



**MARIA AUGUSTA CRIVELANTE HORTA**

**ANÁLISE DO TRANSCRIPTOMA DE *TRICHODERMA HARZIANUM* PARA A BIOPROSPECÇÃO DE  
ENZIMAS HIDROLÍTICAS**

**ANALYSIS OF *TRICHODERMA HARZIANUM* TRANSCRIPTOME FOR BIOPROSPECTING OF  
HYDROLYTIC ENZYMES**

**CAMPINAS, 2014**





UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Biologia

MARIA AUGUSTA CRIVELENTE HORTA

**ANÁLISE DO TRANSCRIPTOMA DE *TRICHODERMA HARZIANUM* PARA A BIOPROSPECÇÃO DE ENZIMAS HIDROLÍTICAS**

**ANALYSIS OF *TRICHODERMA HARZIANUM* TRANSCRIPTOME FOR BIOPROSPECTING OF HYDROLYTIC ENZYMES**

**Tese apresentada ao instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para obtenção do título de Doutora em Genética e Biologia Molecular na área de Genética Vegetal e Melhoramento**

*Thesis presented to the Institute of Biology of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Genetics and Molecular Biology, in the area of Plant Genetics and Genetic Breeding*

**Orientadora: ANETE PEREIRA DE SOUZA**

**Co-orientador: Sindélia Freitas Azzoni**

**Este exemplar corresponde à versão final da tese defendida pela aluna Maria Augusta Crivelente Horta e orientada pela profa. Dra. Anete Pereira de Souza**

A handwritten signature in purple ink, appearing to read "Anete", is written over a horizontal line.

**Anete Pereira de Souza**

**CAMPINAS  
2014**

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Biologia  
Ana Marria Rabetti - CRB 8/2471

C868a Crivelente Horta, Maria Augusta, 1981-  
Análise do transcriptoma de *Trichoderma harzianum* para a bioprospecção de enzimas hidrolíticas / Maria Augusta Crivelente Horta. – Campinas, SP : [s.n.], 2014.

Orientador: Anete Pereira de Souza.

Coorientador: Sindélia Freitas Azzoni.

Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Leveduras (Fungos) - Engenharia genética. 2. Biomassa. 3. Enzimas. 4. Celulose. 5. Hemicelulose. I. Souza, Anete Pereira de, 1962-. II. Azzoni, Sindélia Freitas. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Analysis of *Trichoderma harzianum* transcriptome for bioprospecting of hydrolytic enzymes

**Palavras-chave em inglês:**

Genetic engineering

Biomass

Enzymes

Cellulose

Hemicellulose

**Área de concentração:** Genética Vegetal e Melhoramento

**Titulação:** Doutora em Genética e Biologia Molecular

**Banca examinadora:**

Anete Pereira de Souza [Orientador]

Arnaldo Márcio Ramalho Prata

Roberto Ruller

Fabiana Fanttinnatti Garboggini

Luciana Gonzaga de Oliveira

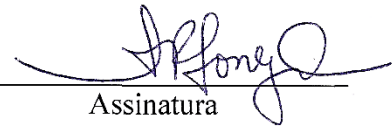
**Data de defesa:** 24-11-2014

**Programa de Pós-Graduação:** Genética e Biologia Molecular

Campinas, 24 de novembro de 2014

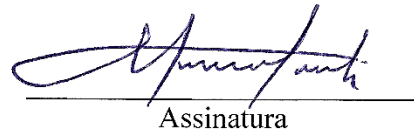
**BANCA EXAMINADORA**

Profa. Dra. Anete Pereira De Souza (orientadora)



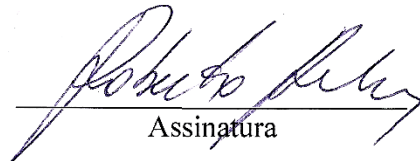
Assinatura

Prof. Dr. Arnaldo Márcio Ramalho Prata



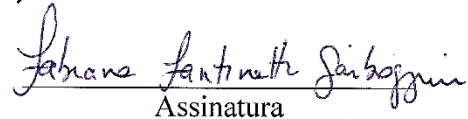
Assinatura

Prof. Dr. Roberto Ruller



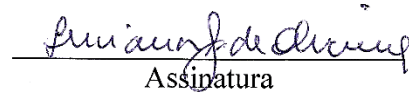
Assinatura

Profª Drª Fabiana Fanttinatti Garboggini



Assinatura

Profª Drª Luciana Gonzaga de Oliveira



Assinatura

Prof. Dr. Marcelo Mendes Brandão

Assinatura

Prof. Dr. George Jackson

Assinatura

Profª Drª Katlin Brauer Massirer

Assinatura



## ABSTRACT

In order to contribute to the development of second-generation ethanol technology, this study analyzes the transcriptome of *T. harzianum* IOC-3844 using high-throughput sequencing techniques. The main objective of this analysis was to identify, characterize and catalog the transcripts expressed by *T. harzianum* related to the degradation of complex substrates such as sugar cane bagasse, revealing the set of genes involved in the degradation of biomass. The analysis of the transcriptome of the fungus *Trichoderma harzianum* under conditions that induce the degradation of biomass allowed the identification of genes potentially effective in the biodegradation process, an essential step for understanding the enzymatic process. Sequencing resulted in 246 million sequences with 72 bp, which corresponds to 14.7 GB analyzed. After assembly, 32,494 contigs were generated, identified and classified according to their identity. All sequence contigs were compared with NCBI database, Gene Ontology (GO terms), Kyoto Encyclopedia of Genes (KEGG), Carbohydrate Active-Enzymes (CAZymes). 487 CAZymes were identified in the transcriptome, including those related to reactions of cellulose and hemicellulose depolymerization. Sequences classified as catalytic activity (6,975) and regulatory activity (143) may be involved with this type of reaction. This analysis defines the set of genes involved in the degradation of cellulose and hemicellulose of *T. harzianum*, and accessory genes related to depolymerization of the biomass. An analysis of expression levels was used to calculate the set of differentially expressed genes in different culture conditions. The results add to knowledge about the composition of the genome and gene expression activity of the fungus *Trichoderma harzianum*, and provides

important information regarding the genetic mechanisms of biomass degradation that the fungus uses. The information obtained may be used for other species of filamentous fungi with potential for biodegradation.

Keywords: biomass, *Trichoderma*, fungus, biodegradation, hydrolytic enzymes, CAZy, cellulose, hemicellulose, transcriptome, high-throughput sequencing, enzymatic hydrolysis, sugarcane bagasse, glycoside hydrolases.



## RESUMO

Buscando contribuir com o desenvolvimento da tecnologia de produção do etanol de segunda geração, o presente estudo analisa o transcriptoma de *T. harzianum* IOC-3844 utilizando técnicas de sequenciamento *high-throughput*. O principal objetivo dessas análises foi identificar, caracterizar e catalogar os transcritos expressos por *T. harzianum* relacionados com a degradação de substratos complexos, como o bagaço de cana de açúcar, revelando o conjunto de genes envolvidos na degradação da biomassa. A análise do transcriptoma do fungo *Trichoderma harzianum* sob condições que induzem a degradação da biomassa permitiu a identificação de sequências de genes potencialmente eficazes no processo de biodegradação, uma etapa essencial à compreensão do processo de hidrólise enzimática. O sequenciamento resultou em 246 milhões de sequências com 72 pb, o que corresponde a 14,7 GPB analisados. Após a montagem, 32.494 contigs foram gerados, submetidos à identificação e classificados de acordo com sua identidade. Todas as sequências de contigs foram comparados com o banco de dados do NCBI, *Gene Ontology* (GO terms), *Enciclopédia de Genes Kyoto* (KEGG), *Carbohydrate Active-Enzymes* (CAZymes). Foram identificados 487 CAZymes no transcriptoma, inclusive aquelas ligadas as reações químicas de despolimerização de celulose e hemicelulose. As sequências classificadas como atividade catalítica (6.975) e atividade reguladora (143) podem estar envolvidas com esse tipo de reação. A análise permitiu definir o principal conjunto de genes envolvidos na degradação da celulose e de hemicelulose do *T. harzianum*, e genes acessórios relativos à despolimerização de biomassa. Uma análise dos níveis de expressão permitiu determinar os conjuntos de genes diferencialmente expressos em diferentes

condições de cultivo. Os resultados obtidos acrescentam conhecimento sobre a constituição do genoma, as atividades de expressão gênica do fungo *Trichoderma harzianum* e fornece informações importantes a respeito dos mecanismos genéticos de degradação de biomassa que o fungo utiliza. As informações obtidas poderão ser utilizadas para outras espécies de fungos filamentosos com potencial para a biodegradação.

Palavras-chave: biomassa, *Trichoderma*, fungos, biodegradação, enzimas hidrolíticas, CAZy, celulose, hemicelulose, transcriptoma, sequenciamento *high-throughput*, hidrólise enzimática, bagaço de cana-de-açúcar, glicosil hidrolases.

## SUMÁRIO

ABSTRACT .....	vii
RESUMO .....	ix
Lista de Figuras .....	xvii
Lista de Tabelas.....	xix
Lista de abreviações.....	xxi
Organização da Tese.....	xxiii
1.Introdução.....	1
2.REVISÃO BIBLIOGRÁFICA.....	7
2.1 Hidrólise Enzimática.....	10
2.2 Classificação das sequências relacionadas à atividade hidrolíticas.....	18
2.3 A Degradação Promovida por Fungos Filamentosos.....	21
2.4 Transcriptoma.....	29
3.OBJETIVO.....	33
3.1 Objetivo Geral.....	35
3.2 Específicos.....	35
4. CAPÍTULO 1 - <i>Fermentação, bibliotecas de cDNA e análise do transcriptoma</i> .....	37
4.1 Perfil de produção enzimática.....	39
4.2 Bibliotecas de cDNA.....	47
4.2.1 Sequenciamento dos clones .....	49
4.2.2 Análise das sequências obtidas a partir das bibliotecas de cDNA .....	50
4.3 Perfil do Transcriptoma e análise das bibliotecas.....	58
5. CAPÍTULO 2 - <i>Perfil do transcriptoma e análise dos níveis de expressão de enzimas relacionadas à biodegradação</i> .....	67
6. CONSIDERAÇÕES FINAIS .....	113
7. RESUMO DOS RESULTADOS .....	119
8. CONCLUSÕES.....	123
9. PERSPECTIVAS.....	127
10. BIBLIOGRAFIA.....	131



***Dedicatória***

***Dedico esta tese à comunidade científica brasileira  
que tanto se esforça  
para colocar o Brasil  
entre os melhores países do mundo.***



## **Agradecimentos**

Agradeço sinceramente à todos que me acompanharam e me incentivaram nestes anos de estudo e pesquisa.

Agradeço à professora Dra Anete Pereira de Souza, que sempre acreditou que poderíamos fazer um ótimo trabalho, e nunca mediu esforços para que tudo fosse feito da melhor maneira. Obrigado professora, além de genética, você me ensinou a fazer pesquisa.

Agradeço aos colegas de laboratório, que tornaram mais fácil esse trabalho, sempre compartilhando o conhecimento, até mesmo nas horas de descontração.

Agradeço à minha família, que, além de me apoiar, sempre me incentivou a ir cada vez mais longe, sem medir os obstáculos.

Agradeço à equipe do Centro de Tecnologia do Bioetanol (CTBE) pelas ideias e indagações que contribuíram bastante com este trabalho.

Agradeço à Fapesp pelo apoio financeiro.





## Lista de Figuras

Figura 1: Resíduo gerado de bagaço de cana-de-açúcar.....	10
Figura 2: Rotas tecnológicas para a produção de etanol.....	12
Figura 3: Estrutura da parede celular vegetal.....	14
Figura 4: Mecanismos de ação das celulases.....	15
Figura 5: Mecanismos de ação das hemicelulases.....	17
Figura 6: Monômeros precursores da lignina.....	18
Figura 7: Culturas de <i>Trichoderma harzianum</i> . A - linhagem T-22(KRL-AG2) (27) e B- linhagem IOC3844 (imagem própria) crescendo em meio batata-dextrose- agar.....	23
Figura 8: Características morfológicas de <i>Trichoderma harzianum</i> . 6–14: Conidiophores, 8 e 14 são atípicas;15: Conidia; 16: Chlamydospores. Escala: 6–9, 12–14, 16 = 25 µm; 10, 11, 15 = 10 µm.....	24
Figura 9: Microscopia eletrônica de transmissão de <i>T. harzianum</i> T-203 inoculado em tecidos da raiz de pepino. Um grande número de hifas fúngicas (T) desenvolvem-se na superfície da raiz. As hifas penetram na epiderme da raiz (Ep) progressivamente até a área cortical (CA), principalmente pelo crescimento intercelular (setas em A). Justaposições de parede (WA) são vistas em células hospedeiras abaixo das áreas colonizadas. A colonização da epiderme e córtex não está associado com alterações da célula hospedeira ou digestão da parede celular (C). Espaços intercelulares (IS), estela vascular (VS). Escala A, 10 m; B, 1,5 mm; C, 5 im.....	26
Figura 10: A- Concentração de proteína total (mg/L) por tempo de fermentação (horas) em BED. B- Perfil da atividade celulolítica (FPU) durante fermentação em BED para as amostras induzidas com BED, celulose e lactose. Os dados foram comparados segundo ANOVA. C: Atividade de beta-glicosidase (UI/ml) por tempo de fermentação (horas). D: Atividade total de xilanase (UI/ml) por tempo de fermentação.....	41
Figura 11: A : Eletroforese unidimensional de cDNA. amostras 1-padrão tamanho 1Kb, 2-BED, 3-celulose, 4-lactose, 5-controle. B: Seleção de colônias transformantes / clones positivos (brancos) e negativos (azuis) para a presença do inserto.....	49
Figura 12: Resultados da análise da biblioteca de cDNA, A- Classificação da anotação e B- classificação segundo similaridade das sequências com espécies.....	52
Figura 13: Classificação segundo a função molecular das sequências anotadas.....	53

Figura 14: Classificação segundo processo biológico das sequências anotadas.....	53
Figura 15: Classificação segundo componente celular das sequências anotadas.....	54
Figura 16: Classificação geral das sequências obtidas na análise do transcriptoma das bibliotecas BED, CEL e LAC segundo os termos do Gene Ontology.....	60
Figura 17: Classificação das sequências possivelmente envolvidas nas reações de degradação de biomassa obtidas através da análise do transcriptoma das bibliotecas BED, CEL e LAC segundo os termos do Gene Ontology.....	61
Figura 18: Distribuição das sequências de acordo com a função molecular.....	62
Figura 19: Distribuição das sequências de acordo com a função biológica.....	63
Figura 20: Distribuição das sequências segundo a localização celular.....	64
Figura 21: Fluxograma descritivo do protocolo de análise do transcriptoma.....	118

## Lista de Tabelas

Tabela 1: Principais enzimas envolvidas na degradação de lignocelulósicos.....	20
Tabela 2- Concentração de proteína total (mg/L) por tempo de fermentação (horas). Os dados foram comparados segundo ANOVA e correspondem ao valor médio $\pm$ desvio padrão.....	40
Tabela 3- Atividade celulolítica detectada e expressa FPU (UI/ml) durante a fermentação em BED das amostras induzidas por diferentes fontes de carbono: celulose, lactose e BED. Os dados foram comparados segundo ANOVA e correspondem ao valor médio $\pm$ desvio padrão; Os valores seguidos de diferentes letras são significantes para $p^* < 0,01$ . .....	44
Tabela 4 : <i>Reads</i> gerados após sequenciamento e testes de qualidade.....	58
Tabela 5: Montagem ( <i>assembly de novo</i> ). Reads: fragmentos analisados 1(DSB), 2 (CEL) 3 (LAC); matched: número reads que alinhou com os contigs; not matched: número de reads sem alinhamento; contigs: número de contigs formados no <i>assembly de novo</i> , reads in pairs: número de reads mapeadas com os respectivos pares; broken paired reads: reads que apresentam somente um par mapeado ou fora da orientação/ distância; a.v: tamanho médio (pb); total bases: número de pares de bases total. ....	59
Tabela 6: Genes que codificam enzimas das três principais classes de celulases amplificados a partir do DNA de <i>T. harzianum</i> IOC-3844. ....	66



## **Lista de Abreviações**

BED- Bagaço Explodido e Deslignificado

CEL- Celulose

LAC- Lactose

DSB- Bagaço Explodido e Deslignificado

FPA- Atividade de Papel Filtro

NGS- Next Generation Sequencing

E.C.- Enzyme Code

GO- Gene Ontology

CAZy - Carbohydrate-Active Enzymes Database

KEGG- Kyoto Encyclopedia of Genes and Genomes

GH- glicosilidrolase

GT- glicosiltransferase

CE-carboidrato-esterase

CBM- módulos de ligação à carboidrato



## Organização da Tese

Inicialmente, será apresentada uma Introdução Geral e uma Revisão Bibliográfica que compreende tópicos relacionados à hidrólise enzimática, classificação das sequências relacionadas à atividade hidrolítica, degradação promovida por fungos filamentosos e análise do Transcriptoma. A seguir será delineado o Objetivo Geral e Objetivos Específicos do trabalho.

Em seguida é apresentado o Capítulo 1, subdividido em: Perfil da produção enzimática, Bibliotecas de cDNA e Perfil do transcriptoma. Os resultados mostram dados sobre o experimento de fermentação, e descreve a abordagem inicial para a determinação de sequências hidrolíticas, as bibliotecas de cDNA. Também descreve a montagem *de novo* das bibliotecas e a análise individual de cada uma delas. O capítulo 2 contém a análise do transcriptoma do fungo em condições de degradação de biomassa, através do artigo "Transcriptome Profile of *Trichoderma harzianum* IOC- 3844 Induced by Sugarcane Bagasse", publicado na revista PlosOne. Neste capítulo são apresentados dados oriundo do sequenciamento *high-throughput*, o perfil de transcrição do fungo e a classificação das enzimas hidrolíticas relacionadas às reações de degradação, além de uma análise sobre os respectivos níveis de expressão.

São apresentadas algumas Considerações Finais, seguido de um breve Resumo dos Resultados, para então determinar as Conclusões. A seguir são apresentadas as Perspectivas para o prosseguimento dos estudos e a Bibliografia do trabalho.





# 1. INTRODUÇÃO



A utilização de microrganismos para a biotransformação da matéria acontece desde a antiguidade. Através da observação do ambiente, o ser humano percebeu que certos processos aconteciam devido à presença de microrganismos no meio. Os babilônios e os sumérios usavam leveduras na produção de álcool antes de 6000 AC (1). Já os egípcios, em 2000 AC, utilizavam leveduras para a produção de pães. Na Ásia, antes do nascimento de Cristo, utilizava-se *Penicillium rouquefortii* na produção de queijos. A 2500 anos atrás, na China, o fungo *Aspergillus oryzae* era usado no processo de fabricação de koji, que foi levado também para o Japão no século VII (2).

Na metade do século XIX, Luis Pasteur estudou a função de microrganismos na produção de vários produtos como alimentos fermentados, vinho, cervejas, queijo, leite, iogurte, combustíveis e química fina. Ele identificou muitos processos microbiológicos e descobriu um dos principais princípios da fermentação: a utilização de substratos por microrganismos para a produção de metabólitos primários e secundários de interesse ao homem (1). Através do isolamento de alguns microrganismos e cultivo em meio adequado, Pasteur deu início ao que hoje recebe o nome de biotecnologia (3).

Atualmente a biotecnologia pode ser entendida como a aplicação de organismos e de sistemas biológicos na produção de bens e serviços (4). Tradicionalmente, a aplicação da biotecnologia na indústria de alimentos se restringia à produção de pães, queijos, bebidas, álcool, vinagre e iogurte. Mais recentemente, houve um aumento do interesse pelo uso dessa tecnologia na extração e produção de ingredientes não nutritivos, biologicamente ativos, como os biofertilizantes, biocombustíveis e a biorremediação. Somente nas última

décadas, impulsionado por uma demanda por processos mais eficientes e menos poluentes, foram observados progressos significativos nas técnicas de produção bioprodutos, por fermentação, por processos enzimáticos e por engenharia genética a partir de sistemas biológicos derivados do DNA recombinante (rDNA) (5).

Durante a Primeira Guerra Mundial iniciou-se a utilização de microrganismos para a produção de substâncias como etanol, acetona e ácido cítrico. Durante esse período foi realizado pela primeira vez um cultivo microbiológico asséptico em larga escala, quando Chaim Weizmann usou um fermentador para a produção de acetona por *Clostridium acetobutylicum* (6). A partir da Segunda Guerra Mundial, os microrganismos tiveram importância na produção em larga escala de antibióticos. (7), e desde então muitos bioprocessos e microrganismos tem sido explorados.

A produção de álcool combustível por fermentação é um exemplo de processo biotecnológico empregado em larga escala. Atualmente, uma solução formada por açúcares fermentescíveis é o substrato para a levedura que produz o álcool. Porém, a biomassa formada depois da extração destes açúcares (resíduos de cana-de-açúcar, milho, trigo, dendê entre outros) também é rica em energia, que pode ser transformada em etanol, o chamado etanol de segunda geração. O primeiro grande desafio para o etanol de segunda geração é converter a biomassa, cuja estrutura lignocelósica precisa ser rompida para o acesso aos açúcares.

Como o tratamento químico da biomassa pode impactar negativamente na qualidade do processo, o tratamento via hidrólise enzimática tem sido uma alternativa bastante promissora. Neste caso, um conjunto de enzimas, denominado de coquetéis enzimáticos, são aplicados na biomassa, catalizando reações de hidrólise e tornando-os metabolizáveis pelas leveduras produtoras de álcool. Estas enzimas devem possuir características como alta especificidade, alta taxa de conversão, alta estabilidade, baixo custo e capacidade de atuar em sinergismo, ou seja, essas enzimas devem atuar concomitantemente, em um substrato complexo dando origem a um composto que pode ser a base para a geração dos chamados biocombustíveis.

Buscando contribuir para o desenvolvimento da hidrólise enzimática da biomassa, o presente trabalho foi buscar no transcriptoma de um microrganismo, especificamente um fungo filamentos, os mecanismos genéticos utilizados para a biodegradação.

O fungo *Trichoderma harzianum*, que apresenta comprovado potencial de biocontrole e biodegradação é uma linhagem pública, disponibilizada pelo Instituto Oswaldo Cruz, unidade do Rio de Janeiro/RJ e pelo CPQBA (Campinas, São Paulo). Dados genéticos e genômicos gerados para a linhagem IOC3844 são de grande utilidade para os diferentes grupos que queiram contribuir para os estudos de regulação da expressão gênica, descoberta de novos genes e mecanismos de degradação de diferentes substratos utilizando esta linhagem como modelo.

Nada mais lógico do que entender um processo natural altamente eficiente, efetuado pelo *Trichoderma harzianum*, para então poder reproduzi-lo no processo de hidrólise enzimática para a produção de etanol. Qual o conjunto de enzimas que este fungo produz responsáveis pelas reações de hidrólise? Esta resposta foi obtida na presente tese e será detalhada nas próximas seções. Este trabalho forneceu informações importantes sobre os genes expressos em condições de degradação de biomassa e o conjunto de enzimas que atuam eficientemente na biomassa lignocelulósica.

# 1. REVISÃO BIBLIOGRÁFICA





Petróleo, gás natural e seus derivados representam 55% do consumo mundial de energia. São esses combustíveis que permitem a existência dos meios de transporte rápidos e eficientes que temos hoje, bem como boa parte das atividades industriais (8).

É preciso encontrar alternativas para esses combustíveis. Nada mais racional do que produzi-los com base em matéria orgânica renovável (biomassa), da qual, no passado distante, os combustíveis fósseis foram produzidos pela natureza.

No Brasil, o etanol, produzido da cana-de-açúcar, já substitui hoje metade da gasolina que seria consumida e seu custo é competitivo sem os subsídios que viabilizaram o programa no seu início. Isso foi conseguido em aproximadamente 30 anos, a partir de ações voltadas para a produção de álcool mais eficientemente, tais como a criação do Proálcool, programa lançado no país em meados da década de 1970 para reduzir a dependência da importação de petróleo e, no fim da década de 90, o desenvolvimento de um programa de pesquisa voltado para o estudo do genoma expresso da cana-de-açúcar (SUCEST, FAPESP). Os programas de pesquisa e desenvolvimento, resultaram em novos cultivares de cana, mais produtivos e mais adaptados ao Brasil, bem como processos fermentativos mais econômicos e rentáveis (8) (9) (10).

Embora todos esses esforços já tenham sido feitos no passado, sabe-se que um aumento considerável na produção de álcool no Brasil só poderá ser conquistado pelo aproveitamento do açúcar proveniente de biomassa, como a o bagaço de cana (8). Estudos mostram que a produção de etanol por hectare de cana, que atualmente é de 6000 L/ha, pode chegar a 10,000 L/ha, se 50% do bagaço produzido for convertido em etanol (11).



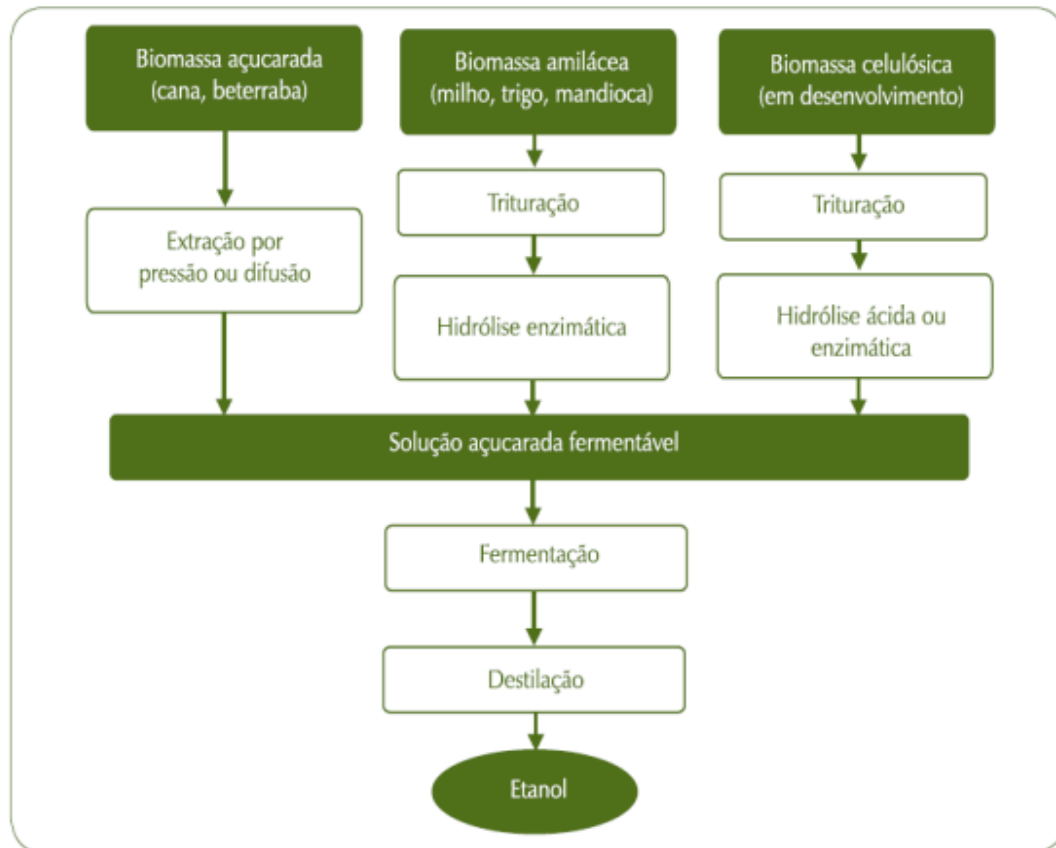
Figura 1 : Resíduo gerado de bagaço de cana-de-açúcar. Fonte: <http://knowtec.com/>

## 2.1- Hidrólise Enzimática

Atualmente as tecnologias industriais para conversão de açúcares e amido em bioetanol, podem ser consideradas maduras e disponíveis, exceto as referentes à hidrólise de materiais lignocelulósicos, ora em desenvolvimento, adotadas para a obtenção de etanol da parte celulósica da cana (8).

Mediante rotas biológicas, o etanol pode ser produzido com base em qualquer biomassa que contenha quantidades significativas de amido ou açúcares. A figura 2 sintetiza as rotas tecnológicas para produção de bioetanol, considerando as diferentes matérias-primas de interesse. Hoje em dia, há um leve predomínio da produção com base em materiais amiláceos (53% do total),

como milho, trigo e outros cereais e grãos. Para esses casos, a tecnologia de conversão geralmente começa com a separação, a limpeza e a moagem do grão. A moagem pode ser úmida, quando o grão é embebido e fracionado antes da conversão do amido a açúcar (via úmida), ou seca, quando isso é feito durante o processo de conversão (via seca). Em ambos os casos, o amido é tipicamente convertido em açúcares por meio de um processo enzimático a altas temperaturas. Os açúcares liberados são, então, fermentados por leveduras, e o vinho resultante é destilado para a purificação do etanol. Além do etanol, esses processos envolvem, geralmente, diversos coprodutos, que variam conforme a biomassa utilizada (8) (11).



Fonte: Elaboração de Luiz Augusto Horta Nogueira.

Figura 2- Rotas tecnológicas para produção de etanol à partir de biomassa (8).

Considerando a produção com base em açúcares, como é o caso da cana e da beterraba, o processo é mais simples e envolve uma etapa a menos, já que os açúcares estão disponíveis na biomassa. Normalmente, o processo se baseia na extração dos açúcares (por meio da moagem ou da difusão), que podem seguir diretamente para a fermentação. Após a fermentação, o vinho resultante é destilado, assim como no caso da produção com base no amido

Como hoje apenas um terço da biomassa contida na cana é aproveitado para a produção de açúcar e de etanol, o grande desafio é transformar a celulose, que

está no bagaço e na palha descartada na colheita desta gramínea, em álcool combustível, e produzir o chamado etanol de segunda geração.

Muitos estudos pretendem viabilizar a produção de álcool via bagaço e palha da cana, e fazem parte do Projeto Bioetanol, que tem como um dos objetivos desenvolver a hidrólise enzimática no Brasil, uma das vias para obtenção do etanol. Dessa forma os pesquisadores esperam obter o etanol por via enzimática utilizando celulasas, enzimas produzidas por microorganismos capazes de quebrar a molécula de celulose em moléculas menores de açúcares que podem ser transformados em álcool combustível após o processo de fermentação.

Para tanto é necessário a desestruturação da parede celular vegetal, uma estrutura rígida e complexa composta principalmente de celulose, hemicelulose e lignina, para uma seguinte depolimerização das cadeias de celulose e hemicelulose para a obtenção de monômeros que podem ser metabolizados pelas leveduras produtoras de etanol (Figura 3).

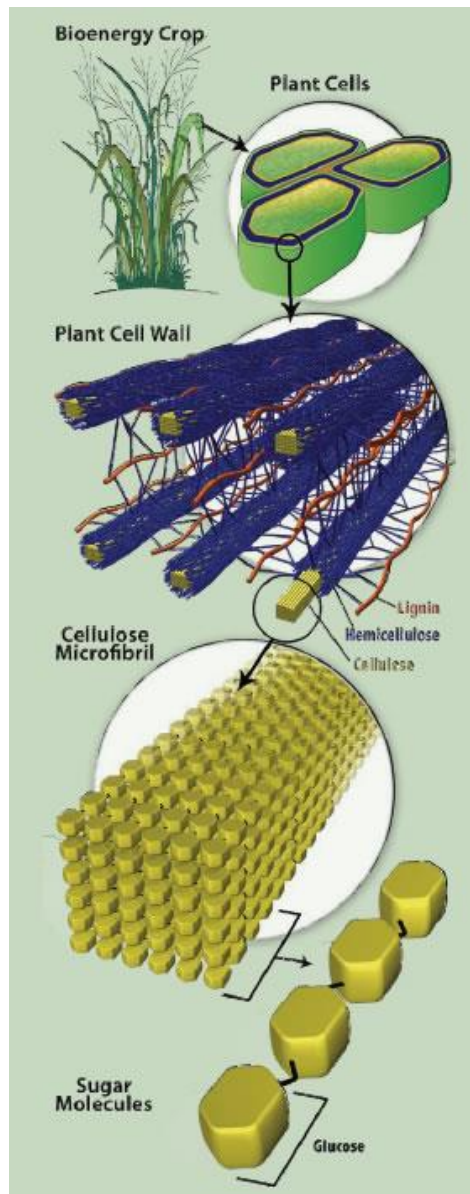


Figura 3: Estrutura da parede celular vegetal. Fonte: <http://www2.lbl.gov/Publications/YOS/Feb/>

Enquanto a conversão de amido em açúcar requer apenas uma enzima, a hidrólise de celulose envolve um complexo arranjo de enzimas que trabalham sinergicamente, as chamadas celulasas (Figura 4).

