



UNIVERSIDADE DE BRASÍLIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA
MOLECULAR

GENÔMICA ESTRUTURAL E FUNCIONAL
DE FUNGOS DO GÊNERO *Trichoderma*

Andrei Stecca Steindorff

Orientador:

Dra. Eliane Ferreira Noronha

Co-orientador: **Dr. Georgios Joannis Pappas Júnior**

Co-orientador estrangeiro: **Igor Grigoriev**

Brasília, Março de 2016



GENÔMICA ESTRUTURAL E FUNCIONAL DE FUNGOS DO GÊNERO *Trichoderma*

Andrei Stecca Steindorff

Orientador:

Dra. Eliane Ferreira Noronha

Co-orientador: **Dr. Georgios Joannis Pappas Júnior**

Co-orientador estrangeiro: **Igor Grigoriev**

Tese de doutorado apresentada ao Instituto de Biologia Celular do Instituto de Ciências Biológicas da Universidade de Brasília como requisito para obtenção do título de doutor em Ciências Biológicas – Biologia Molecular.

Brasília, Março de 2016

Trabalho realizado no laboratório de Enzimologia, Departamento de Biologia Celular do Instituto de Ciências Biológicas da Universidade de Brasília, Laboratório de Enzimologia da Universidade Federal de Goiás e no Grupo de anotação de genomas do *Joint Genome Institute* (JGI), California, Estados Unidos.

Orientador:

Dra. Eliane Ferreira Noronha

Co-orientador: **Dr. Georgios Joannis Pappas Júnior**

Co-orientador estrangeiro: **Igor Grigoriev**

Banca Examinadora:

Prof. Dr. Roberto do Nascimento Silva (USP) - Examinador Externo

Profa. Dra. – Léia Cecilia de Lima Fávaro (Embrapa Agroenergia) - Examinador Externo

Prof. Dr. Fernando Lucas de Melo (UnB) - Examinador Interno

Prof. Dr. Robert Neil Gerard Miller (UnB) – Examinador Interno

Prof. Dr. Eliane Ferreira Noronha (UnB) – Orientador

Membro Suplente:

Prof. Dr. Tatsuya Nagata (UnB)

*“Success consists of going from
failure to failure without loss of
enthusiasm.”*

Winston Churchill

Agradecimentos

Gostaria de agradecer inicialmente à minha família, que sempre me apoiou em todos os momentos de minha vida. Agradeço especialmente minha mãe, que sempre esteve ao meu lado para o que fosse necessário. Agradeço também a Lis, por ter me aguentado todo este período de doutorado, idas e vindas à Brasília e um ano nos Estados Unidos. Apesar disso tudo, foi sempre uma companheira ideal para todos os momentos.

Agradeço a todos que de alguma forma me ajudaram durante todo esse período. Aos meus “pais” acadêmicos: Eliane, que sempre busca ajudar todos os alunos e foi uma “mãezona” - Obrigado por tudo! - E ao Cirano, que sempre me apoiou nos últimos 11 anos que trabalhamos juntos, sem suas oportunidades, com certeza não estaria aqui. Agradeço a todos os alunos dos laboratórios de Enzimologia UnB e UFG, que sempre me receberam de braços abertos. Agradeço a Francilene pelo apoio nos trabalhos, sempre de bom humor e disposta a ajudar. Agradeço também ao professor Roberto pela amizade e por todas as oportunidades de parcerias oferecidas. Obrigado professor Georgios pelo apoio nesse complicado mundo da bioinformática e ao professor Robert, coordenador da pós-graduação e grande parceiro deste trabalho.

Um agradecimento especial aos meus companheiros *roommates* Marcelo e Fabyano. Com certeza com vocês Brasília ficou melhor. Todas as horas de conversas filosóficas, não poderiam ter melhores companhias. Também ao João Paulo, que vez ou outra nos dava o ar da graça em Brasília.

Agradeço ao Igor Grigoriev, que me recebeu da melhor forma no “*gat*” (*genome annotation team*), e que apesar de todas as burocracias envolvidas para me receber, não poupou esforços para isso. Posso dizer que me senti parte do grupo, inclusive dando opiniões nas reuniões de grupo. Me possibilitou também participar da disciplina de genômica comparativa na UC Berkeley, que foi bastante proveitosa em todos os sentidos. Agradeço aos anotadores Robert, Sajeet, Alan, Stephen, Robin e Asaf, que abriram minha mente para o universo da genômica e bioinformática de uma forma que eu não imaginava possível.

Agradeço à UnB, especialmente ao Programa de Pós-graduação em Biologia Molecular pela oportunidade. Com certeza um dos melhores e mais tradicionais programas de pós-graduação do Brasil. Agradeço ao CNPq pela bolsa de doutorado e à CAPES pela bolsa sanduíche, ambas fundamentais para o desenvolvimento deste trabalho.

