

Logo, esses fatores de transcrição, juntamente com outros genes identificados (inclusive os genes não caracterizados) em cada fonte de carbono pela técnica RNA-seq, também serão alvos de futuros estudos para elucidar os reais mecanismos envolvidos na produção de enzimas que degradam a biomassa via fator de transcrição XYR1. Com todos os resultados obtidos, pode-se afirmar que esse regulador tem um impacto significativo sobre a capacidade de *T. reesei* expressar genes associados à degradação e de se desenvolver em presença de polissacarídeos da parede celular vegetal.

CONCLUSÕES



8 CONCLUSÕES

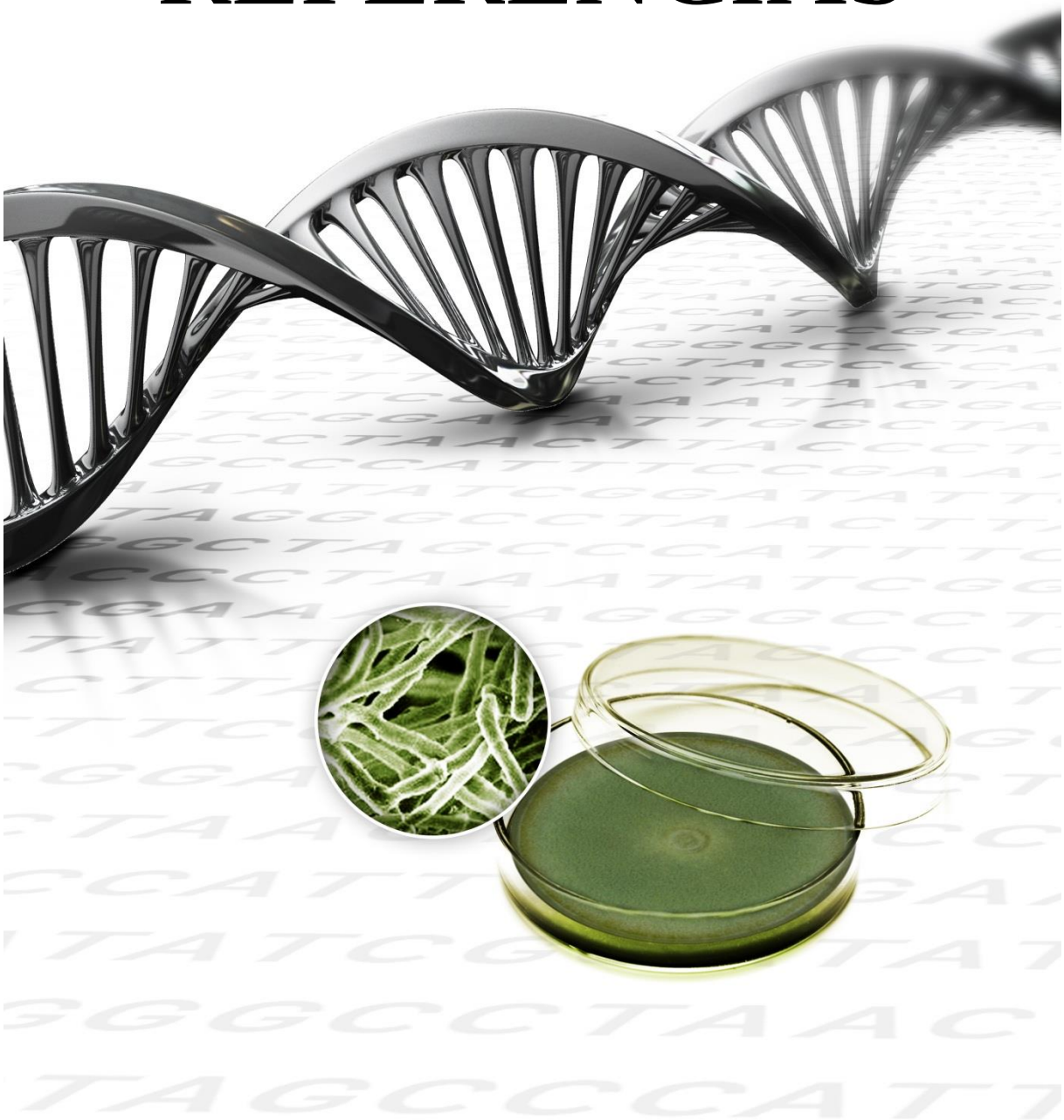
Nesta tese, foi analisada a regulação da produção de enzimas celulolíticas e xilanolíticas em *Trichoderma reesei*, destacando-se a interação de diferentes fontes de carbono e o papel do fator de transcrição XYR1 no controle deste processo. Também destacou-se, pela primeira vez, a identificação quantitativa do sítio de ligação do regulador XYR1, que coordena a expressão dos genes que codificam celulases em *T. reesei*. As análises permitiram definir o PWM para XYR1 que são específicos de *T. reesei*, ofuscando o viés das sequências consenso determinados em outros organismos. Observou-se que os sítios individuais caracterizam o regulador XYR1, porém motivos demonstrados com repetições invertidas e evertidas e com definição dos espaçamentos dos sítios foram mais associados com promotores alvos deste regulador. Diante dos resultados apresentados e do modelo geral para a indução de celulases disponíveis, foi proposto um modelo mecanicista que poderia explicar a presença de um ou dois sítios de ligação XYR1 nos promotores de celulases. O mecanismo de indução sugere que, sob uma condição de repressão, ou seja, na presença de glicose, a produção de celulases é completamente bloqueada, enquanto que em condições de limitação de nutrientes, níveis basais de enzimas (principalmente *cel7a* e *cel6a*) são produzidas. Subsequente, quando o fungo encontra celulose, as enzimas atuam no substrato insolúvel para gerar indutores solúveis como a soforose que, por sua vez, provocaria o sinal de elevado nível de produção de celulase. Em nosso modelo, foi observado uma alta confiança de sítios duplos de ligação XYR1 nos promotores *cel7a* e *cel6a*, em que condições de limitação de nutrientes poderiam aumentar os níveis de XYR1 através da liberação de CCR, mediado por CRE1. Aumentando os níveis de XYR1 permitiria a formação de homodímeros que ativam preferencialmente promotores dotados de elementos de regulação cis dispostos como repetições, como para *cel7a* e *cel6a*. Isso explicaria um aumento na expressão basal dessas proteínas. Assim, essas enzimas convertem celulose em indutores como soforose, e outros fatores adicionais agiriam em sinergia com XYR1, talvez pela formação de heterodímeros, e ativam a formação de celulases por sítios individuais repetidos. Evidentemente, novas abordagens experimentais serão necessárias para obter novos *insights* sobre os mecanismos de integração presentes nos promotores de celulase. Acredita-se que o trabalho apresentado neste estudo contribui significativamente para esta tarefa.

Análises do perfil transcricional de *T. reesei* foram realizadas para a identificação de genes diferencialmente expressos em resposta às condições de crescimentos em que o fungo

filamentoso foi exposto. As principais diferenças de expressão em *T. reesei* em celulose e soforose está associada com *CAZy*mes, proteínas acessórias, transportadores, fatores transcricionais e transporte de elétrons. Os resultados demonstrados neste estudo sugerem que ambos celobiose e soforose atuam como co-indutores na produção de celulases em *T. reesei*. Além disso, futuras investigações genômicas em novos genes identificados no crescimento em celulose abrirá novas linhas de pesquisa na clarificação da regulação de celulases e hemicelulases em *T. reesei*. O fator transcricional XYR1 foi crítico para a expressão de genes em *T. reesei*, de tal modo que a sua ausência afetou a transcrição de enzimas (Genes *CAZy*), fatores transcricionais, transportadores, além de outros genes constituintes do genoma desse fungo filamentoso. Com as análises realizadas, novos genes envolvidos na degradação de celulose foram identificados e *T. reesei* é passível de se adaptar às condições de crescimento aqui estudadas, mesmo na ausência desse principal regulador. Porém, ainda será necessária a realização da caracterização de fatores transcricionais e transportadores associados à degradação da celulose e de componentes adicionais da parede celular das plantas. Todos os resultados obtidos contribuem para entender o complexo mecanismo responsável para a expressão de genes que degradam polissacarídeos constituídos na parede celular vegetal.

O método RNA-seq e a combinação com as ferramentas de bioinformática adequadas oferecem uma nova abordagem para estudar a dinâmica de expressão gênica, em uma escala global, durante o processo de desenvolvimento fúngico em condições indutoras e repressora. Assim, os resultados apresentados fornecem uma base para futuras estratégias na construção de mutantes que produzirão altos níveis de enzimas para serem utilizados na degradação da biomassa e facilitar a aplicação na produção de etanol 2G. Consequentemente, o conhecimento de como os diferentes genes de celulases e xilanases são regulados podem contribuir para o desenvolvimento de uma variedade de coquetéis enzimáticos para a hidrólise da biomassa. Todavia, muitos estudos ainda serão necessários para elucidar toda a rede envolvida na regulação da formação dessas enzimas.

REFERÊNCIAS



REFERÊNCIAS

ABOYOUN, P.; PAGES, H.; LAWRENCE, M. Package " GenomicRanges." **Bioconductor**, v. 1.16.4, p. 1–86, 2014. Disponível em: <http://www.bioconductor.org/packages/release/bioc/manuals/GenomicRanges/man/GenomicRanges.pdf>. Acesso em: 03/09/2014.

AGUIAR, C.; PEREIRA, D.; RODRIGUES, F.; et al. Produção de etanol através do bagaço de cana-de-açúcar. **Instituto Politécnico - Centro Universitário UNA**, p. 1–10, 2013.

AKEL, E.; METZ, B.; SEIBOTH, B.; KUBICEK, C. P. Molecular regulation of arabinan and L-arabinose metabolism in *Hypocrea jecorina* (*Trichoderma reesei*). **Eukaryotic cell**, v. 8, p. 1837–1844, 2009.

ALEIXO, T. R. T. **Aproveitamento da biomassa de cana para produção de bioetanol**, 2012, 63f. Trabalho de Conclusão de Curso - TCC (Curso de Tecnologia em Biocombustíveis) - Centro Estadual de Educação Tecnológica Paula Souza, Faculdade de Tecnologia de Araçatuba, São Paulo.

AMATO, F.; MATOSO, F. **Mistura de etanol na gasolina sobe para 27% a partir de 16 de março**. 2015. Disponível em: <http://g1.globo.com/economia/noticia/2015/03/mistura-de-etanol-na-gasolina-sobe-para-27-partir-de-16-de-marco.html>. Acesso em: 04/03/2015.

ALMEIDA, C. F. DE. **Study of the Role of SETD2 Mutations in clear cell Renal Cell Carcinoma (ccRCC)**, 2013, 96f. Dissertação (Mestrado em Bioestatística) - Faculdade de Ciências, Universidade de Lisboa, Portugal.

ALMEIDA, M. C. DE O. **Indução de celulases e xilanases por *Trichoderma reesei* e *Penicillium* variabile em cultivo em estado sólido a partir de substratos lignocelulósicos**, 2012, Dissertação (Mestrado em Engenharia Química) - Universidade Federal de Santa Catarina, Centro Tecnológico, Desenvolvimento de Processos Químicos e Biotecnológicos, Florianópolis - SC.

ALVES, J. M. B; MACRI, R. DE C. V. Etanol de segunda geração: estudo de materiais lignocelulósicos e aplicações da lignina. **Ciência & Tecnologia**, v. 5, p. 1–12, 2013.

AMORE, A.; GIACOBBE, S.; FARACO, V. Regulation of cellulase and hemicellulase gene expression in fungi. **Current genomics**, v. 14, n. 4, p. 230–49, 2013.

ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome biology**, v. 11, n. 10, p. R106, 2010.

ANDREWS, S. FastQC: A quality control tool for high throughput sequence data. **Babraham bioinformatics**, p. 1, 2010. Disponível em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Acesso em: 21/08/2014.

ANP. Biocombustíveis. **ANP - Agência Nacional de Petróleo, Gás Natural e Biocombustíveis**, p. 2, 2012. Disponível em: <http://www.anp.gov.br/>. Acesso em: 19/05/2014.

ANP. Usinas autorizadas pela ANP para produzir etanol. **novaCana**, p. 1, 2014. Disponível em: <http://www.novacana.com/dados/usinas/autorizadas-pela-anp>. Acesso em: 14/07/2014.

ANTONIÊTO, A. C. C.; SANTOS CASTRO, L. DOS; SILVA-ROCHA, R.; PERSINOTI, G. F.; SILVA, R. N. Defining the genome-wide role of CRE1 during carbon catabolite repression in *Trichoderma reesei* using RNA-Seq analysis. **Fungal Genetics and Biology**, v. 73, p. 93–103, 2014.

ARO, N.; ILMÉN, M.; SALOHEIMO, A.; PENTTILÄ, M. ACEI of *Trichoderma reesei* is a repressor of cellulase and xylanase expression. **Applied and environmental microbiology**, v. 69, p. 56–65, 2003.

ARO, N.; PAKULA, T.; PENTTILA, M. Transcriptional regulation of plant cell wall degradation by filamentous fungi. **FEMS microbiology reviews**, v. 29, p. 719–739, 2005.

ARO, N.; SALOHEIMO, A.; ILMÉN, M.; PENTTILÄ, M. ACEII, a novel transcriptional activator involved in regulation of cellulase and xylanase genes of *Trichoderma reesei*. **The Journal of biological chemistry**, v. 276, p. 24309–24314, 2001.

AZEVEDO, M. S. DE; SANTOS, R. V. DE O.; MAGALHÃES, T. V. Produção de Etanol no Brasil. **Revista de divulgação do Projeto Universidade Petrobras e IF Fluminense**, v. 2, p. 151–154, 2012.

BAILEY, T. L.; WILLIAMS, N.; MISLEH, C.; LI, W. W. MEME: Discovering and analyzing DNA and protein sequence motifs. **Nucleic Acids Research**, v. 34, p. 369–373, 2006.

BASTOS, V. D. Etanol, alcoolquímica e biorrefinarias. **BNDES Setorial**, v. 25, p. 5–38, 2007, 34f. Disponível em: <http://www.ambiente.sp.gov.br/wp-content/uploads/publicacoes/etanol/alcoolquimica.pdf>. Acesso em: 11/07/2014.

BATTAGLIA, E.; KLAUBAUF, S.; VALLET, J.; et al. Xlr1 is involved in the transcriptional control of the pentose catabolic pathway, but not hemi-cellulolytic enzymes in *Magnaporthe oryzae*. **Fungal genetics and biology : FG & B**, v. 57, p. 76–84, 2013.

BATTAGLIA, E.; VISSER, L.; NIJSSEN, A; et al. Analysis of regulation of pentose utilisation in *Aspergillus niger* reveals evolutionary adaptations in Eurotiales. **Studies in mycology**, v. 69, n. 1, p. 31–8, 2011.

BÉGUIN, P. Molecular biology of cellulose degradation. **Annu. Rev. Microbiol**, v. 44, n. 27, p. 219–248, 1990.

BEY, M.; ZHOU, S.; POIDEVIN, L.; et al. Cello-oligosaccharide oxidation reveals differences between two lytic polysaccharide monooxygenases (family GH61) from *Podospora anserina*. **Applied and environmental microbiology**, v. 79, n. 2, p. 488–96, 2013.

BISARIA, V. S.; MISHRA, S. Regulatory aspects of cellulase biosynthesis and secretion. **Critical reviews in biotechnology**, v. 9, p. 61–103, 1989.

BISCHOF, R.; FOURTIS, L.; LIMBECK, A.; et al. Comparative analysis of the *Trichoderma reesei* transcriptome during growth on the cellulase inducing substrates wheat straw and lactose. **Biotechnology for biofuels**, v. 6, n. 1, p. 127, 2013.

BLAKE, J. A.; HARRIS, M. A. The Gene Ontology (GO) Project: Structured vocabularies for molecular biology and their application to genome and expression analysis. **Current Protocols in Bioinformatics**, v. Chapter 7, p. Unit 7.2., 2008.

BNDES E CGEE. **Bioetanol de cana-de-açúcar:** energia para o desenvolvimento sustentável. 2008, 317f. Disponível em: <www.cgee.org.br/atividades/redirect.php?idProduto=5126>. Acesso em: 19/05/2014.

BOYCE, K. J.; ANDRIANOPOULOS, A. Morphogenetic circuitry regulating growth and development in the dimorphic pathogen *Penicillium marneffei*. **Eukaryotic Cell**, 2013.

BROAD-INSTITUTE. Sequencing Module 3: Overview. **Broad Institute**, v. 3, p. 1–73, 2014. Disponível em: <https://www.broadinstitute.org/files/shared/illuminavids/sequencingSlides.pdf>. Acesso em: 21/08/2014.

BRUNNER, K.; LICHTENAUER, A. M.; KRATOCHWILL, K.; DELIC, M.; MACH, R. L. Xyr1 regulates xylanase but not cellulase formation in the head blight fungus *Fusarium graminearum*. **Current Genetics**, v. 52, p. 213–220, 2007.

BÜTTNER, M. The monosaccharide transporter(-like) gene family in *Arabidopsis*. **FEBS letters**, v. 581, n. 12, p. 2318–24, 2007.

CANILHA, L.; KUMAR CHANDEL, A.; SANTOS MILESSI, T. S. DOS; et al. Bioconversion of sugarcane biomass into ethanol: an overview about composition, pretreatment methods, detoxification of hydrolysates, enzymatic saccharification, and ethanol fermentation. **Journal of Biomedicine and bBotechnology**, v. 2012, p. 1–15, 2012.

CANTAREL, B. L.; COUTINHO, P. M.; RANCUREL, C.; et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. **Nucleic acids research**, v. 37, p. D233–D238, 2009.

CARLE-URIOSTE, J. C.; ESCOBAR-VERA, J.; EL-GOGARY, S.; et al. Cellulase induction in *Trichoderma reesei* by cellulose requires its own basal expression. **The Journal of biological chemistry**, v. 272, p. 10169–10174, 1997.

CARLSON, J. M.; CHAKRAVARTY, A.; DEZIEL, C. E.; GROSS, R. H. SCOPE: a web server for practical de novo motif discovery. **Nucleic acids research**, v. 35, n. Web Server issue, p. W259–64, 2007.

CARVALHO, M. C. DA C. G. DE; SILVA, D. C. G. DA. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, v. 40, p. 735–744, 2010.

CASTRO, L. D. S.; ANTONIÊTO, A. C. C.; PEDERSOLI, W. R.; et al. Expression pattern of cellulolytic and xylanolytic genes regulated by transcriptional factors XYR1 and CRE1 are affected by carbon source in *Trichoderma reesei*. **Gene expression patterns: GEP**, v. 14, n. 2, p. 88–95, 2014a.

CASTRO, L. D. S.; PEDERSOLI, W. R.; ANTONIÊTO, A. C. C.; et al. Comparative metabolism of cellulose , sophorose and glucose in *Trichoderma reesei* using high-throughput genomic and proteomic analyses. **Biotechnology for biofuels**, v. 7, n. 1, p. 41, 2014b.

CHAMBERGO, F. S.; BONACCORSI, E. D.; FERREIRA, A. J. S.; et al. Elucidation of the metabolic fate of glucose in the filamentous fungus *Trichoderma reesei* using expressed sequence tag (EST) analysis and cDNA microarrays. **The Journal of biological chemistry**, v. 277, p. 13983–13988, 2002.

CHAMI KHAZRAJI, A.; ROBERT, S. Interaction Effects between Cellulose and Water in Nanocrystalline and Amorphous Regions: A Novel Approach Using Molecular Modeling. **Journal of Nanomaterials**, v. 2013, p. 1–10, 2013.

CHANG, P.-K.; EHRLICH, K. C. Genome-wide analysis of the $Zn(II)_2Cys_6$ zinc cluster-encoding gene family in *Aspergillus flavus*. **Applied microbiology and biotechnology**, v. 97, n. 10, p. 4289–300, 2013.

CHANG, V. S. .; HOLTZAPPLE, M. T. Fundamental Factors Affecting Biomass Enzymatic Reactivity. **Applied Biochemistry and bBotechnology**, v. 84-86, p. 5–37, 2000.

CHEN, X.; LUO, Y.; YU, H.; et al. Transcriptional profiling of biomass degradation-related genes during *Trichoderma reesei* growth on different carbon sources. **Journal of biotechnology**, v. 173, p. 59–64, 2014.

CNI. Indústria automobilística e sustentabilidade. **Brasília: CNI - Confederação Nacional da Indústria**, 2012, 43f. Disponível em: <http://arquivos.portaldaindustria.com.br/app/conteudo_18/2013/09/23/4970/20131002175420378115i.pdf >. Acesso em: 11/06/2014.

CORADETTI, S. T.; CRAIG, J. P.; XIONG, Y.; et al. Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 19, p. 7397–402, 2012.

CORADETTI, S. T.; XIONG, Y.; GLASS, N. L. Analysis of a conserved cellulase transcriptional regulator reveals inducer-independent production of cellulolytic enzymes in *Neurospora crassa*. **MicrobiologyOpen**, v. 2, n. Klich 2002, p. 595–609, 2013.

COSTE, A. T.; RAMSDALE, M.; ISCHER, F.; SANGLARD, D. Divergent functions of three *Candida albicans* zinc-cluster transcription factors (CTA4, ASG1 and CTF1)

complementing pleiotropic drug resistance in *Saccharomyces cerevisiae*. **Microbiology (Reading, England)**, v. 154, n. Pt 5, p. 1491–501, 2008.

DASHTBAN, M.; BUCHKOWSKI, R.; QIN, W. Effect of different carbon sources on cellulase production by *Hypocrea jecorina* (*Trichoderma reesei*) strains. **International journal of biochemistry and molecular biology**, v. 2, p. 274–86, 2011.

DERNTL, C.; GUDYNAITE-SAVITCH, L.; CALIXTE, S.; et al. Mutation of the Xylanase regulator 1 causes a glucose blind hydrolase expressing phenotype in industrially used *Trichoderma* strains. **Biotechnology for biofuels**, v. 6, n. 1, p. 62, 2013.

DIAS, L. P. **Bioacumulação de elementos-traço em *Agaricus bisporus* e *Pleurotus spp*.**, 2013, 53f. Dissertação (Mestrado em Microbiologia Agrícola) - Universidade Federal de Lavras, MG.

DIAS, M. O. S.; CAVALETT, O.; MACIEL-FILHO, R.; BONOMI, A. Integrated First and Second Generation Ethanol Production from Sugarcane. **Chemical Engineering Transactions**, v. 37, p. 445–450, 2014.

DIMAROGONA, M.; TOPAKAS, E.; CHRISTAKOPOULOS, P. Cellulose degradation by oxidative enzymes. **Computational and Structural Biotechnology Journal**, v. 2, n. 3, p. 1–8, 2012.

DODD, D.; CANN, I. K. O. Enzymatic deconstruction of xylan for biofuel production. **Glob Change Biol Bioenergy**, v. 1, n. 1, p. 2–17, 2010.

DRABER, K. **Etanol de segunda geração já é realidade**, 2013, 54f. Monografia (Graduação do Curso Engenharia Bioquímica) - Escola de Engenharia de Lorena da Universidade de São Paulo, Lorena.

DU, J.; LI, S.; ZHAO, H. Discovery and characterization of novel d-xylose-specific transporters from *Neurospora crassa* and *Pichia stipitis*. **Molecular bioSystems**, v. 6, p. 2150–2156, 2010.

DUPONT. Accellerase® 1500. **DuPont Genencor Science**, p. 1–4, 2013. Disponível em: <http://accellerase.dupont.com/fileadmin/user_upload/live/accellerase/documents/DUP-00413_ProdSheet_1500_web.pdf>. Acesso em: 15/07/2014.

EL-GOGARY, S.; LEITE, A; CRIVELLARO, O.; EVELEIGH, D. E.; EL-DORRY, H. Mechanism by which cellulose triggers cellobiohydrolase I gene expression in *Trichoderma reesei*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 86, n. 16, p. 6138–41, 1989.

ELLIOTT, D. C. Biomass, Chemicals from. **Encyclopedia of Energy**, v. 1, p. 163–174, 2004. Elsevier, Inc., Oxford, United Kingdom.

ENSINAS, A. V; CODINA, V.; MARECHAL, F.; ALBARELLI, J.; APARECIDA, M. Thermo-Economic Optimization of Integrated First and Second Generation Sugarcane Ethanol Plant. **Chemical Engineering Transactions**, v. 35, p. 523–528, 2013.

ERIKSEN, D. T.; HSIEH, P. C.; LYNN, P.; ZHAO, H. Directed evolution of a cellobiose utilization pathway in *Saccharomyces cerevisiae* by simultaneously engineering multiple proteins. **Microbial cell factories**, v. 12, n. 1, p. 61, 2013.

EYLEY, S.; THIELEMANS, W. Surface modification of cellulose nanocrystals. **Royal Society of Chemistry**, p. 1–36, 2014.

FEKETE, E.; KARAFFA, L.; KARIMI AGHCHEH, R.; et al. The transcriptome of *lae1* mutants of *Trichoderma reesei* cultivated at constant growth rates reveals new targets of LAE1 function. **BMC genomics**, v. 15, p. 447, 2014.

FELENBOK, B.; FLIPPHI, M.; NIKOLAEV, I. Ethanol catabolism in *Aspergillus nidulans*: a model system for studying gene regulation. **Progress in nucleic acid research and molecular biology**, v. 69, p. 149–204, 2001.

FERREIRA, L. C. C. **Caracterização do potencial energético entre a produção de etanol celulósico e a cogeração a partir do bagaço de cana**, 2012, 68f. Projeto de Graduação (Engenharia Mecânica) - Faculdade de Tecnologia, Universidade de Brasília.

FISHMAN, C. An Introduction to Illumina's Next Gen Sequencing. **Illumina**, p. 1–80, 2010. Disponível em: <http://www.illumina.com/content/dam/illuminamarketing/documents/products/illumina_sequencing_introduction.pdf>. Acesso em: 18/08/2014.

FOREMAN, P. K.; BROWN, D.; DANKMEYER, L.; et al. Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. **The Journal of biological chemistry**, v. 278, p. 31988–31997, 2003.

FRUTUOSO, M. A; MARANA, S. R. A single amino acid residue determines the ratio of hydrolysis to transglycosylation catalyzed by β-glucosidases. **Protein and peptide letters**, v. 20, n. 1, p. 102–6, 2013.

FURUKAWA, T.; SHIDA, Y.; KITAGAMI, N.; et al. Identification of the cis-acting elements involved in regulation of xylanase III gene expression in *Trichoderma reesei* PC-3-7. **Fungal Genetics and Biology**, v. 45, p. 1094–1102, 2008.

FURUKAWA, T.; SHIDA, Y.; KITAGAMI, N.; et al. Identification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in *Trichoderma reesei*. **Fungal Genetics and Biology**, v. 46, p. 564–574, 2009.

GARBER, M.; GRABHERR, M. G.; GUTTMAN, M.; TRAPNELL, C. Computational methods for transcriptome annotation and quantification using RNA-seq. **Nature methods**, v. 8, n. 6, p. 469–77, 2011.

GASCH, A. P.; SPELLMAN, P. T.; KAO, C. M.; et al. Genomic expression programs in the response of yeast cells to environmental changes. **Molecular biology of the cell**, v. 11, p. 4241–4257, 2000.

GEYSENS, S.; PAKULA, T.; UUSITALO, J.; et al. Cloning and Characterization of the Glucosidase II Alpha Subunit Gene of *Trichoderma reesei*: a Frameshift Mutation Results in the Aberrant Glycosylation Profile of the Hypercellulolytic Strain Rut-C30. **Applied and Environmental Microbiology**, v. 71, n. 6, p. 2910–2924, 2005.

GIELKENS, M. M. C.; DEKKERS, E.; VISSER, J.; GRAAFF, L. H. DE. Two cellobiohydrolase-encoding genes from *Aspergillus niger* require D-xylose and the xylanolytic transcriptional activator XlnR for their expression. **Applied and Environmental Microbiology**, v. 65, p. 4340–4345, 1999.

GLASS, N. L.; SCHMOLL, M.; CATE, J. H. D.; CORADETTI, S. Plant cell wall deconstruction by ascomycete fungi. **Annual review of microbiology**, v. 67, p. 477–98, 2013a.

GLASS, N. L.; PALMA-GUERRERO, J.; JONKERS, W.; et al. Revealing fungal communication modules by genomics, population genomics, and genome wide association studies in *Neurospora crassa*. **Phytopathology**, v. 103, p. 186–187, 2013b.

GORSCHE, R.; JOVANOVIC, B.; GUDYNAITE-SAVITCH, L.; MACH, R. L.; MACH-AIGNER, A. R. A highly sensitive in vivo footprinting technique for condition-dependent identification of cis elements. **Nucleic acids research**, v. 42, n. 1, p. e1, 2014.

GRANBIO. Biotecpédia. **GranBio**, p. 1, 2014. Disponível em: <http://www.granbio.com.br/inovacao/biotecpedia/?verbete=348>. Acesso em: 14/07/2014.

GRITZALI, M.; BROWN, R. D. the cellulase systems of *Trichoderma*: relationships between purified extracellular enxymes from induced or cellulose-grown cells. **Advances in Chemistry**. v. 181, p.237–260, 1979.

GRUBISICH, J. C. Etanol de cana tem várias vantagens em relação ao de milho. **Portal Unica**, p. 1, 2011.
Disponível em: <http://www.unica.com.br/convidados/2878140892033148390 8/etanol-de-cana-tem-variasvantagens-em-relacao-ao-de-milho/>. Acesso em: 10/06/2014.

GUELZIM, N.; BOTTANI, S.; BOURGINE, P.; KÉPÈS, F. Topological and causal structure of the yeast transcriptional regulatory network. **Nature genetics**, v. 31, p. 60–63, 2002.

GUO, F.; FANG, Z.; XU, C. C.; SMITH, R. L. Solid acid mediated hydrolysis of biomass for producing biofuels. **Progress in Energy and Combustion Science**, v. 38, n. 5, p. 672–690, 2012.

HAKANPÄÄ, J.; SZILVAY, G. R.; KALJUNEN, H.; et al. Two crystal structures of *Trichoderma reesei* hydrophobin HFBI--the structure of a protein amphiphile with and

without detergent interaction. **Protein science: a publication of the Protein Society**, v. 15, n. Ebbole 1997, p. 2129–2140, 2006.

HÄKKINEN, M.; ARVAS, M.; OJA, M.; et al. Re-annotation of the CAZy genes of *Trichoderma reesei* and transcription in the presence of lignocellulosic substrates. **Microbial Cell Factories**, v. 11, p. 134, 2012.

HÄKKINEN, M.; VALKONEN, M. J.; WESTERHOLM-PARVINEN, A.; et al. Screening of candidate regulators for cellulase and hemicellulase production in *Trichoderma reesei* and identification of a factor essential for cellulase production. **Biotechnology for biofuels**, v. 7, n. 1, p. 14, 2014.

HALLIKAS, O.; PALIN, K.; SINJUSHINA, N.; et al. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. **Cell**, v. 124, p. 47–59, 2006.

HARMAN, G. E.; HOWELL, C. R.; VITERBO, A.; CHET, I.; LORITO, M. *Trichoderma* species--opportunistic, avirulent plant symbionts. **Nature reviews. Microbiology**, v. 2, n. 1, p. 43–56, 2004.

HASPER, A. A.; VISSER, J.; GRAAFF, L. H. DE. The *Aspergillus niger* transcriptional activator XlnR, which is involved in the degradation of the polysaccharides xylan and cellulose, also regulates D -xylose reductase gene expression. **Molecular Microbiology**, v. 36, p. 193–200, 2000.

HEAD, S. R.; KOMORI, H. K.; LAMERE, S. A; et al. Library construction for next-generation sequencing: overviews and challenges. **BioTechniques**, v. 56, p. 61–4, 66, 68, passim, 2014.

HENDERSON, P. J. F.; MAIDEN, M. C. J. Homologous Sugar Transport Proteins in *Escherichia coli* and Their Relatives in Both Prokaryotes and Eukaryotes. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 326, n. 1236, p. 391–410, 1990.

HENDERSON;, P. J. F.; BALDWIN, S. A. Structural biology: Bundles of insights into sugar transporters. **Nature**, v. 490, p. 348–350, 2012.

HEROLD, S.; BISCHOF, R.; METZ, B.; SEIBOTH, B.; KUBICEK, C. P. Xylanase gene transcription in *Trichoderma reesei* is triggered by different inducers representing different hemicellulosic pentose polymers. **Eukaryotic cell**, v. 12, n. 3, p. 390–8, 2013.

HERPOËL-GIMBERT, I.; MARGEOT, A.; DOLLA, A.; et al. Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains. **Biotechnology for biofuels**, v. 1, n. 1, p. 18, 2008.

HIRATA, M. H.; TAVARES, V.; HIDRATA, R. D. C. Da biologia molecular à medicina: métodos comumente utilizados em farmacogenética. **Medicina Ribeirão Preto**, v. 39, n. 4, p. 522–534, 2006.

HORN, S. J.; VAAJE-KOLSTAD, G.; WESTERENG, B.; EIJSINK, V. G. Novel enzymes for the degradation of cellulose. **Biotechnology for biofuels**, v. 5, n. 1, p. 45, 2012.

HRMOVÁ, M.; PETRÁKOVÁ, E.; BIELY, P. Induction of cellulose- and xylan-degrading enzyme systems in *Aspergillus terreus* by homo- and heterodisaccharides composed of glucose and xylose. **Journal of general microbiology**, v. 137, n. 3, p. 541–7, 1991.

ILLUMINA. **HiSeq ™ 2000 Sequencing System**, 2011a, 4f. Disponível em: <http://www.illumina.com/documents/products/datasheets/datasheet_hiseq2000.pdf>. Acesso em: 18/08/2014.

ILLUMINA. **TruSeq ™ RNA and DNA Sample Preparation Kits v2**, 2011b, 4f. Disponível em: <http://res.illumina.com/documents/products/datasheets/datasheet_truseq_sample_prep_kits.pdf>. Acesso em: 18/08/2014.

ILMÉN, M.; SALOHEIMO, A.; ONNELA, M. L.; PENTTILÄ, M. E. Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*. **Applied and environmental microbiology**, v. 63, p. 1298–1306, 1997.

ILMÉN, M.; THRANE, C.; PENTTILÄ, M. The glucose repressor gene cre1 of *Trichoderma*: isolation and expression of a full-length and a truncated mutant form. **Molecular & General Genetics**, v. 251, n. 4, p. 451–60, 1996.

IVANOVA, C.; BÅÅTH, J. A; SEIBOTH, B.; KUBICEK, C. P. Systems analysis of lactose metabolism in *Trichoderma reesei* identifies a lactose permease that is essential for cellulase induction. **PloS one**, v. 8, n. 5, p. 1–10, 2013.

JACKSON, G.; ROCHA, D. M.; LORENA, E. D. E. DE; et al. Validação de metodologia para a caracterização química de bagaço de cana-de-açúcar. **Química Nova**, v. 32, n. 6, p. 1500–1503, 2009.

JALAK, J.; KURAŠIN, M.; TEUGJAS, H.; VÄLJAMÄE, P. Endo-exo synergism in cellulose hydrolysis revisited. **The Journal of biological chemistry**, v. 287, n. 34, p. 28802–28815, 2012.

JOAQUIM, L. M.; EL-HANI, C. N. A genética em transformação: crise e revisão do conceito de gene. **Sci. stud. [online]**, v. 8, n. 1, p. 93–128, 2010.

JUHÁSZ, T.; SZENGYEL, Z.; RÉCZEY, K.; SIIKA-AHO, M.; VIIKARI, L. Characterization of cellulases and hemicellulases produced by *Trichoderma reesei* on various carbon sources. **Process Biochemistry**, v. 40, p. 3519–3525, 2005.

JUN, H.; GUANGYE, H.; DAIWEN, C. Insights into enzyme secretion by filamentous fungi: comparative proteome analysis of *Trichoderma reesei* grown on different carbon sources. **Journal of proteomics**, v. 89, p. 191–201, 2013.

KARAFFA, L.; FEKETE, E.; GAMAUF, C.; et al. D-Galactose induces cellulase gene expression in *Hypocrea jecorina* at low growth rates. **Microbiology**, v. 152, p. 1507–1514, 2006.

KARIMI-AGHCHEH, R.; BOK, J. W.; PHATALE, P. A; et al. Functional analyses of *Trichoderma reesei* LAE1 reveal conserved and contrasting roles of this regulator. **G3 (Bethesda, Md.)**, v. 3, n. 2, p. 369–78, 2013.

KIRCHER, M.; KELSO, J. High-throughput DNA sequencing--concepts and limitations. **BioEssays: news and reviews in molecular, cellular and developmental biology**, v. 32, p. 524–536, 2010.

KLAUBAUF, S.; NARANG, H. M.; POST, H.; et al. Similar is not the same: Differences in the function of the (hemi-)cellulolytic regulator XlnR (Xlr1/Xyr1) in filamentous fungi. **Fungal genetics and biology: FG & B**, v. 72, p. 73–81, 2014.

KONIS, K. **Statistics with R:** Computing and Graphics, 2014, 15f. Disponível em: <http://www.stats.ox.ac.uk/~konis/OUCS/ComputingGraphics.pdf>. Acesso em: 04/09/2014.

KOPPRAM, R.; TOMÁS-PEJÓ, E.; XIROS, C.; OLSSON, L. Lignocellulosic ethanol production at high-gravity: challenges and perspectives. **Trends in biotechnology**, v. 32, n. 1, p. 46–53, 2014.

KUBICEK, C. P. Fungi and Lignocellulosic Biomass. **Fungi and Lignocellulosic Biomass**, 2012, 304f.

KUBICEK, C. P. Systems biological approaches towards understanding cellulase production by *Trichoderma reesei*. **Journal of biotechnology**, v. 163, n. 2, p. 133–42, 2013.

KUBICEK, C. P.; HERRERA-ESTRELLA, A.; SEIDL-SEIBOTH, V.; et al. Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. **Genome biology**, v. 12, n. 4, p. R40, 2011.

KUBICEK, C. P.; MESSNER, R.; GRUBER, F.; MACH, R. L.; KUBICEK-PRANZ, E. M. The *Trichoderma* cellulase regulatory puzzle: From the interior life of a secretory fungus. **Enzyme and Microbial Technology**, 1993.

KUBICEK, C. P.; MIKUS, M.; SCHUSTER, A.; SCHMOLL, M.; SEIBOTH, B. Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. **Biotechnology for biofuels**, v. 2, p. 19, 2009.

KUBICEK, C. P.; MUHLBAUER, G.; KLOTZ, M.; JOHN, E.; KUBICEK-PRANZ, E. M. Properties of a Conidial-bound Cellulase Enzyme System from *Trichoderma reesei*. **Journal of General Microbiology**, v. 134, p. 1215–1222, 1988.

LANE, J. GranBio's Alagoas cellulosic ethanol project, in pictures and video. **Biofuels Digest**, p. 1–8, 2014.

Disponível em: <http://www.biofuelsdigest.com/bdigest/2014/01/14/granbios-alagoas-cellulosic-ethanol-project-in-pictures-and-video/>. Acesso em: 14/07/2014.

LANGMEAD, B. Aligning short sequencing reads with Bowtie. **Curr Protoc Bioinformatics**, 2010. Chapter 11:Unit 11.7.

LANGMEAD, B. Introduction to the Burrows-Wheeler Transform and FM Index. **JHU's**, p. 1–12, 2013, 12f. Disponível em: <http://www.cs.jhu.edu/~langmea/resources/bwt_fm.pdf>. Acesso em: 02/092014.

LAWRENCE, M.; CAREY, V.; GENTLEMAN, R. Package "rtracklayer." **Bioconductor**, v. 1.24.2, p. 1–75, 2014. Disponível em: <http://www.bioconductor.org/packages/release/bioc/manuals/rtracklayer/man/rtracklayer.p>. Acesso em: 03/09/2014.

LEITE, M. S. B. **Estrutura da cromatina:** Ação de detergentes e busca por peptídeo ideal ligante do PATCH acídico, 2013, 64f. Dissertação (Mestrado em Ciências Farmacêuticas) - Universidade de Brasília, Faculdade de Ciências da Saúde, Programa de Pós-Graduação em Ciências Farmacêuticas, Brasília.

LEVASSEUR, A.; DRULA, E.; LOMBARD, V.; COUTINHO, P. M.; HENRISSAT, B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. **Biotechnology for biofuels**, v. 6, n. 1, p. 41, 2013.

LI, C.; YANG, Z.; ZHANG, R. H. C.; et al. Effect of pH on cellulase production and morphology of *Trichoderma reesei* and the application in cellulosic material hydrolysis. **Journal of biotechnology**, v. 168, n. 4, p. 470–7, 2013.

LI, F.; MA, C.; WANG, X.; et al. Characterization of Sucrose transporter alleles and their association with seed yield-related traits in *Brassica napus L*. **BMC plant biology**, v. 11, n. 1, p. 168, 2011.

LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics (Oxford, England)**, v. 25, p. 1754–1760, 2009.

LI, H.; HANDSAKER, B.; WYSOKER, A.; et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics (Oxford, England)**, v. 25, n. 16, p. 2078–9, 2009.

LI, X.-L.; SKORY, C. D.; COTTA, M. A.; PUCHART, V.; BIELY, P. Novel family of carbohydrate esterases, based on identification of the *Hypocrea jecorina* acetyl esterase gene. **Applied and environmental microbiology**, v. 74, p. 7482–7489, 2008.

LICHIUS, A.; SEIDL-SEIBOTH, V.; SEIBOTH, B.; KUBICEK, C. P. Nucleo-cytoplasmic shuttling dynamics of the transcriptional regulators XYR1 and CRE1 under conditions of cellulase and xylanase gene expression in *Trichoderma reesei*. **Molecular microbiology**, v. 94, n. October, p. 1162–1178, 2014.

LIMA, V. F. DE. **Produção e uso de enzimas derivadas do fungo *Pleurotus ostreatus* na hidrólise de bagaço de cana pré-tratado por processo quimiotermomecânico**, 2013, 122f. Dissertação (Mestrado em Ciências - Programa de Pós-Graduação em Biotecnologia Industrial - Àrea de Concentração: Conversão de Biomassa) - Escola de Engenharia de Lorena da Universidade de São Paulo.

LING, M.; QIN, Y.; LI, N.; LIANG, Z. Binding of two transcriptional factors, Xyr1 and ACEI, in the promoter region of cellulase *cbh1* gene. **Biotechnology letters**, v. 31, n. 2, p. 227–31, 2009.

LIU, H.; CHENG, T.; XIAN, M.; et al. Fatty acid from the renewable sources: A promising feedstock for the production of biofuels and biobased chemicals. **Biotechnology advances**, v. 32, n. 2, p. 382–389, 2013.

LIVAK, K. J.; SCHMITTGEN, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. **Methods (San Diego, Calif.)**, v. 25, n. 4, p. 402–8, 2001.

LOMBARD, V.; GOLACONDA RAMULU, H.; DRULA, E.; COUTINHO, P. M.; HENRISSAT, B. The carbohydrate-active enzymes database (CAZy) in 2013. **Nucleic acids research**, v. 42, n. Database issue, p. D490–5, 2014.

LORA, E. S.; ANDRADE, R. V. Biomass as energy source in Brazil. **Renewable and Sustainable Energy Reviews**, v. 13, p. 777–788, 2009.

LUCY, T.; PETRI, D. F. S.; QUÍMICA, I. DE; et al. Revisão. **Química Nova**, v. 33, n. 7, p. 1549–1558, 2010.

MACH, R. L.; ZEILINGER, S. Regulation of gene expression in industrial fungi: *Trichoderma*. **Applied microbiology and biotechnology**, v. 60, p. 515–522, 2003.

MACH-AIGNER, A. R.; PUCHER, M. E.; MACH, R. L. D-Xylose as a repressor or inducer of xylanase expression in *Hypocrea jecorina* (*Trichoderma reesei*). **Applied and environmental microbiology**, v. 76, n. 6, p. 1770–6, 2010.

MACH-AIGNER, A. R.; PUCHER, M. E.; STEIGER, M. G.; et al. Transcriptional regulation of xyr1, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*. **Applied and environmental microbiology**, v. 74, p. 6554–6562, 2008.

MACPHERSON, S.; LAROCHELLE, M.; TURCOTTE, B. A fungal family of transcriptional regulators: the zinc cluster proteins. **Microbiology and molecular biology reviews: MMBR**, v. 70, p. 583–604, 2006.

MANDELS, M.; PARRISH, F. W.; REESE, E. T. Sophorose as an inducer of cellulase in *Trichoderma viride*. **Journal of bacteriology**, v. 83, p. 400–408, 1962.

MARCO, J. DA C. I. DE. **Produção e caracterização de mananase de *Aspergillus foetidus* cultivado em casca do grão da soja**, 2014, 90f. Dissertação (Mestrado em Biologia

Microbiana) - Universidade de Brasília, Instituto de Ciências Biológicas, Pós-Graduação em Biologia Microbiana, Brasília.

MARGEOT, A.; HAHN-HAGERDAL, B.; EDLUND, M.; SLADE, R.; MONOT, F. New improvements for lignocellulosic ethanol. **Current opinion in biotechnology**, v. 20, n. 3, p. 372–80, 2009.

MARGOLLES-CLARK, E.; TENKANEN, M.; SODERLUND, H.; PENTTILA, M. Acetyl xylan esterase from *Trichoderma reesei* contains an active-site serine residue and a cellulose-binding domain. **Eur. J. Biochem**, v. 237, p. 553–560, 1996a.

MARGOLLES-CLARK, E.; SALOHEIMO, M.; SIIKA-AHO, M.; PENTTIKI, M. The a-glucuronidase-encoding. **Gene**, v. 172, p. 171–172, 1996b.

MARGOLLES-CLARK, E.; TENKANEN, M.; NAKARI-SETA, T.; MERJA, P. Cloning of Genes Encoding alpha-L-Arabinofuranosidase and beta-Xylosidase from *Trichoderma reesei* by Expression in *Saccharomyces cerevisiae*. **Applied and Environmental Microbiology**, v. 62, n. 10, p. 3840–3846, 1996c.

MARGOLLES-CLARK, E.; TENKANEN, M.; LUONTERI, E.; PENTTILÄ, M. Three alpha-galactosidase genes of *Trichoderma reesei* cloned by expression in yeast. **European journal of biochemistry/FEBS**, v. 240, n. 1, p. 104–11, 1996d.

MARJOTTA-MAISTRO, M. C.; GUILHOTO, J. J. M. The importance of the sugar cane and alcohol sector and its relationships with the economic struture of Brazil. **MPRA Munich Personal RePEc Archive**, , n. 54226, p. 1–15, 2014.

MARTINEZ, D.; BERKA, R. M.; HENRISSAT, B.; et al. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). **Nature biotechnology**, v. 26, n. 5, p. 553–60, 2008.

MARTÍNEZ-ANTONIO, A.; COLLADO-VIDES, J. Identifying global regulators in transcriptional regulatory networks in bacteria. **Current Opinion in Microbiology**, 2003.

MARYANA, R.; MA'RIFATUN, D.; WHENI, A. I.; SATRIYO, K. W.; RIZAL, W. A. Alkaline Pretreatment on Sugarcane Bagasse for Bioethanol Production. **Energy Procedia**, v. 47, p. 250–254, 2014.

MASARIN, F.; GURPILHARES, D. B.; BAFFA, D. C.; et al. Chemical composition and enzymatic digestibility of sugarcane clones selected for varied lignin content. **Biotechnology for biofuels**, v. 4, n. 1, p. 55, 2011.

MAURYA, D. P.; SINGH, D.; PRATAP, D.; MAURYA, J. P. Optimization of solid state fermentation conditions for the production of cellulase by *Trichoderma reesei*. **Journal of environmental biology/Academy of Environmental Biology, India**, v. 33, n. 1, p. 5–8, 2012.

MELLO-DE-SOUSA, T. M.; GORSCHE, R.; RASSINGER, A.; et al. A truncated form of the Carbon catabolite repressor 1 increases cellulase production in *Trichoderma reesei*. **Biotechnology for biofuels**, v. 7, n. 1, p. 1–12, 2014.

MELLO-DE-SOUSA, T. M.; RASSINGER, A.; PUCHER, M. E.; et al. The impact of chromatin remodelling on cellulase expression in *Trichoderma reesei*. **Artigo científico submetido para publicação**, 2015.

METZ, B.; SEIDL-SEIBOTH, V.; HAARMANN, T.; et al. Expression of biomass-degrading enzymes is a major event during conidium development in *Trichoderma reesei*. **Eukaryotic cell**, v. 10, n. 11, p. 1527–35, 2011.

MIETTINEN-OINONEN, A. *Trichoderma reesei* strains for production of cellulases for the textile industry. **VTT Publications**, 2004, 96p.
Disponível em: <http://www.vtt.fi/inf/pdf/publications/2004/P550.pdf>. Acesso em: 03/01/2015.

MINTY, J. J.; SINGER, M. E.; SCHOLZ, S. A; et al. Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. **Proceedings of the National Academy of Sciences of the United States of America**, v. 110, n. 36, p. 14592–7, 2013.

MOON, R. J.; MARTINI, A.; NAIRN, J.; SIMONSEN, J.; YOUNGBLOOD, J. **Cellulose nanomaterials review:** structure, properties and nanocomposites, 2011, 55p (3941-3994). Disponível em:
<http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1767&context=nanopub>. Acesso em: 25/06/2014.

MOREIRA, F. C.; ITA, A. G.; SANTOS, W. D. DOS. Comparação entre dois métodos de determinação de lignina em bagaço de cana e tegumento de soja. VII EPCC - Encontro Interncional de Produção Científica Cesumar. **Anais...** . p.1–4, 2011.

MORGAN, M. Sequences, Genomes, and Genes in R/Bioconductor. **Bioconductor**, p. 1–46, 2013, 46p. Disponível em:
<http://www.ebi.ac.uk/training/sites/ebi.ac.uk.training/files/materials/2013/131021_HTS/genesandgenomes.pdf>. Acesso em: 03/09/2014.

MORGAN, M. T. Bioconductor for Sequence Analysis. **Bioconductor**, p. 1–15, 2014. Disponível em:
<http://master.bioconductor.org/help/coursematerials/2014/SeattleFeb2014/Bioconductor.pd> Acesso em: 03/09/2014.

MORIKAWA, Y.; OHASHI, T.; MANTANI, O.; OKADA, H. Cellulase induction by lactose in *Trichoderma reesei* PC-3-7. **Applied Microbiology and Biotechnology**, v. 44, p. 106–111, 1995.

MUKHERJEE, P. K.; HORWITZ, B. A; HERRERA-ESTRELLA, A.; SCHMOLL, M.; KENERLEY, C. M. *Trichoderma* research in the genome era. **Annual review of phytopathology**, v. 51, p. 105–29, 2013.

MURRAY, P. G.; COLLINS, C. M.; GRASSICK, A.; TUOHY, M. G. Molecular cloning, transcriptional, and expression analysis of the first cellulase gene (*cbh2*), encoding cellobiohydrolase II, from the moderately thermophilic fungus *Talaromyces emersonii* and structure prediction of the gene product. **Biochemical and Biophysical Research Communications**, v. 301, n. 2, p. 280–286, 2003.

NAKARI-SETÄLÄ, T.; PALOHEIMO, M.; KALLIO, J.; et al. Genetic modification of carbon catabolite repression in *Trichoderma reesei* for improved protein production. **Applied and environmental microbiology**, v. 75, n. 14, p. 4853–60, 2009.

NEVALAINEN, H.; PETERSON, R. Making recombinant proteins in filamentous fungi- are we expecting too much? **Frontiers in microbiology**, v. 5, n. February, p. 75, 2014.

NITTA, M.; FURUKAWA, T.; SHIDA, Y.; et al. A new Zn(II) 2Cys 6-type transcription factor BglR regulates β-glucosidase expression in *Trichoderma reesei*. **Fungal Genetics and Biology**, v. 49, p. 388–397, 2012.

NOGAWA, M.; GOTO, M.; OKADA, H.; MORIKAWA, Y. L-Sorbose induces cellulase gene transcription in the cellulolytic fungus *Trichoderma reesei*. **Current Genetics**, v. 38, p. 329–334, 2001.