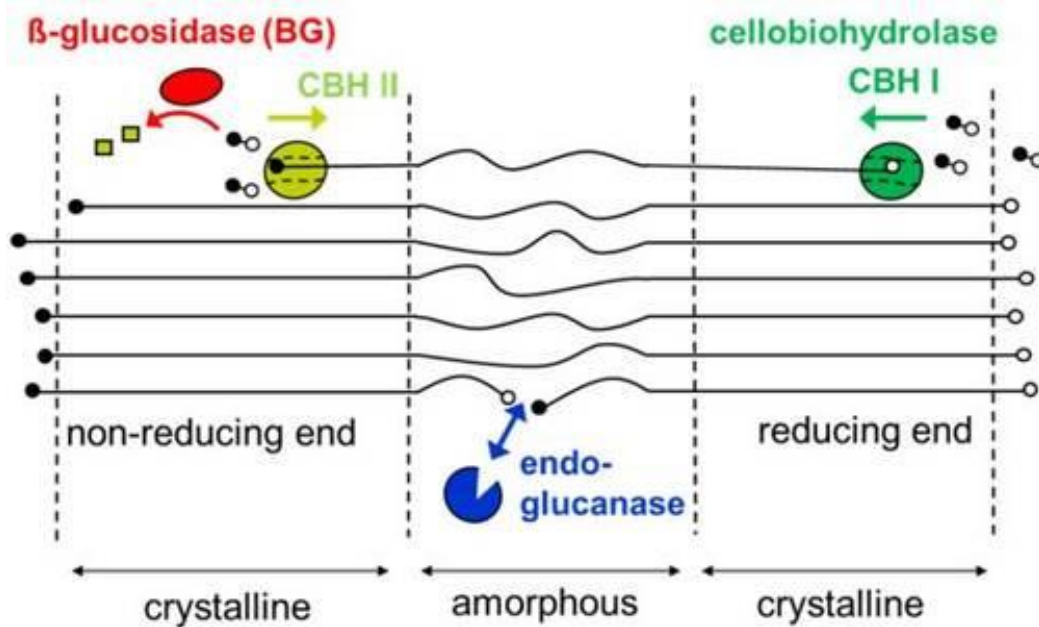


Figura 4: Ação das celulasas. Fonte: <http://www.avt.rwth-aachen.de/AVT/typo3temp/pics/5ea64e8d21.jpg>

As celulasas são normalmente classificadas com base na sua sequência de aminoácidos e na sua especificidade, sendo possível dividi-las em três grupos: (a) endoglucanases ou endo-1,4- $\beta$ -D-glucanases (EC 3.2.1.4), (b) exoglucanases ou celobiohidrolases (EC 3.2.1.91) e (c)  $\beta$ -glucosidases ou  $\beta$ -glucosideglucoidrolases (EC 3.2.1.21) (12).

As exoglucanases quebram progressivamente a celulose em cadeias menores a partir dos terminais redutores e não redutores para liberação de glicose e celobiose (dissacarídeo que resulta da hidrólise parcial da celulose) solúveis. Endoglucanases hidrolisam randomicamente pontes intramoleculares acessíveis de cadeias  $\beta$ -1,4-glicosídica proporcionando múltiplos novos terminais livres para a ação das exoglucanases. A hidrólise é finalizada com a ação das  $\beta$ -glucosidases que hidrolisam celobiose em glicose. A ação dessas enzimas é

cooperativa. A maioria das enzimas quando atuam individualmente não promovem a hidrólise completa e efetiva, embora sejam conhecidas algumas poucas celobioidrolases que consigam degradar completamente celulose cristalina em glicose (8) (12).

Já as hemiceluloses são heteropolissacarídeos formados por vários resíduos de açúcares pentoses (xilose e arabinose) e hexoses (glicose, manose e galactose), ácidos urônicos e grupos acetila. Esses açúcares estão ligados entre si, principalmente por ligações glicosídicas  $\beta$ -1,4, formando uma estrutura principal composta por um tipo específico de resíduo, a partir da qual surgem ramificações laterais de cadeias curtas de outros compostos. As hemiceluloses são classificadas de acordo com o açúcar predominante na cadeia principal e na ramificação lateral.

As principais hemiceluloses encontradas em plantas são os xiloglucanos (XyG), os glucuronoarabinoxilanos (GAX) e os mananos (MN). Em todos os casos, há uma cadeia principal de monossacarídeos de glicose, xilose e manose, respectivamente, que pode ser ramificada com diferentes monossacarídeos.



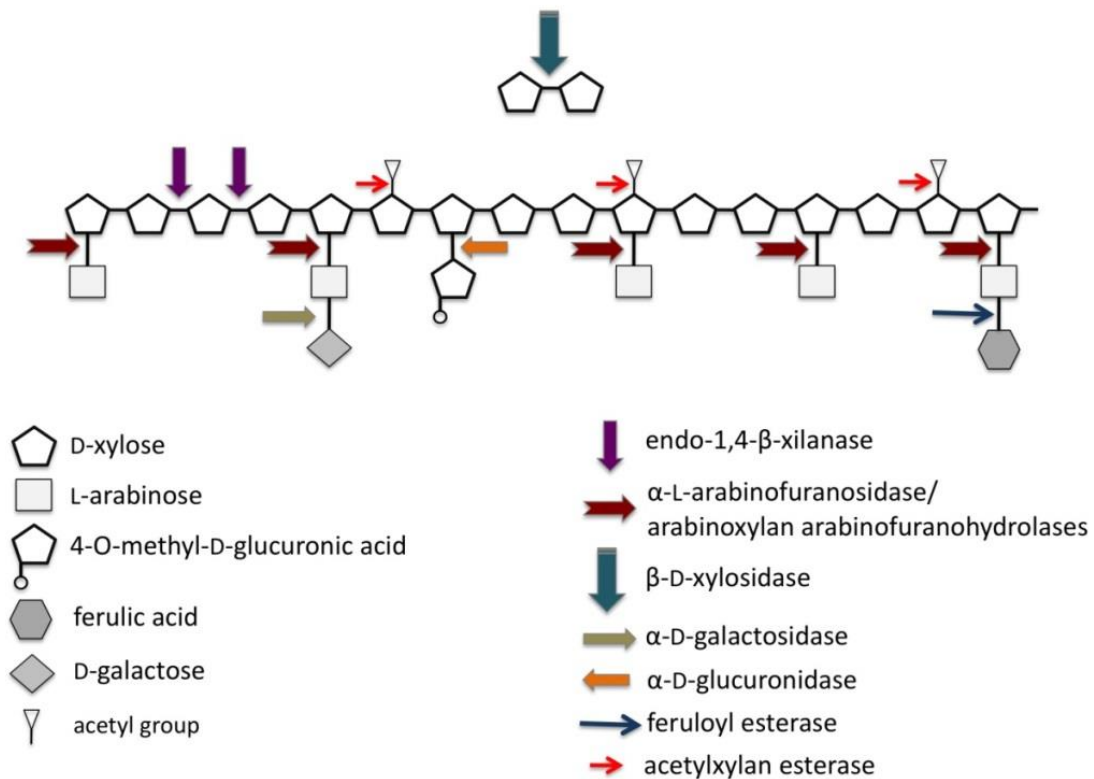


Figura 5: Mecanismo de ação das hemicelulases, degradação de arabinoxilana. As setas representam a atividade enzimática para o respectivo substrato. Fonte: <http://www.intechopen.com/source/html/44319/media/image4.jpeg>

Outro biopolímero presente nas plantas, a lignina, pode representar até 25% de toda a biomassa lignocelulósica produzida no planeta e seu teor nos resíduos vegetais pode atingir até 40% do seu peso seco. A lignina está concentrada em tecidos relacionados com condução de solutos e suporte mecânico e representa um conjunto de polímeros amorfos, de alto peso molecular e muitas ligações cruzadas. Possui natureza química bem distinta dos carboidratos, sendo caracterizada por uma estrutura aromática de natureza eminentemente fenólica. As unidades monoméricas precursoras da lignina são hidroxilas fenólicas dos

alcoóis trans-p-cumarílico, trans-coniferílico e álcool trans-sinapílico, cuja representação química pode ser observada na Figura 6.

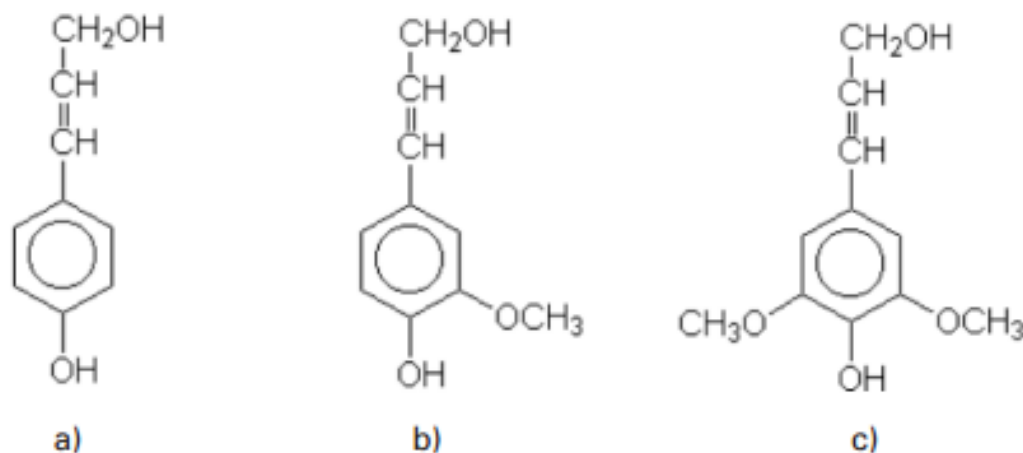


Figura 6: Monômeros precursores da lignina (a) álcool trans-para-cumárico, (b) álcool trans-conefírico e (c) álcool trans-sinapílico.

Devido à natureza fenólica da lignina, essa fração não pode ser diretamente convertida em etanol. A lignina tem sido usada para a produção de energia pelo processo de combustão (8) ou para a conversão em outros bioprodutos de interesse comercial.

## 2.2 Classificações das sequências relacionadas à atividade hidrolítica

Durante o processo enzimático, a hidrólise da biomassa é catalisada por uma grande variedade de enzimas, com diferentes especificidades, necessárias para a degradação dos componentes dos materiais lignocelulósicos. Muitos trabalhos exploram a cooperação sinérgica de diferentes enzimas para a bioconversão de lignocelulósicos (13) (14). A tabela 1 mostra as principais enzimas necessárias para a degradação dos componentes principais de

lignocelulósicos a monômeros. Ainda assim, existem evidências de que muitas outras proteínas podem contribuir para a degradação de lignocelulose, como as glicosilhidrolases da família 61, expansinas e swoleninas (12) (15) (16) (17).

A tabela 1 mostra a classificação das sequências segundo o código E.C. (Enzyme Code), que é uma classificação numérica para enzimas baseadas nas reações que catalisam, organizada pelo *EMBL-EBI* (The European Bioinformatics Institute). Outra classificação bastante utilizada é aquela segundo "*Carbohydrate-Active Enzymes Database*" (CAZy), um banco de dados que descreve famílias baseadas na estrutura catalítica e em módulos de ligação com carboidratos (ou domínios funcionais) de enzimas que degradam, modificam ou criam ligações glicosídicas. Assim, através destas duas classificações é possível determinar exatamente qual tipo de reação é catalisada pelas referidas enzimas.

A seleção das famílias de enzimas relacionadas às reações que promovem a degradação de biomassa pode ser realizada através da identificação das sequências presentes no transcriptoma. A identificação de sequências por similaridade pode apontar genes ou proteínas homólogas, através da detecção de sequências similares, que refletem em um ancestral comum (18).

**Tabela 1:** Principais enzimas envolvidas na degradação de lignocelulósicos (12).

<b>Componente do material lignocelulósico</b>	<b>Enzima ativa</b>	<b>NumeroEnzymeCode</b>
<b>Celulose</b>	Exo-1,4- $\beta$ -glucanases	3.2.1.91 e 3.2.1.176
	Endo-1,4- $\beta$ -glucanases	3.2.1.4
	$\beta$ -glucosidases	3.2.1.21
<b>Hemiceluloses</b>	Endo-xilanases	3.2.1.32
	$\beta$ -xilosidases	3.1.2.37 e 3.2.1.72
	Endo-mannanase	3.2.1.78 e 3.2.1.101
	$\beta$ -manosidase	3.2.1.25
	$\alpha$ -L-arabinofuranosidases	3.2.1.55
	$\alpha$ -glucuronidase	3.2.1.139
	Ferulicacidesterase	3.1.1.73
	$\alpha$ -galactosidase	3.2.1.22
	Feruloylsterase	3.1.1.73
	Acetilxilanesterase	3.1.1.72
	p-coumaricacidesterase	
<b>Lignina</b>	Lacase	1.10.3.2
	Manganes peroxidase	1.11.1.13
	Lignina peroxidase	1.11.1.14
<b>Pectina</b>	Pectina metil esterase	3.1.1.11
	Pectina liase	4.2.2.10
	Polygalacturonase	3.2.1.15
	Rhamnogalacturonanlyase	4.2.2.23

O conhecimento sobre os genes e regiões genômicas tem o potencial de tornar o processo de produção do etanol economicamente praticável. Atualmente há um grande interesse em se reduzir os custos para a produção de celulases já conhecidas e aumentar a sua eficiência, bem como buscar novas enzimas e coquetéis enzimáticos mais específicos e estáveis, que possam ser aplicados no processo de hidrólise enzimática para conversão da biomassa celulósica.

### **2.3 A Degradação Promovida por Fungos Filamentosos**

Os fungos são os organismos mais eficientes na produção de celulases extracelulares, sendo que alguns foram extensivamente estudados e utilizados para aplicações industriais como na indústria de alimentos, na agroindústria, ração animal, na produção de papel, indústria têxtil, tratamento de dejetos, detergentes e nas indústrias químicas (19) (20). Em estudos anteriores, o uso de celulases de origem fúngica já foi testado para a melhoria da produtividade de etanol utilizando-se materiais lignocelulósicos. Porém estas enzimas demonstraram ser lentas e muitas vezes instáveis, sendo assim os esforços se concentraram em implementar a performance das enzimas já conhecidas. Desta forma optou-se por desenvolver enzimas mais estáveis a variações de temperaturas e pH através de mutagênese (evolução dirigida ou *design* racional), fazendo combinações de diversas enzimas para novas rotas de degradação, otimizando essas enzimas para determinados pré-tratamentos ou procurando por novas enzimas específicas a objetivos concretos de produção (21).

Para o futuro a grande promessa é a construção de um organismo capaz de exercer dupla função, a de expressar uma ou várias enzimas especialmente desenhadas, com a capacidade de degradar a celulose e ao mesmo tempo utilizar os açúcares produzidos pela biodegradação. Para que se alcance essa perspectiva é importante entender os mecanismos genéticos relacionados às enzimas capazes de promover um aumento na conversão de biomassa do processo de biodegradação da celulose.

Atualmente as enzimas celulolíticas mais estudadas são encontradas nos fungos filamentosos *Trichoderma reesei* (*Hypocrea jecorina*) e *Aspergillus niger*, as quais já são utilizadas em diversos processos industriais. Esses organismos também são fonte da maioria das preparações (misturas complexas de enzimas) empregadas na conversão de biomassa em etanol (23). Tais fungos já possuem seus genomas sequenciados e análises preliminares destes genomas sugerem a presença de uma ampla gama de enzimas (200 em *T. reesei* e 170 em *A. niger*) envolvidas na degradação de polissacarídeos (23).

*Trichoderma* é um gênero de fungo apto a colonizar diferentes substratos em diferentes condições ambientais. A figura 7 mostra características morfológicas macroscópicas da espécie *Trichoderma harzianum*. Já a figura 8 mostra suas características morfológicas microscópicas.

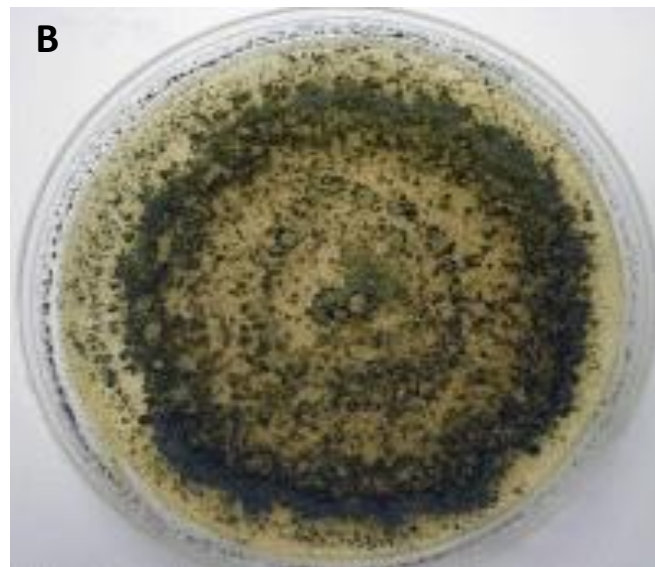


Figura 7: Culturas de *Trichoderma harzianum*. **A** - linhagem T-22(KRL-AG2) (27) e **B**- linhagem IOC-3844 (imagem própria) crescendo em meio batata-dextrose-agar. A área branca não contém esporos enquanto a área verde contém densa massa de esporos.

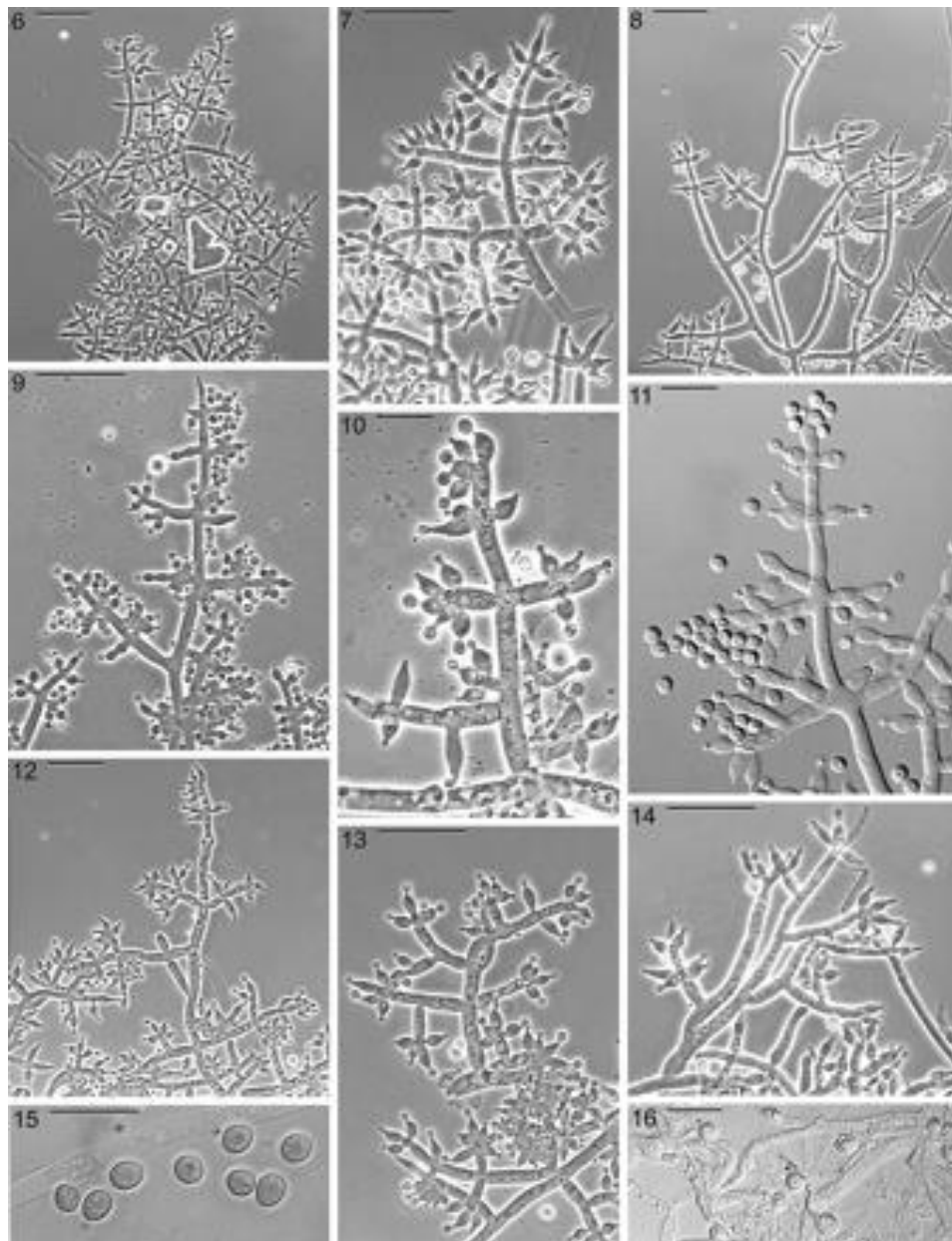


Figura 8: Características morfológicas de *Trichoderma harzianum*. 6–14: Conidiophores, 8 e 14 são atípicas; 15: Conidia; 16: Chlamydospores. Escala: 6–9, 12–14, 16 = 25  $\mu\text{m}$ ; 10, 11, 15 = 10  $\mu\text{m}$ . Fonte: Samuels, 2002(28).

Um dos mais significantes nichos ecológicos ocupados pelo *Trichoderma* é a rizosfera das plantas que é efetivamente colonizada através da capacidade desse



fungo de interagir com as plantas e competir com os microrganismos do solo. Essa habilidade é o resultado de um longo período de evolução no qual desenvolveram-se mecanismos biológicos para ataque de outros microrganismos e para o favorecimento do crescimento da planta. Esse complexo mecanismo pode ser visualizado nas imagens apresentadas na figura 9 (24), onde nota-se a penetração das hifas no tecido vegetal e a formação de calosidades, que pode estar relacionada com a ativação do sistema de defesa e crescimento da planta.

Assim, o potencial de biocontrole do *Trichoderma* depende da sua diversidade metabólica e de seu potencial secretor, que é responsável pela produção de grandes quantidades de diversas enzimas hidrolíticas envolvidas na degradação de paredes celulares (25) (26).

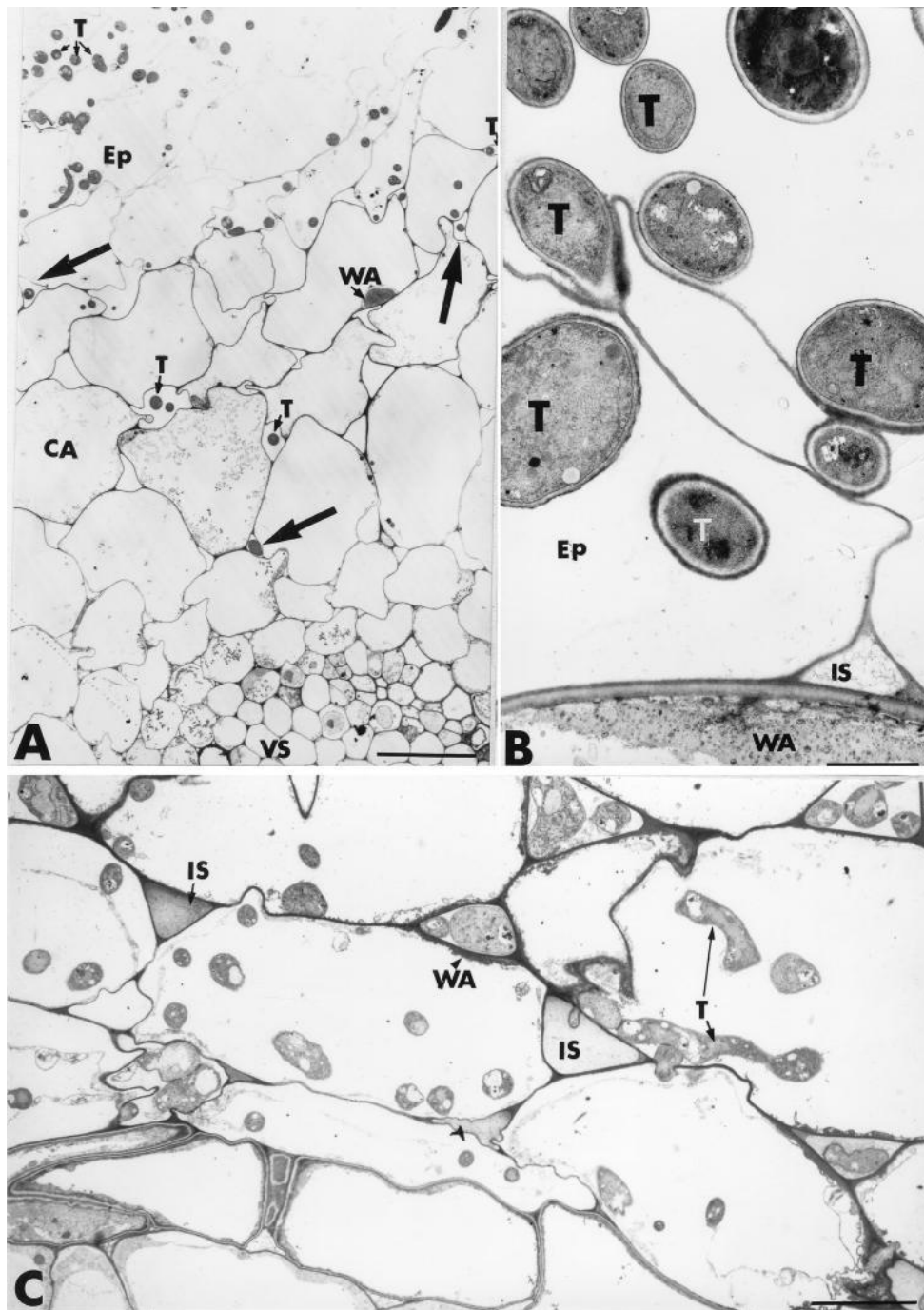


Figura 9: Microscopia eletrônica de transmissão de *T. harzianum* T-203 inoculado em tecidos da raiz de pepino. Um grande número de hifas fúngicas (T) desenvolvem-se na superfície da raiz. As hifas penetram na epiderme da raiz (Ep) progressivamente até a área cortical (CA), principalmente pelo crescimento

intercelular (setas em A). Justaposições de parede (WA) são vistas em células hospedeiras abaixo das áreas colonizadas. A colonização da epiderme e córtex não está associado com alterações da célula hospedeira ou digestão da parede celular (C). Espaços intercelulares (IS), estela vascular (VS). Escala: A, 10  $\mu$ m; B, 1,5  $\mu$ m; C, 5  $\mu$ m. Fonte: Yedidia, 1999 (24).

O genoma do *Trichoderma*, apesar de seu alto valor biotecnológico, tem sido pouco explorado comparado com outros microrganismos. Um projeto de genoma estrutural promoveu a primeira versão da sequência completa do genoma de *T. reesei*, um isolado sem habilidade de biocontrole, mas com interesse industrial (29).

Estudos demonstram que linhagens do fungo *Trichoderma harzianum* apresentam atividades enzimáticas que hidrolisam diferentes tipos de carboidratos (30) (31) (32). Este fungo, assim como o bem estudado *Trichoderma reesei*, expressa uma diversidade de celulasas.

Embora não exista um número tão elevado de publicações envolvendo o estudo de celulasas e hemicelulasas de *T. harzianum* (quando comparado a *T. reesei*, sendo que *T. reesei* foi à primeira espécie onde notou-se a presença desta classe enzimática), estudos recentes (26) (33) (34) (35) têm demonstrado a grande eficiência de *T. harzianum* (especialmente a linhagem brasileira IOC-3844, alvo deste projeto) com relação à hidrólise de materiais lignocelulósicos.

Em termos de produtividade volumétrica e atividade específica, a linhagem IOC-4038 provou produzir altas quantidades de celulasas, mesmo quando comparadas com a linhagem padrão de uso industrial *T. reesei* Rut C30. A linhagem IOC-4038 produziu maiores quantidades de FPAse, Endoglucanase e beta-glucosidase que *T. reesei* Rut C30. O extrato enzimático bruto demonstrou

características apropriadas para sua aplicação na hidrólise da celulose, bem como alta estabilidade termal a 50°C e atividade de xilanase (26).

Algumas celulasas de linhagens de *T. harzianum* já foram isoladas ou seus genes clonados e expressos para estudar sua atividade hidrolítica (36) (37) (38). A linhagem tipo C4 possui quatro endoglucanases, duas celobiohidrolases e uma  $\beta$ -glucosidase, sendo que a maior parte da atividade celulolítica se deve a ação combinada de todas essas enzimas juntas, contudo, a  $\beta$ -glucosidase desempenha um papel central na hidrólise (37). Já a produção de  $\beta$ -xilanases,  $\beta$ -xilosidases e  $\alpha$ -arabinofuranosidases foi realizada em meios de cultura sólido, usando fontes de carbono complexas como o bagaço de cana, e em meios de cultura líquidos e fontes de carbono quimicamente definidas como xilana, mostrando alta atividade enzimática e estabilidade a temperaturas que alcançam 50°C e pH 5,0 (39) (40) (41).

A complexidade do secretoma de *T. harzianum* T4 e suas variações em resposta a diferentes fontes de carbono (glicose, celulose, xilana, bagaço de cana) foram exploradas por meio da análise proteômica dos filtrados obtidos do cultivo do fungo. Observou-se uma diferença notável na colonização do substrato, nas atividades enzimáticas e nos perfis 2D do secretoma. As amostras provenientes do ensaio com bagaço de cana foram as que apresentaram as maiores atividades enzimáticas. Ensaio de espectrometria de massas identificaram duas proteínas associadas ao metabolismo de carboidratos e glicoproteínas (41).

Um trabalho recente (42) estudou a estrutura protéica de celobiohidrolases da família das glicosilhidrolases (GH7) de diferentes fungos degradadores de biomassa, incluindo o *T. harzianum*, e determinou diferenças na estrutura dos domínios catalíticos. Essas enzimas possuem semelhanças na estrutura

tridimensional, porém apresentam diferenças no comprimento e na sequência de *loops* ao longo do domínio de ligação à celulose. Essas diferenças podem variar a acessibilidade dos sítios ativos e afetar a dinâmica dos movimentos dos *loops*, influenciando a interação enzima-substrato, a inibição pelo produto e as reações de iniciação e de liberação do produto (42). Isso explica os diferentes parâmetros enzimáticos entre as enzimas chaves para a degradação de biomassa de diferentes microrganismos (42).

Essas diferenças podem ser detectadas através do estudo das regiões do genoma que codificam tais domínios catalíticos, e que portanto, interferem na expressão gênica. Somente com o conhecimento das sequências genômicas associadas aos genes de interesse, é possível ter-se uma abordagem eficiente para o estudo da regulação da expressão dos genes de interesse nesse trabalho.

