



UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

RAFAELA ROSSI ROSOLEN

**REDES DE CO-EXPRESSÃO GÊNICA PARA IDENTIFICAR
MÓDULOS FUNCIONAIS DE XYR1 E CRE1 EM *TRICHODERMA* SPP.**

**“GENE CO-EXPRESSION NETWORKS TO IDENTIFY FUNCTIONAL
MODULES OF XYR1 AND CRE1 IN *TRICHODERMA* SPP.”**

**CAMPINAS
2020**

RAFAELA ROSSI ROSOLEN

**REDES DE CO-EXPRESSÃO GÊNICA PARA IDENTIFICAR
MÓDULOS FUNCIONAIS DE XYR1 E CRE1 EM *TRICHODERMA* SPP.**

**“GENE CO-EXPRESSION NETWORKS TO IDENTIFY XYR1 AND
CRE1 FUNCTIONAL MODULES IN *TRICHODERMA* SPP.”**

Dissertação apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestra em Genética e Biologia Molecular, na área de Genética de Microorganismos.

Dissertation presented to the Institute of Biology of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Genetics and Molecular Biology, in the area of Genetics of Microorganisms.

Orientador: Profa. Dra. Anete Pereira de Souza

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELA ALUNA RAFAELA ROSSI ROSOLEN E ORIENTADA PELA PROFESSORA DOUTORA ANETE PEREIRA DE SOUZA.

**CAMPINAS
2020**

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

R733r Rosolen, Rafaela Rossi, 1993-
Redes de co-expressão gênica para identificar módulos funcionais de XYR1 e CRE1 em *Trichoderma* spp. / Rafaela Rossi Rosolen. – Campinas, SP : [s.n.], 2020.

Orientador: Anete Pereira de Souza.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Redes reguladoras de genes. 2. Fatores de transcrição. 3. *Trichoderma*.
I. Souza, Anete Pereira de, 1962-. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Gene co-expression networks to identify functional modules of XYR1 and CRE1 in *Trichoderma* spp.

Palavras-chave em inglês:

Gene regulatory networks

Transcription factors

Trichoderma

Área de concentração: Genética de Microorganismos

Titulação: Mestra em Genética e Biologia Molecular

Banca examinadora:

Anete Pereira de Souza [Orientador]

Juliana Velasco de Castro Oliveira

Adriano Rodrigues Azzoni

Data de defesa: 27-02-2020

Programa de Pós-Graduação: Genética e Biologia Molecular

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0002-8964-4034>

- Currículo Lattes do autor: <http://lattes.cnpq.br/6114146809588548>

Campinas, 27 de fevereiro de 2020.

COMISSÃO EXAMINADORA

Profa. Dra. Anete Pereira de Souza

Prof. Dra. Juliana Velasco de Castro Oliveira

Prof. Dr. Adriano Rodrigues Azzoni

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa de Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia.

*Dedico aos meus pais Rosemary e Francisco,
a minha avó Anna e ao meu namorado Matheus,
por todo apoio, incentivo, compreensão
e amor ao longo da minha formação.*

AGRADECIMENTOS

Primeiramente à Deus, pela vida e pela oportunidade de estar concluindo mais essa etapa. À Nossa Senhora Aparecida pelas bênçãos e graças alcançadas através de vossa intercessão. Ao Santo Anjo da Guarda, por sempre me conduzir pelos caminhos de luz.

Aos meus pais e à minha avó, por todo carinho, compreensão e amor ao longo da minha formação. Obrigada pelo apoio emocional e financeiro, os quais me permitiram seguir em frente mesmo nos momentos mais difíceis. Graças a vocês, mais esta conquista está sendo possível. Ao meu namorado, por toda paciência, amor e companheirismo, os quais foram essenciais no findar deste ciclo. Gratidão por estar ao meu lado.

À minha orientadora Profa. Dra. Anete Pereira de Souza por gentilmente ter me acolhido em seu grupo de pesquisa, por todos os ensinamentos, por compartilhar conhecimentos e por me orientar durante o processo de pesquisa dessa dissertação, me instigando a melhorar e a superar os meus limites. Obrigada pela contribuição teórico-científica para minha formação profissional. Por fim, obrigada pela sua paciência e dedicação tal qual pela confiança depositada em mim.

Ao Dr. Juliano Sales Mendes por todo ensino, paciência, empatia e auxílio no início dos experimentos. Ao Dr. Clelton Aparecido dos Santos por todo suporte e orientação na conduta dos experimentos. Ao Dr. Jaire Alves Ferreira Filho por toda ajuda empenhada na realização do meu projeto de pesquisa. Ao doutorando Alexandre Hild Aono por gentilmente contribuir com as análises de bioinformática, além de toda motivação, conselho e amizade. Obrigada pela dedicação e colaboração ativa na escrita do manuscrito. À Ma. Déborah Aires Almeida pela amizade, apoio, discussões e, mesmo à distância, auxílio ao final deste ciclo. Obrigada pela contribuição significativa durante todo o meu projeto de mestrado.

Agradeço aos Prof. Dr. Marcelo Falsarella Carazzolle, Prof. Dr. Adriano Rodrigues Azzoni e Dr. André da Silva Santiago pelo aceite ao convite em participar da minha banca de qualificação, por todas as correções e discussões relacionadas ao meu projeto.

Agradeço à Dra. Rosana Goldbeck por ter se disponibilizado a me ajudar no delineamento experimental de parte do meu projeto e por ser solícita e permitir a realização de

alguns experimentos em seu laboratório. Obrigada ao Laboratório de Engenharia Metabólica e de Bioprocessos (LEMeB) pelo suporte técnico e material.

A todos os colegas do Laboratório de Análise Genética e Molecular (LAGM) por contribuírem de algum modo com minha formação, especialmente ao Juliano, Jaire, Déborah, Alexandre, Maria Lorenza, Clelton, Cleiton, Beatriz, Michele, Carla e Danilo.

Aos técnicos do laboratório Danilo, Aline e Patrícia por toda ajuda.

À Deborah, Alexandre e Michele pela amizade, incentivo, apoio, momentos de descontração e palavras de conforto. Obrigada por tornarem os dias mais leves e alegres.

Ao Lucas e Vivian, pelos anos de amizade. À Bruna pela amizade, companheirismo e ajuda durante as disciplinas da pós-graduação. Gratidão por ter lhe conhecido!

Ao Programa de Pós-Graduação em Genética e Biologia Molecular e à Universidade Estadual de Campinas – UNICAMP.

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

À CAPES (Programa Biologia Computacional – 88887.176241/2018-00 e Programa de Excelência Acadêmica – PROEX – 88882.329483/2019-01) pela bolsa de mestrado concedida. À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (2015/09202-0) pelo auxílio na pesquisa.

Enfim, agradeço aos que tornaram de alguma forma possível a realização do meu projeto de mestrado. Sou grata a todos!

“The LORD is my shepherd, I lack nothing.”

Psalm 23:1

RESUMO

Os biocombustíveis são hoje uma das principais alternativas à substituição dos combustíveis fósseis. Considerando que o Brasil é o maior produtor de cana-de-açúcar, uma das alternativas propostas é a produção de etanol a partir do bagaço da cana, denominado etanol de segunda geração (2G). Esta tecnologia, entretanto, ainda apresenta algumas limitações econômicas, como o custo elevado dos coquetéis enzimáticos necessários à hidrólise da biomassa lignocelulósica. Neste contexto, espécies do gênero ascomiceto filamentosso *Trichoderma* desempenham um papel importante por produzirem enzimas que degradam a parede celular vegetal eficientemente, sendo dessa forma, alvos atrativos na indústria biotecnológica. Tendo como condição ancestral o micoparasitismo, espécies desse gênero, tais como *Trichoderma atroviride* e *Trichoderma harzianum*, são amplamente utilizadas na agricultura como agentes de biocontrole. Contudo, estudos demonstraram o potencial de *T. harzianum* em prospectar enzimas hidrolíticas associadas à degradação do substrato lignocelulósico. Dessa forma, uma melhor compreensão dos mecanismos genéticos relacionados ao principal ativador (XYR1) e repressor (CRE1) de celulasas e hemicelulasas se faz necessária em *T. harzianum* e, para fins comparativos, na espécie basal *T. atroviride*. Diante deste cenário, este trabalho teve como objetivo investigar o perfil de expressão gênica relacionado aos fatores de transcrição (FTs) XYR1 e CRE1 em condição de degradação da celulose em *T. harzianum* (IOC-3844 e CBMAI-0179) e em *T. atroviride* CBMAI-0020. Para isso, redes de co-expressão gênica foram modeladas para cada linhagem. Inicialmente, a análise filogenética indicou que XYR1 e CRE1 estão amplamente distribuídos dentre os fungos ascomicetos e sugeriu um processo de diferenciação para *T. atroviride* em relação à *T. harzianum*. A inferência das redes separou os transcritos em diversos grupos de acordo com o perfil de expressão durante o crescimento em celulose ou em glicose. Portanto, os grupos com os transcritos *xyl1* e *cre1* foram identificados. Nestes grupos, transcritos codificando enzimas ativas em carboidratos (CAZymes), FTs e transportadores de açúcar e íons, bem como proteínas com função desconhecida, foram co-expressos, porém diferenciando qualitativa e quantitativamente entre as linhagens. Transcritos centrais (*hubs*) foram identificados e incluíram transcritos ainda não caracterizados ou ainda não descritos como relacionados à degradação da celulose. Além disso, várias vias metabólicas desencadeadas nos grupos *xyl1* e *cre1* foram identificadas com enriquecimento correspondente ao micoparasitismo em *T. atroviride*, especialmente associado à CRE1. Nossos resultados sugerem diferentes estratégias utilizadas por *Trichoderma* spp. relacionadas à XYR1 e CRE1 durante a degradação da celulose. Também foi observado que os transcritos codificando CAZymes referentes à

degradação da celulose e da hemicelulose em *T. harzianum* não são necessariamente co-expressos com os FTs aqui estudados. Tomados em conjunto, nossas descobertas fornecem novos *insights* sobre os transcritos co-expressos com XYR1 e CRE1 em linhagens de *T. harzianum*, os quais podem ser alvo de novos estudos para avaliar seu papel na degradação da lignocelulose e ser explorados para a melhoria de coquetéis enzimáticos. Finalmente, nossos resultados podem contribuir para o melhor entendimento dos mecanismos genéticos envolvidos na regulação de enzimas hidrolíticas, expandindo o potencial do uso de *T. harzianum* em diversas aplicações industriais.

Palavras-chave: Redes de co-expressão gênica; Fatores de transcrição; *Trichoderma* spp.

ABSTRACT

Biofuels are today one of the main alternatives to the replacement of fossil fuels. Brazil is the leading producer of sugarcane; thus, a proposed alternative is the production of ethanol from sugarcane bagasse, known as second-generation ethanol (2G). However, this technology still has some economic limitations, such as the high cost of enzymatic cocktails necessary for the hydrolysis of lignocellulosic biomass. In this context, species in the filamentous ascomycete genus *Trichoderma* play an important role in producing plant cell wall-degrading enzymes being an attractive target for the biotechnology industry. Having as an ancestral trait the mycoparasitism, species of *Trichoderma*, such as *Trichoderma atroviride* and *Trichoderma harzianum*, are widely used in agriculture as biocontrol agents. Although, studies have demonstrated the potential of *T. harzianum* in producing a set of enzymes acting efficiently in the saccharification of the lignocellulosic substrate. Thus, a better understanding of genetic mechanisms related to the main activator (XYR1) and repressor (CRE1) of cellulases and hemicellulases is necessary for *T. harzianum* and, for comparative purposes, in the basal species *T. atroviride*. In this scenario, this study aimed to investigate the gene expression profile associated to transcription factors (TFs) XYR1 and CRE1 during cellulose degradation for *T. harzianum* (IOC-3844 and CBMAI-0179) and for *T. atroviride* CBMAI-0020. For this, gene co-expression networks were modeled for each strain. Initially, phylogenetic analysis indicated that both regulatory proteins are widely distributed among ascomycete fungi and suggested a differentiation process for *T. atroviride* in relation to *T. harzianum*. The co-expression network analyses separated the transcripts into several groups according to the expression profile during growth in cellulose or glucose. Groups with *xyl1* and *cre1* were identified and within them transcripts encoding carbohydrate-active enzymes (CAZymes), TFs and sugar and ion transporters, as well as proteins with unknown function, were co-expressed. Although, qualitatively and quantitatively differences among groups and strains were observed. Core transcripts from these groups (hubs) were identified, and they included transcripts not yet characterized or described as related to cellulose degradation. In addition, several metabolic pathways triggered in *xyl1* and *cre1* groups have been identified with an enrichment corresponding to mycoparasitism in *T. atroviride*, especially associated with CRE1. Our results suggest different strategies used by *Trichoderma* spp. related to XYR1 and CRE1 during cellulose degradation. In addition, it was observed that the transcripts encoding CAZymes associated with cellulose and hemicellulose degradation in *T. harzianum* are not necessarily co-expressed with the TFs herein studied. Taken together, our findings provide new insights into transcripts co-expressed with XYR1 and CRE1 in *T. harzianum* strains, which could be the

target of new studies to evaluate their role in lignocellulose degradation and that can be exploited for the improvement of enzymatic cocktails. Finally, our results can contribute to a better understanding of the genetic mechanisms involved in hydrolytic enzyme regulation, expanding the potential of *T. harzianum* in several industrial applications.

Keywords: Gene co-expression networks; Transcription factors; *Trichoderma* spp.

LISTA DE ILUSTRAÇÕES

Revisão Bibliográfica

Figura 1: Modelo simplificado da arquitetura da parede celular vegetal e os principais componentes da lignocelulose.....	19
Figura 2: Principais etapas para a produção de etanol 2G a partir de biomassa lignocelulósica da cana	21
Figura 3: Degradação da celulose e da hemicelulose por enzimas hidrolíticas em <i>T. reesei</i> .	22
Figura 4: Perfil da atividade enzimática de celulasas (FPase) das diferentes linhagens de <i>Trichoderma</i> em condição de degradação da biomassa vegetal.....	26
Figura 5: Visão geral da regulação de enzimas hidrolíticas em <i>T. reesei</i>	28
Figura 6: Exemplo de uma análise de rede de co-expressão	31
Figura 7: Teste de expressão de ThX	77
Figura 8: Solubilização de ThX	77

Artigo

Figura 1: Filogenia molecular dos FTs CRE1 e XYR1 em Ascomycota	42
Figura 2: Subgrupos identificados nos grupos <i>cre1</i> e <i>xyr1</i> em <i>Trichoderma</i> spp.....	44
Figura 3: Classificação da função KO dos transcritos identificados nos grupos <i>cre1</i> e <i>xyr1</i> ...	47
Figura 4: Comparação de FTs, transportadores e CAZymes entre linhagens e grupos	49
Figura 5: Distribuição das famílias CAZymes nos grupos <i>cre1</i> e <i>xyr1</i> em <i>Trichoderma</i> spp.	50
Figura 6: Distribuição de FTs e transportadores nos grupos <i>cre1</i> e <i>xyr1</i> em <i>Trichoderma</i> spp.	52

Material Suplementar

Figura S1: Diagramas de Venn das vias metabólicas identificadas nos subgrupos <i>cre1</i> e <i>xyr1</i> compartilhadas por <i>Trichoderma</i> spp.	71
Tabela S7: Número de transcritos co-expressos em cada subgrupo formado a partir dos grupos <i>cre1</i> e <i>xyr1</i> para <i>Trichoderma</i> spp. e sua cor correspondente	72
Tabela S10: Valores de expressão em TPM de <i>cre1</i> e <i>xyr1</i> para <i>Trichoderma</i> spp. sob condições de crescimento em celulose e glicose	74

SUMÁRIO

1.	INTRODUÇÃO GERAL.....	16
2.	REVISÃO BIBLIOGRÁFICA.....	18
2.1.	Etanol de segunda geração.....	18
2.2.	Hidrólise enzimática	22
2.3.	Fungos filamentosos e prospecção de enzimas hidrolíticas.....	24
2.4.	<i>Trichoderma</i> como produtor de enzimas hidrolíticas	24
2.5.	Regulação gênica associada à desconstrução da biomassa vegetal	27
2.6.	Proteínas transportadoras e degradação do material lignocelulósico	29
2.7.	Análise de transcriptoma por RNA-Seq	30
2.8.	Redes de co-expressão gênica como ferramenta biotecnológica.....	30
3.	OBJETIVOS.....	33
3.1.	Objetivo Geral.....	33
3.2.	Objetivos Específicos	33
4.	ARTIGO.....	34
	Comparative gene coexpression networks analysis reveals different strategies of <i>Trichoderma</i> spp. associated with the transcriptional factors XYR1 and CRE1 during cellulose degradation.	34
5.	RESULTADOS COMPLEMENTARES	74
5.1.	Contextualização	74
5.2.	Materiais e métodos	75
5.2.1.	Mapeamento e reconstrução do transcrito <i>gh11</i>	75
5.2.2.	Amplificação e clonagem do transcrito <i>gh11</i>	75
5.2.3.	Expressão em <i>E. coli</i> de ThX.....	75
5.2.4.	Teste de expressão de ThX	76
5.2.5.	Solubilização de ThX pelo método de <i>refolding</i>	76
5.3.	Resultados parciais.....	76
5.4.	Perspectivas	78
6.	CONCLUSÕES GERAIS.....	79
7.	PERSPECTIVAS GERAIS	80

8. REFERÊNCIAS BIBLIOGRÁFICAS.....	81
ANEXO I – DECLARAÇÃO CIBio	94
ANEXO II – DECLARAÇÃO DE AUTORIA.....	95

1. INTRODUÇÃO GERAL

Trichoderma é um gênero de fungo ascomiceto filamentoso capaz de colonizar diversos substratos em diferentes condições ambientais [1]. Em virtude da versatilidade nutricional ampla, os microrganismos que compreendem tal gênero são extensivamente explorados em diferentes contextos biotecnológicos. Tendo como condição ancestral o micoparasitismo, espécies desse gênero, tais como *Trichoderma atroviride* e *Trichoderma virens*, por exemplo, são amplamente utilizadas na agricultura como agentes de biocontrole micoparasitário [2].

Ambos são filogeneticamente distantes de *Trichoderma reesei* [3], o qual possui um arsenal menor de genes relacionados ao micoparasitismo e ao metabolismo secundário [2, 4]. De forma antagônica, este fungo tem um estilo de vida saprotrófico e é um produtor de enzimas hidrolíticas vastamente utilizadas em coquetéis enzimáticos aplicados biotecnologicamente na desconstrução do material lignocelulósico [5-7], dentre outras aplicações [8].

As espécies citadas são filogeneticamente bem definidas [9, 10], em contraste com *Trichoderma harzianum sensu lato*, o qual possui um processo de especiação complexo baseado na coexistência e interação de organismos com diferentes histórias evolutivas e na ausência de fronteiras genéticas estritas entre eles [2, 11]. Embora algumas linhagens são exploradas comercialmente para o controle biológico de fungos patogênicos de plantas [12-14], estudos recentes demonstraram seu papel como potencial produtor de enzimas hidrolíticas. Em termos de aplicação industrial, tais enzimas poderiam ser empregadas ao longo do processo de produção de etanol de segunda geração (etanol 2G) [15-20].

As enzimas utilizadas por esses microrganismos para desconstruir a lignocelulose são definidas na literatura internacional como “*Carbohydrate-Active Enzymes – CAZymes*” [7] e podem ser encontradas no banco de dados CAZy (<http://www.cazy.org/>), sendo estas classificadas em: glicosil hidrolases (GHs), glicosil transferases (GTs), polissacarídeo liases (PLs), carboidrato esterases (CEs), proteínas com atividade auxiliar (AAs) e proteínas com módulos de ligação à carboidrato (CBMs) [21]. Destas, as GHs são responsáveis pela hidrólise e/ou transglicosilação de ligações glicosídicas, e representam quase metade das enzimas classificadas no banco [22].

Em *T. reesei*, a expressão destas enzimas lignocelulolíticas é induzida por diferentes moléculas, tais como sofrorose, celobiose, xilose, xilobiore, xilo-oligossacarídeos, arabitól, galactose e lactose [23, 24] e regulada a nível transcricional por fatores de transcrição (FTs) específicos. Dentre estes, destacam-se os FTs XYR1 e CRE1, que se ligam a região promotora dos genes que codificam celulasas, hemicelulasas, além de outras CAZymes, ativando ou

reprimindo, respectivamente, a expressão dos mesmos, bem como das vias metabólicas necessárias para metabolizar os monômeros resultantes [24-28].

Uma das etapas mais onerosas no processo de conversão de biomassa a etanol 2G é a hidrólise enzimática [29]; assim sendo, a fim de tornar tal processo vantajoso do ponto de vista econômico, a combinação de metagenômica, transcriptômica e proteômica seriam algumas das abordagens utilizadas para a prospecção de novas enzimas a partir de microrganismos, tais como os fungos [30]. Uma abordagem interessante para explorar a transcriptômica seria a construção de redes de co-expressão [31], uma vez que são importantes ferramentas para entender as respostas biológicas desencadeadas frente a determinada situação, sugerindo quais transcritos estariam ativos simultaneamente e, portanto, desempenhando uma função comum [32, 33].

Recentemente, redes de co-expressão foram utilizadas pelo nosso grupo a fim de explorar as proteínas compartilhadas durante o processo de degradação da biomassa lignocelulósica entre *T. harzianum* IOC-3844, *T. atroviride* CBMAI-0020 e *T. reesei* CBMAI-0711 cultivados nas condições de celulose ou glicose [34]. Na mesma direção, Borin e colaboradores (2018) identificaram em *T. reesei* RUT-C30 várias celulasas, transportadores de açúcares, FTs e proteínas hipotéticas agrupados em módulos genéticos altamente conectados. Dentro destes grupos, *hubs* foram identificados e vários destes continham pelo menos um local de ligação à XYR1 na região promotora [35].

Porém, um dos grandes desafios para entender o mecanismo molecular da degradação da biomassa é como os FTs relacionados a esse sistema agem. Considerando que, em relação ao gênero *Trichoderma*, as pesquisas sobre a regulação gênica relacionada às enzimas hidrolíticas têm se concentrado em *T. reesei* [36-38], a fim de expandir o potencial de bioprospecção de enzimas lignocelulolíticas a partir de fungos filamentosos, uma melhor compreensão deste processo em outras espécies de *Trichoderma* faz-se necessária.

Neste cenário, este é o primeiro trabalho que investiga de forma comparativa os mecanismos genéticos desencadeados pelos principais reguladores do processo de degradação da lignocelulose em *Trichoderma* spp. através da construção de redes de co-expressão gênica. Por meio desta metodologia, foi possível explorar os genes e as vias metabólicas subjacentes à regulação transcricional de XYR1 e CRE1 quanto à degradação da celulose considerando um ancestral comum micoparasitário (*T. atroviride* CBMAI-0020) e linhagens com potencial hidrolítico da espécie micoparasita *T. harzianum* (IOC-3844 e CBMAI-0179) [34, 39] advindas de processos de diferenciação e adaptação.

2. REVISÃO BIBLIOGRÁFICA

2.1. Etanol de segunda geração

Estima-se que até 2050 a população mundial será de nove bilhões de pessoas [40]. Diante deste cenário, espera-se um aumento na demanda por matérias-primas e, conseqüentemente, nas emissões de gases do efeito estufa e no uso não racional dos recursos naturais finitos. A possibilidade de esgotamento dessas reservas aliada à crescente demanda mundial por energia e o alto grau de dependência econômica dos combustíveis fósseis, além das mudanças climáticas a estes associadas, emergiu o interesse em fontes energéticas alternativas, dentre elas, os biocombustíveis [41-43].

Diante deste cenário, foi implementado no Brasil o programa Proálcool, o qual surgiu frente à necessidade de desenvolver outro combustível líquido que pudesse substituir a gasolina, cujo preço estava supervalorizado na década de 1970 [41]. Essa nova política energética consolidou o setor sucroenergético brasileiro, rendendo hoje o segundo lugar no *ranking* dos maiores produtores de etanol com a produção de 28% deste biocombustível no mundo, atrás apenas dos EUA [44].

De forma breve, para produzir o etanol brasileiro, a cana é colhida, lavada, moída e o caldo extraído é fermentado por leveduras industriais. Em seguida, o etanol gerado é destilado e convertido em etanol anidro ou hidratado, que pode ser comercializado, principalmente, como combustível ou aditivo para carros [45]. Uma vez que este processo de fabricação envolve a fermentação de açúcares presentes no caldo de cana, o aumento na produtividade deste biocombustível significaria o aumento da área de cultivo de cana-de-açúcar e, a longo prazo, poderia conflitar com a produção de *commodities* alimentares [41, 46, 47].

É válido ressaltar que o Brasil é o maior produtor mundial de cana [44, 45], cujas safras correspondem a mais de 620 milhões de toneladas por ano, com destaque para o estado de São Paulo, o qual produziu 333 milhões de toneladas na safra 2018/2019, o que corresponde a 53,7% do total produzido no país [48, 49]. Dessa forma, levando-se em consideração que, em média, são gerados 280 kg de bagaço através do processamento industrial de 1 tonelada de cana, uma das alternativas propostas é a produção do etanol utilizando como matéria-prima esse subproduto, denominado de etanol de segunda geração (etanol 2G) [50].

A biomassa proveniente para produzi-lo tem como elemento estrutural a lignocelulose, a qual é composta por 30-50% de celulose, 20-30% de hemicelulose e 15-25% de lignina (**Figura 1**) [51, 52]. Entretanto, estes três componentes não são distribuídos uniformemente dentro das paredes celulares, podendo variar de acordo com as espécies, tecidos e maturidade das células vegetais [51].