SUMÁRIO

ÍNDICE DE FIGURAS	7
ÍNDICE DE TABELAS	9
RESUMO	10
ABSTRACT	11
INTRODUÇÃO GERAL	12
Da molécula de DNA aos genomas	12
Sequenciamento de DNA.....	13
<i>Next Generation Sequencing (NGS)</i>	14
REVISÃO BIBLIOGRÁFICA	17
Taxonomia do gênero <i>Trichoderma</i>	17
O gênero <i>Trichoderma</i> e o micoparasitismo.....	19
<i>Sclerotinia sclerotiorum</i> e mecanismos de biocontrole	23
OBJETIVO GERAL	25
Objetivos específicos	26
<u>CAPÍTULO 1: GENÔMICA COMPARATIVA DE FUNGOS DO GÊNERO</u>	
<u>TRICHODERMA</u>	27
INTRODUÇÃO	27
METODOLOGIA	28
Manutenção dos isolados	28
Condições de cultivo para isolamento de DNA	29
Sequenciamento, montagem e anotação do genoma estrutural	29
Comparação dos genomas in silico.....	31
Identificação de genes ortólogos no gênero <i>Trichoderma</i>	31
Construção da filogenia de <i>Trichoderma</i>	32
Categorização das proteínas quanto à função	32
Análise de componentes principais (PCA) utilizando dados filogenéticos.....	33
Análise comparativa das quitinases (GH18) de <i>Trichoderma</i>	33
RESULTADOS E DISCUSSÃO	34

Sequenciamento, montagem e anotação do genoma estrutural do <i>T. harzianum</i> TR274	34
Genômica comparativa do gênero <i>Trichoderma</i>	37
Filogenia e conservação das proteínas	37
Função e evolução de classes protéicas.....	43
Perda das GH18 subgrupo C1 pela seção <i>Longibrachiatum</i>	46
CONCLUSÕES	50
<u>CAPÍTULO 2: ANÁLISE DO TRANSCRITOMA DE <i>T. harzianum</i> CRESCIDO NA PRESENÇA DE PAREDE CELULAR DE <i>S. sclerotiorum</i></u>	52
INTRODUÇÃO	52
CONCLUSÃO	68
<u>CAPÍTULO 3: ANÁLISE DO TRANSCRITOMA DA INTERAÇÃO IN VIVO ENTRE <i>T. harzianum</i> E <i>S. sclerotiorum</i></u>	69
INTRODUÇÃO	69
METODOLOGIA	70
Manutenção dos isolados	70
Condições de cultivo para isolamento de RNA	71
Preparação das amostras de RNA para sequenciamento	72
Sequenciamento e análise dos dados de RNA-seq	72
RESULTADOS E DISCUSSÃO	74
CONCLUSÕES	84
REFERÊNCIAS BIBLIOGRÁFICAS	85
ARTIGOS PUBLICADOS DURANTE O DOUTORADO	95

ÍNDICE DE FIGURAS

- Figura 1.** Esquema atual do Dogma Central da Biologia Molecular demonstrando o fluxo da informação genética..... 12
- Figura 2.** Ensaio de confronto direto entre *Trichoderma* spp. e *R. solani* após 10 dias de crescimento em meio BDA a 25°C. Setas laranjas indicam o quanto a colônia de *Trichoderma* spp. cresceu sobre a de *R. solani*. Figura retirada de Atanasova et al., 2013..... 22
- Figura 3.** Comparação entre dois diferentes programas para montagem de genomas: AllPathsLG e MaSurCa. A comparação foi feita usando o genoma de referência de *T. harzianum* CBS 226.95. As linhas verdes são os alinhamentos com boa qualidade, mas com extremidades não correspondentes ao *scaffold* de referência e as linhas vermelhas são alinhamentos de baixa qualidade também extremidades não correspondentes. As diferenças de altura e tons representam diferentes contigs..... 35
- Figura 4.** Filogenia e características dos genomas de espécies de *Trichoderma*. A – Árvore de máxima-verossimilhança de 9 isolados de *Trichoderma* (Triat: *T. atroviride*; Trias: *T. asperellum*; Trihar: *T. harzianum* TR274; Triha: *T. harzianum* CBS 226.95; Trivi: *T. virens*; Triru: *T. reesei* RUTC30; Trire: *T. reesei* QM6a; Trilo: *T. longibrachiatum*; Trici: *T. citrinoviride* e grupos externos: Fusgr: *Fusarium graminearum*; Chagl: *Chaetomium globosum*). Ramos vermelhos são espécies predominantemente micoparasitas e em azul, espécies predominantemente saprófitas. B – Proporção de regiões repetitivas nos genomas. C – Compartilhamento de proteínas entre as espécies de *Trichoderma*. D – Comparação das categorias de proteínas transportadoras, SSCP, CAZymes, fatores de transcrição, proteases e metabolismo secundário..... 39
- Figura 5.** Diagrama de Venn com 7 diferentes espécies de *Trichoderma*. Nesta figura, não foram apresentadas isolados da mesma espécie. As cores representam o número de espécies comparadas e o número central a quantidade de proteínas compartilhadas entre as espécies de *Trichoderma*. Blocos circulados em vermelho correspondem a proteínas exclusivas das seções e em preto, proteínas compartilhadas entre seções *Trichoderma* e *Pachibasium*..... 41
- Figura 6.** Evolução das famílias genicas em *Trichoderma*. A – Árvore filogenética juntamente com a quantidade de famílias gênicas que sofreram expansão (seta para cima) e contração (seta para baixo) em cada nó relativo a seu ancestral. Cada cor nas setas representam uma classe de proteínas. Os valores representados nos nós basais de cada seção representam a média do tamanho dos genomas