NOGUCHI, Y.; SANO, M.; KANAMARU, K.; et al. Genes regulated by AoXlnR, the xylanolytic and cellulolytic transcriptional regulator, in *Aspergillus oryzae*. **Applied microbiology and biotechnology**, v. 85, p. 141–154, 2009.

NOVACANA. Anidro ou hidratado: diferenças. **novaCana.com**, p. 1–5, 2014a. Disponível em: <http://www.novacana.com/etanol/anidro-hidratado-diferencas/>. Acesso em: 10/06/2014.

NOVACANA. Etanol celulósico: o bagaço da cana-de-açúcar como matéria- prima para a hidrólise. **novaCana.com**, p. 5–9, 2014b.
Disponível em: <http://www.novacana.com/estudos/etanol-celulosico-bagaco-cana-de-acucar-como-materia-prima-para-hidrolise-241013/>. Acesso em: 02/07/2014.

NOVOZYMES. Novozymes Cellic® CTec3. **Novozymes**, p. 1, 2014. Disponível em: <http://www.bioenergy.novozymes.com/en/cellulosicethanol/CellicCTec3/Pages/default.asp> Acesso em: 10/06/2014.

OGATA, B. H. **Caracterização das frações celulose, hemicelulose e lignina de diferentes genótipos de cana-de-açúcar e potencial de uso em biorrefinarias**, 2013, 108p. Dissertação (Mestrado em Ciências) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba-SP.

OGAWA, M.; KOBAYASHI, T.; KOYAMA, Y. ManR, a Transcriptional Regulator of the β-Mannan Utilization System, Controls the Cellulose Utilization System in *Aspergillus oryzae*. **Bioscience, biotechnology, and biochemistry**, v. 77, p. 426–429, 2013.

OLSSON, L.; CHRISTENSEN, T. M. I. E.; HANSEN, K. P.; PALMQVIST, E. A. Influence of the carbon source on production of cellulases, hemicellulases and pectinases by *Trichoderma reesei* Rut C-30. **Enzyme and Microbial Technology**, v. 33, p. 612–619, 2003.

OSHLACK, A.; ROBINSON, M. D.; YOUNG, M. D. From RNA-seq reads to differential expression results. **Genome Biology**, v. 11, n. 220, p. 1–10, 2010.

ÖZCAN, S.; JOHNSTON, M. Function and Regulation of Yeast Hexose Transporters. **Microbiol Mol Biol Rev**, v. 63, n. 3, p. 554–569, 1999.

PABINGER, S.; DANDER, A.; FISCHER, M.; et al. A survey of tools for variant analysis of next-generation genome sequencing data. **Briefings in Bioinformatics**, v. 15, p. 256–278, 2014.

PAO, S. S.; PAULSEN, I. T.; SAIER, M. H. Major facilitator superfamily. **Microbiology and molecular biology reviews: MMBR**, v. 62, p. 1–34, 1998.

PARDO, E.; OREJAS, M. The Aspergillus nidulans Zn(II)2Cys6 transcription factor AN5673/RhaR mediates L-rhamnose utilization and the production of α-L-rhamnosidases. **Microbial Cell Factories**, v. 13, n. 1, p. 161, 2014.

PATIL, R.; GENCO, J. M.; PENDSE, H. CLEAVAGE OF ACETYL GROUPS FROM NORTHEAST HARDWOOD FOR ACETIC ACID PRODUCTION. Cleavage of acetyl groups from northeast hardwood for acetic acid production. PEERS Conference. **Anais...** . p.2177–2266, 2012.

PAULY, M.; GILLE, S.; LIU, L.; et al. Hemicellulose biosynthesis. **Planta**, v. 238, n. 4, p. 627–42, 2013.

PEDRO, N. C. DA R. M. **Avaliação do potencial de produção de etanol de 2ª geração a partir dos resíduos das podas do olival**. Tese (Doutorado em Química), 2013, 222f. Universidade da Beira Interior, Covilhã - Portugal.

PEGOS, V.; SALAZAR, M. Técnicas aplicadas a Genética e Biologia Molecular. **Unicamp**, p. 1–77, 2014.

PEIJ, N. N. M. E. VAN; GIELKENS, M. M. C.; VRIES, R. P. DE; VISSER, J.; GRAAFF, L. H. DE. The transcriptional activator XlnR regulates both xylanolytic and endoglucanase gene expression in *Aspergillus niger*. **Applied and Environmental Microbiology**, v. 64, p. 3615–3619, 1998a.

PEIJ, N. N. VAN; VISSER, J.; GRAAFF, L. H. DE. Isolation and analysis of xlnR, encoding a transcriptional activator co-ordinating xylanolytic expression in *Aspergillus niger*. **Mol Microbiol**, v. 27, n. 1, p. 131–42, 1998b.

PEREIRA, G. AMARANTE G.; CARNAÚBA, M.; FIORI, C. S. Relatório de biossegurança da levedura *Saccharomyces cerevisiae* linhagem RN1016. **BioCelere**, v. 55, n. 11, p. 1–289, 2013.

PETERSON, R.; NEVALAINEN, H. *Trichoderma reesei* RUT-C30--thirty years of strain improvement. **Microbiology (Reading, England)**, v. 158, n. Pt 1, p. 58–68, 2012.

PHILBROOK, A.; ALISSANDRATOS, A.; EASTON, C. J. Biochemical Processes for Generating Fuels and Commodity Chemicals from Lignocellulosic Biomass. **Environmental Biotechnology - New Approaches and Prospective Applications**, p. 39–64, 2013.

POKKULURI, P. R.; DUKE, N. E. C.; WOOD, S. J.; et al. Structure of the catalytic domain of glucuronoyl esterase Cip2 from *Hypocrea jecorina*. **Proteins: Structure, Function and Bioinformatics**, 2011.

PORTNOY, T.; MARGEOT, A.; SEIDL-SEIBOTH, V.; et al. Differential regulation of the cellulase transcription factors XYR1, ACE2, and ACE1 in *Trichoderma reesei* strains producing high and low levels of cellulase. **Eukaryotic cell**, v. 10, p. 262–271, 2011a.

PORTNOY, T.; MARGEOT, A.; LINKE, R.; et al. The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei*: a master regulator of carbon assimilation. **BMC genomics**, v. 12, p. 269, 2011b.

QUIROZ-CASTAÑEDA, R. E.; FOLCH-MALLOL, J. L. Hydrolysis of Biomass Mediated by Cellulases for the Production of Sugars. **Sustainable Degradation of Lignocellulosic Biomass - Techniques, Applications and Commercialization**. p.119–155, 2013.

RABONATO, A. C. **Linhagens fúngicas na hidrólise enzimática de bagaço de cana-de-açúcar**, 2013, 69f. Dissertação (Mestrado em Agronomia) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Ciências Agronômicas, Campus de Botucatu - SP.

RAMIRO, F. Various commercial processes, 2014. Disponível em: <http://www.sbioinformatics.com/design_thesis/Acrylic_Acid/Acrylic-2520acid_Methods 2520of-2520Production.pdf>. Acesso em: 02/07/2014.

RAUSCHER, R.; WÜRLEITNER, E.; WACENOVSKY, C.; et al. Transcriptional regulation of xyn1, encoding xylanase I, in *Hypocrea jecorina*. **Eukaryotic cell**, v. 5, p. 447–456, 2006.

RENEWABLES, B. GranBio. **BetaRenewables**, p. 1–2, 2014. Disponível em: <http://www.betarenewables.com/projects/5/granbio>. Acesso em: 14/07/2014.

REZENDE, C. A.; LIMA, M. A. DE; MAZIERO, P.; et al. Chemical and morphological characterization of sugarcane bagasse submitted to a delignification process for enhanced enzymatic digestibility. **Biotechnology for biofuels**, v. 4, n. 1, p. 54, 2011.

RIES, L.; BELSHAW, N. J.; ILMÉN, M.; et al. The role of CRE1 in nucleosome positioning within the cbh1 promoter and coding regions of *Trichoderma reesei*. **Applied microbiology and biotechnology**, v. 98, n. 2, p. 749–62, 2014.

RIES, L. N. A. **Regulation of genes encoding enzymes involved in plant cell wall deconstruction in *Trichoderma reesei***, 2013, 306p. Thesis (Doctor of Philosophy) - University of Nottingham - England.

RIES, L.; PULLAN, S. T.; DELMAS, S.; et al. Genome-wide transcriptional response of *Trichoderma reesei* to lignocellulose using RNA sequencing and comparison with *Aspergillus niger*. **BMC genomics**, v. 14, p. 541, 2013.

ROSSI, A.; CRUZ, A. H. S.; SANTOS, R. S.; et al. Ambient pH sensing in filamentous fungi: Pitfalls in elucidating regulatory hierarchical signaling networks. **IUBMB Life**, v. 65, p. 930–935, 2013.

ROSTAGNO, M. A; PRADO, J. M.; MUDHOO, A.; et al. Subcritical and supercritical technology for the production of second generation bioethanol. **Critical reviews in biotechnology**, v. 8551, p. 1–11, 2014.

RUBIN, E. M. Genomics of cellulosic biofuels. **Nature**, v. 454, n. 7206, p. 841–5, 2008.

RUIJTER, G. J. G.; VISSER, J. Carbon repression in *Aspergilli*. **FEMS Microbiology Letters**, 1997.

SALOHEIMO, A. Isolation of the ace1 Gene Encoding a Cys2-His2 Transcription Factor Involved in Regulation of Activity of the Cellulase Promoter *cbh1* of *Trichoderma reesei*. **Journal of Biological Chemistry**, v. 275, n. 8, p. 5817–5825, 2000.

SALOHEIMO, M.; NAKARI-SETÄLÄ, T.; TENKANEN, M.; PENTTILÄ, M. cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. **European journal of biochemistry / FEBS**, v. 249, n. 2, p. 584–91, 1997.

SALOHEIMO, M.; PALOHEIMO, M.; HAKOLA, S.; et al. Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. **European journal of biochemistry/FEBS**, v. 269, p. 4202–4211, 2002.

SAMUELS, G. J. *Trichoderma*: a review of biology and systematics of the genus. **Mycological Research**, v. 100, n. 8, p. 923–935, 1996.

SAMUELS, G. J. *Trichoderma*: systematics, the sexual state, and ecology. **Phytopathology**, v. 96, n. 2, p. 195–206, 2006.

SANTOS, F. A.; QUEIRÓZ, J. H. DE; COLODETTE, J. L.; FERNANDES, S. A.; GUIMARÃES, V. M. Potencial da palha de cana-de-açúcar para produção de etanol. **Química Nova**, v. 35, n. 5, p. 1004–1010, 2012.

SANTOS, J. O. DOS; SANTOS, R. M. DE S.; FERNANDES, A. DE A.; et al. Os impactos produzidos pelas mudanças climáticas. **ACSA - Agropecuária Científica no Seminário**, v. 9, n. 1, p. 9–16, 2013.

SCHARR, A. O tripé fundamental em qualquer projeto de etanol celulósico. **novaCana**, p. 1–8, 2013. Disponível em: <http://www.novacana.com/n/etanol/2-geracao-celulose/tripe-fundamental-projeto-etanol-celulosico-251013/>. Acesso em: 01/07/2014.

SCHARR, A. A levedura secreta da GranBio. **novaCana**, p. 2–5, 2014. Disponível em: <http://www.novacana.com/n/etanol/2-geracao-celulose/levedura-secreta-granbio-230513/>. Acesso em: 15/07/2014.

SCHMOLL, M.; KUBICEK, C. P. Regulation of *Trichoderma* cellulase formation: lessons in molecular biology from an industrial fungus. A review. **Acta microbiologica et immunologica Hungarica**, v. 50, p. 125–145, 2003.

SCHMOLL, M.; SCHUSTER, A.; SILVA, R. D. N.; KUBICEK, C. P. The G-alpha protein GNA3 of *Hypocrea jecorina* (Anamorph *Trichoderma reesei*) regulates cellulase gene expression in the presence of light. **Eukaryotic cell**, v. 8, n. 3, p. 410–20, 2009.

SCHNEIDER, T. D.; STEPHENS, R. M. Sequence logos: a new way to display consensus sequences. **Nucleic acids research**, v. 18, n. 20, p. 6097–100, 1990.

SCHUSTER, A.; SCHMOLL, M. Biology and biotechnology of *Trichoderma*. **Applied microbiology and biotechnology**, v. 87, n. 3, p. 787–99, 2010.

SECCON, D. M. **Análise de um clone metagenômico produtor de compostos potencialmente bioativos**, 2012, 79f. Dissertação (Mestrado em Ciências - Bioquímica) - Curso de Pós Graduação em Ciências, Universidade do Paraná, Curitiba.

SEIBOTH, B.; HAKOLA, S.; MACH, R. L.; SUOMINEN, P. L.; KUBICEK, C. P. Role of four major cellulases in triggering of cellulase gene expression by cellulose in *Trichoderma reesei*. **Journal of bacteriology**, v. 179, p. 5318–5320, 1997.

SEIBOTH, B.; IVANOVA, C.; SEIDL-SEIBOTH, V. *Trichoderma reesei*: A Fungal Enzyme Producer for Cellulosic Biofuels. **Intechopen**. p.309–341, 2011.

SEIBOTH, B.; KARIMI, R. A.; PHATALE, P. A; et al. The putative protein methyltransferase LAE1 controls cellulase gene expression in *Trichoderma reesei*. **Molecular microbiology**, v. 84, n. 6, p. 1150–64, 2012.

SEIBOTH, B.; MESSNER, R.; GRUBER, F.; KUBICEK, C. P. Disruption of the *Trichoderma reesei cbh2* gene coding for cellobiohydrolase II leads to a delay in the triggering of cellulase formation by cellulose. **Journal of General Microbiology**, 1992.

SEIDL, V.; GAMAUF, C.; DRUZHININA, I. S.; et al. The *Hypocrea jecorina* (*Trichoderma reesei*) hypercellulolytic mutant RUT C30 lacks a 85 kb (29 gene-encoding) region of the wild-type genome. **BMC genomics**, v. 9, n. 327, p. 1–15, 2008.

SHANNON, P.; MARKIEL, A.; OZIER, O.; et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. **Genome research**, v. 13, n. 11, p. 2498–504, 2003.

SHIDA, Y.; FURUKAWA, T.; OGASAWARA, W.; et al. Functional analysis of the *egl3* upstream region in filamentous fungus *Trichoderma reesei*. **Applied Microbiology and Biotechnology**, v. 78, p. 515–524, 2008.

SIFAEG. Etanol celulósico. Etanol de resíduos da cana-de-açúcar deve ser produzido este semestre. **SIFAEG - Sindicato da Indústria de Fabricação de Etanol do Estado de Goiás**, p. 1–2, 2014. Disponível em: <http://www.sifaeg.com.br/noticias/etanol-celulosico/>. Acesso em: 14/07/2014.

SILVA, F. V. Panorama e perspectivas do etanol lignocelulósico. **Revista Liberato, Novo Hamburgo**, v. 13, p. 01–134, 2012a.

SILVA, J. T. DA. **Genetic transcript analyzer - ferramenta computacional para análise de transcrição gênica por RNA-seq**, 2012b, 64f. Dissertação (Mestrado em Bioinformática) - Universidade do Paraná, Curitiba.

SILVA, N. L. C. Produção de bioetanol de segunda geração a partir de biomassa residual da indústria de celulose, 2010, 109f. Dissertação (Mestrado em Tecnologia de Processos Químicos e Bioquímicos) - Universidade Federal do Rio de Janeiro, Escola de Química, Rio de Janeiro.

SILVA, R. DO N. **Estudos de sinalização celular em *Hypocrea jecorina* (*Trichoderma reesei*) durante a expressão dos genes de celulase (*cbh1* e *cbh2*) em presença de celulose e soforose e durante o antagonismo contra *Pythium ultimum***, 2008, 129f. Tese (Doutorado em Biologia Molecular) - Instituto de Ciências Biológicas, Universidade de Brasília, Brasília.

SIQUEIRA, L. Maior usina de etanol 2G do mundo será construída na China. **novaCana**, p. 7–10, 2014. Disponível em: <http://www.novacana.com/n/etanol/2-geracao-celulose/china-maior-usina-etanol-2g-mundo-310714/>. Acesso em: 31/07/2014.

SMYTH, G.; RITCHIE, M.; SILVER, J.; et al. Package "limma" , 2014, 145p. Disponível em: <http://www.bioconductor.org/packages/release/bioc/vignettes/limma/inst/doc/usersguide.pdf>. Acesso em: 04/09/2014.

SOUZA, L. G. A. DE. **Redes de inovação em etanol de segunda geração**, 2013a, 215p. Tese (Doutorado em Ciências) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba - SP.

SOUZA, W. R. DE. Microbial Degradation of Lignocellulosic Biomass. **InTech**, p. 207–248, 2013b. Disponível em: <http://cdn.intechopen.com/pdfs-wm/44319.pdf>. Acesso em: 24/07/2014.

STÅLBRAND, H.; SALOHEIMO, A.; VEHMAANPERÄ, J.; HENRISSAT, B.; HENRISSAT, B. Cloning and expression in *Saccharomyces cerevisiae* of a *Trichoderma reesei* beta-mannanase gene containing a cellulose binding domain. **Applied and Environmental Microbiology**, v. 61, p. 1090–1097, 1995.

STANGL, H.; GRUBER, F.; KUBICEK, C. P. Characterization of the *Trichoderma reesei cbh2* promoter. **Current Genetics**, v. 23, p. 115–122, 1993.

STEIGER, M. G.; MACH, R. L.; MACH-AIGNER, A. R. An accurate normalization strategy for RT-qPCR in *Hypocrea jecorina* (*Trichoderma reesei*). **Journal of Biotechnology**, v. 145, p. 30–37, 2010.

STERNBERG, D.; MANDELS, G. R. Induction of cellulolytic enzymes in *Trichoderma reesei* by sophorose. **Journal of Bacteriology**, v. 139, n. 3, p. 761–769, 1979.

STRAUSS, J.; MACH, R. L.; ZEILINGER, S.; et al. Crel, the carbon catabolite repressor protein from *Trichoderma reesei*. **FEBS Letters**, 1995.

STRICKER, A.; MACH, R.; DE-GRAAFF, L. Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*). **Applied microbiology and biotechnology**, v. 78, n. 2, p. 211–20, 2008.

STRICKER, A. R.; GROSSTESSNER-HAIN, K.; WÜRLEITNER, E.; MACH, R. L. Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. **Eukaryotic cell**, v. 5, n. 12, p. 2128–37, 2006.

STRICKER, A. R.; STEIGER, M. G.; MACH, R. L. Xyr1 receives the lactose induction signal and regulates lactose metabolism in *Hypocrea jecorina*. **FEBS Letters**, v. 581, p. 3915–3920, 2007.

STRICKER, A. R.; TREFFLINGER, P.; ARO, N.; PENTTILÄ, M.; MACH, R. L. Role of Ace2 (Activator of Cellulases 2) within the xyn2 transcriptosome of *Hypocrea jecorina*. **Fungal Genetics and Biology**, v. 45, p. 436–445, 2008.

SUN, J.; TIAN, C.; DIAMOND, S.; GLASS, N. L. Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in *Neurospora crassa*. **Eukaryotic cell**, v. 11, n. 4, p. 482–93, 2012.

SUTO, M.; TOMITA, F. Induction and catabolite repression mechanisms of cellulase in fungi. **Journal of bioscience and bioengineering**, v. 92, n. 4, p. 305–11, 2001.

SZAKMARY, K.; WOTAWA, A.; KUBICEK, C. P. Origin of oxidized cellulose degradation products and mechanism of their promotion of cellobiohydrolase I biosynthesis in *Trichoderma reesei*. **Journal of General Microbiology**, 1991.

TAMURA, K.; DUDLEY, J.; NEI, M.; KUMAR, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. **Molecular biology and evolution**, v. 24, n. 8, p. 1596–9, 2007.

TANSEY, W. P. Transcriptional activation: risky business. **Genes & development**, v. 15, p. 1045–1050, 2001.

TENKANEN, M.; PULS, J.; POUTANEN, K. Two major xylanases of *Trichoderma reesei*. **Enzyme and Microbial Technology**, v. 14, n. 7, p. 566–574, 1992.

TEUGJAS, H.; VÄLJAMÄE, P. Selecting β-glucosidases to support cellulases in cellulose saccharification. **Biotechnology for biofuels**, v. 6, n. 1, p. 105, 2013.

THORVALDSDÓTTIR, H.; ROBINSON, J. T.; MESIROV, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. **Briefings in bioinformatics**, v. 14, n. 2, p. 178–92, 2013.

TIWARI, P.; MISRA, B. N.; SANGWAN, N. S. β-Glucosidases from the fungus *Trichoderma*: an efficient cellulase machinery in biotechnological applications. **BioMed research international**, v. 2013, p. 1–10, 2013.

TOMME, P.; TILBEURGH, H. VAN; PETTERSSON, G.; et al. Studies of the cellulolytic system of *Trichoderma reesei* QM9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. **European journal of biochemistry/FEBS**, v. 170, n. 3, p. 575–81, 1988.

TÖRRÖNEN, A.; MACH, R. L.; MESSNER, R.; et al. The Two Major Xylanases from *Trichoderma reesei*: Characterization of Both Enzymes and Genes. **Bio/Technology**, v. 10, n. 11, p. 1461–1465, 1992.

TRAPNELL, C.; WILLIAMS, B. A; PERTEA, G.; et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nature biotechnology**, v. 28, n. 5, p. 511–5, 2010.

UNICA. **Biomassa tem potencial para diversificar matriz energética.** 2013. Disponível em: <http://www.unica.com.br/unica-na-midia/21424699920311080579/biomassa-tem-potencial para-diversificar-matriz-energetica-por-cento0D-por-cento0A/>. Acesso em: 2/07/2014.

UNICA. Quatro usinas inauguram era do etanol celulósico em escala comercial nos EUA. **Portal do agronegócio**, p. 6–7, 2014.
Disponível em: <http://www.portaldoagronegocio.com.br/noticia/quatro-usinas-inauguram-era-do-etanol-celulosico-em-escala-comercial-nos-eua-110763>. Acesso em: 18/07/2014.

VANHOLME, B.; DESMET, T.; RONSSE, F.; et al. Towards a carbon-negative sustainable bio-based economy. **Frontiers in plant science**, v. 4, p. 174, 2013.

VARUZZA, L. **Introdução à análise de dados de sequenciadores de nova geração Versão 2.0.1**. 2013, 76p. Disponível em: <http://lvaruzza.com/files/apostila_bioinfo_2.0.1.pdf>. Acesso em: 18/08/2014.

VENABLES, W. N.; SMITH, D. M. An Introduction to R. **R Core Team**, v. 3.1.1, p. 1–105, 2014. Disponível em: <http://cran.r-project.org/doc/manuals/R-intro.pdf>. Acesso em: 03/09/2014.

VÊNCIO, R. Z. N.; KOIDE, T.; GOMES, S. L.; PEREIRA, C. A. DE B. BayGO: Bayesian analysis of ontology term enrichment in microarray data. **BMC bioinformatics**, v. 7, p. 86, 2006.

VERBEKE, J.; COUTINHO, P.; MATHIS, H.; et al. Transcriptional profiling of cellulase and expansin-related genes in a hypercellulolytic *Trichoderma reesei*. **Biotechnology letters**, v. 31, n. 9, p. 1399–405, 2009.

VIEIRA, M. C. A.; LIMA, J. F.; BRAGA, N. M. Setor Sucroalcooleiro Brasileiro : Evolução e Perspectivas. **BNDES**, p. 209–245, 2007.

VIIKARI, L.; VEHMAANPERÄ, J.; KOIVULA, A. Lignocellulosic ethanol: From science to industry. **Biomass and Bioenergy**, v. 46, p. 13–24, 2012.

VITIKAINEN, M.; ARVAS, M.; PAKULA, T.; et al. Array comparative genomic hybridization analysis of *Trichoderma reesei* strains with enhanced cellulase production properties. **BMC genomics**, v. 11, p. 441, 2010.

WANG, J.; QUIRK, A.; LIPKOWSKI, J.; DUTCHER, J. R.; CLARKE, A. J. Direct in situ observation of synergism between cellulolytic enzymes during the biodegradation of crystalline cellulose fibers. **Langmuir: the ACS journal of surfaces and colloids**, v. 29, n. 48, p. 14997–15005, 2013.

WEINHANDL, K.; WINKLER, M.; GLIEDER, A.; CAMATTARI, A. Carbon source dependent promoters in yeasts. **Microbial cell factories**, v. 13, p. 5, 2014.

WENDLAND, A.; SOUZA, A. DE; HENRISSAT, B.; et al. Anais do Simpósio Micro-organismos em Agroenergia: da Prospecção aos Bioprocessos. **EMBRAPA - Anais do Simpósio Micro-organismos em Agroenergia: da Pospecção aos Bioprocessos**, , n. 2177-4439, p. 161p, 2013.

WIECZORKE, R.; KRAMPE, S.; WEIERSTALL, T.; et al. Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. **FEBS Letters**, v. 464, p. 123–128, 1999.

WILSON-O'BRIEN, A. L.; PATRON, N.; ROGERS, S. Evolutionary ancestry and novel functions of the mammalian glucose transporter (GLUT) family. **BMC evolutionary biology**, v. 10, p. 152, 2010.

WU, F.-C.; WU, J.-Y.; LIAO, Y.-J.; WANG, M.-Y.; SHIH, I.-L. Sequential acid and enzymatic hydrolysis in situ and bioethanol production from *Gracilaria* biomass. **Bioresource technology**, v. 156, p. 123–31, 2014.

WÜRLEITNER, E.; PERA, L.; WACENOVSKY, C.; et al. Transcriptional regulation of *xyn2* in *Hypocrea jecorina*. **Eukaryotic cell**, v. 2, p. 150–158, 2003.

XIONG, Y.; CORADETTI, S. T.; LI, X.; et al. The proteome and phosphoproteome of Neurospora crassa in response to cellulose , sucrose and carbon starvation. **Fungal Genetics and Biology**, v. 72, p. 21–33, 2014.

XU, J.; NOGAWA, M.; OKADA, H.; MORIKAWA, Y. Regulation of *xyn3* gene expression in *Trichoderma reesei* PC-3-7. **Applied microbiology and biotechnology**, v. 54, p. 370–375, 2000.

XU, J.; TAKAKUWA, N.; NOGAWA, M.; OKADA, H.; MORIKAWA, Y. A third xylanase from *Trichoderma reesei* PC-3-7. **Applied Microbiology and Biotechnology**, v. 49, n. 6, p. 718–724, 1998.

YADAV, R. K. P.; TIMILSINA, A.; YADAWA, R. K.; POKHREL, C. P. Potential Cellulosic Ethanol Production from Organic Residues of Agro-Based Industries in Nepal. **ISRN Renewable Energy**, v. 2014, p. 1–6, 2014.

YAN, N. Structural advances for the major facilitator superfamily (MFS) transporters. **Trends in biochemical sciences**, v. 38, n. 3, p. 151–9, 2013.

YAO, J.; BAJJALIEH, S. M. SVOP is a nucleotide binding protein. **PLoS ONE**, v. 4, 2009.

ZAPAROLLI, D. Duas usinas de etanol celulósico entrarão em operação este ano no Brasil. **novaCana**, p. 8–10, 2014. Disponível em: <http://www.novacana.com/n/etanol/2-geracao-celulose/usinas-etanol-celulosico-operacao-brasil-280414/>. Acesso em: 14/07/2014.

ZEILINGER, S.; EBNER, A.; MAROSITS, T.; MACH, R.; KUBICEK, C. P. The *Hypocrea jecorina* HAP 2/3/5 protein complex binds to the inverted CCAAT-box (ATTGG) within the *cbh2* (cellobiohydrolase II-gene) activating element. **Molecular genetics and genomics: MGG**, v. 266, p. 56–63, 2001.

ZEILINGER, S.; MACH, R. L. .; KUBICEK, C. P. Two Adjacent Protein Binding Motifs in the *cbh2* (Cellobiohydrolase II-encoding) Promoter of the Fungus *Hypocrea jecorina* (*Trichoderma reesei*) Cooperate in the Induction by Cellulose. **Journal of Biological Chemistry**, v. 273, n. 51, p. 34463–34471, 1998.

ZEILINGER, S.; MACH, R. L.; SCHINDLER, M.; HERZOG, P.; KUBICEK, C. P. Different inducibility of expression of the two xylanase genes *xyn1* and *xyn2* in *Trichoderma reesei*. **The Journal of biological chemistry**, v. 271, p. 25624–25629, 1996.

ZEILINGER, S.; SCHMOLL, M.; PAIL, M.; MACH, R. L.; KUBICEK, C. P. Nucleosome transactions on the *Hypocrea jecorina* (*Trichoderma reesei*) cellulase promoter *cbh2* associated with cellulase induction. **Molecular Genetics and Genomics**, v. 270, p. 46–55, 2003.

ZHANG, W.; KOU, Y.; XU, J.; et al. Two major facilitator superfamily sugar transporters from *Trichoderma reesei* and their roles in induction of cellulase biosynthesis. **The Journal of biological chemistry**, v. 288, n. 46, p. 32861–72, 2013.

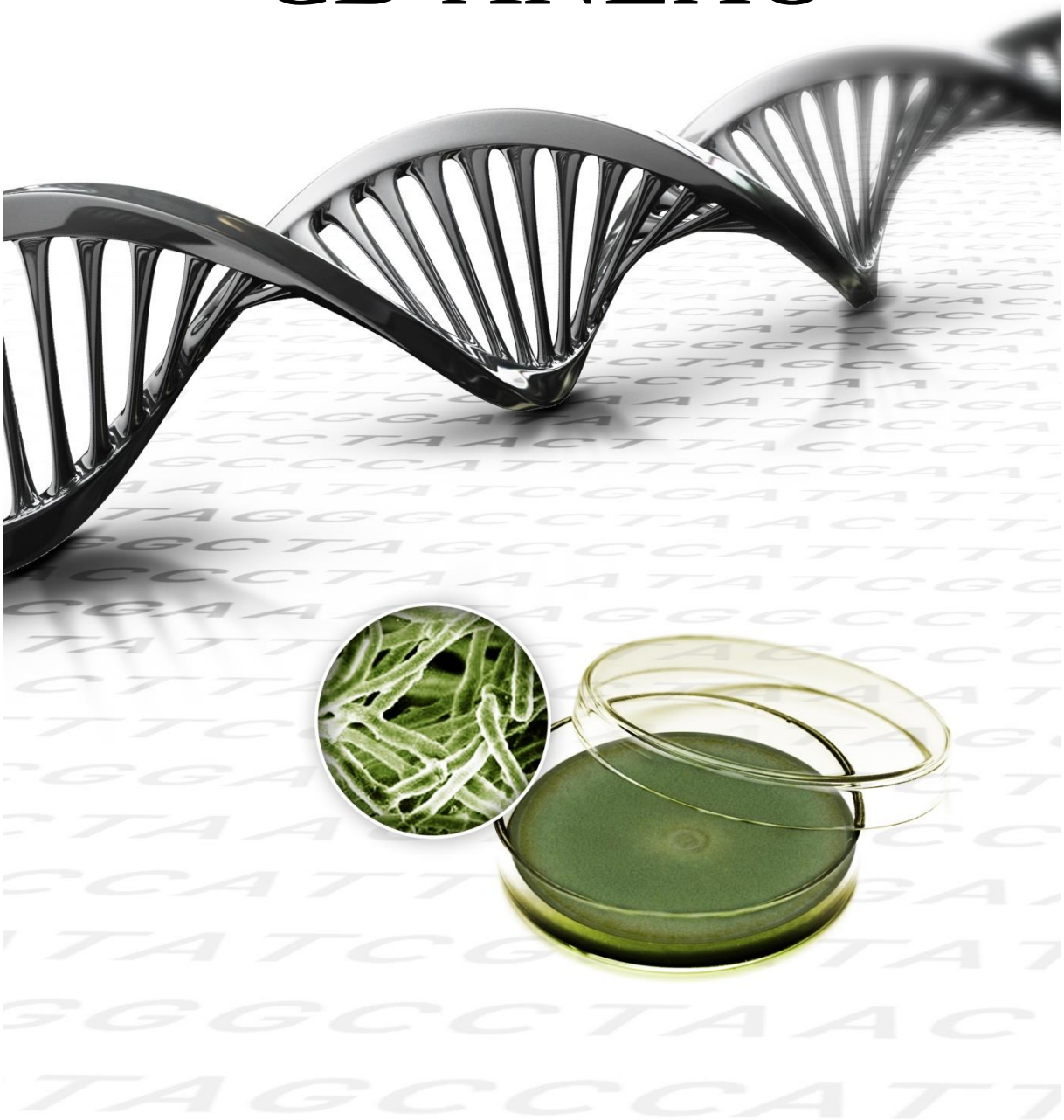
ZHAO, C.; WAALWIJK, C.; WIT, P. J. G. M. DE; LEE, T. VAN DER; TANG, D. EBR1, a novel Zn(2)Cys(6) transcription factor, affects virulence and apical dominance of the hyphal tip in *Fusarium graminearum*. **Molecular plant-microbe interactions : MPMI**, v. 24, p. 1407–18, 2011.

ZHOU, Q.; XU, J.; KOU, Y.; et al. Differential involvement of β-glucosidases from *Hypocrea jecorina* in rapid induction of cellulase genes by cellulose and cellobiose. **Eukaryotic cell**, v. 11, n. 11, p. 1371–81, 2012.

ZNAMEROSKI, E. A.; CORADETTI, S. T.; ROCHE, C. M.; et al. Induction of lignocellulose-degrading enzymes in *Neurospora crassa* by cellodextrins. **Proceedings of the National Academy of Sciences**, 2012.

ZNAMEROSKI, E. A.; GLASS, N. L. Using a model filamentous fungus to unravel mechanisms of lignocellulose deconstruction. **Biotechnology for Biofuels**, v. 6, n. 1, p. 1, 2013.

CD ANEXO



CD ANEXO

Arquivo 1: Tabelas 1.1 a 1.3

Tabela 1.1 – Fatores Transcricionais up-regulados na condição de crescimento celulose

Tabela 1.2 – Fatores Transcricionais up-regulados na condição de crescimento soforose

Tabela 1.3 – Fatores Transcricionais up-regulados na condição de crescimento glicose

Arquivo 2: Tabela 2 – 9129 genes analisados por RNA-seq em *T. reesei* QM9414

Arquivo 3: Tabela 3 - 9129 genes analisados por RNA-seq em *T. reesei Δxyr1* comparado ao *T. reesei* QM9414

Arquivo 4: Tabela 4 – Predição de todo o genoma de sítios de ligação de XYR1 em *T. reesei*

Arquivo 5: Tabela 5 – Predição de sítios de ligação XYR1 em genes down-regulados induzidos por celulose no mutante *Δxyr1*

Arquivo 6: Tabela 6 – Predição de sítios de ligação XYR1 em genes up-regulados induzidos por celulose no mutante *Δxyr1*

Arquivo 7: Tabela 7 – Predição de sítios de ligação XYR1 em genes down-regulados induzidos por soforose no mutante *Δxyr1*

Arquivo 8: Tabela 8 – Predição de sítios de ligação XYR1 em genes up-regulados induzidos por soforose no mutante *Δxyr1*

Arquivo 9: Tabelas 9.1 a 9.3

Tabela 9.1. Genes do regulon de Celulose

Tabela 9.2. Genes do regulon de Glicose

Tabela 9.3. Genes do regulon de Soforose

Arquivo 10: Tabela 10 – Genes diferencialmente expressos em celulose e soforose.

Arquivo 11: Tabela 11 – Comparação dos níveis de expressão de genes avaliados por RNA-seq e qPCR.

Arquivo 12: Tabelas 12.1 a 12.6

Tabela 12.1 – Genes regulon glicose (Down-regulados em *Δxyr1*/QM9414)

Tabela 12.2 – Genes regulon soforose (Down-regulados em *Δxyr1*/QM9414)

Tabela 12.3 – Genes regulon celulose (Up-regulados em *Δxyr1*/QM9414)

Tabela 12.4 – Genes regulon soforose (Up-regulados em *Δxyr1*/QM9414)

Tabela 12.5 – Genes regulon celulose (Down-regulados em *Δxyr1*/QM9414)

Tabela 12.6 – Genes regulon glicose (Up-regulados em *Δxyr1*/QM9414)

Arquivo 13: Tabelas 13.1 a 13.10

Tabela 13.1 – Rede regulatória: genes up-regulados em Celulose (*Δxyr1*/QM9414)

Tabela 13.2 – Rede regulatória: genes down-regulados em Celulose (*Δxyr1*/QM9414)

Tabela 13.3 – Rede regulatória: genes up-regulados em Soforose (*Δxyr1*/QM9414)

Tabela 13.4 – Rede regulatória: genes down-regulados em Soforose (*Δxyr1*/QM9414)

Tabela 13.5 – Rede regulatória: genes up-regulados em Glicose (*Δxyr1*/QM9414)

Tabela 13.6 – Rede regulatória: genes down-regulados em Glicose (*Δxyr1*/QM9414)

Tabela 13.7 – Rede regulatória: Celulose versus Soforose (*Δxyr1*/QM9414)

Tabela 13.8 – Rede regulatória: Soforose versus Glicose (*Δxyr1*/QM9414)

Tabela 13.9 – Rede regulatória: Celulose versus Glicose (*Δxyr1*/QM9414)

Tabela 13.10 – Rede regulatória: todas as condições (*Δxyr1*/QM9414)

Arquivo 14: Tabelas 14.1 a 14.3

Tabela 14.1 – Genes *CAZy* up- ou down-reguladas em celulose (*Δxyr1*/QM9414)

Tabela 14.2 – Genes *CAZy* up- ou down-reguladas em soforose (*Δxyr1*/QM9414)

Tabela 14.3 – Genes *CAZy* up- ou down-reguladas em glicose (*Δxyr1*/QM9414)

Arquivo 15: Tabela 15 – Transportadores up- e down-regulados em presença de celulose, soforose e glicose (*Δxyr1*/QM9414)

Arquivo 16: Tabela 16 – Comparação dos níveis de expressão de genes testados por RNA-seq e qPCR-RT em Δ*xyr1*/QM9414.

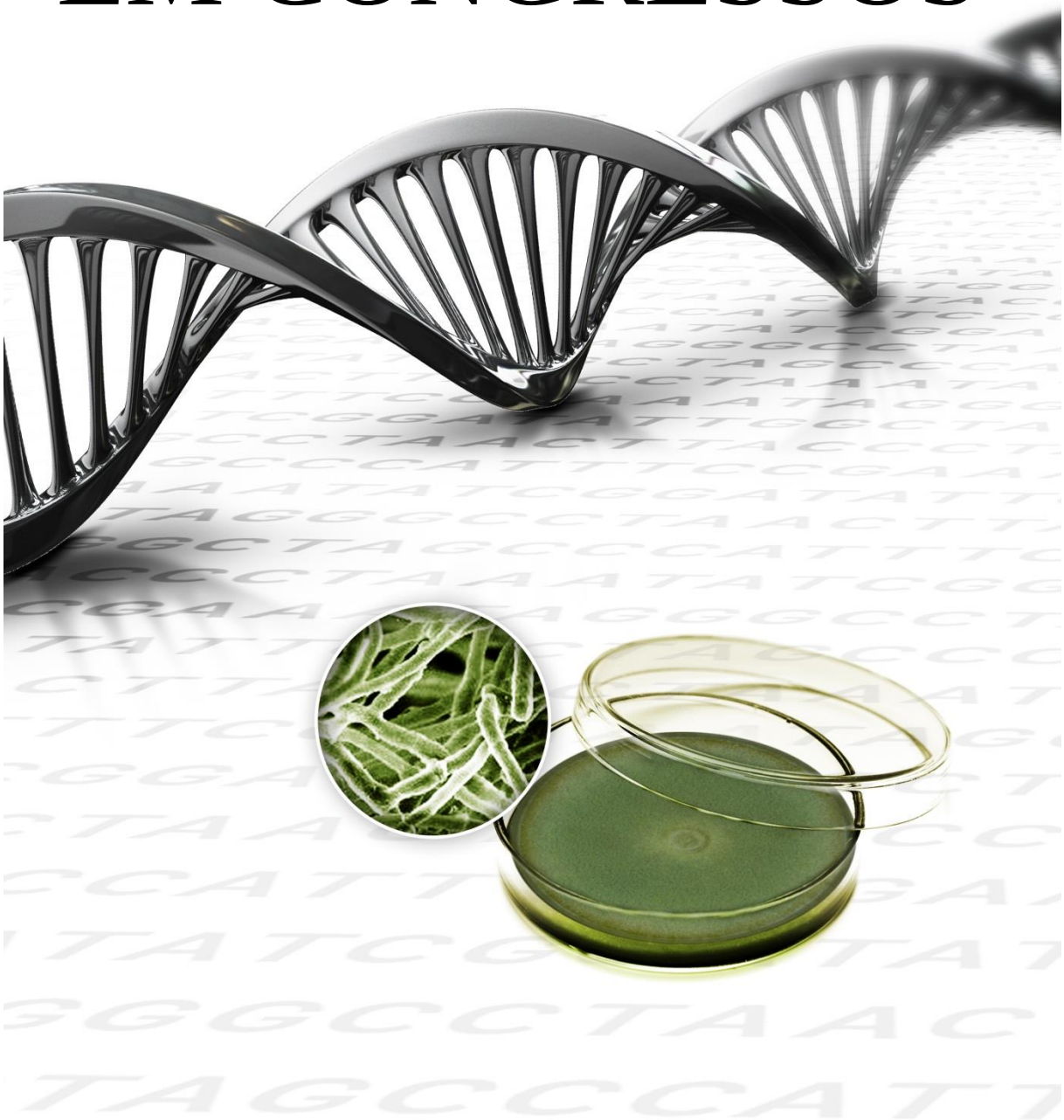
Arquivo 17: Tabela 17.1 a 17.3

Tabela 17.1 – GHs específicas das linhagens QM9414 e Δ*xyr1*

Tabela 17.2 – Transportadores específicos das linhagens QM9414 e Δ*xyr1*

Tabela 17.3 – Fatores específicos das linhagens QM9414 e Δ*xyr1*

PARTICIPAÇÕES EM CONGRESSOS



PARTICIPAÇÕES EM CONGRESSOS

Resumos publicados em anais de congressos nacionais

- **CASTRO L dos S.**; ANTONIETO, A. C. C.; PEDERSOLI, W. R.; CIPRIANO, A. K. A. L.; COSTA, J. S.; TRAMONTINA, R.; SILVA, R. N. Expression Analysis of Cellulase Genes of *Trichoderma reesei* (*Hypocrea Jecorina*) Regulated by *Gαi* Protein in Presence of Cellulose. In: **XLI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq, 2012, Foz do Iguaçu - PR.**
- PEDERSOLI, W. R.; **CASTRO L dos S.**; ANTONIETO, A. C. C.; SILVA, R. N. Analysis of Differentially Secretome of *Trichoderma reesei* Regulated by *Gαi* Protein Using Two-dimensional Differential Gel Electrophoresis. In: **XLI Reunião Annual da Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq, 2012, Foz do Iguaçu - PR.**
- PEDERSOLI, W. R.; **CASTRO L dos S.**; ANTONIETO, A. C. C.; SILVA, R. N. Analysis of Secretome of *Trichoderma reesei* by Two-Dimensional Differential Gel Electrophoresis During Cellulase Formation. In: **XLI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq, 2012, Foz do Iguaçu - PR.**

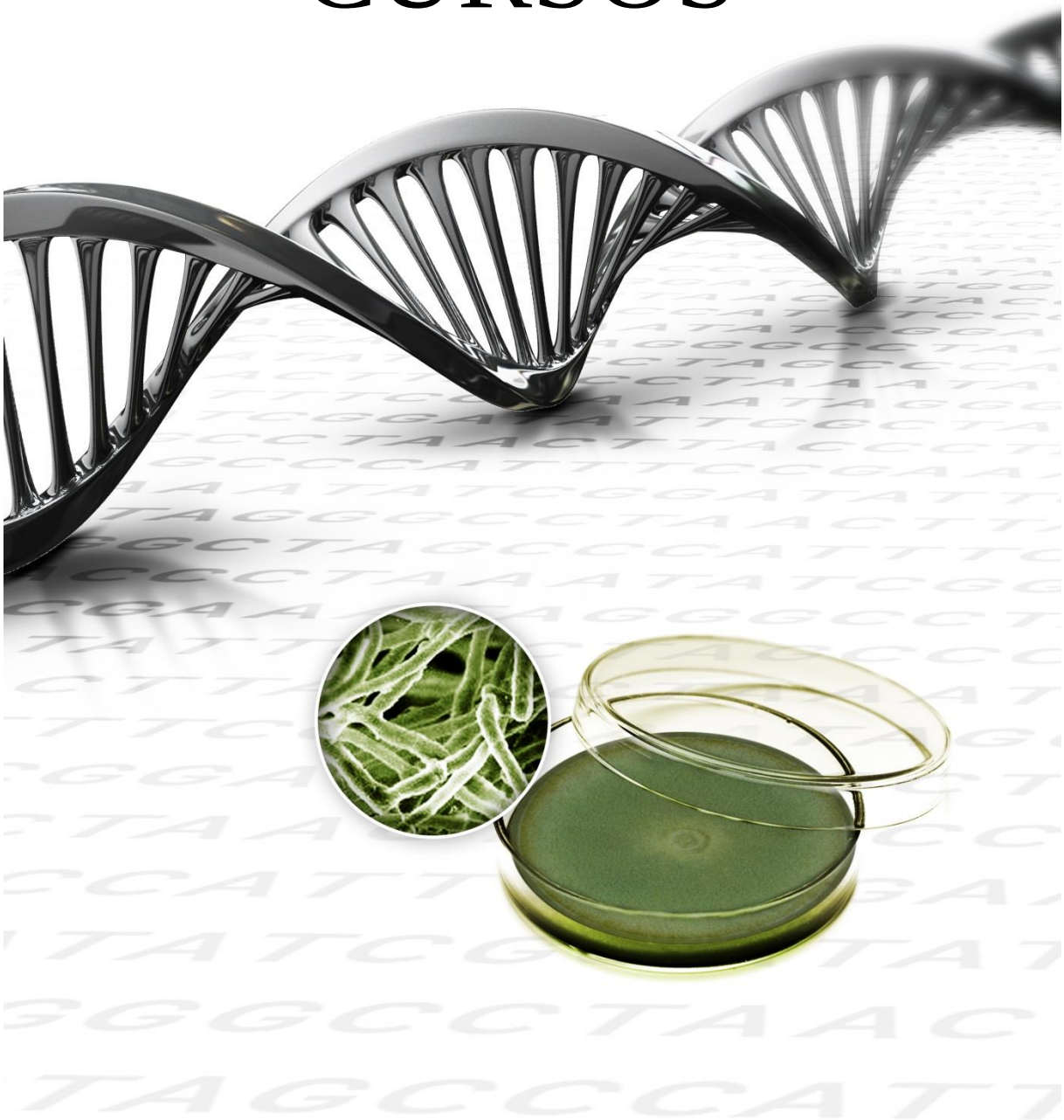
Resumos publicados em anais de congressos internacionais

- **CASTRO L dos S.**; PERSINOTI, G. F.; ANTONIETO, A. C. C.; SILVA-ROCHA, R.; SILVA, R. N. New insights into regulation of biomass degrading enzymes by XYR1 in *Trichoderma reesei*. In: **12th European Conference on Fungal Genetics, 2014, Seville (Spain).**
- SILVA, R. N.; ANTONIETO, A. C. C.; **CASTRO L dos S.**; PERSINOTI, G. F. Analysis of global regulation of gene expression of carbon catabolite repression during the formation on cellulases by *Trichoderma reesei* (*Hypocrea Jecorina*). In: **12th European Conference on Fungal Genetics, 2014, Seville (Spain).**
- ANTONIETO, A. C. C.; **CASTRO L dos S.,** PEDERSOLI, W. R.; SILVA, R. N. The transcriptional factors XYR1 and CRE1 regulate the expression of Cellulolytic and Xylanolytic genes at carbon source dependent-manner in *Hypocrea jecorina*

(*Trichoderma reesei*). In: **27th Fungal Genetics Conference, 2013, Pacific Grove, California.**

- PEDERSOLI, W. R.; **CASTRO L dos S.**; ANTONIETO, A. C. C.; FACA, V. M.; SILVA, R. N. Analysis of carbon catabolite repression (CCR) during cellulase formation by *Trichoderma reesei* (*Hypocrea jecorina*) using two-dimensional differential gel electrophoresis (2D-DIGE). In: **27th Fungal Genetics Conference, 2013, Pacific Grove - California.**
- ANTONIETO, A. C. C.; **CASTRO L dos S.**; STEINDORFF, A. S.; PERSINOTI, G. F.; SILVA, R. N. Decoding the black box of cellulase formation in *Hypocrea jecorina* by RNA-SEQ analysis. In: **FEMS 2013 5th Congress of European Microbiologists, 2013, Leipzig - Germany.**

CURSOS



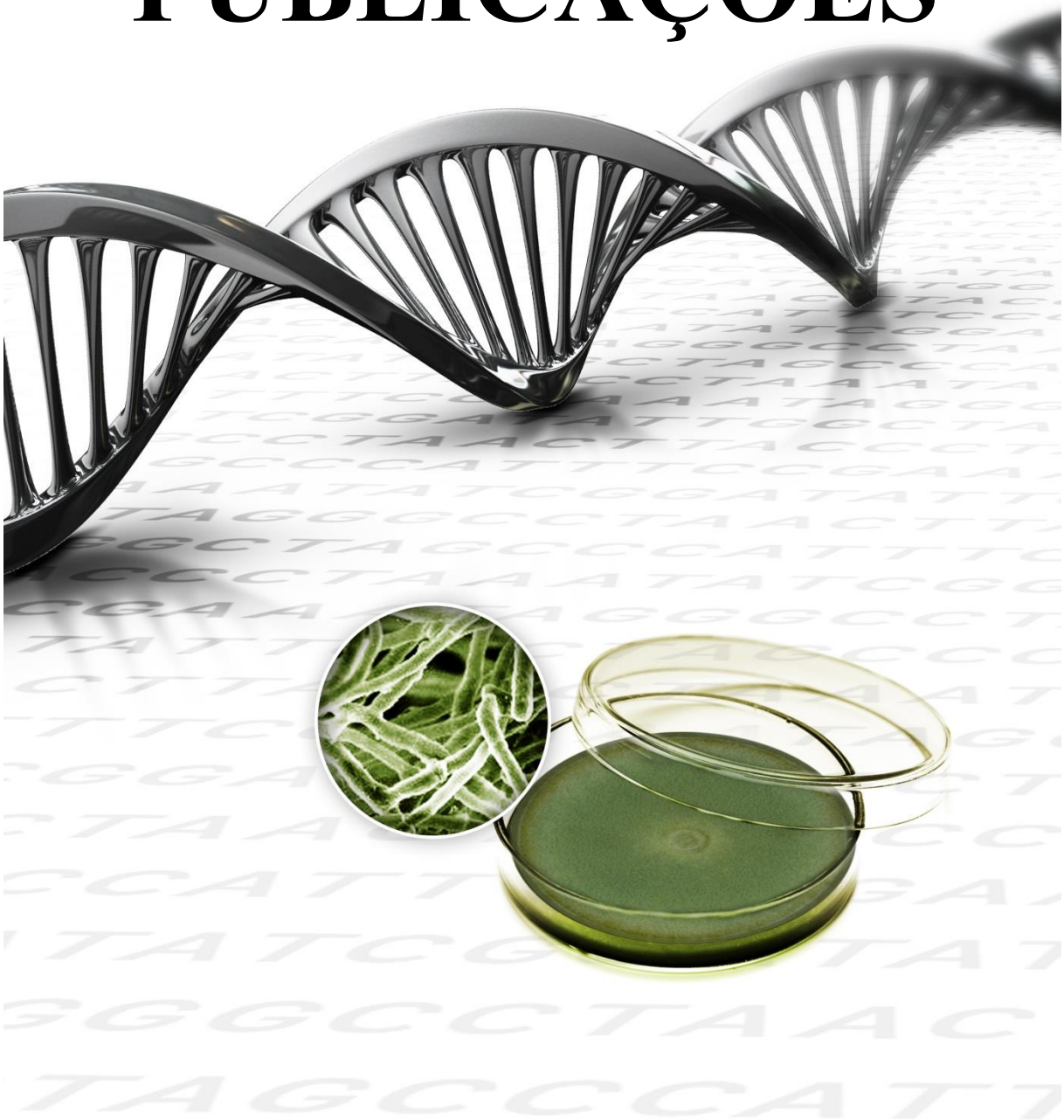
CURSOS

- Workshops de capacitação para pesquisadores da USP em publicação científica. 4º Workshop - FMRP – 10 e 11 de Dezembro; Ribeirão Preto – SP (16 horas);

- A Bioinformática no Século 21. 2Bio Instituto de Bioinformática e Biotecnologia. 19 a 23 de novembro de 2012 – Natal, – RN (30 Horas);

- Using R/Bioconductor to decode 'OMICS' data. 10th summer bioinformatics course. Ribeirão Preto Medical School (FMRP). University of São Paulo (USP); Jan 27 - Feb 7, 2014; Ribeirão Preto – SP (70 horas).