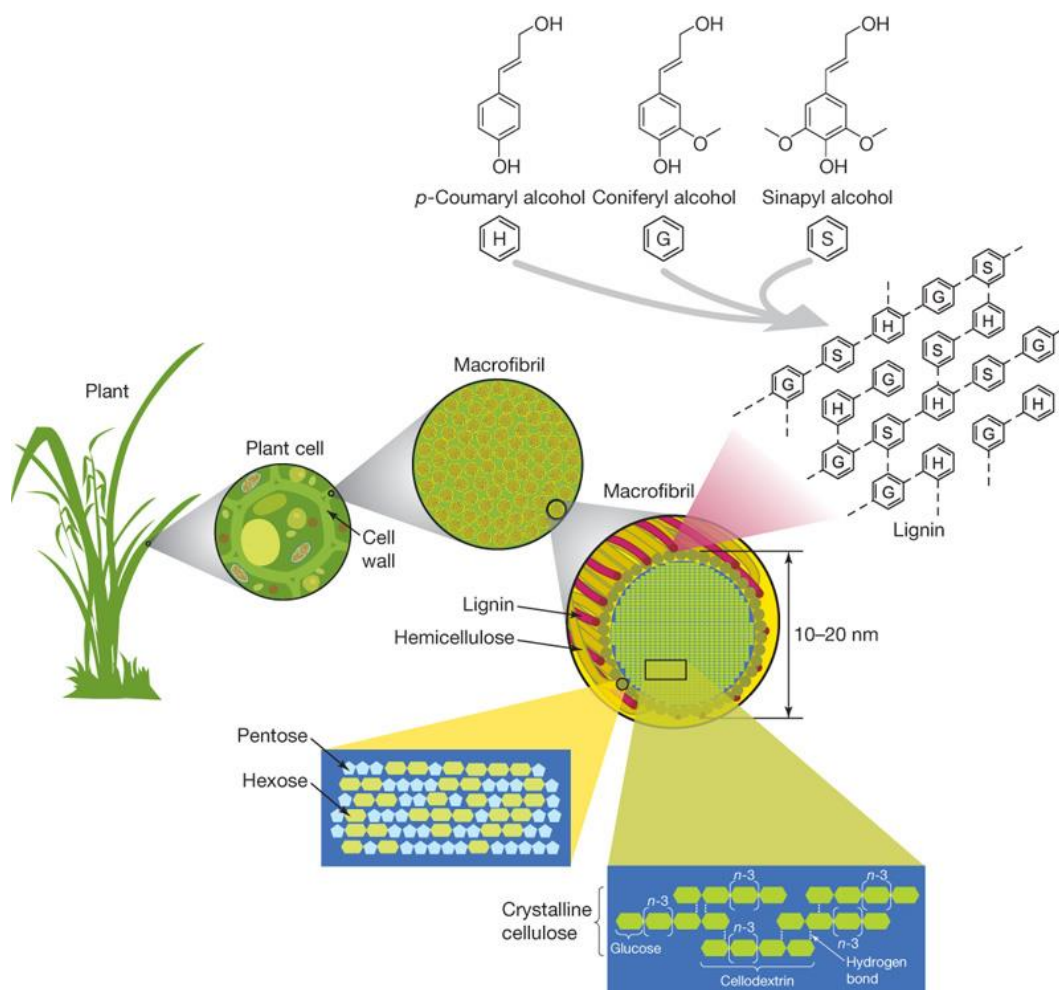


Figura 1. Modelo simplificado da arquitetura da parede celular vegetal e os principais componentes da lignocelulose. A celulose, principal componente da parede celular vegetal; a hemicelulose, segundo maior componente; e a lignina formam estruturas chamadas microfibrilas, as quais são organizadas em macrofibrilas e medeiam a estabilidade estrutural na parede celular da planta [52].

A celulose, principal componente da parede celular vegetal, é um polímero linear composto por unidades de D-glicose ligadas umas às outras por ligações glicosídicas do tipo β -1,4 (**Figura 1**) [53]. Sua porção cristalina, cuja estrutura compactada impede o acesso de enzimas, bem como de pequenas moléculas, como água e outros solventes, é altamente resistente à hidrólise [54, 55]. Por outro lado, sua região amorfa está exposta de tal forma que permite a formação de pontes de hidrogênio com a água, sendo hidrolisada mais rapidamente [56, 57]. Aliada a celulose, tem-se a hemicelulose (**Figura 1**), o segundo polímero mais abundante da parede celular vegetal. Trata-se de um heteropolímero altamente ramificado composto por pentoses (xilose e arabinose), hexoses (manose, glicose e galactose) e por

açúcares ácidos (ácido glucurônico, galacturônico e metilgalacturônico) [58, 59], cuja composição é variável e depende da biomassa em questão [60].

O terceiro componente da lignocelulose é a lignina, um heteropolímero aromático complexo altamente resistente à degradação química e biológica. Sua composição consiste em três álcoois aromáticos – p-coumaril, coniferil e sinapil – derivados dos aminoácidos tirosina e fenilalanina. Estes álcoois são polimerizados pela ação de lacases e peroxidases, dando origem às três unidades principais da lignina: p-hidroxifenil (H), guaiacil (G), e siringil (S) (**Figura 1**) [61, 62]. Ao contrário da degradação da celulose e da hemicelulose, as quais liberam açúcares fermentescíveis, a degradação da lignina produz compostos fenólicos com diversas aplicações de alto valor agregado [63].

Para a produção do etanol 2G a partir da biomassa lignocelulósica são necessárias quatro etapas principais: pré-tratamento, sacarificação, fermentação e destilação [64], como mostrado abaixo na **Figura 2**.

A etapa de pré-tratamento reduz a recalcitrância da parede celular vegetal e permite o acesso de enzimas hidrolíticas e oxidativas aos componentes da lignocelulose. Após o pré-tratamento adequado, a celulose e a hemicelulose estão mais acessíveis à hidrólise e passam pela etapa de sacarificação, onde são produzidos açúcares monoméricos (pentoses e hexoses). A etapa de hidrólise pode ser feita quimicamente pela adição de H_2SO_4 ou bioquimicamente pelo uso de celulases e hemicelulases produzidas por microrganismos [65].

Os açúcares liberados são então fermentados por leveduras industriais na etapa de fermentação, obtendo-se como produto final o etanol 2G [64, 66, 67]. Porém, uma vez que este pode estar contaminado com polissacarídeos não convertidos, cinzas e compostos resultantes da quebra da lignina, além de presente em frações de baixo volume, é necessária a etapa de destilação. Neste processo, o etanol é separado dos outros componentes, o que garante a purificação do etanol 2G e o aumento na concentração deste biocombustível [68].

Dentre os processos citados, a etapa de hidrólise enzimática é um dos principais gargalos para a produção do etanol 2G em virtude do alto custo dos coquetéis enzimáticos e a necessidade de uma grande quantidade de enzimas para conversão completa da biomassa em açúcares fermentescíveis, tornando o processo de fabricação economicamente inviável [29]. A fim de tornar a conversão da biomassa em etanol 2G vantajosa do ponto de vista econômico, a combinação de metagenômica, genômica, transcriptômica e proteômica seriam algumas das abordagens utilizadas para a prospecção de novas enzimas a partir de microrganismos, tais como os fungos (**Figura 2**) [30].

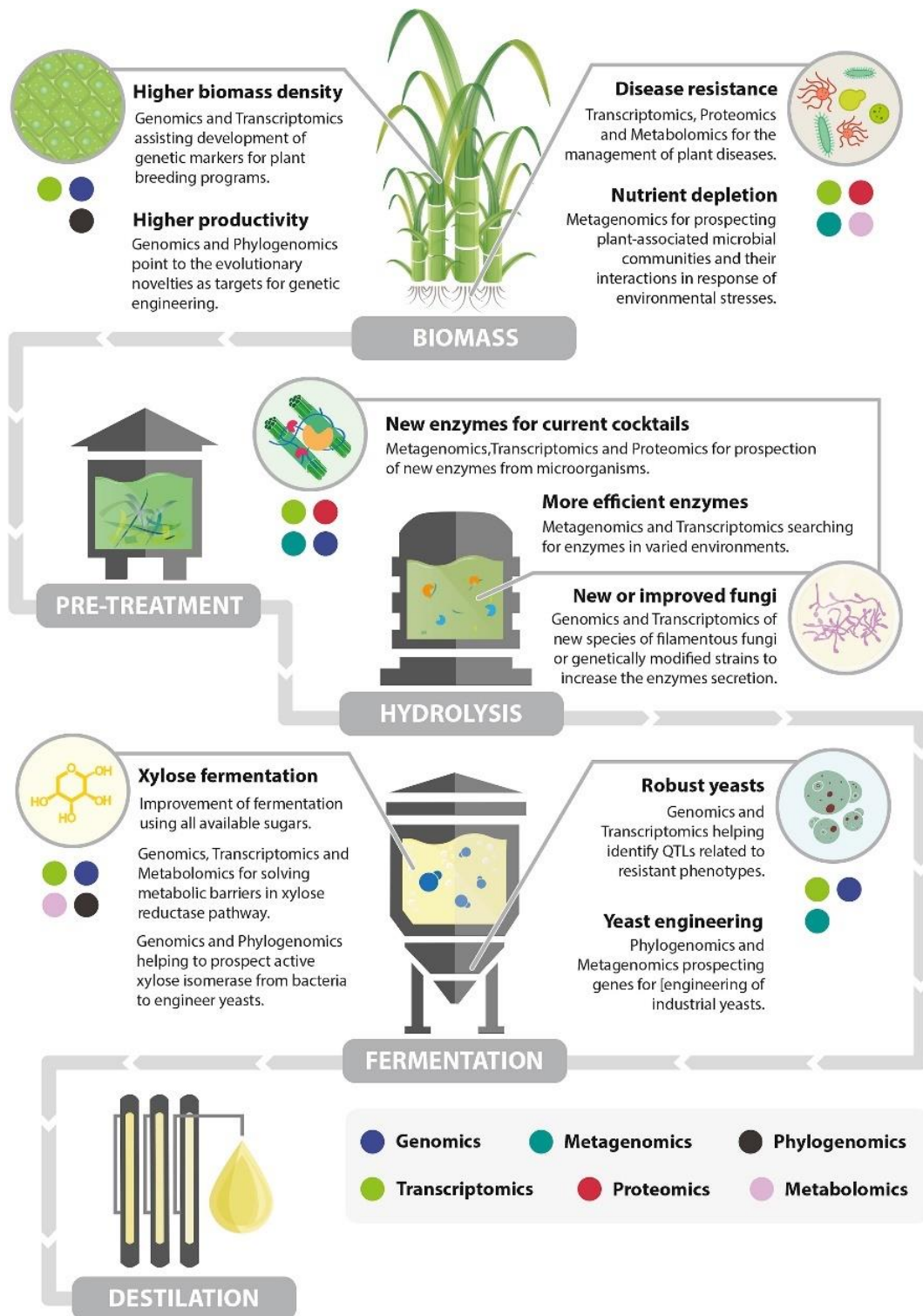


Figura 2. Principais etapas para a produção de etanol 2G a partir de biomassa lignocelulósica da cana. O bagaço de cana deve ser desconstruído com as etapas de pré-tratamento e hidrólise enzimática. Os açúcares liberados neste último processo são então fermentados por leveduras industriais e, após a etapa de destilação, o etanol 2G é obtido [30]. Ademais, a fim de diminuir os custos de produção e aumentar a produtividade, as análises

“ômica” vêm sendo amplamente exploradas ao longo das várias etapas de produção do etanol 2G.

2.2. Hidrólise enzimática

A hidrólise enzimática é um processo complexo afetado por muitos fatores. Dentre estes, podem ser citados os diferentes métodos de pré-tratamento aplicados, os quais podem resultar em diferenças na composição e na estrutura do material lignocelulósico; portanto, influenciando significativamente o rendimento de etanol 2G [69]. Além disso, o acesso das celulasas ao polímero da celulose mostrou estar relacionado ao conteúdo de lignina presente na parede celular vegetal, afetando consideravelmente a sacarificação enzimática [69].

Em virtude da estrutura recalcitrante (**Figura 3a**), a completa degradação da parede celular vegetal requer a ação coordenada de uma ampla gama de enzimas atuando sinergicamente (**Figuras 3b e 3c**). Devido à natureza altamente heterogênea da hemicelulose, a completa hidrólise desse polímero requer um grupo complexo de enzimas, tais como: endo-1,4- β -xilanases (EC 3.2.1.8), β -1,4-xilosidases (EC 3.2.1.37), β - mananases (EC 3.2.1.78), α -L-arabinofuranosidases (EC 3.2.1.55), acetil xilano esterases (EC 3.1.1.72), feruloil esterases (EC 3.1.1.73), α -glucuronidasases (EC 3.2.1.-), entre outras [64] (**Figura 3b**).

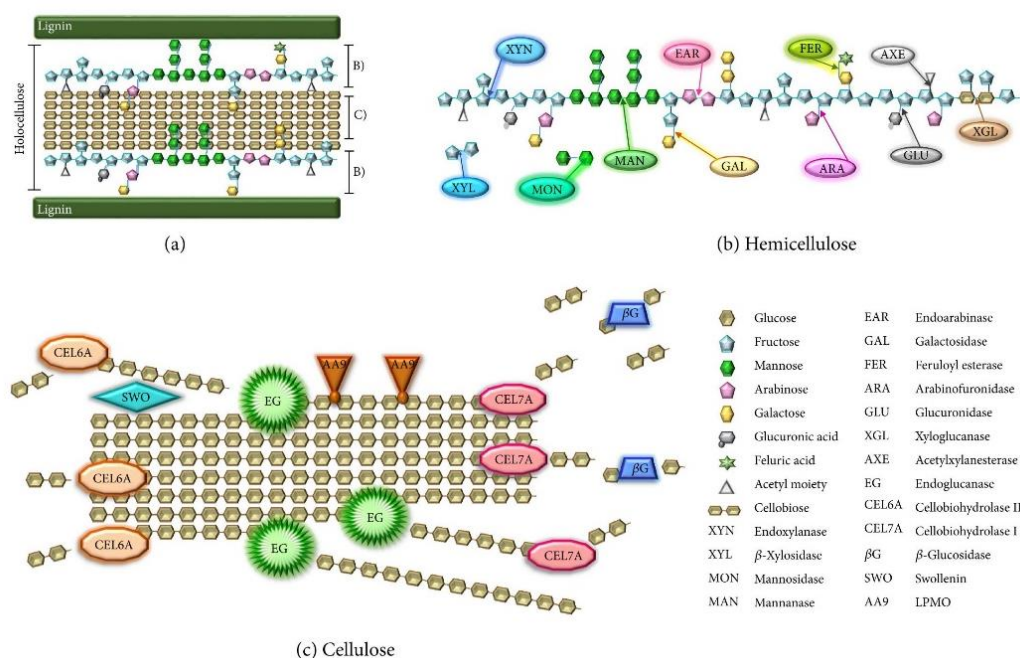


Figura 3. Degradação da celulose e da hemicelulose por enzimas hidrolíticas em *T. reesei*.

Em virtude da composição heterogênea, a hemicelulose é despolimerizada pela ação sinérgica de um conjunto variável de enzimas permitindo, dessa forma, o acesso das celulasas ao polímero da celulose. A degradação enzimática da celulose tem início com a clivagem das regiões amorfas da cadeia pela ação das endoglucanases (EG), enquanto as exoglucanases CEL6A e

CEL7A clivam as extremidades não redutora e redutora, respectivamente. Posteriormente, as β -glicosidases (BG) clivam as moléculas de celobiose em glicose, um açúcar facilmente metabolizável. As AA9 facilitam a ação das celulases por meio de um mecanismo oxidativo, enquanto as proteínas suoleninas (SWO) expandem a cadeia celulolítica, tornando-a mais exposta e suscetível à ação das celulases [37].

No que diz respeito à degradação da celulose, esta envolve a ação sinérgica de três classes de enzimas: endo- β -1,4-glucanases (EG) (EC 3.2.1.4), exoglucanases (CEL6A e CEL7A) (EC 3.2.1.91; EC 3.2.1.176) e β -glicosidases (BG) (EC 3.2.1.21) [37, 64] (**Figura 3c**).

As EGs iniciam a degradação da celulose clivando randomicamente as ligações glicosídicas localizadas na região interna da cadeia de celulose. As CEL6A (CBH2, GH6) e CEL7A (CBH1, GH7) atacam progressivamente as extremidades não redutora e redutora, respectivamente, da celulose, as quais foram geradas anteriormente pelas EGs (**Figura 3c**). Como produto, as CEL6A e CEL7A liberam oligossacarídeos, principalmente celobiose, um dissacarídeo de glicose que inibe a ação dessas enzimas. Por fim, as BGs hidrolisam a ligação glicosídica β -1,4 da celobiose, liberando unidades de glicose, o que, consequentemente, diminui o produto de inibição das CEL6A e CEL7A [59, 70] (**Figura 3c**).

Além do mais, enzimas típicas celulolíticas provenientes de fungos possuem um domínio catalítico (CD) e módulos de ligação à carboidratos (CBMs). Estes últimos têm um papel importante em aproximar os domínios catalíticos ao substrato para potencializar a atividade enzimática, facilitando, assim, a hidrólise da celulose [71].

Quanto a degradação da lignina, esta ocorre por reações oxidativas atribuídas principalmente ao metabolismo secundário. As principais enzimas envolvidas são: lacases (Lac) (EC 1.10.3.2), peroxidases dependentes de manganês (MnP) (EC 1.11.1.13), peroxidases de lignina (LiP) (EC 1.11.1.14), peroxidases versáteis (VP) (EC 1.11.1.16) e, mais recentemente, *dye-decolorizing peroxidases* (DyP) [70, 72].

Além das enzimas acima descritas, enzimas de funções acessórias e proteínas também são importantes na despolimerização da lignocelulose (**Figura 3c**). Dentre estas, as monooxigenases líticas de polissacarídeos (LPMOs), tais como AA9, facilitam a ação das celulases por meio de um mecanismo oxidativo, enquanto as proteínas suoleninas modificam a superfície da celulose, tornando-a mais exposta e suscetível à ação de celulases [37]. Recentemente, foi demonstrado que uma proteína dessa classe apresenta um efeito sinérgico com uma xilanase comercial durante a degradação de xilano [17].

2.3. Fungos filamentosos e prospecção de enzimas hidrolíticas

Os fungos estão entre os sistemas biológicos mais bem estudados para degradação de biomassa [73]. Como decompositores, têm como capacidade inata realizar a conversão eficiente do material lignocelulósico em açúcares facilmente metabolizáveis através da secreção de enzimas que atuam na despolimerização do substrato recalcitrante [74].

Os fungos filamentosos dos filos Ascomycota (*T. reesei*, *Aspergillus* spp., *Fusarium* spp., *Penicillium* spp., *Neurospora crassa*) e Basidiomycota (*Phanerochaete chrysosporium*, *Oligoporus (Postia) placenta*) são de fundamental importância biotecnológica e vêm sendo estudados quanto ao seu potencial celulolítico e na produção de químicos de alto valor agregado [75-77].

As enzimas utilizadas por esses microrganismos para desconstruir a lignocelulose são agrupadas em coquetéis enzimáticos e devem apresentar as seguintes características: alta estabilidade térmica, tolerância a resíduos do processo de conversão, atuação sinérgica e custo viável, propiciando, dessa forma, a aplicação das mesmas em escala industrial [64, 78, 79]. As principais empresas que produzem esses coquetéis com enzimas fúngicas incluem Novozymes (<http://www.novozymes.com>), DSM (<http://www.dsm.com>), Roal Oy (<http://www.roal.fi>), AB Enzymes (<http://www.abenzymes.com>) e DuPont (<http://www.dupont.com>) [28].

Entretanto, esforços vêm sendo realizados para reduzir os custos de produção e aumentar a eficiência dos mesmos [1], uma vez que os coquetéis que já são utilizados em escala industrial ainda não são capazes de degradar completamente a biomassa vegetal, havendo a necessidade de explorar novas enzimas e de entender melhor seus mecanismos de ação [80].

2.4. *Trichoderma* como produtor de enzimas hidrolíticas

Composto por fungos filamentosos amplamente distribuídos na natureza e que podem ser encontrados em ecossistemas que variam desde a tundra até o tropical, o gênero *Trichoderma* abrange microrganismos cujas características principais estão relacionadas à ampla capacidade metabólica e à agressão competitiva natural, despertando, portanto, grande interesse biotecnológico [1, 81].

Algumas espécies de *Trichoderma*, por exemplo, são utilizadas na biorremediação de ambientes contaminados em virtude da alta tolerância de tais microrganismos frente à uma variedade de poluentes recalcitrantes, incluindo metais pesados, pesticidas e hidrocarbonetos poliaromáticos [82-85]. Além das espécies altamente benéficas e frequentemente usadas, o gênero *Trichoderma* também inclui patógenos humanos oportunistas, os quais representam uma ameaça séria e muitas vezes letal em pessoas infectadas pelo HIV e em outros pacientes imunocomprometidos [86, 87]. Ademais, devido à versatilidade nutricional única, espécies

desse gênero são largamente exploradas na agricultura como agentes de controle biológico, bem como na produção de enzimas para indústrias têxtil, de alimentos e de biocombustíveis. [37, 88-90].

Dentre os fungos do gênero *Trichoderma*, *T. reesei* é o fungo celulolítico e hemicelulolítico mais bem estudado [5, 91-94]. Conhecido pela alta capacidade de produzir e secretar celulasas e hemicelulasas, *T. reesei* foi isolado originalmente nas Ilhas Salomão durante a Segunda Guerra Mundial [95]. A cepa isolada, nomeada como QM6a [96], produz pelo menos duas celobiohidrolases (CBH1/Cel7A e CBH2/Cel6A), cinco endoglucanases (EG1/Cel7B, EG2/Cel5A, Cel5B, EG3/Cel12A, EG45/Cel45A), duas β -glicosidases já caracterizadas (BG1/Cel3A e BG2/Cel1A) e mais cinco preditas (Cel3B, Cel3D, Cel1B, Cel3C, Cel3E), além de inúmeras hemicelulasas [97]. Posteriormente, a linhagem de *T. reesei* QM6a foi submetida a uma série de experimentos de mutagênese induzida dando origem a linhagem hipercelulolítica conhecida como *T. reesei* RUT-C30, uma das principais cepas industriais destinadas à produção de enzimas hidrolíticas [94].

Enquanto *T. reesei* é adaptado a um estilo de vida saprotrófico, *T. harzianum* e *T. atroviride*, por exemplo, são descritos como fungos micoparasitários e, portanto, utilizados como agentes de biocontrole contra fungos fitopatogênicos [98, 99]. Porém, ao longo dos últimos anos, as linhas de pesquisa com *T. harzianum* também têm focado no potencial biotecnológico desta espécie em prospectar enzimas hidrolíticas relacionadas à degradação da lignocelulose [15, 16, 19, 100].

Em um estudo prévio, diferentes espécies de *Trichoderma* foram submetidas a um cultivo cuja principal fonte de carbono era o bagaço de cana-de-açúcar. Para um tempo de fermentação igual a 96 horas, as linhagens de *T. harzianum* CBMAI-0179 e IOC-3844 apresentaram alta atividade enzimática de celulase (FPase) quando comparadas com outras linhagens da mesma espécie e com *T. reesei* CBMAI-0711 (**Figura 4**) [101], motivo pelo qual foram definidas como promissoras quanto a produção de enzimas hidrolíticas em condição de degradação da biomassa vegetal. Por outro lado, a linhagem *T. atroviride* CBMAI-0020 não apresentou atividade enzimática para a condição analisada (**Figura 4**), sendo, portanto, utilizada como parâmetro nos experimentos subsequentes.

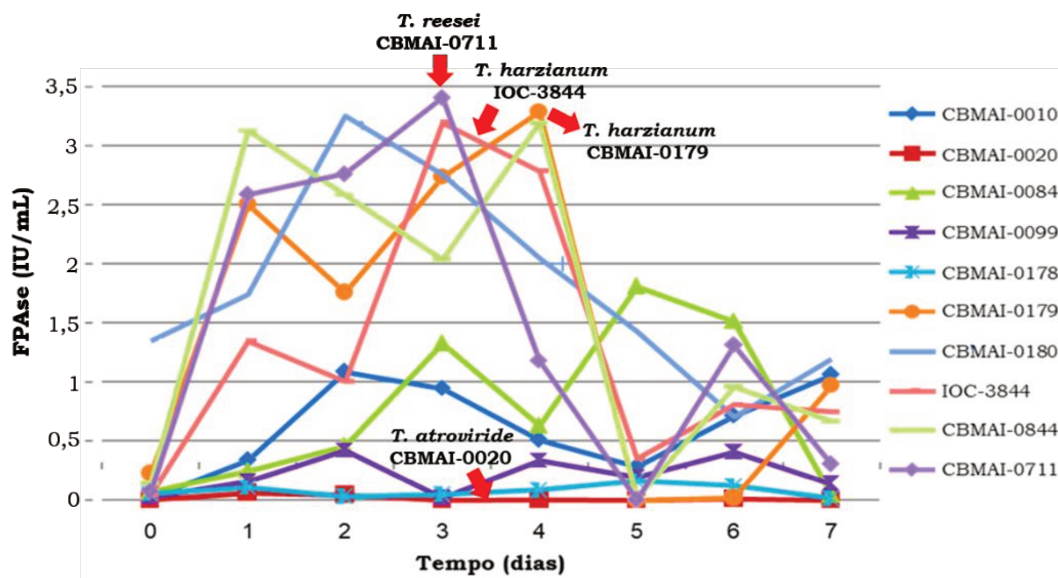


Figura 4: Perfil da atividade enzimática de celulases (FPase) das diferentes espécies de *Trichoderma* em condição de degradação da biomassa vegetal [101]. As linhagens de *T. harzianum* (IOC-3844 e CBMAI-0179) apresentaram atividade enzimática semelhante a *T. reesei* CBMAI-0711, fungo modelo na produção de enzimas hidrolíticas. Em contraste, *T. atroviride* CBMAI-0020 não apresentou desempenho enzimático satisfatório na condição analisada.

Em vista do exposto, foi publicado o primeiro transcriptoma de *T. harzianum* IOC-3844 em diferentes condições (bagaço, celulose e lactose) buscando investigar o potencial desse fungo quanto à degradação do substrato lignocelulósico [15]. Posteriormente, o transcriptoma, o exoproteoma e as atividades enzimáticas dos extratos de *T. harzianum* IOC-3844 foram comparados aos daqueles de *T. reesei* CBMAI-0711 e de *T. atroviride* CBMAI-0020, revelando um sistema comum de biodegradação das espécies estudadas, o que possibilitou a prospecção de novos alvos chaves no processo de sacarificação enzimática [34].

Mais recentemente, foi analisado (dados não publicados) o primeiro transcriptoma de *T. harzianum* CBMAI-0179 em diferentes condições (celulose e glicose) [39]. Os resultados obtidos revelaram o potencial de tal microrganismo como produtor de enzimas chaves na conversão de polissacarídeos complexos em açúcares simples e reiteraram o baixo desempenho enzimático de *T. atroviride* CBMAI-0020 em condição de degradação de celulose [39]. Além do mais, outros trabalhos têm sido realizados com linhagens de *T. harzianum* mostrando o potencial dessa espécie na produção de enzimas hidrolíticas [17-20, 102].

De acordo com a filogenia proposta por Horta e colaboradores (2018), a qual considerou a região ITS (*Internal Transcribed Spacer*), *T. atroviride* CBMAI-0020 apresentou-se

distantemente relacionado de *T. harzianum* IOC-3844 e de *T. reesei* CBMAI-0711 [34]. Embora a semelhança entre as espécies *T. atroviride* e *T. harzianum* mostre que elas compartilham um fundo genético amplamente comum, pequenas diferenças genéticas podem ser detectadas ao comparar seus desempenhos enzimáticos, os quais mostraram-se discrepante, como pode ser observado na **Figura 4**.

Tais diferenças poderiam ser condizentes à regulação de genes responsáveis pela expressão de enzimas que atuam na degradação da parede celular vegetal. Assim, uma vez que dentro do gênero *Trichoderma*, o modelo de regulação gênica disponível para este processo é descrito para *T. reesei* [25, 103-106], o estudo deste tópico em outros fungos seria de fundamental importância para entender como ocorre a expressão de enzimas hidrolíticas nestes microrganismos. Diante deste cenário e para fins de comparação, neste estudo consideramos um ancestral comum micoparasitário, *T. atroviride* CBMAI-0020, e duas linhagens de *T. harzianum* (CBMAI-0179 e IOC-3844) com potencial celulolítico e hemicelulolítico.

2.5. Regulação gênica associada à desconstrução da biomassa vegetal

De maneira geral, em fungos, a regulação de genes relacionados à utilização da biomassa vegetal ocorre pela ativação da expressão gênica via indutores específicos em equilíbrio com a repressão catabólica de carbono (CCR). Na presença de indutores, tais como soforose, celobiose, xilose, xilobiase, xilo-oligossacarídeos, arabitol, galactose e lactose [23, 24], a regulação é feita por FTs específicos que se ligam a região promotora dos genes que codificam celulasas, hemicelulasas e outras CAZymes, ativando a expressão dos mesmos (**Figura 5**), bem como das vias metabólicas necessárias para a utilização da fonte de carbono disponível [28].