Outro estudo interessante (43) mostrou como a atividade celulolítica dos domínios anquirina (*ankyrin-cellulase scaffolding*) interferem na estabilidade de região consenso, termoestável. Assim um domínio consenso anquirina pode ser manipulado para promover um dobramento no formato da proteína, que torna enzimas celulasas ativas. Estes esqueletos de consenso de anquirina podem ser utilizados como uma plataforma versátil para combinar os domínios de celulase de diversas reações. Tal abordagem pode fornecer informações sobre como celulasas podem se comportar quando ligadas a outros domínios encontrados na natureza, tais como *dockerins* e módulos de ligação de celulose (43).

## 2.4 Transcriptoma

O transcriptoma é o conjunto completo dos transcritos (RNAs) em uma célula, para um estágio específico de desenvolvimento ou condição fisiológica.

Conhecer e compreender o transcriptoma é um passo essencial para interpretar os elementos funcionais do genoma, revelar os constituintes moleculares de células e tecidos, e conhecer os mecanismos moleculares de desenvolvimento e adaptação (44).

Os principais objetivos de análises transcriptômicas são: catalogar todas as espécies de transcritos, incluindo mRNAs, RNAs não codificantes e *small* RNAs; determinar a estrutura transcrita dos genes, em termos de sítios de iniciação, 5´ e 3´ ends; determinar *splicing* alternativo e outras modificações pós-transcricionais. A análise do transcriptoma também permite quantificar as mudanças nos níveis de expressão de cada transcrito em determinadas condições (44).

Recentemente, o desenvolvimento de novos métodos de sequenciamento de DNA *high-throughput*, ou *next generation sequencing (NGS)* possibilitaram tanto o mapeamento como a quantificação de um número muito superior de sequências. Esses métodos têm sido amplamente utilizados, principalmente devido ao menor custo na obtenção de genomas e transcriptomas. Denominado RNA-seq tais métodos apresentam vantagens em relação aos métodos já existentes e devem revolucionar a maneira com a qual os transcriptomas eucariotos são analisados. Os avanços nesta área, tornaram possível a análise de RNA-seq do fungo *Trichoderma harzianum* durante o desenvolvimento desta tese, onde foram analisadas 14,7 Gpb de sequências transcritas.

Apesar dos trabalhos existentes que comprovam o potencial de produção de enzimas degradativas da linhagem *T. harzianum* IOC3844, faltam estudos genômicos desta linhagem, bem como de qualquer outra de *T. harzianum*. Não há dados suficientes disponíveis na literatura que permitem uma análise complexa do genoma desse microrganismo, para se elucidar as vias de regulação

e expressão de genes relacionados às enzimas com atividade degradativa. Desta forma o trabalho realizado nesta tese, foi pioneiro na obtenção de dados genômicos para *T. harzianum* na extensão alcançada pelo sequenciamento NGS realizado, bem como pela análise do transcriptoma de *Trichoderma harzianum* efetuada.





### 3. OBJETIVO



### **3.1 Objetivo Geral**

Verificar o perfil de expressão das enzimas de *Trichoderma harzianum* (linhagem IOC-3844) que atuam sobre material lignocelulósico, principalmente celulose e hemicelulose, por meio do seu transcriptoma.

### **3.2 Objetivos Específicos**

- I. Realizar a fermentação em diferentes fontes de carbono (celulose, lactose e bagaço de cana).
- II. Verificar as atividades enzimáticas presentes no extrato;
- III. Determinar todos os genes expressos e seus respectivos níveis de expressão;
- IV. Analisar a expressão diferencial de genes relacionados à atividade degradativa;



## 4. CAPÍTULO 1:

*Fermentação, bibliotecas de cDNA e análise do transcriptoma.*



#### 4.1 Perfil de produção enzimática

Para comprovar o potencial de produção de enzimas hidrolíticas foram determinadas as atividades de FPA, (Filter Paper Activity), xilanase, proteínas totais e beta-glicosidase dos extratos aquosos. O micélio que foi produzido durante a fermentação foi utilizado para extração de DNA e RNA e armazenado a  $-80^{\circ}\text{C}$  para extração do mRNA.

A atividade em papel filtro (FPA) foi determinada como descrito por Ghose (1987) (45), a atividade de xilanase foi determinada como descrito por Bailey e Poutanen (1989) (46). A atividade de beta-glicosidase foi medida utilizando p-nitrophenol-b-D-glucoside (Sigma–Aldrich, USA), de acordo com Zhang (2009) (47). As proteínas totais foram mensuradas em microplacas utilizando kit específico da Bio-Rad (BioRad Laboratories, USA), através de um procedimento que se baseia no método de Bradford (48).

Já está bem estabelecido que as diferentes condições de cultura afetam significativamente a produção de celulasas e hemicelulasas. Os carboidratos e seus derivados induzem a maioria das enzimas celulolíticas. Assim a fonte de carbono desempenha um papel importante para a produção de enzimas (49). No presente trabalho foram utilizadas diferentes fontes de carbono como indutores de crescimento do micélio (BED, celulose e lactose). Os indutores utilizados na fase de crescimento promovem a expressão de diferentes conjuntos de genes uma vez que o fungo procura se adaptar aos diferentes meios.

Mesmo em condições semelhantes de fermentação é detectável a influência do indutor, como observado nos gráficos apresentados a seguir (Figura 10, A-D).

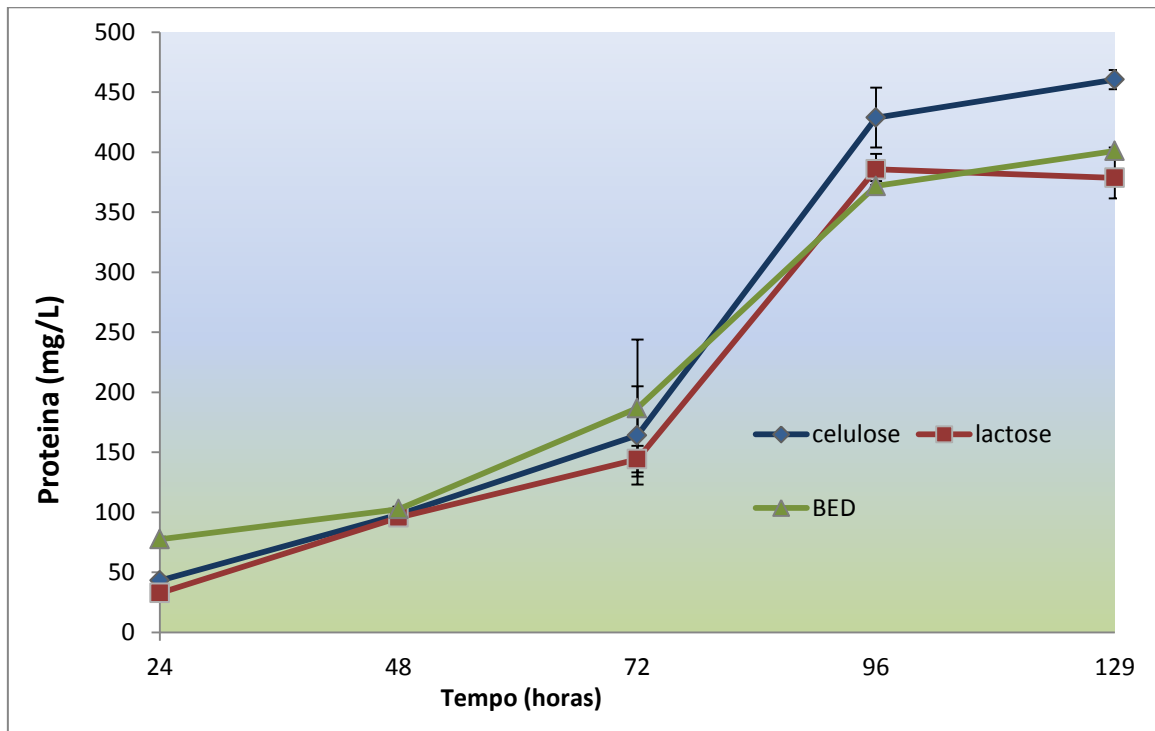
O monitoramento da concentração de proteínas permitiu observar um perfil crescente ao longo do tempo (Tabela 2, Figura 10 A).

Tabela 2- Concentração de proteína total (mg/L) por tempo de fermentação (horas). Os dados foram comparados segundo ANOVA e correspondem ao valor médio  $\pm$  desvio padrão. Os valores seguidos de diferentes letras são significantes para  $p^* < 0,01$  e  $p < 0,05$ .

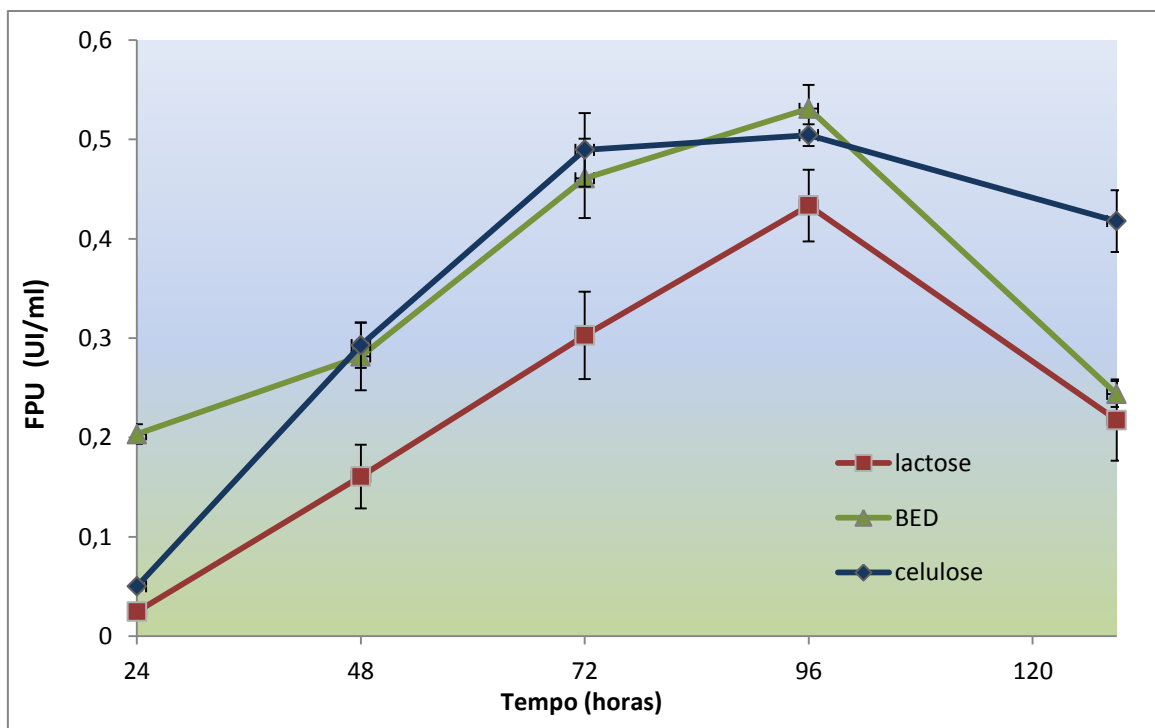
<b>Proteína total (mg/L)</b>				
<b>Tempo</b> (h)	<b>Bed</b>	<b>Celulose</b>	<b>Lactose</b>	<b>p *</b>
24	77 <sup>c</sup> $\pm$ 2	43 <sup>a</sup> $\pm$ 3	32 <sup>b</sup> $\pm$ 3	<b>p &lt; 0,001</b>
48	102 <sup>a</sup> $\pm$ 2	99 <sup>a</sup> $\pm$ 6	96 <sup>a</sup> $\pm$ 3	<b>p &gt; 0,05</b>
72	186 <sup>a</sup> $\pm$ 57	164 <sup>a</sup> $\pm$ 41	144 <sup>a</sup> $\pm$ 11	<b>p &gt; 0,05</b>
96	371 <sup>b</sup> $\pm$ 4	428 <sup>a</sup> $\pm$ 25	385 <sup>b</sup> $\pm$ 13	<b>0,0178</b>
129	401 <sup>c</sup> $\pm$ 3	460 <sup>a</sup> $\pm$ 8	378 <sup>b</sup> $\pm$ 17	<b>0,0003</b>

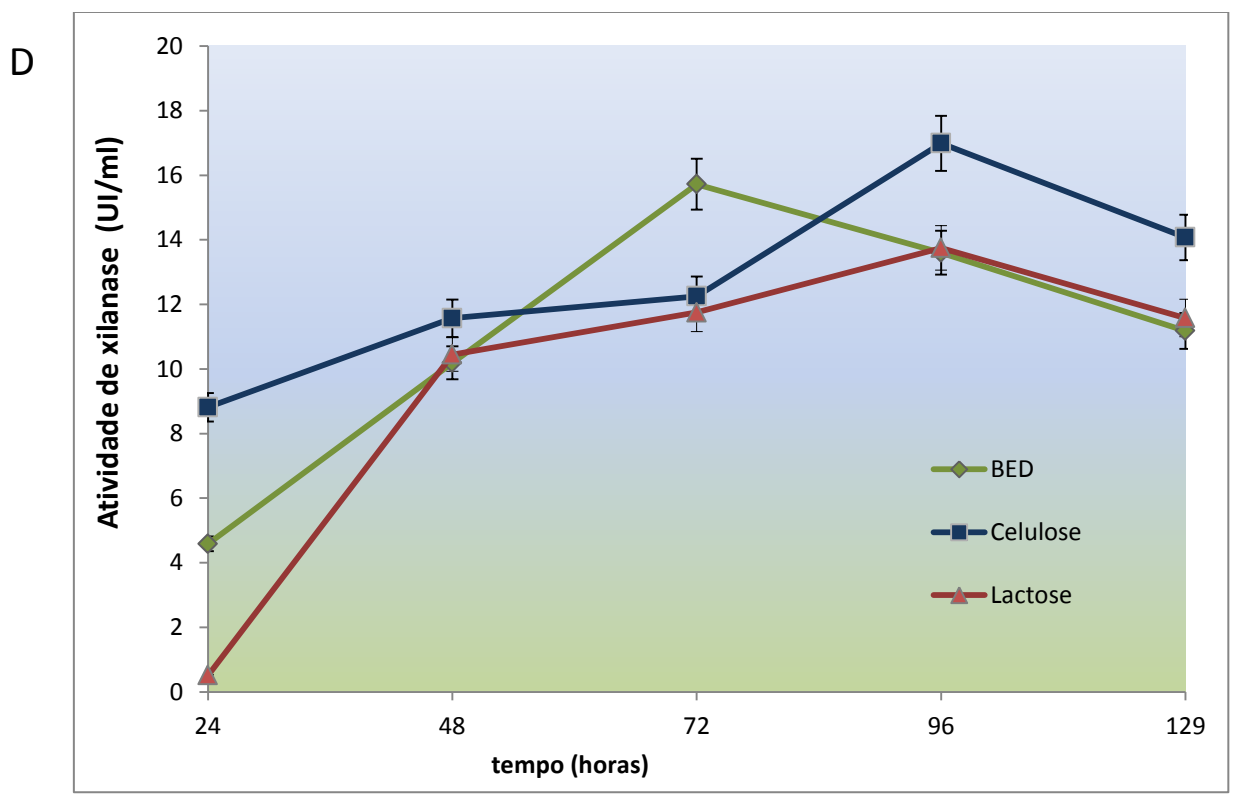
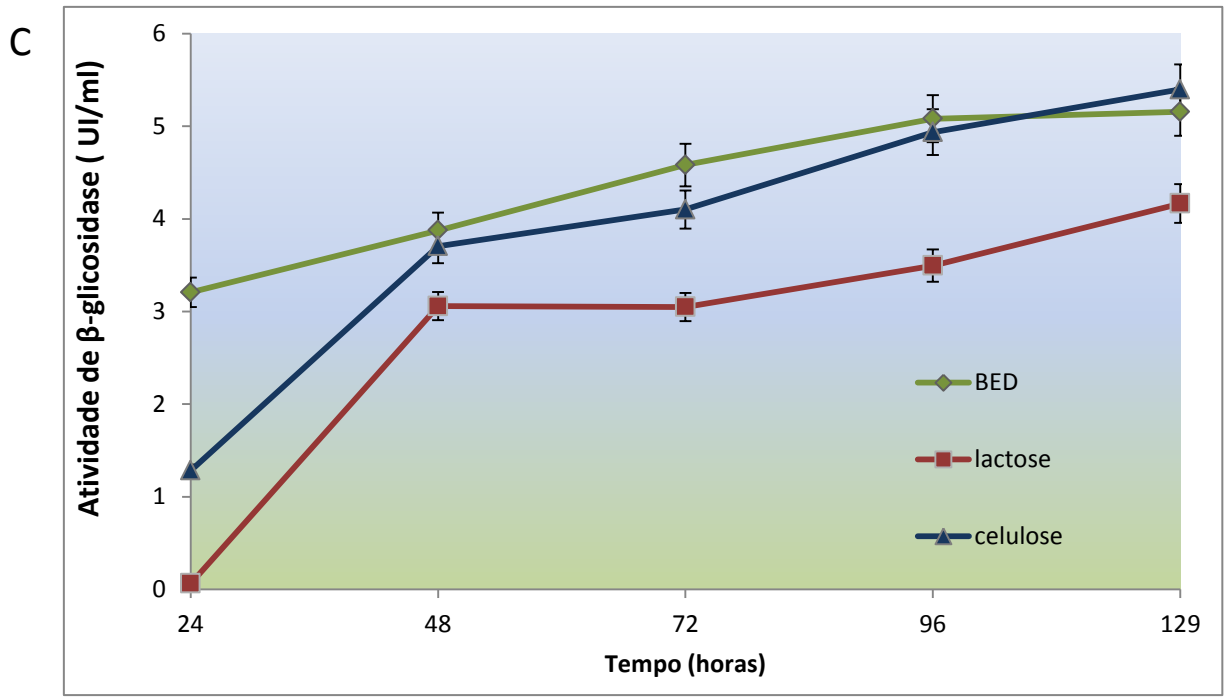


A



B





**Figura 10:** **A:** Concentração de proteína total (mg/L) por tempo de fermentação (horas) em BED. **B:** Perfil da atividade celulolítica (FPA) durante fermentação em BED para as amostras induzidas com BED, celulose e lactose. Os dados foram comparados segundo ANOVA. **C:** Atividade de beta-glicosidase (UI/ml) por tempo de fermentação (horas). **D:** Atividade total de xilanase (UI/ml) por tempo de fermentação.

As proteínas detectadas correspondem às moléculas produzidas pelo fungo (enzimas, co-fatores), produtos de degradação, proteínas do substrato fonte de carbono, etc. O perfil crescente já era esperado devido ao acúmulo de moléculas no meio e ao crescimento do micélio.

É importante notar, que mesmo sob as mesmas condições de fermentação, a produção de proteína só atinge níveis semelhantes segundo teste estatístico ANOVA no tempo de 48 e 72 horas entre todas as amostras. No tempo de 24 e 129 horas as três amostras são consideradas estatisticamente diferentes, com 96h as amostras que utilizaram BED e lactose são semelhantes entre si porém diferentes da amostra que utilizou celulose como indutor do crescimento do micélio. Após esse período a amostra induzida por celulose teve maior concentração de proteínas totais.

A amostra que utilizou lactose como indutor foi a que apresentou o menor desempenho em relação à concentração de proteínas e outras atividades analisadas.

A atividade celulolítica foi mensurada através da atividade em papel de filtro (FPA) e expressa em FPU/ml. FPA não mede a atividade enzimática específica de uma enzima, mas sim a atividade de todo um complexo multienzimático

ativo para a hidrólise da celulose, aqui designado como atividade celulolítica (Tabela 3, figura 10-B).

Tabela 3- Atividade celulolítica detectada e expressa FPU (UI/ml) durante a fermentação em BED das amostras induzidas por diferentes fontes de carbono: celulose, lactose e BED. Os dados foram comparados segundo ANOVA e correspondem ao valor médio  $\pm$  desvio padrão; Os valores seguidos de diferentes letras são significantes para  $p^* < 0,01$ .

Tempo (h)	BED	Celulose	Lactose	p *
24	0,2 <sup>b</sup> $\pm$ 0,01	0,05 <sup>a</sup> $\pm$ 0,004	0,02 <sup>c</sup> $\pm$ 0,002	P<0,0001
48	0,28 <sup>a</sup> $\pm$ 0,034	0,29 <sup>a</sup> $\pm$ 0,023	0,15 <sup>c</sup> $\pm$ 0,032	0,0017
72	0,46 <sup>a</sup> $\pm$ 0,04	0,49 <sup>a</sup> $\pm$ 0,037	0,30 <sup>c</sup> $\pm$ 0,044	0,0029
96	0,53 <sup>a</sup> $\pm$ 0,024	0,50 <sup>a</sup> $\pm$ 0,011	0,43 <sup>c</sup> $\pm$ 0,036	0,007
129	0,24 <sup>b</sup> $\pm$ 0,013	0,42 <sup>a</sup> $\pm$ 0,031	0,22 <sup>b</sup> $\pm$ 0,041	0,0004

Em relação a essa atividade, a amostra que utilizou a celulose como indutor em sua forma cristalina e pura apresentou um perfil crescente até 72h de cultivo, quando atingiu seu máximo pico de atividade celulolítica. A amostra que utilizou BED também como indutor inicia a fermentação com nível 4 vezes maior de atividade celulolítica, provavelmente porque o fungo já havia se adaptado ao substrato na fase de produção de micélio. Neste caso o conjunto de genes ativado durante a indução do crescimento do micélio foi o mesmo utilizado na fermentação, o que proporcionou uma atividade muito superior nas

primeiras 24h. No intervalo de 48h a 96h o perfil de atividade celulolítica das amostras induzidas com celulose e BED mantêm um perfil crescente e a atividade de FPA semelhante às amostras que utilizaram celulose como indutor, segundo análise de variância (ANOVA) presentes na Tabela 3. Isso indica que o conjunto de genes ativos após a fase de adaptação a fermentação pode ser parecido, porém a amostra induzida com BED deve expressar algum gene diferencial pois, atingiu maior pico de atividade celulolítica nas primeiras 24 horas.

Outras enzimas do complexo celulolítico necessárias para a completa degradação da celulose são as beta-glicosidases. Esse grupo de enzimas hidrolisa celobiose e outras celodextrinas em glicose, produzindo açúcares livres passíveis de serem metabolizados pelo fungo. A figura 10-C mostra a atividade de beta-glicosidases nos extratos analisados.

A atividade de beta-glicosidase apresentou um perfil crescente ao longo do tempo, o que pode ser explicado pelo acúmulo de produtos de degradação da celulose, como moléculas compostas por dois ou mais monômeros de glicose. Essas moléculas sofrem a ação das beta-glicosidases e são degradadas de modo que os monômeros de glicose formados podem ser metabolizados pelo fungo. As amostras induzidas com BED e celulose iniciaram a fermentação com maior atividade, e atingiram maior pico indicando que o conjunto de genes necessários para a ação das beta-glicosidases já vinha se expressando. A diferença detectada entre as duas amostras no tempo de 24h deve-se a expressão diferencial de algum gene na fase anterior a fermentação de crescimento do micélio. A adaptação do fungo é observada após 48 horas quando as duas amostras apresentam a mesma tendência para a atividade de beta-glicosidase. A amostra induzida com lactose iniciou a fermentação com baixa atividade pois a lactose

não forma o substrato para esse conjunto de enzimas, por isso não houve indução para produção de beta-glicosidases, o que ocorreu somente a partir da fermentação, que continha BED capaz de formar substrato para esse conjunto de enzimas.

Para a completa degradação de biomassa é necessária a ação conjunta de enzimas que degradem também a hemicelulose, o componente mais abundante da parede celular das plantas. Nesse caso, foi determinada a atividade enzimática dos extratos para xilanases, um conjunto de enzimas que quebram a ligação  $\beta 1 \rightarrow 4$  da cadeia de polissacarídeo xilana, maior componente da hemicelulose presente na parede celular das plantas (26) (Figura 10-D).

A atividade total de xilanases para a amostra induzida com lactose apresentou um perfil crescente no primeiro estágio de fermentação devido a adaptação do fungo a fonte de carbono. A amostra induzida com BED atinge o ponto máximo de atividade de xilanase com 72h enquanto que as amostras induzidas com celulose e lactose atingem o pico com 96h. A diferença detectada deve-se mais uma vez a influência do indutor na fase de crescimento do micélio. A amostra que foi induzida com BED já vinha expressando um conjunto de genes correspondentes a ação das xilanases e por isso precisou de menos tempo para atingir o pico de atividade. Neste caso a celulose se mostrou um bom indutor da atividade de xilanases pois, apesar de demandar um tempo maior, essa amostra atingiu o máximo de produção de xilanases detectado no experimento.

Através desses experimentos foi possível verificar a influência da fonte de carbono utilizada como indutor na fase de crescimento do micélio. A amostra que utilizou lactose atingiu, com 96 horas de fermentação, níveis de proteínas totais semelhantes ao da amostra que utilizou BED, porém manteve as

atividades enzimáticas analisadas abaixo das obtidas para as amostras induzidas com BED e com celulose. Isso indica que as proteínas produzidas na fase de indução podem influenciar a fermentação até mesmo em tempos longos de fermentação, após a fase de adaptação do fungo às condições de fermentação.

No caso da amostra induzida com celulose, a fase de adaptação parece acontecer nas primeiras 48 horas de fermentação, quando a amostra atinge níveis das atividades enzimáticas próximos aos obtidos pela amostra induzida com BED. Esses resultados indicam que o conjunto de genes ativos para a degradação de BED é influenciado pela fonte de carbono, na fase de crescimento do micélio, e pelo tempo de fermentação, e que apesar da tendência do fungo a se adaptar ao meio pela expressão e repressão de certos genes pode-se observar a influência do indutor em tempos longos de fermentação.

Assim, a influência do tipo de indutor nos estágios iniciais de biodegradação pode ser uma ferramenta útil para um aumento da eficiência da produção de enzimas celulolíticas, pois pode favorecer a produção de enzimas específicas para o substrato, com maior eficiência de conversão e capazes de atingir o pico de atividade em intervalos de tempo menores.

## **4.2 Bibliotecas de cDNA**

Inicialmente a abordagem do projeto para a determinação dos transcritos era por meio da utilização das bibliotecas de cDNA. Essa abordagem foi utilizada e as bibliotecas foram montadas, porém durante o desenvolvimento do projeto houve a possibilidade de realizar o sequenciamento *high-throughput*. Esse tipo de sequenciamento é mais rápido e barato, o que nos levou a interromper a utilização das bibliotecas de cDNA.

Para a construção da biblioteca de cDNA foi utilizado o kit In-Fusion®SMARTer™ Directional cDNA Library Construction Kit. O kit inclui protocolo completo para a síntese de cDNA e construção da biblioteca. Para a síntese de cDNA foi utilizado RNA total correspondente ao tempo de fermentação de 96 horas, que apresentou maiores atividades celulolíticas, e seguiu-se metodologia do fabricante, conforme manual específico (50). Assim obteve-se uma solução que contém os fragmentos de cDNA purificados (Figura 11-A) e clonados no vetor pSMART2IFD. Esta solução é considerada nossa biblioteca estoque (armazenada a -20°C), para preparo dos clones com inserto através da transformação em células eletrocompetentes de *E. coli* (DH10B) (Figura 11-B).

Os testes de verificação de tamanho do inserto foram feitos por PCR utilizando *primer* específico do vetor de clonagem. Pela observação dos géis e comparação com o padrão de peso molecular foi possível verificar que tamanho dos insertos da amostra lactose 96h e celulose 96h correspondem a fragmentos que variam de 1650 a 650 pb. Já para a amostra BED 96h as bandas correspondem a fragmentos que variam de 1650 a 850 pb. Porém a maior parte das bandas observadas e a região de maior *smear* é aquela entre 1000 e 850 pb (Figura 11A). Essa faixa de tamanho é considerada ideal para o sequenciamento e identificação de ESTs. A porcentagem média de colônias recombinantes positivas para a presença do inserto variou de 25% a 35% (figura 11-B).

Para a amostra que utilizou BED como indutor do crescimento do micélio foram plaqueados 3199 clones positivos em placa ELISA de 96 poços, armazenados a -70°C, em meio 2YT/HMFM com Ampicilina 100µg/ml. Para a amostra que utilizou lactose como indutor foram 135 clones e para a amostra induzida com celulose foram armazenados 3085 clones nas mesmas condições.



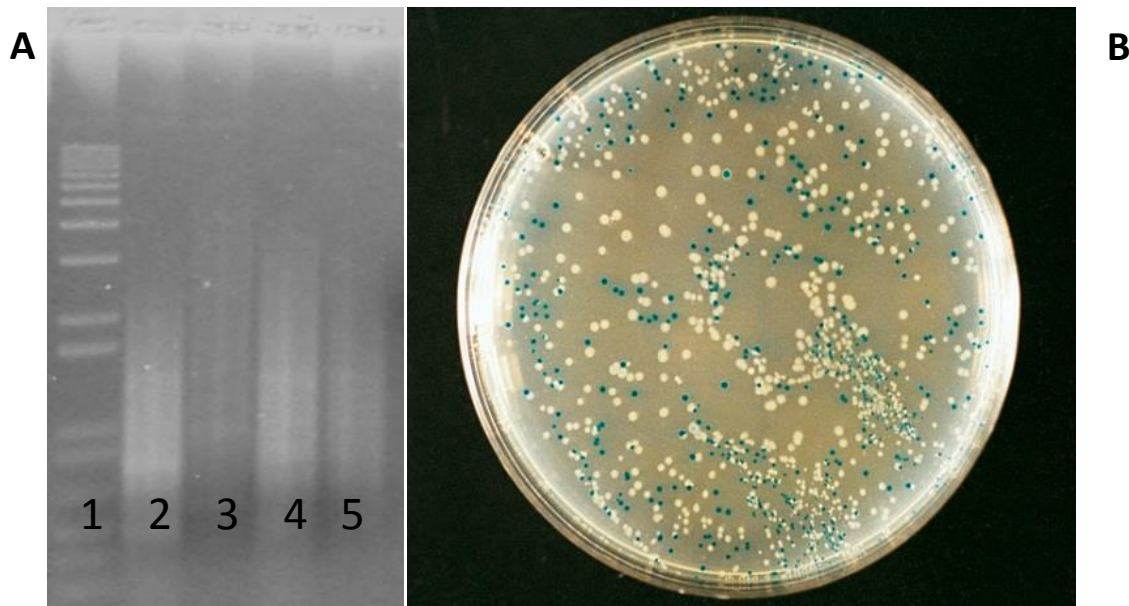


Figura 11: **A** : Eletroforese unidimensional de cDNA: amostras 1-padrão tamanho 1Kb, 2-BED, 3-celulose, 4-lactose, 5-controle. **B**: Seleção de colônias transformantes: clones positivos (brancos) e negativos (azuis) para a presença do inserto.

#### 4.2.1 Sequenciamento dos clones

Os clones foram sequenciados por eletroforese capilar no equipamento ABI 3500 Genetic Analyzer (Applied Biosystems), preparado para sequenciar placas de 96 poços.

Para o sequenciamento foi necessário isolar o DNA plasmidial de cada um dos clones positivos das bibliotecas armazenados em placas de 96 poços. Para isso foi preciso inocular 2 $\mu$ l dos clones em meio Circle Grow com 100  $\mu$ g/ml de ampicilina e incuba-los a 37°C, 300 rpm por 22h. As células foram centrifugadas e ressuspensas com 240 $\mu$ l de solução Tris-glicose-EDTA pH

8,0, centrifugadas novamente e o sobrenadante descartado. Aos poços foram adicionados 60 µl de solução Tris-glicose-EDTA pH 8,0 e 5µl de RNase 10 mg/ml. A cada poço, então, foi adicionado 60 µl de NaOH 0,2M – SDS 1% e as placas incubadas por 10 min antes da centrifugação. Depois da centrifugação, à cada poço foram adicionados 60µl de acetato de potássio 3M e as placas incubadas a 90°C por 30 min. As placas foram resfriadas e centrifugadas. Todo o volume das placas foi transferido para as placas de filtro, e após a centrifugação foram adicionados aos filtrados 90 µl de isopropanol, e centrifugadas a 4°C, 4000 rpm por 45 min. O sobrenadante foi descartado, seguindo-se a adição de 160 µl de etanol 70%. As placas foram centrifugadas novamente, o sobrenadante descartado e secas por 60 min a temperatura ambiente para a ressuspensão do conteúdo em 30 µl de água MilliQ.