ANEXOS DE PUBLICAÇÕES



Gene Expression Patterns 14 (2014) 88–95

Contents lists available at ScienceDirect

Gene Expression Patterns

journal homepage: www.elsevier.com/locate/gep

Expression pattern of cellulolytic and xylanolytic genes regulated by transcriptional factors XYR1 and CRE1 are affected by carbon source in *Trichoderma reesei*

Lilian dos Santos Castro [a,1], Amanda Cristina Campos Antoniêto [a,1], Wellington Ramos Pedersoli [a], Rafael Silva-Rocha [a], Gabriela F. Persinoti [b], Roberto Nascimento Silva [a,*]

[a] Department of Biochemistry and Immunology, Ribeirao Preto Medical School, University of Sao Paulo, 14049-900 Ribeirão Preto, SP, Brazil
[b] Department of Genetic, Ribeirao Preto Medical School, University of Sao Paulo, 14049-900 Ribeirão Preto, SP, Brazil

ARTICLE INFO

Article history:
Received 6 August 2013
Received in revised form 9 December 2013
Accepted 7 January 2014
Available online 27 January 2014

Keywords:
Cellulases
Gene expression
Transcriptional factor
Trichoderma reesei
Xylanases

ABSTRACT

Trichoderma reesei is the most important fungus for the industrial production of enzymes to biomass deconstruction. Most of the genes encoding cellulases and hemicellulases are regulated by the transcription factors CRE1 and XYR1. In this work, the regulation of 22 genes of cellulases and xylanases by these transcription factors was investigated under three different carbon sources. Analysis of gene expression and enzymatic profiles of CMCase, β-glucosidase, and xylanases showed different regulation that was depended of the carbon source in both Δ*xyr1* and Δ*cre1* mutants. In the presence of glucose, the majority of genes evaluated (82%) showed increased expression levels in the Δ*cre1* mutant compared to the parental QM9414 strain. In the Δ*xyr1* mutant, it was observed that expression of cellulase and xylanase genes was reduced compared to the parental QM9414 strain, when cultured in the presence of cellulose or sophorose. Interesting, in the presence of glucose, approximately 60% of the analyzed genes had increased expression in the Δ*xyr1* mutant compared to parental strain. Furthermore, no correlation between gene expression and the number of putative binding sites of XYR1 and CRE1 to promoter region of cellulolytic and xylanolytic studied genes was observed. Therefore, these results demonstrated that the regulation of cellulase and xylanase by the transcription factors CRE1 and XYR1 is influenced by different carbon sources.

© 2014 Elsevier B.V. All rights reserved.

The ascomycete *Trichoderma reesei* (anamorph of *Hypocrea jecorina*) is one of the most well-studied fungi used in biotechnology as a potential source of industrial cellulases and xylanases for biomass degradation (Saloheimo and Pakula, 2012). The cellulolytic system of this fungus consists of at least 3 different types of enzymes: exoglucanases (cellobiohydrolases EC 3.2.1.91), endoglucanases (EC 3.2.1.4), and β-glucosidases (EC 3.2.1.21), which may present under different isoforms (Saloheimo et al., 1994). Additionally, hemicellulases are also necessary for the hydrolysis of vegetal biomass. Theses enzymes act in the hydrolysis of hemicellulose, which xylan is the most abundant component, and increase the efficiency of cellulases action (Cabero et al., 2012). The fungus *T. reesei* has 5 genes that encode for xylanases, designated *xyn1* to *xyn5*, and 16 genes that encode hemicellulases (Foreman et al., 2003; Herold et al., 2013). Regulation of genes that encode cellulases and hemicellulases has been extensively studied in *T. reesei* (Schmoll and Kubicek, 2003; Stricker et al., 2008). Five transcription factors (TFs) important in this process have been previously described; positive regulators such as XYR1, ACE2, and the HAP2/3/5 complex, and negative regulators such as ACE1 and the carbon catabolic repressor CRE1 (Portnoy et al., 2011b).

XYR1 is the main transcriptional activator of genes that encode hydrolases such as *xyn1*, *xyn2*, *bxl1*, *cbh1*, *cbh2*, *egl1*, and *bgl1*. Functional XYR1-binding sequences have been described to be not only 5′-GGCTAA-3′ motifs arranged as an inverted repeat (Rauscher et al., 2006), but also as a single motif which have A or T substitutions in the 3′-proximal three bases, 5′-GGC(A/T)$_3$-3′ (Furukawa et al., 2009). Deletion of *xyr1* eliminates the induction of cellulases in the presence of cellulose and sophorose, highlighting its essential role in induction (Stricker et al., 2008). The regulation of cellulases and hemicellulases expression is also a target of carbon catabolic repression (CCR). In *T. reesei* CCR is mediated by CRE1 (Strauss et al., 1995), a Cys2His2 zinc finger protein homologous to the CreA from *Aspergillus nidulans* (Dowzer and Kelly, 1991). This

* Corresponding author. Tel.: +55 1636023260; fax: +55 (16) 3633 6840.
E-mail address: rsilva@fmrp.usp.br (R.N. Silva).
[1] Both authors contributed equally to this paper.

http://dx.doi.org/10.1016/j.gep.2014.01.003
1567-133X/© 2014 Elsevier B.V. All rights reserved.

protein has a consensus sequence at its binding site, 5′-SYGGRG-3′, that was identified in the promoters of *cbh1*, *xyn1*, *cbh2*, *egl1*, and *egl2* in *T. reesei* (Ilmen et al., 1997). It has been postulate that CCR would only occur if CRE1 bind in double active sites and this binding is mediated by specific serine phosphorylation (Portnoy et al., 2011a). An analysis of *cre1* deletion mutants demonstrated that catabolic derepression was not sufficient to increase cellulase and xylanase production, suggesting that overexpression of these enzymes depends on inducers (Nakari-Setala et al., 2009).

The sequencing of *T. reesei* genome allowed the identification of new targets for XYR1 and CRE1 (Kubicek, 2012). However, the effects of regulation of XYR1 and CRE1 on expression of cellulases or xylanases are described for a limited number of genes. These studies were restricted in the presence of only cellulose/lactose as carbon sources (Ilmen et al., 1997; Portnoy et al., 2011a). In this report, we used three different carbon sources to analyze 22 genes of cellulases and xylanases regulated by XYR1 and CRE1 in *T. reesei*. We also investigate a possible relation between the number of putative binding sites for XYR1 and CRE1 and gene expression regulation. The results demonstrated that the regulation of cellulase and xylanase by the transcription factors CRE1 and XYR1 is carbon source dependent and no correlation between gene expression and the number of putative binding sites of XYR1 and CRE1 to promoter region of 22 studied genes was observed. Our findings provide new insights into the regulation of cellulase and xylanase biosynthesis in *T. reesei* and contribute for improvement of biomass-degrading enzymes production for applications in the biofuel industry.

1. Results

1.1. In silico analysis of putative binding sites for XYR1 and CRE1 in promoter regions of cellulolytic and xylanolytic genes

In order to establish a possible relationship between the expression of cellulolytic and xylanolytic genes and the number of putative binding sites of XYR1 and CRE1, we looked for potential binding motifs in the 5′-upstream region (1.5 kbp) of 22 cellulolytic and xylanolytic genes. Among several possibilities generated by the sequence 5′-GGCWWW-3′, we compared the frequency presence of the motifs in the 1.5 kbp 5′-upstream region of the 22 XYR1-regulated genes with to the whole *T. reesei* genes. The 5′-GGCTAA-3′, 5′-GGCTAT-3′ and, 5′-GGCAAA-3′ motifs were found at 5.3-, 2.3-, and 1.6-fold higher frequencies in the XYR1-regulated genes compared to the other ORFs. In this way, we identified 78 single binding sites and 16 sites arranged as an inverted repeat sequence in the promoter region of 22 analyzed genes (Fig. 1). On the other hand, for the catabolic repressor CRE1, 31 double binding motifs were found in the promoter region of 22 analyzed genes. Among these motifs, 18 were identified to bind directly and 13 arranged as an inverted repeat (Fig. 1).

The genes that showed more putative binding sites for XYR1 were *cel7a*, *xyn3*, *cel6a*, and *cel45a* (18, 9, and 7 sites respectively) (Fig. 1). Among them, *cel7a*, *cel6a* showed increased expression in QM9414 when cellulose (Fig. 3A) and sophorose (Fig. 3B) were used as inducers and reduced expression in presence of glucose (Fig. 3C), whereas *cel45a* and *xyn3* showed reduced expression in presence of cellulose (Fig. 3A) and sophorose (Fig. 3B) and *cel45a* showed increased expression in presence of glucose (Fig. 3C). Opposite results were found for the mutant Δ*xyr1* when the same genes were analyzed (Fig. 3). Among the genes with low numbers of XYR1 motifs binding *cel74a* and *cel3a* (only one site for each gene) presented similar behavior to *cel7a*. Furthermore, for each carbon source, genes were grouped into different clusters, suggesting dependence between XYR1 and the carbon source in the regulation of some cellulolytic and xylanolytic genes (Fig. 3). For CRE1, only *cel1b* and *cel3e*, encoding for β-glucosidases, showed 6 putative binding sites. Six genes (*cel45a*, *cel3a*, *cel5b*, *xyn4*, *cel7b* and *cel12a*) did not present any double binding motif for CRE1 (Fig. 1). Analyzing the mutant Δ*cre1*, most genes with the high number of motifs for CRE1 binding showed an increase of expression on cellulose, sophorose and glucose (Fig. 3). On the other hand, genes with no binding site for this factor were also up regulated in the three carbon sources (Fig. 3). Interestingly, even in the Δ*xyr1* mutant, *cel1b* and *cel3e* also showed high expression in cellulose, sophorose, and glucose compared with QM9414 (Fig. 3), suggesting cooperation between XYR1 and CRE1 in the regulation of these genes. Together, *in silico* analysis results provide no evidence of a correlation between the expression of cellulolytic and xylanolytic genes and the number of putative binding sites for both XYR1 and CRE1.

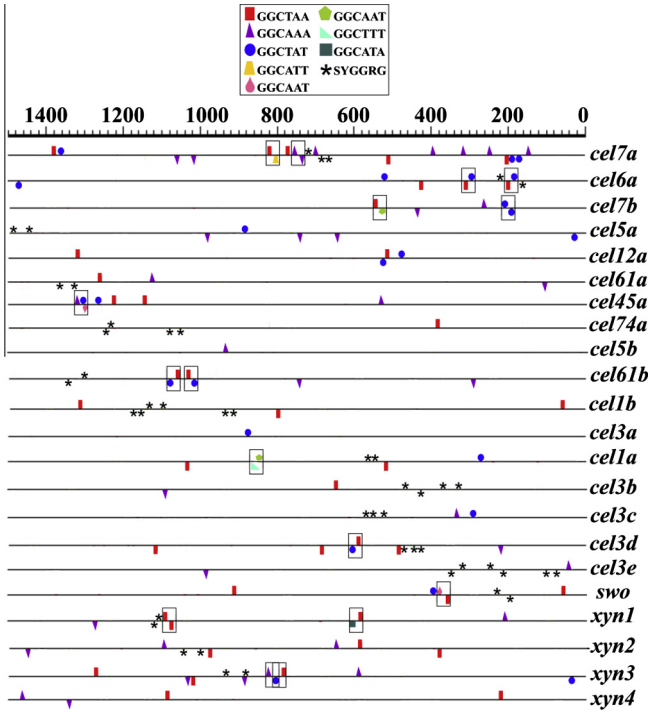


Fig. 1. Schematic representation of putative binding sites of XYR1 (5′-GGCWWW-3′) and CRE1 (5′-SYGGRG-3′) binding in the 1.5-kbp upstream region of cellulolytic and xylanolytic genes. The position of the motifs is relative to the translation initiation codon (ATG). The symbols above and below the lines indicate binding motifs in the sense and nonsense DNA strand respectively. The binding motifs for CRE1 are represented by inverted and direct repeats. XYR1 inverted repeats sites are highlighted (rectangle) and simple sites are indicated at the legend. W = A/T, S = C/G, Y = C/T e R = A/G, according to IUPAC norms.

1.2. Analysis of the influence of carbon source on regulation of cellulolytic and xylanolytic gene expression by XYR1 and CRE1

To guarantee comparisons between strains, we first examined the growth pattern of strains on cellulose and glucose. As shown in Fig. 2A and B, only Δ*cre1* showed a discrete difference ($P > 0.05$) in growth when compare to the parental QM9414. No differences was observed in glucose consuming at 24 and 48 h and at the end of experiment about 25% of glucose was still available for strains, excluding the possibility of stress by nutrient exhaustion (Fig. 2C). Furthermore, the pH behavior was the same for all strains during glucose cultivation, ranged from 4.85 ± 0.12 at 0 h to 3.88 ± 0.075 at 48 h of cultivations.

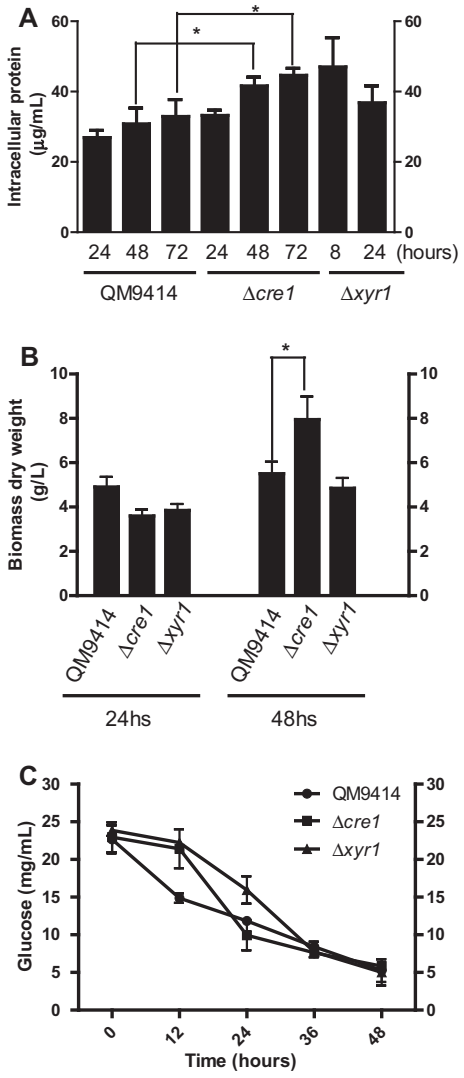


Fig. 2. Growth profiles of *T. reesei*, QM9414, Δ*cre1* and Δ*xyr1* on cellulose (A) and glucose (B) as carbon source, and glucose consuming during glucose experiments (C). The biomass on cellulose was estimated by intracellular protein content as described in material and methods. Error bars are represented from three biological replicates and asterisk (*) represent statistical difference at $P < 0.05$.

In order to evaluate the influence of XYR1 and CRE1 on the regulation of cellulolytic and xylanolytic gene expression in *T. reesei*, QM9414, Δ*xyr1*, and Δ*cre1* were cultured in the presence of cellulose, sophorose, and glucose for the indicated times. The expression of 22 genes encoding for 18 cellulases and 4 xylanases (Table 1) was assessed by RT-qPCR. In the presence of cellulose, gene expression of cellulases and xylanases in the Δ*xyr1* mutants (except for *xyn3* and *cel3e*) was generally reduced, whereas an increase was observed in Δ*cre1* compared to QM9414 (Fig. 3A). Hierarchical clustering of genes revealed the formation of 2 major clusters (Fig. 3A) in presence of cellulose, showing synergistic regulation of exoglucanases/endoglucanases/β-glucosidase but by different sets of genes, including xylanases. Interestingly, *xyn3*, *cel3b*, and *xyn4* did not form a cluster with other genes, suggesting a different form of regulation (Fig. 3A). In presence of sophorose three clusters were observed (Fig. 3B). The first contained the genes *cel3c*, *cel1b*, and *xyn3*, the second contained *cel45a*, *cel7a*, *cel61a*, and *xyn1*, and the third cluster contained the most analyzed genes. *cel3b* was the only gene not clustered with any gene (Fig. 3B). On the other hand, in presence of the repressor glucose, 18 of the 22 genes evaluated showed a considerably increased expression in the Δ*cre1* mutant compared with parental QM9414 (Fig. 3C) and 3 distinct clusters were observed using the same similarity criteria described above (Fig. 3C). Although both cellulose and sophorose are known as inducers of cellulases and xylanases, the observed differences in gene clustering demonstrate that the regulation of expression of these genes by XYR1 and CRE1 depends on the carbon source.

To complement the transcriptional profile of QM9414, Δ*cre1*, and Δ*xyr1*, we constructed a Venn diagram showing genes that exhibited a at least 2-fold $\log_2$ increase in expression (Fig. 4A), as well as a 2-fold $\log_2$ decrease in expression (Fig. 4B) in Δ*cre1*/QM9414 and Δ*xyr1*/QM9414 strains grown on cellulose, sophorose, or glucose.

When we compared Δ*cre1*/QM9414 (Fig. 4A), 7 genes (*cel6a*, *cel3a*, *cel7b*, *cel3c*, *cel3e*, *xyn2*, and *swo*) had at least a 2-fold increase in expression in all 3 conditions whereas 8 genes (*cel7a*, *cel1a*, *cel1b*, *cel3b*, *cel3d*, *cel45a*, *cel61a*, and *xyn1*) had increased expression in cellulose and glucose condition. The genes *cel74a*, *xyn3*, and *xyn4* had increased expression in sophorose and glucose conditions, whereas the expression of genes *cel5a*, *cel5b*, and *cel12a* increased in the presence of sophorose and cellulose. Only *cel61b*, encoding for a copper-dependent polysaccharide monooxygenase, had high expression in the presence of sophorose (Fig. 4A). Comparison between Δ*xyr1*/QM9414 (Fig. 4A) revealed that *xyn3* and *cel5b* had a >2-fold increase in all 3 conditions, whereas none of the genes had a >2-fold expression increase in both cellulose and sophorose conditions. Three genes had increased expression in cellulose (*cel45a*, *cel3e*, and *xyn2*) and sophorose (*cel1b*, *cel3b* and *cel3c*), separately. Interestingly, 10 genes (*cel7a*, *cel6a*, *cel7b*, *cel5a*, *cel12a*, *cel74a*, *xyn1*, *xyn4*, and *swo*), most of them encoding cellobiohydrolases and endoglucanases, had a >2-fold increase in expression in the presence of glucose, suggesting a strong relationship between XYR1 and CCR.

Analysis of genes with a <2-fold decrease in expression is shown in Fig. 4B. Comparison of Δ*cre1*/QM9414 shows that none of the genes had decreased expression in the 3 evaluated conditions. When we compared cellulose and glucose, only 1 gene, *cel61b* (encoding to an endoglucanase), was repressed. A larger number of repressed genes, 6 in total, were found in the presence of sophorose, followed by cellulose, and glucose with 3 genes each (Fig. 4B). In the Δ*xyr1*/QM9414 comparison, most of the repressed genes, 10 in total, were found in the presence of cellulose/sophorose. *Cel1a*, *cel3d*, and *cel61b* were repressed in all 3 conditions, whereas *cel1b*, *cel3b*, and *cel3c*, encoding β-glucosidases, were repressed in cellulose. Furthermore, *xyn1* and *cel3e* were repressed in the presence of sophorose and glucose, respectively (Fig. 4B).

1.3. Cellulolytic and xylanolytic activities

To determine whether increased/decreased cellulase/xylanase gene expression was also followed by changes in cellulase/xylanase activity in Δ*cre1* and Δ*xyr1*, compared with parental QM9414, we measured endoglucanase (CMCase), β-glucosidase, and xylanase activities (Fig. 5). In the presence of cellulose, a significant increase in CMCase activity (approximately 4-fold, $P < 0.001$) was observed after 72 h of incubation of Δ*cre1* compared with QM9414 (Fig. 5A), whereas in the presence of sophorose, although the increase was significant ($P < 0.001$), it was lower than that observed in cellulose (Fig. 5B). β-Glucosidases seem to be more influenced by CCR/CRE1 because β-glucosidase activities were approximately 8-fold and 6-fold higher ($P < 0.001$) in Δ*cre1* compared with parental QM9414 after growth in cellulose (Fig. 5A) and sophorose (Fig. 5B), respectively. Xylanase activities in Δ*cre1*, compared with parental QM9414, were approximately 1.5-fold higher ($P < 0.001$) in cellulose (Fig. 5A), and 6-fold higher after growth in the presence of sophorose (Fig. 5B). In the presence of

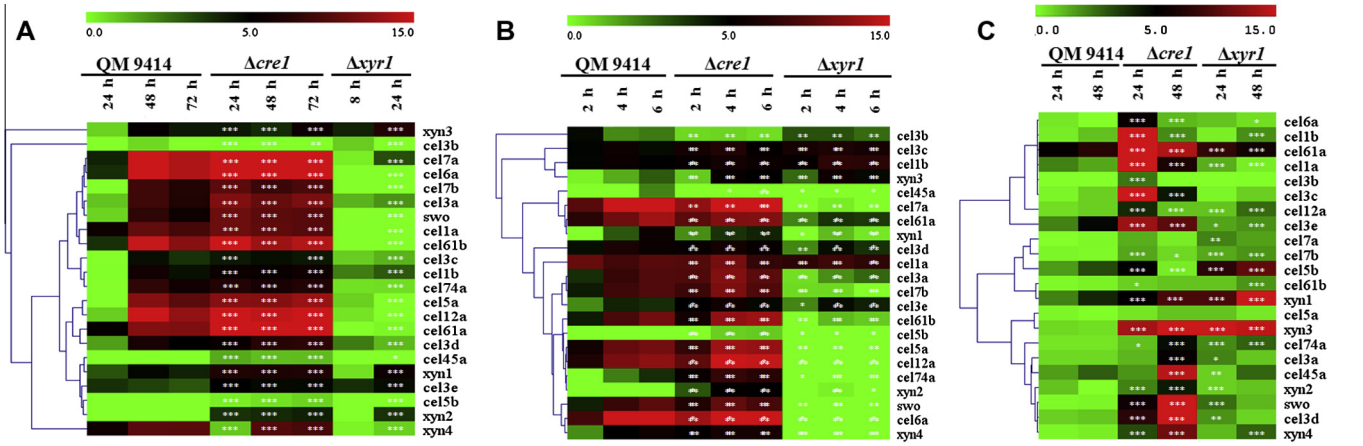


Fig. 3. Heat map of cellulolytic and xylanolytic gene expression patterns of *T. reesei*, QM9414 parental strain, Δ*cre1* and Δ*xyr1* grown in the presence of cellulose (A), sophorose (B) and glucose (C). The expression scale is represented in $\log_2$ and was calculated relative to transcription levels of QM9414 grown on glycerol at 24 h. All values were normalized using β-actin gene. The values represent the mean of 3 repetitions with standard deviation less than 0.5. Hierarchical clustering analysis was performed using MeV v.4.6.1, with Pearson's product-moment correlation coefficient as a similarity statistic and average linkage clustering as the linkage method. A Euclidian distance metric was used to measure the differences in expression. Levels in expression between the genes and groups were created based on the distance between the centroids of the groups. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

Table 1
Primers used in this study.

No.	Code	Protein ID	5′ → 3′
1	cel6a R	72567	TGT TCC ACC CGT TGT AGT TG
2	cel6a F		ACA AGA ATG CAT CGT CTC CG
3	actin R	44504	GGT ACC ACC AGA CAT GAC AAT GTT
4	actin F		TGA GAG CGG TGG TAT CCA CG
5	cel3a R	76672	TAG CTG AGA TCT CGT CGT C
6	cel3a F		CTG TAC ATC ACC TAC CCA TC
7	cel7a R	123989	GGT AGC CTT CTT GAC TGA GT
8	cel7a F		CCG AGC TTG GTA GTT ACT CTG
9	swo R	123992	GAG TGA ATG TCT TGA TGG
10	swo F		CCA AAC TAT ACG AGT AGC C
11	cel1a R	120749	AAT CAG CTC GTC AAA CAG CG
12	cel1a F		TTT GCC TGG TCG CTC ATG
13	cel61a R	73643	ACC GCT GCC ACC ACA CTG
14	cel61a F		GCG CCA CTG TTC CTG GAG
15	cel45a R	49976	TGG TCC AGA ATG CAC TCG
16	cel45a F		CAG CGA CGT CTA CAT TGG
17	cel7b R	122081	AGG TCT TGG AGG TGT CAA CG
18	cel7b F		CCC TCA ACA CTA GCC ACC AG
19	cel1b R	22197	TCC AAG TGC GAG TCA AAG TAG
20	cel1b F		CCA TCT ACA TCA CCG AGA ACG
21	cel3b R	121735	ATG TGG AGG TTG GAG AAC TTG
22	cel3b F		CCA GGA TAA CTT CAA CGA GGG
23	cel3c R	82227	ATC CCA ACC CCA TTC CTT TC
24	cel3c F		GCG TAC AAT GGC ATC AAT GG
25	cel5a R	120312	GTA CAG CCA AAG TCA AAA CCC
26	cel5a F		GCC ACT ACT ATC ACC ACT TCG
27	cel5b R	82616	TGG TAA CAG CAT TGG GAG TG
28	cel5b F		AAA GTA CCG TCC AAC ACC C
29	cel12a R	123232	CCC GAG TAC GTG ACA TGA TTC
30	cel12a F		AGA TTG CCA TTC CCC AGA AG
31	cel61b R	120961	CAC GCA CTG AGG ATA GTT CTG
32	cel61b F		ACT ATG TCT TCC GCC ATG AAC
33	cel3d R	46816	CTT GAA GGT AGC GTA AGA CAG G
34	cel3d F		AAC CCA GCA TAT CTC AAC TAC G
35	cel3e R	76227	TCG TGA GTC CAA AGT GAA CAG
36	cel3e F		ATG TCT GGA AGT GAG GTT GC
37	cel74a R	49081	TGA TGT CTT TCC AAG TTC CCC
38	cel74a F		GCC TTG TAT CTG ACC TAT TCC G
39	xyn1 R	74223	TCT GTC TTT TGG GCT TGG AG
40	xyn1 F		GGC CAA ATT ATC GTC AAC TGT C
41	xyn2 R	123818	TCT GCA CAG TAA CAG TTC CG
42	xyn2 F		TGT CAA CGA GCC TTC CAT C
43	xyn3 R	120229	GTT CAA AAC TCA CCC AAG CAC
44	xyn3 F		AAG TCA TCC GCA CCC ATG
45	xyn4 R	111849	GAC AGG TTG GCA AAA TGG TG
46	xyn4 F		TGT CAG CAA TTC GGG TCT TC

glucose, deletion of CRE1 was not sufficient to lead to high cellulase or xylanase production independent of induction (data not shown), indicating that CRE1 is not the only protein that participates in the process of CCR. Furthermore, compared with QM9414, the Δ*xyr1* mutant showed no significant changes in cellulase or xylanase activities in any of the conditions (Fig. 5A and B).

2. Discussion

Due to their ability to produce large amounts of cellulases and xylanases, *T. reesei* have been extensively studied for use in cellulosic ethanol production (Portnoy et al., 2011b). In this study, we demonstrated that cellulase and xylanase genes respond differently, depending on the available carbon source, to regulation by CRE1 and XYR1.

CRE1 has been shown to bind in the repressor domain of certain genes coding for hydrolases (Mach-Aigner et al., 2008), not only in *Trichoderma* and *Aspergillus* species, but also in *Humicola grisea*, *Botrytis cinerea*, and *Sclerotinea sclerotiorum*, indicating that this factor is well conserved among species. In the presence of glucose, we demonstrated that 82% of the genes evaluated in the mutant Δ*cre1* had an increased expression level compared to the parental QM9414, which is consistent with the literature (Ilmen et al., 1996; Mach et al., 1996). Our data also showed that some genes encoding cellulases, especially endoglucanases, β-glucosidases, and a xylanase (*xyn1*) in the parental QM9414 strain are more sensitive to derepression by depletion of glucose (Fig. 3C). This phenomenon has been previously observed, but only for cellobiohydrolases (*cel6A* and *cel7a*) and the endoglucanase *cel7b* (Ilmen et al., 1996). Although it was possible to detect cellulase and xylanase transcripts at low level, no enzymatic activity was detected in the presence of glucose.

CRE1 function is more complex than postulated and probably goes beyond CCR alone. Portnoy et al. (2011a) have reported that genes encoding cellulolytic and xylanolytic enzymes were not up-regulated in the Δ*cre1* mutant in the presence of glucose. Our data are consistent with this work. In fact, in the presence of glucose, these genes did not have a >2-fold increase in expression (Fig. 3A). However, the same result was not observed in the presence of cellulose and sophorose (Fig. 3A). These results indicate that, irrespective of the CCR, *Trichoderma* can recognize if glucose was released from a polysaccharide (e.g., cellulose or xylan) or is

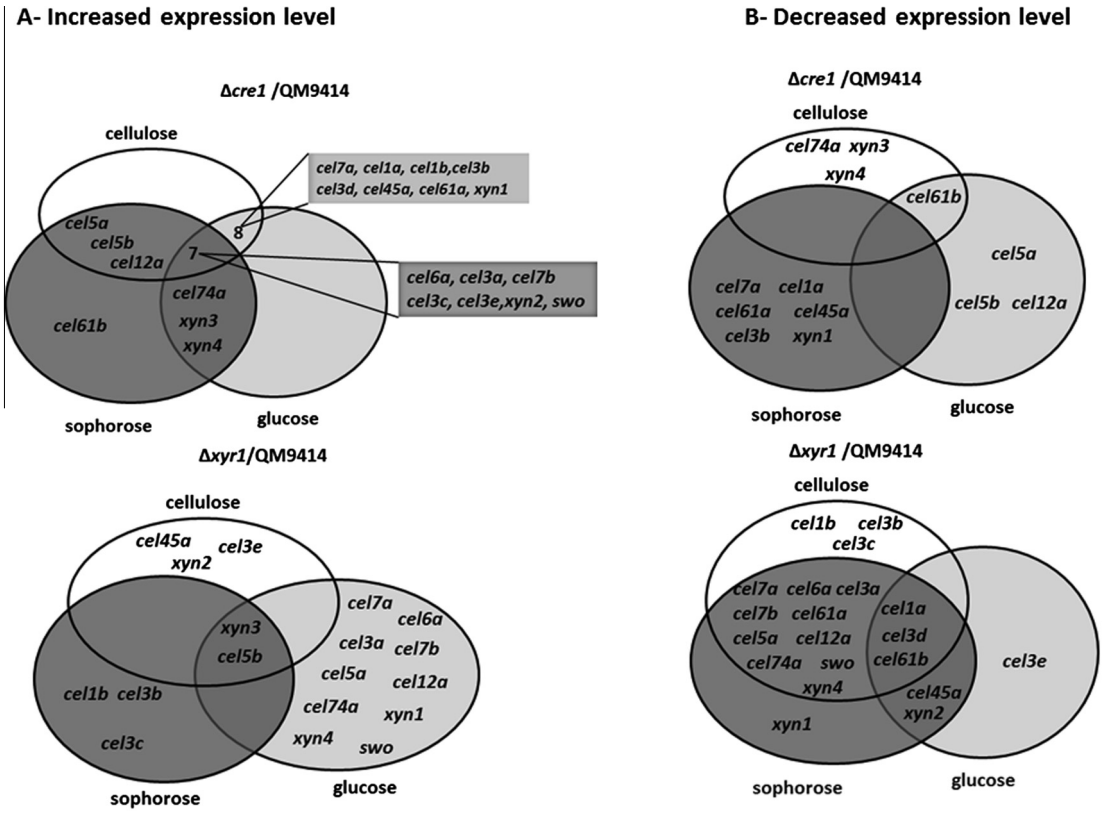


Fig. 4. Venn diagram of differentially expressed cellulolytic and xylanolytic genes in *T. reesei*, QM9414 parental strain, Δ*cre1* and Δ*xyr1*, grown in the presence of cellulose, sophorose or glucose. Overlap among genes that exhibited a statistically significant increase (A) and decrease (B) in expression levels in Δ*cre1*/QM9414 and Δ*xyr1*/QM9414, respectively as indicated. For each condition, we considered the high expression independent of the time and a fold change down ⩽ 2-fold ⩾ up.

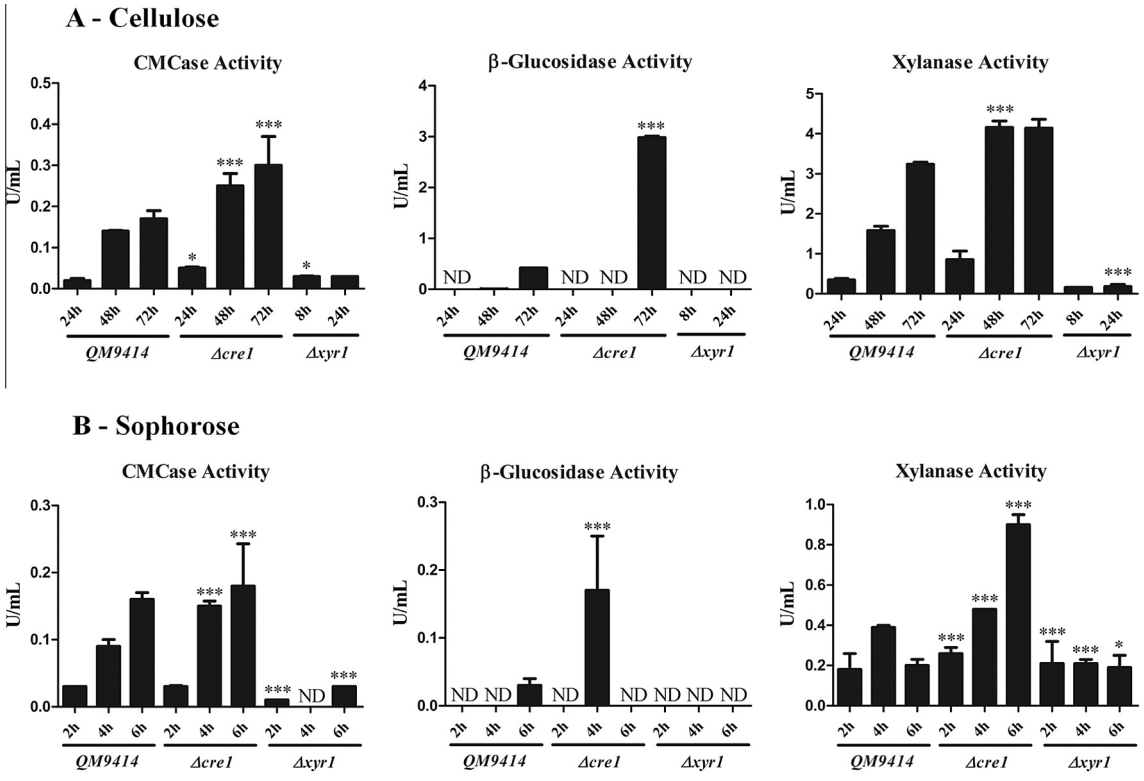


Fig. 5. Cellulase (CMCase), β-glucosidase, and xylanase activities from culture supernatant of *T. reesei*, QM9414 parental strain, Δ*cre1* and Δ*xyr1* grown in the presence of cellulose (A) or sophorose (B) for different times as indicated. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

present as a free compound, and accordingly modify gene regulation. Furthermore, compared with QM9414, various genes encoding cellulolytic and xylanolytic enzymes were down-regulated in the Δ*cre1* mutant, even in presence of glucose, suggesting that CRE1 can act positively on these genes in a direct or indirect manner. Our data is supported by the fact that genes related to carbohydrate metabolism and TFs were found to be positively influenced by CRE1 (Portnoy et al., 2011a). Furthermore, there is evidence that CCR in *T. reesei* acts preferentially at the entry of substrates into the cell and can influence hexose permeases expression (Ilmen et al., 1996; Portnoy et al., 2011a).

In our studies with the Δ*xyr1* mutant, we found a general reduction in the expression of cellulase and xylanase genes compared to the parental QM9414 strain, when cultured in the presence of cellulose or sophorose (Fig. 3). These findings corroborate the data from Furukawa et al. (2009) that also observed a significant reduction in the expression of cellulase and xylanase genes in the PC-3–7 mutant in the presence of L-sorbose or sophorose. Unlike our observations in cellulose and sophorose, approximately 60% of the analyzed genes had increased expression in the Δ*xyr1* mutant compared to QM9414, in the presence of glucose at 24 and 48 h (Fig. 3C). In addition to XYR1 and CRE1, 2 other TFs, ACE1 and ACE2, are known to regulate the expression of cellulase and xylanase genes by repression or activation, respectively (Rauscher et al., 2006). A plausible explanation for this observation may is be a partial constitutive expression of ACE2. Although *xyr1* is subject to CCR, *ace2* is not, supporting our results (Portnoy et al., 2011a).

In addition to classical cellulases, new types of cellulases have recently been described in *T. reesei*, including swollenin (SWO) (Saloheimo et al., 2002). SWO has an expansin-like domain in the N-terminal region, which is normally found in plants during their growth phase. It has been suggested that SWO makes cellulose fibers more accessible for attack by cellulases (Saloheimo et al., 2002). SWO is positively regulated when either cellulose or sophorose is present and negatively regulated in the presence of glucose (Verbeke et al., 2009). Our results suggest that *swo* is a target of regulation by CRE1 and XYR1, independent of the carbon source.

It is known that hydrolases act synergistically in the degradation of biopolymeric substrates such as cellulose and xylan. It has been reported that *cel3a* and *bxl1* are regulated by XYR1 in the presence of sophorose, but *cel1a* is not (Stricker et al., 2008). In this study, we found that *cel3a* and *cel1a* were differentially regulated and that this regulation was dependent not only on XYR1 and CRE1, but also on the carbon source. However, β-glucosidases also require a specific TF (BglR) for its expression (Nitta et al., 2012).

The genes encoding xylanases also appeared to be dependent on XYR1 and the negative regulator, CRE1 (Fig. 4A and B). D-Xylose can induce most xylanases, with the exception of XYN3, whereas L-arabinose induces the xylanases XYN1, XYN4, and XYN5 (Herold et al., 2013). Whatever the form of regulation of gene expression of cellulase and xylanase, our data suggest that this regulation may occur directly or indirectly by a factor or a set of transcription factors and that this process is dependent of carbon source.

In conclusion, our work describes the regulation of cellulolytic and xylanolytic enzymes formation, highlighting the interaction of carbon sources and CRE1/XYR1 in the control of this process. Thus, the data presented here could form the basis for future strategies in the construction of mutants to be used in the deconstruction of plant biomass for the production of second-generation biofuels. The knowledge of how different genes of cellulase and xylanase are regulated may contribute to the development of various enzyme cocktails to the hydrolysis of biomass. However, many studies are still needed to elucidate the entire network involved in the regulation of the formation of cellulases and xylanases.

3. Experimental procedures

3.1. Strains and culture conditions

T. reesei strains QM9414 (ATCC 26921), Δ*cre1* (Portnoy et al., 2011a), and Δ*xyr1* (Stricker et al., 2006) were obtained from the Institute for Chemical Engineering (Vienna University of Technology, Research Area Gene Technology and Applied Biochemistry, Vienna, Austria). The strains were maintained on MEX medium (3% malt extract and agar 2% w/v). For all experiments, QM9414 was used as the parental strain. Approximately 10^8 conidia/l (final concentration) was used as inoculum. Strains were grown in 1-L Erlenmeyer flasks at 28 °C on a rotary shaker (200 rpm) in 200 ml of Mandels and Andreotti medium (MA) (Schmoll et al., 2009) supplemented with 0.1% (w/v) peptone to induce germination and 1% (w/v) of microcrystalline cellulose (CA435236 Sigma), 2% (w/v) of glucose or 1 mM of sophorose (CA35208, Serva), as indicated in the respective experiments. For experiments with sophorose, mycelia were precultured in 1-L Erlenmeyer flasks for 24 h in the same conditions as above, except that 1% (w/v) of glycerol was used as the carbon source. Pre-grown mycelia were washed, and equal amounts were resuspended in the MA medium without peptone containing 1 mM of sophorose. The flasks were incubated at 200 rpm in an orbital shaker at 28 °C for 24, 48, and 72 h for cellulose experiments; 24 and 48 h for glucose experiments; and 2, 4, and 6 h for sophorose experiments. Δ*xyr1* strains, cultured in the presence of cellulose, were previously grown in glycerol for 24 h and then transferred to a medium containing cellulose for 8–24 h according to Stricker et al. (2006). Experiments were conducted in triplicate for each sample. After induction, mycelia were collected by filtration, frozen and stored at −80 °C, and used for total RNA extraction. Culture supernatants were used for cellulase and xylanase activities assays.

3.2. RNA extraction and transcript analysis by quantitative PCR (RT-qPCR)

Total RNA was isolated from mycelia by grinding with a mortar and pestle under liquid nitrogen, followed by extraction using TRIZOL reagent (Invitrogen, USA) according to the manufacturer's instructions. RNA was quantified by spectrophotometric DO 260/280, and integrity was checked using an Agilent 2100 Bioanalyzer and also by gel electrophoresis in 1% agarose. RT-qPCR was used to evaluate gene expression. Total RNA (1 μg) from each pooled sample was first digested with DNase I (Fermentas) to remove genomic DNA. cDNA synthesis was carried out using the Maxima™ First Strand cDNA Synthesis kit (Fermentas) according to manufacturer's instructions. Synthesized cDNA was diluted 1:50 and used as a template for real-time PCR. Reactions were performed in the Bio-Rad CFX96™ by using SsoFast™ EvaGreen® Supermix (Bio-Rad) for detection according to the manufacturer's instructions. Each reaction (20 μl) contained 10 μl of SsoFastTM EvaGreen® Supermix (Bio-Rad), forward and reverse primers (500 nm each; Table 1), cDNA template, and nuclease-free water. PCR cycling conditions: 10 min at 95 °C (1 cycle), 10 s at 95 °C followed by 30 s at 60 °C (40 cycles), and a melting curve of 60–95 °C with an increment of 0.5 °C for 10 s to test for primer dimers and nonspecific amplification. β-Actin transcript was used as an internal references to normalize the amount of total RNA present in each reaction (Verbeke et al., 2009). Expression levels of genes were calculated from the threshold cycle according to the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001) relative to transcription levels of QM9414 grown in non induced condition glycerol for 24 h (Verbeke et al., 2009). The experiment was repeated 3 times for each sample. The results were analyzed using MeV v.4.6.1. software, and Heat

Maps were constructed to access the differences in expression between the strains in each indicated condition.

3.3. Activity assay

Endoglucanase (CMCase) activity was performed in microplates (96-well PCR plate) following the protocol described by Xiao et al. (2005) with some modifications: 30 μl of 1% carboxymethylcellulose (CMC), previously prepared in sodium acetate buffer at pH 4.8, was added to each well with 30 μl of enzyme. After a 30-min incubation at 50 °C, 60 μl of DNS (3,5-dinitrosalicylic acid) was added, followed by heating at 95 °C for 5 min to allow color development. Then, the samples were transferred to a flat-bottomed microplate and read at an absorbance of 540 nm. One enzyme unit was defined as the amount of enzyme capable of releasing 1 μmol of reducing sugar per minute.

β-Glucosiadase activity was also carried out in a microplate assay format described earlier (Lopes et al., 2012). Protein concentrations of the samples were determined using the Bradford micro assay, with bovine serum albumin (Sigma Aldrich, St. Louis, USA) used as a standard.

Xylanase activity was determined by mixing 25 μl of enzyme solution with 50 μl of xylan from beechwood (1.0 mg ml^{-1}) in 100 mM sodium acetate buffer (pH 5.0) for 30 min at 50 °C. After incubation, 75 μl of DNS was added and heated at 95 °C for 5 min for color development. After this, 100 μl of the solution was transferred to a microliter plate, and the absorbance was read using a spectrophotometer at 540 nm. One unit of enzyme was defined as the amount of enzyme capable of releasing 1 μmol of reducing sugar per minute.

3.4. In silico analysis of putative binding sites for XYR1 and CRE1 in promoter regions of cellulolytic and xylanolytic genes

The complete genome sequence of *T. reesei* (Martinez et al., 2008) were downloaded from Joint Genome Institute *T. reesei* database (http://genome.jgi-psf.org/Trire2/Trire2.home.html) and was used to obtain the 5′-upstream regions (1.5-kbp) for each gene. The occurrence of the motifs GGCWWW-5′-3 "(Furukawa et al., 2009) and 5′-SYGGRG-3" (Portnoy et al., 2011a) of XYR1 and CRE1 transcription factors, were determined in both strands of the 22 genes analyzed in this study, as well as for the 9117 genes in *T. reesei* genome, using a Perl script. The frequency of each motif was calculated as the total number of base pairs of the analyzed upstream sequences per total number of motifs, as described by Furukawa et al. (2009). Potential binding motifs were identified in the promoter region 1.5 kbp upstream of predicted translational start sites of selected genes using the algorithm Scope version 2.1.0 (http://genie.dartmouth.edu/scope/) (Carlson et al., 2007).

3.5. Analytical methods

Biomass concentrations were determined by gravimetric analysis for glucose culture. In each time point, mycelia was filtrated on filter paper and incubated at 70 °C for 3 h and then weighed. For cellulose culture, biomass concentrations were indirectly measured by the amount of intracellular protein quantified by the Quick Start Bradford protein assay kit (Bio-Rad) with bovine serum albumin (BSA) as a standard. The glucose concentration during the cultivation on glucose was assessed by enzymatic reaction using Glucose Liquiform (LABTEST) kit. The pH measurements were determined directly using pH-metro. The results are means of the three independent experiments cultivations for each carbon source.

3.6. Statistical analysis

All tests of statistical significance, both in the analysis of gene expression and enzyme activities, were performed with one-way ANOVA (nonparametric) followed by the Bonferroni test (comparing all pairs of columns) available in the Prism software v. 5.0. The parental QM9414 strain was used as a standard for comparison with the 2 mutants under all conditions and times. Cultivation of mutants at 24, 48, and 72 h (cellulose); 2, 4, and 6 h (sophorose); and 24 and 48 h (glucose) was compared with that of the parental strains. As an exception, cultivation of mutant Δ*xyr1* at 48 h (cellulose) was compared with the parental strain at 24 h (cellulose) because its cultivation time was different.

Acknowledgements

This work was supported by The State of São Paulo Research Foundation (FAPESP) (proc. 2010/15683-8).

References

Cabero, K., Pozzo, T., Liden, G., Karlsson, E.N., 2012. A cellulolytic Hypocrea strain isolated from South American brave straw produces a modular xylanase. Carbohydr. Res. 356, 215–223.
Carlson, J.M., Chakravarty, A., DeZiel, C.E., Gross, R.H., 2007. SCOPE: a web server for practical de novo motif discovery. Nucleic Acids Res. 35, W259–264.
Dowzer, C.E.A., Kelly, J.M., 1991. Analysis of the CreA gene, a regulator of carbon catabolite repression in *Aspergillus nidulans*. Mol. Cell. Biol. 11, 5701–5709.
Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J., Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C., Olivares, H.A., Teunissen, P.J., Yao, J., Ward, M., 2003. Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. J. Biol. Chem. 278, 31988–31997.
Furukawa, T., Shida, Y., Kitagami, N., Mori, K., Kato, M., Kobayashi, T., Okada, H., Ogasawara, W., Morikawa, Y., 2009. Identification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in *Trichoderma reesei*. Fungal Genet. Biol. 46, 564–574.
Herold, S., Bischof, R., Metz, B., Seiboth, B., Kubicek, C.P., 2013. Xylanase gene transcription in *Trichoderma reesei* is triggered by different inducers representing different hemicellulosic pentose polymers. Eukaryot. Cell 12, 390–398.
Ilmen, M., Saloheimo, A., Onnela, M.L., Penttila, M.E., 1997. Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*. Appl. Environ. Microbiol. 63, 1298–1306.
Ilmen, M., Thrane, C., Penttila, M., 1996. The glucose repressor gene cre1 of Trichoderma: isolation and expression of a full-length and a truncated mutant form. Mol. Gen. Genet. 251, 451–460.
Kubicek, C.P., 2012. Systems biological approaches towards understanding cellulase production by *Trichoderma reesei*. J. Biotechnol. 163 (2), 133–142.
Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) method. Methods 25, 402–408.
Lopes, F.A.C., Steindorff, A.S., Geraldine, A.M., Brandao, R.S., Monteiro, V.N., Lobo, M., Coelho, A.S.G., Ulhoa, C.J., Silva, R.N., 2012. Biochemical and metabolic profiles of Trichoderma strains isolated from common bean crops in the Brazilian Cerrado, and potential antagonism against *Sclerotinia sclerotiorum*. Fungal Biol. 116, 815–824.
Mach, R.L., Strauss, J., Zeilinger, S., Schindler, M., Kubicek, C.P., 1996. Carbon catabolite repression of xylanase I (xyn1) gene expression in *Trichoderma reesei*. Mol. Microbiol. 21, 1273–1281.
Mach-Aigner, A.R., Pucher, M.E., Steiger, M.G., Bauer, G.E., Preis, S.J., Mach, R.L., 2008. Transcriptional regulation of xyr1, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*. Appl. Environ. Microbiol. 74, 6554–6562.
Martinez, D., Berka, R.M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S.E., Chapman, J., Chertkov, O., Coutinho, P.M., Cullen, D., Danchin, E.G., Grigoriev, I.V., Harris, P., Jackson, M., Kubicek, C.P., Han, C.S., Ho, I., Larrondo, L.F., de Leon, A.L., Magnuson, J.K., Merino, S., Misra, M., Nelson, B., Putnam, N., Robbertse, B., Salamov, A.A., Schmoll, M., Terry, A., Thayer, N., Westerholm-Parvinen, A., Schoch, C.L., Yao, J., Barabote, R., Nelson, M.A., Detter, C., Bruce, D., Kuske, C.R., Xie, G., Richardson, P., Rokhsar, D.S., Lucas, S.M., Rubin, E.M., Dunn-Coleman, N., Ward, M., Brettin, T.S., 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. Hypocrea jecorina). Nat. Biotechnol. 26, 553–560.
Nakari-Setala, T., Paloheimo, M., Kallio, J., Vehmaanpera, J., Penttila, M., Saloheimo, M., 2009. Genetic modification of carbon catabolite repression in *Trichoderma reesei* for improved protein production. Appl. Environ. Microbiol. 75, 4853–4860.

Nitta, M., Furukawa, T., Shida, Y., Mori, K., Kuhara, S., Morikawa, Y., Ogasawara, W., 2012. A new Zn(II)(2)Cys(6)-type transcription factor BglR regulates beta-glucosidase expression in *Trichoderma reesei*. Fungal Genet. Biol. 49, 388–397.

Portnoy, T., Margeot, A., Linke, R., Atanasova, L., Fekete, E., Sandor, E., Hartl, L., Karaffa, L., Druzhinina, I.S., Seiboth, B., Le Crom, S., Kubicek, C.P., 2011a. The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei*: a master regulator of carbon assimilation. BMC Genomics 12, 269.

Portnoy, T., Margeot, A., Seidl-Seiboth, V., Le Crom, S., Ben Chaabane, F., Linke, R., Seiboth, B., Kubicek, C.P., 2011b. Differential regulation of the cellulase transcription factors XYR1, ACE2, and ACE1 in *Trichoderma reesei* strains producing high and low levels of cellulase. Eukaryot. Cell 10, 262–271.