Dentre estes, XYR1 (*xylanase regulator 1*) é o ativador transcricional mais importante dos genes que codificam celulasas e hemicelulasas (**Figura 5**). Trata-se de um FT do tipo Zn₂Cys₆ que se liga não apenas ao motivo GGCTAA, disposto em locais duplos e invertidos (separados por 10-12 pb), mas também a um único motivo GGC (A/T)₃ [106, 107]. Ortólogos estão presentes em outros fungos desconstrutores de biomassa, como XLnR em *A. niger* e XLR-1 em *N. crassa*, embora o conjunto de genes controlados por este regulador é altamente específico da espécie e parece estar ligado ao estilo de vida dela [108]. Via de regra, este FT controla o catabolismo da d-xilose e o sistema xilanolítico. Porém, em *T. reesei*, o sistema celulolítico [108] e alguns genes do sistema arabinanolítico [109, 110] também são controlados por XYR1. Ainda para essa espécie, a mudança entre o citoplasma e o núcleo é o mecanismo chave para ativar/desativar a expressão de genes alvo. Sob condições de indução, ocorrem a

biossíntese *de novo* e a importação nuclear rápida de XYR1, enquanto o término da indução resulta em sua rápida degradação nuclear [111].

Analisando a região promotora de *cbh1*, o FT ACE2 foi identificado como sendo uma proteína ativadora de celulasas que se liga a um motivo encontrado no promotor deste gene [112]. Este motivo também é reconhecido por XYR1 [106], o que sugere uma interação entre ambos FTs em mediar a expressão gênica de enzimas hidrolíticas (**Figura 5**). Juntamente com o repressor ACE1, acredita-se ainda que ACE2 faça a fina regulação da expressão de *xyr1* em *T. reesei* [24, 104].

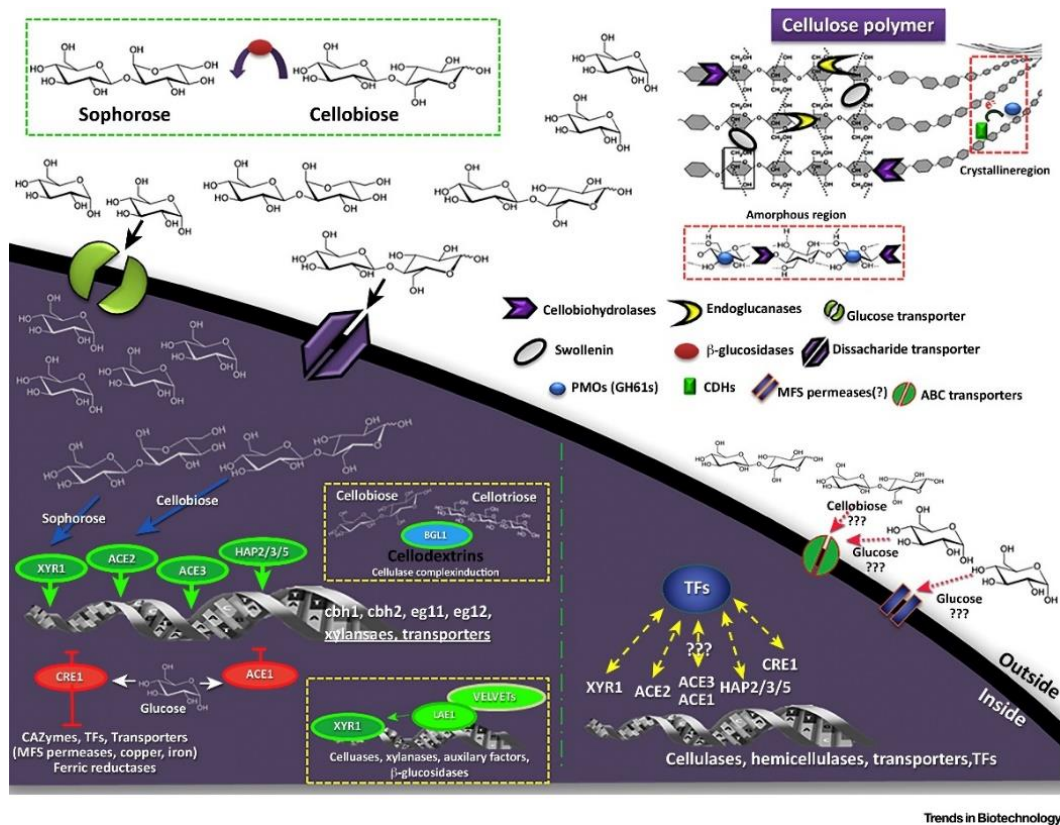


Figura 5. Visão geral da regulação de enzimas hidrolíticas em *T. reesei*. A expressão de celulasas e hemicelulasas é controlada principalmente por FTs, os quais ativam (XYR1, ACE2, ACE3, LAE1, componentes VELVET e o complexo HAP2/3/5) ou reprimem (ACE1 e CRE1) a expressão de enzimas do complexo hidrolítico de acordo com a fonte de carbono presente no meio [7]. Os transportadores, tais como *MFS permeases* e *ABC transporters*, também são essenciais na utilização da lignocelulose pelo fungo, pois sinalizam mudanças ambientais a nível celular através do transporte de moléculas e, consequentemente, controlam a expressão de genes que codificam CAZymes.

Em *T. reesei* e outros fungos, o principal regulador da CCR é o FT CRE1, o qual reprime a expressão dos genes que codificam celulasas e hemicelulasas quando há um açúcar facilmente

metabolizável, tal como a glicose e/ou a frutose (**Figura 5**). Trata-se de um FT do tipo Cys₂His₂ cujos sítios de ligação são compostos por dois motivos SYGGRG e acredita-se que a repressão direta só ocorra quando o fator se liga aos dois sítios [113, 114]. No entanto, dados de estudos *in vitro* sugeriram que apenas um local é necessário, como foi mostrado no promotor *cbh2* em *T. reesei* [115].

A deleção de *cre1* leva à desrepressão da transcrição de genes celulolíticos e hemicelulíticos [26, 116-118]. Entretanto, em termos de aplicação industrial, a deleção completa desse gene não é uma alternativa atrativa para aumentar a expressão de celulasas e hemicelulasas, uma vez que a deleção de *cre1* em *T. reesei* ou sua substituição por um mutante inibe o crescimento das hifas e afeta o acúmulo final de biomassa, levando à diminuição da produção de enzimas hidrolíticas [7, 26, 119]. CRE1 também está envolvido em muitos outros processos, como micoparasitismo em *T. harzianum*, transporte/metabolismo de nitrogênio e aminoácidos em *A. nidulans* e *T. reesei* e controle ambiental do pH em fungos patogênicos [28]. Assim sendo, sua deleção completa causa efeitos pleiotrópicos [105].

Dessa forma, novas abordagens vêm sendo estabelecidas a fim de reduzir a CCR mediada por CRE1 em *T. reesei*, tal como a construção de FTs de fusão [120]. Com essa abordagem, *cre1* é mantido intacto para não prejudicar o crescimento da hifa e, consequentemente, a eficiência na produção de enzimas hidrolíticas.

Outros FTs também atuam na rede regulatória associada a degradação da parede lignocelulósica. Dentre estes podem ser citados os reguladores positivos ACE2, ACE3, complexo HAP2/3/5, LAE1 e componentes VELVET (**Figura 5**) [28]. Quanto aos reguladores negativos, além de CRE1, ACE1 reprime a expressão de celulasas e hemicelulasas, como ilustrado na **Figura 5**, e ambos regulam transcricionalmente a expressão de *xyl1* em condição de repressão [104].

2.6. Proteínas transportadoras e degradação do material lignocelulósico

Aproximadamente 5% do genoma de *T. reesei* compreende genes que codificam proteínas envolvidas no transporte [43]. As proteínas transportadoras desempenham um papel importante durante o processo de degradação da biomassa, uma vez que podem transportar pequenas moléculas indutoras do ambiente extracelular para o meio intracelular do fungo, influenciando a expressão dos genes que codificam as CAZymes (**Figura 5**) [121, 122]. Entre estas, o maior grupo de transportadores identificados pertence à superfamília dos facilitadores maioritários (MFS), classe de transportadores de açúcares, tais como, hexose e pentose, seguida pelos transportadores ABC (*ATP-binding cassette*), os quais estão envolvidos com a secreção de metabólitos secundários, resistência a compostos tóxicos e sinalização celular [99]. Além

disso, em *T. reesei*, foi demonstrado que tais transportadores são modulados pelos FTs XYR1 e CRE1, sugerindo que ambos estejam envolvidos durante o processo de desconstrução do substrato lignocelulósico nesse fungo [37].

2.7. Análise de transcriptoma por RNA-Seq

As moléculas de RNA são componentes vitais de todas as células vivas devido ao seu papel fundamental na intermediação entre a informação genômica e o proteoma (RNAs mensageiros ou mRNAs), como elementos funcionais que atuam na regulação da expressão gênica (RNAs não codificadores ou ncRNAs), ou como catalisadores de reações bioquímicas (ribozimas). Quando tais moléculas são expressas em conjunto tem-se o transcriptoma, o qual pode ser investigado usando tecnologias de sequenciamento de RNA (RNA-Seq), dentre outros métodos [30, 123].

Resumidamente, a primeira etapa do RNA-Seq refere-se à construção de uma biblioteca de cDNA, a qual compreende a síntese de fragmentos de DNA complementares a partir de RNA isolado e, posteriormente, a ligação de adaptadores. Em seguida, os fragmentos podem ser amplificados e depois sequenciados, gerando leituras curtas de 30 a 400 pb a partir de uma extremidade (*single-end sequencing*) ou de ambas as extremidades (*paired-end sequencing*) [30].

Os principais objetivos da realização da análise de RNA-Seq incluem: descoberta e classificação de transcritos, anotação da estrutura gênica e quantificação da mudança da abundância de transcritos ao comparar diferentes condições, amostras ou estágios de desenvolvimento [30]. No entanto, é válido ressaltar que a descoberta de genes diferencialmente expressos entre duas ou mais condições biológicas contrastantes é, sem dúvida, o principal objetivo da maioria dos estudos de RNA-Seq em biotecnologia [34, 35]. Ademais, os dados de RNA-Seq podem ser usados para descobrir outros padrões interessantes de expressão gênica por meio da construção de redes de co-expressão [124].

2.8. Redes de co-expressão gênica como ferramenta biotecnológica

Com o rápido acúmulo de grandes conjuntos de dados “ômicos” gerados a partir da transcriptômica, por exemplo, um dos desafios está em como extrair significados biológicos relevantes a partir dos resultados obtidos. Uma abordagem interessante para explorar os conjuntos de dados de transcriptoma é a construção de redes de co-expressão, também chamadas de redes de associação, correlação e influência [31].

Na rede de co-expressão gênica, os nós representam os genes e as arestas as interações entre esses elementos. Dessa forma, parte-se do pressuposto de que genes envolvidos em uma mesma via metabólica – degradação da parede vegetal, por exemplo –, são expressos de maneira

semelhante sob diferentes tratamentos, além de contemplarem funções análogas [33]. Tais elementos potencialmente co-regulados na resposta do organismo à determinada condição de crescimento podem ser agrupados em *clusters* de genes com alta conectividade [125]. Portanto, espera-se que genes com funções desconhecidas estejam agrupados com genes já caracterizados, abrindo perspectivas para a identificação e caracterização de novos alvos [33] (**Figura 6**).

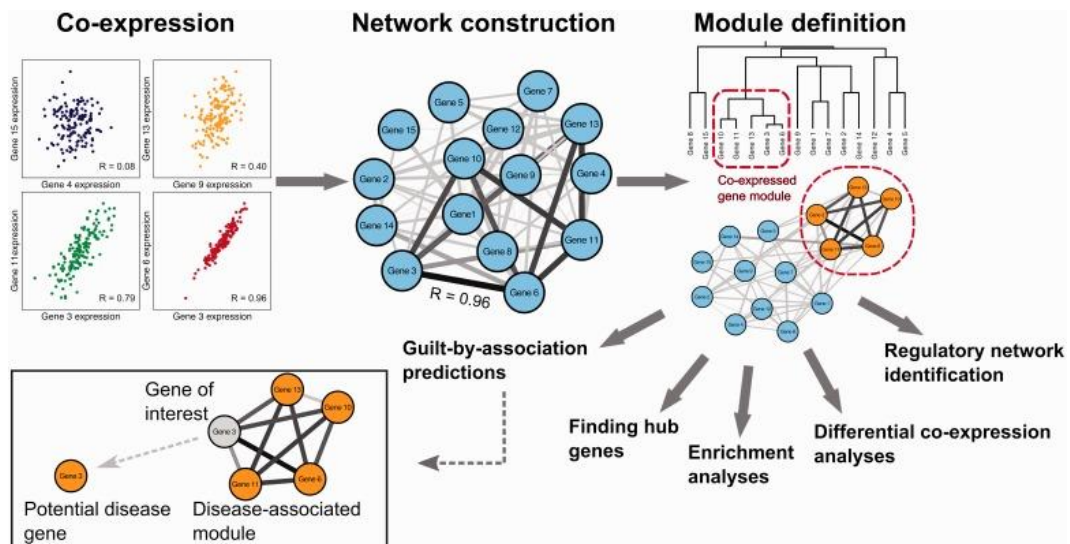


Figura 6. Exemplo de uma análise de rede de co-expressão. Primeiro, a correlação pareada é determinada para cada par de genes a partir dos dados de expressão. Em seguida, essas correlações aos pares podem ser representadas como uma rede, onde módulos podem ser definidos usando a análise de *cluster*. Em seguida, são realizados o enriquecimento funcional e a busca de genes *hub*. Os genes potenciais podem então ser identificados em módulos onde diversos genes associados são co-expressos [32].

Uma outra abordagem utilizada a fim de buscar genes chaves dentro de um processo biológico é a detecção de *hubs*. Esses genes altamente conectados desempenham um papel central na estabilidade da rede contra perturbações e são muito importantes em diversos processos celulares dentro do contexto estudado [126, 127] (**Figura 6**).

Na área de biocombustíveis, são poucos os trabalhos explorando redes de co-expressão gênica em espécies do gênero *Trichoderma* [34, 35, 43]. Na literatura, até onde sabemos, não há trabalhos abordando a análise comparativa e integrativa de diferentes espécies e linhagens de *Trichoderma* tendo como enfoque os principais reguladores positivo e negativo envolvidos na desconstrução da parede celular vegetal. Aqui, é importante ressaltar que a regulação da expressão gênica afeta diretamente a composição das misturas enzimáticas resultantes e é, portanto, altamente relevante para aplicações [128].

Deste modo, este estudo buscou investigar o perfil de expressão gênica relacionado aos FTs XYR1 e CRE1 em condição de degradação da celulose para *T. harzianum* (IOC-3844 e CBMAI-0179) e *T. atroviride* CBMAI-0020. Para isso, foi utilizada a Análise de Rede de Correlação Ponderada (WGCNA) [124], a qual estima a similaridade da co-expressão entre os genes e constrói uma matriz de correlação ponderada seguindo uma topologia sem escala. Através dessa abordagem foi possível identificar e comparar os transcritos co-expressos com *xyr1* e *cre1* bem com os *hubs* destes módulos. Do mesmo modo, as vias metabólicas desencadeadas por tais reguladores transcricionais em resposta à utilização da fonte de carbono presente no meio também foram investigadas.

3. OBJETIVOS

3.1. Objetivo Geral

Caracterizar e comparar a partir de abordagens *in silico* os FTs XYR1 e CRE1 em espécies do gênero *Trichoderma*.

3.2. Objetivos Específicos

- Utilizar a sequência específica de *xyl1* e *cre1* de *T. harzianum* T6776 e de *T. atroviride* IMI206040 como referência para o alinhamento dos *reads* de RNA-Seq de *T. harzianum* (CBMAI-0179 e IOC-3844) e *T. atroviride* (CBMAI-0020), respectivamente;
- Investigar o perfil de expressão de *xyl1* e *cre1* nos experimentos de RNA-Seq para *T. harzianum* (CBMAI-0179 e IOC-3844) e *T. atroviride* (CBMAI-0020) cultivados em celulose ou em glicose;
- Realizar análise filogenética para as sequências proteicas de XYR1 e CRE1 em diferentes fungos ascomicetos, incluindo as linhagens em estudo;
- Construir redes de co-expressão gênica para *T. harzianum* (IOC-3844 e CBMAI-0179) e *T. atroviride* (CBMAI-0020) a partir dos dados de RNA-Seq e, por meio de análises de agrupamento, identificar os grupos de transcritos co-expressos com *xyl1* e *cre1*, bem como os genes centrais (*hubs*) de tais grupos;
- Caracterizar as funções biológicas encontradas para os transcritos associados aos FTs XYR1 e CRE1 e inferir possíveis vias metabólicas desencadeadas por esses reguladores durante o processo de degradação da celulose em *Trichoderma* spp.;

4. ARTIGO

Comparative gene coexpression networks analysis reveals different strategies of *Trichoderma* spp. associated with the transcriptional factors XYR1 and CRE1 during cellulose degradation.

Rafaela Rossi Rosolen^{1,2}, Alexandre Hild Aono^{1,2}, Déborah Aires Almeida^{1,2}, Jaire Alves Ferreira Filho^{1,2}, Maria Augusta Crivelente Horta^{1,2} and Anete Pereira de Souza^{1,3*}

¹Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas (UNICAMP), Campinas, SP, Brazil

²Graduate Program in Genetics and Molecular Biology, Institute of Biology, UNICAMP, Campinas, SP, Brazil

³Department of Plant Biology, Institute of Biology, UNICAMP, Campinas, SP, Brazil

***Corresponding author**

Prof Anete Pereira de Souza

Dept. de Biologia Vegetal

Universidade Estadual de Campinas

CEP 13083-875

Campinas, São Paulo, Brazil

Tel.: +55-19-3521-1132

E-mail: anete@unicamp.br

Abstract

Background: *Trichoderma* species have attracted attention as alternative plant polysaccharide-degrading enzymes sources, among other reasons. Most *Trichoderma* spp., such as *Trichoderma atroviride* and *Trichoderma harzianum*, are described as mycoparasitic fungi and widely used in agriculture as biocontrol agents. *T. harzianum* is also a potential candidate for hydrolytic enzyme production in which gene expression is tightly controlled. Thus, a better understanding of transcriptional regulation related to the main activator (XYR1) and repressor (CRE1) of cellulases and hemicellulases is necessary. Herein, we explore the genetic mechanisms of the key regulators of genes encoding plant cell wall-degrading enzymes by inferring a gene coexpression network for *T. harzianum* (IOC-3844 and CBMAI-0179) and, for comparative purposes, for *T. atroviride* CBMAI-0020.

Results: Phylogenetic analysis indicated that both regulatory proteins are widely distributed among ascomycete fungi and suggested how *T. atroviride* is differentiated from *T. harzianum*. The coexpression network analyses separated the transcripts into several groups according to the expression profile during growth in cellulose or glucose. Groups with *xyl1* or *cre1* were identified, and within them, transcripts encoding carbohydrate-active enzymes (CAZymes), TFs, sugar and ion transporters, and proteins with unknown function were coexpressed. However, qualitative and quantitative differences among groups and strains were observed. Core transcripts from these groups (hubs) were identified, and they included transcripts not yet characterized or described as related to cellulose degradation. Several metabolic pathways triggered were recognized with high similarity between both regulators during cellulose degradation, but they differed according to the strains. An enrichment related to mycoparasitism in *T. atroviride*, especially that associated with CRE1, was observed. It was also noticed that the transcripts encoding CAZymes related to polysaccharide degradation in *T. harzianum* are not necessarily coexpressed with the TFs studied herein.

Conclusion: Our results suggest that different strategies related to XYR1 and CRE1 are used by *Trichoderma* spp. during cellulose degradation and provide new insights into transcripts coexpressed with both TFs in *T. harzianum*. This knowledge can be exploited to improve the understanding of the genetic mechanisms involved in hydrolytic enzyme regulation, expanding the potential of *T. harzianum* use in several industrial applications.

Keywords: *Trichoderma harzianum*, *Trichoderma atroviride*, cellulose, XYR1, CRE1, weighted gene co-expression network analysis.

Background

Due to the expanding population and industrialization, the demand for energy is increasing throughout the world. Currently, the primary energy source is fossil fuels, i.e., nonrenewable sources such as natural gas, oil, and coal [1]. To reduce the dependency on these finite resources, the development of new sustainable alternatives has become critical. In this context, increasing attention has been given to the development of biofuels from biomass [2, 3].

The United States and Brazil are the world's largest bioethanol producers and are responsible for 56 and 28% of its global production, respectively [4]. In Brazil, the sugar and first-generation biofuel (1G ethanol) production system uses sugarcane as feedstock, which generates large amounts of lignocellulosic materials as residue [5]. On average, 280 kg of bagasse is produced through industrial processing of 1 ton of sugarcane [6]. Thus, new technologies based on converting agricultural residues into second-generation biofuel (2G ethanol) are emerging [7-9].

Despite advances, the cost of biomass-degrading enzymes remains the main bottleneck for commercial biofuel production [10-12]. In nature, bacteria and fungi produce and secrete a combination of enzymes that act synergistically to breakdown complex polysaccharides [13]. In order to establish the 2G ethanol as a competitive technology, the application of these enzymes for the improvement of enzymatic cocktails is earning great attention in the biofuel industry [14, 15].

Because of the exclusive nutritional versatility, species of the filamentous ascomycete genus *Trichoderma* have been used in a wide range of biotechnological applications [16-19]. While fungal cellulases from *Trichoderma reesei* are the most applied enzymes in the biotechnology industry [17, 20], the use of *Trichoderma atroviride* and *Trichoderma virens* have been explored by examining their biocontrol capacity against plant pathogenic fungi [21, 22]. Both species are distantly related to *T. reesei* [23] and represent well-defined phylogenetic species [24, 25]. In contrast, *Trichoderma harzianum sensu lato* is also commonly used as a biocontrol agent but is part of a complex of several cryptic species [26]. More recently, studies on *T. harzianum* strains have shown their potential to produce a set of enzymes that can degrade the residual biomass of sugarcane [27-32].

Lignocellulosic biomass is a complex recalcitrant structure that requires a consortium of enzymes for its complete depolymerization. For cellulose degradation, three types of cellulases are required: cellobiohydrolases, endoglucanases and β -glucosidases. For hemicellulose hydrolysis, several enzymes are needed, such as endo-xylanases, β -xylosidases,

β -mannanases, arabinofuranosidases, α -glucuronidases and esterases [33]. Furthermore, auxiliary enzymes and proteins are involved in the degradation of the plant wall, such as, respectively, lytic polysaccharide monooxygenases (LPMOs) [34] and swollenin [35], which can improve the overall performance of biomass hydrolysis [29].

These hydrolytic enzymes and other accessory proteins related to complete breakdown of cellulose and hemicellulose are collectively named carbohydrate-active enzymes (CAZymes) [36]. The expression of CAZyme-encoding genes is regulated at the transcriptional level in a coordinated manner that depends on the availability of the carbon sources [37-39]. In *T. reesei*, cellulase and hemicellulase gene expression is regulated by the action of several transcription factors (TFs) [40]. The main TF involved in the expression of most cellulase and hemicellulase genes is xylanase regulator 1 (XYR1) [40, 41]. Moreover, the expression of these genes is subject to carbon catabolite repression (CCR) [42], which is regulated by carbon catabolite repressor 1 (CRE1) [43].

Although XYR1 orthologs are present in almost all filamentous ascomycete fungi, the molecular mechanisms triggered by this regulator depend on the species [44]. XYR1 was first described in *T. reesei* as the main activator of enzymes involved in d-xylose metabolism [41]. However, for *T. atroviride*, the induction of genes encoding cell wall-degrading enzymes considered relevant for mycoparasitism, such as *axe1* and *swol*, is influenced by XYR1 [45]. CRE1 is the only conserved TF throughout the fungal kingdom, suggesting a conserved mechanism for CCR in fungi [46], and in many cases, it has a negative effect on the production of lignocellulose-degrading enzymes [47].

One of the great challenges in understanding the molecular mechanisms of expression of CAZymes and other proteins required for biomass deconstruction is how the main TFs act. Given the fact that inside the genus *Trichoderma*, the research about gene regulation of lignocellulose degrading enzymes has focused on *T. reesei* [48-50], a better understanding of the transcriptional regulation on another *Trichoderma* spp. has shown necessary in order to optimize the use of these fungi as enzyme producers. One interesting approach to explore the regulation of lignocellulolytic enzymes *in silico* is through gene coexpression networks, which have proven to be a fundamental tool to indicate the transcripts that are active simultaneously; thus, may have a common function [51, 52].

To date, only a few studies have employed the gene coexpression network in studies on *Trichoderma* spp. [53-55]. Through this methodology, no research has been reported exploring comparatively the genetic mechanisms triggered by the main regulators involved in plant biomass degradation in *Trichoderma* spp. In this context, we considered a common

mycoparasitic ancestor species (*T. atroviride* CBMAI-0020) and two mycoparasitic strains with hydrolytic potential (*T. harzianum* CBMAI-0179 and *T. harzianum* IOC-3844) [54, 55] to model a network for each strain. This approach allowed us to identify and compare modules, hub genes and metabolic pathways associated with CRE1 and XYR1 under cellulose degradation. A better understanding of the regulation of cell wall degrading enzymes will help to elucidate the events that trigger cellulose degradation in mycoparasite *Trichoderma* species. This will be important to increase the efficiency of the production of these enzymes and expanding the potential of *T. harzianum* in several industrial applications, including in biofuel industry.

Materials and Methods

Fungal Strains and Culture Conditions

The species originated from the Brazilian Collection of Environment and Industry Microorganisms (CBMAI), located in the Chemical, Biological and Agricultural Pluridisciplinary Research Center (CPQBA) at the University of Campinas (UNICAMP), Brazil. Briefly, *T. harzianum* CBMAI-0179 (Th0179), *T. harzianum* IOC-3844 (Th3844) and *T. atroviride* CBMAI-0020 (Ta0020) strains were cultivated on solid medium for 8 days at 37 °C as described in a previous work and scraped with a washing solution [54]. Fermentation process was performed in biological triplicates and was initiated with the inoculation of 10⁷ spores/mL per 200 mL of pre-inoculum composed of crystalline cellulose (Celuflok, São Paulo, Brazil, degree of crystallinity 0.72 g/g, composition 0.857 g/g cellulose and 0.146 g/g hemicellulose) or glucose as the carbon source [54]. Glucose was used as a control in RNA-Seq experiments. After 72 h of incubation, 50 mL of the pre-inoculum was used to inoculate 450 mL of fermentation solution in a 2 L Erlenmeyer flask. The fermentation process continued for 96 h. Then, mycelial samples from cellulose and glucose conditions were extracted from the Th0179, Th3844 and Ta0020 strains, stored at -80 °C and ground in liquid nitrogen. Frozen material was used for RNA extraction, and the sequencing experiment was carried out on the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA). These data were obtained and described in a previous study [54].