Após a extração plasmidial, as reações de sequenciamento foram preparadas através da amplificação do inserto por reações de PCR. Após amplificação do inserto, a amostra foi purificada pela metodologia EDTA/etanol. As amostras secaram por aproximadamente 45 min e foram estocadas a 4°C prontas para o sequenciamento.

#### **4.2.2 Análise das sequências obtidas a partir das bibliotecas de cDNA**

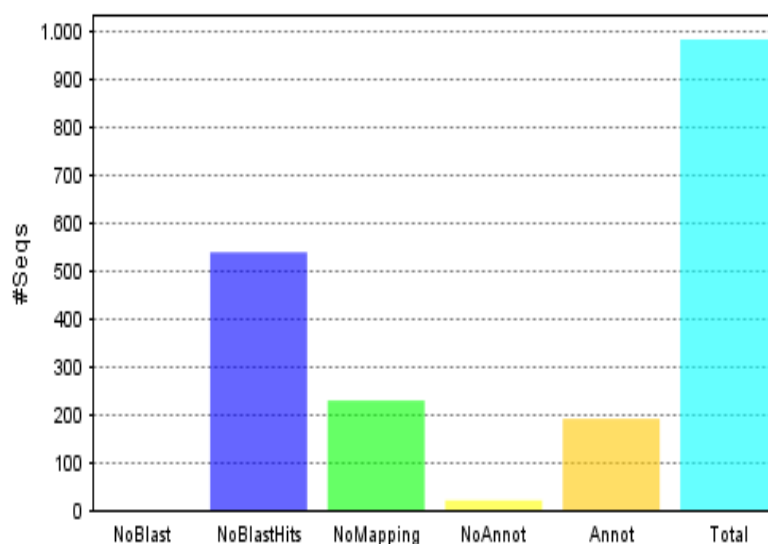
As sequências foram obtidas com a extensão .ab1. Esses arquivos gerados foram pré-analisados quanto a qualidade e tamanho no software de análise de cromatogramas [BioEdit](#), e as sequências com qualidade foram analisadas no NCBI *nr*, comparando a sequência de nucleotídeos obtida com sequências de proteínas depositadas. As sequências também foram analisadas através do

software de análise [Blast2Go](#), Neste caso os arquivos foram convertidos para FASTA e concatenados em um único arquivo para posterior análise.

Resultados apontaram para sucesso desta metodologia para a detecção de genes. Novecentos e oitenta e três clones foram sequenciados (biblioteca 1, indução e fermentação em BED) o que permitiu identificar clones com sequências que apresentaram 94% de identidade com uma beta-1,6-glucanase [AEZ66647.1](#), 63% de identidade com alfa/beta hidrolase ([gb|EFY86157.1](#)), 69% com esterase [EGX95927.1](#), além de sequências próximas a peptidases, catalases, malatodesidrogenase, ubiquitinas, hidrofobinas, etc.

Uma análise conjunta das sequências foi obtida e está apresentada nas figuras 12, 13, 14 e 15.

A



B

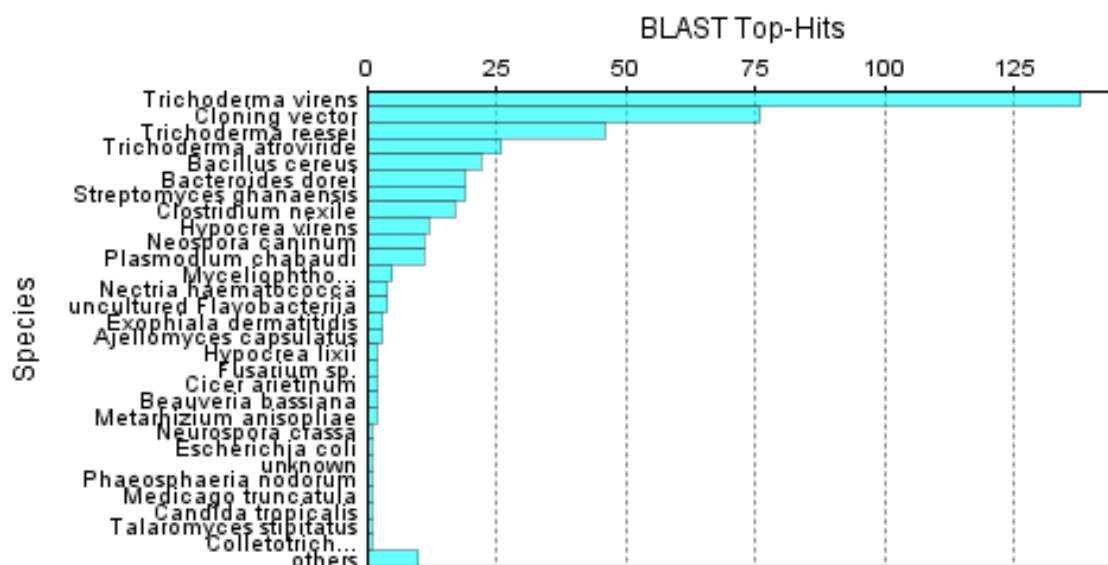


Figura 12: Resultados da análise da biblioteca de cDNA, A- Classificação da anotação e B-classificação segundo similaridade das seqüências com outras espécies.

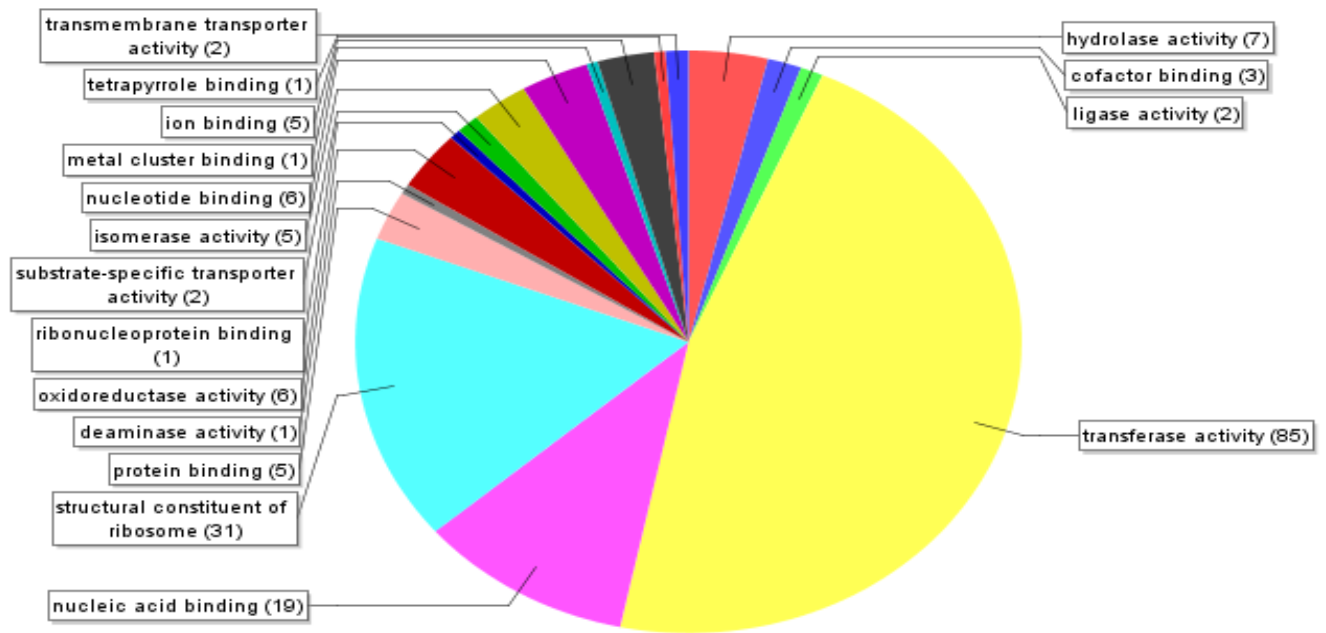


Figura 13: Classificação segundo a função molecular das sequências anotadas.

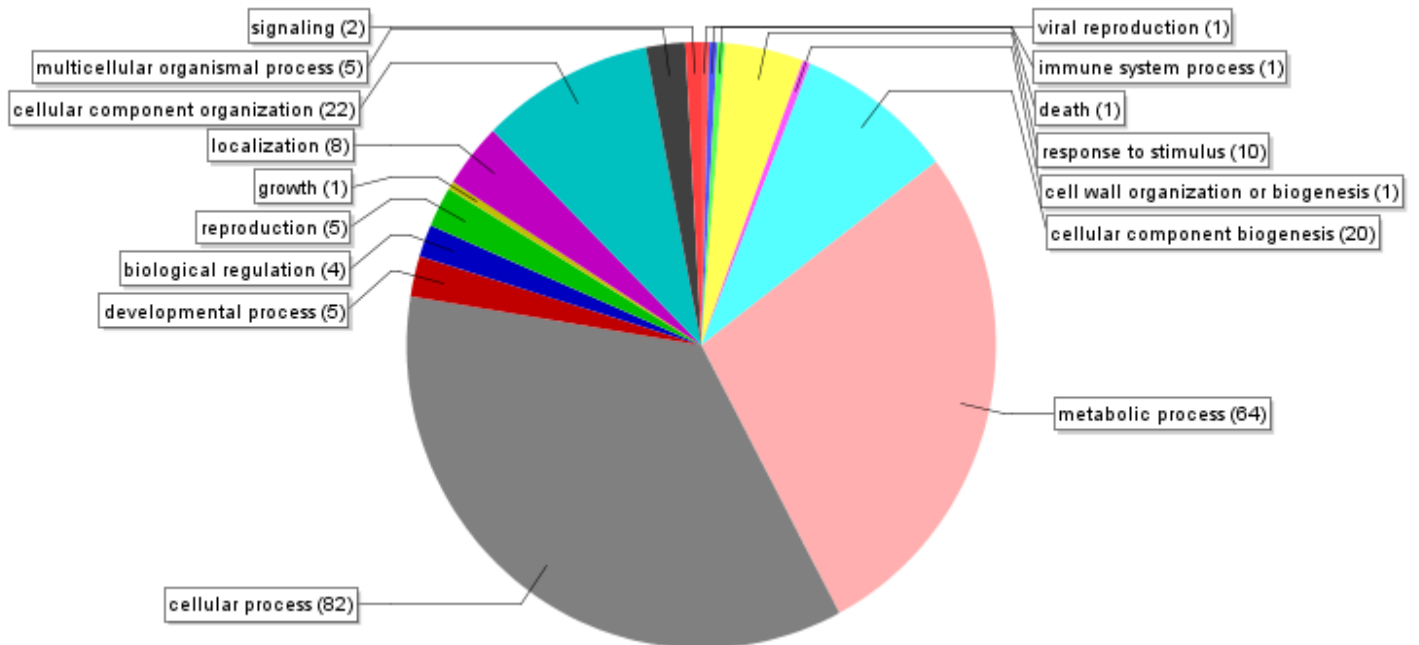


Figura 14: Classificação segundo processo biológico das sequências anotadas.

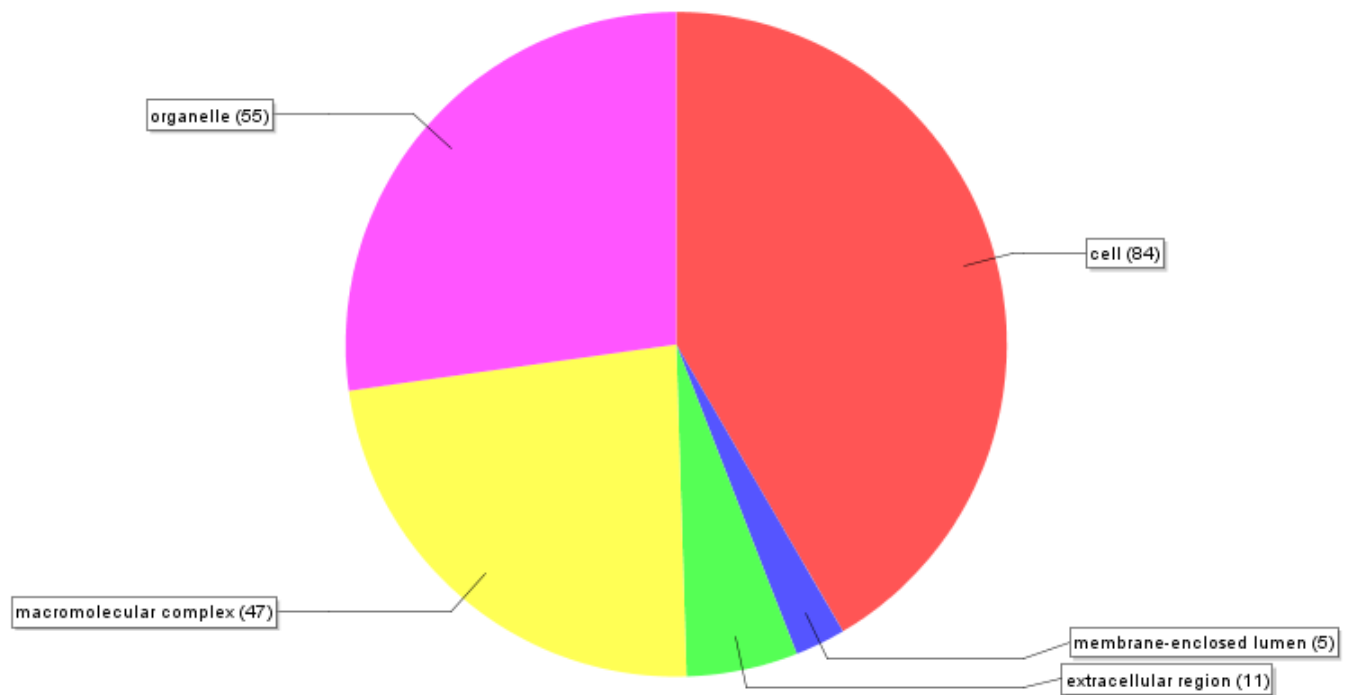


Figura 15: Classificação segundo componente celular das sequências anotadas.

A figura 12 mostra que das 983 sequências analisadas 539 (54,8%) não foram passíveis de identificação, classificadas como *No Blast Hits* e 230 (23,4%) apresentaram semelhança com sequências já descritas nos bancos de dados sem anotação, 192 (19,5%) foram identificadas e anotadas e 22 (2,2%) não foram anotadas.

Muitas das 54,8% das sequências apresentaram baixa qualidade, sequências de poucos pares de bases ou até mesmo ausência de qualquer sequência no respectivo arquivo. Falhas nas etapas de preparação e extração plasmidial podem ter gerado baixas concentrações de DNA para a amplificação, o que gera uma amostra com baixo sinal, insuficiente para que o sequenciador

atribua bases com qualidade suficiente à sequência analisada. Essa deve ser a principal causa das falhas, pois foi observada a variação do aproveitamento médio das sequências de cada placa, a fim de determinar as etapas críticas do processo de preparação das amostras para o sequenciamento.

Outro ponto importante é a obtenção de placas com concentrações de DNA homogêneas entre os poços para que as reações para o preparo das amostras ocorram nas concentrações ideais em cada um dos poços das placas. Alguns parâmetros foram estabelecidos para melhorar esse aproveitamento como: o volume da solução plasmidial a ser filtrado foi estabelecido, diferentes selos para as placas foram testados para evitar a perda do material, a concentração de DNA para a reação de amplificação foi aumentada assim como o número de ciclos da PCR. Assim foi possível melhorar o aproveitamento das sequências por placa analisada que passou de 30 para 80%.

No total 444 sequências foram identificadas, analisadas, quanto a função e processo biológico e anotadas (Figuras 13, 14 e 15). Ressalta-se que muitas sequências que apresentaram identidade com outras já depositadas não foram anotadas pois, atualmente, a comunidade científica busca determinar inúmeras classificações de sequências e proteínas. É um esforço conjunto que envolve não só o depósito de novas sequências em banco de dados, mas também um estudo aprofundado sobre a função, estrutura, atividade, interações e modificações dos genes codificados pelas sequências determinadas. A figura 13 mostra a distribuição das 182 sequências que puderam ser classificadas de acordo com a função molecular, a figura 14 mostra a classificação de 232 sequências de acordo com o processo biológico que participa e a figura 15 mostra as 202 sequências que tiveram seu componente celular identificado. As sequências têm em 315 e 940pb.

De acordo com a análise de função molecular 30 sequências são relacionadas com reações *binding*, ou seja, interagem seletivamente e não-covalentemente com alvo, 104 estão relacionadas atividade catalítica, 31 relacionam-se com a atividade estrutural molecular, 2 com as atividades de transporte dentro ou fora da célula, ou entre células e 2 com a atividade de transporte de elétrons. Dentre as 104 sequências com atividade catalítica foram identificadas 7 sequências correspondentes a enzimas que desempenham atividade hidrolítica, promovendo a hidrólise de várias ligações, *e.g.* C-O, C-N, C-C, etc. Hidrolase é o nome sistêmico de qualquer enzima da classe EC 3 (GO:0016787).

Uma análise comparativa mostra que existem sequências com alta similaridade com sequências já depositadas que podem indicar um gene alvo. A presença de regiões conservadas e homólogas entre a sequência proveniente do *T.harzianum* e aquela depositada nos bancos de dados é fator importante que pode indicar a presença de um gene ou região genômica com a mesma característica/função. Alguns dos principais alinhamentos das sequências determinadas foram analisados e detectou-se 55% de identidade com amidohidrolase [XM\_002790928.1], 96% com glicosilhidrolase [EHK24192.1], 66% com alfa/beta hidrolase [EFZ00770.1], 67% com esterase (EFY999501), 60% com glicosilhidrolase [XP003649436.1] e 61% com endo,1,4-beta-xilanase, 98% com glicosilhidrolase, 77% com endo 1,4-beta-xilanase [CAA76571.1], 93% com glicosilhidrolase [EHK16138.1] de outros fungos.

Estas sequências podem estar envolvidas com as reações de hidrólise enzimática necessárias para a degradação do substrato complexo que o fungo utilizou como fonte de carbono. A presença de enzimas que degradam ligações glicosídicas, ligadas ao metabolismo de carboidratos (GO:0005975,



GO:0004553) está relacionada com a despolimerização das cadeias de hemicelulose presente no bagaço de cana-de-açúcar.

Uma análise geral do experimento mostra que do total de sequências submetidas à análise, 39,7% foram identificadas e classificadas no banco de dados [InterPro](#) das quais 50% também foram identificadas quanto a função e processo biológico que participam segundo a classificação do [Gene Ontology](#) (GO terms). Entre elas as sequências classificadas como hidrolíticas são as que o presente projeto se propõe a analisar a fim de determinar as que são passíveis de clonagem.

Determinar o ponto em que há maior probabilidade de encontrar as enzimas alvo foi de fundamental importância para o sucesso da abordagem. O crescimento do fungo no bagaço de cana induz a produção de uma série de enzimas hidrolíticas, a produção dessas enzimas aumenta com o tempo de fermentação e conseqüentemente a transcrição desses genes em RNAm. É no tempo de maior atividade enzimática para FPU, Beta-glicosidase e xilanase (96h) que está aumentada a chance de que um fragmento de RNAm correspondente ao conjunto de enzimas hidrolíticas seja clonado com sucesso, armazenado em uma biblioteca cDNA, corretamente sequenciado, identificado e anotado.

Apesar das bibliotecas de cDNA não terem sido utilizadas com toda a sua potencialidade, pode-se prever o sucesso do experimento em termos do alto número de sequências identificadas.

Durante a execução do projeto surgiu a oportunidade de efetuar o sequenciamento *high-throughput*, e assim optamos por mudar a abordagem das bibliotecas de cDNA para as bibliotecas montadas à partir do sequenciamento

tendo a certeza da obtenção de uma quantidade maior e melhor de sequências de RNA.

### 4.3 Perfil do Transcriptoma e análise das bibliotecas

O RNA total das amostras foi sequenciado no Genome Analyser Iix, um sequenciador NGS, seguindo os protocolos do equipamento, com sucesso. A amostra que foi induzida com bagaço de cana (BED) deu origem a biblioteca 1 (*Delignified Sugarcane Bagasse*), aquela induzida com celulose cristalina deu origem a biblioteca 2 e aquela que cresceu em lactose, a biblioteca 3. Todas elas com 96 horas de fermentação em BED.

No total foram gerados pelo sequenciador automático 246.476.390 *reads* com 72pb (tabela 4). Todos os *reads* foram submetidos aos testes de qualidade (*quality trimming*), resultando em um total de 205.276.761 de sequências utilizáveis. Na média 83,31% das sequências passaram por algum tipo de correção e 41.199.629 foram excluídas, resultando em um total de 14,7Gpb analisadas e prontas para a montagem (*assembly*).

Tabela 4: *Reads* gerados após sequenciamento e testes de qualidade.

Biblioteca	Reads	Tamanho médio(pb)	Números de reads pós trim	Porcentagem trimmed	Tamanho médio pós trim (pb)
1	81,705,758	72.0	68,720,401	84.11%	71.8
2	84,301,646	72.0	68,644,205	81.43%	71.7
3	80,468,986	72.0	67,912,155	84.4%	71.8

As sequências foram montadas *de novo* (sem genoma de referência) separadamente, utilizando parâmetros que tornam a montagem bastante precisa.

Um total de 31.096 *contigs* foram gerados para a biblioteca 1, o total de 29.433 para a biblioteca 2 e o total de 27.271 para a biblioteca 3. As *reads* de cada análise também foram unidas em uma única montagem (1+2+3) formando 32.494 *contigs* com tamanho médio de 759pb (tabela 5):

Tabela 5: Montagem (*assembly de novo*). **Reads**: fragmentos analisados 1(DSB), 2 (CEL) 3 (LAC);**matched**: número reads que alinhou com os contigs;**not matched**: número de reads sem alinhamento;**contigs**: numero de contigs formados no *assembly de novo*, **reads in pairs**: número de reads mapeadas com os respectivos pares;**broken paired reads**: reads que apresentam somente um par mapeado ou fora da orientação/ distância; **a.v**: tamanho médio (pb); **total bases**: número de pares de bases total.

BIBLIOTECA	1			2			3			1+2+3		
		a.v	total bases		a.v	total bases		a.v	total bases		a.v	total bases
<b>reads</b>	68720401	71,77	4932003465	68644205	71,71	4923494328	67912155	71,79	4875313083	205276761	71,76	14,7Gbp
<b>matched</b>	34617349	71,76	2484044408	41619703	71,72	2984981847	43287174	71,75	3106051098	113363236	71,72	8130919472
<b>not matched</b>	34103052	71,78	2447959057	27024502	71,73	1938512481	24624981	71,85	1769261985	91913525	71,81	6599891404
<b>contigs</b>	31096	579	18016415	29433	731	21541636	27271	798	21777376	32494	759	24690373
<b>reads in pairs</b>	2613350	202,6		3944792	201,7		4808028	202,2		10252298	203,8	
<b>broken paired reads</b>	27951498	71,81		32586584	71,82		34049592	71,83		93428467	71,78	
<b>n75</b>	395			527			595			553		
<b>n50</b>	796			1196			1395			1251		
<b>n25</b>	1480			2130			2395			2280		

Visando uma análise detalhada da estrutura do transcriptoma, todos os *contigs* foram submetidos a identificação por similaridade no banco de dados do NCBI *Inr*. Após a identificação das sequências estas foram anotadas através do programa [Blast2Go](#), utilizando o servidor disponível em nosso centro de pesquisas. Uma análise mostra a classificação das sequências das bibliotecas 1(BED), 2(CEL) e 3(LAC) segundo os termos do Gene Ontology (51) (Figuras 16 e 17).

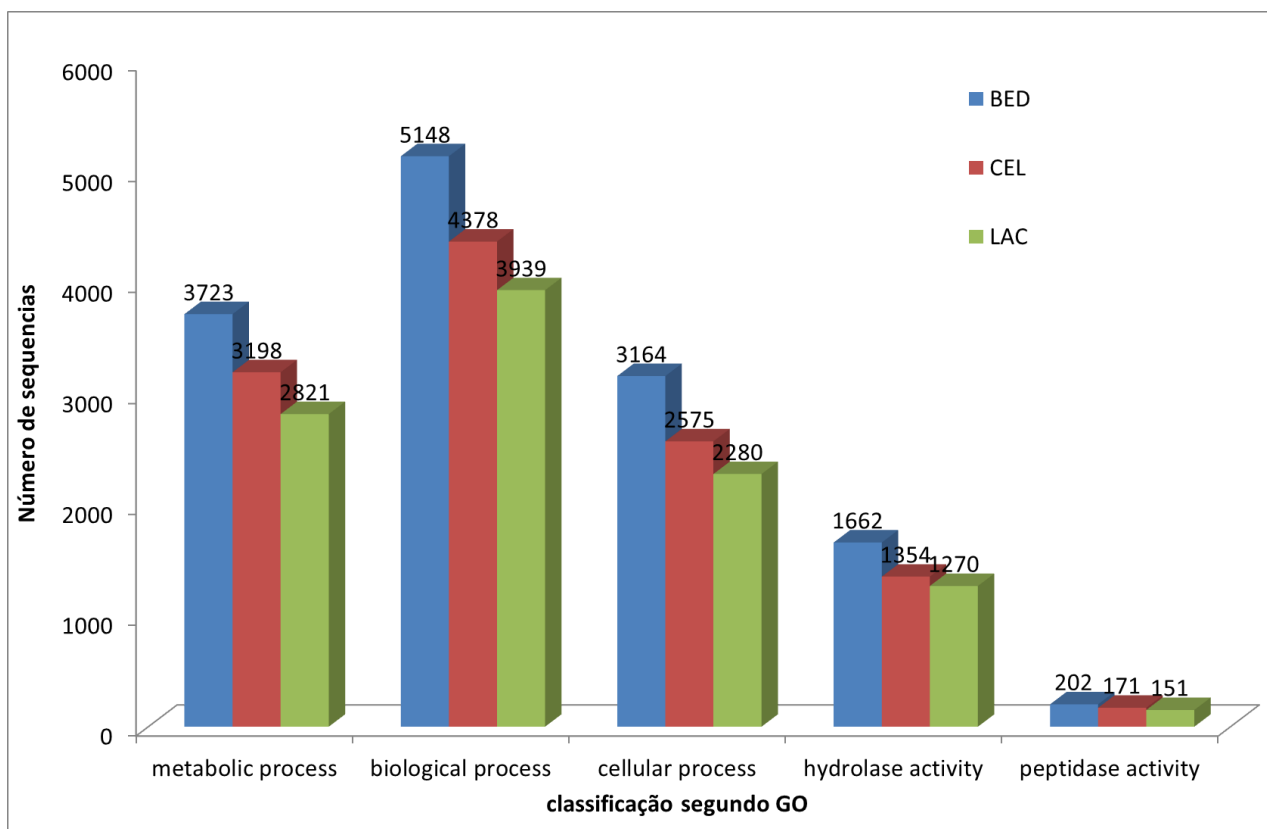


Figura 16: Classificação geral das sequências obtidas na análise do transcriptoma das bibliotecas BED,CEL e LAC segundo os termos do Gene Ontology (49).

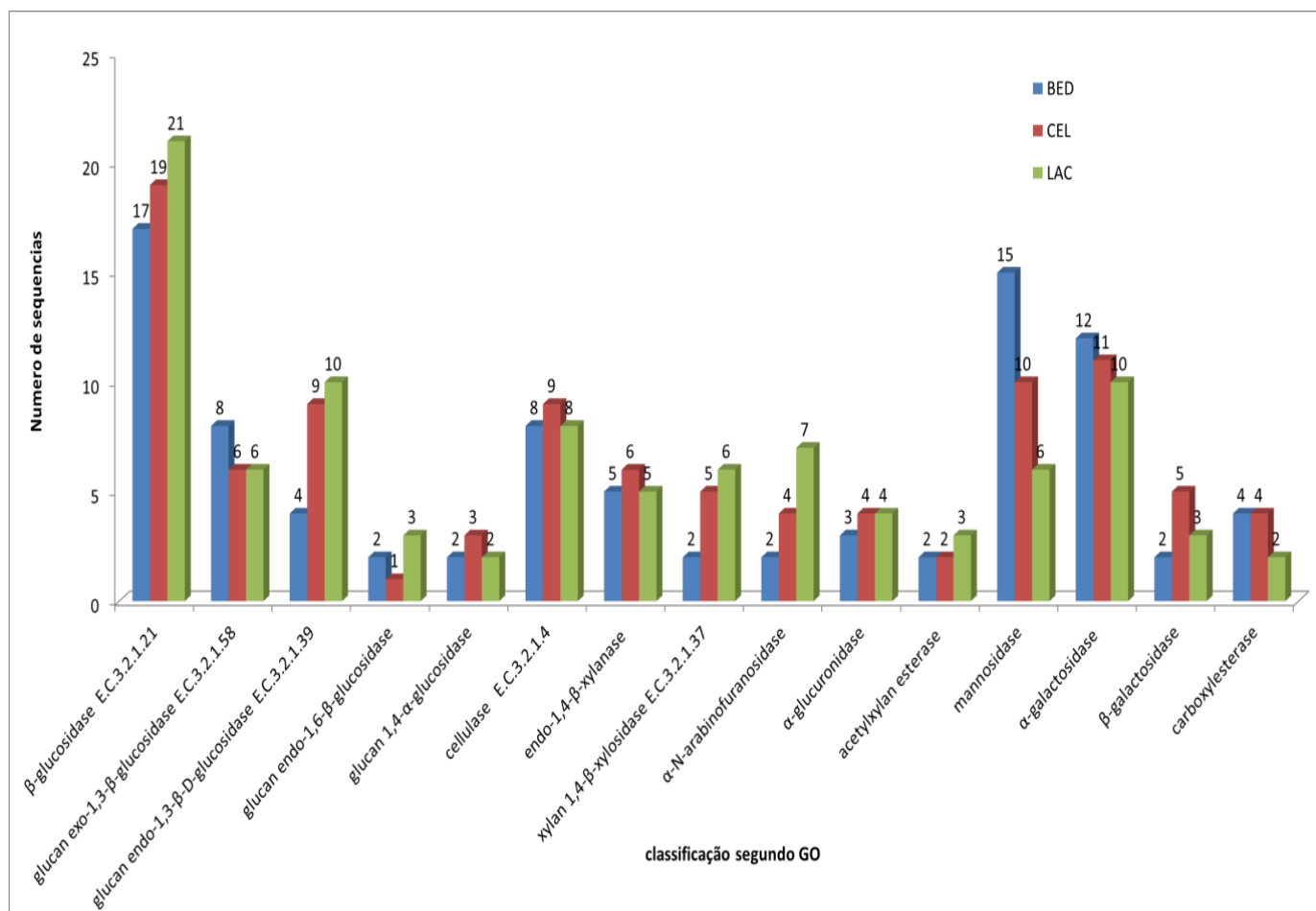


Figura 17: Classificação das sequências possivelmente envolvidas nas reações de degradação de biomassa obtidas através da análise do transcriptoma das bibliotecas BED, CEL e LAC segundo os termos do Gene Ontology (49).