Rauscher, R., Wurleitner, E., Wacenovsky, C., Aro, N., Stricker, A.R., Zeilinger, S., Kubicek, C.P., Penttila, M., Mach, R.L., 2006. Transcriptional regulation of xyn1, encoding xylanase I, in *Hypocrea jecorina*. Eukaryot. Cell 5, 447–456.

Saloheimo, A., Henrissat, B., Hoffren, A.M., Teleman, O., Penttila, M., 1994. A novel, small endoglucanase gene, egl5, from *Trichoderma reesei* isolated by expression in yeast. Mol. Microbiol. 13, 219–228.

Saloheimo, M., Pakula, T.M., 2012. The cargo and the transport system: secreted proteins and protein secretion in *Trichoderma reesei* (*Hypocrea jecorina*). Microbiology 158, 46–57.

Saloheimo, M., Paloheimo, M., Hakola, S., Pere, J., Swanson, B., Nyyssonen, E., Bhatia, A., Ward, M., Penttila, M., 2002. Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. Eur. J. Biochem. 269, 4202–4211.

Schmoll, M., Kubicek, C.P., 2003. Regulation of Trichoderma cellulase formation: lessons in molecular biology from an industrial fungus. A review. Acta Microbiol. Immunol. Hung. 50, 125–145.

Schmoll, M., Schuster, A., Silva, R.D., Kubicek, C.P., 2009. The G-Alpha Protein GNA3 of *Hypocrea jecorina* (Anamorph *Trichoderma reesei*) regulates cellulase gene expression in the presence of light. Eukaryot. Cell 8, 410–420.

Strauss, J., Mach, R.L., Zeilinger, S., Hartler, G., Stoffler, G., Wolschek, M., Kubicek, C.P., 1995. Cre1, the carbon catabolite repressor protein from *Trichoderma reesei*. FEBS Lett. 376, 103–107.

Stricker, A.R., Grosstessner-Hain, K., Wurleitner, E., Mach, R.L., 2006. Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. Eukaryot. Cell 5, 2128–2137.

Stricker, A.R., Mach, R.L., de Graaff, L.H., 2008. Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*). Appl. Microbiol. Biotechnol. 78, 211–220.

Verbeke, J., Coutinho, P., Mathis, H., Quenot, A., Record, E., Asther, M., Heiss-Blanquet, S., 2009. Transcriptional profiling of cellulase and expansin-related genes in a hypercellulolytic *Trichoderma reesei*. Biotechnol. Lett. 31, 1399–1405.

Xiao, Z., Storms, R., Tsang, A., 2005. Microplate-based carboxymethylcellulose assay for endoglucanase activity. Anal. Biochem. 342, 176–178.

OPEN ACCESS Freely available online

PLOS ONE

Deciphering the *Cis*-Regulatory Elements for XYR1 and CRE1 Regulators in *Trichoderma reesei*

Rafael Silva-Rocha[1], Lilian dos Santos Castro[1], Amanda Cristina Campos Antoniêto[1], María-Eugenia Guazzaroni[2], Gabriela Felix Persinoti[1], Roberto Nascimento Silva[1]*

1 Department of Biochemistry and Immunology, FMRP - University of São Paulo, Ribeirao Preto, São Paulo, Brazil, **2** Department of Biochemistry, FFCLRP - University of São Paulo, Ribeirao Preto, São Paulo, Brazil

Abstract

In this work, we report the *in silico* identification of the *cis*-regulatory elements for XYR1 and CRE1 proteins in the filamentous fungus *Trichoderma reesei*, two regulators that play a central role in the expression of cellulase genes. Using four datasets of condition-dependent genes from RNA-seq and RT-qPCR experiments, we performed unsupervised motif discovery and found two short motifs resembling the proposed binding consensus for XYR1 and CRE1. Using these motifs, we analysed the presence and arrangement of putative *cis*-regulatory elements recognized by both regulators and found that shortly spaced sites were more associated with XYR1- and CRE1-dependent promoters than single, high-score sites. Furthermore, the approach used here allowed the identification of the previously reported XYR1-binding sites from *cel7a* and *xyn1* promoters, and we also mapped the potential target sequence for this regulator at the *cel6a* promoter that has been suggested but not identified previously. Additionally, seven other promoters (for *cel7b*, *cel61a*, *cel61b*, *cel3c*, *cel3d*, *xyn3* and *swo* genes) presented a putative XYR1-binding site, and strong sites for CRE1 were found at the *xyr1* and *cel7b* promoters. Using the *cis*-regulatory architectures nearly defined for XYR1 and CRE1, we performed genome-wide identification of potential targets for direct regulation by both proteins and important differences on their functional regulons were elucidated. Finally, we performed binding site mapping on the promoters of differentially expressed genes found in *T. reesei* mutant strains lacking *xyr1* or *cre1* and found that indirect regulation plays a key role on their signalling pathways. Taken together, the data provided here sheds new light on the mechanisms for signal integration mediated by XYR1 and CRE1 at cellulase promoters.

Citation: Silva-Rocha R, Castro LdS, Antoniêto ACC, Guazzaroni M-E, Persinoti GF, et al. (2014) Deciphering the *Cis*-Regulatory Elements for XYR1 and CRE1 Regulators in *Trichoderma reesei*. PLoS ONE 9(6): e99366. doi:10.1371/journal.pone.0099366

Editor: Joel S. Bader, Johns Hopkins University, United States of America

Received February 13, 2014; **Accepted** May 13, 2014; **Published** June 18, 2014

Copyright: © 2014 Silva-Rocha et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by The State of São Paulo Research Foundation (FAPESP, 2010/15683-8). RS-R holds a FAPESP post-doctoral Fellowship (FAPESP, 2013/04125-2). MEG benefits from a Young Talent Fellowship of the National Counsel of Technological and Scientific Development (CNPq, Ref. Number 370630/2013-0). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rsilva@fmrp.usp.br

Introduction

Trichoderma reesei is a filamentous fungus extremely relevant to biotechnology due to its remarkable capability to produce a wide number of cellulolytic enzymes [1,2]. This mesophilic organism is endowed with a tremendous repertoire of hydrolytic enzymes related to the deconstruction of lignocellulosic biomass from plants that are of high importance for biotech processes such as paper industry or fuel production [3,4,5]. Due to its elevated biotechnological potential, *T. reesei* has been extensively studied in the past decades as a model of cellulases and hemicellulases producing organism [6,7,8,9]. This organism is endowed with different classes of biomass-related hydrolytic enzymes (here, collectively referred as cellulases), and special attention has been placed on enzymes such as endoglucanases (Cel7b, Cel5a, Cel12a, Cel61a and Cel45a), sGH61 polysaccharide monoxygenase (PMOs, Cel61a and Cel61b), exoglucanases (Cel7a and Cel6a), β-glucosidases (Cel3a and Cel1a), endo-β-1,4-xylanases (XYN1 and XYN2) and β-xylosidase (BXL1). From this particular set of enzymes, Cel7a, Cel6a, Cel7b and Cel5a are the most abundantly produced under inducing conditions such as growth in the presence of cellulose or sophorose (a glucose disaccharide produced during cellulose degradation; [6]).

In order to allow the ration engineering of new strains of *T. reesei* with enhance enzyme production levels, a great interest has been to elucidate the molecular mechanisms operating at the transcriptional network that controls the expression of cellulase genes in response to the cognate environmental conditions [6,10,11,12,13,14]. These efforts have allowed the identification of many regulatory proteins and signalling pathways that are responsible for the coordination of cellulase expression in this fungus [15,16,17,18,19,20]. For some of the enzymes mentioned above, at least three mechanistic steps take place at the promoter regions: chromatin reorganization, de-repression and induction [2,6]. Chromatin reorganization is related to the dynamic positioning of nucleosomes in response to environmental or physiologic signals [21]. In Eukaryotes, nucleosomes are important players in gene regulation since their binding to DNA segments is able to lock the chromatin in a blocked state, where transcriptional factors (TFs) cannot interact with the *cis*-regulatory elements located in the region occupied by the nucleosome [22,23,24]. In fact, this process has been shown to regulate the basal expression

levels of cellulase promoters in *T. reesei* [25,26]. As a second step, de-repression is related to the increase in the basal promoter expression level in response to the removal of a repressive signal [6,10,27,28]. In the case of cellulases, carbon catabolite repression (CCR) is mediated by alternative carbon source of easy degradation, such as glucose [10,20,27,28]. In this sense, the modulation of the promoter activity during CCR has been postulated to occur mainly through changing of the chromatin state of the target promoter [25,26]. Finally, the third mechanism involves the induction of high promoter activities in response to some signals (in the case of cellulases, the enzyme substrates) that is mediated by general and specific TFs [6,13,17,29].

The investigation of the regulatory network for cellulase expression in *T. reesei* has allowed the identification of several TFs related to each of the mechanisms described above. For instance, the XYR1 (*xylanase regulator 1*) transcriptional factor is the main positive regulator of cellulase expression in *T. reesei* [11,19], and homologues of this protein performs the same role in other cellulase producing organisms such as *Aspergillus niger* and *Neurospora crassa* [30,31]. XYR1 is a zinc binuclear cluster protein that is able to bind to several cellulase promoters and is virtually essential for full expression of these genes during growth under inducing conditions [11,19]. Moreover, XYR1 production is also regulated at the transcriptional level by the carbon catabolite repressor CRE1 protein [17] and is repressed by the specific transcriptional factor ACE1 [32,33]. CRE1 is a Cys2His2 type transcriptional factor that is responsible to mediate glucose dependent CCR at several cellulase promoters [10,20]. This regulator is an homologue of the CCR protein Mig1 from *Saccharomyces cerevisiae* [34] and affects chromatin organization at target promoters in response to glucose [25,26]. In turn, ACE1 contains three Cys2His2-type zinc fingers and also regulate other cellulase genes such as Cel7a and XYN1 apparently through the interference with the binding of XYR1 at targets promoters [29,32]. In addition to the above cited proteins, additional regulators required for the expression of cellulase genes are the HAP2/3/5 complex, which is necessary to generate an open chromatin structure that is essential for full promoter activation, [26,29,35,36] and the zinc binuclear cluster protein ACE2, which is a specific cellulase activator that only occurs in *T. reesei* [17,36,37].

Current available experimental data related to cellulase regulation in *T. reesei* clearly evidence sophisticate interplay between the characterized and unknown TFs at the target promoters to provide fine-tuning of enzyme production levels in this organism [6,38]. Yet, the mechanisms by which the regulators at stake interact with the target *cis*-regulatory elements at each particular promoter are only beginning to be elucidated [11,18,36,39,40]. For instance, the consensus binding sequences of the two main cellulase regulators XYR1 and CRE1 (5′-GGCWWW-3′ and 5′-SYGGRG-3′, respectively) have been proposed on the basis of the comparison with homologous regulators form other organisms [41] or cannot be used to distinguish between genes regulated or not by these regulators [11]. Thus, quantitative information on the *cis*-regulatory elements associated with the interaction of XYR1 and CRE1 with their target promoters is crucial to improve engineering attempts to construct new cellulase responsive promoters [12,42,43] and to understand the role of these regulators in *T. reesei* at the global scale [10,11]. In this work, we analysed four sets of co-regulated genes identified using RNA deep sequencing (RNA-seq) and Real Time quantitative PCR (RT-qPCR) from *T. reesei* cells grown in the presence of inducers (cellulose and sophorose) and a repressor (glucose) carbon sources. Separated analysis of the different datasets allowed the identification of new DNA motifs that were associated with the regulation by XYR1 and CRE1 in this organism. The nearly revealed motifs were used to generate, by the first time, Position Weight Matrixes (PWMs) representing the putative *cis*-regulatory elements recognized by both regulators. Furthermore, these PWMs were used to inspect the architecture of target promoters in co-expressed genes at the genomic scale, allowing the identification of the functional regulons of XYR1 and CRE1 in *T. reesei*. The PWMs generated here were able to successfully identify two previously characterized *cis*-regulatory elements at the *cel7a* and *xyn1* promoters [25,29], as well as to find a new regulatory region at the *cel6a* promoter that agrees with previous *in vivo* data [26]. Taken together, the results provided here add valuable information on the regulatory scope of XYR1 and CRE1 in *T. reesei* at the genomic scape, revealing some important differences with other filamentous fungus [8,30,44] and providing new clues on the molecular mechanisms of promoter recognition by these regulators.

Materials and Methods

Experimental Datasets

The analysis of *cis*-regulatory elements was performed using four groups of co-regulated genes identified using different experimental setups. Three groups represent genes encoding TFs that were identified using RNA-seq [45] from cells growing in cellulose (7

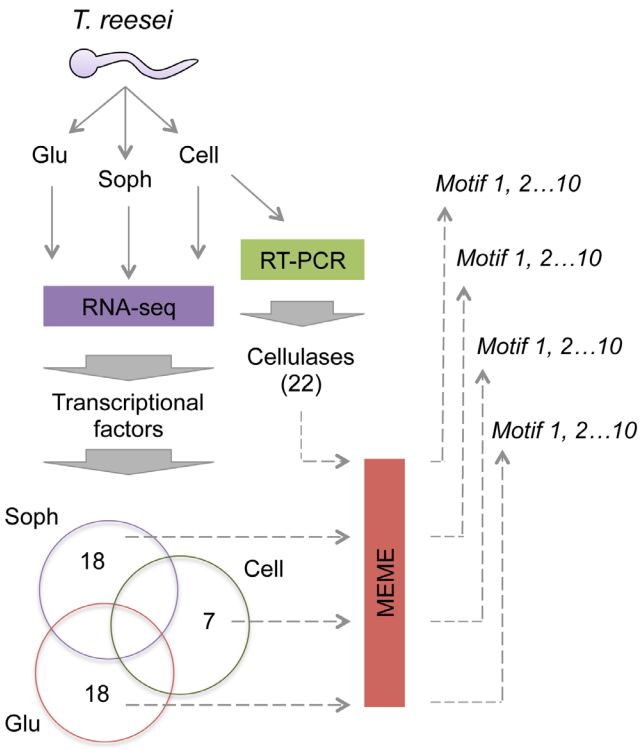


Figure 1. Schematic representation of the approach used for motif discovery in *T. reesei*. Three sets of co-regulated genes were retrieved from RNA-seq experiments of *T. reesei* cells grown on cellulose, sophorose or glucose [45]. Only up-regulated genes encoding for putative TFs were selected. A fourth set of genes are 22 cellulases analysed through RT-PCR experiments [46]. A 1.5 kb promoter region of each gene from the four groups was retrieved from the complete genome sequence [49] and used for motif discovery using MEME. From the resulting identified motifs, those sharing similarities with the reported binding consensus for XYR1 and CRE1 were selected for further analysis.
doi:10.1371/journal.pone.0099366.g001

genes), sophorose (18 genes) or glucose (18 genes) as sole carbon sources and that were specifically up regulated. The lists of genes from these groups are given in **Tables S1 to S3** from Supporting Information. The fourth group is represented by 22 cellulases-encoding genes whose promoters are regulated by the XYR1 regulator [46]. For each of the 65 analysed genes, a 1.5 kb DNA sequence immediately upstream of ATG codons was retrieved from the complete genome sequence of *T. reesei* available at the JGI homepage (http://genome.jgi-psf.org/Trire2/Trire2.home.html) using *ad hoc Perl* scripts. These sequences are expected to contain the *cis*-regulatory elements for the different TFs acting on each gene. The four groups of fasta sequences where then used to identify conserved DNA motifs as described below. The overall approach used in this work is represented in **Fig. 1**.

De novo Motif Discovery

In order to identify new *cis*-regulatory elements in the four regulons assayed, we analysed the promoter sequences using the MEME tool [47]. For MEME analysis, we set the parameters to search for short DNA motifs (6 to 10 nt in length) expected to occur zero or one time per sequence at forward or reverse strand, allowing a maximal of ten different motifs to be reported by the program. From the resulting outputs, motifs displaying similarities with the DNA binding sites of XYR1 and CRE1 from *T. reesei* were selected. In cases where similar DNA consensus for the regulators appeared on different motif outputs, the aligned sequences were merged to create a single motif dataset. The resulting datasets were used to construct Position Weight Matrixes (PWM) by extracting the information content of the sequence alignments as described previously [48]. For XYR1, a PWM representing the first 8-nt of the motif (**Fig. 2B**) was used, while for CRE1, a PWM containing the complete 10-nt motif (**Fig. 2D**) was constructed. The two resulting PWMs (named PWM_{XYR1} for XYR1 and PWM_{CRE1} for CRE1 regulators) were used for further analysis.

Genome-wide Analysis of *Cis*-regulatory Elements

The two PWMs generated using the motifs discovered with MEME were used to analyse the promoters of all annotated genes in the genome of *T. reesei* [49]. For this, promoters of 1.5 kb in length for the ~9,000 genes of *T. reesei* were analysed to identify the best motif for both PWMs on each promoter. Next, the same dataset was re-analysed to identify multiple motifs per promoter with a score above a specific threshold, which was set to 6.2 for PWM_{XYR1} and 8.0 for PWM_{CRE1}. The resulting identified sites were then analysed to identify adjacent *cis*-regulatory elements located within short distances (lower than 30 bp) with architectures similar to previously related functional sites for XYR1 and CRE1 regulators [11,18,25,29,37,50,51]. Additionally, the same work-flow (i.e., identification of the best site, mapping of multiple sites

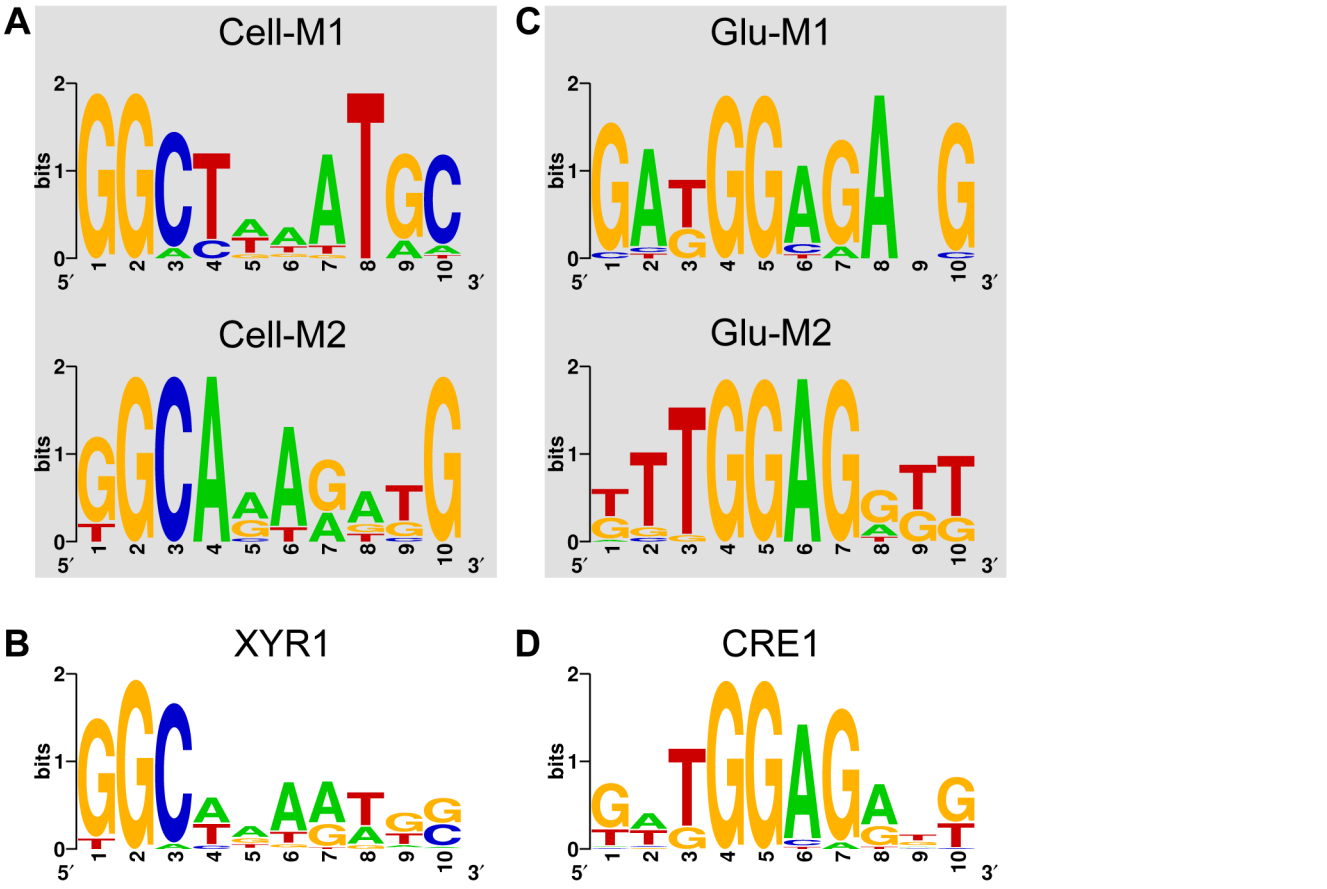


Figure 2. Motifs identified in the cellulose and glucose regulated genes. A) The two motifs identified in the promoter dataset of 22 cellulases that resemble the XYR1 consensus (5′-GGCWWW-3′) are shown. **B)** Combination of the Cell-M1 and Cell-M2 motifs to create the XYR1 consensus used to search for XYR1 binding sites in *T. reesei*. **C)** The two motifs identified using the promoters of TFs up regulated under glucose growth that share similarity to the proposed CRE1-binding consensus (5′-SYGGRG-3′) are shown. **D)** Representation of the consensus resulting from the combination of Glu-M1 and Glu-M2 motifs.
doi:10.1371/journal.pone.0099366.g002

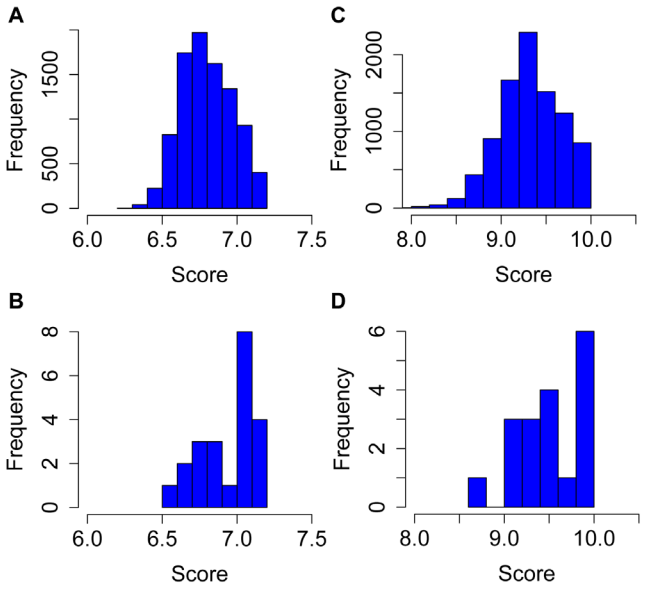


Figure 3. Search for single XYR1 and CRE1 binding sites on different promoter datasets. For the analysis, only the best site was retrieved for each studied promoter. **A**) Distribution of XYR1-binding sites score for all genes from the *T. reesei* genome. **B**) Score of XYR1-binding sites at the 22 cellulase promoters. **C**) Distribution of CRE1-bindind sites scores at the genome scale. **D**) Scores of CRE1-bindind sites found at the 18 promoters from the glucose dataset.
doi:10.1371/journal.pone.0099366.g003

and the search for adjacent elements) was applied to the promoters of the 18 genes from the Glucose dataset and the 22 cellulases promoters using both PWMs. Finally, two additional datasets, representing genes differentially expressed by growth on cellulose, sophorose and glucose in *Δxyr1* and *Δcre1* mutants of *T. reesei*, were inspected as described above to determine genes potentially and directly regulated by XYR1 (in the case of cellulose and sophorose growth conditions) and CRE1 (for glucose growth).

Results and Discussion

Discovery of Putative *Cis*-regulatory Elements in Co-regulated Genes in *T. reesei*

In order to gain quantitative information on the *cis*-regulatory elements of XYR1 and CRE1 in *T. reesei*, we used four different datasets of co-regulated genes to search for short repetitive DNA motifs potentially recognized by these regulators. For this, we used RNA-seq data from *T. reesei* cells grown on cellulose, sophorose and glucose as sole carbon sources [45]. Raw sequence data and count data for all samples are available at Gene Expression Omnibus (GEO database) under the accession number GSE53629. Within the differentially expressed genes identified in each condition, we selected only those encoding for TFs and that were up regulated in the different carbon sources. This procedure leads to the identification of 7, 18 and 18 TF-encoding genes on cellulose, sophorose and glucose growth conditions, respectively (**Table S1–S3**). It is worth mentioning that we focussed on TF-encoding genes since they could mediate the regulation by XYR1 or CRE1 in an indirect way [6]. In addition to these three datasets, we used a fourth group formed by 22 cellulase-encoding genes whose expression was impaired in a strain of *T. reesei* lacking a functional *xyr1* gene [46]. Using these datasets, we could then search for DNA motifs that are similar between the different group of genes and those which are specific to each

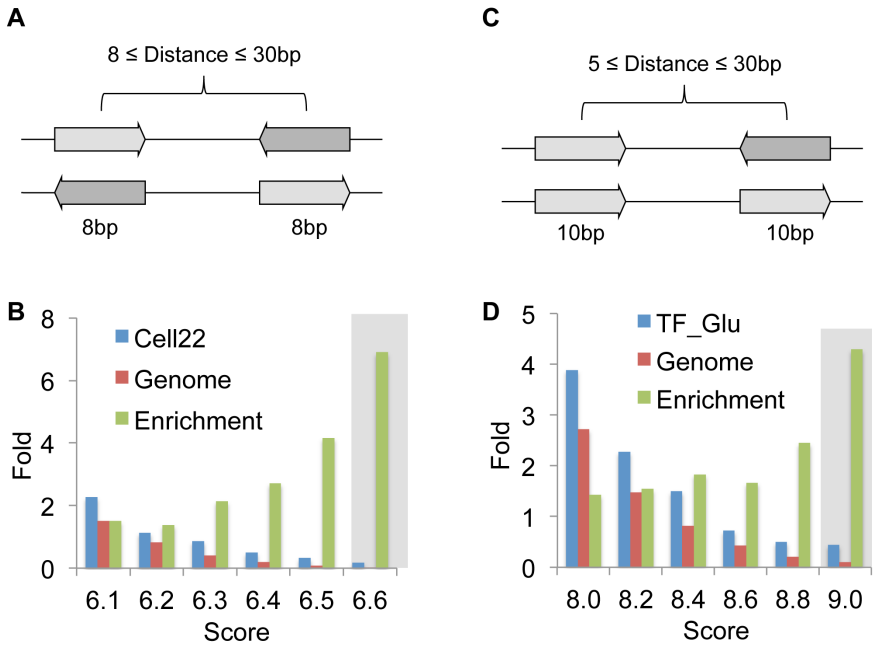


Figure 4. Search for repeated XYR1 and CRE1 binding sites on different datasets. A) In the XYR1, both inverted and everted sites were considered and only sites within a distance between 8 and 30 bp were taken. **B**) Representation of repeated binding sites at the cellulase promoters and at the genome scale. The y-axis (fold) represents the number of sites identified relative to the number of promoters from the datasets. The enrichment group represents the rate between sites per promoters from the cellulase promoters and the corresponding valued from the genome group. Grey shaded region highlight the score with higher enrichment. **C**) For the prediction of CRE1-binding sites, both inverted and direct repeats spaced between 5 to 30 bp were considered. **D**) Representation of sites per promoters and the enrichment at the glucose dataset vs. the genome, calculated as in **B**.
doi:10.1371/journal.pone.0099366.g004

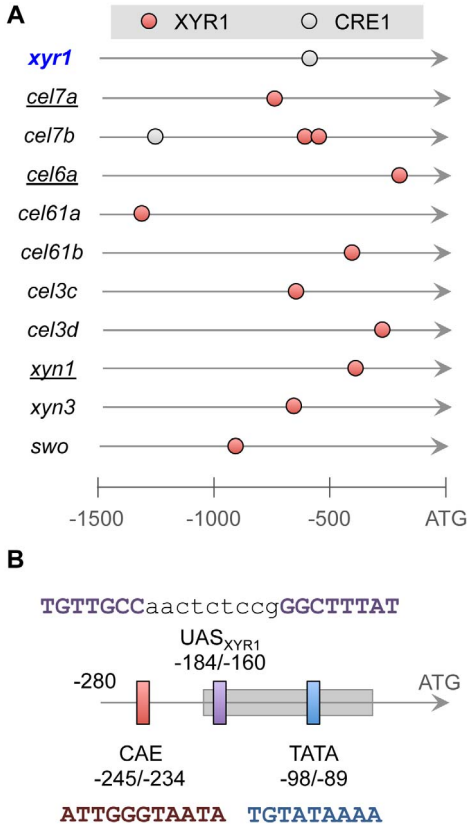


Figure 5. Identification of XYR1 and CRE1-binding sites at target promoters. A) Representation of the binding sites found at the promoters of 10 cellulase encoding-genes and at promoters of the *xyr1* gene. Each circle represent a binding site formed by repeats of the core sequences recognized by the two regulators. **B**) Zoom in at the promoter region of *cel6a* gene, showing the CAE (vertical red bar), the TATA-box (vertical blue bar), the nucleosome −1 binding region (horizontal grey rectangle) and the nearly identified binding site for XYR1, labelled as UAS_{XYR1}. The DNA sequences of each regulatory element are shown [26]. Shown positions are relative to the start codon (ATG) of the *cel6a* gene.
doi:10.1371/journal.pone.0099366.g005

experimental condition. The overall strategy used here is described schematically in **Fig. 1**. For motif discovery, we used MEME software [47] set to find short DNA segments (from 6 to 10 nt in length) that occurred zero or one time in each promoter. From the group of retrieved motifs (10 per dataset), we searched for those resembling the reported consensus for XYR1 (5′-GGCWWW-3′, where W stands for "A" or "T") and CRE1 (5′-SYGGRG-3′, where refers to S is for "G" or "C", Y for "C" or "T" and R for "A" or "G") [11,41]. In the case of the genes from RNA-seq experiments related to cells grown on cellulose or sophorose, we found no motif that resembled either consensus sequence. However, in the case of the 22 cellulases genes dependent of XYR1 [46] we were able to identify two motifs (named Cell-M1 and Cell-M2, **Fig. 2A**) showing a highly conserved GGC core followed by an AT rich region [11]. Each of these motifs was present in all the 22-cellulase promoters analysed, and they mainly diverged in the bases that were conserved at the 3′-end of the sequences (**Fig. 2A**). Then, the DNA sequences of the Cell-M1 and Cell-M2 motifs were merged to generate consensus that potentially represents the binding site for XYR1 in *T. reesei* (**Fig. 2B**).

On the other hand, when the short DNA motifs discovered by MEME on the Glucose dataset were analysed, we again could find two motifs (Glu-M1 and Glu-M2) that resembled the reported consensus sequence for CRE1 (**Fig. 2C**). These motifs have a core GGAG sequence at positions 4 to 7 that matches the expected GGRG consensus. However, although position 3 was expected to be occupied by a T or a C base, the most frequent base found at this location was a T, followed by a G, which was the second most represented base nucleotide (**Fig. 2C**). The region formed by these 5 bases (from 3 to 7) was the most conserved over the motifs, as shown by the consensus generated by the joint of their sequences (**Fig. 2D**). The position 2, which was expected to be occupied by a G or a C, was less conserved compared to the core region but displayed a slight preference for A and T bases (**Fig. 2D**). Finally, since these motifs were only detected on the Glucose promoter dataset and due to their high homology with the consensus for CRE1 binding site, we suggest that the motif in **Fig. 2D** in fact represents the *cis*-regulatory element recognized by this protein in *T. reesei*. Taken together, XYR1 and CRE1 motifs (**Fig. 2B and 2D**) display the first high-resolution representation of the binding sites recognized by XYR1 and CRE1, respectively, in *T. reesei*, and are valuable tools to investigate their regulons in this organism.

Determining the Architecture of *Cis*-regulatory Elements for XYR1 and CRE1

After identifying the putative binding consensus of XYR1 and CRE1 in *T. reesei*, we decided to decipher how these regulators recognize their target promoters. While several works have tried to understand this process at the global scale, a unified model for protein-DNA interaction for these regulators is still not available [6,38]. For instance, several reports support the notion that functional XYR1-binding sites have to be arranged in a specific way, such as inverted repeats [29,52]. Alternatively, Furukawa and colleagues (2009) suggested that XYR1 regulated promoters are endowed with a higher number of single DNA sequences matching the GGCWWW consensus than the background genome. In the same way, promoter recognition by CRE1 has been explained in terms of single sites or repeated sites (either inverted or direct) spaced shortly from one another [10,20,25,27,39]. Since the two motifs generated here (**Fig. 2B and 2D**) represents a higher resolution description of the binding sites for both regulators, we resolved to investigate the role of single and multi-sites for the regulation of target promoters in their respective datasets. Following first the assumption that single sites would be enough to direct the regulators to their targets, one would expect that regulated promoters would harbour binding sites with higher affinity than un-regulated ones [53]. Using the score obtained from the specific Position Weight Matrixes (PWMs) representing XYR1 and CRE1 binding consensus as indicative of the relative TF-binding affinity [54], we accessed the score of identified sites at the genome scale and on the specific datasets (i.e., the group of 22 cellulases for XYR1 sites and the group of TFs up-regulated under Glucose growth for CRE1). For this, we identified only the best hit per promoter according to the two PWMs.

As shown in **Fig. 3A**, the putative binding sites for XYR1 at the genome scale followed a normal-like distribution with scores ranging from ~6.2 to 7.3 and a peak near 6.7. However, when the promoter of the 22 cellulase genes were analysed, we found a much arrowed distribution with two peaks, one close to 6.7 and the other around 7 (**Fig. 3B**). In fact, this analysis reveals that nearly half of the cellulase promoters are endowed with a site with score above 7, while a considerably lower portion of the entire genome presents sites above this score. In the case of CRE1, the same pattern was observed, since a large portion of the Glucose

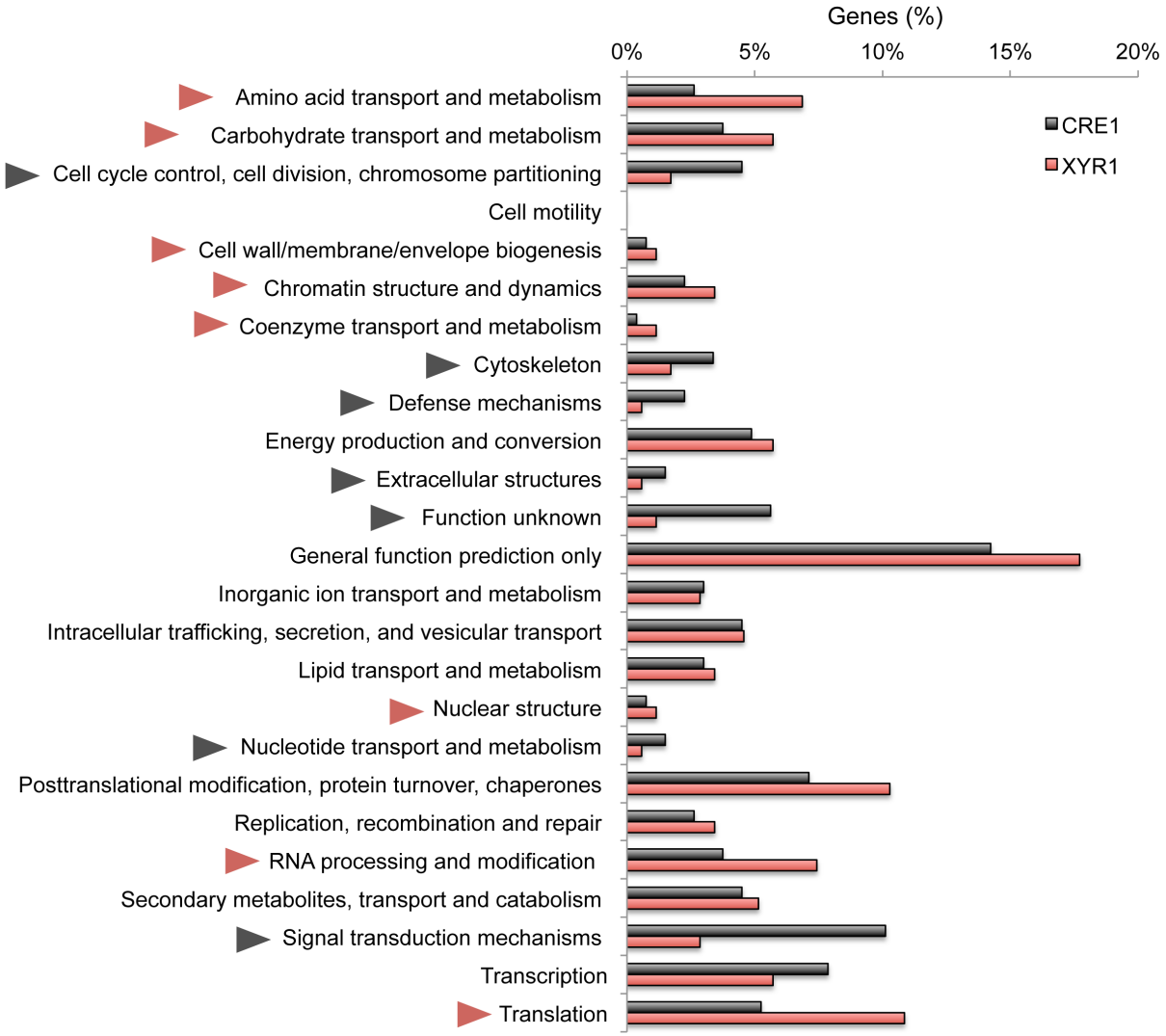


Figure 6. Defining the categories of genes potentially regulated directly by XYR1 and CRE1. Percentages of genes belonging to each functional category for both regulons are shown. Red triangles indicate functional categories enriched in the XYR1 regulon, while grey triangles point to those more abundant on the CRE1 regulon.
doi:10.1371/journal.pone.0099366.g006

promoter dataset displayed binding sites with scores between 9.8 and 10, while sites in this range were less frequent at the genome scale (**Fig. 3C–D**). In this way, these data strengthen the notion that single, high-affinity sites could be used as good descriptors of promoters targeted by XYR1 and CRE1 [11]. Yet, since the high score sites were still very abundant at the genomic level, we decided to investigate binding sites arranged in specific architectures proposed previously [6,11,18,29,55]. In the case of XYR1 binding sites, we focused on both inverted and everted repeats located within in a distance of between 8 to 30 nt from each other (**Fig. 4A**). We then searched for sites that fulfilled this requirement and used different thresholds in terms of score of both sites (ranging from 6.1 to 6.6). As shown in **Fig. 4B**, as expected, increasing threshold values generated the identification of fewer sites per promoters at both the cellulase group and at the genome scale. However, at higher thresholds we observed an enrichment of sites at the cellulase promoters in comparison to genes of *T. reesei* genome (**Fig. 4B**). In fact, one of the promoters that passed the 6.6 threshold value is from the Cel7a gene, which encodes one of the most abundant cellulase genes produced by *T. reesei* [6]. For the inspection of CRE1 binding sites, we searched for sites spaced within a distance of between 5 to 30 nt and both inverted and direct repeats were considered ([56]; **Fig. 4C**). As in the case of XYR1, the same tendency for enrichment of sites at the target promoters was observed, with an augmentation level of more than 4 when the score 9.0 was applied (**Fig. 4D**). These results show that searching for dual binding sites allowed a better definition of XYR1 and CRE1 targets. It is worth to notice that the disposition of the binding sites seemed to be more important than the PWM score itself, since thresholds near the average values of the genome distribution still provided a high enrichment in the target promoters (**Fig. 4B and 4D**). All together, the data provided here suggest that dual binding sites are more relevant for the recognition of the target promoters by XYR1 and CRE1 than single sites.

Inspection of Cellulase Promoters for XYR1 and CRE1 Binding Sites

Once we defined the relevant architecture of the *cis*-regulatory elements potentially recognized by XYR1 and CRE1, we decided

to search for the presence of these elements in the promoters of the cellulase-encoding genes. As discussed before, it is well known that cellulase-encoding genes are controlled at the transcriptional level by induction in the presence of the substrates (such as cellulose and sophorose) and repression mediated by a preferred carbon source such as glucose [6,38]. Whereas the participation of XYR1 and CRE1 regulators have been characterized for the induction and repression of cellulase promoters, respectively, remains an open question about which promoters are directly recognized by these proteins and which ones are regulated through indirect mechanisms involving yet unknown TFs [6,38]. Up to now, *in vivo* and *in vitro* evidence for direct interaction between XYR1 and CRE1 have been reported for promoters such as *cel7a* [18,19], *xyn1* [13,19,29] and proposed for *cel6a* [26,40,57]. In this way, we proceeded to the identification of potential XYR1 and CRE1 binding sites on 22 cellulase promoters using the architectures defined in the previous section, and we contrasted results with the information available in literature for the three characterized promoters. In addition, we inspected the promoter region of the *xyr1* gene, since CRE1 has been reported to affect its expression [17,58]. As represented in **Fig. 5A**, using a threshold value of 6.4 for XYR1 binding sites, we were able to identity dual elements in 10 cellulase promoters. While most of the promoters presented a single dual site, the *cel7b* promoter displayed two sites that were shortly spaced and located about 600 bp upstream of the gene start codon (**Fig. 5A**). In addition, most of the identified promoters (9 out of 10) presented a putative XYR1 binding site located less than 1 kb from the ATG codon. Another interesting finding was that the promoters of three (Cel7a, Cel6a, Cel7b) of the four most efficient cellulolytic enzymes produced by *T. reesei* presented a putative dual XYR1 binding site as detected using the searching approach presented here.

Comparison of the binding sites found *in silico* with those previously characterized at some cellulase promoters showed a remarkable level of agreement. First, in the case of the *xyn1* promoter, a GGCTAA-box formed by two inverted repeats of the GGCWWW element and located around position −410 was shown to be required for XYR1 interaction *in vivo* and *in vitro* [29]. Using the searching methodology described here, the same sequence was identified as the putative XYR1 binding site (**Fig. 5A**). In the case of the *cel7a* promoter, two potential single binding sites for XYR1 have been proposed as functional at positions −320 and −733, but no direct evidence for their role was provided yet [25]. Our *in silico* analysis of the *cel7a* promoter revealed a high-score direct repeat site that includes at its 3′ region the −733 site previously reported. Finally, the case of the *cel6a* promoter represents a more interesting example. Previous analysis have revealed the existence of a region named CAE (for *cbh2* activating element) between positions −234 and −245 that is essential for the induction of this promoter in response to the presence of cellulose and sophorose [51], and this region was found to be regulated by the HAP2/3/5 complex and a yet uncharacterized protein [26]. By studying the nucleosome occupancy of the *cel6a* promoter, Zeilinger and colleagues (2003) found that the CAE is located in a nucleosome-free region and that proteins binding to this element controls the assemble of a nucleosome (named nucleosome −1) covering a region from −192 to −49 that includes the TATA-box [26]. These authors proposed an induction model where the removal of the nucleosome −1 is necessary to allow the interaction of proteins with TATA-box to allow the induction of the *cel6a* promoter in response to cellulose and sophorose. However, the target sequence of the XYR1 regulator at this promoter has not yet been identified, although the available data suggest that it would bind somewhere downstream the CAE region [26,51]. The search for XYR1 binding sites used here allowed the identification of an everted repeat at the *cel6a* promoter at the position −160 to −184, which agrees perfectly with the current available information on the regulation of this promoter. It is worth to notice that everted binding sites are recognized by other zinc finger proteins [56], but XYR1 has not been associated with these elements so far. **Fig. 5B** summarizes the putative *cel6a* promoter architecture, including the nearly identified potential XYR1 binding site (named UAS_{XYR1}) and the *cis*-regulatory elements characterized previously [26,51].

When CRE1 binding sites were investigated in the cellulase dataset, only the *cel7a* promoter revealed an element that passed the stringent used criteria (**Fig. 5A**). However, most of the CRE1 binding sites proposed or demonstrated in literature have a poly-G at the "GGRG" part of the consensus sequence, while the PWM_{CRE1} identified here has a clear preference for the GGAG sequence (**Fig. 2D**). Yet, direct interaction between CRE1 and target promoters has been demonstrated for *cel7a* promoter but not for *cel6a* [26], clearly suggesting that CCR mediated by CRE1 on cellulase genes should be exerted through indirect mechanisms. In this sense, the clear candidate to be the mediator of CRE1 regulation is XYR1 itself, which has been reported as affected by CRE1 [17,58]. In fact, *xyr1* was found to be the top one up-regulated gene under growth on glucose in a strain of *T. reesei* lacking a functional CRE1 protein (Antonieto *et al.* 2014, manuscript in preparation). Quite surprisingly, the searching for CRE1 binding sites at the *xyr1* promoter retrieved just a *cis*-regulatory element (**Fig. 5A**), contrasting the previous prediction of 10 single sites found using the degenerated consensus for this regulator [58]. Taken together, these analyses suggest that the binding sites identified here represent high confidence binding sites for XYR1 and CRE1 in *T. reesei*.

Genome-wide Identification of Potential XYR1 and CRE1 Targets

Once we defined the *cis*-regulatory architectures potentially recognized by XYR1 and CRE1, we performed a genome-wide inspection of potential target promoters for both regulators. For this, a list of 9,115 promoters relative to the annotated genes of *T. reesei* were analysed using the search criteria described in **Fig. 4**. Using a stringent threshold of 6.6 and 9.2 for XYR1 and CRE1 binding sites, respectively, we identified 233 genes potentially regulated by the former (**Table S4**) and 310 candidates for the latter (**Table S5**). Next, we classified the identified genes according to their functional categories (KOG), and then we compared the regulons to identify the difference in the functional scope of both regulators. As shown in **Fig. 6**, the potential XYR1 regulon is enriched mainly with genes related to the metabolization of carbohydrates and amino acids, chromatin structure and dynamics, RNA processing and modification and translation, among others. On the other hand, the putative CRE1 regulon showed a strong augmentation for genes related to signal transduction mechanisms and genes with unknown functions, cytoskeleton, cell cycle control and signal transduction mechanisms. While this analysis provided some clues about the potential targets of XYR1 and CRE1, it should be notice that it does not provide a full description of their regulons since additional TFs could mediate indirect regulation at target promoters [6,59]. In this way, a better understanding of the functional scope of these proteins requires the integration of *in vivo* expression data, as is described below.

Defining the Direct Role of XYR1 and CRE1 under Different Growth Conditions

In order to get an insight into the functioning of the regulon of XYR1 and CRE1 in *T. reesei*, we carried out a search for binding sites of these proteins in 6 sets of genes differentially regulated under different growth conditions. For this, we searched for XYR1 sites in the promoters of genes up and down regulated under growth in the presence of cellulose and sophorose in a strain of *T. reesei* lacking the functional *xyr1* gene, as determined through RNA-seq experiments (Castro *et al.*, manuscript in preparation). In the same way, we surveyed CRE1 sites in promoters of genes up and down regulated during growth on glucose in a strain lacking the *cre1* gene (Antonieto *et al.*, manuscript in preparation). Using these datasets, we found that between 13.6 and 15.8% of the genes differentially regulated in the wild type and *xyr1* minus strain presented a putative XYR1 binding site. In the case of the *cre1* mutant experiments, between 8.9 and 13.9% of the promoters were endowed with a putative binding site for CRE1. The list of genes identified using this analysis along with their expression values determined using RNA-seq is provided in the Supporting Information (**Tables S6-S11**). Taken together, these data strongly indicated that indirect regulation plays an important role on control of target genes by XYR1 and CRE1 proteins in *T. reesei.*

Conclusions

The data provide here addressed for the first time the quantitative identification of binding sites for XYR1 and CRE1 proteins, two general regulators that coordinate the expression of cellulase-encoding genes in *T. reesei* [10,11,58]. The main advantage of the approach used here was the utilization of sets of co-regulated genes to allow the unsupervised discovery of DNA motifs potentially related to the binding of TFs acting at the target group of genes. This analysis allowed us to define PWMs for XYR1 and CRE1 that are specific to *T. reesei*, eluding the bias generated by using consensus sequences determined in other organisms [41]. With these tools on hand, we could observe that while single sites worked generally well as descriptors of XYR1- and CRE1-regulated genes, repeated motifs shortly spaced and with different arrangements seemed to be more associated with promoters targeted by these regulators [42]. So, why some promoters would be endowed with high-score single sites while others presented repeated sites? In the face of the results presented here and those from the general model for cellulase induction currently available [6,38], we propose a mechanistic model for XYR1 binding that could explain the presence of single or dual sites on cellulase promoters. The current proposed induction mechanism suggests that under a repression condition (i.e., in the presence of glucose) the production of cellulases is completely blocked, while under starvation conditions basal levels of these enzymes (mainly Cel7a and Cel6a) are produced. Subsequently, when the fungus finds cellulose, the produced enzymes act on this insoluble substrate to generate soluble inducers such as sophorose, which in turn would trigger the signal for high level of cellulase production [60,61]. In our model, since we observed high-confidence dual binding sites on the *cel7a* and *cel6a* promoters as well as a strong CRE1 site in the *xyr1* regulatory region (**Fig. 5A**), starvation conditions would increase the levels of the XYR1 protein through the release of CCR mediated by CRE1 on its promoter. Next, increasing XYR1 levels would allow the formation of homodimers that would preferentially activate promoters endowed with *cis*-regulatory elements arranged as repeats, such as those for Cel7a and Cel6a. This would account for an increase in the basal expression of these proteins [60,61]. Therefore, when these enzymes convert cellulose into the inducers such as sophorose, additional TFs able to sense this molecule would act in synergy with XYR1, perhaps through the formation of heterodimers, to active cellulase promoters formed by single of repeated sites, allowing thus the production of high levels of cellulases [6,13]. Candidates for such promoter specific regulation include the nearly characterized BglR that regulates some β-glucosidase genes [16]. Evidently, new experimental approaches are required to get further insights into the mechanisms of signal integration present in the cellulase promoters in *T. reesei*, and we believe that the work reported here will contribute significantly for this task.

Supporting Information

Tables S1 Dataset of TFs up regulated in cellulose growth condition.
(PDF)

Tables S2 Dataset of TFs up regulated in sophorose growth condition.
(PDF)

Tables S3 Dataset of TFs up regulated in glucose growth condition.
(PDF)

Tables S4 Genome-wide prediction of XYR1 binding sites *in T. reesei*.
(PDF)

Tables S5 Genome-wide prediction of CRE1 binding sites *in T. reesei*.
(PDF)

Tables S6 Prediction of XYR1 binding sites on genes down regulated in a *Δxyr1* mutant induced with cellulose.
(PDF)

Tables S7 Prediction of XYR1 binding sites on genes up regulated in a *Δxyr1* mutant induced with cellulose.
(PDF)

Tables S8 Prediction of XYR1 binding sites on genes down regulated in a *Δxyr1* mutant induced with sophorose.
(PDF)

Tables S9 Prediction of XYR1 binding sites on genes up regulated in a *Δxyr1* mutant induced with sophorose.
(PDF)

Tables S10 Prediction of CRE1 binding sites on genes down regulated in a *Δcre1* mutant induced with sophorose.
(PDF)

Tables S11 Prediction of CRE1 binding sites on genes up regulated in a *Δcre1* mutant induced with sophorose.
(PDF)

Acknowledgments

We are thankful to lab colleagues for insightful discussion and assistance.