Transcription Profiling

Sequencing quality filters were applied using Trimmomatic v0.36 [56] as described by [55]. Reads were mapped against *T. harzianum* T6776 [57] and *T. atroviride* IMI206040 [58] reference genomes using CLC Genomics Workbench software (CLC bio – 6.5.2 v;

Finlandsgade, Dk) [59]. Gene expression values were calculated in transcripts per million (TPM), and genes with p-values ≤ 0.05 and fold changes ≥ 1.5 (upregulated) or ≤ -1.5 (downregulated) were considered differentially expressed, as in [55].

Phylogenetic Analysis of CRE1 and XYR1 Proteins

The ascomycete fungus CRE1 and XYR1 protein sequences with similarity $\geq 60\%$ and $\geq 70\%$, respectively, to the reference genome of *T. harzianum* T6776 were retrieved from the NCBI database (<https://www.ncbi.nlm.nih.gov/>). Additionally, from the selected genome, CRE1 and XYR1 sequences were recovered and used as a reference for Th3844, Th0179 and Ta0020 TF consensus sequences in RNA-Seq read mapping. These sequences were also included in the phylogenetic analyses. Multiple sequence alignment was performed using ClustalW [60], and a phylogenetic tree was created using Molecular Evolutionary Genetics Analysis (MEGA) software v7.0 [61]. The maximum likelihood (ML) [62] method of inference was used based on the Dayhoff model with a freqs (F^+) matrix-based model for CRE1 and based on the Jones-Taylor-Thornton (JTT) model with freqs (F^+) for XYR1. We used 1,000 bootstrap replicates [63] for each analysis. Pairwise deletion was employed to address alignment gaps and missing data. Trees were visualized and edited using the FigTree program v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Gene Coexpression Network Construction

The gene coexpression networks were modeled for Th3844, Th0179 and Ta0020 using transcriptome data ([PRJNA336221](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=PRJNA336221)) from a previous work that were obtained with biological triplicates using cellulose or glucose as the carbon source [54]. For the network construction, TPM values were used along with the R package for weighted correlation network analysis (WGCNA) [64]. To decrease noise and to remove residuals from the analysis, WGCNA filters were applied to remove transcripts with zero variance for any replicate under different experimental conditions and a minimum relative weight of 0.1. Pearson's correlation values among all pairs of transcripts across all selected samples for each strain were used for network construction. A softpower β was chosen for each network using *pickSoftThreshold* to fit the signed network to a scale-free topology to obtain an adjacency matrix where nodes correspond to genes and edges to the strength of their connection. To obtain a dissimilarity matrix, we built a topological overlap matrix as implemented in the package, with the exponentiation performed on each edge value.

Identification of Groups, Subgroups and Hubs

To identify groups of transcripts densely connected in the gene coexpression network, we used simple hierarchical clustering on the dissimilarity matrix with R [65]. From the acquired dendrogram, the *dynamicTreeCut* package [66] was used to obtain the ideal number of groups and the respective categorization of the transcripts. According to the functional annotation performed by [55], groups containing the desired TFs XYR1 and CRE1 were identified and named the *xyr1* and *cre1* groups, respectively. Within each detected group associated with different TFs, we used Python 3 programming language [67] together with the *networkx* library [68] to identify subgroups using the Louvain method [69] and hubs based on the degree distribution. We used different thresholds for keeping edges on the networks and recognizing hubs: for the *cre1* group, (I) 0.9 (all strains) was used, and for the *xyr1* group, (II) 0.9 (Th3844), (III) 0.5 (Th0179) and (IV) 0.7 (Ta0020) were used. Additionally, we tested the Molecular Complex Detection (MCODE) (v1.4.2) plugin [70] using Cytoscape software v3.7.2 with the parameters set to default (degree cutoff = 2, node score cutoff = 0.2, K-core = 2, maximum depth = 100, haircut method) for subgroup identification.

Functional Annotation of Transcripts within *xyr1* and *cre1* Groups

Functional annotation of the *xyr1* and *cre1* groups was performed based on an analysis of their transcript composition. Gene Ontology (GO) [71] terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) [72] Orthologies (KOs) related to the selected groups were chosen to explore the biological functions and metabolic pathways in which these transcripts participate. All identified GO categories were used to create a treemap to visualize possible correlated categories in the dataset caused by cellulose degradation. This step was performed using the REVIGO tool [73]. The metabolic pathways related to the KOs obtained and found in *T. reesei* were then selected due the high number of annotations correlated to this species in the KEGG database. To obtain the pathways related to the enzymes identified as belonging to the different groups, we used the Python 3 programming language [67] together with the BioPython library [74]. We selected metabolic pathways linked to the TFs XYR1 and CRE1 that act in categories related to biomass degradation in *Trichoderma* species.

Results

Molecular Phylogeny of CRE1 and XYR1 in Ascomycota

The phylogenetic analysis of the amino acid sequences related to CRE1 and XYR1 showed a clear diversification of these regulators inside the ascomycete fungi (**Fig. 1** and

Additional file 1: Table S1). The genus *Trichoderma* is a member of the family Hypocreaceae, and we observed in both phylogenetic trees the separation of this family from the other families classified in the phylum Ascomycota (with bootstrap support of 100% in both cases). However, the families observed in the CRE1 phylogenetic tree (**Fig. 1a**) were not the same as those identified in XYR1 (**Fig. 1b**) as well the phylogenetic relationships of the complete protein sequences related to both TFs of the studied *Trichoderma* species/strains.

The CRE1 phylogenetic tree showed a high phylogenetic proximity between Th3844 and Th0179 strains (with bootstrap support of 51%). For both *T. harzianum* strains, the protein sequence related to CRE1 presented a close relationship with the corresponding sequence in other species of the same genus, such as *T. virens* (XP_013956509.1), *Trichoderma guizhouense* (OPB38360.1) and *Trichoderma lixii* (CAA64656.1) as well in other *T. harzianum* (PNP52054.1, KKO97123.1 and XP_024780493.1) (with bootstrap support of 95%) (**Fig. 1a**). On the other hand, Ta0020 shared closer affinity with two other *T. atroviride* strains (XP_013941427.1 and AKM12665.1), with 99% bootstrap support (**Fig. 1a**).

We observed in the phylogenetic tree of XYR1 a high phylogenetic proximity between Th3844 and *T. harzianum* CBS 226.95 (XP_024780881.1) (with bootstrap support of 96%) (**Fig. 1b**), while Th0179 was closer to *T. harzianum* T6776 (KKO98630.1) (with bootstrap support of 50%) (**Fig. 1b**) and Ta0020 was closer to *T. atroviride* IMI206040 (XP_013941705.1) (with bootstrap support of 88%) (**Fig. 1b**).

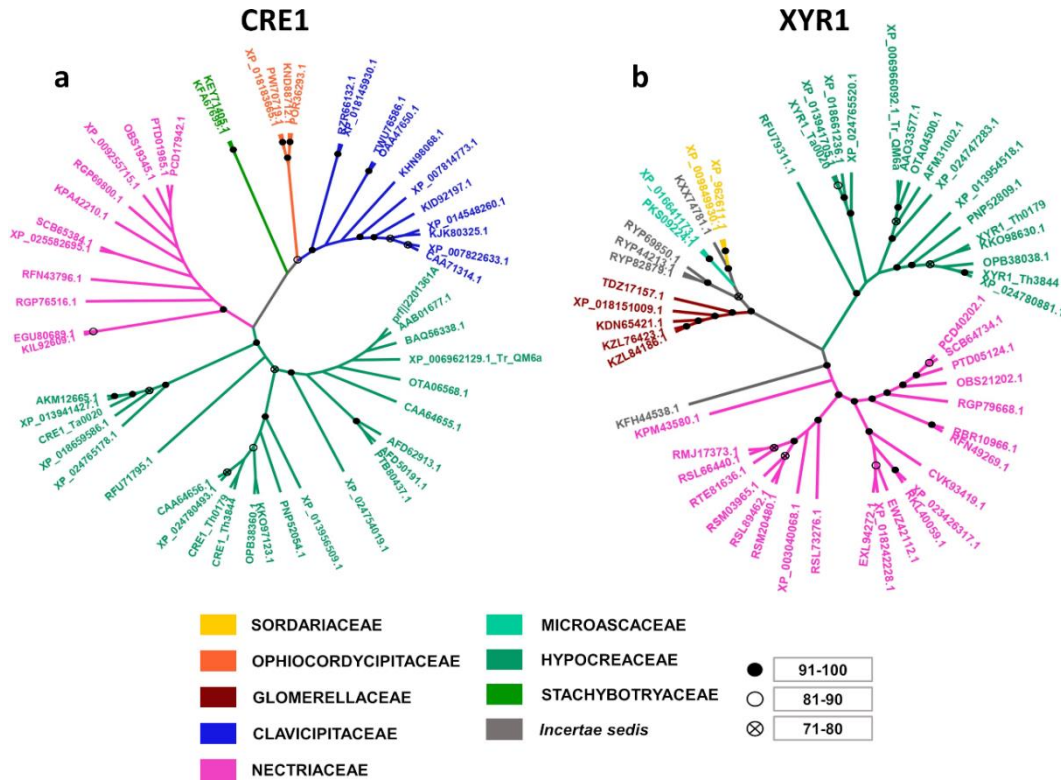


Fig. 1. Molecular phylogeny of CRE1 and XYR1 in Ascomycota. The complete protein sequences related to CRE1 (a) and XYR1 (b) were used to analyze and to compare the phylogenetic relationships of the studied *Trichoderma* species/strains inside the ascomycete fungi.

Construction of Gene Coexpression Networks for *Trichoderma* spp.

We used transcriptome data obtained in a previous work [55] for Th3844, Th0179 and Ta0020. Through this dataset, gene coexpression networks were modeled for each strain. Transcript identification was obtained from comparative alignments of sequencing data and the following reference sequences: the *T. harzianum* T6776 genome for Th3844 and Th0179 and the *T. atroviride* IMI206040 genome for Ta0020 [55]. Using these data and applying the described filters to the WGCNA package, we obtained (I) 11,050 transcripts (Th3844), (II) 11,105 transcripts (Th0179), and (III) 11,021 transcripts (Ta0020) (Additional file 2: Table S2). Based on these transcripts, we calculated Pearson's correlation matrices. Different softpower β values were chosen to fit the scale independence of the different networks to a scale-free topology (27 for Ta0020, 8 for Th0179, and 48 for Th3844) by considering the connectivity between the transcripts based on their expression correlation. Based on transcript expression levels, the networks were partitioned into manageable groups to explore the putative

coregulatory relationships. In total, we identified 87 groups for Th3844, 75 groups for Th0179 and 100 groups for Ta0020 (Additional file 3: Table S3).

Coexpressed Transcripts with XYR1 and CRE1

Our focus in *Trichoderma* spp. coexpression networks were the TFs XYR1 and CRE1, two regulators that play a central role in the expression of cellulase and hemicellulase genes during biomass degradation [46]. Therefore, transcripts encoding these proteins were sought within each module generated using the WGCNA package. Transcripts having correlated expression patterns with these regulators were identified in the Th3844, Th0179 and Ta0020 coexpression networks and grouped together (Additional file 4: Table S4). In relation to the number of transcripts, Ta0020 presented the highest coexpression with *cre1* and the lowest coexpression with *xyl1* compared to the number of transcripts identified in the other strains, which showed the opposite profile (**Table 1**). Between *T. harzianum* strains, the amount of coexpressed transcripts related to *xyl1* and *cre1* were differed (**Table 1**).

Table 1. Total numbers of coexpressed transcripts identified in *Trichoderma* spp. distributed in the *cre1* and *xyl1* groups.

Groups	Strains		
	Th3844	Th0179	Ta0020
<i>cre1</i>	53	116	251
<i>xyl1</i>	167	153	50
Total	220	269	301

Each group previously identified containing the *xyl1* and *cre1* transcripts was partitioned, and subgroups were formed for all strains. Thus, we could identify the transcripts that are most correlated inside the modules. We tested the MCODE methodology as previously used for gene coexpression networks [53]. The number of subgroups identified was (I) 5 for Ta0020 (the subgroups contained 23, 10, 23, 16 and 6 transcripts), (II) 4 for Th0179 (the subgroups contained 3, 3, 5 and 4 transcripts), and (III) 3 for Th3844 (the subgroups contained 20, 14, and 9 transcripts) for subgroups related to the *cre1* groups, and (IV) 1 for Ta0020 (8 transcripts), (V) 4 for Th0179 (the subgroups contained 56, 18, 16 and 7 transcripts), and (VI) 2 for Th3844 (the subgroups contained 124 and 43 transcripts) for subgroups related to the *xyl1* groups (Additional file 5: Table S5).

We sought the transcripts encoding CRE1 and XYR1 inside these subgroups and found them for only Th3844 (*cre1* and *xyr1*) and Th0179 (*cre1*). Although we found transcripts encoding CAZymes, transporters, TFs and other interesting proteins, the MCODE methodology was not useful for all the modeled coexpression networks. Thus, we tested the Louvain algorithm [69] for detecting the most correlated niches in the *cre1* and *xyr1* groups. This approach enabled the identification of the main subgroups for the molecular reactions occurring during cellulose degradation (**Fig. 2** and Additional file 6: Table S6).

As can be seen in **Fig. 2**, each strain presented a different coexpression subnetwork topology for *cre1* and *xyr1* groups based on the distribution of the subgroups. For *cre1* groups, the Th0179 subnetwork is more tightly interconnected than the Ta0020 subnetwork and is less than Th3844 subnetwork. For *xyr1* groups, the same profile was observed for all strains. For each subgroup formed inside a group, a color was attributed to the transcripts that constituted the respective subgroup (Additional file 7: Table S7).

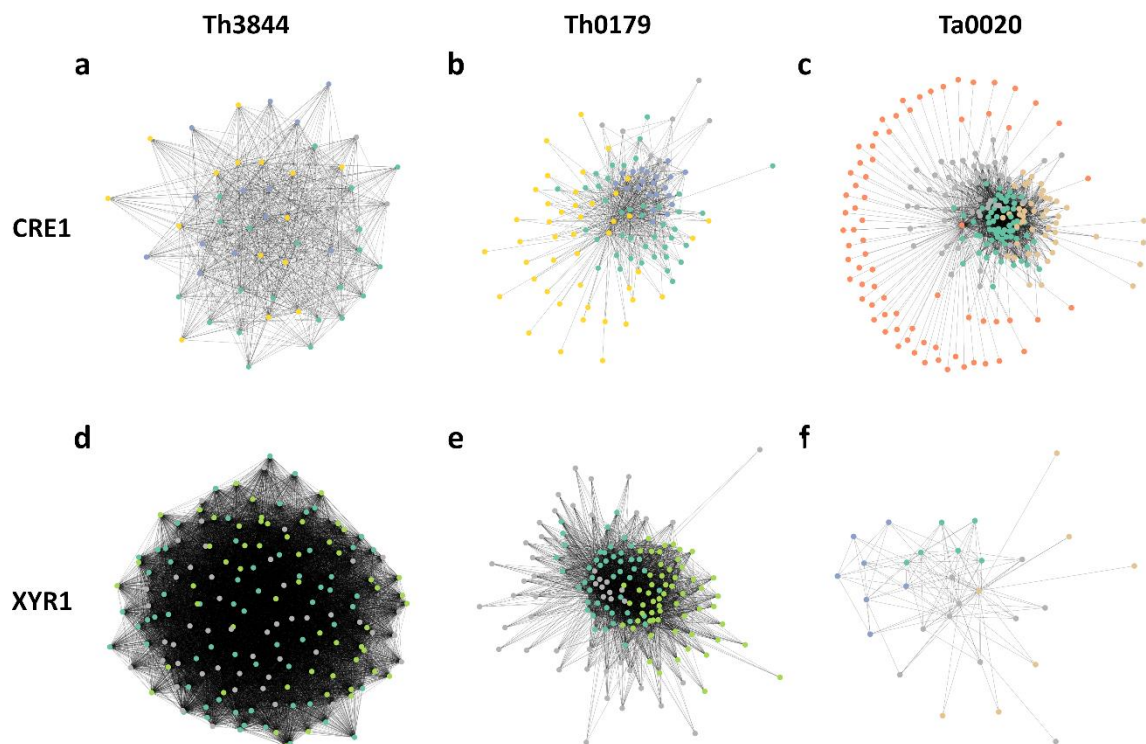


Fig. 2. Subgroups identified in the *cre1* and *xyr1* groups among *Trichoderma* spp. The groups containing CRE1 and XYR1 were partitioned according to Louvain algorithm, and subgroups were formed for the Th3844 (a), Th0179 (b) and Ta0020 (c) groups related to *cre1* and for the Th3844 (d), Th0179 (e) and Ta0020 (f) groups related to *xyr1*. Different colors represent different subgroups inside a group. Among groups, equal colors do not represent the

same transcripts. Th3844: *T. harzianum* IOC-3844; Th0179: *T. harzianum* CBMAI-0179; Ta0020: *T. atroviride* CBMAI-0020.

For each subgroup formed from the coexpressed transcripts in the *cre1* and *xyl1* groups, a denotation was attributed according to the strain (Additional file 7: Table S7). The *cre1* transcript was in the **1.0**, **2.1** and **3.5** subgroups for Th3844, Th0179 and Ta0020, respectively (Additional file 7: Table S7). We found the *xyl1* transcript associated with the transcripts that formed the **4.1**, **5.2** and **6.5** subgroups for Th3844, Th0179 and Ta0020, respectively (Additional file 7: Table S7). In addition, we observed the presence of disconnected elements in the subgroups formed for Th0179 and Ta0020 in the *cre1* group. This same pattern was observed in the *xyl1* group for Ta0020 (Additional file 7: Table S7). For *cre1* subgroup, Ta0020 showed more than double of transcripts compared to identified in the *T. harzianum* strains and for *xyl1* subgroup, presented a few numbers of transcripts (Additional file 7: Table S7).

Hubs Transcripts with Similar Expression Patterns to XYR1 and CRE1

The predictive potential of transcripts with high connectivity in gene coexpression networks involved in cellulose degradation by *Trichoderma* spp. remains to be thoroughly investigated. In this way, transcripts with a high number of coexpressed neighbors, defined here as hubs, were sought in groups in which the transcripts *cre1* and *xyl1* were identified (Additional file 8: Table S8). In some cases, we had to decrease the size of the correlation coefficient threshold to identify highly connected transcripts within a group. Thus, different thresholds representing cases from a moderate correlation (0.5) to a very high correlation (0.9) [75] were used. To avoid extrapolating inferences, only transcripts with a degree of connectivity were selected to improve understanding.

In the *cre1* group, the transcripts encoding a concentrative nucleoside transporter (CNT) (TRIATDRAFT_46358) for Ta0020, a mitochondrial ribosomal protein (MRP) family L3 (THAR02_01324) for Th0179 and a proteasome component PRE3 precursor (THAR02_10705) and CRE1 (THAR02_10775) for Th3844 were identified as hubs, with the last having a second candidate. In the *xyl1* group, transcripts encoding unknown protein were found to be hubs for Ta0020 (TRIATDRAFT_40550), for Th0179 (THAR02_07681) and for Th3844 (THAR02_00323). A transcript encoding XYR1 (TRIATDRAFT_78601) was also defined as hub for Ta0020.

Functional Annotation of Transcripts within *xyl1* and *cre1* Groups

GO enrichment analysis was performed with the transcripts within each *creI* and *xylI* group. For the *creI* group, under the molecular function category, catalytic activity (GO: 0003824) was the main annotated function for Th3844, Th0179 and Ta0020. Other functions that were enriched among these species under this category were hydrolase activity (GO: 0016787), nucleic acid binding (GO: 0003676), oxidoreductase activity (GO: 0016491) and metal ion binding (GO: 0046872). For the biological process category, the main function was metabolic process (GO: 0008152). Other functions observed in all strains that deserve special attention under this category were oxidation-reduction process (GO: 0055114) and transcription/DNA-templated (GO: 0006351). In addition, for the cellular component category, the main function was integral component of membrane (GO: 0016021), followed by cytoplasm (GO: 0005737) and nucleus (GO: 0005634).

For the *xylI* group, catalytic activity was the main annotated function for all strains under the molecular function category. Other enriched functions found in *T. harzianum* strains were ATP binding (GO: 0005524), DNA binding (GO: 0003677), zinc ion binding (GO: 0008270) and oxidoreductase activity. It is important to note that hydrolase activity was the second main annotated function under the molecular function category for Th3844, whereas for Th0179 and Ta0020, this activity was not enriched. For the biological process category, the main function was metabolic process. Other functions shared by Th3844, Th0179 and Ta0020 enriched under this category were: oxidation-reduction process, transcription/DNA-templated and regulation of transcription/DNA-templated (GO: 0006355). For the cellular component, the main function was integral component of membrane, followed by nucleus.

KO functional annotation of the transcripts encoding enzymes and unknown proteins with enzymatic activity enriched in the *creI* and *xylI* groups was also investigated (**Fig. 3**). For the 14 pathway classes observed for each TF, 13 were shared between the two regulators. Only two pathway classes, metabolism of other amino acids and lipid metabolism, were exclusive to the TFs CRE1 and XYR1, respectively. However, the number of identified enzymes for a given pathway class was different for each strain (**Fig. 3**).

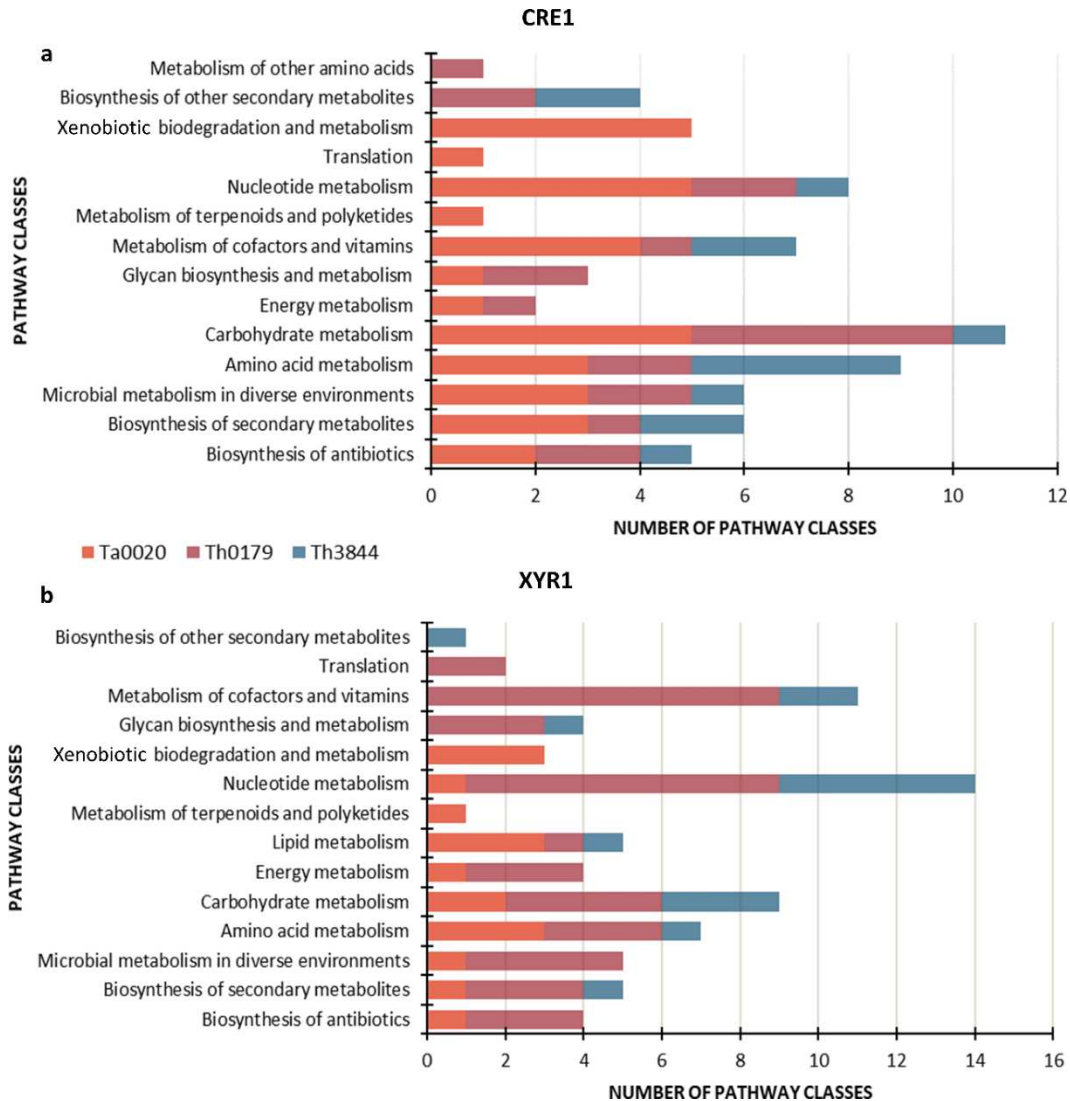


Fig. 3. KO functional classification of the transcripts identified in the *creI* and *xyrI* groups. The sequences related to enzymes and coexpressed with *creI* (a) and *xyrI* (b) were annotated according to the main KO functions for Th3844, Th0179 and Ta0020. Ta0020: *T. atroviride* CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844; KO: Kyoto Encyclopedia of Genes and Genomes Orthology.

In relation to the pathway classes triggered along with *creI*, carbohydrate metabolism and amino acid metabolism were highly activated for Th3844 and Th0179, respectively, whereas for Ta0020, xenobiotic biodegradation and metabolism, nucleotide metabolism and carbohydrate metabolism were highly activated (**Fig. 3a**). The pathway classes enriched in the *xyrI* groups were identified as nucleotide metabolism and metabolism of cofactors and vitamins for Th3844, nucleotide metabolism for Th0179 and xenobiotic biodegradation and metabolism, lipid metabolism and amino acid metabolism for Ta0020 (**Fig. 3b**).

The number of metabolic pathways for the *cre1* and *xyl1* groups shared among all strains was larger than that observed exclusively for each strain or between only two strains/species (Additional file 7: Fig. S1). In addition, compared to the other two strains, Ta0020 presented the highest number of exclusive pathway classes related to both TFs (Additional file 7: Fig. S1).

Differentially Expressed Transcripts from *xyl1* and *cre1* Groups

We determined which of the transcripts present in the *cre1* and *xyl1* groups were differentially expressed considering the statistical analyses performed by [55]. We found different quantities of differentially expressed transcripts for Ta0020 (78) and for Th0179 (6) (Additional file 9: Table S9). For Ta0020, most of the upregulated transcripts under cellulose growth (70) were associated with the *cre1* group, including *cre1* ($p\text{-value} = 0$) whereas *xyl1* showed a potentially significant expression level under glucose growth ($p\text{-value} = 0.05$) (Additional file 7: Table S10). However, for Th3844, *cre1* showed a significant modulation in expression ($p\text{-value} = 0.03$) between the two carbon sources, with a high expression level under glucose growth (Additional file 7: Table S10). Even with the reduced differentiation shown (-1.13), this phenomenon has already been considered by other studies [76-78]. For Th0179, *xyl1* ($p\text{-value} = 0.08$) and *cre1* ($p\text{-value} = 0.27$) showed no significant modulation in transcript expression between growth under cellulose or glucose, but *xyl1* had a higher transcript expression level under glucose whereas *cre1* under cellulose growth conditions (Additional file 7: Table S10). For Th3844, the same was observed for *xyl1* ($p\text{-value} = 0.29$) but with a higher transcript expression level under glucose (Additional file 7: Table S10).