O alto número de seqüências classificadas, possível através do uso de tecnologia de sequenciamento *high-throughput*, permitiu uma visão precisa (Figura 16) dos diferentes processos biológicos que ocorriam no exato momento de maior atividade enzimática do extrato durante a fermentação (96 h). Uma classificação mais refinada das seqüências permitiu a análise das seqüências que podem estar envolvidas na degradação de biomassa (Figura 17). Os resultados mostram que existem diferenças entre as bibliotecas, pois em alguns casos foi observada diferença entre o número de seqüências anotadas para a B-

glucosidase, por exemplo. Neste caso, como a classificação se refere aos *contigs* formados é possível concluir que, possivelmente existe algum gene expresso na biblioteca LAC, que apresentou 21 classificações, que não está presente na biblioteca BED, com 17 classificações.

Uma análise dos níveis de expressão pode indicar quais genes estão presentes nas três bibliotecas e seu comportamento no que se refere aos níveis de expressão. Para tanto foi realizada uma única montagem com as sequências derivadas das três bibliotecas. Aqui estão apresentados os resultados da análise dos 32.494 *contigs* formados. Através da busca por homologia (18) foi possível determinar a distribuição das sequências encontradas, relativas a função molecular, função biológica e localização celular (Figuras 18, 19 e 20).

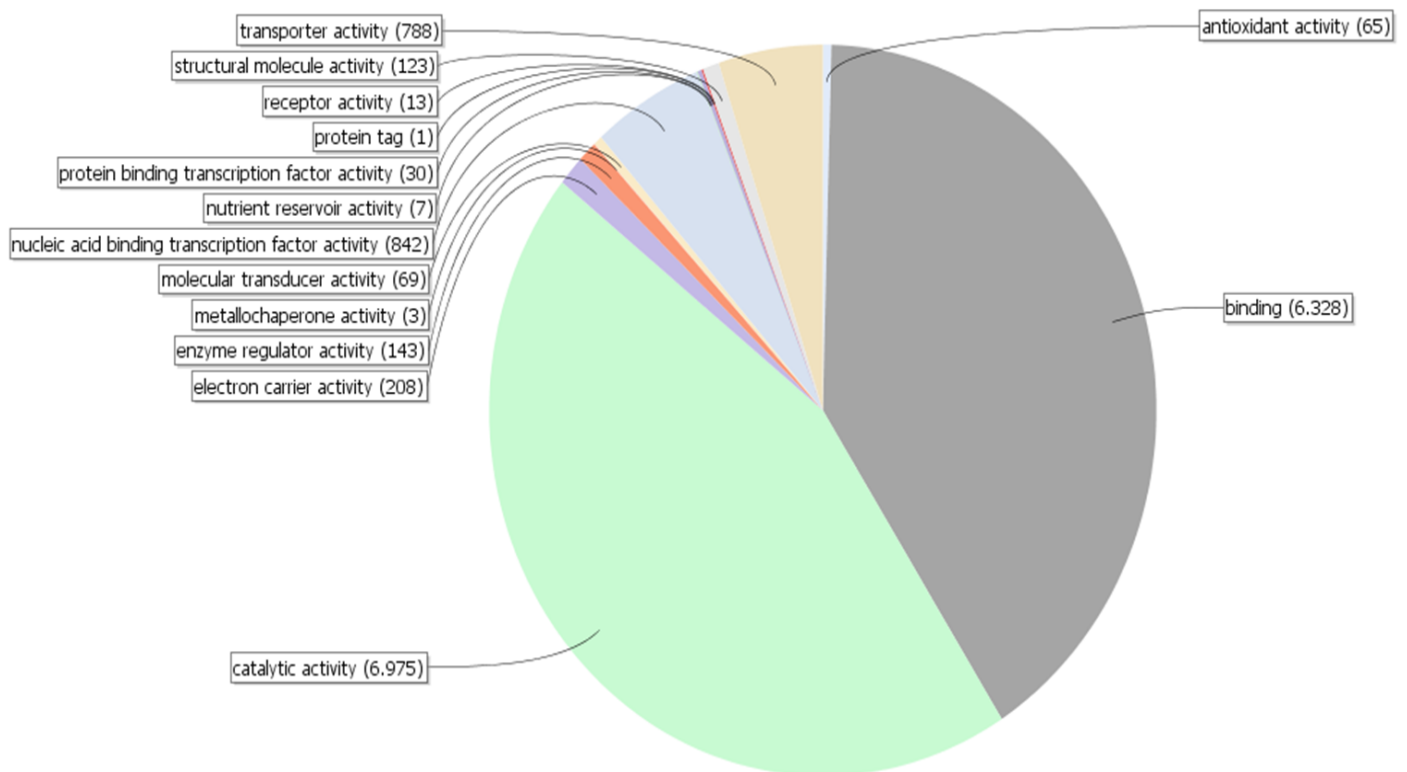


Figura 18: Distribuição das sequências de acordo com a função molecular.

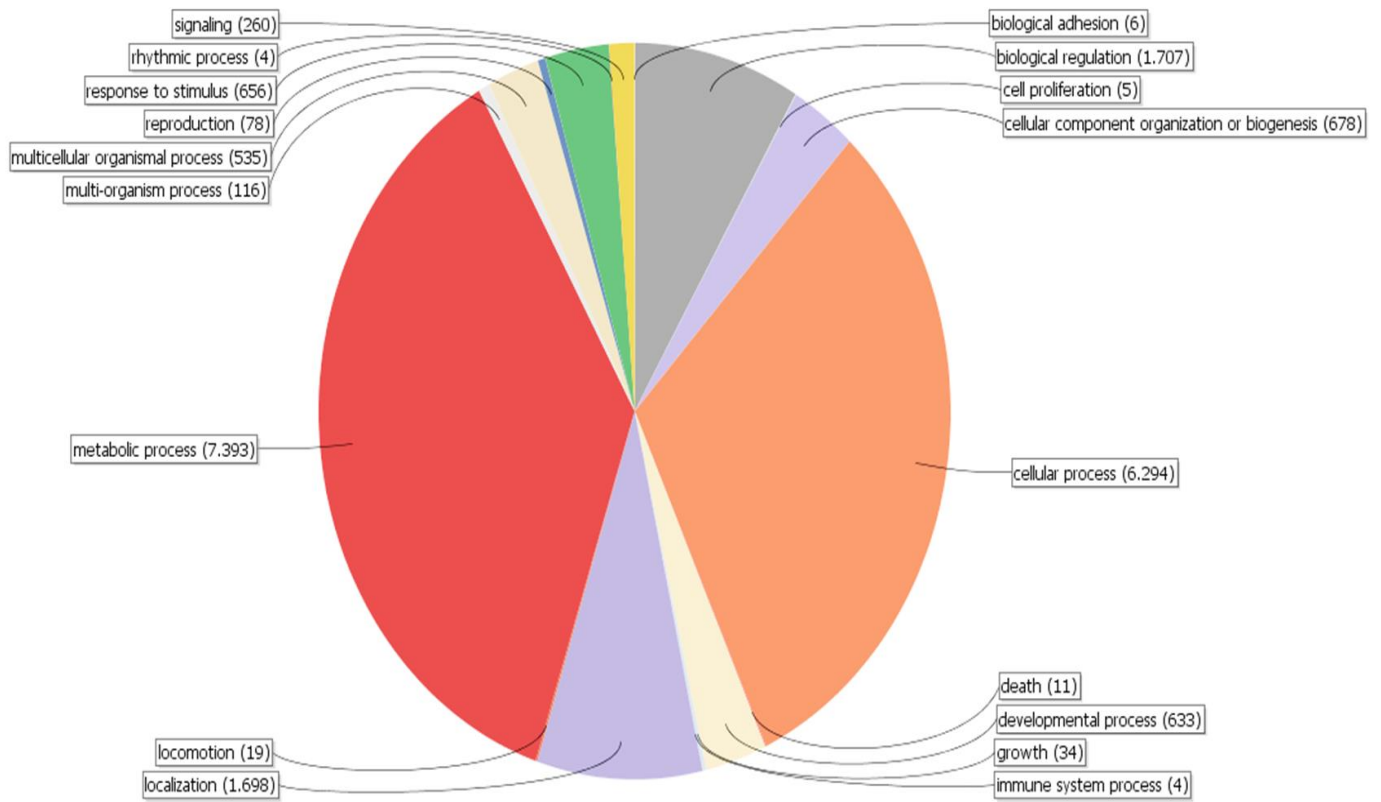


Figura 19: Distribuição das seqüências de acordo com a função biológica.

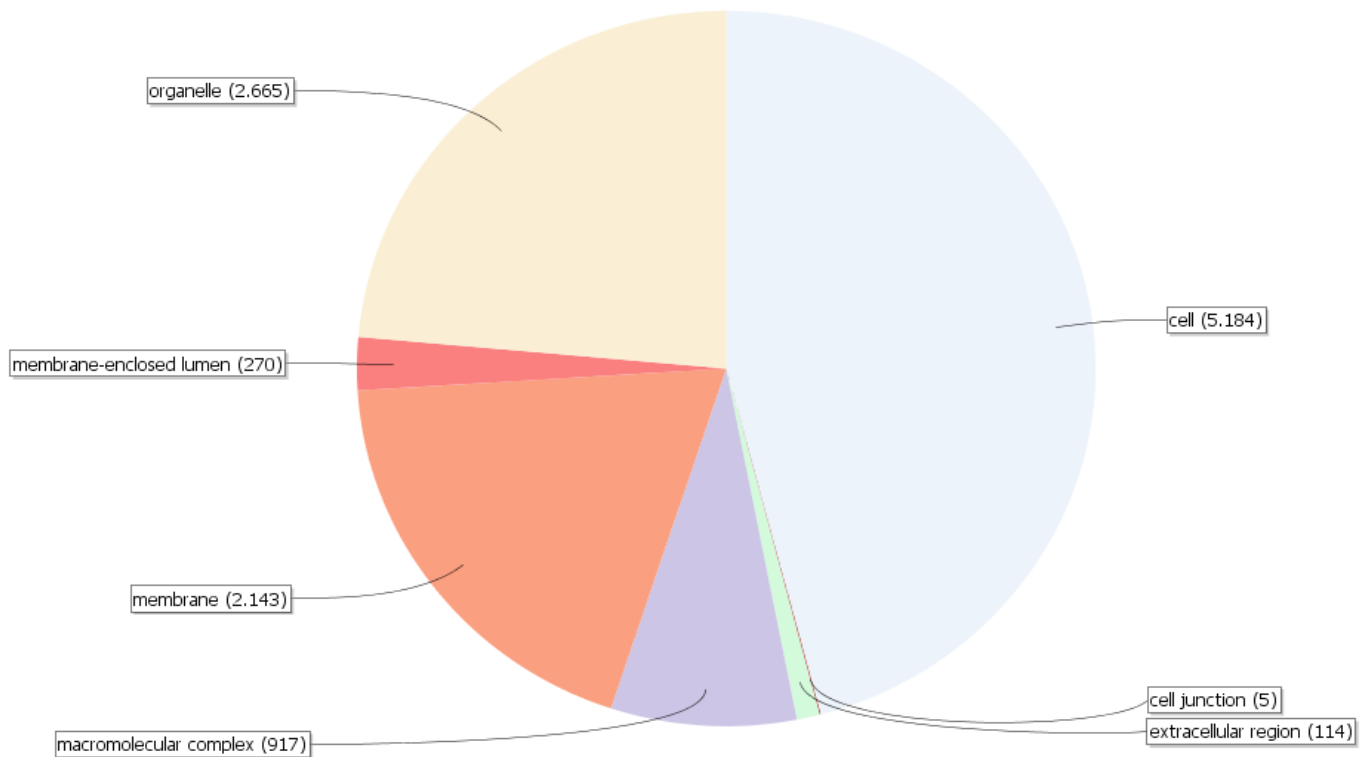


Figura 20: Distribuição das seqüências segundo a localização celular.

As seqüências classificadas como responsáveis por atividade catalítica (6.975) e regulatória (143) podem estar envolvidas nas reações de degradação de lignocelulósicos. Quando são analisados os processos biológicos envolvidos na ação predita da seqüência classificada, processos metabólicos (7.393) seguidos por processos celulares (6.294) são os que apresentam o maior número de classificações. Em relação a função molecular, os *contigs* envolvidos em reações de *binding* e atividade catalítica são mais frequentes. Em termos de componente celular, *contigs* envolvidos em partes celulares (5.184), organelas (2.665) e membrana (2.143) são mais abundantes.

Para elucidar a influência de um substrato complexo como o bagaço de cana no conjunto de genes transcritos pelo fungo, nós analisamos os perfis de



transcrição das amostras. Estes são os primeiros resultados publicados que elucidaram o perfil de expressão deste fungo em substrato complexo (Figuras 18, 19 e 20).

Das 6.975 sequências classificadas possuindo atividade catalítica (21.46% do total de *contigs*), 2.555 apresentaram atividade de hidrolase, agindo sobre ligações químicas tipo éster, carbono-nitrogênio, carbono-carbono, entre outras. O alto número de hidrolases sequenciadas e identificadas nos permitiu determinar exatamente quais sequências gênicas são relacionadas com as reações de degradação foco deste estudo. Um perfil similar de anotação, porém oriundo de um número muito menor de amostras, gerado a partir do Gene Ontology (GO) foi descrito por Steindorff (52), para uma biblioteca de ESTs com 2.927 sequências. Em ambos os experimentos, a maioria das sequências classificadas, segundo a função molecular, correspondem a atividade catalítica e de ligação (*binding*), relativas ao processo que participam foram classificadas com maior número processos metabólicos e celulares e a respeito da localização celular, célula e organelas foram as classificações mais numerosas.

Para a hidrólise de substratos complexos, como o bagaço de cana, o fungo produz diferentes enzimas especializadas necessárias para promover a hidrólise do material lignocelulósico. O efeito da interação entre diferentes grupos de enzimas tem sido estudado (12)(13)(53) e pode ser observado neste experimento.

Através dessa análise foi possível desenhar *primers* específicos para a seleção de regiões genômicas completas (tabela 6). Genes que codificam enzimas das três principais classes de celulasas já foram amplificados a partir do DNA de *T. harzianum* IOC-3844, bem como genes que codificam uma hemicelulase e uma proteína auxiliar de celulasas:

Tabela 6: Genes que codificam enzimas das três principais classes de celulasas amplificados a partir do DNA de *T. harzianum* IOC-3844.

<b>Enzima</b>	<b>Tipo</b>	<b>Função</b>
<b>Endoglucanase I (EGI)</b>	Endocelulase	Clivam aleatoriamente ligações internas em regiões amorfas da molécula de celulose, liberando novos terminais de cadeia.
<b>Endoglucanase II (EGII)</b>		
<b>Endoglucanase III (EGIII)</b>		
<b>Endoglucanase IV (EGIV)</b>		
<b>Celobiohidrolase I (CBHI)</b>	Exocelulase	Quebram de duas a quatro unidades dos terminais de cadeia expostos, produzidos pelas endocelulasas. CBHI atua em terminais redutores.
<b>b-Glucosidase</b>	Celobiase	Hidrolisa o produto das exocelulasas em monossacarídeos.
<b>Swollenina</b>	Similar a Expansinas de plantas	Afrouxam as fibras de celulose, sem produção de açúcares redutores, o que permite que as enzimas hidrolíticas tenham sua ação potencializada.
<b>Xilanase</b>	Hemicelulase	Quebram xilana em xilose, decompondo desta forma a hemicelulose, um dos principais componentes da parede celular de plantas.

Estes genes foram amplificados a partir de uma biblioteca de BACs do fungo em estudo (processo FAPESP 2009/13185-3), cujos clones apresentaram insertos de tamanhos que variam de 80 a 200 Kpb, havendo a possibilidade de, nesses grandes insertos, encontrarmos genes e suas regiões regulatórias completas. Os mecanismos de regulação destes genes ainda são desconhecidos porém, o estudo destas regiões fornecerá informações sobre como acontece a expressão destes genes, a produção destas proteínas e consequentemente, as reações de hidrólise enzimática.

## 5. CAPÍTULO 2:

*Perfil do transcriptoma e análise dos níveis de expressão de enzimas relacionadas à biodegradação.*



## **"Transcriptome Profile of *Trichoderma harzianum* IOC-3844 Induced by Sugarcane Bagasse"**

Autores: Maria Augusta Crivelente Horta, Renato Vicentini, Priscila da Silva Delabona, Prianda Laborda, Aline Crucello, Sindélia Freitas, Reginaldo Massanobu Kuroshu, Igor Polikarpov, José Geraldo da Cruz Pradella, Anete Pereira Souza.

Publicado na revista PlosOne  
Fevereiro 2014 | Volume 9 | Issue 2 | e88689

# Transcriptome Profile of *Trichoderma harzianum* IOC-3844 Induced by Sugarcane Bagasse

Maria Augusta Crivelente Horta<sup>1</sup>, Renato Vicentini<sup>1</sup>, Priscila da Silva Delabona<sup>2</sup>, Prianda Laborda<sup>1</sup>, Aline Crucello<sup>1</sup>, Sindélia Freitas<sup>2</sup>, Reginaldo Massanobu Kuroshu<sup>3,4</sup>, Igor Polikarpov<sup>3</sup>, José Geraldo da Cruz Pradella<sup>2</sup>, Anete Pereira Souza<sup>1,5\*</sup>

**1** Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas (UNICAMP), Campinas, SP, Brazil, **2** Brazilian Bioethanol Science and Technology Laboratory (CTBE), Brazilian Center of Research in Energy and Materials (CNPEM), Campinas, SP, Brazil, **3** Physics Institute of São Carlos, University of São Paulo (USP), São Carlos, SP, Brazil, **4** Institute of Science and Technology, Federal University of São Paulo (UNIFESP), São José dos Campos, SP, Brazil, **5** Department of Plant Biology, Biology Institute, University of Campinas (UNICAMP), Campinas, SP, Brazil

## Abstract

Profiling the transcriptome that underlies biomass degradation by the fungus *Trichoderma harzianum* allows the identification of gene sequences with potential application in enzymatic hydrolysis processing. In the present study, the transcriptome of *T. harzianum* IOC-3844 was analyzed using RNA-seq technology. The sequencing generated 14.7 Gbp for downstream analyses. *De novo* assembly resulted in 32,396 contigs, which were submitted for identification and classified according to their identities. This analysis allowed us to define a principal set of *T. harzianum* genes that are involved in the degradation of cellulose and hemicellulose and the accessory genes that are involved in the depolymerization of biomass. An additional analysis of expression levels identified a set of carbohydrate-active enzymes that are upregulated under different conditions. The present study provides valuable information for future studies on biomass degradation and contributes to a better understanding of the role of the genes that are involved in this process.

**Citation:** Horta MAC, Vicentini R, Delabona PdS, Laborda P, Crucello A, et al. (2014) Transcriptome Profile of *Trichoderma harzianum* IOC-3844 Induced by Sugarcane Bagasse. PLoS ONE 9(2): e88689. doi:10.1371/journal.pone.0088689

**Editor:** Vladimir N. Uversky, University of South Florida College of Medicine, United States of America

**Received:** August 1, 2013; **Accepted:** January 9, 2014; **Published:** February 18, 2014

**Copyright:** © 2014 Horta et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo- FAPESP (<http://www.fapesp.br/>) and Conselho Nacional de Desenvolvimento Científico e Tecnológico -CNPq (<http://www.cnpq.br/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: anete@unicamp.br

## Introduction

The fungus *Trichoderma harzianum* is a well-known biocontrol agent [1],[2]. Most previously published genetic studies concerning this organism have explored its molecular mechanisms of biocontrol. This biocontrol ability enables the fungus to identify and degrade cell walls, and the mechanisms that underlie these processes were explored in the present study.

Several studies have suggested that *T. harzianum* may be utilized for the production of hydrolytic enzymes from a cellulolytic complex [3],[4],[5],[6], due to its ability to produce high levels of both  $\beta$ -glucosidase and endoglucanases [7]. These studies have demonstrated that this fungus is a potential source of hydrolytic enzymes and may aid in understanding the transcriptional regulation of biomass degradation by filamentous fungi. The utilization of sugarcane bagasse as a biomass for the production of second-generation ethanol requires its degradation into mono-oligosaccharides and small oligosaccharides that may be metabolized by ethanol-producing yeast. The major bottleneck for this process is the enzymatic hydrolysis of sugarcane bagasse [8]. The hydrolytic effectiveness of an enzymatic mixture is highly dependent on the feedstock and any pretreatment it has received [9]. A strategic issue to be considered during the development of enzymatic mixtures optimized for second-generation ethanol production is the cultivation of microorganisms utilizing the lignocellulosic material that will be hydrolyzed. This cultivation method may select for enzymes that are optimal for the hydrolysis of a specific feedstock [9],[10]. One of the primary mechanisms of the adaptive processes of cells in a complex medium is the alteration of transcription levels, which can lead to the production of specialized proteins, differences in membrane composition and other changes in cellular machinery [11].

A large variety of enzymes with different specificities are required to degrade the components of lignocellulose [10],[12],[13],[14]. However, many other proteins may also contribute to lignocellulose degradation in ways that are not yet clearly understood, such as the glycoside hydrolase family 61 proteins, the expansins and the swollenins [10],[14],[15]. Three types of enzymes are required to hydrolyze cellulose into glucose monomers: exo-1,4- $\beta$ -glucanases, such as EC 3.2.1.91 and EC 3.2.1.176 (cellobiohydrolase); endo-1,4- $\beta$ -glucanases, such as EC 3.2.1.4; and  $\beta$ -glucosidases, such as EC 3.2.1.21 (cellobiases) [10],[16]. Cellobiohydrolases attack the reducing or nonreducing ends of the cellulose chains, whereas endo-glucanases cleave these chains in the middle and reduce the degree of polymerization [10],[17]. The composition of hemicellulose is more variable than that of cellulose; therefore, more enzymes are required for its effective hydrolysis. The enzymes that degrade hemicellulose can be divided into depolymerizing enzymes, which cleave the backbone of the molecule, and enzymes that remove the substituent of the molecule, which may sterically hinder the depolymerizing enzymes. The core enzymes for the degradation of xylan to monomers are the endo-xylanases, which cleave the xylan backbone into shorter oligosaccharides, and  $\beta$ -xylosidase, which cleaves short xylo-oligosaccharides into xylose. Similarly, the core enzymes for the degradation of mannan are endo-mannanase and  $\beta$ -mannosidase. However, xylans and mannans generally contain a number of different substituents linked to their main backbones, including arabinose, acetyl groups, galactose and glucose. A host of ancillary enzymes are required to remove these substituents and allow the core enzymes to degrade the xylan and mannan backbones. These ancillary enzymes include the  $\alpha$ -L-arabinofuranosidases,  $\alpha$ -glucuronidase, ferulic acid esterase,  $\alpha$ -galactosidase,



feruloyl esterase, acetyl xylanesterase and acetyl mannan esterase. The ferulic acid esterases specifically cleave the linkages between hemicellulose and lignin. The  $\alpha$ -L-arabinofuranosidases also possess different specificities; some cleave 1,2 linkages or 1,3 linkages, whereas others cleave doubly substituted arabinose residues from arabinoxylan [10],[18].

Fungi from the genera *Trichoderma*, *Penicillium*, *Aspergillus* and *Humicola grisea var. thermoidea* degrade lignocellulose components, including sugarcane bagasse [8]. These fungi can degrade cellulose, hemicellulose and lignin in decaying plants using a complex set of excreted hydrolytic and oxidative enzymes, including glycosyl hydrolases from different families [10]. Although many studies have been conducted to characterize the action of the enzymes involved in lignocellulose degradation, little is known regarding the transcription and genomic regulation of the genes that encode these enzymes. *Trichoderma reesei* is the major industrial source of the cellulases and hemicellulases that are utilized in the depolymerization of biomass to simple sugars, which are then further converted into chemical intermediates and biofuels. Unexpectedly, despite the industrial utility and effectiveness of the carbohydrate-active enzymes of *T. reesei*, the genome of this species encodes fewer cellulases and hemicellulases than that of any other sequenced fungus that can hydrolyze plant cell wall polysaccharides [19],[20]. Thus, a better understanding of the genetic mechanisms of this fungus is necessary to explore its extraordinary biotechnological potential. The present study analyzes the transcriptome of *T. harzianum* IOC-3844 grown in a sugarcane bagasse-based culture medium and the induction of hydrolytic activity in this medium, with particular emphasis on the potential contributions of the fungus to fuel biotechnology and other industrial applications. This organism is available in

public collections, and studies addressing the mechanisms of regulating and gene expression in this fungus are important to make its use in biotechnological processes viable. This work seeks to contribute to the understanding of the reactions involved in biomass degradation at the enzymatic level and will serve as the basis for other studies exploring the biotechnological potential presented by *T. harzianum*. The primary goal of these analyses was to identify, characterize and catalog the transcripts expressed by *T. harzianum* that are involved in the degradation of complex substrates, thereby revealing the complexity of the hydrolytic pathways involved in biomass degradation.

## **Materials and Methods**

Regarding the Ethics Statement, we confirm that no specific permits were required for the present studies. Additionally, we confirm that the field studies did not involve endangered or protected species.

### **Strain and Culture Media**

The *T. harzianum* IOC-3844 strain used in this study was provided by Professor Dr. Nei Pereira Jr. (Federal University of Rio de Janeiro, Rio de Janeiro, Brazil). The species was confirmed by comparing its ITS1 and ITS2 sequences with those of standard strains of *T. harzianum*. (available at [Institute Oswaldo Cruz](#), Rio de Janeiro, RJ, Brazil and Centro de Pesquisas Químicas e Biológicas na Agricultura (CPQBA) - [CBMAI](#), UNICAMP, Campinas, SP, Brazil). The stock cultures were stored at 4°C on potato dextrose agar (PDA) slants. The fungi were grown on PDA plates (90 x 15 mm) at 29°C for 8 days.

The composition of the basal medium was adapted from Mandels and Weber (1969) [21] and included ( $\text{g L}^{-1}$ )  $\text{KH}_2\text{PO}_4$  (2.0),  $\text{NH}_4\text{SO}_4$  (1.4),  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  (0.3),  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  (0.3),  $\text{CoCl}_2$  (0.002),  $\text{MnSO}_4 \cdot \text{H}_2\text{O}$  (0.0016),  $\text{ZnSO}_4 \cdot \text{H}_2\text{O}$  (0.0014),  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$  (0.005) and urea (0.3). The pH was adjusted to 5.2. Three different preculture media for mycelial production were prepared from the basal medium through the addition of  $2 \text{ g L}^{-1}$  glucose,  $1 \text{ g L}^{-1}$  peptone,  $1 \text{ ml L}^{-1}$  Tween 80 and  $10 \text{ g L}^{-1}$  of a carbon source. The carbon sources used in the three preculture media were lactose, crystalline cellulose and delignified sugarcane bagasse (DSB, from a local mill, Usina Vale do Rosário, Orlandia, SP, Brazil), which was prepared and characterized according to Rocha *et al.* (2012) [22]. The percentage composition of the DSB was  $89.5 \pm 1.6$  cellulose,  $3.4 \pm 0.3$  hemicellulose and  $5.5 \pm 0.2$  lignin [4]. The preculture media were sterilized at  $121^\circ\text{C}$  for 20 min.

The production medium was composed of the basal medium,  $10 \text{ g L}^{-1}$  DSB as a unique carbon source,  $1 \text{ g L}^{-1}$  peptone and  $1 \text{ ml L}^{-1}$  Tween 80; the medium was then sterilized at  $121^\circ\text{C}$  for 20 min. All other chemicals were of at least analytical grade. The following libraries were classified based on the preculture media: the "DSB" library was generated from a preculture medium that contained DSB, the "CEL" library was generated from a preculture medium that contained crystalline cellulose, and the "LAC" library was generated from a preculture medium that contained lactose. This latter condition was designated as the control.

### **Preculture and fermentation**

Conidial suspensions were prepared through the addition of sterilized distilled water and Tween 80 to the PDA plates, which resulted in conidial suspensions

of  $9 \times 10^5$  spores  $\text{ml}^{-1}$ . After preparation, 4.0 ml of each conidial suspension was transferred to Erlenmeyer flasks containing 600ml of each preculture medium, and the flasks were incubated for 72 h at  $29^\circ\text{C}$  on a rotary shaker at 200 rpm. A volume of 30 ml of each medium was transferred to individual Erlenmeyer flasks containing 270 ml of the production medium. The flasks were incubated at  $29^\circ\text{C}$  for 129h on a rotary shaker at 200 rpm. Samples of the mycelia and the fermentation extracts were removed to determine the enzymatic activity and to conduct the transcriptome analyses.

### **Analytical measurements**

The filter paper activity (FPase) was determined as described by Ghose (1987)[23], with modifications to diminish the scale of the procedure by a factor of 10. All statistical comparisons were done using Student's t test ( $P < 0.05$ ).

### **RNA extraction and transcriptome sequencing**

The mycelial samples from the LAC, CEL and DSB conditions were extracted after 96 h of fermentation, stored at  $-70^\circ\text{C}$  and used for RNA extraction. The fungal mRNA was isolated according to Jones *et al.* (1985) [24] with some modifications. Two grams of each mycelial sample was frozen using liquid nitrogen in a mortar and ground with a pestle into a fine powder. Next, NTES buffer (4.5 ml, 0.1 M NaCl, 0.01 M Tris-HCl at pH 7.5, 1 mM EDTA and 1% SDS) and phenol/chloroform/isoamyl alcohol [3 ml of a 1:1:1 mixture] were added, and the sample was ground until the mixture had thawed. After vortexing for 10 min, the solution was centrifuged at 8,000 rpm for 10 min at

4°C. To the aqueous phase, 3 ml of phenol/chloroform/isoamyl alcohol [1:1:1 mixture] was added, and the solution was centrifuged at 8,000 rpm for 10 min. The aqueous phase was then removed, and the nucleic acid was precipitated through the addition of a 0.1 volume aliquot of 2 M NaAc, pH 4.5, and two volumes of 100% ethanol. The precipitate was centrifuged at 8,000 rpm for 10 min, and the pellet was resuspended in 2.5 ml of sterile water. To remove the DNA, 2.5 ml of 4 M LiAc was added, and the solution was incubated for 48 h at -20°C. The precipitate was collected by centrifugation at 8,000 rpm for 10 min, then washed with 70% ethanol, resuspended in 50 µL of sterile water and stored at -70°C.

The RNA samples were quantitated using a fluorescence-based method, and their quality was determined using a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA).

The libraries were constructed using 4 µg of each RNA sample and the TruSeq RNA sample preparation kit (Illumina Inc., San Diego, CA) according to the manufacturer's instructions. The expected target sizes were confirmed using a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA), and the libraries were quantified using qPCR. The average insertion size was 260 bp. The clustering was conducted using 10 µM of each library and a TruSeq PE Cluster Kit on cBot (Illumina Inc., San Diego, CA). The sequencing was performed on the Illumina Genome AnalyzerIIx, which is a next-generation high-throughput sequencer (Illumina Inc., San Diego, CA), according to the manufacturer's specifications for paired-end reads of 72 bp in individual lanes.