Author Contributions

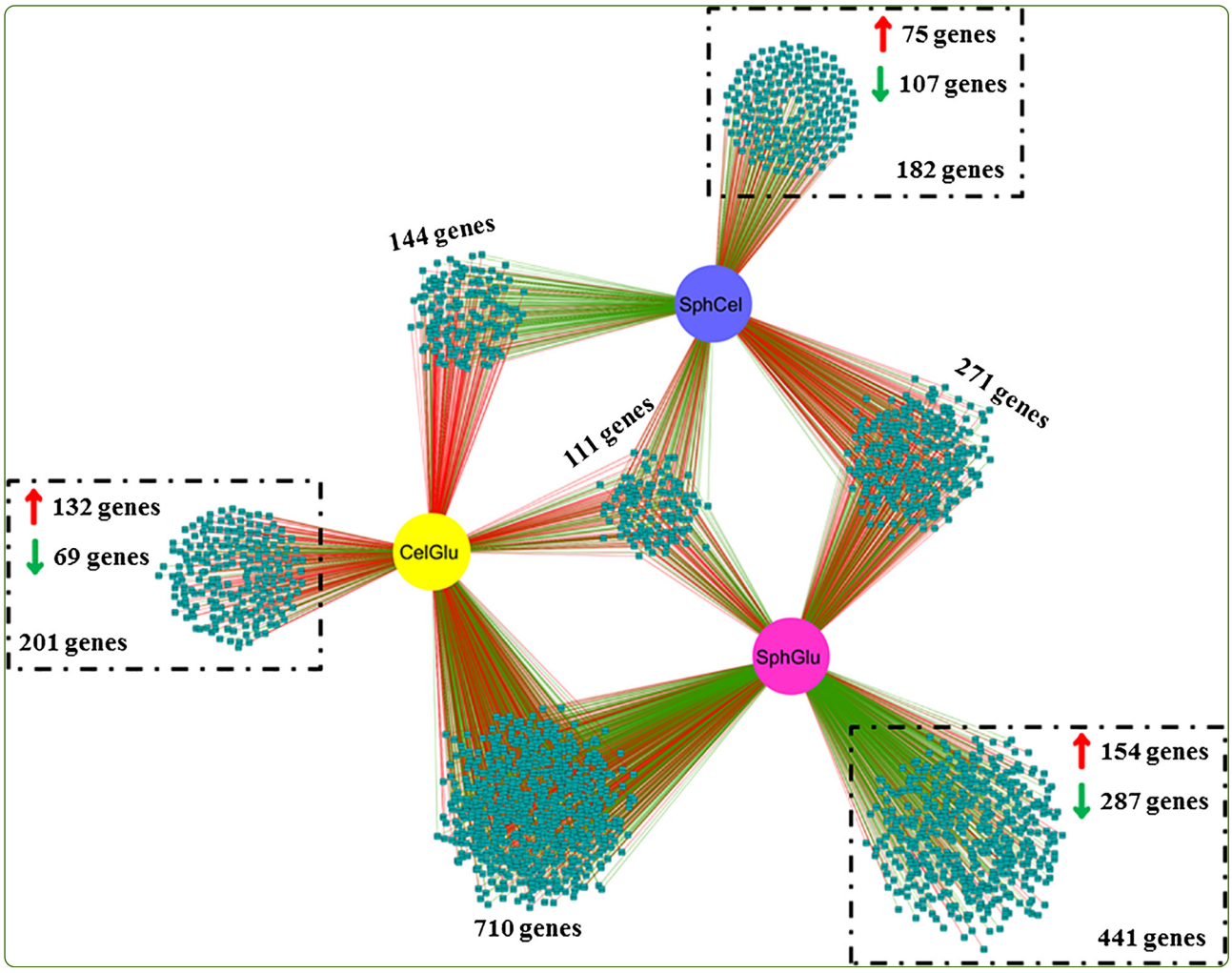
Conceived and designed the experiments: RSR RNS. Performed the experiments: RSR LSC ACCA GFP. Analyzed the data: RSR MEG. Wrote the paper: RSR RNS.

References

1. Tomme P, Van Tilbeurgh H, Pettersson G, Van Damme J, Vandekerckhove J, et al. (1988) Studies of the cellulolytic system of Trichoderma reesei QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. European journal of biochemistry/FEBS 170: 575–581.
2. Sternberg D, Mandels GR (1979) Induction of cellulolytic enzymes in Trichoderma reesei by sophorose. Journal of bacteriology 139: 761–769.
3. Minty JJ, Singer ME, Scholz SA, Bae CH, Ahn JH, et al. (2013) Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. Proceedings of the National Academy of Sciences of the United States of America 110: 14592–14597.
4. Maurya DP, Singh D, Pratap D, Maurya JP (2012) Optimization of solid state fermentation conditions for the production of cellulase by Trichoderma reesei. Journal of environmental biology/Academy of Environmental Biology, India 33: 5–8.
5. Schuster A, Schmoll M (2010) Biology and biotechnology of Trichoderma. Applied microbiology and biotechnology 87: 787–799.
6. Amore A, Giacobbe S, Faraco V (2013) Regulation of cellulase and hemicellulase gene expression in fungi. Current genomics 14: 230–249.
7. Adav SS, Chao LT, Sze SK (2012) Quantitative secretomic analysis of Trichoderma reesei strains reveals enzymatic composition for lignocellulosic biomass degradation. Molecular & cellular proteomics : MCP 11: M111 012419.
8. Stricker AR, Mach RL, de Graaff LH (2008) Regulation of transcription of cellulases- and hemicellulases-encoding genes in Aspergillus niger and Hypocrea jecorina (Trichoderma reesei). Applied microbiology and biotechnology 78: 211–220.
9. Ouyang J, Yan M, Kong D, Xu L (2006) A complete protein pattern of cellulase and hemicellulase genes in the filamentous fungus Trichoderma reesei. Biotechnology journal 1: 1266–1274.
10. Portnoy T, Margeot A, Linke R, Atanasova L, Fekete E, et al. (2011) The CRE1 carbon catabolite repressor of the fungus Trichoderma reesei: a master regulator of carbon assimilation. BMC genomics 12: 269.
11. Furukawa T, Shida Y, Kitagami N, Mori K, Kato M, et al. (2009) Identification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in Trichoderma reesei. Fungal genetics and biology : FG & B 46: 564–574.
12. Wang S, Liu G, Wang J, Yu J, Huang B, et al. (2013) Enhancing cellulase production in Trichoderma reesei RUT C30 through combined manipulation of activating and repressing genes. Journal of industrial microbiology & biotechnology 40: 633–641.
13. Herold S, Bischof R, Metz B, Seiboth B, Kubicek CP (2013) Xylanase gene transcription in Trichoderma reesei is triggered by different inducers representing different hemicellulosic pentose polymers. Eukaryotic cell 12: 390–398.
14. Gyalai-Korpos M, Nagy G, Mareczky Z, Schuster A, Reczey K, et al. (2010) Relevance of the light signaling machinery for cellulase expression in Trichoderma reesei (Hypocrea jecorina). BMC research notes 3: 330.
15. Seiboth B, Karimi RA, Phatale PA, Linke R, Hartl L, et al. (2012) The putative protein methyltransferase LAE1 controls cellulase gene expression in Trichoderma reesei. Molecular microbiology 84: 1150–1164.
16. Nitta M, Furukawa T, Shida Y, Mori K, Kuhara S, et al. (2012) A new Zn(II)(2)Cys(6)-type transcription factor BglR regulates beta-glucosidase expression in Trichoderma reesei. Fungal genetics and biology : FG & B 49: 388–397.
17. Portnoy T, Margeot A, Seidl-Seiboth V, Le Crom S, Ben Chaabane F, et al. (2011) Differential regulation of the cellulase transcription factors XYR1, ACE2, and ACE1 in Trichoderma reesei strains producing high and low levels of cellulase. Eukaryotic cell 10: 262–271.
18. Ling M, Qin Y, Li N, Liang Z (2009) Binding of two transcriptional factors, Xyr1 and ACEI, in the promoter region of cellulase cbh1 gene. Biotechnology letters 31: 227–231.
19. Stricker AR, Grosstessner-Hain K, Wurleitner E, Mach RL (2006) Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in Hypocrea jecorina. Eukaryotic cell 5: 2128–2137.
20. Ilmen M, Thrane C, Penttila M (1996) The glucose repressor gene cre1 of Trichoderma: isolation and expression of a full-length and a truncated mutant form. Molecular & general genetics : MGG 251: 451–460.
21. Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. Annual review of biochemistry 78: 273–304.
22. Berger SL (2007) The complex language of chromatin regulation during transcription. Nature 447: 407–412.
23. Narlikar GJ, Fan HY, Kingston RE (2002) Cooperation between complexes that regulate chromatin structure and transcription. Cell 108: 475–487.
24. Struhl K (1999) Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. Cell 98: 1–4.
25. Ries L, Belshaw NJ, Ilmen M, Penttila ME, Alapuranen M, et al. (2014) The role of CRE1 in nucleosome positioning within the cbh1 promoter and coding regions of Trichoderma reesei. Applied microbiology and biotechnology 98: 749–762.
26. Zeilinger S, Schmoll M, Pail M, Mach RL, Kubicek CP (2003) Nucleosome transactions on the Hypocrea jecorina (Trichoderma reesei) cellulase promoter cbh2 associated with cellulase induction. Molecular genetics and genomics : MGG 270: 46–55.
27. Mach RL, Strauss J, Zeilinger S, Schindler M, Kubicek CP (1996) Carbon catabolite repression of xylanase I (xyn1) gene expression in Trichoderma reesei. Molecular microbiology 21: 1273–1281.
28. Messner R, Kubicek CP (1991) Carbon Source Control of Cellobiohydrolase I and II Formation by Trichoderma reesei. Applied and environmental microbiology 57: 630–635.
29. Rauscher R, Wurleitner E, Wacenovsky C, Aro N, Stricker AR, et al. (2006) Transcriptional regulation of xyn1, encoding xylanase I, in Hypocrea jecorina. Eukaryotic cell 5: 447–456.
30. Sun J, Tian C, Diamond S, Glass NL (2012) Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in Neurospora crassa. Eukaryotic cell 11: 482–493.
31. van Peij NN, Visser J, de Graaff LH (1998) Isolation and analysis of xlnR, encoding a transcriptional activator co-ordinating xylanolytic expression in Aspergillus niger. Molecular microbiology 27: 131–142.
32. Aro N, Ilmen M, Saloheimo A, Penttila M (2003) ACEI of Trichoderma reesei is a repressor of cellulase and xylanase expression. Applied and environmental microbiology 69: 56–65.
33. Saloheimo A, Aro N, Ilmen M, Penttila M (2000) Isolation of the ace1 gene encoding a Cys(2)-His(2) transcription factor involved in regulation of activity of the cellulase promoter cbh1 of Trichoderma reesei. The Journal of biological chemistry 275: 5817–5825.
34. Klein CJ, Olsson L, Nielsen J (1998) Glucose control in Saccharomyces cerevisiae: the role of Mig1 in metabolic functions. Microbiology 144 (Pt 1): 13–24.
35. Ogasawara W, Shida Y, Furukawa T, Shimada R, Nakagawa S, et al. (2006) Cloning, functional expression and promoter analysis of xylanase III gene from Trichoderma reesei. Applied microbiology and biotechnology 72: 995–1003.
36. Wurleitner E, Pera L, Wacenovsky C, Cziferszky A, Zeilinger S, et al. (2003) Transcriptional regulation of xyn2 in Hypocrea jecorina. Eukaryotic cell 2: 150–158.
37. Aro N, Saloheimo A, Ilmen M, Penttila M (2001) ACEII, a novel transcriptional activator involved in regulation of cellulase and xylanase genes of Trichoderma reesei. The Journal of biological chemistry 276: 24309–24314.
38. Glass NL, Schmoll M, Cate JH, Coradetti S (2013) Plant cell wall deconstruction by ascomycete fungi. Annual review of microbiology 67: 477–498.
39. Takashima S, Iikura H, Nakamura A, Masaki H, Uozumi T (1996) Analysis of Cre1 binding sites in the Trichoderma reesei cbh1 upstream region. FEMS microbiology letters 145: 361–366.
40. Gorsche R, Jovanovic B, Gudynaite-Savitch L, Mach RL, Mach-Aigner AR (2014) A highly sensitive in vivo footprinting technique for condition-dependent identification of cis elements. Nucleic acids research 42: e1.
41. Cubero B, Scazzocchio C (1994) Two different, adjacent and divergent zinc finger binding sites are necessary for CREA-mediated carbon catabolite repression in the proline gene cluster of Aspergillus nidulans. The EMBO journal 13: 407–415.
42. Rahman Z, Shida Y, Furukawa T, Suzuki Y, Okada H, et al. (2009) Evaluation and characterization of Trichoderma reesei cellulase and xylanase promoters. Applied microbiology and biotechnology 82: 899–908.
43. Mach RL, Zeilinger S (2003) Regulation of gene expression in industrial fungi: Trichoderma. Applied microbiology and biotechnology 60: 515–522.
44. Lockington RA, Rodbourn L, Barnett S, Carter CJ, Kelly JM (2002) Regulation by carbon and nitrogen sources of a family of cellulases in Aspergillus nidulans. Fungal genetics and biology : FG & B 37: 190–196.
45. Dos Santos Castro L, Pedersoli WR, Antonieto AC, Steindorff AS, Silva-Rocha R, et al. (2014) Comparative metabolism of cellulose, sophorose and glucose in Trichoderma reesei using high-throughput genomic and proteomic analyses. Biotechnology for biofuels 7: 41.
46. Castro LD, Antonieto AC, Pedersoli WR, Rocha RS, Persinoti GF, et al. (2014) Expression pattern of cellulolytic and xylanolytic genes regulated by transcriptional factors XYR1 and CRE1 are affected by carbon source in Trichoderma reesei. Gene expression patterns : GEP.
47. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic acids research 34: W369–373.
48. Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. Nucleic acids research 18: 6097–6100.
49. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, et al. (2008) Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina). Nature biotechnology 26: 553–560.
50. Murray PG, Collins CM, Grassick A, Tuohy MG (2003) Molecular cloning, transcriptional, and expression analysis of the first cellulase gene (cbh2), encoding cellobiohydrolase II, from the moderately thermophilic fungus Talaromyces emersonii and structure prediction of the gene product. Biochemical and biophysical research communications 301: 280–286.
51. Zeilinger S, Mach RL, Kubicek CP (1998) Two adjacent protein binding motifs in the cbh2 (cellobiohydrolase II-encoding) promoter of the fungus Hypocrea jecorina (Trichoderma reesei) cooperate in the induction by cellulose. The Journal of biological chemistry 273: 34463–34471.
52. Furukawa T, Shida Y, Kitagami N, Ota Y, Adachi M, et al. (2008) Identification of the cis-acting elements involved in regulation of xylanase III gene expression

in Trichoderma reesei PC-3-7. Fungal genetics and biology : FG & B 45: 1094–1102.
53. Wasserman WW, Sandelin A (2004) Applied bioinformatics for the identification of regulatory elements. Nature reviews Genetics 5: 276–287.
54. Hallikas O, Palin K, Sinjushina N, Rautiainen R, Partanen J, et al. (2006) Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. Cell 124: 47–59.
55. Shida Y, Furukawa T, Ogasawara W, Kato M, Kobayashi T, et al. (2008) Functional analysis of the egl3 upstream region in filamentous fungus Trichoderma reesei. Applied microbiology and biotechnology 78: 515–524.
56. MacPherson S, Larochelle M, Turcotte B (2006) A fungal family of transcriptional regulators: the zinc cluster proteins. Microbiology and molecular biology reviews : MMBR 70: 583–604.
57. Stangl H, Gruber F, Kubicek CP (1993) Characterization of the Trichoderma reesei cbh2 promoter. Current genetics 23: 115–122.
58. Mach-Aigner AR, Pucher ME, Steiger MG, Bauer GE, Preis SJ, et al. (2008) Transcriptional regulation of xyr1, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in Hypocrea jecorina. Applied and environmental microbiology 74: 6554–6562.
59. Guelzim N, Bottani S, Bourgine P, Kepes F (2002) Topological and causal structure of the yeast transcriptional regulatory network. Nature genetics 31: 60–63.
60. Carle-Urioste JC, Escobar-Vera J, El-Gogary S, Henrique-Silva F, Torigoi E, et al. (1997) Cellulase induction in Trichoderma reesei by cellulose requires its own basal expression. The Journal of biological chemistry 272: 10169–10174.
61. el-Gogary S, Leite A, Crivellaro O, Eveleigh DE, el-Dorry H (1989) Mechanism by which cellulose triggers cellobiohydrolase I gene expression in Trichoderma reesei. Proceedings of the National Academy of Sciences of the United States of America 86: 6138–6141.

Biotechnology for Biofuels



Comparative metabolism of cellulose, sophorose and glucose in *Trichoderma reesei* using high-throughput genomic and proteomic analyses

dos Santos Castro *et al.*

BioMed Central

dos Santos Castro *et al. Biotechnology for Biofuels* 2014, **7**:41
http://www.biotechnologyforbiofuels.com/content/7/1/41

Biotechnology for Biofuels

RESEARCH Open Access

Comparative metabolism of cellulose, sophorose and glucose in *Trichoderma reesei* using high-throughput genomic and proteomic analyses

Lilian dos Santos Castro[1†], Wellington Ramos Pedersoli[1†], Amanda Cristina Campos Antoniêto[1], Andrei Stecca Steindorff[2], Rafael Silva-Rocha[1], Nilce M Martinez-Rossi[3], Antonio Rossi[1], Neil Andrew Brown[4], Gustavo H Goldman[4], Vitor M Faça[1], Gabriela F Persinoti[3] and Roberto Nascimento Silva[1*]

Abstract

Background: The filamentous fungus *Trichoderma reesei* is a major producer of lignocellulolytic enzymes utilized by bioethanol industries. However, to achieve low cost second generation bioethanol production on an industrial scale an efficient mix of hydrolytic enzymes is required for the deconstruction of plant biomass. In this study, we investigated the molecular basis for lignocellulose-degrading enzyme production *T. reesei* during growth in cellulose, sophorose, and glucose.

Results: We examined and compared the transcriptome and differential secretome (2D-DIGE) of *T. reesei* grown in cellulose, sophorose, or glucose as the sole carbon sources. By applying a stringent cut-off threshold 2,060 genes were identified as being differentially expressed in at least one of the respective carbon source comparisons. Hierarchical clustering of the differentially expressed genes identified three possible regulons, representing 123 genes controlled by cellulose, 154 genes controlled by sophorose and 402 genes controlled by glucose. Gene regulatory network analyses of the 692 genes differentially expressed between cellulose and sophorose, identified only 75 and 107 genes as being specific to growth in sophorose and cellulose, respectively. 2D-DIGE analyses identified 30 proteins exclusive to sophorose and 37 exclusive to cellulose. A correlation of 70.17% was obtained between transcription and secreted protein profiles.

Conclusions: Our data revealed new players in cellulose degradation such as accessory proteins with non-catalytic functions secreted in different carbon sources, transporters, transcription factors, and CAZymes, that specifically respond in response to either cellulose or sophorose.

Keywords: Trichoderma reesei, RNA-seq, DIGE, Cellulases, Bioethanol

Background

The growing worldwide demand for energy and the desire to reduce dependency on finite fossil fuels has increased interest in alternative energy sources, especially liquid biofuels such as bioethanol and biodiesel. Ethanol obtained from lignocellulosic, non-food, feedstocks (for example, sugarcane bagasse or wheat straw) represents an attractive alternative due to its applicability in existing motor vehicles. In addition, the combustion of lignocellulosic-derived ethanol is considered cleaner than oil-based fuels [1]. Lignocellulosic biomass locks away approximately half of the energy produced by plants during photosynthesis and is the most abundant renewable organic carbon resource on Earth. Lignocellulose predominately consists of three polymers that are tightly interlinked, cellulose, hemicellulose and lignin, which correspond to approximately 98% of lignocellulose dry weight [2].

The production of fuel ethanol from lignocellulose requires biomass pretreatment, cellulose hydrolysis, hexose fermentation, separation, effluent treatment, and depending on the raw material, additional costs may occur [3].

* Correspondence: rsilva@fmrp.usp.br
†Equal contributors
[1]Department of Biochemistry and Immunology, Ribeirão Preto Medical School, University of São Paulo, 14049-900 Ribeirão Preto, SP, Brazil
Full list of author information is available at the end of the article

BioMed Central

© 2014 dos Santos Castro et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

In recent years, new technologies have been developed for the pretreatment of sugarcane bagasse such as the application of novel enzymes to increase the saccharification of cellulose/hemicellulose and specialized fermentation technologies, aiding in the development of second-generation (2G) bioethanol [4].

The filamentous fungus *Trichoderma reesei* is one of the main producers of cellulases and hemicellulases used in industrial scale [5] and is especially important for the production of 2G biofuels from lignocellulose [6]. Despite *T. reesei* being the most prominent lignocellulosic degrader among the genus *Trichoderma*, this species has a reduced number of cellulolytic enzymes compared to other lignocellulosic fungi [7]. This ability is attributed to *T. reesei* possessing efficient systems for the transport of nutrients and the induction/secretion of cellulases. Subsequently, the study of the cellulolytic system in *T. reesei* is of substantial interest to industrial biotechnology.

The *T. reesei* cellulolytic system consists of at least three different types of enzymes: exoglucanases (cellobiohydrolases EC 3.2.1.91), endoglucanases (EC 3.2.1.4) and β-glucosidases (EC 3.2.1.21), which occur in various isoforms [8]. *T. reesei* produces at least two exo-acting cellobiohydrolases (CEL7A and CEL6A), five endo-acting cellulases (CEL7B, CEL5A, CEL5B, CEL12A, CEL45A), two characterized β-glucosidases (CEL3A and CEL1A), and an additional five predicted β-glucosidases (CEL3B, CEL3D, CEL1B, CEL3C, CEL3E) [9]. Besides the classic cellulases, new players involved in cellulose degradation were recently described in *T. reesei*, such as the expansin-like proteins swollenin (SWOI) and expansin/family 45 endoglucanase-like (EEL1, EEL2, and EEL3) [7,10]. In addition, GH61 polysaccharide monooxygenases (PMOs), which were recently re-annotated as AA9 (Auxiliary family activity 9, www.cazy.org), have been shown to enhance lignocellulose degradation by an oxidative mechanism [9].

The production of the main cellulases by *T. reesei* is controlled by a sophisticated regulation system that avoids energy expenditure on unrequired processes when readily metabolisable carbon sources are present [6]. Since the 1960s when Mary Mandels and Elwyn T. Reese [11], raised the question 'Cellulases are adaptive enzymes, but the natural substrate – cellulose – is insoluble. So how does induction occur?' many studies have been conducted in an attempt to discover the natural inducer of cellulase formation [12-14]. It is now known that expression of cellulolytic genes in *T. reesei* are induced in the presence of cellulose and several disaccharides such as cellobiose (β-D-glucopyranosyl-(1 → 4)-β-D-glucopyranose), δ-cellobiono-1,5-lactone (β-D-glucopyranosyl-(1 → 4)-D-glucono-1,5-lactone), lactose (β-D-galactopyranosyl-(1 → 4)-D-glucose) and sophorose (2-O-β-D-glucopyranosyl-α-D-glucose)[15]. Sophorose is the strongest cellulase inducer and is considered to be a possible natural inducer. It is assumed that sophorose is formed by *T. reesei* during cellulose hydrolysis by a transglycosylation reaction [16]. However, additional low-molecular weight compounds have been reported to promote cellulase gene expression, such as l-arabitol and l-sorbose [17]. In contrast, the presence of easily metabolisable carbon sources such as glucose and fructose, represses the expression of cellulolytic genes [18].

The regulation of cellulase gene expression occurs at the transcriptional level in a coordinated manner and is dependent on the presence of the inducer [19]. This regulation is driven by specific transcriptional factors (TFs) that bind to cellulase gene promoters acting either in a positive or a negative way. So far, at least three transcriptional activators XYR1, ACE2, the HAP2/3/5 complex, as well as the two repressors CRE1 and ACE1 are involved in the regulation of cellulase gene expression in *T. reesei* [20].

Despite extensive studies attempting to answer the question raised by Mandels and Reese, neither the nature of the inducer nor how *T. reesei* senses the inducer and relays the cellulase induction signal, have been elucidated. In this study we report a comparison of the transcriptome (RNAseq) and secretome (two-dimensional Fluorescence Difference Gel Electrophoresis (2-D DIGE)) of *T. reesei* grown on cellulose, sophorose and glucose, in attempt to understand the molecular basis of lignocellulose-degrading enzyme induction. Our results provide new insights and revealed new players in cellulose degradation such as proteins with non-catalytic functions secreted in different carbon sources, transporters, transcription factors, carbohydrate active enzyme (CAZymes), and the regulatory network of *T. reesei* in response to cellulose and sophorose. These data will contribute to the development of industrial *T. reesei* strains by engineering its metabolism to produce high levels of cellulases for plant cell-wall degradation.

Results

Global gene expression profiles of *Trichoderma reesei* grown in three different carbon sources

T. reesei QM9414 was grown directly in three different carbon sources; glucose, sophorose, and cellulose as described (see Methods). We previously demonstrated the growth profiles and glucose consumption of *T. reesei* QM9414 in the presence of cellulose and glucose [21]. Based on these data, we designed the strategy to pool the time points of each condition before the sequencing. Nine barcoded libraries were sequenced using the Illumina Hiseq 2000 System, generating approximately 117 million 100-bp paired-end reads corresponding to 23.32 GB of sequence data (Additional file 1: Table S2). Reads were mapped to the *T. reesei* QM6a reference genome available

from JGI (*Trichoderma reesei* v2.0) using the Bowtie aligner. Overall, 68% of reads mapped to the reference genome (Additional file 1: Table S2). There was a high correlation (Pearson correlation, $r^2 \geq 0.71$) between the three biological replicates of each condition used in the transcriptional analysis (Additional file 2: Figure S1A-B). After sample normalization, boxplots were constructed in order to determine if the conditions are comparable and the results are shown (see Additional file 2, Figure S1 C-D). The boxplots showed that both normalized samples and raw data displayed the same plot profile and no significant statistical difference (P <0.05), demonstrating that the samples are comparable.

The *T. reesei* gene expression profiles obtained from the different carbon sources were analyzed using R Bioconductor DESeq. Of the 9,129 genes encoded by the *T. reesei* genome, 1,788 genes were identified as being differentially expressed (P <0.05) on glucose/cellulose, 2,545 genes on sophorose/cellulose and 2,481 genes on sophorose/glucose (Figure 1A-C).

Appling a two-fold change (that is, $\log_2$ fold change ≥1 or ≤ −1) and an adjusted P-value ≤0.05 as a threshold, 2,060 genes were identified as differentially expressed in at least one of the respective carbon source comparisons. Figure 2A shows that 1,886 genes were differentially expressed in glucose and in cellulose, as represented by 703 and 491 genes being up- and downregulated exclusively in glucose, and 254 and 102 genes being up- and downregulated exclusively in cellulose, respectively. On the other hand, 1,889 genes were differentially expressed in sophorose and in cellulose, with 321 and 405 being up- or downregulated in sophorose, and 262 and 97 being up- or downregulated in cellulose, respectively (Figure 2B). Yet, in glucose and sophorose 1,670 genes were modulated, with 262 and 245 being up- or downregulated in sophorose, and 505 and 473 up- or downregulated genes in glucose, respectively (Figure 2C). Interestingly, the number of transcriptionally modulated genes in sophorose (726) was greater than that in cellulose (359), when both were compared to glucose.

Hierarchical clustering of the 2,060 differentially expressed genes identified in the comparisons cellulose versus glucose (cel/glu), sophorose versus cellulose (soph/cel) and sophorose versus glucose (soph/glu), allowed the identification of three possible regulons, representing 123 genes modulated by cellulose, 154 genes modulated by sophorose and 402 genes modulated by glucose, totaling 679 genes (Figure 3A; Additional file 3: Table S3). Gene Ontology (GO) annotation of the 679 carbon source-specific genes revealed that 46%, 34% and 39% of the genes from the cellulose, glucose and sophorose regulons respectively were genes of unknown function. These results emphasize the potential for the discovery of genes involved in the cellulase production in *T. reesei* during growth under inducing or repressing conditions.

In order to further evaluate the carbon source-specific regulons shown in Figure 3, the top 10 genes differentially expressed on cellulose, glucose, and sophorose were identified (Table 1). The top 10 upregulated genes in cellulose included the glycoside hydrolases (GH) GH5, GH31 and GH16, the carbohydrate esterase CE5, an oxidase, a specific Major facilitator superfamily (MFS) permease and five proteins of unknown function (Table 1). The top 10 upregulated genes in sophorose included a GH76 hydrolase, four oxidoreductases, two MFS permeases and three proteins of unknown function (Table 1). It is interesting to note that there are more GHs in the top 10 upregulated genes of cellulose than in sophorose. As expected, the top 10 differentially expressed genes in glucose did not show any genes encoding hydrolytic enzymes (Table 1). These results indicate a specific gene expression in response to the available carbon source in *T. reesei*.

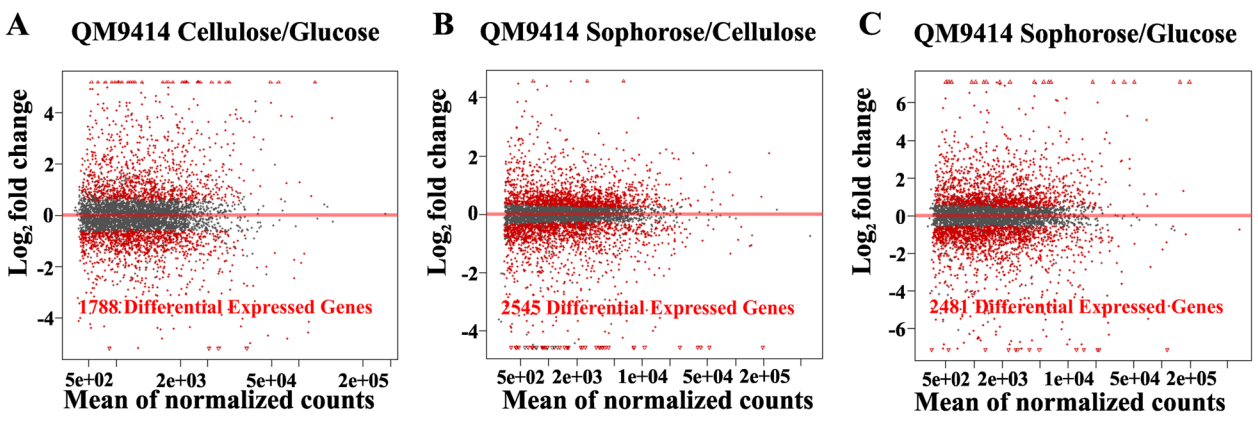


Figure 1 Comparison of full-genome expression profiles of QM9414 strain grown in cellulose, sophorose, and glucose as the carbon source, measured by RNA-seq. (A) Cellulose/glucose. **(B)** Cellulose/sophorose. **(C)** Sophorose/glucose. Differentially expressed genes identified by DESeq package are plotted in red (P <0.05).

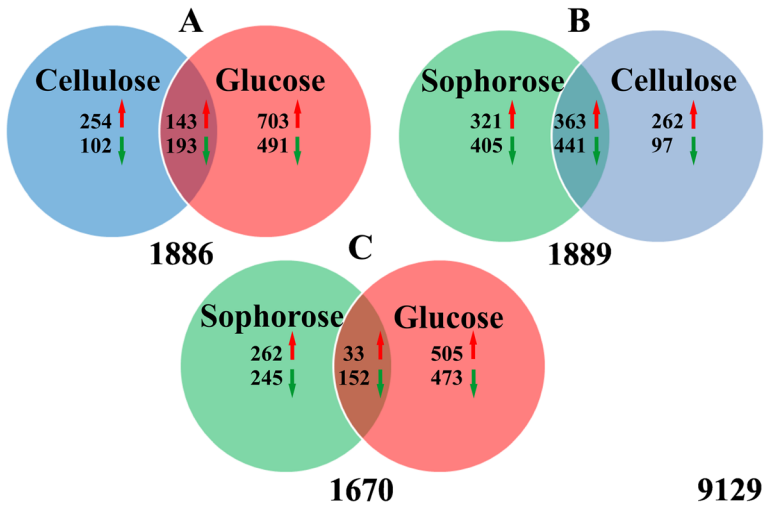


Figure 2 Venn diagram representing the number of differentially expressed genes in the QM9414 strain. (A) QM9414 strain growth in cellulose and glucose as carbon source. **(B)** QM9414 strain growth in cellulose and sophorose. **(C)** QM9414 strain growth in sophorose and glucose. The red arrows indicate the number of upregulated genes and green arrows the number of downregulated genes in the conditions analyzed. The numbers below the given Venn diagram represent the total number of regulated genes. The number in the lower right of rectangle indicates the number of transcripts in the *T. reesei* genome. Thresholds for calling differentially expressed genes were ($P \leq 0.05$) and ≥ two-fold change, that is, $\log_2$-fold change ≥1 or ≤ −1.

CAZYome

The mean FPKM (fragments per kilobase of exon per million fragments mapped) for all the genes within a single GH family were calculated. The total of all the FPKM means for each GH family when cultured in glucose, cellulose and sophorose were utilized to demonstrate the overall enzymatic potential and global transcriptional response (Figure 4). During growth in glucose the overall transcription of GH encoding genes was low, whereas growth in the presence of cellulose or sophorose resulted in a dramatic induction of a wide array of GH families, reflecting the transcriptional induction of the CAZYome. Similarly, cellulose and sophorose resulted in a greater transcriptional induction of cellobiohydrolase members from the GH6 and GH7 families.

Looking in more detail, Table S4 (Additional file 4) shows the enzymes that are differentially upregulated in sophorose and cellulose. Twenty GHs and one CE were upregulated in response to the presence of sophorose whereas 23 GHs and two CEs were induced in the presence of cellulose (Additional file 4: Table S4A and B). Interestingly, genes of enzymes involved in xylan degradation, such as xylanases *(xyn2*, *xyn3*, *xyn4*), acetyl xylan esterase (*axe1*), xyloglucanase (*cel74a*), α-xylosidase (GH31) and arabinoxylans degradation, such as arabinofuranosidase (ABF1 and ABF2) were preferentially expressed in the presence of cellulose but not in sophorose (Additional file 4: Table S4B). Furthermore, a polysaccharide monooxygenase (*cel61A*) was upregulated only in cellulose, in accordance with a role in the cellulose oxidation process. These results were also observed by Bischof *et al.* [22] when transcriptional data from wheat straw was compared to lactose. On the other hand, eight genes encoding α - and β-glucosidases (including *cel3c*, *cel3b* and *cel1b*), and a candidate for α-amylase and α-1,6-mannanase (while on cellulose a β-mannanase was expressed), were upregulated in sophorose (Additional file 4: Table S4A). Interestingly, in both sophorose and cellulose, enzymes that degrade trehalose were induced indicating that the fungus may catabolize stored trehalose, producing glucose, during cellulase production.

When comparing the fold change in gene expression among the three conditions we observed that even in the presence of glucose, 17 GHs were upregulated (Additional file 4: Table S4C). These genes encoded for enzymes such as endoglucanase (*cel5b*), β-1,4-glucanase (GH5), β-1,3-glucanosyltransferase (GH72), and an uncharacterized GH (Trire2_121136) appeared not to be subject to carbon catabolite repression.

Transcription factors

Table S5 (see Additional file 5) shows the TF encoding genes that were induced in each condition. In this analysis, 7 TF encoding genes were upregulated in cellulose, 18 in sophorose and 18 other TF genes were specific to glucose. Within this group, we focused on the TF genes that were upregulated depending on the carbon source. For example, Trire2_105269 showed a high level of expression in the presence of cellulose, whereas Trire2_123881 showed a high level of expression in sophorose, and Trire2_ID 112499 was upregulated by glucose (Additional file 5: Table S5). TF encoding genes from the Zn2Cys6

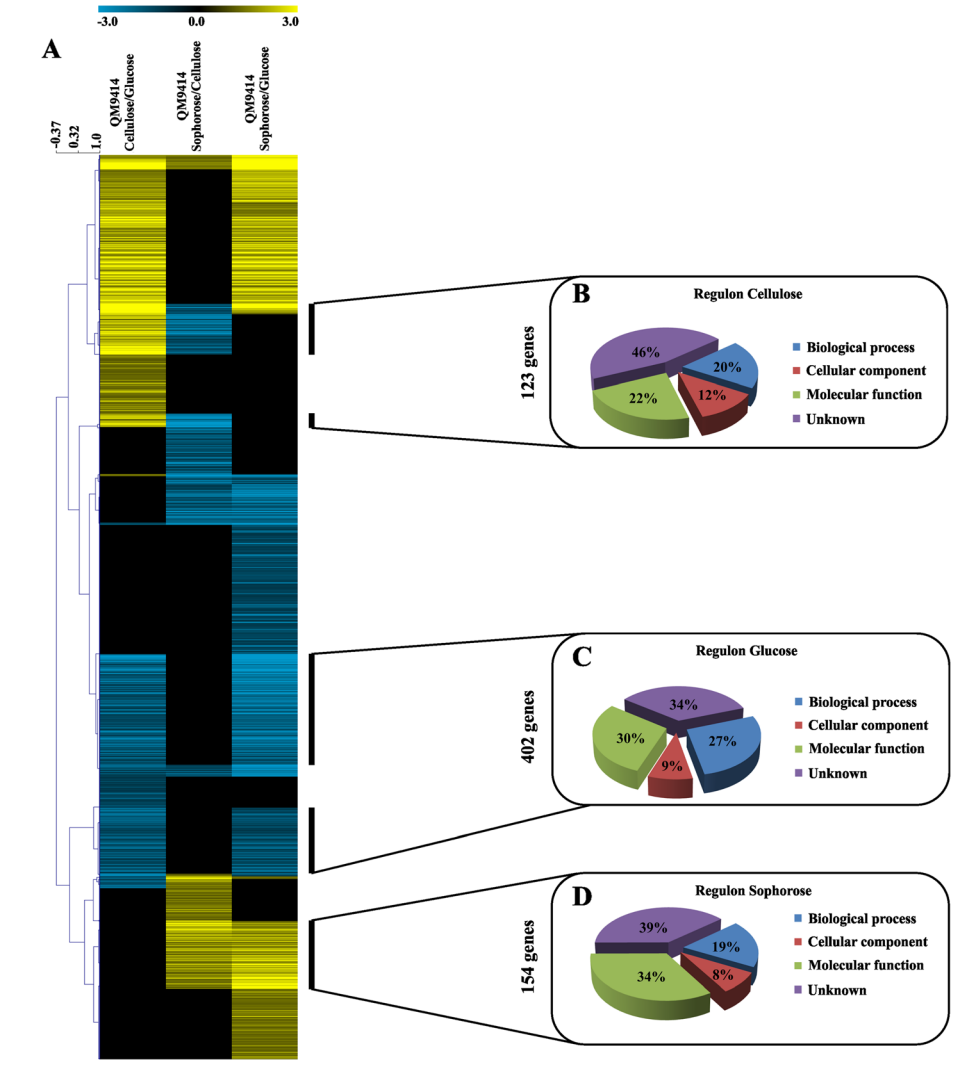


Figure 3 Gene expression profile of *T. reesei*, QM9414 strain, during grown in the presence of cellulose, sophorose and glucose as the carbon source. Expression scale is represented as Log_2 Fold Change. **(A)** Hierarchical clustering analysis was performed using Mev v.4.6.1, with the average linkage method for cluster generation, and uncentered correlation as the similarity metric. Euclidian distance was used to measure the differences in gene expression among the 2,060 genes and the groups (QM9414 Cellulose/Glucose; QM9414 Sophorose/Cellulose and QM9414 Sophorose/Glucose) based on the distance between the centroids of the groups, $P < 0.05$. **(B)** Genes upregulated by cellulose, **(C)** genes upregulated by glucose and **(D)** genes upregulated by sophorose. In **(B)**, **(C)** and **(D)** the summary of the Gene Ontology annotation are also represented.

subgroup known as C2H2 (one of the most common type of transcription factors found in eukaryotes) were only induced in the presence of cellulose and sophorose, but not in glucose-grown cells (Additional file 5: Table S5), suggesting a specific response of C2H2 to the presence of cellulase inducer molecules. TF encoding genes from the bZIP family, on the other hand, showed higher expression in cellulose (Trire2_110152), but were also present in sophorose (Trire2_73654) and glucose (Trire2_119759) (Additional file 5: Table S5).

The expression of TF encoding genes already characterized as being involved in the regulation of the expression of cellulases and hemicellulases is shown in Table 2. Among the positively acting TFs (XYR1, ACE2, CLR-1, CLR-2, and BglR), the gene for XYR1 showed the highest expression level, followed by CLR-1 and BglR (Table 2). The TF genes ACE2 and CLR-2 showed no significant modulation in expression ($P < 0.05$) between the various carbon sources. These results reinforce the hypothesis that XYR1 is the major positive regulator of cellulases and hemicellulase gene expression. On the other hand, the TFs that negatively regulate hydrolytic enzyme gene transcription, such as ACE1 and CREI, showed a lower level of expression, compared to the positive-acting TFs, or were not transcriptionally modulated depending upon the carbon source, suggesting that these TFs may act in

Table 1 Log_2 fold change (FC) of the top 10 genes differentially expressed in cellulose, glucose and sophorose

Condition	Protein ID	Name	GO term	FC Cel/Glu	FC Sph/Cel	FC Sph/Glu
Cellulose	69957	MFS permease	Cellular component	11.001	−3.673	0
	56996	GH5 β-Mannanase MAN1	Biological process	10.849	−5.403	0
	69944	GH31 α-xylosidase/α-glucosidase	Biological process	10.500	−2.007	0
	73632	CE5 acetyl xylan esterase AXE1	Molecular function	8.060	−1.025	0
	108642	Unknown protein	Unknown	3.644	−8.621	0
	112258	Unknown protein	Unknown	4.665	−8.069	0
	124079	Multicopper oxidases	Molecular function	3.787	−7.762	0
	123236	SSCRP	Unknown	2.666	−7.713	0
	119552	Unique protein	Unknown	3.884	−7.090	0
	55886	GH16 glucan endo-1,3(4)-β-D-glucosidase	Unknown	4.086	−7.067	0
Glucose	79816	Unknown protein; secreted	Molecular function	−5.793	0	−9.354
	70520	short chain dehydrogenase/reductase	Molecular function	−4.333	0	−9.263
	23382	Aldehyde reductase AKR7	Molecular function	−4.028	0	−9.213
	30759	Zinc-containing alcohol dehydrogenase superfamily	Molecular function	−3.684	0	−8.508
	123084	Chloroperoxidase	Molecular function	−4.587	0	−8.369
	122998	Unknown protein	Unknown	−3.640	0	−7.637
	69115	Dienelactone hydrolase	Unknown	−4.827	0	−7.496
	81525	Isoflavone reductase	Molecular function	−4.430	0	−7.448
	81586	Unknown protein	Unknown	−2.903	0	−7.338
	76641	MFS permease	Cellular component	−4.454	0	−7.309
Sophorose	106164	short chain dehydrogenase/reductase	Molecular function	0	2.026	10.007
	59628	Unknown protein	Unknown	0	3.846	5.711
	48444	MFS maltose permease	Cellular component	0	2.109	5.524
	5345	FAD-containing oxidoreductase	Molecular function	0	2.579	5.443
	122087	Unknown protein	Cellular component	0	2.898	5.225
	21876	Zinc-binding oxidoreductase	Molecular function	0	5.034	3.828
	22915	Glucose oxidase	Molecular function	0	1.156	4.544
	110267	Unknown protein	Biological process	0	4.486	1.287
	60945	MFS permease	Cellular component	0	1.341	4.451
	55802	GH76 α-1,6-mannanase	Molecular function	0	1.956	4.441

GO, Gene Ontology; MFS, Major facilitator superfamily; GH, glycoside hyrolase; CE, carbohydrate esterase; AXE, acetyl xylan esterase; SSCRP, Small secreted cysteine-rich protein; Cel, cellulose; Glu, glucose; Sph, sophorose.

a cooperative manner or have a more effective mode of action. In addition, the gene for PacC (pH-responsive transcription factor) was regulated in a carbon source manner, showing a higher expression level in the presence of cellulose (Table 2). Other TFs that have been shown to have a regulatory role, such as HAP2/3 and AreA, showed no significant modulation in gene expression in any condition ($\log_2 > 1$ and $P < 0.05$). Taken together, our results depict a complex system of TFs that regulate the expression of hydrolytic enzymes, while also revealing additional, uncharacterized, TFs that appear to play a role.

Transporters

Genes that encode proteins involved in transport comprise about 5% (459 genes) of the *T. reesei* genome. Our results show that among these genes, 14 were regulated exclusively by cellulose, 14 by sophorose, and 30 by glucose, applying an adjusted *P*-value <0.05 as thresholds (Additional file 6: Table S6).

The MFS (Major facilitator superfamily) permeases are the most abundant proteins in the three analyzed conditions. These proteins enable the transport of essential nutrients and ions, plus the excretion of end products of metabolism and cell-environment communication [23].

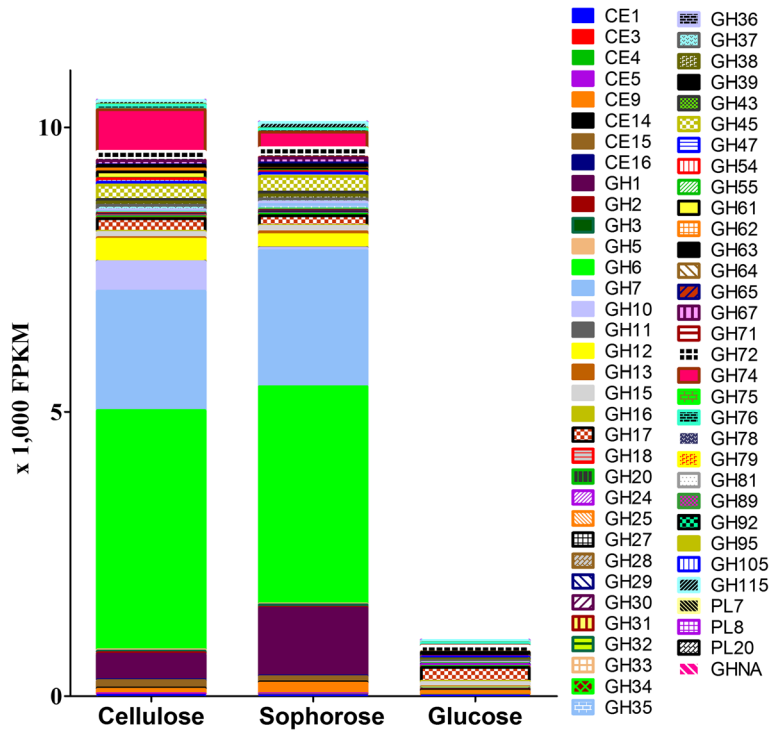


Figure 4 Carbohydrate active enZymes (CAZy) genes and expression data from RNA-seq analysis. Fragments per kilobase of exon per million fragments mapped **(**FPKM) means for each glycolic hydrase (GH) family when cultured in glucose, cellulose and sophorose. CAZy classification was performed based on re-annotation of CAZy genes of *T. reesei* according to Hakkinen *et al*. [9].

The gene encoding for the MFS permease (Trire2_69957) that was specifically highly upregulated in cellulose may be involved in the transport of disaccharides, due to a high similarity with a putative maltose permease of the human pathogenic fungus *Talaromyces marneffei* [24]. Another maltose permease encoding gene (Trire2_48444) was also highly induced by sophorose. Conversely, the MFS permease gene Trire2_76641 was expressed at a higher level in glucose than on sophorose or cellulose (Table S6). A BlastP analysis of this MFS permease showed 85% sequence identity to a synaptic vesicle transporter SVOP and also shared structural similarity to the human glucose transporter 1 (Glut1) [25]. Interestingly, a gene encoding a potential galactose permease (Trire2_62380) that was specifically expressed in glucose and a MFS permease encoding gene (Trire2_76800) that was induced by cellulose, both resembled the 19 *Saccharomyces cerevisiae* transporters that when deleted, contribute to the total loss of hexose uptake [26].

Table 2 Log_2 fold change of characterized transcriptional factor genes involved in the regulation of cellulase and hemicellulase genes

Protein ID	Name	Cel/Glu	Soph/Cel	Soph/Glu
122208	XYR1	6.062	2.087	8.131
78445	ACEII	NS	0.069	NS
27600	CLR-1	2.263	0.516	2.766
26163	CLR2	NS	NS	NS
52368	BglR	1.492	NS	1.570
120117	ACEI	0.807	−0.674	NS
120117	CREI	0.807	−0.674	NS
124286	HAP2	NS	0.455	NS
121080	HAP3	NS	NS	NS
76817	AreA	NS	0.458	NS
120698	PacC	2.123	−0.538	NS

NS, non-significant at P <0.05; Cel, cellulose; Glu, glucose; Soph, sophorose

In order to identify the MFS permeases shared by cellulose and sophorose, the expression results were normalized with the glucose condition (Table 3). From 85 MFS permeases annotated in the *T. reesei* genome, 22 of them seem to be shared by cellulose and sophorose (Table 3). Among them, the most expressed were: *crt1*, which has been shown to be required by *T. reesei* for growth in cellulose and lactose, but not in xylan [27]; *hxt1*, a glucose permease; the MFS gene Trire2_50894, a high affinity glucose transporter [28]; and an MFS gene related to cellulose signaling (Trire2_79202) [29]. Interestingly, the recently described *stp1*, which is involved in cellobiose and glucose transport [27], showed a higher level of expression in sophorose than in cellulose (Table 3),

Table 3 The log_2 fold change of sharing MFS permeases genes in cellulose and sophorose conditions

Protein ID	Description	Cellulose	Sophorose
3405	MFS permease (Crt1)	7.802	9.893
22912	MFS permease (glucose permease HXT1)	7.142	6.368
50894	MFS permease	6.683	7.069
79202	MFS permease, associated with cellulose signalling	5.742	7.646
44175	MFS H + sugar transporter	4.413	4.102
121482	MFS permease	4.364	3.887
67752	MFS permease	4.085	3.919
68812	MFS permease	3.921	3.002
104549	MFS permease	3.871	4.482
47710	MFS permease (Stp1)	3.629	5.065
78833	MFS permease (fucose permease)	3.340	4.967
62502	MFS permease	2.971	3.137
54632	MFS permease	2.841	2.940
80026	MFS permease	2.718	3.073
50618	MFS permease	2.439	2.932
119789	MFS permease	2.135	2.055
106330	MFS permease	1.875	3.465
67334	MFS permease	1.740	2.234
61350	MFS permease	1.665	2.995
65153	MFS permease	1.663	2.731
68813	MFS permease	1.392	1.102
64874	MFS toxin efflux pump	1.155	1.017

indicating a complex regulation on cellobiose/sophorose uptake by *T. reesei*.

Another family of proteins that showed carbon source-dependent transcriptional regulation were the ABC (ATP binding cassette) transporters, which were highly upregulated in cellulose and sophorose. The AAA family (ATPases associated with a variety of cellular activities) and aquaglyceroporin genes were highly expressed in sophorose, whereas the ADP/ATP carrier genes were highly expressed in glucose (Additional file 6: Table S6). In addition, amino acids, oligopeptide, and ion transporter genes were identified as being regulated by the three carbon sources, with a larger number of genes expressed in cellulose.

Deciphering the regulatory network of *T. reesei* in response to cellulose/sophorose

Using the experimental setup described above, we were able to identify a specific set of genes differentially regulated by the analyzed carbon sources. Using these data, the regulatory network of the genes identified as being modulated in a carbon source-dependent manner was reconstructed (Figure 5). Extensive overlapping between the differentially expressed genes in cellulose and sophorose (710 genes) was observed. Additionally, genes specifically associated with each condition were identified, as exemplified by the large number of genes (441) whose expression was specifically modulated during growth in sophorose compared to glucose. These genes represent either genes silenced in glucose but induced by sophorose (upregulated: 154 genes), or genes that are necessary for growth in glucose but dispensable for growth in sophorose (downregulated, 287 genes). Accordingly the analysis of genes specific to cellulose showed an over-representation of upregulated genes (132 genes) that are related to the expression of cellulase genes as compared to glucose (201 in total) (Figure 5).

The comparison between the two inducing conditions provided additional information by revealing new differentially expressed genes that were not identified via the comparison with glucose. From the 692 genes differentially expressed between cellulose and sophorose, only 75 and 107 genes were assigned specifically to sophorose or cellulose respectively (Figure 5).