For Ta0020, the differentially expressed transcripts identified under cellulose growth in the *cre1* group included transcripts encoding CAZymes, sugar transporters, uncharacterized proteins and proteins related to secondary metabolism. Under glucose growth conditions, few differentially expressed transcripts were found. The differentially expressed transcripts identified in the *xyl1* group for Th0179 were not described yet as related to cellulose degradation (Additional file 9: Table S9).

CAZymes, TFs and Transporters Distribution Among *cre1* and *xyl1* Groups

Due to the importance of CAZymes, TFs and transporters for plant biomass degradation, we sought transcripts encoding these types of proteins within the *cre1* and *xyl1* groups (**Fig. 4**). Within the *cre1* group, Ta0020 presented the highest number of transcripts encoding these proteins (**Fig. 4a**). Between the *T. harzianum* strains, only Th0179 showed transcripts encoding

TFs, transporters and CAZymes, whereas Th3844 presented transcripts encoding only CAZymes (**Fig. 4a**).

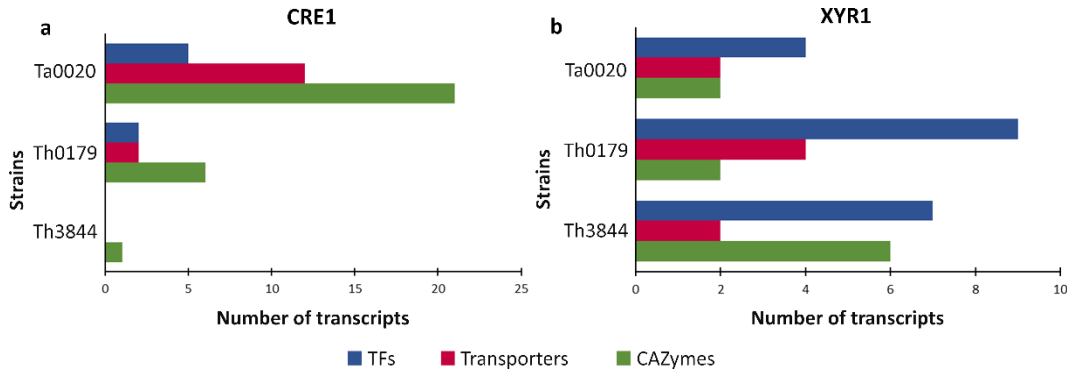


Fig. 4. TFs, transporters and CAZymes comparison among strains and groups. Number of transcripts encoding TFs, transporters and CAZymes for Ta0020, Th0179 and Th3844 distributed among *cre1* (a) and *xyr1* (b) groups. TFs: Transcription factors; CAZymes: Carbohydrate-active enzymes. Ta0020: *T. atroviride* CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844.

The number of transcripts encoding TFs, transporters and CAZymes associated with *xyr1* were more distributed among all strains than those observed in the *cre1* group (**Fig. 4b**). However, comparing with the other two strains, Th0179 showed the highest number of TFs and transporters. While, Th3844 presented the highest number of coexpressed CAZymes (**Fig. 4b**).

We also investigated transcripts encoding TFs, transporters and CAZymes within the *cre1* and *xyr1* subgroups for all analyzed strains. Within the *cre1* subgroup, Th3844 did not present transcripts encoding these types of proteins, whereas Th0179 presented transcripts encoding CAZymes, TFs and transporters, and Ta0020 showed transcripts encoding TFs and transporters. This profile was not observed in the *xyr1* subgroup, where Th3844 presented transcripts encoding CAZymes, TFs and transporters; Th0179 presented transporters and TFs; and Ta0020 presented a TF and a CAZyme.

CAZymes

Considering that hydrolytic enzymes play an important role in plant biomass degradation, a deep investigation of the CAZymes present in the *xyr1* and *cre1* groups was performed considering all strains. We identified the main CAZyme classes, such as carbohydrate esterases (CEs), glycoside hydrolases (GHs), glycosyltransferases (GTs) and

polysaccharide lyases (PLs) in these groups, and their profiles were compared to detect similarities and differences between the strains (**Fig. 5**).

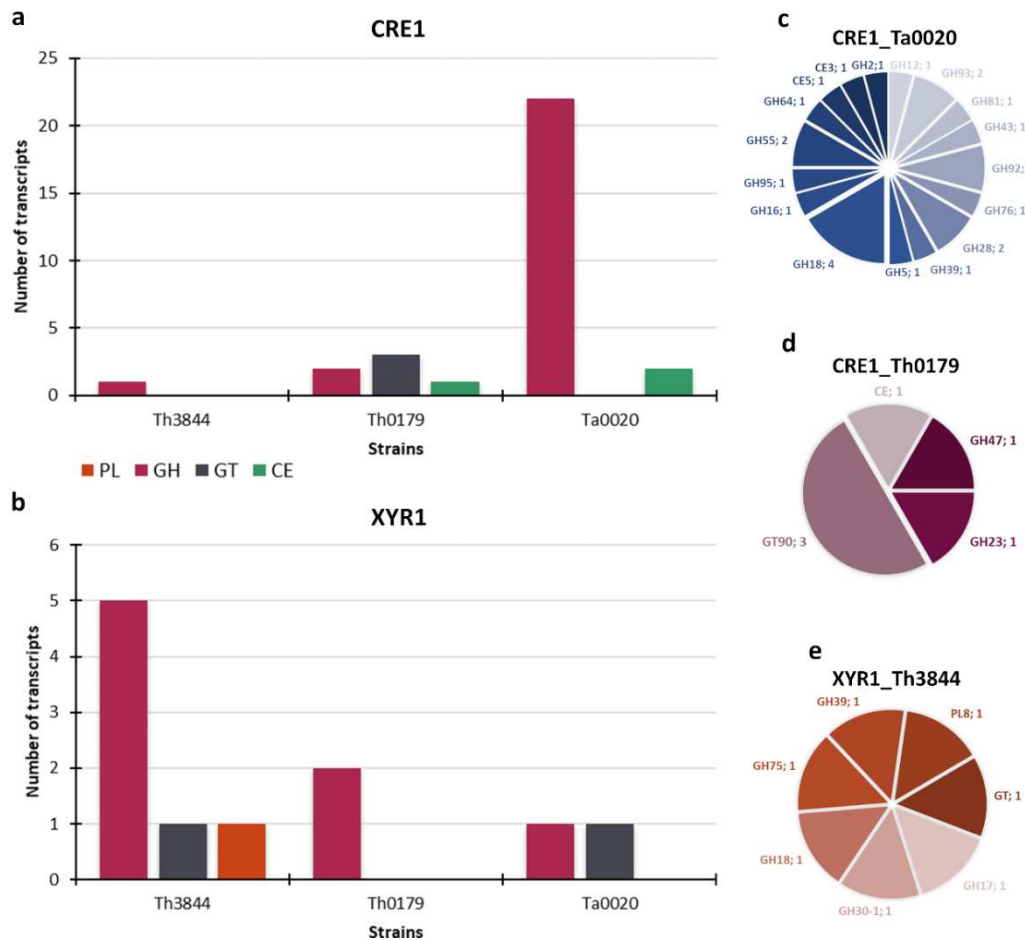


Fig. 5. Distribution of CAZyme families within *cre1* and *xyr1* groups for *Trichoderma* spp. Classification of CAZyme families coexpressed with *cre1* (**a**) and *xyr1* (**b**) for Th3844, Th0179 and Ta0020 and quantification of each CAZyme family in the *cre1* group for Ta0020 (**c**) and Th0179 (**d**) and in the *xyr1* group for Th3844 (**e**). PL: Polysaccharide lyases; GH: Glycoside hydrolases; GT: Glycosyltransferases; CE: Carbohydrate esterases; Ta0020: *T. atroviride* CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844.

For the *xyr1* and *cre1* groups, the GH family was identified in all evaluated strains, with different numbers (**Fig. 5a, b**). These enzymes constitute the main catalytic machinery for the breakdown of glycosidic bonds between carbohydrate molecules [79]. Several GHs were also identified during mycoparasitism process whose activities can be related to fungal cell wall degradation in mycoparasites species of *Trichoderma* [80]. Most GHs related to mycoparasitic activity were associated with *cre1* in Ta0020 (**Fig. 5c**), which presented the highest number of classified transcripts from the GH family for the *cre1* group (**Fig. 5a**). Whereas Th0179 showed transcripts encoding GHs (**Fig. 5d**), Th3844 presented only one encoding a GH28 coexpressed

with *cre1*. For the *xyl1* group, this last strain presented a higher number of transcripts from the GH family (**Fig. 5e**) than did Th0179 (GH16 and GH76) and Ta0020 (GH75).

In addition to GHs, the CE family is also important for the efficient enzymatic hydrolysis of xylan [81] and was associated with the *cre1* transcript in Ta0020 (**Fig. 5c**) and Th0179 (**Fig. 5d**). The PL family, which cleaves certain activated glycosidic linkages in acidic polysaccharides [82], was coexpressed with the *xyl1* transcript in only Th3844 (**Fig. 5e**). In the *cre1* group, GT family, which includes enzymes that act in the synthesis of the cell wall of microorganisms by the transference of chemical groups [83], was present in only Th0179 (**Fig. 5a**), while in the *xyl1* group in Th3844 (**Fig. 5e**) and Ta0020 (GT22).

To obtain more information about the hydrolytic enzymes associated with CRE1 and XYR1, we looked out for transcripts encoding CAZymes in the *cre1* and *xyl1* subgroups. Within the *cre1* subgroup, Ta0020 and Th3844 did not show transcripts encoding CAZymes, but we identified GT90 (THAR02_09987), GH23 (THAR02_00630) and CE1 (THAR02_01920) transcripts in Th0179. In the *xyl1* subgroup, Ta0020 presented a GT22 (TRIATDRAFT_129991) and Th3844 presented GH18 (THAR02_04791), GH75 (THAR02_05160), PL8 (THAR02_00385) and GT (THAR02_10636) transcripts. Th0179 did not present transcripts encoding CAZymes in this subgroup.

Transcription Factors

Considering that the induction or repression of the genes encoding CAZymes, transporters and catabolic pathway enzymes needs the released monosaccharides and the action of many TFs [40, 47], we sought transcripts encoding these regulators coexpressed with *cre1* and *xyl1*. Zn₂Cys₆ was the most abundant coexpressed for the strains that showed transcripts encoding TFs in the *cre1* and *xyl1* groups. In the *cre1* group for Ta0020, the number of transcripts encoding Zn₂Cys₆ was higher than that observed for Th0179 (**Fig. 6a**). For the *xyl1* group, both Th3844 and Th0179 produced many more transcripts encoding Zn₂Cys₆ than Ta0020 did (**Fig. 6b**). Most Zn₂Cys₆ proteins also contain a fungal-specific TF domain, known as the middle homology region, that has a proposed role in regulation [84]. We found transcripts encoding this fungal domain for Ta0020 (**Fig. 6a**) and Th3844 (**Fig. 6b**) in the *cre1* and *xyl1* groups, respectively.

Transcripts encoding C₂H₂, a TF that was already described involved in plant cell wall depolymerization for filamentous fungi [46], were identified in the *cre1* group for Ta0020 (**Fig. 6a**) and in the *xyl1* group for *T. harzianum* strains (**Fig. 6b**). Furthermore, we found a transcript encoding the SteA regulator, which is present in nearly all filamentous ascomycetes associated

with the development and/or morphology of fungi [85], only in the *xyl1* group for Th0179 (**Fig. 6b**).

Considering that the gene regulation involved in plant biomass degradation is crucial, we also investigated transcripts encoding TFs in the *cre1* and *xyl1* subgroups. In the *cre1* subgroup, Ta0020 showed transcripts encoding Zn₂Cys₆ (TRIATDRAFT_182375), C₂H₂ (TRIATDRAFT_36276), and a fungal-specific TF domain (TRIATDRAFT_258080). For Th0179, we only found a transcript encoding a Zn₂Cys₆ (THAR02_10668). Within the *xyl1* subgroup, transcripts encoding Zn₂Cys₆ were found for Th0179 (THAR02_01165 and THAR02_07523) and for Th3844 (THAR02_02597), the latter of which also presented a C₂H₂ transcript (THAR02_02590).

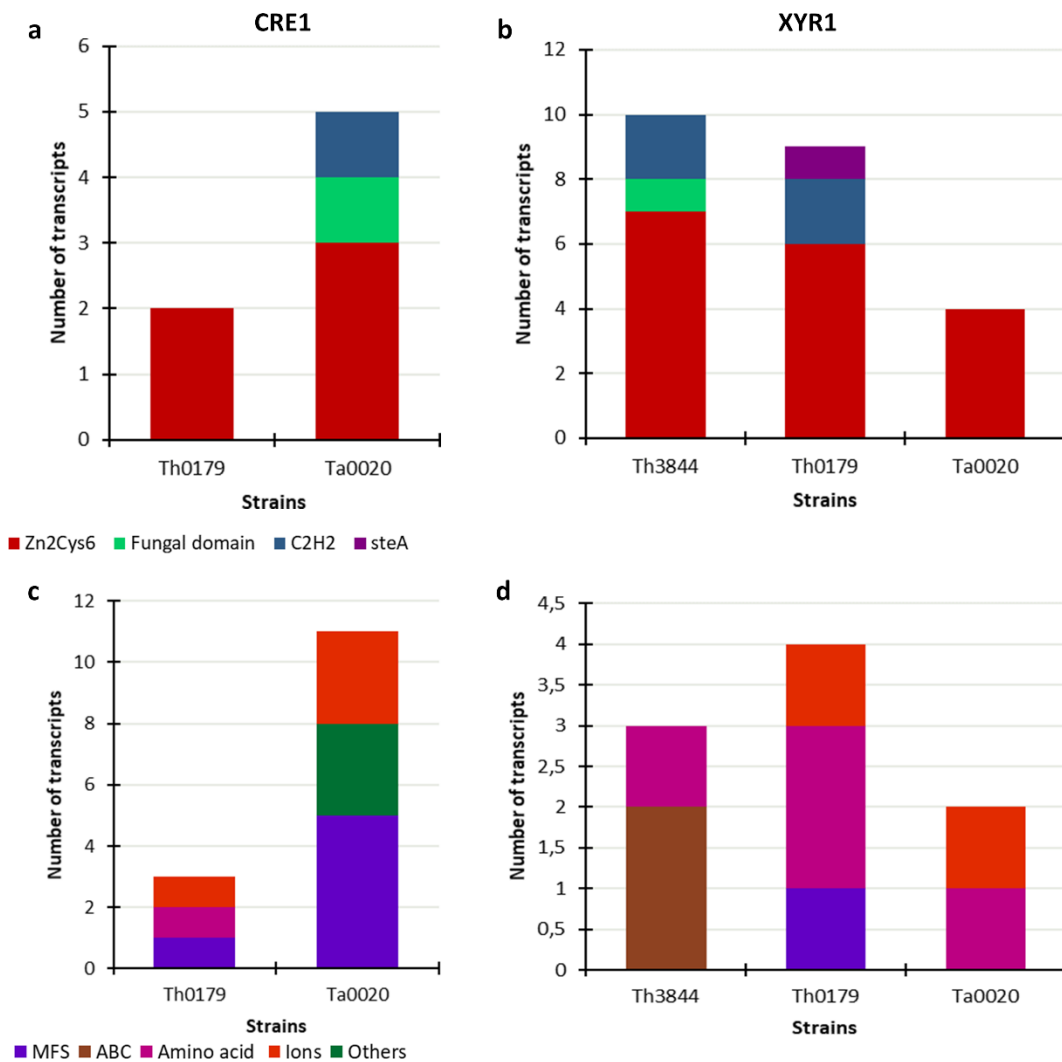


Fig. 6. Distribution of TFs and transporters within *cre1* and *xyl1* groups for *Trichoderma* spp. Classification of TFs coexpressed with *cre1* (a) and *xyl1* (b) and of transporters

coexpressed with *cre1* (c) and *xyl1* (d) for Th3844, Th0179 and Ta0020. Th3844: *T. harzianum* IOC-3844; Th0179: *T. harzianum* CBMAI-0179; Ta0020: *T. atroviride* CBMAI-0020.

Transporters

Transporters are also important players in the degradation of plant biomass. It is well known that approximately 5% of the *T. reesei* genome comprises genes that encode proteins involved in transport [66]. Thus, we sought transcripts encoding transport proteins in the *cre1* and *xyl1* groups. For the *cre1* group, Th3844 did not coexpress transcripts encoding transporters (**Fig. 6c**). On the other hand, this strain showed transcripts with a similar expression pattern as *xyl1* (**Fig. 6d**). These transcripts encode ABC transporters, which act as ATP-driven transporters for several substrates, including lipids, sugars, drugs, heavy metals, and auxin [49], and also amino acid transporters, which probably constitutes a major nutritional source for all fungi [86]. In addition, all strains presented transcripts encoding amino acid transporters coexpressed with *xyl1* in different amounts (**Fig. 6d**). For the *cre1* group, only Th0179 showed transcripts encoding this type of transport protein.

Another important class of protein transport involved in plant biomass degradation is the MFS permeases. In filamentous fungi, these proteins transport a wide range of substrates including small molecule inducers from the extracellular environment into the fungus, which could influence the expression of CAZyme-encoding genes [78]. In the *cre1* group, MFS are the most abundant proteins among the transporters for Ta0020 (**Fig. 6c**). For the *xyl1* group, only Th0179 showed transcripts encoding this type of protein transport (**Fig. 6d**).

Transcripts encoding ion transporters were present in the *cre1* group (**Fig. 6c**), including transcripts for zinc-regulated transporter 1 (TRIATDRAFT_147496) and a magnesium ion transporter (TRIATDRAFT_315729) for Ta0020 and for potassium transporter 8 (THAR02_02027) for Th0179. In the same way, transcripts encoding ion transporters were also present in the *xyl1* group (**Fig. 6d**), such as transcripts for a $\text{Cd}^{2+}\text{Zn}^{2+}$ transporter (TRIATDRAFT_320571) for Ta0020 and for a Ca^{2+} transporter (THAR02_01350) for Th0179.

In relation to the *cre1* subgroup, for Ta0020, we found transcripts encoding an MFS permease transporter (TRIATDRAFT_89305 and TRIATDRAFT_317997), whereas for Th0179, a transcript encoding a potassium transporter (THAR02_02027) was found. In the *xyl1* subgroup, transcripts encoded a $\text{Cd}^{2+}\text{Zn}^{2+}$ transporter (TRIATDRAFT_320571) for Ta0020, an amino acid transporter (THAR02_07655) and a Ca^{2+} transporter (THAR02_01350) for Th0179 and ABC transporters (THAR02_04735 and THAR02_04149) for Th3844.

Discussion

Due to the ability to mycoparasitize other fungi, *T. atroviride* and *T. harzianum* are widely used in agriculture as biocontrol agents against phytopathogens [87-89]. *T. harzianum* has also demonstrated considerable enzymatic potential for application in the lignocellulosic biomass hydrolysis [27, 90, 91]. However, the transcriptional regulation involved in the expression of CAZyme-encoding genes in this species is still poorly explored when compared to that of the major producer of cellulases, *T. reesei*. In this study, we chose to investigate and to compare the genetic mechanisms involved with XYR1- and CRE1-mediated gene regulation in *Trichoderma* spp. For this, we used transcriptome data [55] for Th3844 and Th0179, both mycoparasite strains with hydrolytic potential, and Ta0020, a common mycoparasite ancestor, to model the gene coexpression networks for each. We used this RNA-Seq dataset since a difference in gene expression among these three strains was established, showing that it was possible to capture differences among them [55].

The diversification observed for CRE1 and XYR1 inside Ascomycota may indicate a high functional diversity for these proteins, which can reflect changes in their functions [92]. Furthermore, for both regulators, a differentiation process was observed for Ta0020 compared to Th3844 and Th0179 among *Trichoderma* spp. These phylogenetic relationships may be related to the different types of enzymatic performance previously reported, in which Ta0020 had a lower cellulase activity profile during growth on cellulose compared to *T. harzianum* strains [55].

Analyzing the gene coexpression data using a clustering method allowed us to identify correlations among genes with similar expression patterns, which often indicates that they are active in the same biological processes [51, 52, 93]. Our gene coexpression networks analysis revealed that *T. harzianum* and *T. atroviride* have developed different strategies related to gene regulation involved in cellulose and hemicellulose degradation, which were already described for other fungi [41, 76, 94-96]. We suggest that since Th3844 and Th0179 had a larger number of transcripts associated with *xyl1* than with *cre1*, there were also more metabolic pathways related to XYR1, the main cellulase and hemicellulase activator. At a more specific level, the coexpression network topology of the subgroups exposed that the three strains presented different profiles of coexpressed transcripts with *xyl1* and *cre1*, which were more accentuated in *T. atroviride*.

Assuming that hubs are frequently more relevant to the functionality of networks than other nodes are [47], they were sought out within the *cre1* and *xyl1* groups, and new targets that were coexpressed with the main regulators involved in cellulose degradation were discovered.

Although transcripts encoding uncharacterized proteins were identified as potential hub nodes in the *xyl1* groups, indicating their importance in the degradation process, characterized proteins were found in the *cre1* groups. Transcripts encoding proteins related to proteasome, which is a highly sophisticated protease complex responsible for the degradation of the majority of intracellular proteins [97]; oxidative phosphorylation, which is responsible for mitochondrial ATP production related to the energy required for cell growth, differentiation, and development [98]; and nucleoside import into fungal cells, which can be associated with nucleotide synthesis via salvage pathways in cells lacking the capacity for *de novo* synthesis [99] were found to be hubs in Th3844, Th0179 and Ta0020, respectively.

For *Trichoderma* spp., GO enrichment in the *cre1* group showed that terms related to lignocellulose breakdown and to regulation of gene expression were the main enriched terms. For the *xyl1* group, all these terms or correlated terms were also enriched, except for hydrolase activity, which was present in only Th3844. In a recent study, it was reported that several terms related to lignocellulose breakdown were enriched in the *xyl1* group for *T. reesei* RUT-C30 grown on bagasse or fructose as a carbon source [53]. Furthermore, XYR1 and CRE1 are zinc finger proteins, which encompass a wide variety of compact protein domains stabilized by a structural zinc ion [100]. Several classes of zinc finger motifs are present in TFs and act as parts of DNA-binding and protein-protein interaction domains [101]. In the *cre1* and *xyl1* groups, we found transcripts enriched with metal ion binding and zinc ion binding, respectively. In this way, we can suppose that transcripts encoding zinc finger proteins were coexpressed with them, as we found.

To obtain more information about the metabolic pathways triggered by CRE1 and XYR1 in *Trichoderma* spp., we investigated the enzymatic activity in the *cre1* and *xyl1* groups. Our data showed high similarity between metabolic pathways triggered by both regulators during cellulose degradation. However, comparing Ta0020 and Th0179, they had distinct metabolic processes occurring within *cre1* and *xyl1* groups, which suggests that these two TFs have developed different functional regulation in each strain. In addition, the exclusive pathway classes for Ta0020 were not related to the biomass degradation process. For instance, the metabolism of terpenoids and polyketides, two secondary metabolite compounds involved in the antibiosis process during mycoparasitic activity [102, 103], was identified.

We also investigated the expression profile of the transcripts within *cre1* and *xyl1* groups, including *xyl1* and *cre1*, for all evaluated strains under cellulose and glucose growth. In 2016, Castro et al. (2016) observed that in the presence of glucose, the expression level of the transcript *xyl1* for *T. reesei* was minimal [76]. Our findings did not show this for Ta0020.

In this case, the *xyl1* expression level was higher under glucose growth than under cellulose growth. In addition, *cre1* was upregulated in the presence of cellulose, which was not expected due to the role of CRE1 in the repression of cellulolytic and hemicellulolytic enzymes when an easily metabolizable sugar, i.e., glucose, is available in the environment [46]. Conversely, the *cre1* expression level was higher in the presence of glucose than in the presence of cellulose for Th3844. Upregulated transcripts within the *cre1* group for Ta0020 were also not expected. Here it is important to mention that the gene coexpression networks were modeled based on the transcript expression level under two sets of conditions. Considering that expression of the *cre1* transcript for Ta0020 was upregulated under cellulose growth, the expression of other transcripts into *cre1* group was also expected to be upregulated under these conditions, which we found.

We also investigated the distribution of transcripts encoding CAZymes, TFs, and transporters within the *cre1* and *xyl1* groups among *Trichoderma* spp. Initially, it was believed that CRE1 directly repressed the transcription of cellulase and hemicellulase genes when a readily metabolizable carbon source was present in the medium [43]. While this remains correct, studies in *T. reesei* have demonstrated that CRE1 may also control genes related to several other functions, and only a small number of CAZyme genes are direct targets of this TF during CCR [77, 104, 105]. In this way, we did not expect to find transcripts encoding CAZymes in *cre1* groups. Herein, we had only a few transcripts encoding CAZymes in *cre1* groups for Th3844 and Th0179, whereas others encode proteins with diverse functions. In relation to TFs and transporters, Antoniêto et al. (2014) observed that transcripts encoding these proteins were strongly repressed by CRE1 [77] in *T. reesei*. Thus, we expected to find a large number of transcripts encoding TFs and transporters coexpressed with the *cre1* transcript for all strains, but it was observed only for Ta0020, especially for the transcripts encoding proteins related to transport.

Castro et al. (2016) demonstrated for *T. reesei* that the expression of CAZymes, TFs, and transporters were affected by the TF XYR1 in the context of plant biomass degradation [76]. Furthermore, Reithner et al. (2014) have suggested that XYR1 was also involved in the mediation of the plant-fungus recognition processes in *T. atroviride* [45]. It was also reported that the role of XYR1 differs strongly among ascomycete fungi [44]. We found transcripts encoding CAZymes, TFs and transporters coexpressed with *xyl1* for the three analyzed strains. Although these transcripts were present in Ta0020, they may induce different mechanisms there than in Th3844 and Th0179.

Our results showed that *T. atroviride* has an increased number of GHs coexpressed with *cre1* that are related to mycoparasitism and not to cellulose and hemicellulose degradation. For instance, GH18, whose members have chitinase activity, is the largest family of CAZymes in mycoparasitic *Trichoderma* species [80], whereas GH28, which is the main family involved in the degradation of pectin [106], is also involved in mycoparasitism. Moreover, β -glucanases have been described for mycoparasitic species within *Trichoderma* spp. in GH families 55, 64 and 81 [58]. Among the CAZymes responsible for cellulose degradation, we found transcripts encoding only GH5 and GH12, and for hemicellulose degradation, we found transcripts for only CE5, GH95 and GH43 [31].

Almeida et al. (2020) analyzed the main CAZyme classes detected in the transcriptome of Th3844, Th0179 and Ta0020 under cellulose or glucose growth conditions. They found the main CAZyme families responsible for cellulose and hemicellulose degradation for all analyzed strains [55]. Herein, the transcripts encoding cellulases and hemicellulases were not found in the *cre1* and *xyl1* groups for *T. harzianum*. It is possible that the transcripts encoding these hydrolytic enzymes were grouped in other formed groups within the coexpression networks, mainly because their expression patterns were different from that observed for the transcripts encoding XYR1 and CRE1.