## **Downloading and processing the sequence data**

After the sequencing was completed, the data were transferred to a local high-performance computing server at CBMEG (University of Campinas, Campinas, Brazil). The results were submitted to NCBI under accession number SRX189214, and the raw sequences (archives of paired and paired-end sequences) were submitted to the NCBI Sequence Read Archive (SRA) under accession numbers SRR579379, SRR631745 and SRR631746 for the DSB, CEL and LAC libraries, respectively. Initially, all sequences were trimmed. We utilized the CLC Genomics Workbench (v4.0; Finlandsgade, Dk) to perform the reads trimming, and parameters were set to: quality limit: 0,03; ambiguous limit: 2; minimum final number of nucleotides in reads: 65; phred scale: 15. *De novo* assembly was conducted using CLC Genomics Workbench (v4.0; Finlandsgade, Dk) with the following parameters: similarity=0.98 and length fraction=0.9. The resulting contigs were compared with the NCBI non-redundant protein database (NR) using [BLAST](#) to identify homologous sequences [25], with an *E*-value cutoff of  $\leq 1 e^{-5}$ . The sequences were functionally annotated according to Gene Ontology terms [26], and the annotations were compared with the Kyoto Encyclopedia of Genes and Genomes [28] to establish biochemical pathway associations using [Blast2Go](#), which is a universal web-based annotation application [27] The sequences were aligned against the Carbohydrate-Active Enzymes (CAZymes) database to identify glycosyl hydrolases [29],[30],[31],[32],[33], glycosyltransferases [34],[35], carbohydrate-binding modules [36] and carbohydrate esterases [37]. The *T. harzianum* IOC-3844 genome was provided by Dr. Reginaldo M. Kuroshu (University of São Paulo, São Carlos, Brazil). Only the best alignments showing expectation values lower than  $1 \times 10^{-5}$  were considered for functional gene annotation. To compare the transcript sequences with the

genome and CAZyme datasets, CLC Genomics Workbench was used. The archive of the assembly scaffolds for *T. harzianum* CBS 226.95, which is available on JGI (sequence data produced by the US Department of Energy Joint Genome Institute in collaboration with the user community) [38], was used to calculate the similarity between the data.

### **Expression pattern**

A paired Kal's *t*-test was conducted on the  $\log_2$ -transformed data to determine whether significant differences existed between the expression ratios found in each treatment and the control. Contigs were identified as being differentially expressed in upregulated groups when significance was detected with a false discovery rate lower than  $1 \times 10^{-3}$ . Hierarchical clustering analysis and K-means clustering were performed on the CAZymes that were identified as being differentially expressed. Clustering was performed using Euclidean distance as the distance metric in three partitions according to the cluster features, on the transformed expression values.

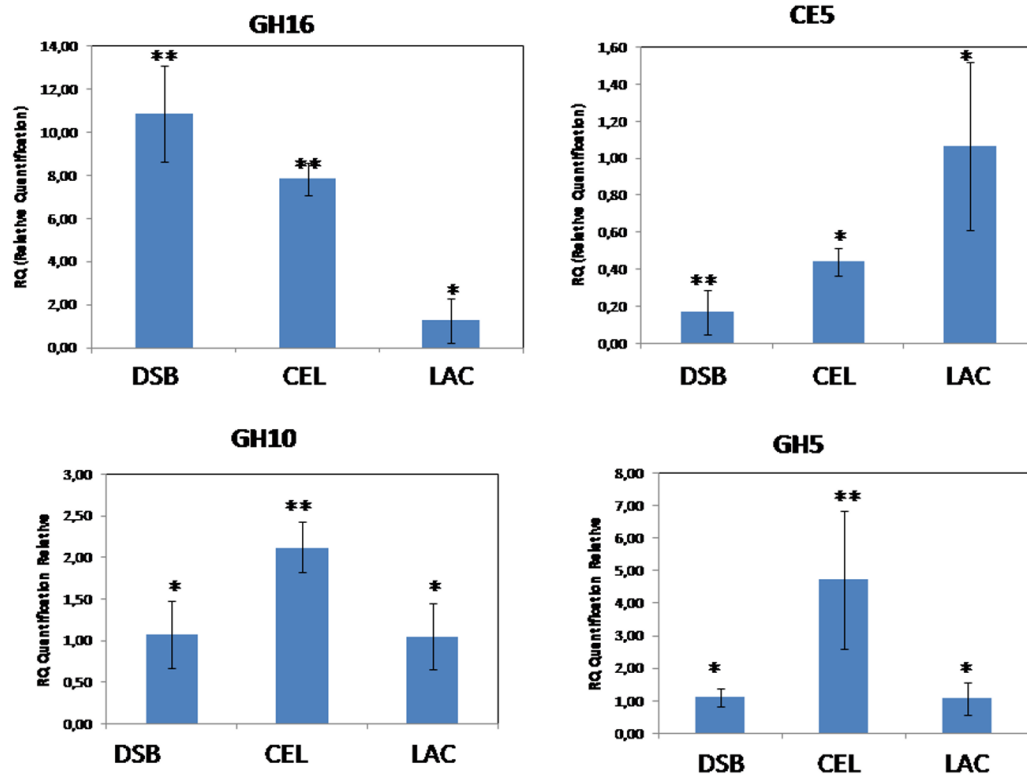
### **Real-Time PCR analysis**

To validate the expression profiles of the assembled genes obtained through sequencing data analysis, quantitative real-time (RT-qPCR) was performed for selected genes. Genes associated with biomass degradation processes were selected and are shown in Supplementary Table S1, together with the primers and annealing temperatures.

Quantification of gene expression was performed by continuously monitoring SYBR Green fluorescence. The reactions were performed in triplicate in a total volume of 6.25  $\mu$ l. Each reaction included 3.12  $\mu$ l of SYBR Green Master Mix

(Invitrogen, Carlsbad, CA), 1.0  $\mu$ l of direct and reverse primers, 0.5  $\mu$ L of cDNA and 1.6  $\mu$ l of water. The reactions were assembled in 384-well plates. PCR amplification-based expression profiling of the selected genes was performed using a gene for squalene-epoxidase as endogenous control. Four genes were tested as endogenous control: genes for actin, beta-tubulin, glyceraldehyde 3-phosphate dehydrogenase, and squalene-epoxidase. The last one had the best performance in RT-qPCR analysis, remaining constant in all treatments. The enzyme squalene-epoxidase catalyses the conversion of squalene to 2,3-(S) oxidosqualene, which is an intermediate in the synthesis of the fungal cell membrane component ergosterol. RT-qPCR was conducted in an ABI PRISM 7500 HT (Applied Biosystems, Foster City, CA). Gene expression was calculated via the Delta-Delta cycle threshold method [47]. All statistical comparisons were done using Student's t test ( $P < 0.05$ ). The obtained RT-qPCR results were in agreement with the RNA expression analyses of the generated assemblies. The same expression profile was observed for the genes encoding GH16, GH10, CE5, and GH5. Figure 1 shows the expression of the selected genes.





**Figure 1.** Expression profiles of selected glycosyl hydrolase genes determined by RT-qPCR. The squalene-epoxidase gene was used as endogenous control. The differences between groups were considered significant at  $P < 0.05$  (Student's t test) and are indicated by \*.

## Results

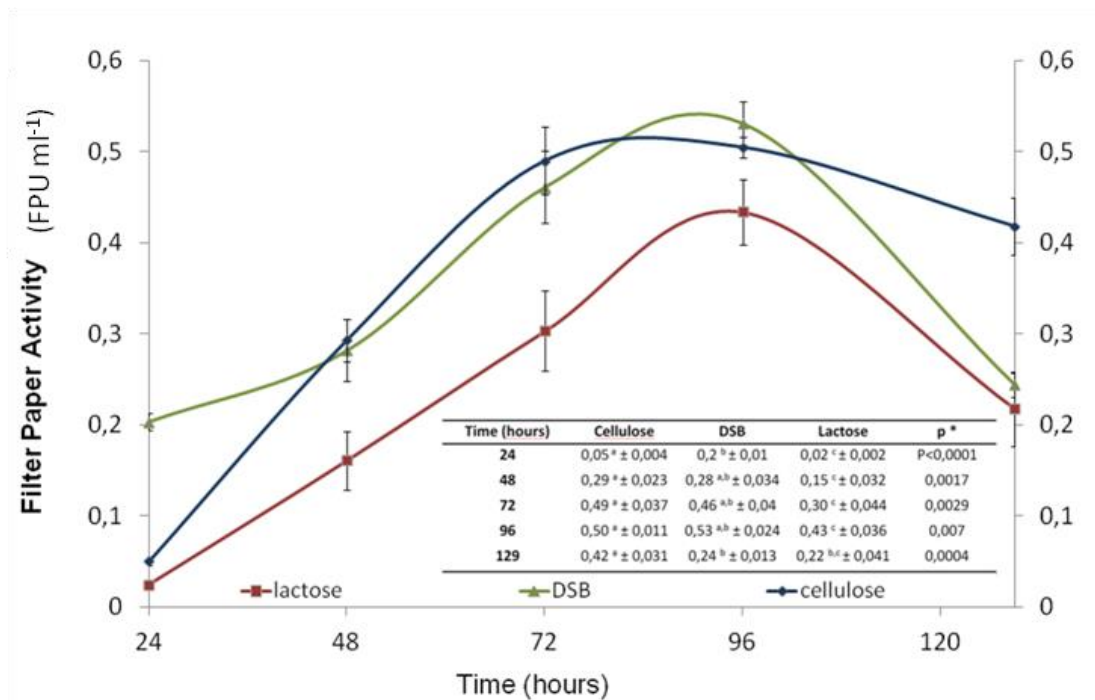
### Enzymatic Activity Profile

The FPase was evaluated to determine the enzymatic activity profile of the cellulases during 129 h of fermentation (Figure 2) using DSB as a carbon source. RNA was isolated from the mycelia at 96 h of cultivation; this time point was associated with a significant production of FPase ( $0.53 \text{ FPU ml}^{-1}$ ) (Figure 2). This 96 h cultivation period included a 48 h adaptation phase.

Previous studies have indicated that the enzymatic activity of this fungus, as measured by cellulase production, is lower when grown on soluble carbon sources than when grown on DSB, and these results are in agreement with the present study [4]. The substrate acts as both an adhesion surface and as fermentable biomass for the fungi, and it activates the synthesis of hydrolytic complexes.

The increased FPase activity indicated the enhanced ability of the fungus to metabolize and degrade compounds in the biomass. The maximum activity was observed between 72 and 96 h, after which the activity was repressed due to the formation of degradation products.

To identify the origin of the enzymatic activity in the extracts, we evaluated the transcriptome expression at 96 h of fermentation.



**Figure 2. Filter paper activity enzymatic profiles (FPU ml<sup>-1</sup> h<sup>-1</sup>).** *T. harzianum* fermentation on a complex substrate (DSB) from extracts grown on preculture media using DSB ( $\Delta$ ), cellulose ( $\diamond$ ) or lactose ( $\square$ ) as the carbon source. The differences were considered significant at  $P < 0.05$  (Student's t-test) and are indicated by (a) for cellulose, (b) for DSB and (c) for lactose samples.

### Sequencing assembly

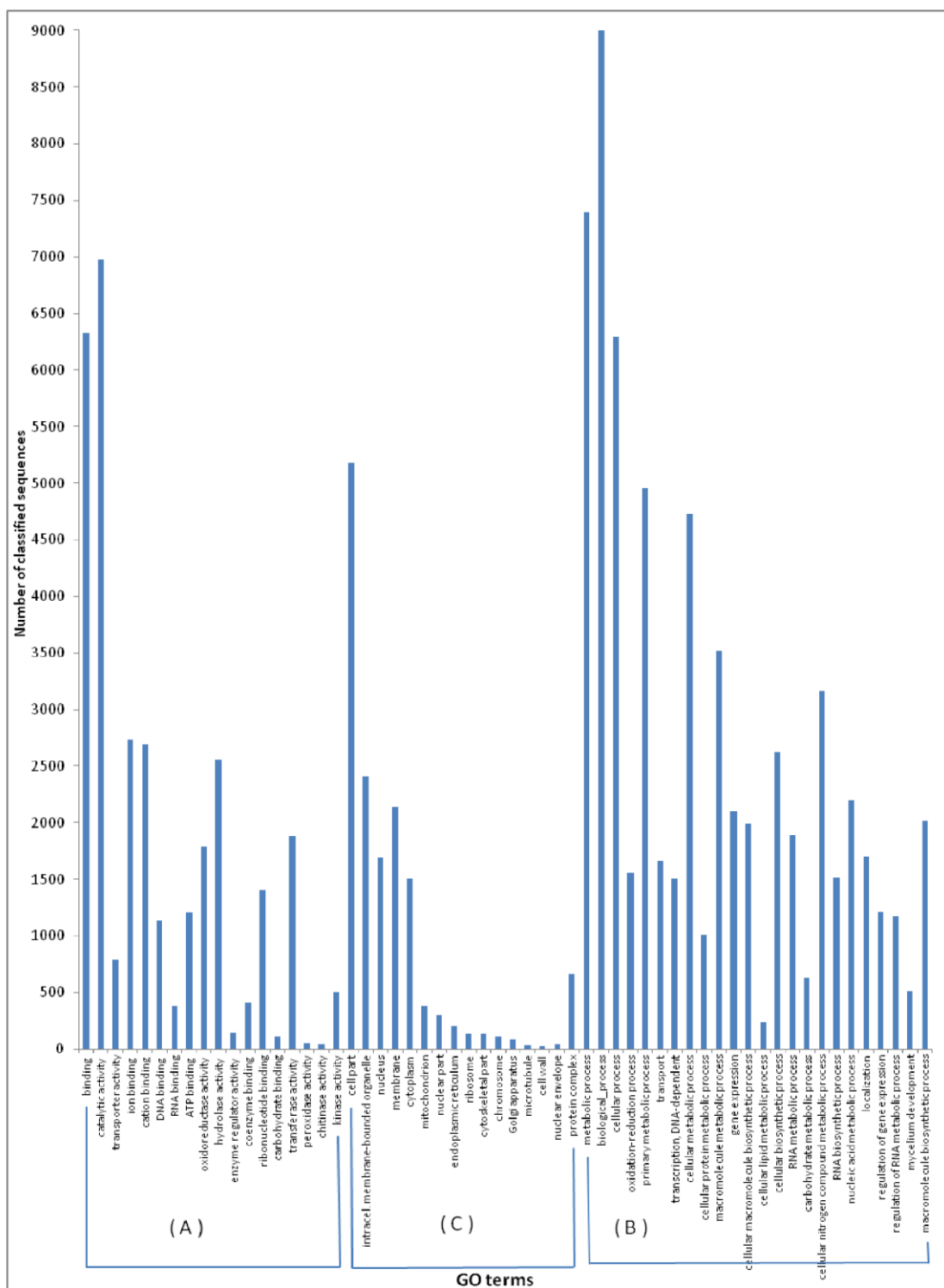
In total, 246 million raw sequencing reads were generated with a target length of 72 bp (Table 1). After quality trimming, 84.11% of the data were retained for a total of 14.7 Gbp of sequencing data. *De novo* assembly using trimmed reads from all libraries resulted in 32,494 contigs, with an N50 of 1,251 bp. The assembled transcripts redundancy was determined through CD-HIT-EST. After this analysis, the final number of contigs was 32,396.

**Table 1. Results of next-generation sequencing, trimming analysis, *de novo* assembly and mapping.** The raw reads were *de novo* assembled to generate contigs for further analysis and annotation.

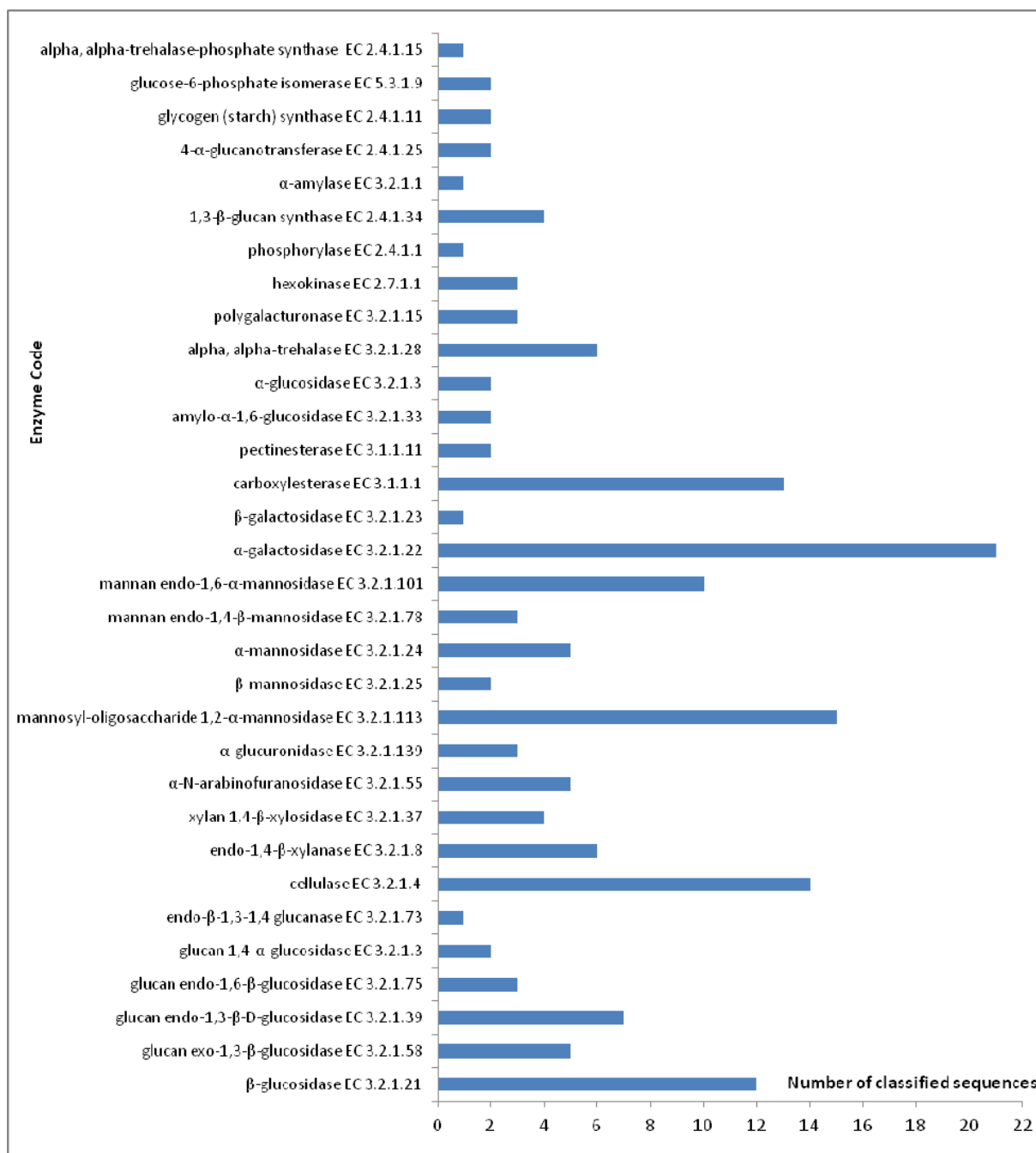
Library	DSB	CEL	LAC	Length (bp)	N50
<b>Number of Raw reads</b>	81,705,758	84,301,646	80,468,986	<b>72.0</b>	
<b>Number of trimmed reads</b>	<b>68,720,401</b>	<b>68,644,205</b>	<b>67,912,155</b>	<b>71.8</b>	
Single reads	5,820,501	6,451,047	5,586,409		
Paired reads	62,899,900	62,193,158	62,325,746		
<b>Number of Mapped reads</b>	<b>28,377,065</b>	<b>28,663,344</b>	<b>29,343,868</b>		
Single reads	2,463,163	2,803,982	2,477,029		
Paired reads	25,913,902	25,859,362	26,866,839		
<b>Unmapped reads</b>	<b>40,343,336</b>	<b>39,980,861</b>	<b>38,568,287</b>		
Single reads	3,357,338	3,647,065	3,109,380		
Paired reads	36,985,998	36,333,796	35,458,907		
<b>Contigs</b>	<b>32,494</b>				<b>1,251</b>
<b>CD-HIT-EST contigs</b>	<b>32,369</b>				

### **Analysis of the transcriptome under the influence of sugarcane bagasse as a substrate**

Transcript profiling is an important strategy for studying the expression of large gene sets under particular conditions. To determine the influence of the complex sugarcane bagasse substrate on gene expression, the contigs generated from the *de novo* assembly of the transcriptome were analyzed. The generated assembly was compared with the archive of the assembly scaffolds for *T. harzianum* CBS 226.95 [38], which demonstrated a similarity of 96% with the contigs derived from the assembly of the transcriptome. To identify the responses of the transcriptome under the different conditions tested, the contigs were annotated and classified according to their predicted functions (Figure 3).



**Figure 3. Molecular functions, biological process distribution and cellular localization of the transcriptome assembly.** Contigs were assigned putative classifications based on homology and evaluated for their predicted involvement in molecular functions (A), biological processes (B) and cellular localization (C).



**Figure 4. Identified sequences that catalyze reactions that are potentially involved in biomass degradation.** The results of identification based on homology using the NCBI NR database indicate the presence of genes that are related to the depolymerization of biomass in the transcriptome.

The high number of sequences generated in this study, which were produced only through the use of next-generation sequencing, allows a precise overview of the different biological processes that occur in an organism at a given moment, and classifying these sequences allows for analysis of the genes that may be involved in biomass degradation. Sequences that were classified as possessing catalytic activity (6,975) or regulating enzymatic activity (143) may be involved in biomass degradation. When analyzed according to biological processes, the majority of the annotations were classified as participating in metabolic processes (7,393), followed by cellular processes (6,294). Regarding molecular functions, binding and catalytic activity were the most frequent classifications. Concerning cellular components, genes involved in cellular (5,184) and organelle (2,665) components and the membrane (2,143) were the most abundant.

For the hydrolysis of complex substrates such as sugarcane bagasse, a microorganism must produce an array of specialized enzymes that can hydrolyze lignocelluloses. The interaction between different classes of enzymes has been extensively studied [12],[13] and was observed in the present study. Figure 4 summarizes several of the enzyme classes that are potentially involved in biomass degradation and the number of contigs assigned to each of them. Of the contigs formed after assembly, 36.18% were classified according to GO terms, with 21.46% being involved in catalytic reactions. Approximately 164 contigs were classified as being potentially involved in metabolic reactions related to biomass degradation.

The  $\beta$ -glucosidase classification are specifically involved in the hydrolysis of cellulose. These sequences catalyze the hydrolysis of terminal, nonreducing  $\beta$ -D-glucose residues through the release of  $\beta$ -D-glucosidase (EC 3.2.1.21) and



glucan 1,4- $\alpha$ -glucosidase (EC 3.2.1.3), which in turn catalyze the hydrolysis of terminal (1 $\rightarrow$ 4)-linked  $\alpha$ -D-glucose residues from the nonreducing ends of the chains. Both of these steps release  $\beta$ -D-glucose, which is the monomer that is further metabolized.

Hemicellulose possesses a more varied composition than cellulose and requires enzymes to be effectively hydrolyzed. Sequences were classified as being involved in the degradation of xylan to monomers, including both endo-xylanases (EC 3.2.1.8), which cleave the xylan backbone into shorter oligosaccharides, and  $\beta$ -xylosidase (EC 3.2.1.37), which cleaves short xylo-oligosaccharides into xylose. Similarly, sequences were related to mannan degradation (EC 3.2.1.113, EC 3.2.1.25, EC 3.2.1.24, EC 3.2.1.78 and EC 3.2.1.101). Several ancillary enzymes were also identified, including  $\alpha$ -glucuronidase (EC 3.2.1.139),  $\alpha$ -galactosidase (EC 3.2.1.22) and arabinofuranosidase (EC 3.2.1.55).

Systematic synergisms between the different enzyme classes could be observed for specific metabolic pathways in the *T. harzianum* transcriptome. These pathways included the metabolism of different sugars, which are associated with the depolymerization of biomass and were classified according to specific criteria of the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Table 2) [28].

**Table 2. Metabolic pathways expressed in different libraries.** The classification of the contigs according to the NCBI NR and KEGG databases [46],[28] indicates which metabolic pathway is active under a specific culture condition.

Metabolic Pathway	Library	Enzyme Code Classification
Fructose and mannose metabolism	DSB	EC 1.1.1.17: 5-dehydrogenase
	CEL	EC 3.2.1.78: endo-1,4- $\beta$ -mannosidase
Lysine biosynthesis	DSB	EC 1.2.1.31: dehydrogenase
Lysine degradation	DSB	EC 1.2.1.31: dehydrogenase
Purine metabolism	CEL	EC 2.7.7.7: DNA polymerase
	CEL	EC 3.6.1.3: adenylypyrophosphatase
	CEL	EC 3.6.1.15: nucleoside triphosphate phosphohydrolase
Thiamine metabolism	LAC	EC 3.6.1.3: adenylypyrophosphatase
	CEL	EC 3.6.1.15: nucleoside triphosphate phosphohydrolase
Methane metabolism	CEL	EC 1.14.13.8: monooxygenase
	CEL	EC 1.11.1.7: lactoperoxidase
Phenylpropanoid biosynthesis	CEL	EC 1.11.1.7: lactoperoxidase
Phenylalanine metabolism	CEL	EC 1.11.1.7: lactoperoxidase
Glycerophospholipid metabolism	CEL	EC 4..1.1.65: decarboxylase
	LAC	EC 1.1.1.8: dehydrogenase (NAD <sup>+</sup> )
Pyrimidine metabolism	CEL	EC 2.7.7.7: DNA polymerase
Starch and sucrose metabolism	CEL	EC 3.2.1.1: endo-1,4- $\beta$ -D-glucanase
	LAC	EC 3.2.1.37: 1,4- $\beta$ -xylosidase
Drug metabolism: cytochrome P450	CEL	EC 1.14.13.8: monooxygenase
Drug metabolism: other enzymes	LAC	EC 3.1.1.1: aliesterase
Riboflavin metabolism	CEL	EC 1.1.1.193: reductase
Fructose and mannose metabolism	CEL	EC 3.2.1.78: endo-1,4- $\beta$ -mannosidase
Amino sugar and nucleotide sugar metabolism	LAC	EC 3.2.1.37: 1,4- $\beta$ -xylosidase
		EC 3.2.1.55: arabinosidase
Pyruvate metabolism	LAC	EC 2.3.3.9: synthase
Nitrogen metabolism	LAC	EC 1.9.3.1: oxidase
Other glycan degradation	LAC	EC 3.2.1.45: psychosine hydrolase
Glycerolipid metabolism	LAC	EC 1.1.1.72: dehydrogenase (NADP <sup>+</sup> )
Oxidative phosphorylation	LAC	EC 1.9.3.1: oxidase
Glyoxylate and dicarboxylate metabolism	LAC	EC 2.3.3.9: synthase

Sphingolipid metabolism	LAC	EC 3.2.1.45: psychosine hydrolase
Fructose and mannose metabolism	DSB	EC 1.1.1.17: 5-dehydrogenase
	CEL	EC 3.2.1.78: endo-1,4- $\beta$ -mannosidase
Lysine biosynthesis	DSB	EC 1.2.1.31: dehydrogenase
Lysine degradation	DSB	EC 1.2.1.31: dehydrogenase
Purine metabolism	CEL	EC 2.7.7.7: DNA polymerase
	CEL	EC 3.6.1.3: adenylypyrophosphatase
	CEL	EC 3.6.1.15: nucleoside triphosphate phosphohydrolase
	LAC	EC 3.6.1.3: adenylypyrophosphatase
Thiamine metabolism	CEL	EC 3.6.1.15: nucleoside triphosphate phosphohydrolase
Methane metabolism	CEL	EC 1.14.13.8: monooxygenase
	CEL	EC 1.11.1.7: lactoperoxidase
Phenylpropanoid biosynthesis	CEL	EC 1.11.1.7: lactoperoxidase
Phenylalanine metabolism	CEL	EC 1.11.1.7: lactoperoxidase
Glycerophospholipid metabolism	CEL	EC 4.1.1.65: decarboxylase
	LAC	EC 1.1.1.8: dehydrogenase (NAD <sup>+</sup> )
Pyrimidine metabolism	CEL	EC 2.7.7.7: DNA polymerase
Starch and sucrose metabolism	CEL	EC 3.2.1.1: endo-1,4- $\beta$ -D-glucanase
	LAC	EC 3.2.1.37: 1,4- $\beta$ -xylosidase
Drug metabolism: cytochrome P450	CEL	EC 1.14.13.8: monooxygenase
Drug metabolism: other enzymes	LAC	EC 3.1.1.1: aliesterase
Riboflavin metabolism	CEL	EC 1.1.1.193: reductase
Fructose and mannose metabolism	CEL	EC 3.2.1.78: endo-1,4- $\beta$ -mannosidase
Amino sugar and nucleotide sugar metabolism	LAC	EC 3.2.1.37: 1,4- $\beta$ -xylosidase
		EC 3.2.1.55: arabinosidase
Pyruvate metabolism	LAC	EC 2.3.3.9: synthase
Nitrogen metabolism	LAC	EC 1.9.3.1: oxidase
Other glycan degradation	LAC	EC 3.2.1.45: psychosine hydrolase
Glycerolipid metabolism	LAC	EC 1.1.1.72: dehydrogenase (NADP <sup>+</sup> )
Oxidative phosphorylation	LAC	EC 1.9.3.1: oxidase
Glyoxylate and dicarboxylate metabolism	LAC	EC 2.3.3.9: synthase
Sphingolipid metabolism	LAC	EC 3.2.1.45: psychosine hydrolase

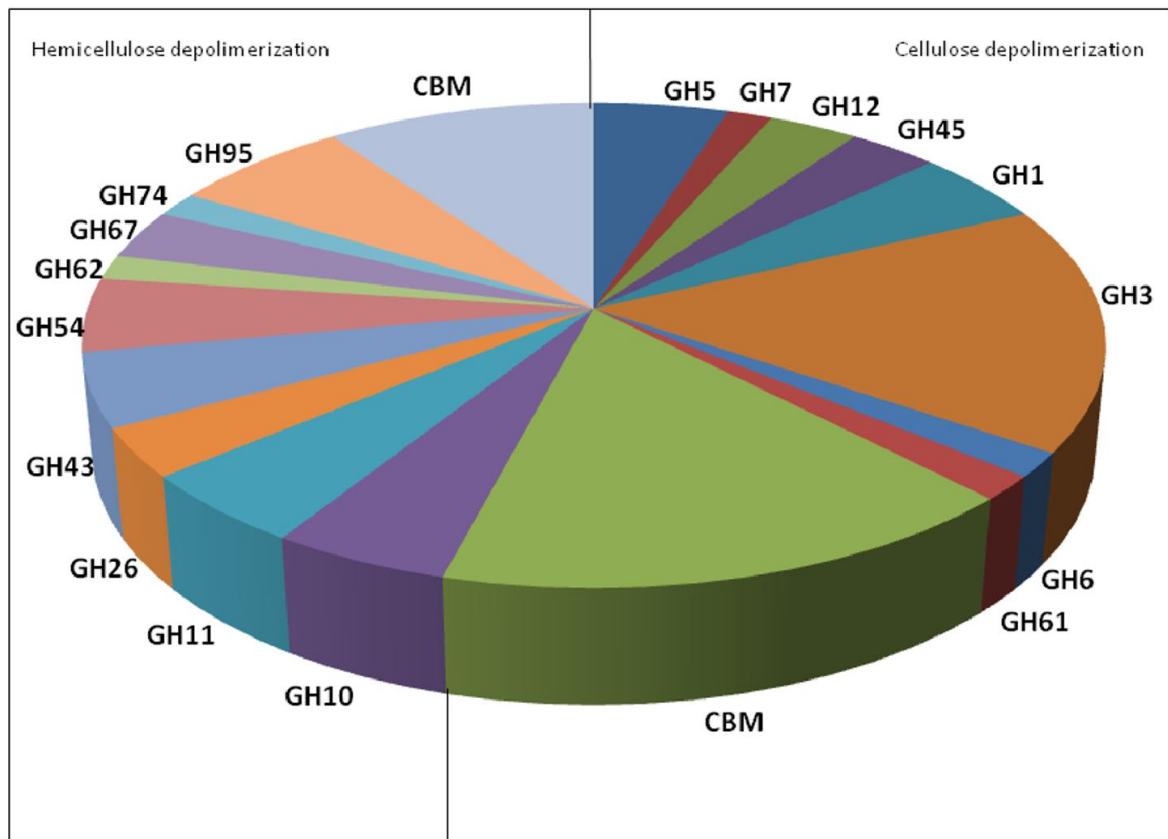
In this manner, different classes of enzymes that act together to degrade the cellulose backbone were identified. The application of enzymes to catalyze the degradation of cellulose to glucose and heteroxylans to pentose is now

considered to be the most viable strategy for providing cost-efficient second-generation ethanol processes [39],[18], and the present study confirms that a variety of different metabolic pathways are necessary for sugar degradation in this yeast.