The majority of the genes identified from this network analysis were of unknown function. Importantly, a cellulose- or sophorose-specific enrichment of different gene classes was observed (Figure 6). During growth in cellulose, there was an enrichment of CAZy encoding genes (that is, GH64, GH 62, GH81, GH76, GH54), accessory proteins (Small secreted cysteine-rich protein (SSCRP), OOC1, and Epl1), transporters (most of them related to iron and metal transporters), TFs (*lae1*, C2H2 and Zn2Cys6 TFs) and a variety of proteins related to electron transport (Table S7). In contrast, there were only three CAZy encoding genes specific to growth in sophorose. However, four genes encoding *Trichoderma* species-specific proteins were only induced on sophorose, suggesting that *Trichoderma* possesses a specialized sophorose metabolism system (the complete list of differentially expressed genes is shown in Additional file 7: Table S7). The substantial overlap between the cellulose and sophorose transcriptomes supports the hypothesis that sophorose is a natural inducer of cellulase transcription, while the cellulose-specific enrichment for additional CAZymes and accessory proteins reflects the difficulty in the deconstruction of this insoluble substrate.

Quantitative real-time PCR (RT-qPCR) analysis

The RNA-seq data were validated using 20 genes with mRNA accumulation that was modulated when the following comparisons were performed: cellulose versus glucose; sophorose versus cellulose; and sophorose versus glucose. The 10 upregulated genes were predominantly glycoside hydrolases and 10 downregulated genes were randomly chosen (see Additional file 8: Table S8).

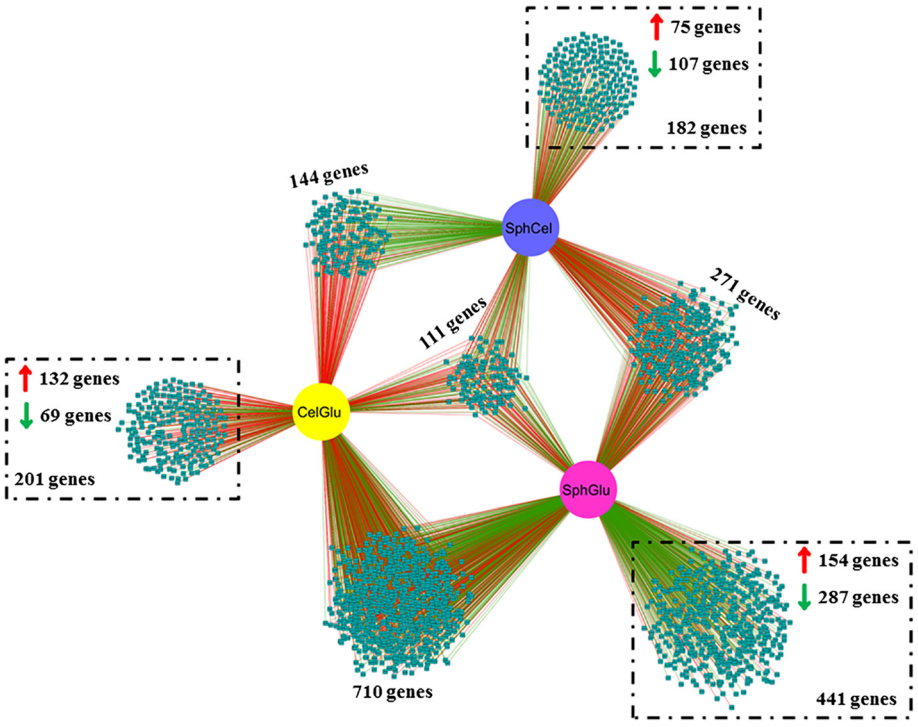


Figure 5 Gene regulatory network (GRN) of 2,060 differentially expressed genes in *T. reesei* QM9414 in each tested condition. Cellulose versus glucose (CelGlu), sophorose versus cellulose (SphCel) and sophorose versus glucose (SphGlu). Genes are represented as nodes (shown as squares), and interactions are represented as edges (shown as lines, that is, red indicates upregulated interactions and green indicates downregulated interactions), that connect the nodes: 3,385 interactions.

The $\log_2$ fold change in gene expression between the three comparisons obtained by RNA-seq and RT-qPCR demonstrated significant Pearson correlation ($r^2 = 0.8882$), indicating the reliability of the RNA-seq analysis (Figure 7).

Secretome analysis by two-dimensional DIGE

The *T. reesei* secretome when grown in glucose, sophorose and cellulose were analyzed by quantitative proteomics (two-dimensional DIGE), followed by liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis. The gels shown in Figure 8 are representative of all three independent gels and three biological replicates. The distribution of the spots indicates that most of the secreted proteins have isoelectric points <6.0 and a molecular weight >30 kDa. In some cases, the molecular weights and isoelectric points observed in the two-dimensional gels were higher than expected, probably due to post-translational changes. Another observation was that various different spots were assigned to the same protein, suggesting the presence of a number of isoforms

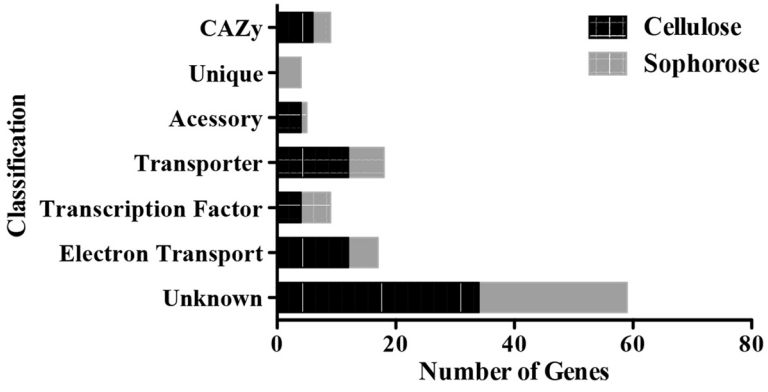


Figure 6 Gene Ontology (GO) enrichment analysis of different classes of genes upregulated in cellulose and sophorose in *T. reseei*. Significantly enriched categories ($P \leq 0.05$) are shown. The complete list of differentially expressed genes is shown in Additional file 7: Table S7.

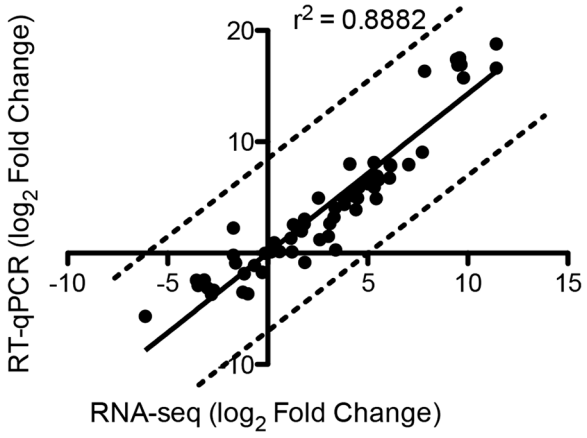


Figure 7 Correlation between RNAseq and quantitative real-time PCR (RT-qPCR). Comparison of $\log_2$ fold change of 20 genes obtained by RNA-seq and RT-qPCR. Real-time PCR was performed using the amplified cDNA from each RNA-seq sample. Strong, statistically significant Pearson correlation is shown between the expression levels measured using real-time PCR and RNA-seq.

or possibly degraded forms of the protein (Table S9.1 and S9.2, see Additional file 9 and Additional file 10, respectively).

The comparative analysis between the cellulose and glucose secretomes showed a total of 170 spots automatically detected by the software PDQuest (BioRad). Among these, 130 spots were statistically validated using a differential abundance ratio ≥2.0-fold (P ≤0.05). In total 89 spots were exclusively expressed in cellulose (Figure 8C) however, only 36 spots could be identified by mass spectrometry (MS) (Table S9.1). The identified proteins included classical cellulases (CEL7A and CEL6A), such as β-1-3-glucanosyltransferase, α-L-arabinofuranosidase, β-1,3-glucanase, α-1,2-mannosidase, Xylanase 4, β-xylosidase and isoforms, as well as a non-hydrolytic CIPI (Cellulose-induced protein). Furthermore, some proteases were also identified as being expressed during growth in cellulose, as well as an protein of unknown function (Trire2_55887) and a novel protein isoamyl alcohol oxidase (Trire2_73631), (see Additional file 9: Table S9.1). On the other hand, 41 spots were exclusively expressed during growth in glucose (Figure 8C). Among them, 28 spots were identified by MS (Additional file 9: Table S9.1): an acid phosphatase-like protein, various isoforms of isoamyl alcohol oxidase, subtilisin-like protease PPRC1, cell-wall glucanosyltransferase, amidase, a cerato-platanin, Epl1/Sm1 as well an SSCRP, and a unique protein (Trire2_121136).

When the sophorose secretome was compared to cellulose, the software PDQuest detected 183 spots, with 30 spots being exclusive to sophorose (≥2.0-fold; P ≤0.05), 37 exclusive to cellulose (≥2.0-fold; and P ≤0.05), and 116 spots common between the two conditions (Figure 8D). After MS analysis 20 spots from sophorose and 31 spots from cellulose were identified (see Additional file 10: Table S9.2). Among the enzymes classified as glycoside hydrolases, families GH3 (BXL1 β-xylosidase, β-glucosidase BGL1/CEL3a), GH64 (endo-β-1,3-glucanase), GH7 (Cellobiohydrolase CBH1/CEL7a, Endo-β-1,4-glucanase EGL1/CEL7b), GH72 (β-1-3-glucanosyltransferase), GH17 (glucan

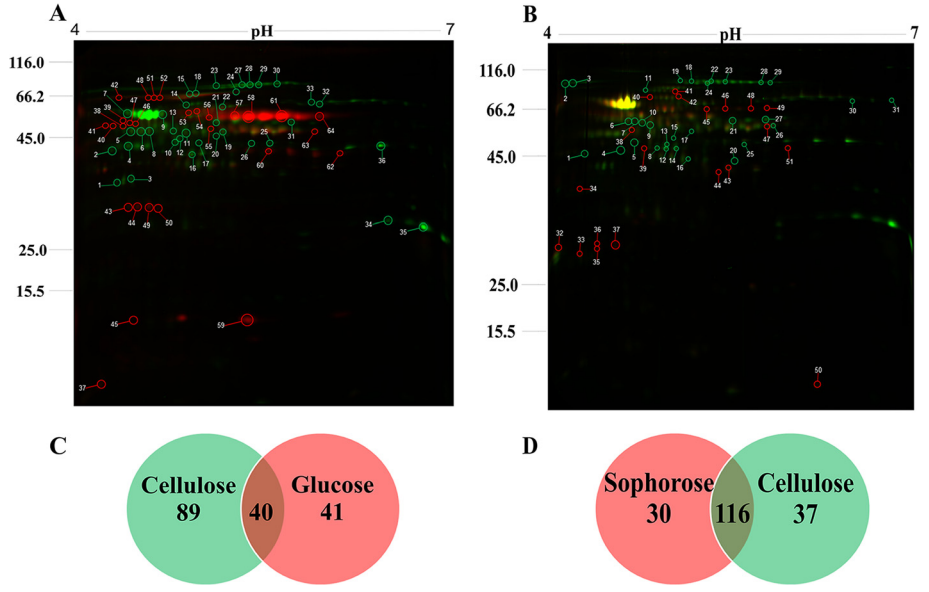


Figure 8 Differential gel electrophoresis (DIGE) analysis of *T. reesei* secretome grown in different carbon sources. (A) DIGE of cellulose (green spots) versus glucose (red spots). **(B)** DIGE of cellulose (green spots) versus sophorose (red spots). **(C)** Venn diagram from analysis of cellulose versus glucose. **(D)** Venn diagram from analysis of cellulose versus sophorose. The numbers in white indicate the spots subjected to liquid chromatography tandem mass spectrometry analysis. Protein IDs of identified spots are listed in Additional files 9 and 10: Table S.9.1 and S9.2).

endo-1,3-β-glucosidase), GH28 (exo-rhamnogalacturonase RGX1), GH61 (polysaccharide monooxygenase CEL61a), GH74 (Xyloglucanase CEL74a), and GH30 (endo-β-1,4-xylanase XYN4) were found to be exclusive in cellulose cultures, whereas only the GH13 (α-amylase) was detected in sophorose. Furthermore, a larger number of proteins not related to glycoside hydrolase functions were detected in sophorose than in cellulose, such as SSCRP, lipoate-binding, ribonuclease T2, ubiquinol cytochrome reductase, DNAse, cell wall glucanosyltransferase, CIP1, ceramidase family protein, amidase, isoamyl alcohol oxidase, and ubiquitin fusion protein. Another interesting feature of the secretome was that both unknown proteins and *Trichoderma* species-specific proteins with described functions were found in both cellulose and sophorose, demonstrating the complexity and uniqueness of the *Trichoderma* secretome during cellulose degradation.

In order to correlate the gene expression data with the secretome, the fold change from cellulose versus glucose in both datasets was compared. There was 70.17% correlation between gene expression and secreted protein profiles (see Additional file 9: Table S9.1). When cellulose and sophorose cultures were compared, lower correlation was observed (47%), possibly due to many non-significant differences in expression (P <0.05) in RNAseq analysis (see Additional file 10: Table S9.2).

Discussion

In natural environments, free-living organisms are continuously challenged with rapidly changing conditions that have a considerable impact on their lifestyle. Genomic and post-genomic techniques have revealed that free-living organisms dedicate a large percentage of their genes to sensing environmental signals and the subsequent coordination of gene expression in response to such cues. How the fungus *T. reesei* recognizes its substrate and activates the transcription of genes encoding transporters and TFs that culminate in the production of hydrolytic enzymes has been a subject of speculation since the 1960s. By using high-throughput genomic and proteomic approaches we describe both repressing (glucose) and de-repressing (cellulose or sophorose) conditions, identifying new players in cellulose degradation in *T. reesei*. In addition, the comparison between cellulose and sophorose, the hypothesized natural inducers of cellulase production, revealed a striking similarity in the global profiles.

The transcriptome study of *T. reesei* identified 123 genes that were specifically induced by cellulose, 154 by sophorose and 402 by glucose (Figure 3). Within these gene sets, 8 permease/transporter genes were induced in cellulose, 6 in sophorose and 11 in glucose (see Additional file 3: Table S3 and Additional file 8: Table S8 respectively). Of these 25 transporters, 10 showed possible homology to *N. crassa* homologues and one of them, MFS permease (Trire2_76800) (highly induced in the presence of cellulose), allowed *S. cerevisiae* to transport xylose [30,31]. Furthermore, a gene encoding a putative galactose permease (Trire2_62380) found in glucose and another MFS permease (Trire2_76800) regulated by cellulose, showed similarity to *S. cerevisiae* transporters involved in hexose uptake [26]. Additional transporters were induced by both cellulose and sophorose, suggesting that sophorose could be the natural inducer of cellulase gene transcription in *T. reesei*. Despite that, the functions of these transporters in *T. reesei* remain obscure. For instance, the transporter Trire2_3405 was recently identified to be specifically involved in cellulase induction by lactose [29], but has also been described as being involved in cellobiose transport [27]. Furthermore, the same transporter was upregulated during growth on wheat straw [22,28], cellulose or sophorose (Table 3). This lack of specificity by transporters could be explained by the close structure of cellobiose/lactose/sophorose or by the fact that some transporters can act as transporters and nutrient sensors. However, more detailed studies will be needed to characterize these transporters and generate a better understanding of the inducer/repressor transport system in *T. reesei.*

Global gene expression analysis by RNA-seq enabled the construction of gene regulatory networks (GRN) that enhanced the understanding of the interaction between different genes during the degradation and metabolism of cellulose. Studies on the control of catabolic genes related to the metabolism of simple substrates (such as glucose) performed in model organisms have revealed very complex GRN, thus, an even more sophisticated network controlling catabolic functions related to the metabolism of complex substrates, such as cellulose, could have been anticipated [32]. In the *T. reesei* model for cellulose degradation, the deep knowledge of the catabolic activities related to cellulose metabolism is accompanied by a very limited understanding of the regulatory pathways responsible for controlling gene expression [22,32,33]. In fact, despite the TFs, XYR1 and CRE1 [34,35], which regulate the induction or repression of the cellulolytic enzymes respectively, and a few more specific regulators (ACE1, ACE2, BGLR) [36,37] that have been experimentally characterized, there remains a lack of information on how, and to what extent, the expression of these enzymes are connected to the core GRN of *T. reesei* [32]. This is important as GRN in free-living organisms are usually densely connected and the final decision on the expression of a particular gene set is generally controlled by many different external/internal signals [38]. The collection of *omics* data provided here tries to fill this gap by providing a global analysis of *T. reesei* grown in three different substrates (cellulose, sophorose and glucose). From the analysis provided, we

started building a bona fide regulatory network for this organism through the identification of 43 TF genes specifically induced in some particular growth conditions (see Additional file 5: Table S5). The GRN revealed that some of factors are exclusively induced in response to cellulose or sophorose (see Additional file 7: Table S7). For instance, the methyltransferase LAE1 has already been described as controlling the expression of cellulases, auxiliary factors for cellulose degradation, β-glucosidases and xylanases [39], proteins commonly found in response to inducers, cellulose, lactose and wheat straw [22,28]. However, our results showed that LAE1 is preferentially expressed in response to cellulose, indicating that the fungus has specific signaling for the metabolism of cellulose. This hypothesis is supported by the fact that recent study showed that LAE1 affects other components of cellulose degradation, such as non-ribosomal peptide synthases, ankyrin-repeat proteins, iron uptake, PTH11-receptors, and oxidases/monoxygenases [40], genes that were also upregulated in the presence of cellulose in our data and in the presence of wheat straw [22]. Another TF gene upregulated in response to cellulose (Trire2_120698) showed homology to the *Aspergillus nidulans* pH-responsive transcription factor *pacC*. It is known that this TF controls a range of functions in filamentous fungi [41]. Although studies have shown that pH is involved in cellulase production in *T. reesei* [42], the regulation of cellulase genes by any pH-responsive TF is still unknown.

The expression level of the *cre1* gene was low even in the presence of glucose. One explanation for this result is the fact that some TFs can act either directly on CAZyme encoding genes or indirectly by regulating other TFs that in turn regulate the expression of CAZyme genes. Here we identified some TF genes that are candidates for the indirectly transcriptional regulation, in a carbon source-dependent manner (see Additional file 5: Table S5). Some of these TFs could play an important role in the coordination of gene expression downstream in the network, either in association with the previously identified general factors at the target promoters or in isolation, in a sort of cascade signaling pathway. Additionally, the identified TFs could work as check points for the integration of different physiological/environmental signals, such as metabolic status of the cell, levels of light, presence of stresses, et cetera. [43,44]. The TFs identified here are candidates for further investigation into the mechanisms of signal integration in this biotechnologically relevant fungus. Understanding these missing regulatory interactions is pivotal for future attempts to synthetically engineer *T. reesei* for enhanced cellulolytic functions.

Analyses of the *T. reesei* secretome has commonly focused on growth in cellulose or lactose [45-47]. Besides the classical cellulases already described, our differential secretome showed the presence of polysaccharide monooxygenase, xyloglucanase CEL74a, and xylanases, induced by cellulose, whereas in sophorose, amidase, amylase and isoamyl alcohol oxidase they were described for the first time. The strong correlation between transcriptome and secretome data in the presented study is consistent with other comparable studies [29,48,49]. Furthermore, a comparison of the cellulose and sophorose transcriptome and differential secretome data did not detect a massive difference in any analyzed category of proteins. This observation suggest that the signaling for cellulose and sophorose to induce cellulase formation is very conserved and thus sophorose still remains a strong candidate as natural inducer.

Despite extensive work related to the regulation of cellulases in *T. reesei*, the real identity of the natural inducer is not yet established. New evidence has recognized cellobiose and cellodextrins as strong candidates for natural inducers [50]. Indeed, studies with *N. crassa* [51] and *A. niger* [52] have discredited sophorose as the natural inducer. It is known that *T. reesei* possesses a different mechanism for the regulation of cellulase production in response to sophorose when compared to other lignocellulose-degrading fungi [5]. Our GRN data showed little differences in the regulation of gene expression by the inducers cellulose and sophorose, suggesting that sophorose could be a natural cellulase inducer. But how did this divergence between *T. reesei* and other fungi occur? Comparative genomics between *T. atroviride, T. virens* and *T. reesei* suggest that the ancestral state of *Hypocrea*/*Trichoderma* was indeed a mycoparasitic, possibly of wood-degrading basidiomycetes [5]. *T. reesei* subsequently may have kept the mycoparasitic characteristic for substrate competition, converting cellobiose to sophorose by a transglycosylation reaction and then metabolizing sophorose. This hypothesis can be supported by the fact that new species-specific proteins were upregulated only in sophorose and by the fact that cellobiose and sophorose are transported and metabolized at different rates [50]. For this reason, we propose that both cellobiose and sophorose act as co-inducers of cellulase formation in *T. reesei*. These facts could explain why among lignocellulose-degrading fungi, *T. reesei* is the more efficient degrader, despite its smaller enzymatic arsenal.

Conclusions

Our study shows little difference between gene expression and the secretome during the growth of *T. reesei* in cellulose and sophorose. The difference in gene expression is associated with CAZymes, accessory proteins, transporters, TFs, and electron transport. Together with recent literature, the results shown here suggest that both cellobiose and sophorose act as co-inducers of cellulase production in *T. reesei*. Further functional genomic investigations of the new players identified to be

involved in growth in cellulose will open up new lines of research into clarifying cellulase and hemicellulase regulation in *T. reesei*. In addition, the data shown in this study will contribute to the construction of industrial strains of *T. reesei* that produce high levels of cellulase for plant cell-wall degradation thus facilitating its application in 2G-bioethanol production.

Methods

Strain and growth conditions

T. reesei strain QM9414 (ATCC 26921) was obtained from the Molecular Biotechnology Laboratory, Institute, TU Vienna, Austria. The strain was maintained on MEX medium (malt extract 3% (w/v) and agar-agar 2% (w/v)) at 4°C. QM9414 was grown on MEX medium at 28°C for 7 to 10 days to complete sporulation. For gene expression assays, a spore suspension containing approximately 10^7 cells mL^{-1} was inoculated into 200 mL of Mandels-Andreotti medium [53] containing 1% (w/v) of cellulose (Avicel), or 2% (w/v) of glucose, or 1 mM of sophorose, as the sole carbon source. The cultures were incubated on an orbital shaker (200 rpm) at 28°C for 24, 48 and 72 hours using cellulose; for 24 and 48 hours with glucose; and 2, 4 and 6 hours with sophorose, as the carbon source. In the latter, the mycelium was previously grown on glycerol 1% (w/v) for 24 hours. After this time, the mycelium was washed with Mandels-Andreotti medium without peptone and then transferred to 20 mL of Mandels-Andreotti medium without peptone containing sophorose 1 mM. All experiments were performed in three biological replicates. The resulting mycelia were collected by filtration, frozen and stored at −80°C until RNA extraction and the supernatants were used for secretome analysis.

RNA extraction

Total RNA was extracted from mycelia of each sample using TRIzol® RNA kit (Invitrogen Life Technologies, Carlsbad, CA, USA), according to the manufacturer's instructions. RNAs concentrations were determined by spectrophotometric OD 260/280 and RNA integrity was verified by both the Agilent 2100 Bioanalyzer (Agilent Technologies, Waldbroon, Germany) and gel electrophoresis in 1% agarose.

High-throughput sequencing (RNA-seq)

Total RNA of three biological replicates, cellulose (24, 48 and 72 hours), sophorose (2, 4 and 6 hours) and glucose (24 and 48 hours) were time points that were pooled, resulting in nine samples for the preparation of next-generation sequencing libraries using the TruSeq RNA Sample Prep kit (Illumina, San Diego, CA, USA). The total RNA samples obtained from *T. reesei* were lyophilized and stored using the RNAstable tube kit (Biomatrica, San Diego, CA, USA) in order to maintain the RNAs integrity for sequencing. Nine barcoded libraries (cel1-3, gluc 1–3 and soph 1–3) were prepared and sequenced by LGC Genomics GmbH (Berlin/Germany) using the Illumina Hiseq 2000 platform.

Data analysis

The Illumina Hiseq 2000 system was used to sequence approximately 117 million 100 bp paired-end reads. These sequences were quality-filtered and mapped to the *Trichoderma reesei* 2.0 reference genome, available from the JGI Genome Portal (http://genome.jgi-psf.org/Trire2/Trire2.home.html), using the Bowtie aligner version 0.12.8 [54], allowing for two mismatches and only unique alignments. After alignment, Samtools version 0.1.18 [55] was used to process the alignments files, which were visualized using the Integrative Genomics Viewer [56]. The genes were annotated using *Trichoderma reesei* 2.0 reference genome and a local database provided by Professor CP Kubicek (TU, Vienna). Unknown proteins were defined as proteins that have yet to be assigned a function in any ascomycete and *T. reesei* species-specific proteins were defined to be proteins that did not occur in any other Pezizomycotina [57]. Bioconductor DESeq package version 1.10.1 [58] was utilized for the differential expression analysis, using two-fold change cutoff, that is, $\log_2$ fold change ≥1 or ≤ −1 and an adjusted *P*-value ≤0.05 as thresholds. Samples were normalized using median log deviation DESeq, available in the Bioconductor package. Cluster analysis was carried out using the software Mev v.4.6.1 to identify cellulose, sophorose and glucose regulons. The average linkage method was used for cluster generation, with uncentered correlation as the similarity metric. Functional enrichment analysis of differentially expressed genes was performed using GO terms was performed using the BayGO algorithm [59]. GO terms significantly enriched, (that is, with *P*-values ≤0.05) were analyzed further. Raw sequence data and count data for all samples are available at [GEO: GSE53629]. CAZy classification was performed based upon the re-annotation of CAZy genes of *T. reesei* according to Hakkinen *et al.* [9].

Regulation network of *T. reesei*

In order to reconstruct the regulatory network of *T. reesei* under the experimental condition analyzed, a table using the following information was generated: inducing condition (QMCelGlu, QMSphCel and QMSphGlu, selecting differentially expressed genes, up- and downregulated in each condition, $P \leq 0.05$), the interaction type (up- or downregulated) and the target gene (that is, the protein ID of each gene affected). This analysis provides a network representation for all the genes (2,060 in total) shown in the heat map of Figure 2. The regulatory network was then generated using the Cytoscape 3.0.1 software [60].

Quantitative qRT-PCR analysis

Differentially expressed genes identified by the RNA-seq analysis were further analyzed by qRT-PCR in order to validate their expression. In this analysis, the same RNA samples, utilized for the RNA sequencing experiments were re-used. Approximately, 1 μg of RNA was treated with DNAseI (Thermo scientific) and reverse-transcribed to cDNA using the First Strand cDNA kit Maxima™ Synthesis according to manufacturer's instructions. The cDNA was diluted to 1/50 fold and used for real-time PCR analysis in the Bio-Rad CFX96™ System, using SsoFast™EvaGreen®Supermix (Bio-Rad, San Francisco, CA, USA) for signal detection in accordance with the manufacturer's instructions. Genes encoding actin (*act*) and a small GTPase SAR/ARF-type (*sar1*) were used as endogenous controls according to [61]. Twenty genes, including up- and downregulated genes in cellulose compared to glucose samples (see Additional file 11: Table S1), were used for qRT-PCR analysis. The following amplification reaction was used: 95°C for 10 minutes followed by 39 cycles of 95°C for 10 seconds, 60°C for 30 seconds followed by a dissociation curve of 60°C to 95°C with an increment of 0.5°C for 10 seconds. Gene expression values were calculated according to the $2^{-\Delta\Delta CT}$ method [62] using the QM9414 strain growth on glucose as the reference sample. Data analysis was performed using GraphPad Prism v 5.1 software.

Sample preparation for proteomic analysis

The protein concentration was determined using the kit Bio-Rad Protein Assay, based on the Bradford method. Protein concentration was adjusted to 1 μg /μl, and 150 μg used for in two-dimensional DIGE and 300 μg in two-dimensional SDS-PAGE. Samples were precipitated using 10% tricarboxylic acid (TCA) in acetone and incubated at −20°C overnight. Samples were centrifuged at 10,000 g for 10 minutes at 4°C and the supernatant removed. β-mercaptanol was added (0.07%) in acetone and centrifuged at 10,000 g for 10 minutes at 4°C. This was repeated three times, discarding each supernatant after centrifugation. After precipitation, the pellet was purified using Ettan2D Clean-Up Kit (GE Healthcare, Waukesha, WI, USA).

Two-dimensional differential gel electrophoresis

The proteins (150 μg) secreted by *T. reesei* under different conditions were labeled with 400 pmol CyDyes (Cy3 or Cy5) according to the manufacturer's instructions (GE Healthcare, Waukesha, WI, USA). An internal pool generated by equal amounts of all samples was labeled with Cy2. The isoelectric focusing was carried out on 18-cm linear IPG strips, pH 4–7, with the addition of 1.2% DeStreak and 1% IPG buffer 4–7 (GE Healthcare). Isoelectric focusing was performed on IPGphor III in four steps: 500 V for 60 minutes, 1000 V for 60 minutes, 8000 V for four hours and 8000 V for six hours. The strips were reduced (1.5% w/v dithioerythritol) and alkylated (2.5% w/v iodocetamide) in equilibration buffer (6 M urea, 50 mM Tris–HCl, pH 6.8, 30% glycerol, 2% SDS). Equilibrated strips were run on homogeneous 12.5% polyacrylamide gels using an Ettan DALTsix electrophoresis (GE Healthcare). All the experiments resulted in three independent replicates for each experimental condition. The preparative gels were stained using colloidal Coomassie and destained with Milli-Q water to remove excess Coomassie particles. Gels were scanned using the laser scanner Pharos FX Plus (Bio-Rad) and Quantite One software (Bio-Rad) using a resolution of 100 μm and the appropriate wavelength. The images were analyzed with the software PDQuest Advanced 2-D Analysis Software (Bio-Rad). Differential expression was determined by statistical analyses using the *t*-test, as the parameter of significance ($P \leq 0.05$).

Protein identification by mass spectrometry

Spots which increased or decreased in volume (protein content) by two-fold or more were manually excised from the gels and washed four times with 50 mM NH_4HCO_3 containing 50% v/v acetonitrile (ACN) to remove SDS and dye. They were then washed with ACN and completely dried in a SpeedVac (Savant Instrument, Farmingdale, NY, USA). Each spot was rehydrated with 20 μl 50 mM NH_4HCO_3 containing 0.3 μg of sequencing grade modified trypsin (Promega, Madison, WI, USA). After 30 minutes of rehydration with the trypsin solution, spots were covered with 50 mM NH_4HCO_3. The hydrolysis reaction was carried out at 37°C for 24 hours and stopped by the addition of 10 μl formic acid 1%. Peptides were extracted twice from the gel with 40 μl 0.1% v/v formic acid solution containing 50% v/v ACN for 1 hour. Extracts were dried in a SpeedVac and resuspended in 35 μl 0.1% v/v formic acid solution containing 5% v/v ACN for MS injection. Samples were then analyzed in an XEVO-TQS mass spectrometer (Waters) coupled with a UPLC chromatography system (Waters). Liquid chromatography separation was performed in a 15 cm column (ACQUITY UPLC HSS C18, 100 Å, 1.8 μm, 1 mm × 150 mm, Waters) using a 30-minute linear gradient from 5 to 30% of ACN in 0.1% formic acid at 150 μl/minute. The spectra were acquired in a data-dependent mode in an m/z range of 400 to 1,500, with selection of the two most abundant ions of each MS spectrum for MS/MS analysis. MS parameters were as follows: capillary voltage of 3.5 KV and capillary temperature of 400°C. Acquired raw data were converted to mzXML and automatically processed by an in-house installation of Labkey Server v12, using theX!Tandem search algorithm [63]. The minimum criterion for

peptide matching was performed using a Peptide Prophet [64] score greater than 0.8. Peptides that met these criteria were further grouped to protein sequences using the Protein Prophet [65] algorithm and only proteins with an error rate of 5% or less and two peptides sequences identified were considered as valid identifications.

Additional files

Additional file 1: Table S2. Summary RNA-seq reads obtained (Illumina Hiseq 2000) in this study.

Additional file 2: Figure S1. Biological replicates used for the RNA-seq analysis. **(A)** Graphs representing the Pearson correlation between biological replicates of each sample. **(B)** Principal component analysis (PCA) of the samples analyzed. **(C)** boxplot of all normalized samples and **(D)** boxplots of raw data.

Additional file 3: Table S3.1. Genes from cellulose regulon. **Table S3.2.** Genes from glucose regulon. **Table S3.3.** Genes from sophorose regulon protein.

Additional file 4: Table S4A. Carbohydrate active enzyme (CAZy) genes that are upregulated in sophorose. **Table S4B.** CAZy enzymes that are upregulated in celulose. **Table S4C.** CAZy enzymes that are upregulated in glucose.

Additional file 5: Table S5. The main transcription factors genes induced in presence of cellulose, sophorose and glucose.

Additional file 6: Table S6. Upregulated transporters genes in presence of cellulose, sophorose and glucose. Values are expressed in log2 fold change.

Additional file 7: Table S7. Differentially expressed genes in cellulose and sophorose. Numbers are expressed as $\log_2$ fold change.

Additional file 8: Table S8. Comparison of the gene expression levels assayed by RNA-seq and RT-qPCR. The numbers highlighted in red did not correlate.

Additional file 9: Table S9. Identified proteins from differential gel electrophoresis (DIGE) analysis between cellulose and glucose.

Additional file 10: Table S10. Identified proteins from differential gel electrophoresis (DIGE) analysis between cellulose and sophorose.

Additional file 11: Table S1. Primers used in the validation of differentially expressed genes.

Abbreviations

2G: second-generation; AA9: Auxiliary family activity 9; ABC: ATP binding cassette; ABF: arabinofuranosidase; ACN: acetonitrile; AXE: acetyl xylan esterase; bp: base pairs; CAZy: carbohydrate active enzyme; CE: carbohydrate esterase; CIP: Cellulose-induced protein; DIGE: differential gel electrophoresis; FC: fold-change; FPKM: fragments per kilobase of exon per million fragments mapped; GH: glycoside hydrolase; GH31: α-xylosidase; GRN: gene regulatory network; LC/MS/MS: liquid chromatography tandem mass spectrometry; MEX: malt extract medium; MFS: Major facilitator superfamily; MS: mass spectrometry; PMO: polysaccharide monooxygenase; RT-qPCR: quantitative real-time PCR; SSCRP: Small secreted cysteine-rich protein; SWOI: swollenin; TF: transcriptional factor; xylanase: xyn.

Competing interests

The authors declare that there are no competing interests.

Authors' contributions

LSC performed the experimental design, laboratory experiments, performed the bioinformatics analysis, and drafted the manuscript. WRP performed the experimental design, DIGE experiments, and drafted the manuscript. ACCA performed the bioinformatics analysis, and drafted the manuscript. ASS performed the bioinformatics analysis, drafted the manuscript, and interpreted the data for the work. RS-R, NMM-R, AR performed the bioinformatics analysis, and drafted the manuscript. NB and GHG drafted and revised the manuscript and discussion, and interpreted the data for the work. VMF performed the secretome analysis and drafted the manuscript. GFP supervised the bioinformatics analysis and drafted the manuscript. RNS designed the project, supervised the research study, prepared/drafted the manuscript, and final approved of the version to be published. All the authors have read and approved the final manuscript.

Acknowledgments

This work was supported by The State of São Paulo Research Foundation (FAPESP) (proc. 2010/15683-8). We are grateful to Professor Christian Kubicek (TU-Wien) for helping in the experimental design of this study.

Author details

[1]Department of Biochemistry and Immunology, Ribeirão Preto Medical School, University of São Paulo, 14049-900 Ribeirão Preto, SP, Brazil. [2]Departamento de Biologia Celular, Universidade de Brasília, Asa Norte, 70910-900 Brasília, DF, Brazil. [3]Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, 14049-900 Ribeirão Preto, SP, Brazil. [4]Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, São Paulo, and Laboratório Nacional de Ciência e Tecnologia do Bioetanol, Campinas, Brazil.

Received: 14 January 2014 Accepted: 26 February 2014
Published: 21 March 2014

References

1. Pessoa-Jr A, Roberto IC, Menossi M, dos Santos RR, Filho SO, Penna TC: **Perspectives on bioenergy and biotechnology in Brazil.** *Appl Biochem Biotechnol* 2005, **121–124:**59–70.
2. Mosier N, Wyman C, Dale B, Elander R, Lee YY, Holtzapple M, Ladisch M: **Features of promising technologies for pretreatment of lignocellulosic biomass.** *Bioresour Technol* 2005, **96:**673–686.
3. Ojeda K, Kafarov V: **Exergy analysis of enzymatic hydrolysis reactors for transformation of lignocellulosic biomass to bioethanol.** *Chem Eng J* 2009, **154:**390–395.
4. Soccol CR, Vandenberghe LPD, Medeiros ABP, Karp SG, Buckeridge M, Ramos LP, Pitarelo AP, Ferreira-Leitao V, Gottschalk LMF, Ferrara MA, Bon EPD, de Moraes LMP, Araujo JD, Torres FAG: **Bioethanol from lignocelluloses: Status and perspectives in Brazil.** *Bioresource Technology* 2010, **101:**4820–4825.
5. Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, Mukherjee M, Kredics L, Alcaraz LD, Aerts A, Antal Z, Atanasova L, Cervantes-Badillo MG, Challacombe J, Chertkov O, McCluskey K, Coulpier F, Deshpande N, von Dohren H, Ebbole DJ, Esquivel-Naranjo EU, Fekete E, Flipphi M, Glaser F, Gomez-Rodriguez EY, Gruber S, *et al*: **Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of Trichoderma.** *Genome biology* 2011, **12:**R40.
6. Schuster A, Schmoll M: **Biology and biotechnology of Trichoderma.** *Appl Microbiol Biotechnol* 2010, **87:**787–799.
7. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EG, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, *et al*: **Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina).** *Nat Biotechnol* 2008, **26:**553–560.
8. Saloheimo M, Nakari-Setala T, Tenkanen M, Penttila M: **cDNA cloning of a Trichoderma reesei cellulase and demonstration of endoglucanase activity by expression in yeast.** *Eur J Biochem* 1997, **249:**584–591.
9. Hakkinen M, Arvas M, Oja M, Aro N, Penttila M, Saloheimo M, Pakula TM: **Re-annotation of the CAZy genes of Trichoderma reesei and transcription in the presence of lignocellulosic substrates.** *Microb Cell Fact* 2012, **11:**134.
10. Verbeke J, Coutinho P, Mathis H, Quenot A, Record E, Asther M, Heiss-Blanquet S: **Transcriptional profiling of cellulase and expansin-related genes in a hypercellulolytic Trichoderma reesei.** *Biotechnol Lett* 2009, **31:**1399–1405.
11. Mandels M, Parrish FW, Reese ET: **Sophorose as an inducer of cellulase in Trichoderma viride.** *J Bacteriol* 1962, **83:**400–408.

12. Kubicek CP, Messner R, Gruber F, Mach RL, Kubicek-Pranz EM: **The Trichoderma cellulase regulatory puzzle: from the interior life of a secretory fungus.** *Enzyme Microb Technol* 1993, **15:**90–99.
13. Seiboth B, Hakola S, Mach RL, Suominen PL, Kubicek CP: **Role of four major cellulases in triggering of cellulase gene expression by cellulose in Trichoderma reesei.** *J Bacteriol* 1997, **179:**5318–5320.
14. Foreman PK, Brown D, Dankmeyer L, Dean R, Diener S, Dunn-Coleman NS, Goedegebuur F, Houfek TD, England GJ, Kelley AS, Meerman HJ, Mitchell T, Mitchinson C, Olivares HA, Teunissen PJ, Yao J, Ward M: **Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus Trichoderma reesei.** *The Journal of biological chemistry* 2003, **278:**31988–31997.
15. Schmoll M, Kubicek CP: **Regulation of trichoderma cellulase formation: lessons in molecular biology from an industrial fungus.A review.** *Acta Microbiol Immunol Hung* 2003, **50:**125–145.
16. El-Gogary S, Leite A, Crivellaro O, Eveleigh DE, El-Dorry H: **Mechanism by which cellulose triggers cellobiohydrolase I gene expression in Trichoderma reesei.** *Proc Natl Acad Sci USA* 1989, **86:**6138–6141.
17. Nogawa M, Goto M, Okada H, Morikawa Y: **L-Sorbose induces cellulase gene transcription in the cellulolytic fungus Trichoderma reesei.** *Curr Genet* 2001, **38:**329–334.
18. Chambergo FS, Bonaccorsi ED, Ferreira AJ, Ramos AS, Ferreira Junior JR, Abrahao-Neto J, Farah JP, Farah JP, El-Dorry H: **Elucidation of the metabolic fate of glucose in the filamentous fungus Trichoderma reesei using expressed sequence tag (EST) analysis and cDNA microarrays.** *J Biol Chem* 2002, **277:**13983–13988.
19. Kubicek CP: **Systems biological approaches towards understanding cellulase production by Trichoderma reesei.** *J Biotechnol* 2013, **163:**133–142.
20. Kubicek CP, Mikus M, Schuster A, Schmoll M, Seiboth B: **Metabolic engineering strategies for the improvement of cellulase production by Hypocrea jecorina.** *Biotechnol Biofuels* 2009, **2:**19.
21. Castro LD, Antonieto AC, Pedersoli WR, Silva-Rocha R, Persinoti GF, Silva RN: **Expression pattern of cellulolytic and xylanolytic genes regulated by transcriptional factors XYR1 and CRE1 are affected by carbon source in Trichoderma reesei.** *Gene Expr Patterns* 2014, **14:**88–95.
22. Bischof R, Fourtis L, Limbeck A, Gamauf C, Seiboth B, Kubicek CP: **Comparative analysis of the Trichoderma reesei transcriptome during growth on the cellulase inducing substrates wheat straw and lactose.** *Biotechnol Biofuels* 2013, **6:**127.
23. Pao SS, Paulsen IT, Saier MH: **Major facilitator superfamily.** *Microbiol Mol Biol R* 1998, **62:**1–34.
24. Boyce KJ, Andrianopoulos A: **Morphogenetic circuitry regulating growth and development in the dimorphic pathogen Penicillium marneffei.** *Eukaryot Cell* 2013, **12:**154–160.
25. Yao J, Bajjalieh SM: **SVOP is a nucleotide binding protein.** *PLoS One* 2009, **4:**e5315.
26. Wieczorke R, Krampe S, Weierstall T, Freidel K, Hollenberg CP, Boles E: **Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in Saccharomyces cerevisiae.** *FEBS Lett* 1999, **464:**123–128.
27. Zhang W, Kou Y, Xu J, Cao Y, Zhao G, Shao J, Wang H, Wang Z, Bao X, Chen G, Liu W: **Two major facilitator superfamily sugar transporters from trichoderma reesei and their roles in induction of cellulase biosynthesis.** *J Biol Chem* 2013, **288:**32861–32872.
28. Ries L, Pullan ST, Delmas S, Malla S, Blythe MJ, Archer DB: **Genome-wide transcriptional response of Trichoderma reesei to lignocellulose using RNA sequencing and comparison with Aspergillus niger.** *BMC Genomics* 2013, **14:**541.
29. Ivanova C, Baath JA, Seiboth B, Kubicek CP: **Systems analysis of lactose metabolism in Trichoderma reesei identifies a lactose permease that is essential for cellulase induction.** *PLoS One* 2013, **8:**e62631.
30. Du J, Li S, Zhao H: **Discovery and characterization of novel d-xylose-specific transporters from Neurospora crassa and Pichia stipitis.** *Mol Biosyst* 2010, **6:**2150–2156.
31. Sun J, Tian C, Diamond S, Glass NL: **Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in Neurospora crassa.** *Eukaryot Cell* 2012, **11:**482–493.
32. Glass NL, Palma-Guerrero J, Jonkers W, Leeder A, Hall C, Kowbel D, Taylor JW, Brem R: **Revealing fungal communication modules by genomics, population genomics, and genome wide association studies in Neurospora crassa.** *Phytopathology* 2013, **103:**186–187.
33. Mukherjee PK, Horwitz BA, Herrera-Estrella A, Schmoll M, Kenerley CM: **Trichoderma research in the genome era.** *Annu Rev Phytopathol* 2013, **51:**105–129.
34. Furukawa T, Shida Y, Kitagami N, Mori K, Kato M, Kobayashi T, Okada H, Ogasawara W, Morikawa Y: **Identification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in Trichoderma reesei.** *Fungal Genet Biol* 2009, **46:**564–574.
35. Strauss J, Mach RL, Zeilinger S, Hartler G, Stoffler G, Wolschek M, Kubicek CP: **Cre1, the carbon catabolite repressor protein from Trichoderma reesei.** *FEBS Lett* 1995, **376:**103–107.
36. Portnoy T, Margeot A, Seidl-Seiboth V, Le Crom S, Ben Chaabane F, Linke R, Seiboth B, Kubicek CP: **Differential regulation of the cellulase transcription factors XYR1, ACE2, and ACE1 in Trichoderma reesei strains producing high and low levels of cellulase.** *Eukaryot Cell* 2011, **10:**262–271.
37. Nitta M, Furukawa T, Shida Y, Mori K, Kuhara S, Morikawa Y, Ogasawara W: **A new Zn(II)(2)Cys(6)-type transcription factor BglR regulates beta-glucosidase expression in Trichoderma reesei.** *Fungal Genet Biol* 2012, **49:**388–397.
38. Martinez-Antonio A, Collado-Vides J: **Identifying global regulators in transcriptional regulatory networks in bacteria.** *Curr Opin Microbiol* 2003, **6:**482–489.
39. Seiboth B, Karimi RA, Phatale PA, Linke R, Hartl L, Sauer DG, Smith KM, Baker SE, Freitag M, Kubicek CP: **The putative protein methyltransferase LAE1 controls cellulase gene expression in Trichoderma reesei.** *Mol Microbiol* 2012, **84:**1150–1164.
40. Karimi-Aghcheh R, Bok JW, Phatale PA, Smith KM, Baker SE, Lichius A, Omann M, Zeilinger S, Seiboth B, Rhee C, Keller NP, Freitag M, Kubicek CP: **Functional analyses of Trichoderma reesei LAE1 reveal conserved and contrasting roles of this regulator.** *G3 (Bethesda)* 2013, **3:**369–378.
41. Rossi A, Cruz AH, Santos RS, Silva PM, Silva EM, Mendes NS, Martinez-Rossi NM: **Ambient pH sensing in filamentous fungi: pitfalls in elucidating regulatory hierarchical signaling networks.** *IUBMB Life* 2013, **65:**930–935.
42. Li C, Yang Z, He Can Zhang R, Zhang D, Chen S, MA L: **Effect of pH on cellulase production and morphology of Trichoderma reesei and the application in cellulosic material hydrolysis.** *J Biotechnol* 2013, **168:**470–477.
43. Mach RL, Zeilinger S: **Regulation of gene expression in industrial fungi: Trichoderma.** *Appl Microbiol Biotechnol* 2003, **60:**515–522.
44. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO: **Genomic expression programs in the response of yeast cells to environmental changes.** *Mol Biol Cell* 2000, **11:**4241–4257.
45. Herpoel-Gimbert I, Margeot A, Dolla A, Jan G, Molle D, Lignon S, Mathis H, Sigoillot JC, Monot F, Asther M: **Comparative secretome analyses of two Trichoderma reesei RUT-C30 and CL847 hypersecretory strains.** *Biotechnol Biofuels* 2008, **1:**18.
46. Jun H, Kieselbach T, Jonsson LJ: **Enzyme production by filamentous fungi: analysis of the secretome of Trichoderma reesei grown on unconventional carbon source.** *Microb Cell Fact* 2011, **10:**68.
47. Adav SS, Chao LT, Sze SK: **Quantitative secretomic analysis of Trichoderma reesei strains reveals enzymatic composition for lignocellulosic biomass degradation.** *Mol Cell Proteomics* 2012, **11:**M111 012419.
48. Tisch D, Kubicek CP, Schmoll M: **New insights into the mechanism of light modulated signaling by heterotrimeric G-proteins: ENVOY acts on gna1 and gna3 and adjusts cAMP levels in Trichoderma reesei (Hypocrea jecorina).** *Fungal Genet Biol* 2011, **48:**631–640.
49. Druzhinina IS, Shelest E, Kubicek CP: **Novel traits of Trichoderma predicted through the analysis of its secretome.** *FEMS Microbiol Lett* 2012, **337:**1–9.
50. Zhou QX, Xu JT, Kou YB, Lv XX, Zhang X, Zhao GL, Zhang WX, Chen GJ, Liu WF: **Differential involvement of beta-glucosidases from hypocrea jecorina in rapid induction of cellulase genes by cellulose and cellobiose.** *Eukaryot Cell* 2012, **11:**1371–1381.
51. Znameroski EA, Coradetti ST, Roche CM, Tsai JC, Iavarone AT, Cate JHD, Glass NL: **Induction of lignocellulose-degrading enzymes in Neurospora crassa by cellodextrins.** *Proc Natl Acad Sci USA* 2012, **109:**6012–6017.
52. Gielkens MM, Dekkers E, Visser J, de Graaff LH: **Two cellobiohydrolase-encoding genes from Aspergillus niger require D-xylose and the xylanolytic transcriptional activator XlnR for their expression.** *Appl Environ Microbiol* 1999, **65:**4340–4345.
53. Schmoll M, Schuster A, Silva Rdo N, Kubicek CP: **The G-alpha protein GNA3 of Hypocrea jecorina (Anamorph Trichoderma reesei) regulates cellulase gene expression in the presence of light.** *Eukaryot Cell* 2009, **8:**410–420.

54. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25:**1754–1760.
55. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD: **The sequence alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25:**2078–2079.
56. Thorvaldsdottir H, Robinson JT, Mesirov JP: **Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration.** *Brief Bioinform* 2013, **14:**178–192.
57. Atanasova L, Jaklitsch WM, Komon-Zelazowska M, Kubicek CP, Druzhinina IS: **Clonal species Trichoderma parareesei sp. nov. likely resembles the ancestor of the cellulase producer Hypocrea jecorina/T. reesei.** *Appl Environ Microbiol* 2010, **76:**7259–7267.
58. Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11:**R106.
59. Vencio RZ, Koide T, Gomes SL, Pereira CA: **BayGO: bayesian analysis of ontology term enrichment in microarray data.** *BMC Bioinformatics* 2006, **7:**86.
60. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13:**2498–2504.
61. Steiger MG, Mach RL, Mach-Aigner AR: **An accurate normalization strategy for RT-qPCR in Hypocrea jecorina (Trichoderma reesei).** *J Biotechnol* 2010, **145:**30–37.
62. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(−Delta Delta C) method.** *Methods* 2001, **25:**402–408.
63. MacLean B, Eng JK, Beavis RC, McIntosh M: **General framework for developing and evaluating database scoring algorithms using the TANDEM search engine.** *Bioinformatics* 2006, **22:**2830–2832.
64. Keller A, Nesvizhskii AI, Kolker E, Aebersold R: **Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.** *Anal Chem* 2002, **74:**5383–5392.
65. Nesvizhskii AI, Keller A, Kolker E, Aebersold R: **A statistical model for identifying proteins by tandem mass spectrometry.** *Anal Chem* 2003, **75:**4646–4658.

doi:10.1186/1754-6834-7-41
Cite this article as: dos Santos Castro *et al.*: Comparative metabolism of cellulose, sophorose and glucose in *Trichoderma reesei* using high-throughput genomic and proteomic analyses. *Biotechnology for Biofuels* 2014 **7**:41.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

Fungal Genetics and Biology 73 (2014) 93–103

Contents lists available at ScienceDirect

Fungal Genetics and Biology

journal homepage: www.elsevier.com/locate/yfgbi

Regular Articles

Defining the genome-wide role of CRE1 during carbon catabolite repression in *Trichoderma reesei* using RNA-Seq analysis

Amanda Cristina Campos Antoniêto [a], Lílian dos Santos Castro [a], Rafael Silva-Rocha [a], Gabriela Felix Persinoti [b], Roberto Nascimento Silva [a,*]

[a] Department of Biochemistry and Immunology, Ribeirão Preto Medical School, University of São Paulo, 14049-900 Ribeirão Preto, SP, Brazil
[b] Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, São Paulo, Brazil

ARTICLE INFO

Article history:
Received 7 July 2014
Accepted 10 October 2014
Available online 18 October 2014

Keywords:
Trichoderma reesei
CRE1
Carbon catabolite repression
RNA-Seq

ABSTRACT

The ascomycete *Trichoderma reesei* is one of the most well-studied cellulolytic fungi and is widely used by the biotechnology industry in the production of second generation bioethanol. The carbon catabolite repression (CCR) mechanism adopted by *T. reesei* is mediated by the transcription factor CRE1. CCR represses genes related to cellulase production when a carbon source is readily available in the medium. Using RNA sequencing, we investigated CCR during the synthesis of cellulases, comparing the *T. reesei* Δ*cre1* mutant strain with its parental strain, QM9414. Of 9129 genes in the *T. reesei* genome, 268 genes were upregulated and 85 were downregulated in the presence of cellulose (Avicel). In addition, 251 genes were upregulated and 230 were downregulated in the presence of a high concentration of glucose. Genes encoding cellulolytic enzymes and transcription factors and genes related to the transport of nutrients and oxidative metabolism were also targets of CCR, mediated by CRE1 in a carbon source-dependent manner. Our results also suggested that CRE1 regulates the expression of genes related to the use of copper and iron as final electron acceptors or as cofactors of enzymes that participate in biomass degradation. As a result, the final effect of CRE1-mediated transcriptional regulation is to modulate the access of cellulolytic enzymes to cellulose polymers or blocks the entry of cellulase inducers into the cell, depending on the glucose content in the medium. These results will contribute to a better understanding of the mechanism of carbon catabolite repression in *T. reesei*, thereby enhancing its application in several biotechnology fields.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Trichoderma reesei (*Hypocrea jecorina*) is a saprophyte fungus widely used in the biotechnology industry because of its large cellulolytic potential (Schuster and Schmoll, 2010). To ensure its survival in different habitats, *T. reesei* must detect cellulose in its environment, produce different cellulases to degrade the insoluble substrate, carry the soluble products across the cytoplasmic membrane, and assimilate the sugars. In addition, *T. reesei* must respond to changes in the nutritional composition of the environment in order to compete with other microorganisms (Seiboth et al., 2011). An important mechanism for controlling metabolic processes in prokaryotes and eukaryotic microorganisms is carbon catabolite repression (CCR), in which the expression of genes necessary for the use of alternative carbon sources is repressed in the presence of easily metabolized carbon sources such as glucose (Strauss et al., 1995).