The production of cellulases and hemicellulases in *T. reesei* occurs by activation of gene expression via specific inducers in balance with the repression of gene expression via CCR [47]. The sensing of inducers starts a signaling pathway resulting in the activation of TFs, which is followed by the production of plant biomass-degrading enzymes and the activation of metabolic pathways to use the accumulating sugar monomers [46, 107]. Transporters also have an important role in the utilization of lignocellulose by fungi, promoting the transport of molecules between the intracellular and extracellular spaces and, consequently, controlling CAZyme-encoding gene expression [108]. For *T. reesei*, MFS, amino acids and ABC transporters were modulated by XYR1 and CRE1 [76, 77, 104]. Our results suggest that the expression of transporters and TFs during cellulose degradation was affected by XYR1 in *T. harzianum*, as it was demonstrated for *T. reesei* [76]. In addition, *T. harzianum* and *T. atroviride* developed different strategies related to the import of molecules from the environment that can influence the production of plant cell wall-degrading enzymes, which may result in contrasting enzymatic performances during cellulose degradation.

Conclusions

Our results propose that the coexpressed transcripts with *xyl1* and *cre1* are different for each strain and suggest that different strategies related to XYR1 and CRE1 are used by *Trichoderma* spp. during cellulose degradation and glucose metabolization. It was observed that the transcripts encoding CAZymes related to cellulose and hemicellulose degradation in *T. harzianum* are not necessarily coexpressed with the TFs studied herein. Taken together, our findings provide new insights into transcripts coexpressed with XYR1 and CRE1 in *T. harzianum* strains, which could be the target of new studies to evaluate their role in lignocellulose degradation. Finally, our results can contribute to a better understanding of the genetic mechanisms involved in hydrolytic enzyme regulation for mycoparasite *Trichoderma* species. This will be important to expand the potential of *T. harzianum* not only in the development of better enzyme cocktails but also in many other industrial applications.

Additional files

Additional file 1: Table S1. Description of the species used for the phylogenetic analysis of the transcription factors CRE1 and XYR1 in Ascomycota.

Additional file 2: Table S2. Filtered transcripts used to model the gene coexpression networks among *Trichoderma* spp.

Additional file 3: Table S3. Transcripts identified in the cluster analysis of gene coexpression networks of *Trichoderma* spp.

Additional file 4: Table S4. Transcripts identified as coexpressed with *cre1* and *xyl1* in gene coexpression networks of *Trichoderma* spp.

Additional file 5: Table S5. Subgroups identified among *cre1* and *xyl1* groups using the MCODE methodology for all evaluated strains.

Additional file 6: Table S6. Subgroups identified among the *cre1* and *xyl1* groups using the Louvain algorithm for all evaluated strains.

Additional file 7: Table S7. Number of transcripts coexpressed in each subgroup formed from *cre1* and *xyl1* groups for *Trichoderma* spp. and its corresponding color. **Fig. S1.** Venn diagrams of pathway classes shared between *Trichoderma* spp. with expression levels associated with *cre1* (a) and *xyl1* (b) transcript levels. **Table 10.** Expression values in TPM of *cre1* and *xyl1* for *Trichoderma* spp. under cellulose and glucose growth conditions.

Additional file 8: Table S8. Hubs identified within the *cre1* and *xyl1* groups of *Trichoderma* spp.

Additional file 9: Table S9. Differentially expressed transcripts established in the *cre1* and *xyl1* groups for all evaluated strains.

List of Abbreviations

1G ethanol: First-Generation Ethanol; **2G ethanol:** Second-Generation Ethanol; **ABC:** ATP-binding cassette; **CAZymes:** Carbohydrate-Active Enzymes; **CBMAI:** Brazilian Collection of Environment and Industry Microorganisms; **CCR:** Carbon Catabolite Repression; **CEs:** Carbohydrate Esterases; **CNT:** Concentrative Nucleoside Transporter; **CPQBA:** Chemical, Biological and Agricultural Pluridisciplinary Research Center; **CRE1:** Carbon Catabolite Repressor 1; **GHs:** Glycoside Hydrolases; **GO:** Gene Ontology; **GTs:** Glycosyltransferases; **JTT:** Jones-Taylor-Thornton; **KEGG:** Kyoto Encyclopedia of Genes and Genomes; **KO:** KEGG Orthology; **LPMOs:** Lytic Polysaccharides Monooxygenases; **MCODE:** Molecular Complex Detection; **MEGA:** Molecular Evolutionary Genetics Analysis; **MFS:** Major Facilitator Superfamily; **MRP:** Mitochondrial Ribosomal Protein; **ML:** Maximum Likelihood; **PLs:** Polysaccharide Lyases; **RNA-Seq:** RNA Sequencing; **Ta0020:** *Trichoderma atroviride* CBMAI-0020; **TFs:** Transcription Factors; **Th0179:** *Trichoderma harzianum* CBMAI-0179; **Th3844:** *Trichoderma harzianum* IOC-3844; **TPM:** Transcripts Per Million; **UNICAMP:** University of Campinas; **WGCNA:** Weighted Gene Coexpression Network Approach; **XYR1:** Xylanase Regulator 1.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analyzed during this study are included in this published article as Additional files 1, 2, 3, 4, 5, 6, 7, 8 and 9.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2015/09202-0, 2018/19660-4), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Computational Biology Program and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). RRR received a partial MS fellowship from the CAPES Computational Biology Program (88887.176241/2018-00) and partial support from CAPES – PROEX (88882.329483/2019-01); AHA received a PhD from the FAPESP (2019/03232-6); DAA received a partial MS fellowship from the FAPESP (2017/17782-2) and partial support from the CAPES Computational Biology Program (88887.336686/2019-00); JAFF received a PhD fellowship from the CNPq (170565/2017-3) and a PhD fellowship from the CAPES (CBP – 88887.334235/2019-00); MACH received a PD fellowship from FAPESP (2018/18856-1); and APS is the recipient of a research fellowship from the CNPq.

Author contributions

RRR and APS conceived and designed the study. JAFF contributed to the study performance. RRR, AHA and DAA performed the data analysis. MACH designed and achieved the fermentation experiments. RRR and AHA drafted the manuscript, which was critically revised by DAA, JAFF and APS. All authors read and approved the final manuscript.

Acknowledgments

We would like to acknowledge funding from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2015/09202-0 and 2018/19660-4), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Computational Biology Program) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We thank the Brazilian Biorenewables National Laboratory (LNBR), Campinas – SP, for conducting the fermentation experiments and the Center of Molecular Biology and Genetic Engineering (CBMEG) at the University of Campinas, SP, for the center and laboratory space.

References

1. Gupta A, Verma JP: **Sustainable bio-ethanol production from agro-residues: A review.** *Renewable and Sustainable Energy Reviews* 2015, **41**:550-567.
2. Kucharska K, Słupek E, Cieśliński H, Kamiński M: **Advantageous conditions of saccharification of lignocellulosic biomass for biofuels generation via fermentation processes.** *Chemical Papers* 2019.
3. Sharma HK, Xu C, Qin W: **Biological Pretreatment of Lignocellulosic Biomass for Biofuels and Bioproducts: An Overview.** *Waste and Biomass Valorization* 2019, **10**:235-251.
4. Association RF: **Renewable Fuels Association.** 2019.
5. Alonso Pippo W, Luengo CA, Alonsoamador Morales Alberteris L, Garzone P, Cornacchia G: **Energy Recovery from Sugarcane-Trash in the Light of 2nd Generation Biofuels. Part 1: Current Situation and Environmental Aspects.** *Waste and Biomass Valorization* 2011, **2**:1-16.
6. Sun JX, Sun XF, Sun RC, Su YQ: **Fractional extraction and structural characterization of sugarcane bagasse hemicelluloses.** *Carbohydrate Polymers* 2004, **56**:195-204.
7. Méndez J, Passos DdF, Wischral D, Modesto LF, Pereira N: **Second-generation ethanol production by separate hydrolysis and fermentation from sugarcane bagasse with cellulose hydrolysis using a customized enzyme cocktail.** *Biofuels* 2019:1-7.
8. Robak K, Balcerek M: **Review of Second Generation Bioethanol Production from Residual Biomass.** *Food technology and biotechnology* 2018, **56**:174-187.
9. Ruchala J, Kurylenko OO, Dmytruk KV, Sibirny AA: **Construction of advanced producers of first- and second-generation ethanol in *Saccharomyces cerevisiae* and selected species of non-conventional yeasts (*Scheffersomyces stipitis*, *Ogataea polymorpha*).** *Journal of Industrial Microbiology & Biotechnology* 2019.
10. Ferreira JA, Brancoli P, Agnihotri S, Bolton K, Taherzadeh MJ: **A review of integration strategies of lignocelluloses and other wastes in 1st generation bioethanol processes.** *Process Biochemistry* 2018, **75**:173-186.
11. Alvira P, Negro MJ, Ballesteros M: **Effect of endoxylanase and α -l-arabinofuranosidase supplementation on the enzymatic hydrolysis of steam exploded wheat straw.** *Bioresource Technology* 2011, **102**:4552-4558.

12. Le Costaouëc T, Pakarinen A, Várnai A, Puranen T, Viikari L: **The role of carbohydrate binding module (CBM) at high substrate consistency: Comparison of *Trichoderma reesei* and *Thermoascus aurantiacus* Cel7A (CBHI) and Cel5A (EGII).** *Bioresource Technology* 2013, **143**:196-203.
13. Bomble YJ, Lin C-Y, Amore A, Wei H, Holwerda EK, Ciesielski PN, Donohoe BS, Decker SR, Lynd LR, Himmel ME: **Lignocellulose deconstruction in the biosphere.** *Current Opinion in Chemical Biology* 2017, **41**:61-70.
14. Sartori T, Tibolla H, Prigo E, Colla LM, Costa JAV, Bertolin TE: **Enzymatic saccharification of lignocellulosic residues by cellulases obtained from solid state fermentation using *Trichoderma viride*.** *BioMed Research International* 2015.
15. Peciulyte A, Pisano M, de Vries RP, Olsson L: **Hydrolytic potential of five fungal supernatants to enhance a commercial enzyme cocktail.** *Biotechnology letters* 2017, **39**:1403-1411.
16. Zafra G, Cortés-Espinosa DV: **Biodegradation of polycyclic aromatic hydrocarbons by *Trichoderma* species: a mini review.** *Environmental Science and Pollution Research* 2015, **22**:19426-19433.
17. Bischof RH, Ramoni J, Seiboth B: **Cellulases and beyond: the first 70 years of the enzyme producer *Trichoderma reesei*.** *Microbial Cell Factories* 2016, **15**:106.
18. Sridevi A, Ramanjaneyulu G, Suvarnalatha Devi P: **Biobleaching of paper pulp with xylanase produced by *Trichoderma asperellum*.** *3 Biotech* 2017, **7**:266-266.
19. Kumar M, Ashraf S: **Role of *Trichoderma* spp. as a Biocontrol Agent of Fungal Plant Pathogens.** In *Probiotics and Plant Health*. Edited by Kumar V, Kumar M, Sharma S, Prasad R. Singapore: Springer Singapore; 2017: 497-506
20. Peterson R, Nevalainen H: ***Trichoderma reesei* RUT-C30 – thirty years of strain improvement.** *Microbiology* 2012, **158**:58-68.
21. Mukherjee AK, Sampath Kumar A, Kranthi S, Mukherjee PK: **Biocontrol potential of three novel *Trichoderma* strains: isolation, evaluation and formulation.** *3 Biotech* 2014, **4**:275-281.
22. Atanasova L, Crom SL, Gruber S, Coulpier F, Seidl-Seiboth V, Kubicek CP, Druzhinina IS: **Comparative transcriptomics reveals different strategies of *Trichoderma* mycoparasitism.** *BMC Genomics* 2013, **14**:121.
23. Druzhinina IS, Kopchinskiy AG, Kubicek CP: **The first 100 *Trichoderma* species characterized by molecular data.** *Mycoscience* 2006, **47**:55.

24. Chaverri P, Samuels GJ, Stewart EL: **Hypocrea virens sp. nov., the Teleomorph of Trichoderma virens.** *Mycologia* 2001, **93**:1113-1124.
25. Dodd SL, Lieckfeldt E, Samuels GJ: **Hypocrea atroviridis sp. nov., the teleomorph of Trichoderma atroviride.** *Mycologia* 2003, **95**:27-40.
26. Druzhinina IS, Kubicek CP, Komoń-Zelazowska M, Mulaw TB, Bissett J: **The Trichoderma harzianum demon: complex speciation history resulting in coexistence of hypothetical biological species, recent agamospecies and numerous relict lineages.** *BMC Evolutionary Biology* 2010, **10**:94.
27. Horta MAC, Vicentini R, Delabona PdS, Laborda P, Crucello A, Freitas S, Kuroshu RM, Polikarpov I, Pradella JGdC, Souza AP: **Transcriptome profile of Trichoderma harzianum IOC-3844 induced by sugarcane bagasse.** *PloS one* 2014, **9**:e88689-e88689.
28. Li J-X, Zhang F, Li J, Zhang Z, Bai F-W, Chen J, Zhao X-Q: **Rapid production of lignocellulolytic enzymes by Trichoderma harzianum LZ117 isolated from Tibet for biomass degradation.** *Bioresource technology* 2019, **292**:122063.
29. Santos CA, Ferreira-Filho JA, O'Donovan A, Gupta VK, Tuohy MG, Souza AP: **Production of a recombinant swollenin from Trichoderma harzianum in Escherichia coli and its potential synergistic role in biomass degradation.** *Microbial Cell Factories* 2017, **16**:83.
30. Santos CA, Morais MAB, Terrett OM, Lyczakowski JJ, Zanthorlin LM, Ferreira-Filho JA, Tonoli CCC, Murakami MT, Dupree P, Souza AP: **An engineered GH1 β -glucosidase displays enhanced glucose tolerance and increased sugar release from lignocellulosic materials.** *Scientific Reports* 2019, **9**:4903.
31. Ferreira Filho JA, Horta MAC, Beloti LL, dos Santos CA, de Souza AP: **Carbohydrate-active enzymes in Trichoderma harzianum: a bioinformatic analysis bioprospecting for key enzymes for the biofuels industry.** *BMC Genomics* 2017, **18**:779.
32. Crucello A, Sforça DA, Horta MAC, dos Santos CA, Viana AJC, Beloti LL, de Toledo MAS, Vincentz M, Kuroshu RM, de Souza AP: **Analysis of Genomic Regions of Trichoderma harzianum IOC-3844 Related to Biomass Degradation.** *PLOS ONE* 2015, **10**:e0122122.
33. Binod P, Gnansounou E, Sindhu R, Pandey A: **Enzymes for second generation biofuels: Recent developments and future perspectives.** *Bioresource Technology Reports* 2019, **5**:317-325.

34. Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen J-CN, Johansen KS, Krogh KBRM, Jørgensen CI, Tovborg M, Anthonsen A, et al: **Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**:15079-15084.
35. Saloheimo M, Paloheimo M, Hakola S, Pere J, Swanson B, Nyysönen E, Bhatia A, Ward M, Penttilä M: **Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials.** *European Journal of Biochemistry* 2002, **269**:4202-4211.
36. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B: **The carbohydrate-active enzymes database (CAZy) in 2013.** *Nucleic acids research* 2014, **42**:D490-D495.
37. Castro LdS, Antoniêto ACC, 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 Expression Patterns* 2014, **14**:88-95.
38. Amore A, Giacobbe S, Faraco V: **Regulation of cellulase and hemicellulase gene expression in fungi.** *Current genomics* 2013, **14**:230-249.
39. 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* 2011, **2**:274-286.
40. 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.** *Eukaryotic cell* 2011, **10**:262-271.
41. 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*.** *Eukaryotic cell* 2006, **5**:2128-2137.
42. Portnoy T, Margeot A, Linke R, Atanasova L, Fekete E, Sándor 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**:269-269.

43. Strauss J, Mach RL, Zeilinger S, Hartler G, Stöffler G, Wolschek M, Kubicek CP: **Crel, the carbon catabolite repressor protein from *Trichoderma reesei***. *FEBS Letters* 1995, **376**:103-107.
44. Klaubauf S, Narang HM, Post H, Zhou M, Brunner K, Mach-Aigner AR, Mach RL, Heck AJR, Altelaar AFM, de Vries RP: **Similar is not the same: Differences in the function of the (hemi-)cellulolytic regulator XlnR (Xlr1/Xyr1) in filamentous fungi**. *Fungal Genetics and Biology* 2014, **72**:73-81.
45. Reithner B, Mach-Aigner AR, Herrera-Estrella A, Mach RL: ***Trichoderma atroviride* transcriptional regulator Xyr1 supports the induction of systemic resistance in plants**. *Applied and environmental microbiology* 2014, **80**:5274-5281.
46. Benocci T, Aguilar-Pontes MV, Zhou M, Seiboth B, de Vries RP: **Regulators of plant biomass degradation in ascomycetous fungi**. *Biotechnology for biofuels* 2017, **10**:152-152.
47. Alazi E, Ram AFJ: **Modulating Transcriptional Regulation of Plant Biomass Degrading Enzyme Networks for Rational Design of Industrial Fungal Strains**. *Frontiers in bioengineering and biotechnology* 2018, **6**:133-133.
48. Shida Y, Furukawa T, Ogasawara W: **Deciphering the molecular mechanisms behind cellulase production in *Trichoderma reesei*, the hyper-cellulolytic filamentous fungus**. *Bioscience, Biotechnology, and Biochemistry* 2016, **80**:1712-1729.
49. de Paula RG, Antoniêto ACC, Ribeiro LFC, Carraro CB, Nogueira KMV, Lopes DCB, Silva AC, Zerbini MT, Pedersoli WR, Costa MdN, Silva RN: **New Genomic Approaches to Enhance Biomass Degradation by the Industrial Fungus *Trichoderma reesei***. *International Journal of Genomics* 2018, **2018**:17.
50. Zhang F, Bunterngsook B, Li J-X, Zhao X-Q, Champreda V, Liu C-G, Bai F-W: **Chapter Three - Regulation and production of lignocellulolytic enzymes from *Trichoderma reesei* for biofuels production**. In *Advances in Bioenergy. Volume 4*. Edited by Li Y, Ge X: Elsevier; 2019: 79-119
51. van Dam S, Vösa U, van der Graaf A, Franke L, de Magalhães JP: **Gene co-expression analysis for functional classification and gene-disease predictions**. *Briefings in bioinformatics* 2018, **19**:575-592.
52. Pirim H: **Construction of Gene Networks Using Expression Profiles**. In *Soft Computing for Biological Systems*. Edited by Purohit HJ, Kalia VC, More RP. Singapore: Springer Singapore; 2018: 67-89

53. Borin GP, Carazzolle MF, dos Santos RAC, Riaño-Pachón DM, Oliveira JVdC: **Gene Co-expression Network Reveals Potential New Genes Related to Sugarcane Bagasse Degradation in *Trichoderma reesei* RUT-30.** *Frontiers in Bioengineering and Biotechnology* 2018, **6**.
54. Horta MAC, Filho JAF, Murad NF, de Oliveira Santos E, dos Santos CA, Mendes JS, Brandão MM, Azzoni SF, de Souza AP: **Network of proteins, enzymes and genes linked to biomass degradation shared by *Trichoderma* species.** *Scientific Reports* 2018, **8**:1341.
55. Almeida DA, Horta MAC, Ferreira Filho JA, Murad NF, de Souza AP: **The synergistic actions of hydrolytic genes in coexpression networks reveal the potential of *Trichoderma harzianum* for cellulose degradation.** *bioRxiv* 2020:2020.2001.2014.906529.
56. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics* 2014, **30**:2114-2120.
57. Baroncelli R, Piaggieschi G, Fiorini L, Bertolini E, Zapparata A, Pè ME, Sarrocco S, Vannacci G: **Draft Whole-Genome Sequence of the Biocontrol Agent *Trichoderma harzianum* T6776.** *Genome announcements* 2015, **3**:e00647-00615.
58. Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, et al: **Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*.** *Genome Biology* 2011, **12**:R40.
59. CLC: **Genomics Workbench (QIAGEN).** 6.5.2 edition. Aarhus, Denmark; 2016.
60. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic acids research* 1994, **22**:4673-4680.
61. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.** *Molecular Biology and Evolution* 2016, **33**:1870-1874.
62. Jones DT, Taylor WR, Thornton JM: **The rapid generation of mutation data matrices from protein sequences.** *Bioinformatics* 1992, **8**:275-282.
63. Felsenstein J: **Confidence Limits on Phylogenies: An Approach Using the Bootstrap.** *Evolution* 1985, **39**:783-791.

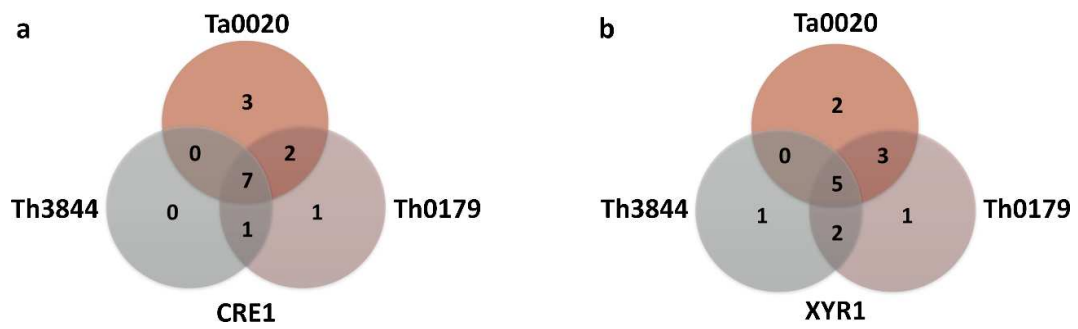
64. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC Bioinformatics* 2008, **9**:559.
65. Team RC: **R: A language and environment for statistical computing.** In *R Foundation for Statistical Computing*. Vienna, Austria; 2018.
66. Langfelder P, Zhang B, Horvath S: **Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R.** *Bioinformatics* 2007, **24**:719-720.
67. Sanner MF: **Python: a programming language for software integration and development.** *Journal of molecular graphics & modelling* 1999, **17**:57-61.
68. Hagberg A, Swart P, S Chult D: **Exploring network structure, dynamics, and function using networkx.** In; 2008-01-01; *United States*. 2008
69. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E: **Fast unfolding of communities in large networks.** *Journal of Statistical Mechanics: Theory and Experiment* 2008, **2008**:P10008.
70. Bader GD, Hogue CWV: **An automated method for finding molecular complexes in large protein interaction networks.** *BMC Bioinformatics* 2003, **4**:2.
71. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: **Gene Ontology: tool for the unification of biology.** *Nature Genetics* 2000, **25**:25-29.
72. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic acids research* 2000, **28**:27-30.
73. Supek F, Bošnjak M, Škunca N, Šmuc T: **REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms.** *PLOS ONE* 2011, **6**:e21800.
74. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL: **Biopython: freely available Python tools for computational molecular biology and bioinformatics.** *Bioinformatics* 2009, **25**:1422-1423.
75. Mukaka MM: **Statistics corner: A guide to appropriate use of correlation coefficient in medical research.** *Malawi medical journal : the journal of Medical Association of Malawi* 2012, **24**:69-71.
76. dos Santos Castro L, de Paula RG, Antoniêto ACC, Persinoti GF, Silva-Rocha R, Silva RN: **Understanding the Role of the Master Regulator XYR1 in *Trichoderma reesei* by Global Transcriptional Analysis.** *Frontiers in Microbiology* 2016, **7**.
77. Antoniêto ACC, dos Santos Castro L, Silva-Rocha R, Persinoti GF, Silva RN: **Defining the genome-wide role of CRE1 during carbon catabolite repression in**

- Trichoderma reesei using RNA-Seq analysis.** *Fungal Genetics and Biology* 2014, **73**:93-103.
78. Antoniêto ACC, de Paula RG, Castro LDS, Silva-Rocha R, Persinoti GF, Silva RN: **Trichoderma reesei CRE1-mediated Carbon Catabolite Repression in Response to Sophorose Through RNA Sequencing Analysis.** *Current genomics* 2016, **17**:119-131.
 79. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics.** *Nucleic acids research* 2009, **37**:D233-D238.
 80. Gruber S, Seidl-Seiboth V: **Self versus non-self: fungal cell wall degradation in Trichoderma.** *Microbiology* 2012, **158**:26-34.
 81. Malgas S, Mafa MS, Mkabayi L, Pletschke BI: **A mini review of xylanolytic enzymes with regards to their synergistic interactions during hetero-xylan degradation.** *World Journal of Microbiology and Biotechnology* 2019, **35**:187.
 82. Linhardt RJ, Galliher PM, Cooney CL: **Polysaccharide lyases.** *Applied Biochemistry and Biotechnology* 1987, **12**:135-176.
 83. Ardèvol A, Rovira C: **Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations.** *Journal of the American Chemical Society* 2015, **137**:7528-7547.
 84. MacPherson S, Laroche M, Turcotte B: **A fungal family of transcriptional regulators: the zinc cluster proteins.** *Microbiology and molecular biology reviews : MMBR* 2006, **70**:583-604.
 85. Todd RB, Zhou M, Ohm RA, Leeggangers HACF, Visser L, de Vries RP: **Prevalence of transcription factors in ascomycete and basidiomycete fungi.** *BMC Genomics* 2014, **15**:214.
 86. Gournas C, Prévost M, Krammer E-M, André B: **Function and Regulation of Fungal Amino Acid Transporters: Insights from Predicted Structure.** In *Yeast Membrane Transport*. Edited by Ramos J, Sychrová H, Kschischo M. Cham: Springer International Publishing; 2016: 69-106
 87. Brunner K, Zeilinger S, Ciliento R, Woo SL, Lorito M, Kubicek CP, Mach RL: **Improvement of the fungal biocontrol agent Trichoderma atroviride to enhance both antagonism and induction of plant systemic disease resistance.** *Applied and environmental microbiology* 2005, **71**:3959-3965.

88. Abdel-Fattah GM, Shabana YM, Ismail AE, Rashad YM: **Trichoderma harzianum: a biocontrol agent against Bipolaris oryzae.** *Mycopathologia* 2007, **164**:81-89.
89. Adnan M, Islam W, Shabbir A, Khan KA, Ghramh HA, Huang Z, Chen HYH, Lu G-d: **Plant defense against fungal pathogens by antagonistic fungi with Trichoderma in focus.** *Microbial Pathogenesis* 2019, **129**:7-18.
90. Delabona PdS, Farinas CS, da Silva MR, Azzoni SF, Pradella JGdC: **Use of a new Trichoderma harzianum strain isolated from the Amazon rainforest with pretreated sugar cane bagasse for on-site cellulase production.** *Bioresource Technology* 2012, **107**:517-521.
91. da Silva Delabona P, Rodrigues GN, Zubieta MP, Ramoni J, Codima CA, Lima DJ, Farinas CS, da Cruz Pradella JG, Seiboth B: **The relation between xyr1 overexpression in Trichoderma harzianum and sugarcane bagasse saccharification performance.** *Journal of Biotechnology* 2017, **246**:24-32.
92. Te Velthuis AJW, Bagowski CP: **Linking fold, function and phylogeny: a comparative genomics view on protein (domain) evolution.** *Current genomics* 2008, **9**:88-96.
93. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**:14863-14868.
94. van Peij NN, Gielkens MM, de Vries RP, Visser J, de Graaff LH: **The transcriptional activator XlnR regulates both xylanolytic and endoglucanase gene expression in Aspergillus niger.** *Applied and environmental microbiology* 1998, **64**:3615-3619.
95. Kowalczyk JE, Benoit I, de Vries RP: **Chapter Two - Regulation of Plant Biomass Utilization in Aspergillus.** In *Advances in Applied Microbiology. Volume 88.* Edited by Sariaslani S, Gadd GM: Academic Press; 2014: 31-56
96. Sun J, Tian C, Diamond S, Glass NL: **Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in Neurospora crassa.** *Eukaryotic cell* 2012, **11**:482-493.
97. Tanaka K: **The proteasome: overview of structure and functions.** *Proceedings of the Japan Academy Series B, Physical and biological sciences* 2009, **85**:12-36.
98. Sylvester JE, Fischel-Ghodsian N, Mougey EB, O'Brien TW: **Mitochondrial ribosomal proteins: Candidate genes for mitochondrial disease.** *Genetics in Medicine* 2004, **6**:73-80.