### **Classification according to CAZymes**

To determine the number of encoded genes related to biomass degradation in the transcriptome, we searched for the following carbohydrate-active enzyme groups: glycoside hydrolases (GH), glycosyltransferases (GTs), carbohydrate esterases (CEs) and the corresponding carbohydrate-binding modules (CBMs). We compared the transcriptome with a group of annotated sequences from the CAZymes database, including the annotated sequences for 17 cellulases from *T. reesei*. To identify CAZymes in *T. harzianum* IOC-3844, the assembled transcriptome was aligned against the specific CAZy dataset, and only the best alignment was considered for each gene sequence. A total of 527 CAZymes were identified in the *T. harzianum* IOC-3844 predicted gene set (unpublished results). We identified 487 predicted CAZymes in the transcriptome using a cutoff *E*-value of  $1 \times 10^{-5}$ . From the predicted CAZymes, we identified 23 genes that encoded proteins of the glycoside hydrolase families that are involved in cellulose depolymerization: three genes were classified as GH5, one gene as GH7, two genes as GH12, two genes as GH45, three genes as GH1, 10 genes as GH3, one gene as GH6 and one gene as GH61. In the cellulose depolymerization group, we found 10 different sequences that encoded carbohydrate-binding modules. Regarding genes involved in hemicellulose depolymerization, 22 genes were identified from the following glycoside hydrolase families: three genes from GH10, three genes from GH11, two genes

from GH26, three genes from GH43, three genes from GH54, one gene from GH62, two genes from GH67, one gene from GH74 and four genes from GH95. Six carbohydrate-binding module sequences were classified as belonging to enzymes that degrade hemicellulose (Figure 5).



**Figure 5. Encoded genes related to cellulose and hemicellulose depolymerization.** Genes classified in the transcriptome analysis of *T. harzianum* fermentation on sugarcane bagasse.

### Comparative expression analysis

To analyze the differences in expression levels among the tested growth conditions, we compared the total assembly generated from all of the

sequenced transcriptome libraries (DSB+CEL+LAC library) with each individual transcriptome assembly. The mapping results are shown in Table 1. To identify the transcriptomic responses under each condition, we analyzed the distribution of the genes that were identified as being differentially expressed. Pairwise comparisons of the subsets indicated the total number of genes that were overexpressed under each condition. The classification of differentially expressed contigs allowed us to determine the set of genes for carbohydrate-active enzymes that were upregulated in each group (Table 3, Figure 6 and 7).

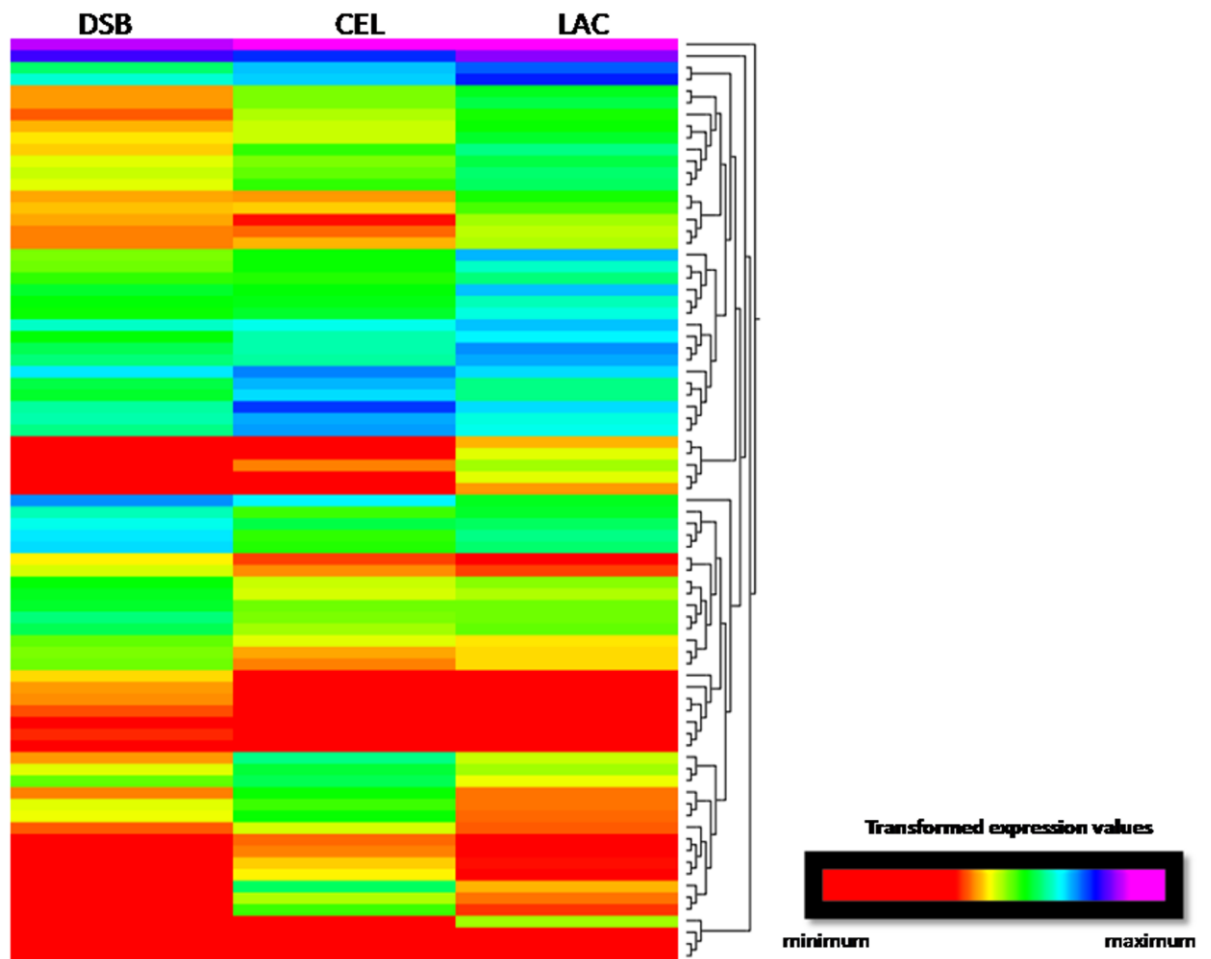
**Table 3. Classification of sequences present in upregulated groups, according to the CAZyme database.** The sequences presenting expectation values lower than  $1 \times 10^{-3}$ , and the best alignment scores are summarized. The clusters are classified according to Figure 6 and 7.

Upregulated group	Cluster	Contig	Length (bp)	Lowest E-value	CAZy	RPKM		
						DSB	CEL	LAC
DSB	1	6765	1,521	0.00	GT69	58.25	28.90	29.42
		7294	1,155	3.72E-33	GT39	84.76	51.53	42.58
		14029	3,726	0.00	GT2	97.16	35.91	58.69
		15954	665	3.74E-13	GH28	64.39	28.07	32.29
		16314	1,044	0.00	GH76	144.84	88.02	98.48
		19203	218	1.86E-34	GH28	293.39	163.56	212.91
		19252	218	1.37E-47	GT2	81.63	42.09	33.30
		19677	250	1.52E-24	GT2	58.78	31.54	24.40
		20009	209	5.02E-35	GH16	103.86	68.61	58.86
		21947	1,631	0.00	GT2	194.11	123.70	142.81
		22621	1,184	0.00	GH125	151.61	80.77	97.54
		23959	898	0.00	GT20	251.35	144.52	151.95
		24020	330	2.94E-173	GT4	359.77	167.88	260.13
		24053	1,141	0.00	GT48	210.54	152.82	151.10
		24197	823	1.37E-36	CE11	156.59	115.13	101.66
		24491	696	0.00	GH128	332.95	226.17	240.71
		25186	416	2.98E-6	CE11	228.13	132.09	154.75

		27900	309	5.9E-42	GH78	373.50	176.02	247.26
		28105	267	2.72E-97	GH72	201.54	119.69	131.08
		28257	262	1.35E-19	GH1	70.75	37.95	31.09
		29726	1,876	0.00	GH5	118.25	83.09	68.79
		31930	361	3.11E-150	GH43	447.09	355.16	205.51
<b>CEL</b>	<b>2</b>	15484	1,124	0.00	GT24	37.62	129.12	77.71
		15510	596	0.00	GH5/CBM1	49.83	239.69	88.81
		15808	642	1.32E-64	GH5	54.56	95.81	60.99
		17441	1,93	5.04E-34	GH10	83.44	257.14	120.21
		19410	217	2.19E-39	GH3	24.20	49.84	17.90
		19509	227	3.56E-37	GT2	287.20	425.57	320.97
		19636	212	5.34E-29	GT2	38.56	103.35	54.65
		19662	220	7.85E-39	GT2	115.65	220.27	135.39
		20286	203	2.77E-26	GH11	954.08	2,397.6	1,835.5
		20620	1,993	9.99E-31	CE15	16.83	58.90	22.72
		23663	557	0.00	CBM20	257.18	433.50	330.20
		23934	1,045	0.00	GT2	221.66	416.88	256.88
		24061	1,256	0.00	GT2	115.31	164.11	77.22
		24114	526	0.00	GT48	363.92	467.54	372.60
		24258	1,118	0.00	GT2	111.80	187.01	75.47
		24717	1,283	0.00	GT2	279.09	550.73	376.59
		25735	222	5.84E-52	GT2	211.60	375.59	260.66
		26247	262	2.93E-56	GH18	153.47	228.37	111.86
		28387	203	2.13E-33	GH18	71.69	119.10	73.19
		31105	817	1.17E-141	GH10	47.92	166.37	67.24
		32239	223	3.49E-37	GT48	36.66	74.47	42.48
		5331	2,885	0.00	GH2/GH27	34.36	80.30	42.64
		6707	1,307	0.00	GH10	80.29	185.68	77.10
<b>LAC</b>	<b>3</b>	7215	1,907	0.00	GH43	78.81	91.29	130.49
		10859	1,914	0.00	GH47	193.40	198.30	294.08
		18143	324	2.36E-87	GH76	79.40	74.62	126.43
		18902	221	3.46E-37	CE5/CBM1	84.67	146.97	206.94
		19230	225	7.7E-45	GH18/CBM1	90.84	121.41	184.33
		19615	210	9.28E-38	CE5/CBM1	101.02	123.10	210.96
		20453	1,983	0.00	GH5	9.49	10.24	135.54
		22045	1,22	0.00	GH18	13.66	30.17	53.55
		22732	1,036	0.00	GH43/CBM1	73.44	129.52	179.90
		23357	1,136	0.00	CE5	32.57	49.05	88.74
		23867	1,884	0.00	GH55	187.33	207.94	318.03

24034	1,109	0.00	GH2	192.75	286.72	354.68
24118	1,48	0.00	GH3	122.62	153.22	248.20
24121	392	5.3E-90	GH72/CBM43	242.90	397.74	516.40
24529	582	2.79E-147	GH31	171.29	177.98	249.27
24859	367	3.48E-109	GH62	148.35	184.33	413.59
25106	622	0.00	CE5/CBM1	228.89	287.24	445.60
25478	230	6.87E-92	GH3	87.18	84.79	179.73
25634	386	1.23E-143	GH18	209.61	190.08	393.49
26359	317	7.11E-105	GH18/CBM1	115.17	144.83	221.46
26916	303	1.61E-158	GH31	113.16	169.49	238.66
27290	226	2.98E-32	GH11	86.07	60.98	133.60
27342	313	2.27E-116	GH62	35.46	47.15	115.85
27456	265	4.58E-54	GH71/CBM24	49.47	79.91	132.71
28080	593	2.34E-32	GH6	702.66	565.85	859.84
28132	364	0.00	GH31	315.80	392.58	588.33
28344	358	2.73E-110	GH11	83.67	144.72	220.94
28732	330	5.55E-31	GH92	93.23	96.63	159.65
29910	1,762	0.00	GH54/CBM42	48.08	58.71	115.27
29947	1,601	0.00	CE5/CBM1	96.47	169.32	259.20
30067	1,876	0.00	GH64	47.94	54.27	84.11
31154	1	0.00	GH13/CBM48	307.18	336.14	398.07
31183	505	0.00	GH62/CBM1	253.87	274.20	429.39





**Figure 6. K-means clustering of differentially expressed genes identified as CAZymes (summarized in Table 3).**

Different genes corresponded to different glycoside hydrolase families involved in carbohydrate metabolism in the different upregulated groups. According to the **Carbohydrate-Active Enzymes database [40]**, the glycoside hydrolases of family 1 include enzymes that possess  $\beta$ -glucosidase (EC 3.2.1.21),  $\beta$ -galactosidase (EC 3.2.1.23) and  $\beta$ -mannosidase (EC 3.2.1.25) activities; the glycoside hydrolases of family 18 possess chitinase (EC 3.2.1.14) activity; the glycoside hydrolases of family 55 exhibit exo-b-1,3-glucanase (EC 3.2.1.58)

and endo-b-1,3-glucanase (EC 3.2.1.39) activities; the glycoside hydrolases of family 3 exhibit  $\beta$ -glucosidase (EC 3.2.1.21) and xylan 1,4- $\beta$ -xylosidase (EC 3.2.1.37) activities; the glycoside hydrolases of family 5 possess chitosanase (EC 3.2.1.132), b-mannosidase (EC 3.2.1.25), endo-b-1,4-glucanase / cellulase (EC 3.2.1.4) and glucan b-1,3-glucosidase (EC 3.2.1.58) activities; the glycoside hydrolases of family 11 present endo-1,4- $\beta$ -xylanase (EC 3.2.1.8) activity; and the glycoside hydrolases of family 16 exhibit endo-1,3- $\beta$ -glucanase (EC 3.2.1.39) or endo-1,3(4)- $\beta$ -glucanase (EC 3.2.1.6) activity. **The LAC library contained 33 classified genes, whereas the CEL library contained 23 genes and the DSB library contained 22 genes. These gene classifications included glycosyltransferases (GTs), which catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules to form glycosidic bonds; carbohydrate esterases (CEs); and the corresponding carbohydrate-binding modules (CBMs). Glycosyltransferases can be classified as either retaining or inverting enzymes according to the stereochemistry of their substrates and reaction products. The glycosyltransferases of family 2 (GT2) exhibit cellulose synthase (EC 2.4.1.12) and chitin synthase (EC 2.4.1.16) activities and appear in all three libraries. The glycosyltransferases of family 4 (GT4) exhibit sucrose synthase (EC 2.4.1.13) and sucrose-phosphate synthase (EC 2.4.1.14) activities. Therefore, some of the genes that are responsible for biomass degradation reactions are highly expressed, whereas others, though not highly expressed, may also confer the ability to degrade organic compounds for energy in this fungus. Thus, the fungus can adapt its cellulolytic system to the composition of its medium by increasing or decreasing the expression of certain genes, as observed in the present study.**

## Discussion

The ability of filamentous fungi to efficiently degrade plant polymers is an important aspect of microbial ecology and may afford many potential industrial applications. The fungal strain *T. harzianum* demonstrates promising results for on-site cellulase production and is a potential candidate for the production of hydrolytic enzymes [4],[6].

To evaluate the cellulase activity of this fungus on pretreated sugarcane bagasse, we measured FPase, which reflects the overall activity of multicomponent enzyme complexes for cellulose hydrolysis [41]. An increase in cellulose activity is observed over the course of cultivation until 96 h, which represents the maximum cellulolytic activity (Figure 2). The DSB sample, which was used as the inducer, initiated fermentation at a level 4-fold greater than cellulolytic activity, which is most likely due to previous adaptation of the fungus to the substrate during the production of mycelia (during the preculture). In this case, the set of genes that were activated during the induction of mycelial growth was identical to the set used in fermentation, which allowed for a higher rate of fermentation in the first 24 h. In the first 48 to 96 h, the cellulolytic activity profile of the sample induced with DSB maintained a growth profile and FPase that was statistically similar to that of samples induced with cellulose. This result indicates that the set of genes that were active after the adaptation phase of fermentation may have been similar between the samples; however, the sample induced with DSB must have differentially expressed some genes in the first 24 h of growth because it reached a higher peak of cellulolytic activity ( $0.2 \pm 0.01$  FPU ml<sup>-1</sup>) compared with the samples induced with cellulose and lactose. The sample that used lactose as the inducer of mycelial growth maintained lower levels of activity

throughout the fermentation . Notably, in the first 24 h of fermentation, the CEL and LAC samples both achieved similar ( $0.05 \pm 0.004$  and  $0.02 \pm 0.002$  FPU ml<sup>-1</sup>, respectively) levels of FPase, suggesting that the set of genes that were activated during the preculture phase generated similar rates of cellulose-degrading enzymatic activity.

To elucidate how the complex sugarcane bagasse substrate influences the set of fungal gene transcripts that conferred enzymatic activity, we analyzed the transcription profiles of the samples. The results represent the first characterization of global gene expression in *T. harzianum* grown on a complex substrate (Figure 3). In the analysis of 32,494 contigs from the cDNA library, 6,975 sequences were classified as possessing catalytic activity (21.46% of total contigs), of which 2,555 possess hydrolase activity and act on chemical bonds such as ester, carbon-nitrogen and carbon-carbon bonds (Table 4). The high number of identified hydrolase sequences allowed us to determine the gene sequences that were related to specific degradation reactions. A similar annotation profile, which was generated using Gene Ontology (GO), was described by Steindorff *et al.*(2012) [5] for an EST sequencing library of 2,927 high-quality sequences. In both experiments, catalytic activity and binding represented the major classified molecular functions, with metabolic and cellular processes being the most prevalent classifications, and the cell and organelle category constituting the most common cellular localization.

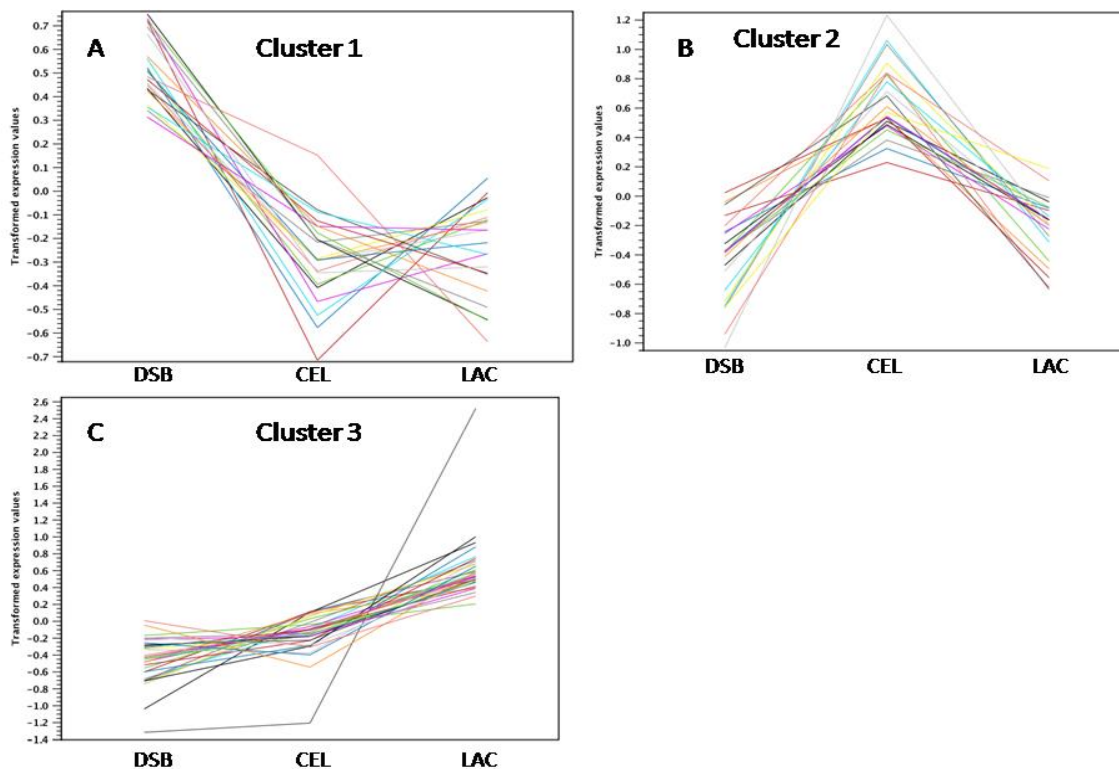
**Table 4. Contig sequences classified according to their putative hydrolytic activity**

<b>GO ID</b>	<b>TERM</b>	<b>Number of classified sequences</b>
GO:0016787	hydrolase activity	2,555
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	336
GO:0016788	hydrolase activity, acting on ester bonds	494
GO:0016798	hydrolase activity, acting on glycosyl bonds	373
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	778
GO:0016798	hydrolase activity, acting on glycosyl bonds	373
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	778
GO:0016817	hydrolase activity, acting on acid anhydrides	789
GO:0016810	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	140
GO:0016820	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	174
GO:0042578	phosphoric ester hydrolase activity	155
GO:0052689	carboxylic ester hydrolase activity	77
GO:0017171	serine hydrolase activity	113
GO:0008081	phosphoric diester hydrolase activity	54
GO:0016811	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds in linear amides	40
GO:0016790	thiolester hydrolase activity	45
GO:0008484	sulfuric ester hydrolase activity	17
GO:0016813	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds in linear amidines	17
GO:0016814	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds in cyclic amidines	25
GO:0047617	acyl-CoA hydrolase activity	9
GO:0033961	cis-stilbene-oxide hydrolase activity	8
GO:0016803	ether hydrolase activity	10
GO:0016289	CoA hydrolase activity	9
GO:0016801	hydrolase activity, acting on ether bonds	11
GO:0019238	cyclohydrolase activity	9
GO:0004416	hydroxyacylglutathione hydrolase activity	4
GO:0016799	hydrolase activity, hydrolyzing N-glycosyl compounds	5

GO:0019120	hydrolase activity, acting on acid halide bonds, in C-halide compounds	3
GO:0003933	GTP cyclohydrolase activity	5
GO:0004848	ureidoglycolate hydrolase activity	3
GO:0003935	GTP cyclohydrolase II activity	3
GO:0004477	methenyltetrahydrofolate cyclohydrolase activity	2
GO:0004045	aminoacyl-tRNA hydrolase activity	2
GO:0003934	GTP cyclohydrolase I activity	2
GO:0008474	palmitoyl-(protein) hydrolase activity	2
GO:0004463	leukotriene-A4 hydrolase activity	2
GO:0016823	hydrolase activity, acting on acid carbon- carbon bonds in ketonic substances	3
GO:0016824	hydrolase activity, acting on acid halide bonds	3
GO:0016822	hydrolase activity, acting on acid carbon- carbon bonds	3
GO:0004039	allophanate hydrolase activity	1
GO:0004635	phosphoribosyl-AMP cyclohydrolase activity	1
GO:0004649	poly(ADP-ribose) glycohydrolase activity	1
GO:0033971	hydroxyisourate hydrolase activity	1
GO:0018738	S-formylglutathione hydrolase activity	1
GO:0004301	epoxide hydrolase activity	1
GO:0033699	DNA 5'-adenosine monophosphate hydrolase activity	1
GO:0003937	IMP cyclohydrolase activity	1
GO:0016802	trialkylsulfonium hydrolase activity	1
GO:0016815	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds in nitriles	1

The current study identified genes that were upregulated by different substrates in the preculture phase (Table 3). The DSB library contained 792 classified contigs, 514 of which were homologous to the *T. harzianum* genome and 22 of which were related to the CAZyme library. Among the 377 classified contigs in the CEL library, 243 were related to the genome, and 23 were identified among the CAZyme. Among the 299 classified contigs in the LAC library, 272 genes were related to the genome, and 33 were identified in the CAZyme dataset

(Table 3). Therefore, according to the CAZyme classification, 79 genes were differentially expressed between two conditions and exhibited an expression level that was measurable in the other conditions. In this analysis, the gene expression values fell into three profiles after K-means clustering (Figure 6). Cluster 1 (Figure 7A) contained the genes (members of the glycoside hydrolase family) that were most highly expressed in the DSB library; cluster 2 (Figure 7B) contained the most highly expressed genes in the CEL library and Cluster 3 (Figure 7C) contained the most highly expressed genes in the LAC library; A difference observed between the set group of GH could be related to the influence of the different substrates. Several contigs were analyzed further in terms of their expression values and similarity.



**Figure 7. The identified genes (summarized in Table 3) were resolved into three clusters. The identified genes were further classified into one of three response profiles based on where they were most highly expressed: (A) in the DSB library, (B) in the CEL library or (C) in the LAC library.**

Among the differentially expressed genes, we identified genes related to extracellular degradative enzymes that play an important role in pathogenesis. These enzymes include the carbohydrate esterase family 5 protein, whose cutinase domain (contig 25106, classified based on CAZ similarity as [EHK47149.1](#), [IPR000675](#)) hydrolyzes cutin and facilitates fungal penetration through the cuticle. Inhibition of this enzyme can prevent fungal infection through intact cuticles. When cutin monomers are released from the cuticle due to small amounts of cutinase on fungal spore surfaces, these monomers can greatly increase the amount of cutinase secreted by the spore, although the mechanism underlying this process remains unknown. Another, more highly expressed, contig was classified as a member of the GH11 family (contig 20286). The overwhelming majority of the glycoside hydrolases of this family are xylanases. These enzymes carry out the endohydrolysis of (1->4)-beta-D-xylosidic linkages in xylans and random hydrolysis of (1->3)-beta-D-glycosidic linkages in (1->3)-beta-D-xylans. Contig 27456 exhibited similarity to family GH71, which includes  $\alpha$ -1,3-glucanase (EC 3.2.1.59). *O*-Glycosyl hydrolases (EC 3.2.1.) are a widespread group of enzymes that hydrolyze glycosidic bonds between two or more carbohydrates or between a carbohydrate and a noncarbohydrate moiety ([IPR005197](#)), and they are also related to CBM24 ( $\alpha$ -1,3-glucan (mutant)-binding function) [42].



In the DSB library, contig 20009 was found to be differentially expressed at a significant RPKM level (103.86) and was similar to a GH16 protein from *Trichoderma virens* ([EHK18881.1](#), [IPR000757](#)). The GH16 family contains a variety of enzymes with a range of known activities. Lichenase (EC 3.2.1.73), xyloglucan xyloglucosyltransferase (EC 2.4.1.207), agarase (EC 3.2.1.81), kappa-carrageenase (EC 3.2.1.83), endo- $\beta$ -1,3-glucanase (EC 3.2.1.39), endo- $\beta$ -1,3-1,4-glucanase (EC 3.2.1.6) and endo- $\beta$ -galactosidase (EC 3.2.1.103) are all members of this family.

In this study, we identified sequences related to different classes of enzymes that act on the cellulose backbone, such as GH5, which exhibits endo- $\beta$ -1,4-glucanase activity in *T. reesei* (EC 3.2.1.4) and is responsible for the hydrolysis of the (1 $\rightarrow$ 4)- $\beta$ -D-glucosidic linkages in cellulose. The GH3 family exhibits  $\beta$ -glucosidase activity in *T. reesei* (EC 3.2.1.21), where it hydrolyzes terminal, nonreducing  $\beta$ -D-glucosyl residues and releases  $\beta$ -D-glucose. This monomer can enter into the eukaryotic energy pathway of glycolysis. Glycolysis produces energy and requires an input of two ATP molecules. This input is used to generate four new ATP molecules, resulting in a net gain of two ATP molecules. Two NADH molecules are also produced; these molecules serve as electron carriers for other biochemical reactions in the cell. The enzymes that are necessary to catalyze the degradation of glucose molecules are expressed throughout the growth of the fungus on the complex substrate, possibly to produce energy through glycolysis and support cell survival and reproduction. The enzymes that act in biomass degradation were the focus of this work, and this analysis allowed us to identify a set of enzymes that are involved in carbohydrate metabolism based on expression profiles.

Regarding expression differences, the LAC library contained numerous genes receiving CAZyme classifications. Lactose, an inexpensive, soluble substrate, leads to reasonably good induction for cellulase production [43],[44]. The fungus does not directly take up lactose but instead hydrolyzes the compound to galactose and glucose. Cellulase synthesis cannot be induced by galactose, and the addition of galactose to the medium decreases FPase levels in the supernatant [45], as reported in this work. Karaffa et al. [43] reported that lactose induces significantly higher cellulase levels compared to galactose, but galactose induces cellulase gene expression at low growth rates in *T. reesei*. In this study, the highest degradation rate would have occurred in the fungi precultured on the lactose medium, followed by cellulose and DSB, due to the complexity of the substrate and the stability of the organic chains involved. The presence of lactose in the early stages of the experiment would have induced genes that are sensitive to lactose and galactose, which may explain the low level of FPase observed in the sample that used lactose as the inducer of hydrolytic systems (even in the fermentation step). As shown in the results, sugarcane bagasse was able to activate the expression of a different set of genes that were differentially expressed compared with the control, and this difference was associated with an increase in cellulase enzymatic activity during fermentation.

This strain of *T. harzianum* demonstrates a complex and efficient genetic mechanism for biomass degradation. The use of RNA-Seq technology was shown to be an efficient strategy for the discovery and selection of potential target genes. The results reported here are valuable for further studies on the expression, purification and characterization of recombinant enzymes for efficient cellulose degradation.

## References

1. Hermosa R, Viterbo A, Chet I, Monte E (2012) Plant-beneficial effects of *Trichoderma* and of its genes. *Microbiology* (Reading, England) 158: 17–25.
2. Vinale F, Sivasithamparam K, Ghisalberti EL, Marra R, Woo SL, et al. (2008) *Trichoderma*–plant–pathogen interactions. *Soil Biology and Biochemistry* 40: 1–10.
3. Maeda RN, Serpa VI, Rocha VAL, Mesquita RAA, Anna LMMS, et al. (2011) Enzymatic hydrolysis of pretreated sugar cane bagasse using *Penicillium funiculosum* and *Trichoderma harzianum* cellulases. *Process Biochemistry* 46: 1196–1201.
4. Delabona PDS, Farinas CS, da Silva MR, Azzoni SF, Pradella JGDC (2012) Use of a new *Trichoderma harzianum* strain isolated from the Amazon rainforest with pretreated sugar cane bagasse for on-site cellulase production. *Bioresource technology* 107: 517–521.
5. Steindorff AS, Silva RDN, Coelho ASG, Nagata T, Noronha EF, et al. (2012) *Trichoderma harzianum* expressed sequence tags for identification of genes with putative roles in mycoparasitism against *Fusarium solani*. *Biological Control* 61: 134–140.
6. Colussi F (2011) Purification, and Biochemical and Biophysical Characterization of Cellobiohydrolase I from *Trichoderma harzianum* IOC 3844. *Journal of Microbiology and Biotechnology* 21: 808–817.
7. de Castro AM, Pedro KCNR, da Cruz JC, Ferreira MC, Leite SGF, et al. (2010) *Trichoderma harzianum* IOC-4038: A promising strain for the production of a cellulolytic complex with significant  $\beta$ -glucosidase activity from sugarcane bagasse cellulignin. *Applied biochemistry and biotechnology* 162: 2111–2122.
8. Soccol CR, Vandenberghe LPDS, Medeiros ABP, Karp SG, Buckeridge M, et al. (2010) Bioethanol from lignocelluloses: Status and perspectives in Brazil. *Bioresource technology* 101: 4820–4825.
9. Sørensen A, Teller PJ, Lübeck PS, Ahring BK (2011) Onsite enzyme production during bioethanol production from biomass: screening for suitable fungal strains. *Applied biochemistry and biotechnology* 164: 1058–1070.