In *T. reesei* and other fungi, the key regulator of CCR is the Cys2His2-type transcription factor CRE1. The functional binding site for this transcription factor (TF) consists of two closely spaced 5′-SYGGRG-3′ motifs. It has been suggested that CRE1-mediated repression requires both binding sites (Kubicek et al., 2009). Binding of CRE1 to DNA may be regulated by phosphorylation of a serine in a short, conserved region of CRE1 by casein kinase 2 protein (Cziferszky et al., 2002). However, we have demonstrated that there is no correlation between gene expression and the number of putative CRE1 binding sites in the promoter regions of cellulolytic and xylanolytic genes (Castro et al., 2014).

CRE1 is similar to other fungal proteins that mediate CCR, such as CreA in *Aspergillus nidulans* and Mig1 in *Saccharomyces cerevisiae* (Portnoy et al., 2011a). In the hypercellulolytic *T. reesei* strain

Abbreviations: RNA-Seq, RNA sequencing; CCR, carbon catabolite repression; bp, base pair; GB, gigabase; RT-qPCR, real-time quantitative PCR; GO, Gene Ontology; GH, glycosyl hydrolase; CAZy, carbohydrate-active enzyme; TF, transcription factor; MFS, major facilitator superfamily; FC, fold change.

* Corresponding author.
E-mail address: rsilva@fmrp.usp.br (R.N. Silva).

http://dx.doi.org/10.1016/j.fgb.2014.10.009
1087-1845/© 2014 Elsevier Inc. All rights reserved.

Rut-C30, which produces cellulases and hemicellulases in medium containing glucose, a truncated form of *cre1* has been identified (Peterson and Nevalainen, 2012). In the Rut-C30 strain, complementation of this mutation with wild-type *cre1* restored catabolite repression in the presence of glucose, providing further evidence for the repressive role of CRE1 (Ilmen et al., 1996). In the yeast *S. cerevisiae*, Mig1 represses the transcription of about 90 genes associated with the use of alternative carbon sources such as sucrose, galactose, and maltose under conditions when glucose is sufficient (Santangelo, 2006).

CCR presents a major challenge to the biotechnology industry for the production of cellulolytic enzymes because the presence of glucose in the culture medium suppresses the production of these enzymes. Thus, there is an ongoing search for ways to circumvent this mechanism and thereby improve the application of cellulases in several industrial sectors, mainly in the production of second generation bioethanol. Despite the current understanding of CRE1-mediated CCR in readily available carbon sources such as glucose, little is known about how this TF acts in the presence of more complex carbon sources, such as cellulose. In the present study, to evaluate CRE1-mediated CCR, we used an RNA sequencing (RNA-Seq) approach to perform a large-scale comparative analysis of the transcriptomes of *T. reesei* strains QM9414 and Δ*cre1* grown in the presence of cellulose and glucose. Our results showed that deletion of the *cre1* gene in *T. reesei* altered the expression of genes related to oxidoreductase activity, nutrient transport, carbohydrate metabolism, and transcription regulation, among others. Gene expression was regulated in a carbon-dependent manner, mostly indirectly.

2. Material and methods

2.1. Strains and growth conditions

T. reesei strains QM9414 (ATCC 26921) and Δ*cre1* (Portnoy et al., 2011a) were obtained from the Institute for Chemical Engineering (Vienna University of Technology, Research Area Gene Technology and Applied Biochemistry, Vienna, Austria). The strains were maintained on MEX (malt extract 3% (*w*/*v*) and agar–agar 2% (*w*/*v*)) medium at 4 °C. Both strains were grown on MEX at 28 °C from days 7–10 until the complete sporulation. For the RNA-Seq experiments, a spore suspension (NaCl 0.8% (*w*/*v*), Tween 80 0.05% (*v*/*v*)) of each strain was inoculated in a 1-L Erlenmeyer flask containing 200 mL of Mandels-Andreotti medium (Schmoll et al., 2009) containing 1% (*w*/*v*) of cellulose (Avicel) or 2% (*w*/*v*) of glucose, depending of the experiment. The flasks were incubated at 200 rpm and 28 °C for 24, 48, and 72 h when grown in cellulose medium and for 24 and 48 h when grown in glucose medium. For real-time quantitative PCR (RT-qPCR) experiments, a spore suspension of strain QM9414 or Δ*cre1* was inoculated in a 250-mL Erlenmeyer flask containing 25 mL of Mandels-Andreotti medium containing 2% (*w*/*v*) of glucose. After 48 h, the pre-grown mycelia were collected, washed three times with Mandels-Andreotti medium without a carbon source, and transferred to 25 mL of Mandels-Andreotti medium containing 1% (*w*/*v*) of cellulose. The mycelia were collected at 6, 12, and 24 h, or 1% (*w*/*v*) of glucose was added after 24 h of growth, and the mycelia were collected 6, 12, and 24 h later. All experiments were conducted in biological triplicates for each sample. After induction, the mycelia were collected by filtration, frozen, and stored at -80 °C.

2.2. RNA extraction

The mycelia of *T. reesei* strains QM9414 and Δ*cre1*, grown in different carbon sources as described above, were filtered through Miracloth, frozen in liquid nitrogen, and macerated. Total RNA was extracted using TRIzol® RNA reagent (Life Technologies) according to the manufacturer's instructions. RNA was quantified, and the integrity was checked using a 2100 Bioanalyzer (Agilent).

2.3. RNA-Seq

Total RNA of three biological replicates obtained from *T. reesei* strains QM9414 and Δ*cre1* in presence of cellulose (24, 48 and 72 h) and glucose (24 and 48 h) were used for sequencing. The total RNA from three time points were pooled, lyophilized and stored using RNAstable Tube kit (Biomatrica) to stabilize the RNA for sequencing. The barcoded libraries were prepared and sequenced by LGC Genomics GmbH (Berlin/Germany) using the Illumina Hiseq 2000 platform.

2.4. Data analysis

Sequences were mapped based on the reference genome of *Trichoderma reesei* 2.0, obtained from the Joint Genome Institute (JGI) Genome Portal (http://genome.jgi-psf.org/Trire2/Trire2.home.html), using Bowtie version 0.12.8 (Li and Durbin, 2009), with only unique alignments allowed. After alignment, SAMtools version 0.1.18 (Li et al., 2009) was used to process the alignments files, which were visualized using the Integrative Genomics Viewer (Thorvaldsdottir et al., 2013). The Bioconductor DESeq package version 1.10.1 (Anders and Huber, 2010) was used for the differential expression analysis, and a twofold change cut-off ($\log_2$ fold change) of $\geqslant 1$ or $\leqslant -1$ and an adjusted *p*-value of $\leqslant 0.05$ were established as thresholds. Samples were normalized using median log deviation implemented in the DESeq package. Functional categorization was performed with Gene Ontology (GO) terms using BayGO software (Vencio et al., 2006), adopting a *p*-value of $\leqslant 0.05$ as the criterion for significantly enriched categories. The network of genes upregulated and downregulated by CRE1 in different carbon sources was generated using Cytoscape version 3.0.1 (Shannon et al., 2003). Raw sequence data and count data for all samples are available at GEO (http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE57374 and GSE53629. Protein sequences from *T. reesei* and other species used in the phylogenetic analysis were obtained from the JGI Genome Portal and other online databases, respectively. Multiple sequence alignment was performed using ClustalW, and neighbor-joining trees were created with the Mega 4 program (Tamura et al., 2007) using 1000 bootstraps.

2.5. RT-qPCR

To validate the differentially expressed genes, we used the pooled RNA samples from strains QM9414 and Δ*cre1* grown in cellulose. Twenty genes were used for validation, including hydrolytic enzymes and other proteins chosen randomly. In this analysis, we used the sample from QM9414 grown in cellulose as the reference. In experiments evaluating the expression of the genes ID 3405, ID 48444, ID 52315, and ID 111750 in QM9414 and Δ*cre1* strains grown on cellulose and then transferred to glucose (see Section 2.1), we used the absolute gene expression values with the actin gene as an endogenous control. The primers used in these experiments are described in Table S1.

For these experiments, 1 μg of RNA was treated with DNase I (Fermentas) to remove genomic DNA. cDNA was then synthesized using a Maxima First Strand cDNA Synthesis kit (Thermo Scientific) according to the manufacturer's instructions. The cDNA, diluted 1:50, was analyzed with real-time PCR using the CFX96™ Real-Time PCR Detection System (Bio-Rad) and the SsoFast™ EvaGreen® Supermix (Bio-Rad), in accordance with the manufacturer's instructions. The actin gene was used as the endogenous control

to normalize the total amount of cDNA present in each reaction. The amplification program used in this study was: 95 °C for 10 min; 39 cycles of 95 °C for 10 s and 60 °C for 30 s; and a dissociation curve of 60 °C to 95 °C at increments of 0.5 °C for 10 s. Gene expression levels were calculated as described by Livak and Schmittgen (2001).

2.6. Statistical analysis

Statistical analysis of gene expression for the genes ID 3405, ID 48444, ID 52315, and ID 111750 was performed with one-way ANOVA (nonparametric) followed by the Bonferroni test (comparing all pairs of columns), available in the Prism version 5.0 software. The parental QM9414 strain was used as a standard for comparison with the Δ*cre1* mutant under all conditions analyzed.

2.7. In silico analysis of putative CRE1 binding sites

For *in silico* prediction of CRE1 binding sites, 1.5-kb sequences upstream of the ATG start codon in 251 genes from the glucose regulon and 268 genes from the cellulose regulon were retrieved from the genomic sequence of *T. reesei*, available from the JGI *T. reesei* database (http://genome.jgi-psf.org/Trire2/Trire2.home.html). Double binding motifs for CRE1, defined as inverted or direct repeats of the core consensus sequence (Fig. S1), were identified in the promoter regions using 9.0 as the threshold (Silva-Rocha et al., 2014). We also looked for XYR1 binding sites in the 1.5-kb promoter regions of the 251 genes belonging to the glucose regulon. Binding motifs for XYR1, defined as inverted and everted repeats of the core consensus sequence (Fig. S1), were identified in the promoter regions using 6.4 as the threshold (Silva-Rocha et al., 2014). We assumed that genes with one or more binding sites for the transcription factors CRE1 and XYR1 are under direct regulation, while the absence of binding sites for these factors suggests indirect regulation.

3. Results

3.1. Global analysis of the genes regulated by CRE1 in cellulose and glucose

T. reesei strains QM9414 and Δ*cre1* were initially grown in medium containing cellulose or glucose as the sole carbon source, as described in Section 2.1. Three biological replicates of each condition were sequenced, yielding approximately 188 million 100-bp paired-end reads, corresponding to 37 GB of nucleotides. Reads were uniquely mapped to the *T. reesei* QM6a reference genome. Overall, 70% of the reads mapped to the reference genome (Table S2). The three biological replicates of each condition showed a high Pearson correlation ($R \geqslant 0.6$) (Fig. S2), and a principal component analysis demonstrated the reliability of RNA-Seq (Fig. S3).

Expression in the Δ*cre1* and QM9414 strains was compared. Using a *p*-value of $\leqslant 0.05$ as the threshold, 815 and 697 genes were modulated by CRE1 in cellulose and glucose, respectively (Fig. 1). Global analysis of the transcriptome showed that 905 genes were upregulated or downregulated in the Δ*cre1* strain, in at least one of the conditions analyzed, when compared with levels in parental strain QM9414. From these data, we generated a regulatory network of all genes upregulated and downregulated by CRE1 in cellulose and glucose. Of the 905 genes represented, 481 and 353 were exclusively modulated by CRE1 in the presence of cellulose and glucose, respectively, with 71 genes commonly upregulated or downregulated in the two conditions (Fig. 2). We defined the "regulon" as the set of genes that were repressed by CRE1 uniquely in each condition. As such, 251 genes belonged to the cellulose regulon, and 268 to the glucose regulon (Table S3). As shown in Fig. 2, when *T. reesei* was grown in a high concentration of glucose, CRE1 activated and repressed genes in almost equal numbers. In contrast, when *T. reesei* was grown in cellulose, CRE1 mainly repressed transcription, with only a small subset of genes activated in this condition.

To verify our results before proceeding with subsequent analyses, we used RT-qPCR to validate the RNA-Seq data. The expression of 20 genes, including those encoding hydrolytic enzymes and other proteins chosen randomly, were analyzed using Δ*cre1* and QM9414 strains grown in the presence of cellulose. The results of the RT-qPCR experiments are shown in Table S4. A strong correlation was found between the results obtained with the two techniques ($r^2 = 0.94$), reflecting the quality of the chosen method and the manner in which both experiments were conducted (Fig. S4).

3.2. Top genes regulated by CRE1 in the two carbon sources

Upregulated and downregulated genes in the Δ*cre1* mutant were categorized using GO terms. Categories such as "integral to

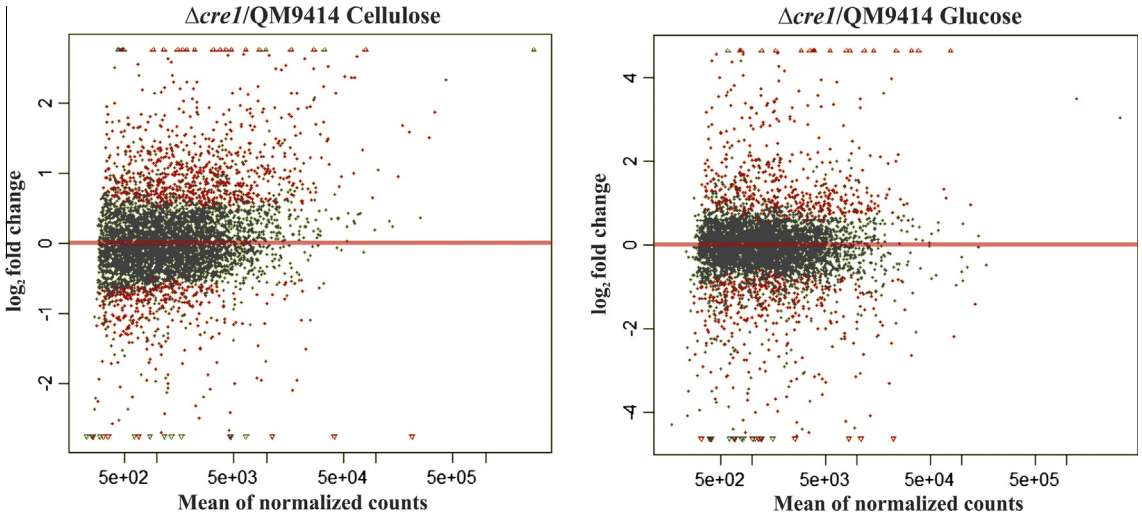


Fig. 1. Genes differentially expressed in the Δ*cre1* mutant relative to expression in the QM9414 parental strain. The expression profile was generated for growth in cellulose or glucose. Differentially expressed genes, identified using the DESeq package, are shown in red. Of the 9129 *T. reesei* genes, 815 and 697 were differentially expressed by growth in cellulose and glucose, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

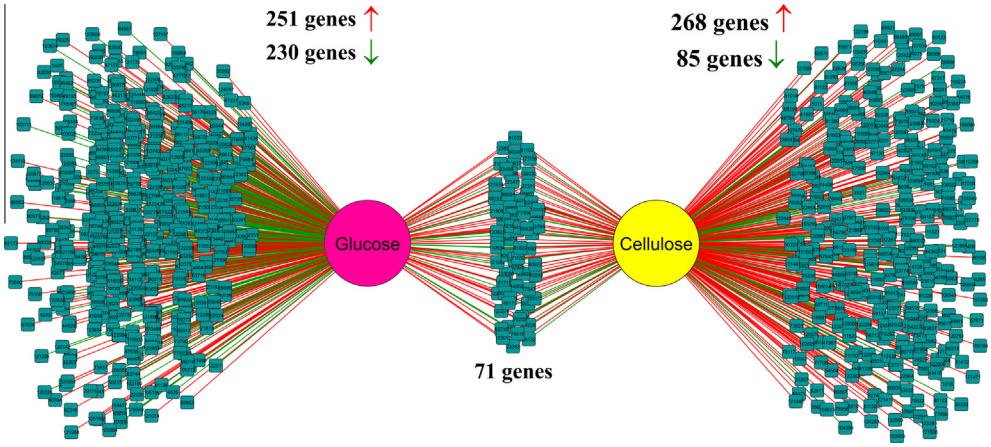


Fig. 2. Regulatory network of the 905 differentially expressed genes. The Δ*cre1* mutant was compared to the QM9414 parental strain after growth in different carbon sources. Genes are represented by blue squares, and interactions are represented by red and green lines that indicate upregulation and downregulation, respectively. The network was created using Cytoscape 3.0. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

membrane," "transporter activity," and "electron transport" were significantly enriched among the upregulated genes in both conditions analyzed, while "integral to membrane," "zinc ion binding," and "electron transport" were overrepresented among the downregulated sets of genes. The analysis suggests that, during CCR, CRE1 mainly represses genes directly or indirectly related to the hydrolysis of lignocellulosic material. During cultivation in glucose, all differentially regulated genes related to metabolism and/or transport of carbohydrates were upregulated in the absence of CRE1. Interestingly, a class of genes related to oxidoreductase activity was highly enriched during cultivation in cellulose. Previous studies have shown that CCR acts preferentially regulating genes related to the entrance of inducer molecules into the cell in the presence of glucose (Portnoy et al., 2011a). During growth in more complex carbon sources, such as cellulose, we observed that genes encoding oxidative metabolism function were the first ones repressed (Fig. 3, Table S5).

The 10 genes whose expression changed the most in the Δ*cre1* strain after growth in each carbon source are listed in Table 1. In cellulose, the gene with the highest expression in the Δ*cre1* mutant encodes a protein (ID 107641) unique to *T. reesei*. Analyzing the structure of this protein, we observe that it has a bromodomain, which represents a large family of evolutionarily conserved protein modules found in many chromatin-associated proteins and play a

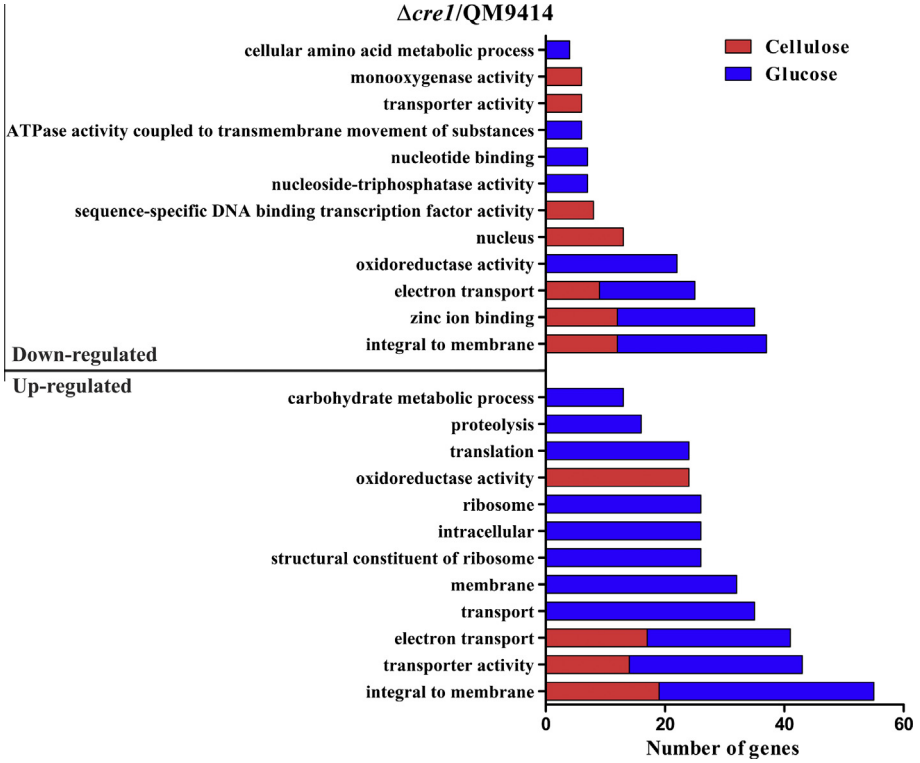


Fig. 3. Enrichment analysis using GO terms for the major upregulated and downregulated genes in the Δ*cre1* strain. *T. reesei* strain Δ*cre1* was compared to the QM9414 parental strain. The red and blue bars correspond to the cellulose and glucose growth conditions, respectively. The categories were considered significantly enriched when the *p*-value was $\leqslant 0.05$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
The 10 most upregulated genes in the Δ*cre1* strain when compared to expression in strain QM9414, after growth in cellulose or glucose. Gene expression values are presented as the $\log_2$ fold change (FC).

Condition	Protein ID	Description	FC	Regulation by CRE1
Cellulose	107641	Unique protein	5.19	Indirect
	23353	Ferric reductase	5.01	Indirect
	111750	Ferric reductase	4.64	Indirect
	69742	Esterase	4.46	Indirect
	62716	Copper transporter	3.95	Indirect
	52315	Copper transporter	3.94	Indirect
	66345	Mn superoxide dismutase	3.83	Indirect
	65695	Unknown protein	3.73	Indirect
	4517	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	3.64	Indirect
	22386	Epl1/Sm1	3.30	Indirect
Glucose	73256	GH81 endo-1,3-β-glucanase	11.05	Indirect
	56418	GH55 1,3-β-glucanase	8.10	Indirect
	59628	Unknown protein	7.79	Indirect
	123456	GH65 α,α-trehalase	7.73	Indirect
	48444	MFS maltose permease	7.49	Indirect
	3405	MFS permease (Ctr1)	6.03	Indirect
	120749	GH1 α-glucosidase BGL2/CEL1a	6.01	Indirect
	105155	Unknown protein	5.91	Indirect
	71532	GH71 1,3-α-glucanase	5.86	Direct
	123718	Amino acid transporter, neutral 11 TM	5.60	Indirect

key role in chromatin remodeling (Zeng and Zhou, 2002). Because of this remodeling function, this gene has the potential to affect transcriptional processes (Sanchez and Zhou, 2009). In this way, the repression of this gene during CCR can result in an alteration of chromatin structure and function and consequently change the expression of a set of genes related to lignocellulose degradation as glucose appears in the medium. Genes that encode ferric reductases (ID 23353 and ID 111750) and copper transporters (ID 62716 and ID 52315) were also upregulated in the Δ*cre1* mutant in the presence of cellulose, as were genes related to stress defense mechanisms, such as superoxide dismutase (ID66345) and cytochrome P450 (ID 4517).

When analyzing CCR in the presence of glucose, we noticed that most of the genes repressed by CRE1 were related to sugar metabolism. This is consistent with the fact that CRE1 is induced when an easy-to-degrade carbon source, such as glucose, is readily available. Under these conditions, there is no need to express genes of hydrolytic enzymes that degrade more complex carbon sources. Therefore, with the exception of two genes with unknown functions (ID 59628 and ID 105155) and an amino acid transporter-encoding gene (ID 123718), seven of the genes with the highest expression in the Δ*cre1* mutant encode proteins involved in the metabolism and transport of sugars (Table 1). Among the genes regulated during growth in cellulose or glucose, most, except the glycosyl hydrolase (GH)71 1,3-α-glucanase-encoding gene, seemed to be indirectly regulated by CRE1. This indicates that the TF does not act alone and that additional regulatory proteins are required to promote CCR and optimize carbon source utilization in *T. reesei*.

3.3. Expression of carbohydrate-active enzyme (CAZy) genes

The CCR mechanism triggered by CRE1 alters the expression of several genes related to the degradation of the lignocellulosic biomass, including those encoding carbohydrate-active enzymes, transporters, and other TFs. Accordingly, we investigated how these genes are affected by CRE1. Recently, Hakkinen and coworkers organized a re-annotation of all genes in the CAZy family in *T. reesei* and identified 201 genes encoding glycosyl hydrolases, 22 encoding carbohydrate esterases, and 5 encoding polysaccharide lyases (Hakkinen et al., 2012). Using this information, we constructed a table containing all genes in the CAZy family that are subject to CCR mediated by CRE1 in the presence of glucose, using a *p*-value of $\leqslant 0.05$ as the threshold (Table 2). The expression of two glucanase genes (ID 73256 and ID 56418), one trehalose gene (ID 123456), and the β-glucosidase *cel1a* gene (ID 120749) was over 64 times higher ($\log_2$ fold change $\geqslant 6$) in the Δ*cre1* mutant, when compared to expression in strain QM9414. Furthermore, four other β-glucosidase genes, including the already characterized *cel3b*, *cel1b*, and *cel3c*, were also upregulated in the mutant strain (Table 2). However, most CAZy genes regulated by CRE1 were indirectly regulated by this TF.

3.4. Expression of transcription factors genes

Many regulatory proteins are required during cellulose degradation. The *T. reesei* genome encodes 494 TFs, six and eight of them were subject to CCR by CRE1 in cellulose and glucose, respectively. No TF gene was commonly regulated in the two conditions (Table 3). Genes encoding zinc finger TFs were upregulated in the Δ*cre1* mutant in both conditions analyzed. In the presence of cellulose, the most highly expressed TF gene in the Δ*cre1* strain, relative to expression in strain QM9414, encodes a Zn2Cys6-type transcriptional regulator (ID 121682). The function of this gene has not yet been described, but the encoded protein contains a Gal4 domain, which is involved in galactose and melibiose metabolism in yeast (Hong et al., 2008). Another gene regulated by CRE1 in cellulose encodes the bZIP transcriptional regulator MeaB (ID 73417). This TF was identified by Polley and Caddick, who described its involvement in nitrogen metabolism in *A. nidulans* (Polley and Caddick, 1996). In *T. reesei*, MeaB gene expression in the mutant strain was 2.7 times higher ($\log_2$ fold change = 1.43) than in the parental strain. Furthermore, the expression of the TF gene *ace2*, encoding a positive regulator of genes encoding cellulase and xylanase in *T. reesei* (Aro, 2001), was over two times higher in the Δ*cre1* strain ($\log_2$ fold change = 1.21) when compared to expression in strain QM9414 (Table 3). On the other hand, during cultivation in glucose, the TF gene *xyr1* (ID 122208) was the main gene targeted by CRE1 during CCR; its expression was 36 times higher ($\log_2$ fold change = 5.19) in the Δ*cre1* mutant than in strain QM9414 (Table 3). XYR1 also participates in the positive regulation of cellulase and xylanase gene expression in *T. reesei* (Mach-Aigner et al., 2008), indicating a role for this TF during CCR. Three other Zn2Cys6-type TF genes (ID 27600, ID 123881

Table 2
CAZy genes upregulated in the Δ*cre1* mutant, relative expression in strain QM9414, after growth in glucose. Gene expression values are presented as the $\log_2$ fold change (FC).

Protein ID	Name	Class	Family	Annotation	FC	Regulation by CRE1
73256		GH[a,*]	81	Candidate endo-1,3-β-glucanase	11.05	Indirect
56418		GH	55	Candidate β-1,3-glucanase	8.10	Indirect
123456		GH	65	Candidate α,α-trehalase	7.73	Indirect
120749	bgl2/cel1a	GH	1	β-Glucosidase	6.01	Indirect
71532		GH	71	Candidate α-1,3-glucanase	5.86	Direct
121735	cel3b	GH	3	Candidate β-glucosidase	5.12	Indirect
82235		GH	31	Candidate α-glucosidase	4.81	Indirect
121746	gluc78	GH	55	Candidate exo-1,3-β-glucanase	3.98	Indirect
1885	gla	GH	15	Glucoamylase	3.49	Indirect
70845		GH	55	Candidate β-1,3-glucanase	3.43	Indirect
22197	cel1b	GH	1	Candidate β-glucosidase	3.28	Indirect
77942		GH	72	Candidate β-1,3-glucanosyl transferase	2.95	Indirect
82227	cel3c	GH	3	Candidate β-glucosidase	2.84	Indirect
3196		GH	38	Candidate α-mannosidase	2.15	Indirect
105956		GH	13	Candidate α-amylase	1.62	Indirect
108477		GH	13	Candidate α-glucosidase/oligo α-glucosidase	1.46	Indirect
66041	chi18-18	GH	18	Candidate chitinase	1.41	Indirect
76266		GH	16	Candidate cell wall glucanosyl transferase	1.36	Indirect
49193		GH	17	Candidate glucan 1,3-β-glucosidase	1.25	Indirect
62166		GH	2	Candidate β-mannosidase	1.12	Indirect

[a] Glycoside hydrolase.

Table 3
Transcription factor genes upregulated in the Δ*cre1* mutant, relative to expression in strain QM9414, after growth in cellulose or glucose. Gene expression values are presented as the $\log_2$ fold-change (FC).

Condition	Protein ID	Description	FC	Regulation by CRE1
Cellulose	121682	Zn2Cys6 transcriptional regulator	1.54	Indirect
	73417	bZIP transcriptional regulator MeaB	1.43	Indirect
	22774	bHLH transcriptional regulator	1.22	Indirect
	78445	Zn2Cys6 transcriptional regulator (ACE2)	1.21	Indirect
	110658	C2HC transcription factor	1.12	Indirect
	47317	Zn2Cys6 transcriptional regulator	1.10	Indirect
Glucose	122208	Zn2Cys6 transcriptional regulator (XYR1)	5.19	Direct
	27600	Zn2Cys6 transcriptional regulator	2.16	Indirect
	123881	Zn2Cys6 transcriptional regulator	2.15	Indirect
	3449	Zn2Cys6 transcriptional regulator	1.65	Indirect
	55422	Elongation factor 2 kinase, putative	1.28	Indirect
	71559	Translation initiation factor eIF-5A gene based on homologyToThe corresponding genes in several filamentous fungi	1.07	Indirect
	48438	Zn2Cys6 transcriptional regulator	1.05	Indirect
	62345	Translation initiation factor	1.00	Indirect

and ID 3449), whose expression was also repressed in the presence of glucose, encode proteins with unknown functions. However, like the TF ID 121682 gene, which was upregulated in the cellulose growth condition, they also contain a Gal4 domain. These data suggest that CCR can be regulated indirectly by different TFs depending on the carbon source available.

3.5. Expression of transporters genes

Genes encoding proteins related to transport comprise approximately 5% of the *T. reesei* genome. Given the large number of transporters, we investigated their involvement in CCR. We found that 18 transporter-encoding genes were upregulated exclusively in cellulose and 33 were upregulated exclusively in glucose in the Δ*cre1* strain, relative to expression in strain QM9414. Five genes were commonly upregulated in cellulose and glucose (Table S6).

In the presence of cellulose, a gene that encode to a copper transporter (ID 108749) was the most strongly upregulated gene in the Δ*cre1* strain, with expression eight times higher ($\log_2$ fold change = 3.02) than in the QM9414 parental strain. Several ion transporter-encoded genes were also repressed by CRE1 during cultivation in cellulose. In both conditions analyzed, the main genes targeted by CRE1-mediated CCR encode to proteins belonging to a major facilitator superfamily (MFS) permeases. In the Δ*cre1* strain, the number of MFS permease genes upregulated was greatest during growth in the repressive carbon source glucose. In the presence of glucose, expression of the maltose permease gene (ID 48444) was almost 180 times higher ($\log_2$ fold change = 7.49) in the Δ*cre1* mutant than in the parental strain. In the Δ*cre1* mutant, copper transport genes were upregulated, relative to expression in strain QM9414, in both conditions. However, CRE1 strongly repressed these genes during growth in cellulose.

To compare the similarity of the target transporters regulated under CCR in each condition, a phylogenetic analysis was performed using the amino acid sequences of the transporters genes upregulated in the Δ*cre1* mutant. Other transport-related proteins from different species were included in the analysis (Table S7). A secreted lipase (ID 57204) from *T. reesei* was used as outgroup (Fig. 4). Seven clusters are evident in the phylogenetic tree: four groups containing MFS permeases, one group formed by maltose transporters, one group consisting of ABC transporters, and one group related to copper transporters.

The largest group of MFS permeases comprised eight *T. reesei* proteins and 11 sugar transporters from different species, suggesting that these MFS permeases are involved in carbohydrate trans-

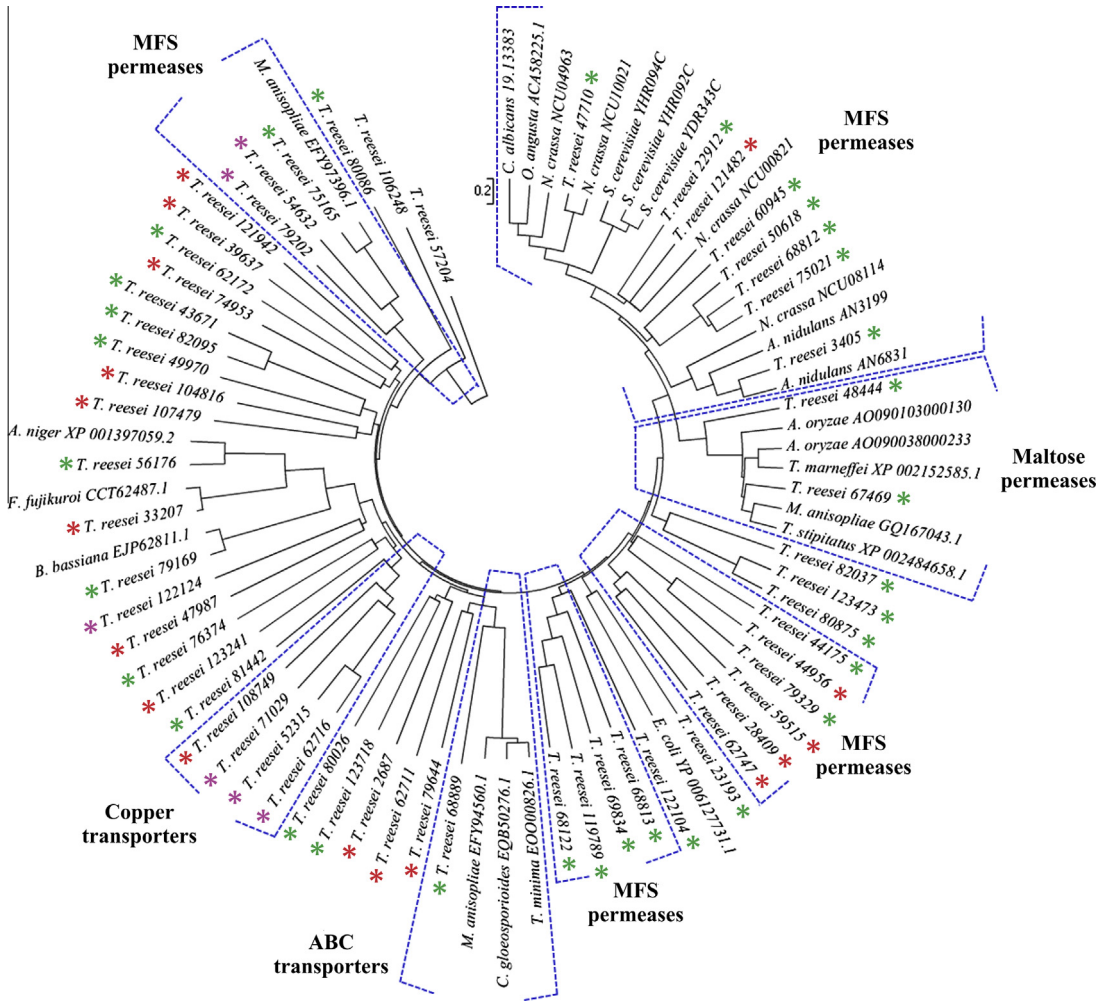


Fig. 4. Phylogenetic analysis of transporters upregulated in the Δ*cre1* mutant during growth in cellulose and glucose. The analysis included other proteins involved in transport from different species. Seven clusters were formed by MFS permeases, maltose permeases, ABC transporters, and copper transporters. Red, green, and purple asterisks represent genes expressed in cellulose, glucose, and both conditions, respectively. The tree was created with the Mega 4 program using the neighbor-joining method with 1000 bootstraps. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

port. *T. reesei* MFS permease genes that were expressed in the same conditions formed two other clusters. The first contained the MFS permease genes ID 68813, ID 69834, ID 119789, and ID 68122, all of which were expressed in the presence of glucose. Similarly, two unique MFS permease genes commonly expressed in cellulose and glucose grouped together with the MFS permease from *Metarhizium anisopliae* EFY97396.1. The fourth cluster of MFS permeases was heterogeneous, containing genes expressed in both conditions. The cluster formed by these maltose permeases comprised proteins of different species, including maltose permeases from *T. reesei*, *Aspergillus oryzae*, and *Talaromyces marneffei*, that, along with sugar transporters from *Talaromyces stipitatus* and *M. anisopliae*, appear to be related to disaccharide transport, not only maltose transport (Lingner et al., 2011). The cluster containing ABC transporters revealed a similarity between protein ID 68889 of *T. reesei* (PDR-type ABC transporter) and ABC transporters from *Colletotrichum gloeosporioides*, *Togninia minima*, and *M. anisopliae*. Finally, we verified the grouping of the four genes that encode copper transporters from *T. reesei*, which were expressed in both cellulose and glucose. In the phylogenetic analysis, most clusters were not specific to a particular carbon source. Thus, although regulation was carbon source-dependent, the structure of the transporters regulated by CCR in cellulose and glucose may be similar.

3.6. Identifying putative binding sites for CRE1 in the promoter regions of genes in the glucose and cellulose regulons

To investigate the mechanisms of CRE1-mediated CCR, we searched for possible CRE1 binding sites in the 1.5-kb promoter regions of 251 genes in the glucose regulon and 268 genes in the cellulose regulon. We searched for double sites and found 33 genes in the glucose regulon and 21 genes in the cellulose regulon with at least one CRE1 binding motif, indicating that most of the CRE1-repressed genes were regulated indirectly (Fig. S5, Table S8). In the glucose regulon, 29 genes contained one double site and four genes contained two doubles sites for CRE1 binding. The genes with two double sites were annotated as tubulin beta chain (ID 122886), unknown protein (ID 120826), MFS permease (ID 79329), and glucokinase (ID 80231). In the cellulose regulon, 19 genes had one double site, one gene had two double sites, and one gene had three double sites for CRE1 binding. The latter was annotated as aspartate/other aminotransferase (ID 75725); its expression was almost three times higher ($\log_2$ fold change = 1.34) in the Δ*cre1* mutant than in strain QM9414.

As noted earlier, the gene encoding the transcription factor XYR1 was directly repressed by CRE1 in the presence of glucose. We looked for XYR1 binding sites in the 1.5-kb promoter regions

of 251 genes in the glucose regulon. Overall, 38 genes contained at least one XYR1 binding site, five of which also had binding sites for CRE1 (Fig. S5, Table S9). Therefore, genes regulated directly by CRE1 during CCR could in turn regulate other genes, accounting for the large number of genes in the glucose and cellulose regulons without potential CRE1 binding sites.

3.7. CCR leads to tight regulation of MFS/copper transporter genes and genes involved in oxidative metabolism for cellulose degradation

The results described above suggest that CCR mediated by CRE1 has different effects depending on the glucose concentration. When glucose is added to the medium or is generated from a more complex polymer, such as cellulose, the fungus responds rapidly to guarantee efficient cellulose metabolism. Accordingly, we performed an experiment similar to one described by Delmas et al. (2012), in which *T. reesei* strains QM9414 and Δ*cre1* were pregrown in 2% glucose for 48 h and then transferred to medium containing 1% cellulose, to which glucose was added to a concentration of 1% 24 h later (see details in Section 2.1). In order to verify the responses of the main genes repressed by CRE1 during growth in the cellulose and glucose condition, we analyzed the following genes by RT-qPCR: a copper transporter (ID 52315); a ferric reductase (ID 111750); and two permeases, MFS permease (ID 3405) and MFS maltose permease (ID 48444). As shown in Fig. 5, the expression of both MFS permease genes increased in the Δ*cre1* mutant in the first 6 h after glucose was added to the medium, demonstrating that sugar transport-related genes are preferentially repressed by CRE1 in the presence of large amounts of free glucose and confirming the RNA-Seq results. Once the added glucose was consumed (12 and 24 h after glucose addition), the expression of both permease genes decreased in the mutant strain, suggesting that their expression depended on high glucose levels. Analyzing the expression of the copper transporter (ID 52315) and the ferric reductase (ID 111750) genes, we observed that expression in the Δ*cre1* mutant was highest after growth in cellulose for 24 h. This overexpression can be explained by the fact that CRE1 senses the low amount of glucose generated from the cellulose polymer and acts preferentially to repress genes related to oxidative metabolism. When a concentrated solution of glucose was added to the medium, CRE1-mediated CCR was quenched, corroborating the results obtained with RNA-Seq (Fig. 5). In summary, CRE1-mediated CCR preferentially affects transport-related genes when the glucose concentration is high and genes related to oxidative processes, such as copper transporters and ferric reductases, when a small amount of glucose is present in the medium.

4. Discussion

The CCR mechanism is essential for the adaptation and survival of some species, such as *T. reesei*. To inhabit a variety of environments, this fungus has evolved ways to recognize nutrients and capture them efficiently. In this context, the transcription factor CRE1 plays a central role. In the presence of carbon sources of varying complexity (cellulose and glucose), different genes were upregulated or downregulated in the Δ*cre1* strain, suggesting that CRE1 regulates the expression of its target genes in a carbon source-dependent manner. Initially, it was thought that CRE1 directly repress the transcription of cellulases and hemicellulases genes when a readily metabolizable carbon source was present (Strauss et al., 1995). However, we showed that other genes with diverse functions were also under the control of CRE1 and that some of the genes encoding hydrolytic enzymes were most likely not directly regulated by CRE1. Only a small number of the transcribed CAZy genes were direct targets of CRE1 during CCR. The same results were observed by Portnoy et al., who found that few cellulolytic genes were upregulated in the Δ*cre1 T. reesei* mutant during growth in the presence of glucose at different rates (Portnoy et al.,

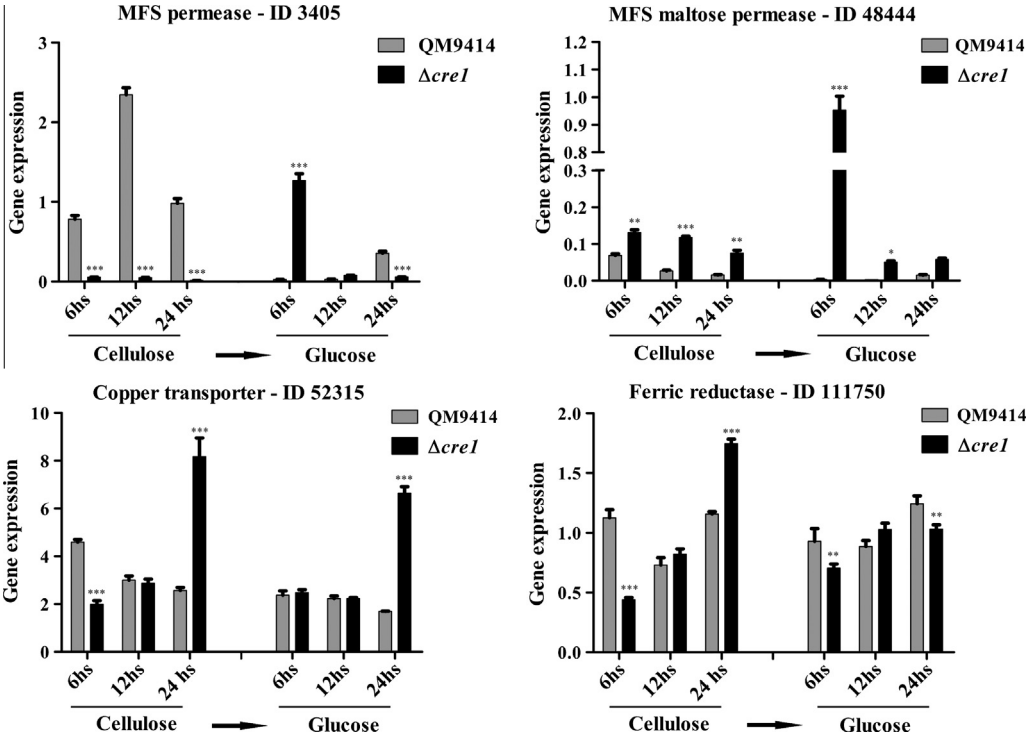


Fig. 5. Gene expression levels of two MFS permeases, a copper transporter, and a ferric reductase. *T. reesei* strains QM9414 and Δ*cre1* were grown in cellulose for 6, 12, and 24 h and for an additional 6, 12, and 24 h after glucose addition. Gray and black bars represent expression in QM9414 and Δ*cre1*, respectively. The absolute expression is shown, with the actin gene used as an endogenous control. Statistically significant differences are marked with ** (*p*-value <0.01) or *** (*p*-value <0.001).

2011a). These results indicate that the function of CRE1 is not straightforward and suggest that regulation of the CAZy genes involves other proteins that are directly or indirectly regulated by CRE1. However, some CAZy genes were strongly repressed by CRE1, such as the gene encoding GH65 α,α-trehalase. This enzyme catalyzes the hydrolysis of trehalose, a disaccharide composed of two glucose molecules that functions as a reserve in filamentous fungi (Best et al., 2011). Accordingly, repression of trehalase gene by CRE1 in the presence of glucose may be a fungal strategy to avoid unnecessary use of its energy stock when a readily available carbon source is present in the medium. This result is in accordance with the literature: the expression of the GH65 α,α-trehalase gene is upregulated in the *T. reesei* Δ*cre1* strain at high growth rates in the presence of glucose (Portnoy et al., 2011a) and in *Neurospora crassa* grown on cellulose (Arkowitz et al., 2011). In addition, the BGL2-encoding gene was upregulated in the Δ*cre1* strain in the presence of glucose (approximately 70-fold; Table 1). BGL2 produces substantial amounts of sophorose and cellobiose from glucose through the transglycosylation reaction (Saloheimo et al., 2002). This finding reinforces the fact that sophorose and cellobiose can act as a cellulase inducers in *T. reesei* (Dos Santos Castro et al., 2014) and shows that the transglycosylation reaction is under the control of CRE1-mediated CCR. On the other hand, in relation to cellulose, the detection of genes related to oxidative metabolism at the top of the list (Table 1) suggests that when *T. reesei* is growing in this complex carbon source, where glucose availability is limited by the cellulose degradation rate, the cells may be under constant semi-starvation conditions, which would induce stress-related metabolism, leading to the expression of genes that encode cellulases and hemicellulases. A similar proposal was put forward by Delmas et al. (2012) for *Aspergillus niger* grown in the presence of wheat straw.

In addition to the CAZy genes, some TF genes were also targets of carbon catabolite repression. The main TF gene upregulated in the presence of cellulose (ID 121682) contains a Gal4 domain, which is related to the growth of yeast in galactose (Traven et al., 2006). Given the similarity between the galactose and lactose structures and the fact that lactose induces cellulase production, repression of this TF coding gene is expected to prevent the induction of cellulases by lactose in the presence of readily available carbon sources. Another cellular process that appeared to be altered by CRE1 was nitrogen metabolism. The repression of the TF MeaB gene demonstrates the close relationship between CCR and the use of nitrogen compounds (Polley and Caddick, 1996) such as amino acids, indicating that proteins and related compounds are important alternatives to easily metabolizable carbon sources. Interestingly, the gene *ace2* (ID 78445), involved in the positive regulation of cellulase and xylanase genes (Aro, 2001), was upregulated in the Δ*cre1* strain in the presence of cellulose (Table 3). Although data suggest that *ace2* is not subject to CCR in the presence of glucose or lactose (Portnoy et al., 2011b), our results showed that it was subject to CCR in the presence of cellulose, suggesting that the response of a given gene to CCR depends on the carbon source available.

The most notable effect of catabolic repression involved the TF Zn2Cys6 *xyr1* (ID 122208) gene during growth in the presence of glucose. It has been postulated that *xyr1* is not regulated by induction but solely by carbon catabolite (de)repression (Stricker et al., 2007). However, Portnoy et al. concluded that *xyr1* is induced by lactose and that its full induction requires the positive action of CRE1 (Portnoy et al., 2011b). There is consensus that XYR1 is the main positive regulator of cellulase and xylanase genes in *T. reesei* (Mach-Aigner et al., 2008). In the presence of a readily available carbon source such as glucose, we observed that that *xyr1* gene was directly repressed by CRE1. It is probable that CRE1 binds to the promoter region of *xyr1* and represses the gene's expression. Thus, all genes regulated by XYR1 become repressed, including most cellulolytic and xylanolytic enzymes. However, because the studies cited above observed different patterns of gene expression regulated by XYR1, we assume that the requirement for CRE1 in XYR1 regulation depends on the available carbon source and that regulation can occur in a direct or indirect manner.

Genes that encode to transporter were also strongly repressed by CRE1. These genes encode proteins that transport molecules (including sugars) between the intracellular and extracellular spaces. Studies with *A. niger* also showed that numerous genes involved in transport were upregulated in the presence of sugarcane bagasse and repressed by glucose (de Souza et al., 2011). Similarly, Ries et al. (2013) demonstrated that several transporter genes in *T. reesei* were highly transcribed in the presence of wheat straw and repressed in the presence of glucose, including seven MFS permeases, a xylose transporter, two oligopeptide transporters, and an iron transporter. Here, we demonstrated that the genes encoding MFS permeases ID 3405, ID 47710 and ID 50618 were subject to CCR by CRE1 (indirectly and directly, respectively) during growth in the presence of glucose. The MFS permeases ID 3405 and ID 47710, named Ctr1 and Stp1, respectively, have also been studied by Zhang et al. (2013), who showed that the proteins play a key role in the cellulolytic signaling process, acting as sophorose and cellobiose transporters, respectively. Our data also demonstrated that both *ctr1* and *stp1* genes were targeted by CRE1 during growth in glucose (Tables 1 and S6) and *ctr1* gene was strongly induced in the presence of cellulose (Fig. 5), indicating that *ctr1* is repressed when there is a readily metabolizable carbon source in the medium, which probably reduces the uptake of cellulase inducer. The *ctr1* gene is also upregulated in the presence of lactose (Ivanova et al., 2013). The similar structures of sophorose and lactose could explain this finding. However, further studies are needed in order to understand this transporter's mechanism of action.