99. Gray JH, Owen RP, Giacomini KM: **The concentrative nucleoside transporter family, SLC28.** *Pflügers Archiv* 2004, **447**:728-734.
100. Hanas JS, Larabee JL, Hocker JR: **Zinc Finger Interactions with Metals and Other Small Molecules.** In *Zinc Finger Proteins: From Atomic Contact to Cellular Function*. Edited by Iuchi S, Kuldell N. Boston, MA: Springer US; 2005: 39-46
101. Takatsuji H: **Zinc-finger transcription factors in plants.** *Cellular and Molecular Life Sciences CMLS* 1998, **54**:582-596.
102. Sivasithamparam KG, Emilio.: **Secondary metabolism in Trichoderma and Gliocladium.** In *Trichoderma and Gliocladium. Volume 1*. Edited by Harman. CPKGE. London, UK; 1998: pp. 139-191
103. Mukherjee PK, Nautiyal CS, Mukhopadhyay AN: **Molecular Mechanisms of Biocontrol by Trichoderma spp.** In *Molecular Mechanisms of Plant and Microbe Coexistence*. Edited by Nautiyal CS, Dion P. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008: 243-262
104. Antoniêto ACC, de Paula RG, Castro LDS, Silva-Rocha R, Persinoti GF, Silva RN: **Trichoderma reesei CRE1-mediated Carbon Catabolite Repression in Response to Sophorose Through RNA Sequencing Analysis.** *Current genomics* 2016, **17**:119-131.
105. Portnoy T, Margeot A, Linke R, Atanasova L, Fekete E, Sándor 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.
106. Martens-Uzunova ES, Schaap PJ: **Assessment of the pectin degrading enzyme network of Aspergillus niger by functional genomics.** *Fungal Genetics and Biology* 2009, **46**:S170-S179.
107. de Vries RP, Visser J: **Aspergillus enzymes involved in degradation of plant cell wall polysaccharides.** *Microbiology and molecular biology reviews : MMBR* 2001, **65**:497-522.
108. Nogueira KMV, de Paula RG, Antoniêto ACC, dos Reis TF, Carraro CB, Silva AC, Almeida F, Rechia CGV, Goldman GH, Silva RN: **Characterization of a novel sugar transporter involved in sugarcane bagasse degradation in Trichoderma reesei.** *Biotechnology for Biofuels* 2018, **11**:84.

Material suplementar



Additional file 7: Fig. S1. Venn diagrams of pathway classes shared between *Trichoderma* spp. with expression levels associated with *cre1* (**a**) and *xyr1* (**b**) transcript levels.

Additional file 7: Table S7. Number of transcripts coexpressed in each subgroup formed from *cre1* and *xyl1* groups for *Trichoderma* spp. and its corresponding color.

<i>cre1</i>				<i>xyl1</i>		
	Subgroup	Number of transcripts	Color	Subgroup	Number of transcripts	Color
Th3844	1.0	21	Jungle green	4.0	64	Jungle green
	1.1	13	Purple	4.1	54	Lime green
	1.2	15	Yellow	4.2	49	Grey
	1.3	4	Grey	-	-	-
Th0179	2.0	41	Jungle green	5.0	46	Jungle green
	2.1	21	Purple	5.1	58	Lime green
	2.2	44	Yellow	5.2	49	Grey
	2.3	8	Grey	-	-	-
	Elements disconnected	2	-	-	-	-
Ta0020	3.0	65	Jungle green	6.1	6	Jungle green
	3.1	73	Orange	6.2	7	Purple
	3.4	47	Beige	6.4	7	Beige
	3.5	50	Grey	6.5	10	Grey
	Elements disconnected	16	-	Elements disconnected	20	-

Additional file 7: Table S10. Expression values in TPM of *cre1* and *xyl1* for *Trichoderma* spp. under cellulose and glucose growth conditions.

Strain	Gene ID	Protein product	FC values	p-value	Cellulose TPM	Glucose TPM	
Th0179	THAR02_10775	KKO971 23.1	1,041	0,271	906,063	854,791	
Th3844	THAR02_10775	KKO971 23.1	-1,136	0,035	519,282	598,932	<i>cre1</i>
Ta0020	TRIATD RAFT_3 01116	XP_0139 41427.1	2,096	0,000	1136,938	537,489	
Th0179	THAR02_09260	KKO986 30.1	-1,525	0,080	38,444	57,553	
Th3844	THAR02_09260	KKO986 30.1	1,225	0,291	89,359	74,078	<i>xyl1</i>
Ta0020	TRIATD RAFT_7 8601	XP_0139 41705.1	-1,649	0,056	23,978	39,190	

5. RESULTADOS COMPLEMENTARES

Foi realizado o estudo *in vitro* de proteínas produzidas por *T. harzianum* IOC-3844 relacionadas ao processo de hidrólise enzimática da biomassa vegetal. Para tanto, foi escolhida como alvo uma endo-xilanase da família GH11 com domínio CBM.

Expressão recombinante e caracterização inicial de uma endo-xilanase de *T. harzianum* em *Escherichia coli*

5.1. Contextualização

As endo-xilanases são os principais componentes do sistema enzimático xilanolítico e atuam na hidrólise das ligações β -1,4 da cadeia de hemicelulose [129], o segundo polímero mais abundante da parede celular vegetal [58]. Tecnicamente, os materiais de origem lignocelulósica têm uma importância de longa data como matérias-primas na indústria de celulose e papel [130]. Mais recentemente, eles também representam uma fonte potencialmente renovável e sustentável de combustíveis, uma vez que não competem com a produção de alimentos e poderiam substituir os derivados de combustíveis fósseis [131]. No entanto, a conversão eficiente de substratos lignocelulósicos em açúcares fermentescíveis requer etapas nas quais as enzimas hidrolíticas desempenham um papel importante [58]. Devido à natureza recalcitrante da biomassa vegetal e objetivando a degradação efetiva do material lignocelulósico, uma das abordagens mais promissoras é o enriquecimento de coquetéis enzimáticos em termos de atividades que lhes faltam [132]. Em escala industrial, as enzimas celulolíticas e hemicelulolíticas são produzidas principalmente por microrganismos, tais como bactérias e fungos [133]. Neste contexto, o fungo ascomiceto filamentoso *T. harzianum* produz uma ampla gama de enzimas ativas à hemicelulose que podem ser exploradas para fins biotecnológicos, principalmente no que se refere à produção de misturas enzimáticas para a degradação da biomassa [15, 17-20], bem como em diversas aplicações industriais [133]. As mais bem caracterizadas das endo-xilanases pertencem a família GH11, as quais, quando comparadas com outras enzimas xilanolíticas, apresentam como características atrativas alta especificidade ao substrato e alta eficiência catalítica [129]. Tais proteínas são excelentes candidatas para a expressão heteróloga a fim de se poder suplementar os coquetéis enzimáticos atualmente comercializados. Ademais, muitas xilanases possuem domínio CBM, o qual tem como função principal aumentar a concentração de enzimas na superfície do substrato, tornando o processo hidrolítico mais eficiente [129, 134]. Assim, este projeto de mestrado teve como tópico complementar a produção e a caracterização inicial de uma endo-xilanase recombinante da família GH11 de *T. harzianum* IOC-3844 (ThX) com domínio CBM em *Escherichia coli*.

Também foi apresentado a otimização de um método de *refolding*, o qual foi anteriormente descrito para uma GH11 de *Bacillus circulans* [131], possibilitando, dessa forma, a solubilização de ThX expressa nos corpos de inclusão.

5.2. Materiais e métodos

5.2.1. Mapeamento e reconstrução do transcrito *gh11*

As bibliotecas de RNA-Seq de *T. harzianum* IOC-3844 [39] foram utilizadas para identificar a principal endo-xilanase presente nesta linhagem durante a degradação da biomassa vegetal. Os *reads* obtidos através do cultivo do fungo em condições de celulose ou de glicose foram mapeados contra a sequência nucleotídica que codifica uma endo-xilanase da família GH11 (número de acesso ao GenBank: [JOKZ01000041.1](https://www.ncbi.nlm.nih.gov/nuclot/JOKZ01000041.1)) disponível para *T. harzianum* T6776 ([PRJNA252551](https://www.ncbi.nlm.nih.gov/nuclot/PRJNA252551)). O mapeamento foi realizado através do *software* CLC Genomics Workbench (CLC bio – 6.5.2 v; Finlandsgade, Dk) [37] utilizando os seguintes parâmetros: configurações de mapeamento (*minimum lenght fraction* = 0.5; *minimum similarity fraction* = 0.8; e *maximum number of hits for a read* = 10) e configurações de alinhamento (*minimum distance* = 150 e *maximum distance* = 300). O transcrito reconstruído para *T. harzianum* IOC-3844 foi denominado de *gh11* e a predição baseada em sequência de domínios proteicos mostrou a presença de um domínio CBM em ThX.

5.2.2. Amplificação e clonagem do transcrito *gh11*

O transcrito *gh11* (819 pb) foi amplificado a partir do cDNA de *T. harzianum* IOC-3844 [34]. Os *primers* 5`- ATTCATATGGCTCCTACGGAGGTAGCC -3` e 5` - TTGGGCTCGAGTTACAGGCACTGAGAATA -3` continham locais de restrição *NdeI* e *XhoI* (sublinhados), respectivamente, e foram projetados usando a sequência de *gh11* obtida neste projeto (ver tópico acima “Mapeamento e reconstrução do transcrito *gh11*”). O produto de amplificação obtido por PCR foi clonado no vetor de expressão pET-28a (+) (Novagen, Madison, WI, EUA), o qual adicionou um marcador N-terminal de seis histidinas e um local de protease de trombina à sequência de codificação. As substituições de nucleotídeos no plasmídeo construído foram analisadas através de sequenciamento de DNA pelo método de Sanger [135].

5.2.3. Expressão em *E. coli* de ThX

ThX foi superexpressa em células de *E. coli* Rosetta (Novagen, Darmstadt, Alemanha). As células foram cultivadas a 37 °C com agitação a 250 rpm em 1 L de meio LB contendo cloranfenicol (34 µg mL⁻¹) e canamicina (30 µg mL⁻¹) até atingir a fase logarítmica - OD₆₀₀ 0,8-1 unidades de absorbância. A expressão proteica recombinante foi induzida pela adição de 0,4 mM de isopropil-β-D-tiogalactopiranosídeo (IPTG), seguido de cultivo em diferentes

tempos (ver tópico abaixo “Teste de expressão de ThX”) à 18 °C e 180 rpm. A cultura foi então centrifugada (3000 g, 15 min, 4 °C) e as células foram ressuspensas em 50 mL de tampão A [HEPES 40 mM pH 7,5, NaCl 150 mM] contendo 1 mg mL⁻¹ de lisozima e 1 mM de PMSF. As células foram rompidas por sonicação e a fração solúvel foi coletada por centrifugação (20.000 g, 40 min, 4 °C). O *pellet* correspondente a fração insolúvel foi armazenado a -20 °C.

5.2.4. Teste de expressão de ThX

Para a avaliação da expressão ao longo do tempo, foram coletadas alíquotas de 40 mL antes da indução (T₀), e de 20 mL e 5 mL nos tempos 3h (T₁) e 20h (T₂) após a indução, respectivamente. As amostras coletadas foram analisadas via eletroforese em gel de poliacrilamida 12% (SDS-PAGE) [136].

5.2.5. Solubilização de ThX pelo método de *refolding*

Os corpos de inclusão obtidos após a expressão da proteína recombinante em *E.coli* e enriquecidos com ThX foram utilizados em um procedimento de *refolding* de proteínas descrito por Santos et al., 2012 [137]. A etapa de diálise foi realizada com os diferentes tampões descritos por Kötzler et al., 2017 [131]. Aos tampões, foram adicionados 1 mM de DTT, uma vez que ThX apresenta seis cisteínas em seu domínio CBM, 20 mM imidazol e 10% v/v de glicerol, como descrito na **Tabela 1**. O pH utilizado em cada tampão foi ajustado de acordo com o ponto isoelétrico (pI) de ThX (pI = 6,4).

Tabela 1. Diferentes tampões utilizados para a diálise de ThX. Modificado de Kötzler et al., 2017 [131].

Tampão A (pH 7)	Tampão B (pH 8,5)
0,06% p/v de PEG 3350	0,06% p/v de PEG 3350
1,0 M de ureia	0,5 M de ureia
0,5 mM de Tween-80	0,5 mM de Tween-80
50 mM de Hepes	50 mM de Tris
1 mM de DTT	1 mM de DTT
20 mM imidazol	20 mM imidazol
10% v/v glicerol	10% v/v glicerol

5.3. Resultados parciais

Após a amplificação do transcrito *gh11* e a inserção do fragmento correspondente no vetor de expressão pET-28a (+), células de *E. coli* Rosetta foram utilizadas para a expressão recombinante de ThX. Inicialmente, foi realizado o teste de expressão a 18 °C e 180 rpm em

dois tempos diferentes após a indução. Através de análise via gel SDS-PAGE 12% (**Figura 7**), o tempo correspondente às três horas (T_1) após a indução foi definido como o mais promissor em comparação aquele referente à indução *overnight*.

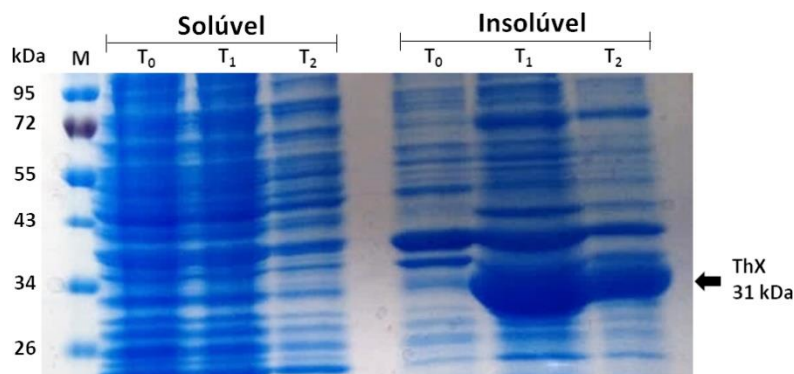


Figura 7. Teste de expressão de ThX. Gel SDS-PAGE 12% onde M, T_0 , T_1 e T_2 indicam, respectivamente, o marcador de massa molecular, zero, três e vinte horas após indução. A banda referente a ThX pode ser vista logo abaixo de 34 kDa.

Uma vez que ThX foi expressa nos corpos de inclusão, o que provavelmente é resultado da natureza altamente hidrofóbica das sequências de GH11 [131], fez-se necessária a otimização de um método de *refolding* de proteínas para a solubilização e purificação da proteína alvo. Como mostrado na **Figura 8**, o método escolhido foi bem-sucedido e grandes quantidades de ThX solúvel foram alcançadas após a etapa de diálise. Ao final desta última etapa, não houve precipitação, no entanto, a purificação de ThX por cromatografia de exclusão por tamanho indicou forte agregação proteica em solução (dados não mostrados).

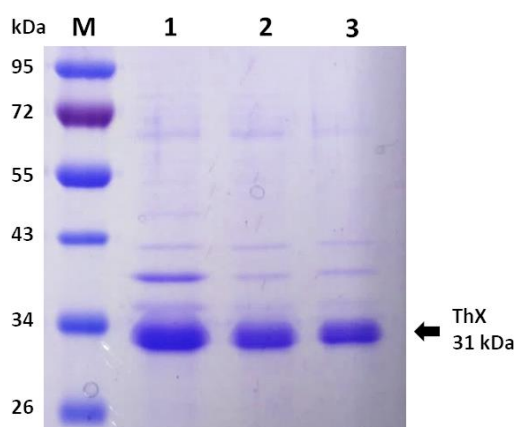


Figura 8. Solubilização de ThX. Gel SDS-PAGE 12% indicando ThX antes da diálise (1) e após a diálise com o tampão A (2) e B (3), respectivamente. M: marcador de massa molecular.

5.4. Perspectivas

A triagem inicial para determinar a atividade enzimática de ThX será realizada com o emprego de xilano comercial obtido de madeira de faia (Sigma-Aldrich®) como substrato. Uma vez confirmada a atividade enzimática de ThX através do método do ácido 3,5-dinitrosalicílico (DNS) [138], o qual estima a quantidade de açúcar redutor liberada sob condições de ensaio, serão realizados ensaios de pH e temperatura a fim de determinar a condição ótima de ThX. Para a determinação do pH ótimo das enzimas, pretende-se empregar os sistemas de tampões: ácido cítrico/citrato de sódio 0,1 M (pH 4,0 – 6,5), fosfato monobásico/fosfato dibásico 0,1 M (pH 7,0 – 8,0), Tris/HCl 0,1 M (pH 8,5) e glicina/NaOH 0,1 M (pH 9 e 9,5). A temperatura escolhida para os ensaios de pH será a de 50 °C. O perfil de atividade enzimática em diferentes temperaturas será obtido no melhor pH de ThX, entre 20 e 80 °C. A caracterização bioquímica também será realizada por meio do cálculo de parâmetros de cinética enzimática empregando o modelo hiperbólico de Michaelis-Menten. Além disso, será testado o efeito da adição de cátions mono e divalentes, tais como MgCl₂, CuCl₂, ZnSO₄, CaCl₂, NiCl₂, KCl, NaCl, MnCl₂, LiCl, CoCl₂, e EDTA sobre a atividade enzimática de ThX. Quanto ao perfil da estrutura secundária e a avaliação da termoestabilidade de ThX, estes serão analisados por dicroísmo circular.

6. CONCLUSÕES GERAIS

De modo geral, as análises revelaram que diferentes estratégias associadas aos FTs XYR1 e CRE1 durante a degradação da celulose são utilizadas por *T. atroviride* e por *T. harzianum*, e que esta diferença pode, inclusive, ser observada entre linhagens da mesma espécie. Além disso, foi observado em *T. harzianum* que os transcritos codificando CAZymes relacionadas à hidrólise da celulose e da hemicelulose não são necessariamente co-expressos com os FTs aqui estudados. Tomados em conjunto, nossos resultados podem contribuir para uma melhor compreensão dos mecanismos genéticos envolvidos na regulação de enzimas hidrolíticas nas espécies micoparasitas de *Trichoderma*. Isso será importante para expandir o potencial de *T. harzianum* não apenas no desenvolvimento de melhores coquetéis enzimáticos, mas também em muitas outras aplicações industriais.

Finalmente, mas não menos importante, os resultados complementares indicaram que ThX foi produzida com sucesso utilizando um hospedeiro procarioto e que o método de *refolding* empregado mostrou ser eficiente para a solubilização da proteína alvo, inicialmente expressa nos corpos de inclusão. Contudo, serão necessários testes empregando substrato específico para avaliar a atividade xilanolítica de ThX. Caso a atividade enzimática seja confirmada, nosso trabalho poderá abrir caminho para estudos adicionais avaliando a estrutura e a função dessa proteína e seu uso para suplementar e melhorar os coquetéis enzimáticos projetados para a conversão da biomassa vegetal em etanol 2G. Ademais, a caracterização bioquímica e estrutural de ThX poderá elucidar sua estabilidade e especificidade em relação à diferentes substratos contribuindo, dessa forma, para a expansão de sua aplicabilidade em termos biotecnológicos.

7. PERSPECTIVAS GERAIS

O presente estudo contribuiu para a identificação de novos transcritos alvos em *T. harzianum* possivelmente relacionados à degradação da celulose e ainda mais importante, associados a nível de expressão com os principais reguladores envolvidos no processo de degradação da parede celular vegetal. Estes transcritos devem ser melhor investigados para que se possa comprovar sua real influência sobre a sacarificação do bagaço através de estudos futuros envolvendo a expressão heteróloga e a caracterização bioquímica e estrutural de proteínas. Porém, em vista da dificuldade de se obter proteínas solúveis e em sua correta conformação original, uma alternativa proposta seria a técnica de imunoprecipitação de cromatina seguida de sequenciamento de DNA em larga escala (ChIP-Seq). Através dessa metodologia, seria possível identificar os sítios de ligação ao DNA associados às proteínas, tais como FTs. Tais experimentos poderiam oferecer estratégias para a melhoria das linhagens de *T. harzianum* estudadas em condição de degradação de celulose visando otimizar os coquetéis enzimáticos e minimizar o custo de produção de etanol 2G a partir da biomassa lignocelulósica disponível no Brasil, bem como explorar o potencial hidrolítico dessa espécie em diversos processos biotecnológicos.

8. REFERÊNCIAS BIBLIOGRÁFICAS

1. Lopes FAC, Steindorff AS, Geraldine AM, Brandão RS, Monteiro VN, Júnior ML, Coelho ASG, Ulhoa CJ, Silva RN: **Biochemical and metabolic profiles of Trichoderma strains isolated from common bean crops in the Brazilian Cerrado, and potential antagonism against Sclerotinia sclerotiorum.** *Fungal Biology* 2012, **116**:815-824.
2. Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, et al: **Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of Trichoderma.** *Genome Biology* 2011, **12**:R40.
3. Druzhinina IS, Kopchinskiy AG, Kubicek CP: **The first 100 Trichoderma species characterized by molecular data.** *Mycoscience* 2006, **47**:55.
4. Xie B-B, Qin Q-L, Shi M, Chen L-L, Shu Y-L, Luo Y, Wang X-W, Rong J-C, Gong Z-T, Li D, et al: **Comparative genomics provide insights into evolution of trichoderma nutrition style.** *Genome biology and evolution* 2014, **6**:379-390.
5. Bischof RH, Ramoni J, Seiboth B: **Cellulases and beyond: the first 70 years of the enzyme producer Trichoderma reesei.** *Microbial Cell Factories* 2016, **15**:106.
6. Druzhinina IS, Kubicek CP: **Chapter Two - Familiar Stranger: Ecological Genomics of the Model Saprotroph and Industrial Enzyme Producer Trichoderma reesei Breaks the Stereotypes.** In *Advances in Applied Microbiology. Volume 95*. Edited by Sariaslani S, Michael Gadd G: Academic Press; 2016: 69-147
7. Gupta VK, Steindorff AS, de Paula RG, Silva-Rocha R, Mach-Aigner AR, Mach RL, Silva RN: **The Post-genomic Era of Trichoderma reesei: What's Next?** *Trends in Biotechnology* 2016, **34**:970-982.
8. Keshavarz B, Khalesi M: **Trichoderma reesei, a superior cellulase source for industrial applications.** *Biofuels* 2016, **7**:713-721.
9. Chaverri P, Samuels GJ, Stewart EL: **Hypocrea virens sp. nov., the Teleomorph of Trichoderma virens.** *Mycologia* 2001, **93**:1113-1124.
10. Dodd SL, Lieckfeldt E, Samuels GJ: **Hypocrea atroviridis sp. nov., the teleomorph of Trichoderma atroviride.** *Mycologia* 2003, **95**:27-40.
11. Druzhinina IS, Kubicek CP, Komoń-Zelazowska M, Mulaw TB, Bissett J: **The Trichoderma harzianum demon: complex speciation history resulting in coexistence of hypothetical biological species, recent agamospecies and numerous relict lineages.** *BMC Evolutionary Biology* 2010, **10**:94.

12. Abdel-Fattah GM, Shabana YM, Ismail AE, Rashad YM: **Trichoderma harzianum: a biocontrol agent against Bipolaris oryzae.** *Mycopathologia* 2007, **164**:81-89.
13. Zhang F, Ge H, Zhang F, Guo N, Wang Y, Chen L, Ji X, Li C: **Biocontrol potential of Trichoderma harzianum isolate T-aloe against Sclerotinia sclerotiorum in soybean.** *Plant Physiology and Biochemistry* 2016, **100**:64-74.
14. Saravanakumar K, Li Y, Yu C, Wang Q-q, Wang M, Sun J, Gao J-x, Chen J: **Effect of Trichoderma harzianum on maize rhizosphere microbiome and biocontrol of Fusarium Stalk rot.** *Scientific Reports* 2017, **7**:1771.
15. Horta MAC, Vicentini R, Delabona PdS, Laborda P, Crucello A, Freitas S, Kuroshu RM, Polikarpov I, Pradella JGdC, Souza AP: **Transcriptome Profile of Trichoderma harzianum IOC-3844 Induced by Sugarcane Bagasse.** *PLOS ONE* 2014, **9**:e88689.
16. Li J-X, Zhang F, Li J, Zhang Z, Bai F-W, Chen J, Zhao X-Q: **Rapid production of lignocellulolytic enzymes by Trichoderma harzianum LZ117 isolated from Tibet for biomass degradation.** *Bioresource technology* 2019, **292**:122063.
17. Santos CA, Ferreira-Filho JA, O'Donovan A, Gupta VK, Tuohy MG, Souza AP: **Production of a recombinant swollenin from Trichoderma harzianum in Escherichia coli and its potential synergistic role in biomass degradation.** *Microbial Cell Factories* 2017, **16**:83.
18. Santos CA, Morais MAB, Terrett OM, Lyczakowski JJ, Zanthorlin LM, Ferreira-Filho JA, Tonoli CCC, Murakami MT, Dupree P, Souza AP: **An engineered GH1 β -glucosidase displays enhanced glucose tolerance and increased sugar release from lignocellulosic materials.** *Scientific Reports* 2019, **9**:4903.
19. Ferreira Filho JA, Horta MAC, Beloti LL, dos Santos CA, de Souza AP: **Carbohydrate-active enzymes in Trichoderma harzianum: a bioinformatic analysis bioprospecting for key enzymes for the biofuels industry.** *BMC Genomics* 2017, **18**:779.
20. Crucello A, Sforça DA, Horta MAC, dos Santos CA, Viana AJC, Beloti LL, de Toledo MAS, Vincentz M, Kuroshu RM, de Souza AP: **Analysis of Genomic Regions of Trichoderma harzianum IOC-3844 Related to Biomass Degradation.** *PLOS ONE* 2015, **10**:e0122122.
21. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B: **The carbohydrate-active enzymes database (CAZy) in 2013.** *Nucleic acids research* 2014, **42**:D490-D495.

22. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics.** *Nucleic acids research* 2009, **37**:D233-D238.
23. Ilmén M, Saloheimo A, Onnela ML, Penttilä ME: **Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*.** *Applied and environmental microbiology* 1997, **63**:1298-1306.
24. Stricker AR, Mach RL, de Graaff LH: **Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*).** *Applied Microbiology and Biotechnology* 2008, **78**:211.
25. 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*.** *Eukaryotic cell* 2006, **5**:2128-2137.
26. Portnoy T, Margeot A, Linke R, Atanasova L, Fekete E, Sándor 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.
27. Antoniêto ACC, dos Santos Castro L, Silva-Rocha R, Persinoti GF, Silva RN: **Defining the genome-wide role of CRE1 during carbon catabolite repression in *Trichoderma reesei* using RNA-Seq analysis.** *Fungal Genetics and Biology* 2014, **73**:93-103.
28. Benocci T, Aguilar-Pontes MV, Zhou M, Seiboth B, de Vries RP: **Regulators of plant biomass degradation in ascomycetous fungi.** *Biotechnology for Biofuels* 2017, **10**:152.
29. Klein-Marcuschamer D, Oleskowicz-Popiel P, Simmons BA, Blanch HW: **The challenge of enzyme cost in the production of lignocellulosic biofuels.** *Biotechnology and Bioengineering* 2012, **109**:1083-1087.
30. de Carvalho LM, Borelli G, Camargo AP, de Assis MA, de Ferraz SMF, Fiamenghi MB, José J, Mofatto LS, Nagamatsu ST, Persinoti GF, et al: **Bioinformatics applied to biotechnology: A review towards bioenergy research.** *Biomass and Bioenergy* 2019, **123**:195-224.
31. Fuller T, Langfelder P, Presson A, Horvath S: **Review of Weighted Gene Coexpression Network Analysis.** In *Handbook of Statistical Bioinformatics*. Edited by Lu HH-S, Schölkopf B, Zhao H. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011: 369-388

32. van Dam S, Vösa U, van der Graaf A, Franke L, de Magalhães JP: **Gene co-expression analysis for functional classification and gene-disease predictions.** *Briefings in bioinformatics* 2018, **19**:575-592.
33. Pirim H: **Construction of Gene Networks Using Expression Profiles.** In *Soft Computing for Biological Systems*. Edited by Purohit HJ, Kalia VC, More RP. Singapore: Springer Singapore; 2018: 67-89
34. Horta MAC, Filho JAF, Murad NF, de Oliveira Santos E, dos Santos CA, Mendes JS, Brandão MM, Azzoni SF, de Souza AP: **Network of proteins, enzymes and genes linked to biomass degradation shared by Trichoderma species.** *Scientific Reports* 2018, **8**:1341.
35. Borin GP, Carazzolle MF, dos Santos RAC, Riaño-Pachón DM, Oliveira JVdC: **Gene Co-expression Network Reveals Potential New Genes Related to Sugarcane Bagasse Degradation in Trichoderma reesei RUT-30.** *Frontiers in Bioengineering and Biotechnology* 2018, **6**.
36. Shida Y, Furukawa T, Ogasawara W: **Deciphering the molecular mechanisms behind cellulase production in Trichoderma reesei, the hyper-cellulolytic filamentous fungus.** *Bioscience, Biotechnology, and Biochemistry* 2016, **80**:1712-1729.
37. de Paula RG, Antoniêto ACC, Ribeiro LFC, Carraro CB, Nogueira KMV, Lopes DCB, Silva AC, Zerbini MT, Pedersoli WR, Costa MdN, Silva RN: **New Genomic Approaches to Enhance Biomass Degradation by the Industrial Fungus Trichoderma reesei.** *International Journal of Genomics* 2018, **2018**:17.
38. Zhang F, Bunterngrsook B, Li J-X, Zhao X-Q, Champreda V, Liu C-G, Bai F-W: **Chapter Three - Regulation and production of lignocellulolytic enzymes from Trichoderma reesei for biofuels production.** In *Advances in Bioenergy. Volume 4*. Edited by Li Y, Ge X: Elsevier; 2019: 79-119
39. Almeida DA, Horta MAC, Filho JAF, Murad NF, de Souza AP: **The synergistic actions of hydrolytic genes in coexpression networks reveal the potential of Trichoderma harzianum for cellulose degradation.** *bioRxiv* 2020:2020.2001.2014.906529.
40. Philp J: **Balancing the bioeconomy: supporting biofuels and bio-based materials in public policy.** *Energy & Environmental Science* 2015, **8**:3063-3068.
41. Cerqueira Leite RCd, Verde Leal MRL, Barbosa Cortez LA, Griffin WM, Gaya Scandiffio MI: **Can Brazil replace 5% of the 2025 gasoline world demand with ethanol?** *Energy* 2009, **34**:655-661.

42. Clark JH, Luque R, Matharu AS: **Green Chemistry, Biofuels, and Biorefinery**. *Annual Review of Chemical and Biomolecular Engineering* 2012, **3**:183-207.
43. dos Santos Castro L, Pedersoli WR, Antoniêto ACC, Steindorff AS, Silva-Rocha R, Martinez-Rossi NM, Rossi A, Brown NA, Goldman GH, Faça VM, 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.
44. RFA: **World Fuel Ethanol Production**. 2019.
45. Amorim HV, Lopes ML, de Castro Oliveira JV, Buckeridge MS, Goldman GH: **Scientific challenges of bioethanol production in Brazil**. *Applied Microbiology and Biotechnology* 2011, **91**:1267.
46. Antizar-Ladislao B, Turrion-Gomez JL: **Second-generation biofuels and local bioenergy systems**. *Biofuels, Bioproducts and Biorefining* 2008, **2**:455-469.
47. T.P. Basso TOB, C.R. Gallo and L.C. Basso: **Towards the Production of Second Generation Ethanol from Sugarcane Bagasse in Brazil**. Biomass Now - Cultivation and Utilization: IntechOpen; 2013.
48. Scheiterle L, Ulmer A, Birner R, Pyka A: **From commodity-based value chains to biomass-based value webs: The case of sugarcane in Brazil's bioeconomy**. *Journal of Cleaner Production* 2018, **172**:3851-3863.
49. UNICA: **Sugarcane, ethanol and sugar production**. UNICA 2019.
50. Sun JX, Sun XF, Sun RC, Su YQ: **Fractional extraction and structural characterization of sugarcane bagasse hemicelluloses**. *Carbohydrate Polymers* 2004, **56**:195-204.
51. Isikgor FH, Becer CR: **Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers**. *Polymer Chemistry* 2015, **6**:4497-4559.
52. Rubin EM: **Genomics of cellulosic biofuels**. *Nature* 2008, **454**:841-845.
53. Meents MJ, Watanabe Y, Samuels AL: **The cell biology of secondary cell wall biosynthesis**. *Annals of botany* 2018, **121**:1107-1125.
54. Arantes V, Saddler JN: **Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis**. *Biotechnology for Biofuels* 2010, **3**:4.
55. Ghasemi M, Singapati AY, Tsianou M, Alexandridis P: **Dissolution of semicrystalline polymer fibers: Numerical modeling and parametric analysis**. *AIChE Journal* 2017, **63**:1368-1383.

56. Mittal A, Katahira R, Himmel ME, Johnson DK: **Effects of alkaline or liquid-ammonia treatment on crystalline cellulose: changes in crystalline structure and effects on enzymatic digestibility.** *Biotechnology for Biofuels* 2011, **4**:41.
57. Gong J, Li J, Xu J, Xiang Z, Mo L: **Research on cellulose nanocrystals produced from cellulose sources with various polymorphs.** *RSC Advances* 2017, **7**:33486-33493.
58. Adsul MG, Singhvi MS, Gaikawari SA, Gokhale DV: **Development of biocatalysts for production of commodity chemicals from lignocellulosic biomass.** *Bioresource Technology* 2011, **102**:4304-4312.
59. Segato F, Damásio ARL, de Lucas RC, Squina FM, Prade RA: **Genomics review of holocellulose deconstruction by aspergilli.** *Microbiology and molecular biology reviews : MMBR* 2014, **78**:588-613.
60. Kricka W, Fitzpatrick J, Bond U: **Metabolic engineering of yeasts by heterologous enzyme production for degradation of cellulose and hemicellulose from biomass: a perspective.** *Frontiers in microbiology* 2014, **5**:174-174.
61. Barros J, Escamilla-Trevino L, Song L, Rao X, Serrani-Yarce JC, Palacios MD, Engle N, Choudhury FK, Tschaplinski TJ, Venables BJ, et al: **4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase.** *Nature Communications* 2019, **10**:1994.
62. dos Santos AC, Ximenes E, Kim Y, Ladisch MR: **Lignin–Enzyme Interactions in the Hydrolysis of Lignocellulosic Biomass.** *Trends in Biotechnology* 2019, **37**:518-531.
63. Mathews SL, Grunden AM, Pawlak J: **Degradation of lignocellulose and lignin by *Paenibacillus glucanolyticus*.** *International Biodeterioration & Biodegradation* 2016, **110**:79-86.
64. Maitan-Alfenas GP, Visser EM, Guimarães VM: **Enzymatic hydrolysis of lignocellulosic biomass: converting food waste in valuable products.** *Current Opinion in Food Science* 2015, **1**:44-49.
65. Lee RA, Lavoie J-M: **From first- to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity.** *Animal Frontiers* 2013, **3**:6-11.
66. Saini JK, Saini R, Tewari L: **Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments.** *3 Biotech* 2015, **5**:337-353.

67. Teixeira ACR, Sodré JR, Guarieiro LLN, Vieira ED, de Medeiros FF, Alves CT: **A Review on Second and Third Generation Bioethanol Production.** SAE International; 2016.
68. Robak K, Balcerek M: **Review of Second Generation Bioethanol Production from Residual Biomass.** *Food technology and biotechnology* 2018, **56**:174-187.
69. Wang M, Li Z, Fang X, Wang L, Qu Y: **Cellulolytic Enzyme Production and Enzymatic Hydrolysis for Second-Generation Bioethanol Production.** In *Biotechnology in China III: Biofuels and Bioenergy*. Edited by Bai F-W, Liu C-G, Huang H, Tsao GT. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012: 1-24
70. Andlar M, Rezić T, Mardetko N, Kracher D, Ludwig R, Šantek B: **Lignocellulose degradation: An overview of fungi and fungal enzymes involved in lignocellulose degradation.** *Engineering in Life Sciences* 2018, **18**:768-778.
71. Li X, Xia J, Zhu X, Bilal M, Tan Z, Shi H: **Construction and characterization of bifunctional cellulases: Caldicellulosiruptor-sourced endoglucanase, CBM, and exoglucanase for efficient degradation of lignocellulose.** *Biochemical Engineering Journal* 2019, **151**:107363.
72. Manavalan T, Manavalan A, Heese K: **Characterization of Lignocellulolytic Enzymes from White-Rot Fungi.** *Current Microbiology* 2015, **70**:485-498.
73. Xie S, Syrenne R, Sun S, Yuan JS: **Exploration of Natural Biomass Utilization Systems (NBUS) for advanced biofuel—from systems biology to synthetic design.** *Current Opinion in Biotechnology* 2014, **27**:195-203.
74. Gupta VK, Kubicek CP, Berrin J-G, Wilson DW, Couturier M, Berlin A, Filho EXF, Ezeji T: **Fungal Enzymes for Bio-Products from Sustainable and Waste Biomass.** *Trends in Biochemical Sciences* 2016, **41**:633-645.
75. Amin FR, Khalid H, Zhang H, Rahman SU, Zhang R, Liu G, Chen C: **Pretreatment methods of lignocellulosic biomass for anaerobic digestion.** *AMB Express* 2017, **7**:72-72.
76. Martínez AT, Ruiz-Dueñas FJ, Camarero S, Serrano A, Linde D, Lund H, Vind J, Tovborg M, Herold-Majumdar OM, Hofrichter M, et al: **Oxidoreductases on their way to industrial biotransformations.** *Biotechnology Advances* 2017, **35**:815-831.
77. Bomble YJ, Lin C-Y, Amore A, Wei H, Holwerda EK, Ciesielski PN, Donohoe BS, Decker SR, Lynd LR, Himmel ME: **Lignocellulose deconstruction in the biosphere.** *Current Opinion in Chemical Biology* 2017, **41**:61-70.

78. Meyer AS, Rosgaard L, Sørensen HR: **The minimal enzyme cocktail concept for biomass processing.** *Journal of Cereal Science* 2009, **50**:337-344.
79. Kubicek CP, Kubicek EM: **Enzymatic deconstruction of plant biomass by fungal enzymes.** *Current Opinion in Chemical Biology* 2016, **35**:51-57.
80. Couturier M, Ladevèze S, Sulzenbacher G, Ciano L, Fanuel M, Moreau C, Villares A, Cathala B, Chaspoul F, Frandsen KE, et al: **Lytic xylan oxidases from wood-decay fungi unlock biomass degradation.** *Nature Chemical Biology* 2018, **14**:306.
81. Kidwai MK, Nehra M: **Biotechnological Applications of Trichoderma Species for Environmental and Food Security.** In *Plant Biotechnology: Recent Advancements and Developments*. Edited by Gahlawat SK, Salar RK, Siwach P, Duhan JS, Kumar S, Kaur P. Singapore: Springer Singapore; 2017: 125-156
82. Kumar V, Dwivedi SK: **Hexavalent chromium stress response, reduction capability and bioremediation potential of Trichoderma sp. isolated from electroplating wastewater.** *Ecotoxicology and Environmental Safety* 2019, **185**:109734.
83. Govarthanan M, Mythili R, Kamala-Kannan S, Selvankumar T, Srinivasan P, Kim H: **In-vitro bio-mineralization of arsenic and lead from aqueous solution and soil by wood rot fungus, Trichoderma sp.** *Ecotoxicology and Environmental Safety* 2019, **174**:699-705.
84. Singh RK, Tripathi R, Ranjan A, Srivastava AK: **Chapter 9 - Fungi as potential candidates for bioremediation.** In *Abatement of Environmental Pollutants*. Edited by Singh P, Kumar A, Borthakur A: Elsevier; 2020: 177-191
85. Hasan S: **Potential of Trichoderma sp. in Bioremediation: A Review.** *Journal of Basic and Applied Engineering Research* 2016.
86. Sandoval-Denis M, Sutton DA, Cano-Lira JF, Gené J, Fothergill AW, Wiederhold NP, Guarro J: **Phylogeny of the clinically relevant species of the emerging fungus Trichoderma and their antifungal susceptibilities.** *Journal of clinical microbiology* 2014, **52**:2112-2125.
87. Schuster A, Schmoll M: **Biology and biotechnology of Trichoderma.** *Applied microbiology and biotechnology* 2010, **87**:787-799.
88. Puranen T, Alapuranen M, Vehmaanperä J: **Chapter 26 - Trichoderma Enzymes for Textile Industries.** In *Biotechnology and Biology of Trichoderma*. Edited by Gupta VK, Schmoll M, Herrera-Estrella A, Upadhyay RS, Druzhinina I, Tuohy MG. Amsterdam: Elsevier; 2014: 351-362

89. Kunamneni A, Plou FJ, Alcalde M, Ballesteros A: **Chapter 24 - Trichoderma Enzymes for Food Industries.** In *Biotechnology and Biology of Trichoderma*. Edited by Gupta VK, Schmoll M, Herrera-Estrella A, Upadhyay RS, Druzhinina I, Tuohy MG. Amsterdam: Elsevier; 2014: 339-344
90. Kumar M, Ashraf S: **Role of Trichoderma spp. as a Biocontrol Agent of Fungal Plant Pathogens.** In *Probiotics and Plant Health*. Edited by Kumar V, Kumar M, Sharma S, Prasad R. Singapore: Springer Singapore; 2017: 497-506
91. Teeri TT, Koivula A, Linder M, Wohlfahrt G, Divne C, Jones TA: **Trichoderma reesei cellobiohydrolases: why so efficient on crystalline cellulose?** *Biochemical Society Transactions* 1998, **26**:173-177.
92. Penttilä M, Lehtovaara P, Nevalainen H, Bhikhabhai R, Knowles J: **Homology between cellulase genes of Trichoderma reesei: complete nucleotide sequence of the endoglucanase I gene.** *Gene* 1986, **45**:253-263.
93. Shoemaker S, Watt K, Tsitovsky G, Cox R: **Characterization and Properties of Cellulases Purified from Trichoderma Reesei Strain L27.** *Bio/Technology* 1983, **1**:687-690.
94. Druzhinina IS, Kubicek CP: **Genetic engineering of Trichoderma reesei cellulases and their production.** *Microbial biotechnology* 2017, **10**:1485-1499.
95. Seidl V, Gamauf C, Druzhinina IS, Seiboth B, Hartl L, Kubicek CP: **The Hypocrea jecorina (Trichoderma reesei) hypercellulolytic mutant RUT C30 lacks a 85 kb (29 gene-encoding) region of the wild-type genome.** *BMC genomics* 2008, **9**:327-327.
96. Mandels M, Reese ET: **Induction of cellulase in Trichoderma viride as influenced by carbon sources and metals.** *Journal of bacteriology* 1957, **73**:269-278.
97. Häkkinen M, Arvas M, Oja M, Aro N, Penttilä M, Saloheimo M, Pakula TM: **Re-annotation of the CAZy genes of Trichoderma reesei and transcription in the presence of lignocellulosic substrates.** *Microbial cell factories* 2012, **11**:134-134.
98. Mukherjee AK, Sampath Kumar A, Kranthi S, Mukherjee PK: **Biocontrol potential of three novel Trichoderma strains: isolation, evaluation and formulation.** *3 Biotech* 2014, **4**:275-281.
99. Atanasova L, Crom SL, Gruber S, Couplier F, Seidl-Seiboth V, Kubicek CP, Druzhinina IS: **Comparative transcriptomics reveals different strategies of Trichoderma mycoparasitism.** *BMC Genomics* 2013, **14**:121.
100. Delabona PdS, Farinas CS, da Silva MR, Azzoni SF, Pradella JGdC: **Use of a new Trichoderma harzianum strain isolated from the Amazon rainforest with**

- pretreated sugar cane bagasse for on-site cellulase production. *Bioresource Technology* 2012, **107**:517-521.
101. Crucello A: **Identificação de regiões genômicas implicadas no catabolismo de biomassa lignocelulósica pelo fungo *Trichoderma harzianum* IOC-3844.** Universidade Estadual de Campinas Instituto de Biologia 2014.
 102. Filho JAF, Horta MAC, dos Santos CA, Almeida DA, Murad NF, Mendes JS, Sforça DA, Silva CBC, Crucello A, de Souza AP: **“Integrative Genomic Analysis for the Bioprospection of Regulators and Accessory Enzymes Associated with Cellulose Degradation in a Filamentous Fungus (*Trichoderma harzianum*)”.** *bioRxiv* 2019:731323.
 103. Mello-de-Sousa TM, Gorsche R, Rassinger A, Poças-Fonseca MJ, Mach RL, Mach-Aigner AR: **A truncated form of the Carbon catabolite repressor 1 increases cellulase production in *Trichoderma reesei*.** *Biotechnology for Biofuels* 2014, **7**:129.
 104. Mach-Aigner AR, Pucher ME, Steiger MG, Bauer GE, Preis SJ, Mach RL: **Transcriptional regulation of *xyl1*, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*.** *Applied and environmental microbiology* 2008, **74**:6554-6562.
 105. Mello-de-Sousa TM, Rassinger A, Derntl C, Poças-Fonseca MJ, Mach RL, Mach-Aigner AR: **The Relation Between Promoter Chromatin Status, Xyr1 and Cellulase Ex-pression in *Trichoderma reesei*.** *Current genomics* 2016, **17**:145-152.
 106. 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 Genetics and Biology* 2009, **46**:564-574.
 107. Rauscher R, Würleitner E, Wacenovský C, Aro N, Stricker AR, Zeilinger S, Kubicek CP, Penttilä M, Mach RL: **Transcriptional regulation of *xyn1*, encoding xylanase I, in *Hypocrea jecorina*.** *Eukaryotic cell* 2006, **5**:447-456.
 108. Klaubauf S, Narang HM, Post H, Zhou M, Brunner K, Mach-Aigner AR, Mach RL, Heck AJR, Altelaar AFM, de Vries RP: **Similar is not the same: Differences in the function of the (hemi-)cellulolytic regulator XlnR (*Xlr1/Xyr1*) in filamentous fungi.** *Fungal Genetics and Biology* 2014, **72**:73-81.
 109. Akeel E, Metz B, Seiboth B, Kubicek CP: **Molecular regulation of arabinan and L-arabinose metabolism in *Hypocrea jecorina* (*Trichoderma reesei*).** *Eukaryotic cell* 2009, **8**:1837-1844.

110. Seiboth B, Metz B: **Fungal arabinan and L-arabinose metabolism.** *Applied microbiology and biotechnology* 2011, **89**:1665-1673.
111. Shelest E: **Transcription factors in fungi.** *FEMS Microbiology Letters* 2008, **286**:145-151.
112. Aro N, Saloheimo A, Ilmén M, Penttilä M: **ACEII, a Novel Transcriptional Activator Involved in Regulation of Cellulase and Xylanase Genes of Trichoderma reesei.** *Journal of Biological Chemistry* 2001, **276**:24309-24314.
113. Cubero B, Scazzocchio C: **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* 1994, **13**:407-415.
114. Mach RL, Strauss J, Zeilinger S, Schindler M, Kubicek CP: **Carbon catabolite repression of xylanase I (xyn1) gene expression in Trichoderma reesei.** *Molecular Microbiology* 1996, **21**:1273-1281.
115. Zeilinger S, Schmoll M, Pail M, Mach RL, Kubicek CP: **Nucleosome transactions on the Hypocrea jecorina (Trichoderma reesei) cellulase promoter cbh2 associated with cellulase induction.** *Molecular Genetics and Genomics* 2003, **270**:46-55.
116. 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.** *Eukaryotic cell* 2011, **10**:262-271.
117. Sun J, Glass NL: **Identification of the CRE-1 cellulolytic regulon in Neurospora crassa.** *PloS one* 2011, **6**:e25654-e25654.
118. Ries LNA, Beattie SR, Espeso EA, Cramer RA, Goldman GH: **Diverse Regulation of the CreA Carbon Catabolite Repressor in Aspergillus nidulans.** *Genetics* 2016, **203**:335-352.
119. Nakari-Setälä T, Paloheimo M, Kallio J, Vehmaanperä J, Penttilä M, Saloheimo M: **Genetic modification of carbon catabolite repression in Trichoderma reesei for improved protein production.** *Applied and environmental microbiology* 2009, **75**:4853-4860.
120. Wang F, Zhang R, Han L, Guo W, Du Z, Niu K, Liu Y, Jia C, Fang X: **Use of fusion transcription factors to reprogram cellulase transcription and enable efficient cellulase production in Trichoderma reesei.** *Biotechnology for biofuels* 2019, **12**:244-244.

121. Porciuncula JdO, Furukawa T, Shida Y, Mori K, Kuhara S, Morikawa Y, Ogasawara W: **Identification of Major Facilitator Transporters Involved in Cellulase Production during Lactose Culture of *Trichoderma reesei* PC-3-7.** *Bioscience, Biotechnology, and Biochemistry* 2013, **77**:1014-1022.
122. Ivanova C, Bååth 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.
123. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nature reviews Genetics* 2009, **10**:57-63.
124. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC Bioinformatics* 2008, **9**:559.
125. Wang YXR, Huang H: **Review on statistical methods for gene network reconstruction using expression data.** *Journal of Theoretical Biology* 2014, **362**:53-61.
126. Luscombe NM, Madan Babu M, Yu H, Snyder M, Teichmann SA, Gerstein M: **Genomic analysis of regulatory network dynamics reveals large topological changes.** *Nature* 2004, **431**:308-312.
127. Han J-DJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJM, Cusick ME, Roth FP, Vidal M: **Evidence for dynamically organized modularity in the yeast protein–protein interaction network.** *Nature* 2004, **430**:88-93.
128. Aghcheh RK, Kubicek CP: **Epigenetics as an emerging tool for improvement of fungal strains used in biotechnology.** *Applied Microbiology and Biotechnology* 2015, **99**:6167-6181.
129. Paës G, Berrin J-G, Beaugrand J: **GH11 xylanases: Structure/function/properties relationships and applications.** *Biotechnology Advances* 2012, **30**:564-592.
130. Viikari L, Kantelinen A, Sundquist J, Linko M: **Xylanases in bleaching: From an idea to the industry.** *FEMS Microbiology Reviews* 1994, **13**:335-350.
131. Kötzler MP, McIntosh LP, Withers SG: **Refolding the unfoldable: A systematic approach for renaturation of *Bacillus circulans* xylanase.** *Protein science : a publication of the Protein Society* 2017, **26**:1555-1563.
132. Liu R, Chen L, Jiang Y, Zou G, Zhou Z: **A novel transcription factor specifically regulates GH11 xylanase genes in *Trichoderma reesei*.** *Biotechnology for Biofuels* 2017, **10**:194.

133. Bhardwaj N, Kumar B, Verma P: **A detailed overview of xylanases: an emerging biomolecule for current and future prospective.** *Bioresources and Bioprocessing* 2019, **6**:40.
134. Zhang J, Moilanen U, Tang M, Viikari L: **The carbohydrate-binding module of xylanase from *Nonomuraea flexuosa* decreases its non-productive adsorption on lignin.** *Biotechnology for Biofuels* 2013, **6**:18.
135. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors.** *Proceedings of the National Academy of Sciences of the United States of America* 1977, **74**:5463-5467.
136. Schagger H, von Jagow G: **Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa.** *Analytical Biochemistry* 1987, **166**:368-379.
137. Santos CA, Beloti LL, Toledo MAS, Crucello A, Favaro MTP, Mendes JS, Santiago AS, Azzoni AR, Souza AP: **A novel protein refolding protocol for the solubilization and purification of recombinant peptidoglycan-associated lipoprotein from *Xylella fastidiosa* overexpressed in *Escherichia coli*.** *Protein Expression and Purification* 2012, **82**:284-289.
138. Miller GL: **Protein Determination of Large Numbers of Samples.** *Analytical Chemistry* 1959, **31**:964-964.

ANEXO I – DECLARAÇÃO CIBio



CIDADE UNIVERSITÁRIA "ZEFERINO VAZ",
29 DE JANEIRO DE 2020.

CIBio: 01 / 2020

IDENTIFICAÇÃO

Mestrado: Rafaela Rossi Rosolen

Programa de Pós-Graduação: Genética e Biologia Molecular – Instituto de Biologia – UNICAMP.

PROJETO

"Redes de co-expressão gênica para identificar módulos funcionais de XYR1 e CRE1 em *Trichoderma* spp."

PARECER

Projeto aprovado pela CIBio / CBMEG sob número 01 / 2010 – Expressão de proteínas de interesse biotecnológico e seu estudo bioquímico, funcional e estrutural.

Coordenador: Profa. Dra. Anete Pereira de Souza.

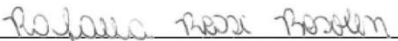
PROFA. DRA. MÔNICA BARBOSA DE MELO
Presidente da CIBio / CBMEG – UNICAMP

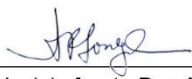
ANEXO II – DECLARAÇÃO DE AUTORIA

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **Redes de co-expressão gênica para identificar módulos funcionais de XYR1 e CRE1 em *Trichoderma spp.***, não infringem os dispositivos da Lei n.º 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 28 de abril de 2020.

Assinatura : 
Nome do(a) autor(a): **Rafaela Rossi Rosolen**
RG n.º 4898990801

Assinatura : 
Nome do(a) orientador(a): **Anete Pereira de Souza**
RG n.º 8680325