10. Van Dyk JS, Pletschke BI (2012) A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-Factors affecting enzymes, conversion and synergy. *Biotechnology advances* 30: 1458–1480.
11. Van Vliet AHM (2010) Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS microbiology letters* 302: 1–7.
12. Banerjee G, Car S, Scott-Craig JS, Borrusch MS, Walton JD (2010) Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations. *Biotechnology for biofuels* 3: 22.
13. Gilbert HJ (2010) The biochemistry and structural biology of plant cell wall deconstruction. *Plant physiology* 153: 444–455.
14. Banerjee G, Scott-Craig JS, Walton JD (2010) Improving Enzymes for Biomass Conversion: A Basic Research Perspective. *BioEnergy Research* 3: 82–92.
15. Saloheimo M, Paloheimo M, Hakola S, Pere J, Swanson B, et al. (2002) Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. *European Journal of Biochemistry* 269: 4202–4211.
16. Erikssoon T, Karlsson J, Tjerneld F (2002) A Model Explaining Declining Rate in Hydrolysis of Lignocellulose Substrates with Cellobiohydrolase I ( Cel7A ) and Endoglucanase I ( Cel7B ) of *Trichoderma reesei*. *Applied biochemistry and biotechnology* 101: 41–60.
17. Teeri T (1997) Crystalline cellulose degradation : new insight into the function of cellobiohydrolases G3. *Trends in Biotechnology* 15: 160–167.
18. Meyer AS, Rosgaard L, Sørensen HR (2009) The minimal enzyme cocktail concept for biomass processing. *Journal of Cereal Science* 50: 337–344.
19. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, et al. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (*syn. Hypocrea jecorina*). *Nature biotechnology* 26: 553–560.
20. Liu H-Q, Feng Y, Zhao D-Q, Jiang J-X (2011) Evaluation of cellulases produced from four fungi cultured on furfural residues and microcrystalline cellulose. *Biodegradation* 23: 465–472.

21. Mandels M, Weber J (1969) The production of cellulases. In: Hajny GJ, Reese ET editors. *Advances in Chemistry*, AMERICAN CHEMICAL SOCIETY: Washington, DC. DOI: 10.1021/ba-1969-0095.ch023.
22. Rocha GJM, Gonçalves AR, Oliveira BR, Olivares EG, Rossell CEV (2012) Steam explosion pretreatment reproduction and alkaline delignification reactions performed on a pilot scale with sugarcane bagasse for bioethanol production. *Industrial Crops and Products* 35: 274–279.
23. Ghose TK (1987) Measurement of cellulase activities. *Pure Appl Chem* 59: 257–268.
24. Jones JD, Dunsmuir P, Bedbrook J (1985) High level expression of introduced chimaeric genes in regenerated transformed plants. *The EMBO journal* 4: 2411–2418.
25. Pearson WR (2013) An introduction to sequence similarity (“homology”) searching. *Current protocols in bioinformatics* 42: 3.1.1–3.1.8.
26. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene Ontology : tool for the unification of biology. *Nature America Inc* 25: 25–29.
27. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
28. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 28: 27–30.
29. Henrissat B, Vegetales M, Grenoble F- (1991) A classification of glycosyl hydrolases based sequence similarities amino acid. *Biochem J* 280: 309–316.
30. Henrissat B, Bairoch A (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* 293: 781–788.
31. Henrissat B, Bairoch A (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochem J* 316: 695–696.
32. Henrissat B, Davies G (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr Op Struct Biol* 7: 637–644.
33. Davies G, Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* 3: 853–859.

34. Campbell J, Davies G, Bulone V, Henrissat B (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *The Biochemical journal* 326: 929–939.
35. Coutinho PM, Deleury E, Davies GJ, Henrissat B (2003) An Evolving Hierarchical Family Classification for Glycosyltransferases. *Journal of Molecular Biology* 328: 307–317.
36. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *The Biochemical journal* 382: 769–781.
37. Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, et al. (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. *The Biochemical journal* 432: 437–444.
38. Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, et al. (2012) The genome portal of the Department of Energy Joint Genome Institute. *Nucleic acids research* 40: D26–32.
39. Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, et al. (2007) Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science* 315 : 804–807.
40. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, et al (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* 37:D233-238
41. Urb K, Szak G, Tengerdy RP (2000) Standardization of the filter paper activity assay for solid substrate fermentation. *Biotechnology Letters* 22: 65–69.
42. Fuglsang CC (2000) Biochemical Analysis of Recombinant Fungal Mutanases. A new family of alpha 1,3-glucanases with novel carbohydrate-binding domains. *Journal of Biological Chemistry* 275: 2009–2018.
43. Karaffa L, Fekete E, Gamauf C, Szentirmai A, Kubicek CP, et al. (2006) D-Galactose induces cellulase gene expression in *Hypocrea jecorina* at low growth rates. *Microbiology* 152: 1507–1514.

44. Lo C-M, Zhang Q, Callow NV, Ju L-K (2010) Roles of extracellular lactose hydrolysis in cellulase production by *Trichoderma reesei* Rut C30 using lactose as inducing substrate. *Process Biochemistry* 45: 1494–1503.
45. Fang X, Yano S, Inoue H, Sawayama S (2008) Lactose enhances cellulase production by the filamentous fungus *Acremonium cellulolyticus*. *Journal of bioscience and bioengineering* 106: 115–120.
46. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research* 40: D109–14.
47. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  method. *Methods* 25: 402–408

## Supporting information

**Supplementary Table S1. Primers used for RT-qPCR detection of glycosyl hydrolase genes.** The squalene-epoxidase gene was used as endogenous control and the sequences analyzed encoded genes of glycoside hydrolases (GH), carbohydrate esterases (CEs) and carbohydrate-binding modules (CBM).

Name	Sequence contig	Primer	T <sub>m</sub>
squalene-epoxidase	7950	5'-GTTGTATGAGACCAGATCGGC-3' 5'-TCACAATGTGTCAATCGAGCA-3'	56°C
CE5,CE5/CBM1	19615	5'-CAGTGGAGTTGGAGAAGCGTC-3' 5'-GTTCCGGGATGAGAGTTGACA-3'	58°C
GH16	20009	5'-GAAATCGAAAAGCCCTGGTT-3' 5'-AGACAACACTCCGCATGCC-3'	57°C
GH10	17441	5'-AACACTCCTTTCCCTGGCCT-3' 5'-GTGCCCCAGTACTTCTTTCCAT-3'	59°C
GH5/CBM1	15510	5'-AGCTCACGAAGAGTTCCCTCC-3' 5'-GGTTCCTGATATCGTCACAAAAC-3'	59°C



## **6. CONSIDERAÇÕES FINAIS**



O fungo *Trichoderma harzianum* já foi e continua sendo alvo de estudos, principalmente no que diz respeito à sua capacidade de biocontrole. Só mais recentemente começou a ser investigada a capacidade degradativa deste microrganismo, porém não existia na literatura estudos genômicos sobre esta linhagem. Assim, o propósito deste trabalho foi determinar os transcritos produzidos pelo fungo em condições de degradação de biomassa.

A primeira abordagem do projeto para a determinação dos transcritos foi por meio da utilização das bibliotecas de cDNA, descrita no capítulo 1. Essa abordagem foi utilizada e as bibliotecas foram montadas, porém durante o desenvolvimento do projeto houve a possibilidade de realizar o sequenciamento *high-throughput*. Esse tipo de sequenciamento é mais rápido e barato, o que nos levou a interromper a utilização das bibliotecas de cDNA. Quando optamos pelo sequenciamento *high-throughput* já havíamos obtido bons resultados referentes a classificação de sequências com atividade hidrolítica, embora poucas comparativamente ao número de genes com atividade hidrolítica presentes no genoma.

Essa mudança de abordagem foi bastante enriquecedora pois permitiu a obtenção de um número muito maior de sequências em menos tempo, o que nos levou a determinar o perfil de transcrição do fungo nas condições estudadas.

A obtenção de 14,7 Gpb sequenciados nos permitiu, após a montagem e anotação uma visão precisa do sistema de expressão de genes que o fungo utiliza em condições de degradação de biomassa. Esses dados, que incluem informações sobre genes do metabolismo basal, genes que podem estar ligados a característica de biocontrole, genes que codificam proteínas acessórias, etc, foram depositados em banco de dados públicos e estão disponíveis para análises aprofundadas sobre o transcriptoma deste microrganismo. Após a análise dos

32.494 *contigs* formados, 6.975 sequências foram classificadas possuindo atividade catalítica (21.46% do total de *contigs*), e 2.555 destes *contigs* apresentaram atividade de hidrolase, agindo sobre ligações químicas do tipo éster, carbono-nitrogênio, carbono-carbono, entre outras. Assim, dentre uma grande quantidade de informação, o foco do projeto manteve-se nos mecanismos genômicos de degradação de biomassa, em termos de número de genes e níveis de expressão. Foram identificadas 487 diferentes CAZymes no transcriptoma, 23 das quais codificam proteínas das famílias de glicosil hidrolases envolvidas especificamente com as reações de despolimerização da celulose: três genes classificados como GH5, um como GH7, dois como GH12, dois como GH45, três como GH1, dez como GH3, um como GH6 e um como GH61. Também foram identificados 10 diferentes sequências gênicas que codificam módulos de ligação à carboidratos. Em relação aos genes envolvidos na despolimerização de hemicelulose foram identificados 22 genes, três genes como GH10, três genes como GH11, dois genes como GH26, três genes como GH43, três genes como GH54, um como GH62, dois como GH67, um como GH74 e quatro genes como GH95, além de seis genes correspondentes a módulos de ligação.

Os níveis de expressão dos genes classificados foram influenciados pelo indutor do crescimento micelial, assim, pode-se observar os conjuntos de genes que foram diferencialmente expresso devido à determinadas condições de cultivo. A biblioteca LAC apresentou 33 genes classificados, enquanto a biblioteca CEL apresentou 23 e a biblioteca DSB obteve 22 genes. Essa classificação inclui as glicosiltransferases (GTs), as carboidrato-esterases (CEs), e os módulos de ligação à carboidratos(CBMs). Esses conjuntos de genes superexpressos podem ter relação com as atividades enzimáticas determinadas

no extrato, porém não é possível inferir a influência específica de cada gene, de modo que podemos concluir que o conjunto de genes superexpressos nas bibliotecas DSB e CEL permite atingir um mesmo nível de atividade FPA e beta-glucosidase, enquanto que para a atividade de xilanase, o conjunto de genes da biblioteca de BED e Lactose permitiu atingir o mesmo nível de atividade.

Uma característica importante deste trabalho diz respeito à precisão das sequências genômicas obtidas. Assim, tornou-se possível determinar, através do desenho de *primers*, os genes e regiões genômicas correspondentes às enzimas estudadas. Outro ponto de real utilidade foi a determinação de um protocolo para sequenciamento, montagem, anotação e seleção de genes de interesse (figura 21), que pode ser útil para outras linhagens e espécies, e variar de acordo com o foco do estudo.

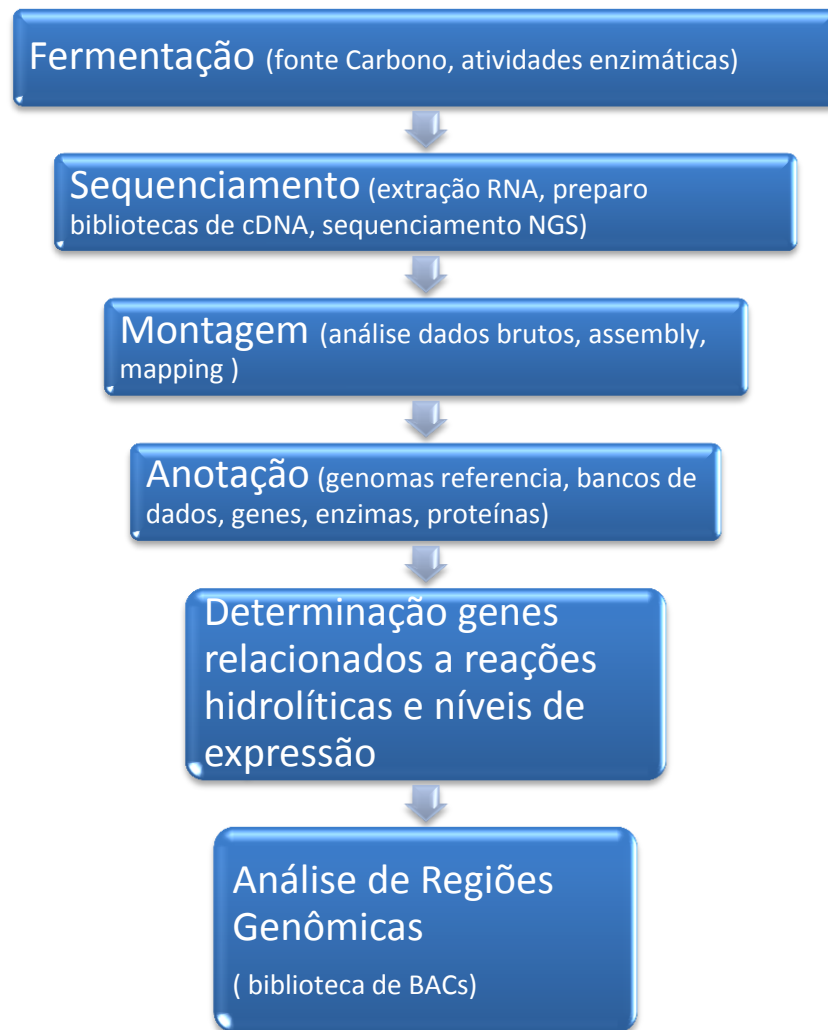


Figura 21: Fluxograma descritivo do protocolo de análise do transcriptoma.

Assim, este estudo inédito do transcriptoma de *T. harzianum* forneceu importantes informações sobre o conjunto de genes responsáveis pelas reações de despolimerização de biomassa.

## **7. RESUMO DOS RESULTADOS**





- i. A cepa de *T. harzianum* IOC-3844 demonstra um mecanismo complexo e eficiente para a degradação de biomassa e, o uso de tecnologia de sequenciamento *high-throughput* para a determinação do transcriptoma deste fungo demonstrou ser uma estratégia eficiente, rápida e barata para a determinação e seleção de genes potenciais para a biodegradação.
- ii. Foi observado o comportamento positivo do fungo frente à utilização de bagaço de cana como substrato, tanto na fase de indução do crescimento do micélio quanto na fermentação. Esse substrato atua como superfície para fixação do fungo e fonte de carbono fermentescível.
- iii. Foi possível determinar as enzimas CAZy presentes no transcriptoma relacionadas a degradação de celulose e hemicelulose, portanto determinou-se com exatidão as sequências gênicas que podem catalisar as reações de despolimerização de biomassa.
- iv. A utilização de diferentes substratos como fonte de carbono permitiu determinar conjuntos de genes que são superexpresso pelas diferentes fontes de carbono.
- v. Através da utilização de tecnologia de sequenciamento *high-throughput* para a determinação do transcriptoma foi elaborado um protocolo de análise, que inclui o sequenciamento, montagem, anotação e classificação das sequências gênicas que permanece válido para outras linhagens.



## **8. CONCLUSÕES**



Os experimentos de sequenciamento do transcriptoma do fungo permitiram delinear todos os genes expressos com seus respectivos níveis de expressão e determinar sequências gênicas ligadas às reações metabólicas e de degradação de biomassa que o fungo utiliza para sobreviver e obter energia. Assim, este foi o primeiro trabalho que traçou o perfil do transcriptoma do *T. harzianum* em condições que estimulem a capacidade do fungo de biodegradação.

Além disso, os experimentos permitiram determinar precisamente o conjunto de enzimas que confere o potencial celulolítico do *T. harzianum* e comprovar sua capacidade de produção de diversas enzimas ligadas à degradação, proteínas acessórias e módulos de ligação com o substrato responsáveis pelas reações de degradação.



# **9. PERSPECTIVAS**





A capacidade dos fungos filamentosos de degradar eficientemente biomassa é um importante tópico cujos estudos podem desvendar os mecanismos de regulação enzimática, o que determina inúmeras possibilidades para a aplicação industrial. Os resultados aqui relatados são valiosos para futuras pesquisas sobre os mecanismos de regulação gênica das enzimas para a degradação eficiente de lignocelulósicos. Estes estudos podem permitir a produção heteróloga de enzimas hidrolíticas para a utilização em coquetéis enzimáticos além da manipulação de regiões do genoma, visando um aumento da produção de enzimas.



## **10.BIBLIOGRAFIA**

1. Najafpour GD. Chapter 1 - Industrial Microbiology. In: Najafpour GDBT-BE and B, editor. Amsterdam: Elsevier; 2007. p. 1–13. Disponível em: <http://www.sciencedirect.com/science/article/pii/B978044452845250001X>
2. Christian, Larroche. Ashok, Pandey.; Carlos Ricardo S. Current Development in Solid State Fermentation. New Delhi: Asiatech; 2007.
3. Schwartz M. The life and works of Louis Pasteur. 2001;(February):597–601.
4. Borém A. Biotecnologia e meio ambiente. 8 v34 . Viçosa, MG; 2005;25–118. Disponível em: <http://www.biotecnologia.com.br/revista/bio34/bio34.pdf>
5. Belem MAF. Application of biotechnology in the product development of nutraceuticals in Canada. Trends Food Sci Technol . 1999 Mar;10(3):101–6. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0924224499000291>
6. STANBURY PF, WHITAKER A, HALL SJ. CHAPTER 3 - The Isolation, Preservation and Improvement of Industrially Important Micro-organisms. In: HALL PFSWJBT-P of FT (Second E, editor. Amsterdam: Pergamon; 1995. p. 35–91. Disponível em: <http://www.sciencedirect.com/science/article/pii/B9780080361314500080>
7. Tortora GJ, Funke BR, Case CL. Microbiologia. Artmed; 2012.
8. CGEE B. Bioetanol de Cana-de-acúcar: Energia para o desenvolvimento sustentável . 2008. Disponível em: <http://www.bioetanoldecana.org/>
9. Pessoa-Jr A, Roberto IC, Menossi M, Dos Santos RR, Ortega Filho S, Penna TCV. Perspectives on bioenergy and biotechnology in Brazil. Appl Biochem Biotechnol. Springer; 2005;121(1-3):59–70.
10. Ramos LP, Wilhelm HM. Current status of biodiesel development in Brazil. Current Status of Biodiesel Development in Brazil . Springer; 2005. p. 807–19. Disponível em: [http://link.springer.com/chapter/10.1007/978-1-59259-991-2\\_68](http://link.springer.com/chapter/10.1007/978-1-59259-991-2_68)
11. Soccol CR, Vandenberghe LPDS, Medeiros ABP, Karp SG, Buckeridge M, Ramos LP, et al. Bioethanol from lignocelluloses: Status and perspectives in Brazil. Bioresour Technol . Elsevier Ltd; 2010 Jul [cited 2012 Mar 17];101(13):4820–5. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/20022746>
12. Van Dyk JS, Pletschke BI. A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-Factors affecting enzymes, conversion and synergy. Biotechnol Adv . Elsevier Inc.; 2012 Mar 13 [cited 2012 Jul 18];30(6):1458–80. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/22445788>
13. Gilbert HJ. The biochemistry and structural biology of plant cell wall deconstruction. Plant Physiol . 2010 Jun [cited 2012 Oct 28];153(2):444–55. Disponível em:

<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2879781&tool=pmcentrez&rendertype=abstract>

14. Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P, et al. Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. *PlosOne*. 2008;3(8).
15. Banerjee G, Scott-Craig JS, Walton JD. Improving Enzymes for Biomass Conversion: A Basic Research Perspective. *BioEnergy Res* . 2010 Jan 9 [cited 2013 May 22];3(1):82–92. Disponible em: <http://link.springer.com/10.1007/s12155-009-9067-5>
16. Harris PV, Welner D, McFarland KC, Re E, Poulsen J-C, Brown K. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase Family 61: structure and function of a large, enigmatic family. *Biochemistry*. 2010;49:3305–16.
17. Merino ST, Cherry J, Ave D. Progress and Challenges in Enzyme Development for Biomass Utilization. *Adv Biochem Engin/Biotechnol*. 2007;108(June):95–120.
18. Pearson WR. An introduction to sequence similarity (“homology”) searching. *Curr Protoc Bioinforma* . 2013 Jun [cited 2013 Jun 21];42:3.1.1–3.1.8. Disponible em: <http://www.ncbi.nlm.nih.gov/pubmed/23749753>
19. Chandra M, Kalra A, Sharma PK, Sangwan RS. Cellulase production by six *Trichoderma* spp. fermented on medicinal plant processings. *J Ind Microbiol Biotechnol* . 2009 Apr [cited 2012 May 15];36(4):605–9. Disponible em: <http://www.ncbi.nlm.nih.gov/pubmed/19277745>
20. Ali, F.U.E.-D. HS. Production and Partial Purification of Cellulase Complex by *Aspergillus niger* and *A. nidulans* Grown on Water Hyacinth Blend. *J Appl Sci Res*. 2008;4(7):875–91.
21. Wonganu, B. et al. Cloning, expression and characterization of a thermotolerant endoglucanase from *Syncephalastrum racemosum* (BCC18080) in *Pichia pastoris*. *Protein Expr Purif*. 2008;58(1):78–86.
22. Schloss, P.D., Handelsman J. Biotechnological prospects from metagenomics. *Curr Opin Biotechnol*. 2003;14:303–10.
23. Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez D a, Druzhinina IS, Thon M, et al. Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. *Genome Biol* . 2011 Jan [cited 2012 Mar 15];12(4):R40. Disponible em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3218866&tool=pmcentrez&rendertype=abstract>
24. Yedidia I, Benhamou N. Induction of Defense Responses in Cucumber Plants (*Cucumis sativus* L.) by the Biocontrol Agent *Trichoderma harzianum*. *Appl Environ Microbiol*. 1999;65(3):1061–70.

25. Vizcaíno JA, Redondo J, Suárez MB, Cardoza RE, Hermosa R, González FJ, et al. Generation, annotation, and analysis of ESTs from four different *Trichoderma* strains grown under conditions related to biocontrol. *Appl Microbiol Biotechnol* . 2007 Jul [cited 2012 Jul 31];75(4):853–62. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/17333173>
26. De Castro AM, Pedro KCNR, da Cruz JC, Ferreira MC, Leite SGF, Pereira N. *Trichoderma harzianum* IOC-4038: A promising strain for the production of a cellulolytic complex with significant  $\beta$ -glucosidase activity from sugarcane bagasse cellulignin. *Appl Biochem Biotechnol* . 2010 Nov [cited 2012 Oct 16];162(7):2111–22. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/20455032>
27. Harman GE (Cornell U. *Trichoderma* spp., including *T. harzianum*, *T. viride*, *T. koningii*, *T. hamatum* and other spp. Deuteromycetes, Moniliales (asexual classification system) . [cited 2014 Feb 5]. Disponível em: <http://www.biocontrol.entomology.cornell.edu/pathogens/trichoderma.php>
28. Samuels GJ, Dodd SL, Gams W, Castlebury L a., Petrini O. *Trichoderma* Species Associated with the Green Mold Epidemic of Commercially Grown *Agaricus bisporus*. *Mycologia* . 2002 Jan;94(1):146. Disponível em: <http://www.jstor.org/stable/3761854?origin=crossref>
29. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, et al. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotechnol* . 2008 May [cited 2012 Mar 12];26(5):553–60. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/18454138>
30. Paulo S, Ruegger MJS, Tauk-tornisielo SM. Atividade da celulase de fungos isolados do solo da Estação Ecológica de. *Revista Brasileira Bot.* ;2004;2:205–11.
31. Silveira FQ de P, Melo IS, Ferreira Filho EX. Carbohydrate-hydrolysing enzyme activity production by solid-state cultures of *Trichoderma harzianum* strains. *Rev microbiol.* 1997;28(3):152–6.
32. Medeiros R, Silva Jr F, Salles B, Estelles R, Filho E. The performance of fungal xylan-degrading enzyme preparations in elemental chlorine-free bleaching for Eucalyptus pulp. *J Ind Microbiol Biotechnol* . Springer; 2002 Apr 1;28(4):204–6. Disponível em: <http://www.springerlink.com/Index/10.1038/sj.jim.7000227>
33. Vizoná Liberato M, Cardoso Generoso W, Malagó Jr W, Henrique-Silva F, Polikarpov I. Crystallization and preliminary X-ray diffraction analysis of endoglucanase III from *Trichoderma harzianum*. *Acta Crystallogr Sect F* . 2012 Mar;68(3):306–9. Disponível em: <http://dx.doi.org/10.1107/S1744309112000838>
34. Maeda RN, Serpa VI, Rocha VAL, Mesquita RAA, Anna LMMS, de Castro AM, et al. Enzymatic hydrolysis of pretreated sugar cane bagasse using *Penicillium funiculosum* and *Trichoderma harzianum* cellulases. *Process Biochem* . Elsevier Ltd; 2011 May [cited 2012 Jun 18];46(5):1196–201. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S1359511311000286>

35. Colussi F. Purification, and Biochemical and Biophysical Characterization of Cellobiohydrolase I from *Trichoderma harzianum* IOC 3844. *J Microbiol Biotechnol* . 2011 Aug [cited 2012 Aug 8];21(8):808–17. Disponível em: <http://www.jmb.or.kr/journal/viewJournal.html?year=2011&vol=21&num=8&page=808>
36. Franco PF, Ferreira HM. Production and characterization of hemicellulase activities from *Trichoderma harzianum* strain T4. *Biotechnol Appl Biochem*. Wiley Online Library; 2004;40(3):255–9.
37. Yun S-I, Jeong C-S, Chhung D-K, Choi H-S. Purification and Some Properties of a B-Glucosidase from *Trichoderma harzianum* Type C-4. *Biosci Biotechnol Biochem*. 2001;65(9):2028–32.
38. Generoso WC, Malagó-Jr W, Pereira N, Henrique-Silva F. Recombinant expression and characterization of an endoglucanase III (cel12a) from *Trichoderma harzianum* (Hypocreaceae) in the yeast *Pichia pastoris*. *Genet Mol Res* . 2012 Jan;11(2):1544–57. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/22653604>
39. 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. *Bioresour Technol* . Elsevier Ltd; 2012 Mar [cited 2012 May 2];107:517–21. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/22221990>
40. Rezende MI, Barbosa ADM, Flora A, Vasconcelos D, Endo AS. XYLANASE PRODUCTION BY TRICHODERMA HARZIANUM RIFAI BY SOLID STATE FERMENTATION ON SUGARCANE BAGASSE. *Brazilian Journal of Microbiology.*; 2002;67–72.
41. Paola Gómez Mendoza D. Variações do secretoma de *Trichoderma harzianum* em resposta a diferentes fontes de carbono. *Dissertação de Mestrado.*; Universidade de Brasília; 2009. p. 1–66.
42. Momeni MH, Payne CM, Hansson H, Mikkelsen NE, Svedberg J, Engström Å, et al. Structural, biochemical, and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus *Heterobasidion irregulare*. *J Biol Chem* . 2013 Feb 22 [cited 2013 Apr 9];288(8):5861–72. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/23303184>
43. Cunha ES, Hatem CL, Barrick D. Insertion of endocellulase catalytic domains into thermostable consensus ankyrin scaffolds: effects on stability and cellulolytic activity. *Appl Environ Microbiol* . 2013 Nov [cited 2014 May 30];79(21):6684–96. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3811507&tool=pmcentrez&rendertype=abstract>
44. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* . 2009 Jan;10(1):57–63. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2949280&tool=pmcentrez&rendertype=abstract>

45. Ghose TK. Measurement of cellulase activities. *Pure Appl Chem*. 1987;59(2):257– 268.
46. Bailey, M.J., Poutanen K. Production of xylanolytic enzymes by strains of *Aspergillus*. *Appl Microbiol Biotechnol*. 1989;30:5–10.
47. Zhang YH, Hong J, Ye X. Cellulase Assays . *Biofuels* . 2009. p. 213–31. Disponível em: [http://www.springerprotocols.com/Abstract/doi/10.1007/978-1-60761-214-8\\_14](http://www.springerprotocols.com/Abstract/doi/10.1007/978-1-60761-214-8_14)
48. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* . 1976 May 7;72(1–2):248–54. Disponível em: <http://www.sciencedirect.com/science/article/pii/0003269776905273>
49. Juhász T, Szengyel Z, Réczey K, Siika-Aho M, Viikari L. Characterization of cellulases and hemicellulases produced by *Trichoderma reesei* on various carbon sources. *Process Biochem* . 2005 Nov [cited 2012 Mar 1];40(11):3519–25. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S1359511305002035>
50. Clontech. Directional cDNA Library Construction Kit User Manual. 2011;1(634933):1–41.
51. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Al E. Gene Ontology : tool for the unification of biology. *Nat Am Inc* . 2000;25(may):25–9. Disponível em: <http://genetics.nature.com>
52. Steindorff AS, Silva RDN, Coelho ASG, Nagata T, Noronha EF, Ulhoa CJ. *Trichoderma harzianum* expressed sequence tags for identification of genes with putative roles in mycoparasitism against *Fusarium solani*. *Biol Control* . Elsevier Inc.; 2012 May [cited 2012 Apr 16];61(2):134–40. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S1049964412000254>
53. Banerjee G, Car S, Scott-Craig JS, Borrusch MS, Walton JD. Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations. *Biotechnol Biofuels* . BioMed Central Ltd; 2010 Jan [cited 2012 Nov 11];3(1):22. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2964541&tool=pmcentrez&rendertype=abstract>





Cidade Universitária "Zeferino Vaz",  
22 de outubro de 2014.

CIBio – Declaração 11/2014.

**Identificação:**

**Doutorado:** Maria Augusta Crivelente Horta

**Projeto:** "ANÁLISE DO TRANSCRIPTOMA DE *TRICHODERMA HARZIANUM*  
PARA A BIOPROSPECÇÃO DE ENZIMAS HIDROLÍTICAS".

**Parecer:**

Projeto aprovado pela CIBio/CBMEG sob número 01/2010.

Coordenador: Profa. Dra. Anete Pereira de Sousa

**Profa. Dra. Edi Lúcia Sartorato**  
Presidente da CIBio/CBMEG - UNICAMP

## DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha dissertação de Mestrado/tese de Doutorado intitulada **ANÁLISE DO TRANSCRIPTOMA DE *TRICHODERMA HARZIANUM* PARA A BIOPROSPECÇÃO DE ENZIMAS HIDROLÍTICAS**

( ) não se enquadra no § 4º do Artigo 1º da Informação CCPG 002/13, referente a bioética e biossegurança.

Tem autorização da(s) seguinte(s) Comissão(ões):

( X ) CIBio – Comissão Interna de Biossegurança , projeto No. \_\_\_\_\_01/2010\_\_\_\_\_, Instituição:  
\_\_\_Instituto de Biologia -UNICAMP\_\_\_\_\_.

( ) CEUA – Comissão de Ética no Uso de Animais , projeto No. \_\_\_\_\_, Instituição:  
\_\_\_\_\_.

( ) CEP - Comissão de Ética em Pesquisa, protocolo No. \_\_\_\_\_, Instituição:  
\_\_\_\_\_.

*\* Caso a Comissão seja externa ao IB/UNICAMP, anexar o comprovante de autorização dada ao trabalho. Se a autorização não tiver sido dada diretamente ao trabalho de tese ou dissertação, deverá ser anexado também um comprovante do vínculo do trabalho do aluno com o que constar no documento de autorização apresentado.*



Aluno: (nome completo)



Orientador: (nome completo)

Profa. Dra. Anete P. de Souza  
Centro de Biologia Molecular e  
Engenharia Genética - UNICAMP

Para uso da Comissão ou Comitê pertinente:

Deferido ( ) Indeferido



Carimbo e assinatura

Profa. Dra. HELJAN COELHO FRANCO DE OLIVEIRA  
Presidente  
CIBioIB-UNICAMP