Genes related to iron homeostasis and copper transport were also upregulated in the *T. reesei* Δ*cre1* strain in the presence of cellulose and glucose (Table 1). These ions participate in essential processes, such as electron transport through the respiratory chain to produce ATP (Dmitriev et al., 2013), suggesting that the metabolism of the fungus decelerates when CRE1 is active. Copper and iron can also act as final electron acceptors for the enzyme cellobiose dehydrogenase (enzyme which in *T. reesei* has possible candidates to its function), which oxidizes several sugars, such as cellobiose, glucose oligomers, and cellulose (Dos Santos Castro et al., 2014). The reduction of Fe^{3+} to Fe^{2+} can generate hydroxyl radicals through the Fenton reaction, promoting the depolymerization of polysaccharides and structural modification of cell wall lignins (Henriksson et al., 2000). Moreover, Quinlan et al. demonstrated that copper is a cofactor for a recently described class of enzymes, termed GH61 glycoside hydrolases, that act directly on cellulose to make it more accessible to traditional cellulases (Quinlan et al., 2011). Consequently, in the presence of cellulose, repression of genes related to copper transport by CRE1 would retard cellulose degradation. Studies by Bischof et al. (2013) have shown that genes related to iron homeostasis, including those encoding iron transporters and iron reductase enzymes, are overexpressed in the presence of wheat straw, relative to expression in the presence of glucose, in *T. reesei*. Thus, the role of these proteins in cellulose degradation may be closely related to their repression by CRE1 in CCR.

The results of this study suggest that *T. reesei* can distinguish the glucose released from more complex carbon sources, such as cellulose, from that readily available in the medium (i.e., medium with a high glucose concentration). In the two cases, the transcription factor CRE1 regulated different sets of genes. Genes encoding proteins related to oxidative metabolism, such as electron trans-

porters and enzymes with oxidoreductase functions, were especially repressed by CRE1 in the presence of cellulose, indicating that the main genes repressed during CCR in this condition are related to the initial access of traditional cellulases to the cellulose polymer. The CCR mechanisms in different carbon sources are not necessarily the same in different species. In *N. crassa*, during the growth of a Δ*cre1* strain in the presence of cellulose, most genes repressed by CRE1 encode cellulases, hemicellulases, and proteins related to plant cell wall degradation (Arkowitz et al., 2011). No copper transporter genes were upregulated in the *N. crassa* Δ*cre1* strain in cellulose. However, a ferric-chelate reductase coding gene was repressed by CRE1 in the presence of cellulose (Arkowitz et al., 2011), as observed in *T. reesei*. It is logical to assume that in the presence of a low concentration of glucose, the fungus does not stop the transcription/translation machinery; restarting it would entail an unnecessary waste of energy. Instead, the fungus prefers to "switch on" or "switch off" the access of hydrolytic enzymes to the cellulose polymer. On the other hand, we found that genes related to the transport of small molecules were the main targets of CRE1-mediated CCR during growth in glucose (high concentration), indicating that CRE1 acts mainly to repress genes related to the entry of cellulase inducers (cellobiose/sophorose) into the cell, thus stopping the transcription of cellulase genes. Moreover, in the presence of glucose and in the presence of cellulose, CRE1 did not appear to act alone. Its regulation can occur indirectly, via other regulatory proteins that interact in a coordinated manner to ensure the adaptation of *T. reesei* to different growth conditions. To our knowledge, this is the first time that such a mechanism, in which multilevel control of cellulase expression and activity is regulated by CRE1, has been described in filamentous fungi. However, because the filamentous fungi that degrade lignocellulose polymers can respond differently to inducers (Dos Santos Castro et al., 2014; Znameroski et al., 2012), additional comparative studies are necessary to understand the CCR process.

5. Conclusions

Our analysis of the *T. reesei* transcriptome identified several genes affected by the transcription factor CRE1 in a carbon source-dependent manner, including genes for cellulolytic enzymes and TFs and genes related to the transport of nutrients and oxidative metabolism. However, most genes were indirectly regulated by CRE1. Our results also suggest a new mechanism for the action of CRE1 in different carbon sources. In cellulose, the repression of genes encoding copper transporters and ferric reductase enzymes impedes the access of traditional cellulases to the cellulose polymer. In contrast, during growth in glucose, CRE1 mainly accounts for the repression of genes related to the entry of cellulase inducers into the cell. We expect that this analysis will contribute to the understanding of the molecular mechanisms involved in the synthesis and regulation of cellulolytic enzymes, thus expanding the potential use of *T. reesei* in several industrial applications.

Acknowledgments

This work was supported by The State of São Paulo Research Foundation (FAPESP) (Proc. 2010/15683-8). We are grateful to Professor Christian Kubicek (TU-Wien) for providing the Δ*cre1* strain and helping with the experimental design of this study. RS-R was supported by a post-doctoral fellowship from FAPESP (2013/04125-2).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fgb.2014.10.009.

References

Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biol. 11, R106.
Arkowitz, R.A. et al., 2011. Identification of the CRE-1 Cellulolytic Regulon in *Neurospora crassa*. PLoS ONE 6, e25654.
Aro, N., 2001. ACEII, a novel transcriptional activator involved in regulation of Cellulase and Xylanase Genes of *Trichoderma reesei*. J. Biol. Chem. 276, 24309–24314.
Best, M. et al., 2011. Inhibition of trehalose breakdown increases new carbon partitioning into cellulosic biomass in *Nicotiana tabacum*. Carbohydr. Res. 346, 595–601.
Bischof, R. et al., 2013. Comparative analysis of the *Trichoderma reesei* transcriptome during growth on the cellulase inducing substrates wheat straw and lactose. Biotechnol. Biofuels 6, 127.
Castro, L.D. et al., 2014. Expression pattern of cellulolytic and xylanolytic genes regulated by transcriptional factors XYR1 and CRE1 are affected by carbon source in *Trichoderma reesei*. Gene Expr. Patterns 14, 88–95.
Cziferszky, A. et al., 2002. Phosphorylation positively regulates DNA binding of the carbon catabolite repressor Cre1 of *Hypocrea jecorina* (*Trichoderma reesei*). J. Biol. Chem. 277, 14688–14694.
de Souza, W.R. et al., 2011. Transcriptome analysis of *Aspergillus niger* grown on sugarcane bagasse. Biotechnol. Biofuels 4, 40.
Delmas, S. et al., 2012. Uncovering the genome-wide transcriptional responses of the filamentous fungus *Aspergillus niger* to lignocellulose using RNA sequencing. PLoS Genet. 8, e1002875.
Dmitriev, O.Y. et al., 2013. Direct regulation of cytochrome c oxidase by calcium ions. PLoS ONE 8, e74436.
Dos Santos Castro, L. et al., 2014. Comparative metabolism of cellulose, sophorose and glucose in *Trichoderma reesei* using high-throughput genomic and proteomic analyses. Biotechnol. Biofuels 7, 41.
Hakkinen, M. et al., 2012. Re-annotation of the CAZy genes of *Trichoderma reesei* and transcription in the presence of lignocellulosic substrates. Microb. Cell Fact. 11, 134.
Henriksson, G. et al., 2000. Is cellobiose dehydrogenase from *Phanerochaete* chrysosporium a lignin degrading enzyme? Biochim. et Biophys. Acta (BBA) – Protein Struct. Mole. Enzymol. 1480, 83–91.
Hong, M. et al., 2008. Structural basis for dimerization in DNA recognition by Gal4. Structure 16, 1019–1026.
Ilmen, M. et al., 1996. The glucose repressor gene cre1 of *Trichoderma*: isolation and expression of a full-length and a truncated mutant form. Mole. General Genet.: MGG 251, 451–460.
Ivanova, C. et al., 2013. Systems analysis of lactose metabolism in *Trichoderma reesei* identifies a lactose permease that is essential for cellulase induction. PLoS ONE 8, e62631.
Kubicek, C.P. et al., 2009. Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. Biotechnol. Biofuels 2, 19.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760.
Li, H. et al., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.
Lingner, U. et al., 2011. Functional characterization of a eukaryotic melibiose transporter. Plant Physiol. 156, 1565–1576.
Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25, 402–408.
Mach-Aigner, A.R. et al., 2008. Transcriptional regulation of xyr1, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*. Appl. Environ. Microbiol. 74, 6554–6562.
Peterson, R., Nevalainen, H., 2012. *Trichoderma reesei* RUT-C30-thirty years of strain improvement. Microbiology 158, 58–68.
Polley, S.D., Caddick, M.X., 1996. Molecular characterisation of meaB, a novel gene affecting nitrogen metabolite repression in *Aspergillus nidulans*. FEBS Lett. 388, 200–205.
Portnoy, T. et al., 2011a. The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei*: a master regulator of carbon assimilation. BMC Genom. 12, 269.
Portnoy, T. et al., 2011b. Differential regulation of the cellulase transcription factors XYR1, ACE2, and ACE1 in *Trichoderma reesei* strains producing high and low levels of cellulase. Eukaryot Cell. 10, 262–271.
Quinlan, R.J. et al., 2011. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. Proc. Natl. Acad. Sci. 108, 15079–15084.
Ries, L. et al., 2013. Genome-wide transcriptional response of *Trichoderma reesei* to lignocellulose using RNA sequencing and comparison with *Aspergillus niger*. BMC Genom. 14, 541.
Saloheimo, M. et al., 2002. Enzymatic properties and intracellular localization of the novel *Trichoderma reesei* beta-glucosidase BGLII (cel1A). Appl. Environ. Microbiol. 68, 4546–4553.

Sanchez, R., Zhou, M.M., 2009. The role of human bromodomains in chromatin biology and gene transcription. Curr. Opin. Drug Discov. Devel. 12, 659–665.
Santangelo, G.M., 2006. Glucose signaling in *Saccharomyces cerevisiae*. Microbiol. Mol. Biol. Rev. 70, 253–282.
Schmoll, M. et al., 2009. The G-alpha protein GNA3 of *Hypocrea jecorina* (*Anamorph Trichoderma reesei*) regulates cellulase gene expression in the presence of light. Eukaryot Cell. 8, 410–420.
Schuster, A., Schmoll, M., 2010. Biology and biotechnology of *Trichoderma*. Appl. Microbiol. Biotechnol. 87, 787–799.
Seiboth, B. et al., 2011. *Trichoderma reesei*: a fungal enzyme producer for cellulosic biofuels. In: Bernardes, M.A.d.S. (Ed.), Biofuel Production – Recent Developments and Prospects. InTech, Rijeka, Croatia, pp. 309–340.
Shannon, P. et al., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.
Silva-Rocha, R. et al., 2014. Deciphering the cis-regulatory elements for XYR1 and CRE1 regulators in *Trichoderma reesei*. PLoS ONE 9, e99366.
Strauss, J. et al., 1995. Crel, the carbon catabolite repressor protein from *Trichoderma reesei*. FEBS Lett. 376, 103–107.
Stricker, A.R. et al., 2007. Xyr1 receives the lactose induction signal and regulates lactose metabolism in *Hypocrea jecorina*. FEBS Lett. 581, 3915–3920.
Tamura, K. et al., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
Thorvaldsdottir, H. et al., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings Bioinform. 14, 178–192.
Traven, A. et al., 2006. Yeast Gal4: a transcriptional paradigm revisited. EMBO Rep. 7, 496–499.
Vencio, R.Z. et al., 2006. BayGO: bayesian analysis of ontology term enrichment in microarray data. BMC Bioinformatics 7, 86.
Zeng, L., Zhou, M.-M., 2002. Bromodomain: an acetyl-lysine binding domain. FEBS Lett. 513, 124–128.
Zhang, W. et al., 2013. Two major facilitator superfamily sugar transporters from *Trichoderma reesei* and their roles in induction of cellulase biosynthesis. J. Biol. Chem. 288, 32861–32872.
Znameroski, E.A. et al., 2012. Induction of lignocellulose-degrading enzymes in *Neurospora crassa* by cellodextrins. Proc. Natl. Acad. Sci. USA 109, 6012–6017.

The impact of chromatin remodelling on cellulase expression in *Trichoderma reesei*

Thiago M Mello-de-Sousa[1], Alice Rassinger[1], Marion E. Pucher[2], Lilian dos Santos Castro[3], Gabriela F Persinoti[4], Rafael Silva-Rocha[3], Marcio J. Poças-Fonseca[5], Robert L Mach[1], Roberto Nascimento Silva[3], Astrid R Mach-Aigner[1§]

[1] Department for Biotechnology and Microbiology, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorfer Str. 1a, A-1060 Wien, Austria

[2] present address:

[3] Department of Biochemistry and Immunology, Ribeirão Preto Medical School, University of São Paulo 14049-900, Ribeirão Preto, SP, Brazil

[4] Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, São Paulo, Brazil

[5] Department of Genetics and Morphology, Institute of Biological Sciences, University of Brasilia, Brasília-DF, Brazil

[§]Corresponding author

Email addresses:

TMMS: tdemello@mail.tuwien.ac.at

AR: alice.rassinger@tuwien.ac.at

MEP: Marion.Pucher@gmx.at

LSC: liliancastro@usp.br

GFP: gabi.felix@gmail.com

RSR: silvarochar@gmail.com

MJPF: mpossas@unb.br

RLM: robert.mach@tuwien.ac.at

RNS: rsilva@fmrp.usp.br

ARMA: astrid.mach-aigner@tuwien.ac.at

Abstract

Background

Trichoderma reesei is used for industry-scale production of plant cell wall-degrading enzymes, in particular cellulases, but also xylanases. The expression of the encoding genes was so far primarily investigated on the level of transcriptional regulation by regulatory proteins. Otherwise, the impact of chromatin remodelling on gene expression received hardly any attention. In this study we aimed to learn if the chromatin status changes in context to the applied conditions (repressing/inducing), and if the presence or absence of the essential transactivator, the Xylanase regulator 1 (Xyr1), influences the chromatin packaging.

Results

Comparing the results of chromatin accessibility real-time PCR analyses and gene expression studies of the two prominent cellulase-encoding genes, *cbh1* and *cbh2*, we found that the chromatin opens during sophorose-mediated induction compared to D-glucose-conferred repression. In the strain bearing a *xyr1* deletion the sophorose mediated induction of gene expression is lost and the chromatin opening is strongly reduced. In all conditions the chromatin got denser when Xyr1 is absent. In case of the xylanase-encoding genes, *xyn1* and *xyn2*, the result was similar concerning the condition-specific response of the chromatin compaction. However, the difference in chromatin status provoked by the absence of Xyr1 is less pronounced. A more detailed investigation of the DNA accessibility in the *cbh1* promoter showed that the deletion of *xyr1* changed the *in vivo* footprinting pattern. In particular, we detected increased hypersensitivity on Xyr1-sites and stronger protection of Cre1-sites. Looking for the players directly causing the observed chromatin remodelling, a whole transcriptome shotgun sequencing revealed that 15 genes encoding putative chromatin

remodelers are differentially expressed in response to the applied condition and two amongst them are differentially expressed in absence of Xyr1.

Conclusions

The regulation of xylanase and cellulase expression in *T. reesei* is not restricted to the action of transcription factors but is clearly related to changes in the chromatin packaging. Both the applied condition and the presence of Xyr1 influence chromatin status. This finding indicates options for an additional level of strain development.

Background

In nature, *Trichoderma reesei* is as a saprophytic fungus an excellent producer of enzymes involved in plant cell wall degradation (PCWD). In industry, these enzymes are used for a number of applications: xylanases are used for example in food industry as a baking agent and for clarification of juice and wine [1] or in the paper industry for de-inking [2]. Cellulases from *T. reesei* are important in textile industry for example for fibre polishing [3] or in the paper industry for recycling processes [2]. Nowadays, it is aimed to produce as one alternative fuel cellulosic ethanol. In this process *T. reesei* cellulases are applied to break down lignocellulose to D-glucose, which can be used subsequently in the sugar-to-ethanol fermentation (*e.g.* [4, 5] and citations therein). Due to the manifold applications of these enzymes many research studies have focused on this organism, its PCWD enzyme expression, and finally, the regulation of the encoding genes. Most of these studies were performed in the wild-type strain QM6a [6] or the high-cellulase producing mutant strain QM9414 [7]. A genome-wide analysis identified 10 cellulolytic and 16 hemicellulolytic enzyme-encoding genes in *T. reesei* [8], of which the most prominent cellulases are the cellobiohydrolases CBHI and CBHII (EC 3.2.1.91) [9] and the most studied xylanases are the endo-ß-1,4-xylanases XYNI and XYNII (EC 3.2.1.8) [10]. The mentioned

research efforts led further to the identification of transcription factors involved in the regulation of the expression of genes coding for PCWD enzymes on the transcriptional level. The most important transactivator is the Xylanase regulator 1 (Xyr1), which is absolutely essential for expression of both, xylanase and cellulase-encoding genes [11]. However, it should be noted that only the cellulase expression strictly follows the induction/repression pattern of the *xyr1* gene [12]. The *xyr1* gene itself is usually expressed at a low level and can be induced by the disaccharide sophorose formed via transglycosylation [12, 13]. Otherwise, the xylanase expression depends on Xyr1, but the transcript levels of these genes do not strictly reflect *xyr1* transcript levels [11, 12]. The most important repressor is the Carbon catabolite repressor 1 (Cre1) [14], which mediates carbon catabolite repression (CCR) in presence of high amounts of easily usable carbon sources, such as D-glucose or D-xylose. Cre1 exerts its repressing function on both, the genes coding for the PCWD enzymes and the gene coding for their activator, *xyr1* (*e.g.* [13, 15]). The different response of *T. reesei*'s transcriptome and secretome to cellulose, sophorose, and D-glucose was just recently investigated in a comparative high-throughput genomic and proteomic study [16]. While a lot is known about the transcriptional regulation of *T. reesei*'s PCWD enzyme-encoding genes by regulatory proteins (reviewed in [17]), so far hardly anything was investigated concerning the impact of the chromatin status on their gene expression. Only for Cre1 it was already earlier suggested that it might influence chromatin remodelling [18]. More recently, it was reported that it is involved in nucleosome positioning [19], and that a truncated version of Cre1, which is present in CCR-released, cellulase hyper-producing strains, supports the opening of chromatin in Cre1-target genes [20]. However, taking into account that chromatin status generally is believed to be a crucial factor in gene expression, this topic did not

receive much attention in *T. reesei* yet. Therefore, in this study, we aimed to learn if the opponent of Cre1, the transactivator Xyr1, is also involved in chromatin remodelling, and if this happens in a condition (inducing/repressing carbon source)-dependent way. We used chromatin accessibility real-time PCR (CHART-PCR) for determining the chromatin status of the genes encoding the mentioned, four major PCWD enzymes and compared this with their gene expression. The results prompted us to have a more detailed investigation of the *cbh1* promoter by *in vivo* footprinting analyses. Finally, we used whole transcriptome shotgun sequencing (WTSS) to identify genes putatively involved in chromatin remodelling that are differentially expressed with regards to the applied condition and/or the absence or presence of Xyr1.

Results

Decreased cellulase gene expression in absence of Xyr1 goes along with denser chromatin

It is well known that Xyr1 is an essential activator of cellulase gene expression [11]. However, so far it was not investigated if the deletion of Xyr1 additionally influences the chromatin status in the fungus. In order to study this, the wild-type strain and the *xyr1* deletion strain were pre-grown and transferred to sophorose (inducing conditions), D-glucose (repressing condition) or no carbon source-containing medium (reference condition) and were incubated for 3 h. By applying CHART-PCR analysis we investigated the chromatin packaging of the core promoter region (bearing the TATA-box) and one upstream regulatory region (URR) bearing Xyr1-binding sites (5'-GGC(T/A)$_3$-3'; [21]) and/or Cre1-binding sites (5'-SYGGRG-3'; [14]) of the *cbh1* and *cbh2* genes each. For overviews on the investigated regions please refer to Fig. 1a and b. Complementary, we investigated the transcript levels of these genes by reverse transcription, quantitative PCR (qPCR) to see if the expression is related to

chromatin accessibility. The expression of *cbh1* and *cbh2* is repressed on D-glucose in both strains and induced by sophorose in the wild-type strain (Fig. 2a, b). The induction is lost in the *xyr1* deletion strain aside from a small increase in gene expression on sophorose compared to D-glucose. Altogether, we observed in both strains a condition-dependent change (*i.e.* sophorose-mediated opening) of chromatin that went along with a change (*i.e.* sophorose-mediated increase) in gene expression. However, comparing the strains under the same condition, the chromatin was always more closed in the *xyr1* deletion strain compared to the wild-type strain (Fig. 2a, b) indicating a contribution of Xyr1 to a general (*i.e.* condition-independent) opening of chromatin in upstream regions of the cellulase-encoding genes.

Xylanase gene repression in absence of Xyr1 is not strictly related to chromatin compaction

In an analogous analysis we investigated the chromatin status of the core promoter and an URR of the *xyn1* and *xyn2* genes each and compared this to the expression of the respective genes. For overviews on the regions investigated by CHART-PCR please refer to Fig. 1c and d. In the wild-type strain the repression on D-glucose, the basal expression on D-xylose, and the induction on sophorose coincided with the increasing opening of chromatin (Fig. 3a, b). Otherwise, in the *xyr1* deletion strain the gene expression was at a similar low level (repressed) independent from the tested condition, while the chromatin packaging differed between the conditions. Interestingly, the chromatin accessibility on sophorose was even similar between the Δ*xyr1*-strain and the wild-type strain (except the URR of *xyn1*) but the sophorose-mediated induction was completely lost in the Δ*xyr1*-strain (Fig. 3a, b). Summarizing, we detected - similar as in case of the cellulase-encoding genes - an induction-specific opening of chromatin together with increasing gene expression in the wild-type strain. However, different from the cellulases, xylanase expression was repressed in the

Δ*xyr1*-strain even if the chromatin status differed condition-dependently. Notably, a comparison between the chromatin compaction between the strains under the same conditions did not point to an involvement of Xyr1 to chromatin opening.

The absence of Xyr1 changes DNA accessibility in the *cbh1* promoter

Since we observed an induction-specific opening of chromatin that went along with increase of gene expression in presence of Xyr1 and a closing of chromatin together with gene repression in absence of Xyr1 in case of the cellulase-encoding genes, we aimed to have a more detailed investigation on the DNA accessibility of the promoter. Therefore, we performed *in vivo* footprinting analyses of the *cbh1* promoter. Two URRs bearing Xyr1-binding sites and/or Cre1-binding sites and the core promoter bearing Xyr1-binding sites close to the TATA-box were investigated (Fig. 4a). The wild-type strain and the Δ*xyr1*-strain were pre-grown on glycerol and then incubated on D-glucose or sophorose for 3 h followed by dimethyl sulphate (DMS)-induced *in vivo* methylation. From figure 4 (b-d) the footprinting pattern of the *xyr1* deletion strain compared to the wild-type strain for the three investigated regions can be inferred. The first investigated URR bears next to a single Cre1-site and a single Xyr1-site, also two Xyr1-sites arranged as inverted repeat with a spacing of 11 bp, which was reported to be the functional binding motif *in vivo* [22]. Under both, repressing and inducing conditions we could detect strong differences in the footprinting pattern of the two strains (Fig. 4b). In particular on sophorose, we observed an increased hypersensitivity towards DNA methylation on the Xyr1-sites in the Δ*xyr1*-strain compared to the wild-type strain, whereas the Cre1-site was stronger protected (Fig. 4b). The second investigated URR bears a functional Cre1 double site [23]. Here, we detected changes in DNA accessibility in the Δ*xyr1*-strain compared to the wild-type strain on D-glucose, but only minor changes on sophorose. The third

investigated URR bears three Xyr1-binding sites arranged *in tandem*. In this case, we detected just a few differences between the two strains, however, most of them on or close to the Xyr1-sites (Fig. 4d).

Identification of differentially expressed genes potentially involved in chromatin remodelling

To learn more about the mechanisms responsible for the chromatin remodelling in context to both, the applied condition and the presence or absence of Xyr1, we used WTSS. Therefore, a *xyr1* deletion strain and its parental strain QM9414 were again exposed to repressing conditions (growth on D-glucose) and inducing conditions (incubation on sophorose). Based on the results obtained by the WTSS, we analyzed the gene expression profiles of 136 candidate genes involved in chromatin structure and dynamics according to the eukaryotic orthologous groups (KOG) in the *T. reesei* genome database (http://genome.jgi-psf.org/Trire2/Trire2.home.html). An overview on these 136 genes is provided here. For the differential expression analysis, a two-fold change cut-off, *i.e.* $\log_2$ fold change ≥ 1 or ≤ -1 and an adjusted p-value ≤ 0.05, was used as threshold. Concerning the first part of our question, *i.e.* the observed differences in chromatin status dependent on the applied condition, we identified 15 genes differentially expressed on sophorose compared to D-glucose in the wild-type strain (listed in Table 1). Concerning, the second part of our question, *i.e.* the influence of Xyr1 on the expression profiles of these genes we examined which ones were differentially expressed in Δ*xyr1*-strain compared to the wild-type strain under sophorose induction. Out of the fifteen genes responding to the applied condition, two genes are additionally differentially expressed in the Δ*xyr1*-strain (transcript ID 53947 and 73708). Notably, the gene with transcript ID 73708, encoding a putative heterochromatin-associated protein, was down-regulated on sophorose compared to D-glucose and up-regulated in absence of Xyr1.

Discussion

The aim of this study was to learn more about the contribution of the chromatin compaction to the regulation of gene expression of PCWD enzymes in *T. reesei*. Altogether, we found for all investigated genes that their induced expression is accompanied by an opening of chromatin. However, we observed differences between cellulase- and xylanase-encoding genes concerning the involvement of Xyr1 in chromatin remodelling: the chromatin of the upstream regions of the cellulase-encoding genes was more compact under all tested conditions when Xyr1 was missing. This indicates a contribution of Xyr1 to chromatin opening in their case. Perhaps this finding is one explanation for the before reported, condition-dependent transcript level pattern of the *cbh1* and *cbh2* genes that exactly follow the one of the *xyr1* gene [12]. When for example under repressing conditions less *xyr1* is present, the positive influence of Xyr1 on chromatin opening might be reduced and this would cause a decrease in cellulase-encoding gene expression. However, the earlier observation and the result from this study about the involvement of Xyr1 in chromatin opening suggests a regulation of the cellulase-encoding genes that is dominated by Xyr1, maybe without important influence of other regulatory proteins (except Cre1). In case of the xylanase-encoding genes we also detected a condition-dependent induction of gene expression that was accompanied by chromatin opening in the wild-type strain. However, the involvement of Xyr1 is different than in case of the cellulase-encoding genes. In absence of Xyr1 the chromatin of the upstream regions of the xylanase-encoding genes did not became more compact, but gene expression decreased under all conditions. For example, in case of *xyn2*, the URR had the same chromatin accessibility under non-repressing conditions (sophorose, D-xylose) in the Δ*xyr1*-strain as in the wild-type strain but the gene expression was strongly repressed

in absence of Xyr1. The fact that gene expression can be repressed simultaneously with enhanced chromatin accessibility might be explained by a generally better access for all kinds of regulatory proteins including repressor proteins. Another possible explanation would be that the absence of Xyr1 simple overrules the regulation by chromatin opening. Anyway, during this study it became obvious that the activating function of Xyr1 on xylanase-encoding gene expression is not exerted on the chromatin level. There are earlier reports on generally different, condition-dependent transcript level patterns of the *xyn1* and *xyn2* genes compared to the *xyr1* gene [12]. One example is the low basal *xyn2* gene expression on D-glucose (*e.g.* [24, 25]) that is not detectable for the *xyr1* gene [13]. All these findings together strongly indicate that additional regulatory factors and mechanisms, which for example are responsible for chromatin opening under inducing conditions, need to be involved.

A whole transcriptome analysis was used to identify genes classified as chromatin remodelers in *T. reesei*, which are differentially expressed dependent on the applied condition (inducing/repressing). Notably, 15 genes are differentially expressed in the wild-type strain (compare Table 1), whereas only 10 genes responded in a condition-dependent manner in the Δ*xyr1*-strain (data not shown). This again supports the assumption that Xyr1 is generally involved in chromatin remodelling mechanisms. The identification of two putative chromatin remodelers, which are under the control of Xyr1 (directly or via expression of other regulatory proteins), point to an indirect role of Xyr1 in chromatin remodelling. Moreover, it can be speculated that Xyr1 additionally recruits chromatin-remodelling proteins in a differential manner towards the promoters of the cellulase- and xylanase-encoding genes. This would be a further explanation for the observed differences concerning the influence of Xyr1 on their chromatin status.

Conclusions

Investigations on the level of chromatin packaging revealed that the transcription factor Xyr1 does exert its activating function on cellulase-encoding gene expression – next to other possible mechanisms - by an opening of chromatin. Such an involvement of Xyr1 in chromatin opening was not found in case of the xylanase-encoding genes. The application of WTSS indentified one chromatin remodeler that is down-regulated under inducing conditions and up-regulated if Xyr1 is missing. According to the results of the present study, this is a target for engineering strains with an enhanced cellulase expression.

Methods

Fungal strains

The following *T. reesei* strains were used throughout this study: the wild-type strain QM6a (ATCC 13631), and a corresponding *xyr1* deletion strain (this study), QM9414 (ATCC 26921), and a QM9414 strain bearing a *xyr1* deletion [11]. All strains were maintained on malt extract agar.

Growth conditions

For carbon source replacement experiments mycelia were pre-cultured in 1-L-Erlenmeyer flasks on a rotary shaker (180 rpm) at 30 °C for 24 h in 250 mL of Mandels-Andreotti (MA) medium [26] supplemented with 1 % (w/v) glycerol as sole carbon source. A total of 10^9 conidia per litre (final concentration) were used as inoculum. Pre-grown mycelia were washed and equal amounts were resuspended in 20 ml MA media containing 1 % (w/v) D-glucose or 2 mM sophorose as sole carbon source or no carbon source and were incubated for 3 h.

For direct cultivation experiments conidia were incubated in 200 mL MA medium containing 2 % (w/v) glucose as the sole carbon source for 24 and 48 hours. Samples were derived from three biological replicates.

Deletion of *xyr1* from the genome of the *T. reesei* wild-type strain

The deletion of the *xyr1* gene was essentially performed as described in [11] using QM6aΔ*tmus53* [27] as a recipient strain in a fungal protoplast transformation following the protocol described in [28]. The *A. nidulans amdS* gene was used as a marker [29].

CHART-PCR

DNase I digestion of chromatin and DNA extraction were carried out as described before [20]. qPCR analysis of the DNase I-treated samples was performed to measure the relative abundance of target regions. PCRs were performed in triplicates in a Rotor-Gene Q system (Qiagen, Hilden, Germany) using the amplification mixture (final volume 20 µL) and cycling conditions described before [20]. Primer sequences are provided in Table 2. The amount of intact input DNA of each sample was calculated by comparing the threshold values of the PCR amplification plots with a standard curve generated for each primer set using serial dilutions of genomic, undigested DNA. The chromatin accessibility index (CAI) was defined as: CAI = 1/(Ds/((Dc1+Dc2)/2)), where Ds is the amount of intact DNA detected for each target region and Dc1 and Dc2 are the amounts of intact DNA detect for the promoter regions of *sar1* and *act* respectively, used as reference genes for normalization.

Analysis of transcript levels

Fungal mycelia were homogenized in 1 mL of peqGOLDTriFast DNA/RNA/protein purification system reagent (PEQLAB Biotechnologie, Erlangen, Germany) using a FastPrep(R)-24 cell disrupter (MP Biomedicals, Santa Ana, CA, USA). RNA was

isolated according to the manufacturer's instructions, and the concentration was measured using the NanoDrop 1000 (Thermo Scientific, Waltham, US). Synthesis of cDNA from mRNA was carried out using the RevertAidTM H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Quantitative PCRs were performed in triplicates in a Rotor-Gene Q system (Qiagen). The amplification mixture (final volume 15 µL) contained 7.5 µL 2 x iQ SYBR Green Mix (Bio-Rad, Hercules, USA), 100 nM forward and reverse primer and 2.5 µL cDNA (diluted 1:20). Primer sequences are provided in Table 2. Cycling conditions and control reactions were performed as described previously [30]. Data normalization using *sar1* and *act* as reference genes and calculations were performed as published previously [30].

In vivo footprinting

In vivo methylation using DMS followed by ligation mediated PCR was performed as described previously [31]. FAM-labelled fragments were generated by a PCR reaction using RG89 and RG90 or RG83 and RG84 for an URR or a TATA-box containing core region within the *cbh1* promoter, respectively. Primer sequences are provided in Table 2. FAM-labelled fragments were analyzed by capillary gel electrophoresis (Microsynth, Balgach, Switzerland) and results were analyzed using the program ivFAST [31].

WTSS

Total RNA was extracted from fungal mycelia using TRIzol® RNA Kit (Life Technologies, part of Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. RNA concentration was determined by spectrophotometry at 260/280 nm and RNA integrity was tested by the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and gel electrophoresis

(1% agarose). The total RNA of the biological replicates was pooled, lyophilized, and stored using the RNAstable® Tube Kit (Biomatrica, San Diego, CA, USA). Barcoded libraries were prepared using the TruSeq RNA Sample Prep kit (Illumina, San Diego, CA, USA) and sequenced by LGC Genomics GmbH (Berlin, Germany) using the Illumina HiSeq 2000 platform.

WTSS data analysis

Sequences from approximately 144 million 100 bp paired-end reads were quality-filtered and mapped to the *Trichoderma reesei* 2.0 reference genome, (http://genome.jgi-psf.org/Trire2/Trire2.home.html) using the Bowtie aligner version 0.12.8 [32] allowing two mismatches and only unique alignments. The SAMtools version 0.1.18 [33] was used to process the alignments files, which were visualized using the Integrative Genomics Viewer [34, 35]. Bioconductor DESeq package version 1.10.1 [36] was utilized for normalization, using the median log deviation, and for the differential expression analysis, applying a two-fold change cut-off, *i.e.* $\log_2$-fold change ≥ 1 or ≤ -1 and an adjusted p-value ≤ 0.05 were used as thresholds. The $\log_2$-fold change was calculated according to the equation:

$$\log_2\text{-fold change} = \log_2 \frac{baseMeanB}{baseMeanA}, \text{ where:}$$

baseMeanB is the base mean (*i.e.* mean of the counts divided by the size factors) for the counts for condition A. Size factors for the columns of the matrix countsB; baseMeanA is the base mean for the counts for condition A.

List of abbreviations

CAI, chromatin accessibility index; CCR, carbon catabolite repression; CHART-PCR, chromatin accessibility real-time PCR; Cre1, Carbon catabolite repressor 1; DMS, dimethyl sulphate; ivFP, *in vivo* footprinting; KOG, eukaryotic orthologous groups;

MA, Mandels Andreotti; PCWD, plant cell wall-degrading / plant cell wall degradation; qPCR, quantitative PCR; URR, upstream regulatory region; WTSS, whole transcriptome shotgun sequencing; Xyr1, Xylanase regulator 1.

Competing interests

The authors declare that they do not have any competing interests.

Authors' contributions

TMMS performed CHART-PCR and transcript analyses and participated in analysis of WTSS data. AR carried out the *in vivo* footprinting analyses. MEP constructed the *xyr1* deletion strain in QM6a. LSC prepared samples for WTSS and participated in analysis of corresponding data. GFP and RSR participated in analysis of WTSS data. RLM and MJPF participated in conception and revision of the manuscript. RNS participated in design of the study and supervision of experiments. ARMA participated in design of study, supervision of experiments, and prepared the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by two grants from the Austrian Science Fund (FWF): V232-B20, P24851 given to ARMA, by a doctoral program of Vienna University of Technology ("CatMat"), and by The State of São Paulo Research Foundation (FAPESP) (proc. 2010/15683-8). We thank Stephanie Steinberger for help with the footprinting analyses.

References

1. Galante YM, De Conti A, Monteverdi R: **Application of *Trichoderma* enzymes in the food and feed industries**. In: *Trichoderma & Gliocladium.*

Edited by Harman GE, Kubicek, C.P., vol. 2. London: Taylor & Francis; 1998: 327-342.
2. Buchert J, Oksanen T, Pere J, Siika-aho M, Suurnäkki A, Viikari L: **Applications of *Trichoderma reesei* enzymes in the pulp and paper industry**. In: *Trichoderma & Gliocladium.* Edited by Harman GE, Kubicek CP, vol. 2. London, UK: Taylor & Francis Ltd.; 1998: 343-357.
3. Galante YM, de Conti, A., Monteverdi, R.: **Application of *Trichoderma* enzymes in the textile industry**. In: *Trichoderma & Gliocladium.* Edited by Kubicek CP, Harman, G.E., vol. 2. London: Taylor & Francis; 1998: 311-325.
4. Gusakov AV: **Alternatives to *Trichoderma reesei* in biofuel production**. *Trends Biotechnol* 2011, **29**(9):419-425.
5. Wilson DB: **Cellulases and biofuels**. *Curr Opin Biotechnol* 2009, **20**(3):295-299.
6. Mandels M, Reese ET: **Induction of cellulase in *Trichoderma viride* as influenced by carbon sources and metals**. *J Bacteriol* 1957, **73**(2):269-278.
7. Mandels M, Weber J, Parizek R: **Enhanced cellulase production by a mutant of *Trichoderma viride***. *Appl Microbiol* 1971, **21**(1):152-154.
8. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D *et al*: **Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*)**. *Nat Biotechnol* 2008, **26**(5):553-560.
9. Teeri T, Salovouri, I., Knowles, J.: **The molecular cloning of the major cellulase gene from *Trichoderma reesei***. *Biotechnology* 1983, **1**:696-699.
10. Törrönen A, Mach RL, Messner R, Gonzalez R, Kalkkinen N, Harkki A, Kubicek CP: **The two major xylanases from *Trichoderma reesei*: characterization of both enzymes and genes**. *Biotechnology (N Y)* 1992, **10**(11):1461-1465.
11. Stricker AR, Grosstessner-Hain K, Würleitner E, Mach RL: **Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina***. *Eukaryot Cell* 2006, **5**(12):2128-2137.
12. Derntl C, Gudynaite-Savitch L, Calixte S, White T, Mach RL, Mach-Aigner AR: **Mutation of the Xylanase regulator 1 causes a glucose blind hydrolase expressing phenotype in industrially used *Trichoderma* strains**. *Biotechnol Biofuels* 2013, **6**(1):62.
13. Mach-Aigner AR, Pucher ME, Steiger MG, Bauer GE, Preis SJ, Mach RL: **Transcriptional regulation of *xyr1*, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina***. *Appl Environ Microbiol* 2008, **74**(21):6554-6562.
14. Strauss J, Mach RL, Zeilinger S, Hartler G, Stoffler G, Wolschek M, Kubicek CP: **Cre1, the carbon catabolite repressor protein from *Trichoderma reesei***. *FEBS Lett* 1995, **376**(1-2):103-107.
15. Mach RL, Strauss J, Zeilinger S, Schindler M, Kubicek CP: **Carbon catabolite repression of xylanase I (*xyn1*) gene expression in *Trichoderma reesei***. *Mol Microbiol* 1996, **21**(6):1273-1281.
16. dos Santos Castro L, Pedersoli WR, Antonieto AC, Steindorff AS, Silva-Rocha R, Martinez-Rossi NM, Rossi A, Brown NA, Goldman GH, Faca VM *et al*: **Comparative metabolism of cellulose, sophorose and glucose in Trichoderma reesei using high-throughput genomic and proteomic analyses**. *Biotechnol Biofuels* 2014, **7**(1):41.

17. Stricker AR, Mach RL, de Graaff LH: **Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*)**. *Appl Microbiol Biotechnol* 2008, **78**(2):211-220.
18. Portnoy T, Margeot A, Linke R, Atanasova L, Fekete E, Sandor E, Hartl L, Karaffa L, Druzhinina IS, Seiboth B *et al*: **The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei*: a master regulator of carbon assimilation**. *BMC Genomics* 2011, **12**:269.
19. Ries L, Belshaw NJ, Ilmén M, Penttilä ME, Alapuranen M, Archer DB: **The role of CRE1 in nucleosome positioning within the *cbh1* promoter and coding regions of *Trichoderma reesei***. *Appl Microbiol Biotechnol* 2014, **98**(2):749-762.
20. Mello-de-Sousa TM, Gorsche R, Rassinger A, Pocas-Fonseca MJ, Mach RL, Mach-Aigner AR: **A truncated form of the Carbon catabolite repressor 1 increases cellulase production in *Trichoderma reesei***. *Biotechnol Biofuels* 2014, **7**(1):129.
21. Furukawa T, Shida Y, Kitagami N, Mori K, Kato M, Kobayashi T, Okada H, Ogasawara W, Morikawa Y: **Identification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in *Trichoderma reesei***. *Fungal Genet Biol* 2009, **46**(8):564-574.
22. Rauscher R, Würleitner E, Wacenovsky C, Aro N, Stricker AR, Zeilinger S, Kubicek CP, Penttilä M, Mach RL: **Transcriptional regulation of *xyn1*, encoding xylanase I, in *Hypocrea jecorina***. *Eukaryot Cell* 2006, **5**(3):447-456.
23. Ilmén M, Onnela ML, Klemsdal S, Keranen S, Penttilä M: **Functional analysis of the cellobiohydrolase I promoter of the filamentous fungus *Trichoderma reesei*** *Mol Gen Genet* 1998, **257**(3):386.
24. Würleitner E, Pera L, Wacenovsky C, Cziferszky A, Zeilinger S, Kubicek CP, Mach RL: **Transcriptional regulation of *xyn2* in *Hypocrea jecorina***. *Eukaryot Cell* 2003, **2**(1):150-158.
25. Zeilinger S, Mach RL, Schindler M, Herzog P, Kubicek CP: **Different inducibility of expression of the two xylanase genes *xyn1* and *xyn2* in *Trichoderma reesei***. *J Biol Chem* 1996, **271**(41):25624-25629.
26. Mandels M: **Applications of cellulases**. *Biochem Soc Trans* 1985, **13**(2):414-416.
27. Steiger MG, Vitikainen M, Uskonen P, Brunner K, Adam G, Pakula T, Penttilä M, Saloheimo M, Mach RL, Mach-Aigner AR: **Transformation system for *Hypocrea jecorina* (*Trichoderma reesei*) that favors homologous integration and employs reusable bidirectionally selectable markers**. *Appl Environ Microbiol* 2011, **77**(1):114-121.
28. Gruber F, Visser J, Kubicek CP, de Graaff LH: **The development of a heterologous transformation system for the cellulolytic fungus *Trichoderma reesei* based on a *pyrG*-negative mutant strain**. *Curr Genet* 1990, **18**(1):71-76.
29. Hynes MJ, Corrick CM, King JA: **Isolation of genomic clones containing the *amdS* gene of *Aspergillus nidulans* and their use in the analysis of structural and regulatory mutations**. *Mol Cell Biol* 1983, **3**(8):1430-1439.
30. Steiger MG, Mach RL, Mach-Aigner AR: **An accurate normalization strategy for RT-qPCR in *Hypocrea jecorina* (*Trichoderma reesei*)**. *J Biotechnol* 2010, **145**(1):30-37.

31. Gorsche R, Jovanovic B, Gudynaite-Savitch L, Mach RL, Mach-Aigner AR: **A highly sensitive *in vivo* footprinting technique for condition-dependent identification of *cis* elements** *Nucleic Acids Res* 2013, **42**(1):e1.
32. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform**. *Bioinformatics* 2009, **25**(14):1754-1760.
33. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.
34. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: **Integrative genomics viewer**. *Nat Biotechnol* 2011, **29**(1):24-26.
35. Thorvaldsdóttir H, Robinson JT, Mesirov JP: **Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration**. *Brief Bioinform* 2013, **14**(2):178-192.
36. Anders S, Huber W: **Differential expression analysis for sequence count data**. *Genome Biol* 2010, **11**(10):R106.

Figures

Figure 1 - Overview on the upstream sequence of the investigated genes encoding PCWD enzymes

The regions investigated by CHART-PCR are indicated by black bars. The core promoter region covering the TATA-box (core) and an URR of the *cbh1* (**a**), *cbh2* (**b**), *xyn1* (**c**), and *xyn2* (**d**) genes each are depicted. DNA-binding sites of Xyr1 and Cre1 are indicated by orange and purple triangles, respectively. The orientation of the triangle represents the orientation of the binding motif. The scale at the top indicates distance from ATG in bp.

Figure 2 - Transcript and CHART analysis of cellulase-encoding genes in presence and absence of Xyr1

The *T. reesei* wild-type strain (dots) and the Δ*xyr1*-strain (triangles) were pre-grown on glycerol and thereafter incubated on D-glucose (G) or sophorose (S) for 3 h. The core promoter region (red) and an URR (blue) of *cbh1* (**a**) and *cbh2* (**b**) genes were investigated. The gene expression analysis was performed by cDNA synthesis followed by qPCR, and transcript levels are depicted on the x-axis. CHART-PCR was performed by DNaseI digestion followed by qPCR, and chromatin accessibility

indices (CAIs) are depicted on the y-axis. In both cases *sar1* and *act* genes were used for data normalization and the wild-type strain incubated without carbon source for 3 h was the reference condition. The dashed line indicates transcript level of the reference condition, *i.e.* levels above are considered induced and levels below are considered repressed. All values are means from measurements in triplicates and three biological experiments (cultivations). Standard deviations were below 5 %. Diagrams are identically scaled.

Figure 3 - Transcript and CHART analysis of xylanase-encoding genes in presence and absence of Xyr1

The *T. reesei* wild-type strain (dots) and the Δ*xyr1*-strain (triangles) were pre-grown on glycerol and thereafter incubated on D-glucose (G), D-xylose (X) or sophorose (S) for 3 h. The core promoter region (red) and an URR (blue) of *xyn1* (**a**) and *xyn2* (**b**) genes were investigated. The gene expression analysis was performed by cDNA synthesis followed by qPCR, and transcript levels are depicted on the x-axis. CHART-PCR was performed by DNaseI digestion followed by qPCR, and CAIs are depicted on the y-axis. In both cases *sar1* and *act* genes were used for data normalization and the wild-type strain incubated without carbon source for 3 h was the reference condition. The dashed line indicates transcript level of the reference condition, *i.e.* levels above are considered induced and levels below are considered repressed. All values are means from measurements in triplicates and three biological experiments (cultivations). Standard deviations were below 5 %. Diagrams are identically scaled.

Figure 4 - *In vivo* footprinting analyses of the *cbh1* promoter in presence and absence of Xyr1

The *T. reesei* wild-type strain QM6a and the Δ*xyr1*-strain were pre-grown on glycerol and then incubated on D-glucose or sophorose for 3 h followed by DMS-induced *in*

vivo methylation. (**a**) Schematic drawing of the *cbh1* promoter and the investigated regions (indicated by green lines). Two URRs (**b**, **c**) bearing functional Xyr1-binding sites (orange) or Cre1-sites (purple) and the core promoter region (**d**) bearing Xyr1-binding sites (orange) were investigated on the forward strand. Numbers indicate the position of the base upstream from ATG. Analysis of data and visualization was performed using ivFAST [31]. Only signals that are statistically different are considered; protected bases are highlighted in red shades and hypersensitive bases are highlighted in blue shades; the three colour intensities each correspond to stronger differences between compared conditions (Δ*xyr1*-strain compared to wild-type strain), *i.e.* increasing colour intensity means more than 1.1-, 1.3-, and 1.5-fold difference.

Tables

Table 1 - Differentially expressed genes that are potentially involved in chromatin remodelling

Transcript ID	Annotation	SO/G	Δ*xyr1*/WT
2648	Predicted component of NuA3 histone acetyltransferase complex	-1.154	0.279
34402	Histone H1	1.300	-0.457
36727	SWI-SNF chromatin-remodeling complex protein	1.493	-0.843
53947	SWI-SNF chromatin-remodeling complex protein	1.196	1.332
56077	SWI-SNF chromatin-remodeling complex protein	3.050	-0.359
65533	Histone deacetylase complex, catalytic component HDA1	1.237	-0.347
73708	Heterochromatin-associated protein HP1 and related CHROMO domain proteins	-3.012	1.253
76872	SWI-SNF chromatin-remodeling complex protein	1.070	-0.708
81517	Sirtuin 5 and related class III sirtuins (SIR2 family)	1.600	-0.160
108909	Nucleosome-binding factor SPN, POB3 subunit	1.050	0.041
110409	Possible homologue of S. cerevisiae SAS10	-1.298	-0.001
110418	SWI-SNF chromatin-remodeling complex protein	1.180	-0.566

110507	Histone acetyltransferase (MYST family)	1.064	-0.562
122943	SWI-SNF chromatin-remodeling complex protein	1.876	-0.127
123327	SWI-SNF chromatin-remodeling complex protein	1.852	0.508

Differential gene expression according to RNAseq analysis comparing either sophorose induction (SO) to glucose repression (G) in the wild-type strain or the *xyr1* deletion strain (Δ*xyr1*) to the wild-type strain (WT) under sophorose induction.

Table 2 - Oligonucleotides used in this study

Name	Sequence (5' - 3')	Usage
RG53	GAATTCAGATC	ivFP, oligo-short
RG54	GCGGTGACCCGGGAGATCTGAATTC	ivFP, oligo-long
RG83	[6-FAM]CCTTTGGGTGTACATGTTTGTGCTCCGG	ivFP, cbh1oligo3fw
RG84	[6-FAM]GGAGAGTGCAGGCCGACTGAGC	ivFP, cbh1oligo3rev
RG89	[6-FAM]GTAGAGGCATGTTGTGAATCTGTGTCGGG	ivFP, cbh1oligo3fw
RG90	[6-FAM]GGTTGTATGCAAAACGCTCCGAGTCAGAC	ivFP, cbh1oligo3rev
actfw	TGAGAGCGGTGGTATCCACG	qPCR
actrev	GGTACCACCAGACATGACAATGTTG	
sar1fw	TGGATCGTCAACTGGTTCTACGA	
sar1rev	GCATGTGTAGCAACGTGGTCTTT	
cbh1f	GATGATGACTACGCCAACATGCTG	
cbh1r	ACGGCACCGGGTGTGG	
cbh2f	CTATGCCGGACAGTTTGTGGTG	
cbh2r	GTCAGGCTCAATAACCAGGAGG	

xyn1f	CAGCTATTCGCCTTCCAACAC	
xyn1r	CAAAGTTGATGGGAGCAGAAG	
taqxyn2f	GGTCCAACTCGGGCAACTTT	
taqxyn2r	CCGAGAAGTTGATGACCTTGTTC	
epiactinTr_f	CTTCCCTCCTTTCCTCCCCCTCCAC	act CHART, region -226 to +24
epiactinTr_r	GCGACAGGTGCACGTACCCTCCATT	
episar1Tr_f	GTCAGGAAATGCCGCACAAGCAAGA	sar1 CHART, region -490 to -224
episar1Tr_r	TGTGTTTTACCGCCTTGGCCTTTGG	
epicbh1_1Tr_f	AAGGGAAACCACCGATAGCAGTGTC	cbh1 CHART, region -902 to -610
epicbh1_1Tr_r	TTTCACTTCACCGGAACAAACAAGC	
epicbh1_2Tr_f	GGATCGAACACACTGCTGCCTTTAC	cbh1 CHART, region -301 to -27
epicbh1_2Tr_r	GGTTTCTGTGCCTCAAAAGATGGTG	
epicbh2_1Tr_f	CGGATCTAGGGCAGACTGGGCATTG	cbh2 CHART, region -587 to -338
epicbh2_1Tr_r	GTGTAGTGTTGCGCTGCACCCTGAG	
epicbh2_2Tr_f	TGCAGCGCAACACTACACGCAACAT	cbh2 CHART, region -355 to -62
epicbh2_2Tr_r	TGCGCCTCATACAGGGTCACAGTCC	

epixyn1_1Tr_f	GCACTCCAAGGCCTTCTCCTGTACT	xyn1 CHART, region -577 to -278
epixyn1_1Tr_r	TAGATTGAACGCCACCCGCAATATC	
epixyn1_3Tr_f	GTCGATATTGCGGGTGGCGTTCAAT	xyn1 CHART, region -306 to -10
epixyn1_3Tr_r	TTTGTGCGTGTTTTCCTTGAAGTCG	
epixyn2_1Tr_f	GTGCCGATGAGACGCTGCTGAGAAA	xyn2 CHART, region -527 to -252
epixyn2_1Tr_r	GATATTGCGCCTTGCAACACCATCG	
epixyn2_2Tr_f	CTCGAGACGGCTGAGACAGCAGCAT	xyn2 CHART, region -311 to -38
epixyn2_2Tr_r	TGTCTTTTGGGCTTGGAGGGGTTGT	