dos isolados constituintes. B – Famílias que sofreram expansão e contração durante a evolução do gênero *Trichoderma*. Cada linha é uma família genica e cada coluna é um nó da árvore filogenética. O nível da cor de cada célula demonstra a quantidade de cópias gênicas predita para cada ancestral de acordo com o programa CAFE. Os números dentro das células mostram a quantidade de cópias que foram significativamente expandidas (azul) e contraídas (vermelho) com p -valor $\leq 0,05$ 44

Figura 7. Análise de componentes principais (PCA) filogenético. Espécies predominantemente micoparasitas são mostradas em vermelho e predominantemente saprófitas em azul. O círculo vermelho mostra a seção ancestral..... 46

Figura 8. Organização dos domínios de estruturais e funcionais de quitinases. SP – Peptídeo Sinal. Domínio conservado típico da família de glicosil hidrolase GH18. BD – Domínio de Ligação. Retirado de Seidl, 2008..... 47

Figura 9. Arvore filogenética de máxima verossimilhança de quitinases do gênero *Trichoderma*. Subgrupos C1 e C2 enfatizados juntamente com sua estrutura de domínios. Cada cor nos nomes representa uma espécie diferente. Nós onde os *bootstraps* não são mostrados apresentam valores maiores que 90..... 49

Figura 10. Comparação dos perfis de expressão genoma completo de *T. harzianum* TR274 na interação com *S. sclerotiorum* antes do contato (BC), durante o contato (C) e após o contato (AC). Genes diferencialmente expressos são mostrados em vermelho ($P < 0.05$)..... 74

Figura 11. Diagrama de Venn dos 3396 genes diferencialmente expressos, divididos em reprimidos (*down-regulated*) e induzidos (*up-regulated*) durante as três condições de interação: Antes do contato (BC), durante o contato (C) e após o contato (AC)..... 75

Figura 12. Quantidade de genes diferencialmente expressos categorizados como proteases, CAZymes, transportadores, fatores de transcrição (TF), metabolismo secundário (SecMet) e pequenas proteínas secretadas e ricas em cisteínas (SSCP). As barras positivas mostram a quantidade de genes induzidos e as barras negativas, reprimidos..... 76

Figura 13. *Heatmap* e categorização dos agrupamentos dos genes diferencialmente expressos mostrados na escala de \log_2 *fold change*. As categorias mostradas nas barras horizontais mostram a proporção de cada classe de CAZymes, proteases, metabolismo secundário, pequenas proteínas secretadas ricas em cisteínas (SSCP), Fatores de transcrição (TF) e transportadores..... 78

ÍNDICE DE TABELAS

Tabela 1. Comparação das características do genoma do isolado de <i>T. harzianum</i> TR274 com os genomas de outros isolados e espécies de <i>Trichoderma</i>	36
Tabela 2. Total de proteínas ortólogas com anotação PFAM.....	40
Tabela 3. Domínios PFAM diferencialmente expressos pelo teste exato de Fisher ($P < 0,05$) antes do contato.....	80
Tabela 4. Domínios PFAM diferencialmente expressos pelo teste exato de Fisher ($P < 0,05$) durante o contato.....	81
Tabela 5. Domínios PFAM diferencialmente expressos pelo teste exato de Fisher ($P < 0,05$) após o contato.....	82

RESUMO

O controle biológico é um processo complexo que inclui diferentes mecanismos e uma diversidade de vias metabólicas. Espécies de *Trichoderma harzianum* são conhecidas por sua atividade de biocontrole contra patógenos de plantas. Para melhor entender os mecanismos utilizados por *T. harzianum* no controle biológico, no presente trabalho foi sequenciado o genoma do isolado TR274 usando sequenciamento Illumina, assim como seu respectivo transcrito na interação direta com *Sclerotinia sclerotiorum* ou na presença de sua parede celular. A montagem do genoma feita utilizando o programa AllPaths-LG cobertura máxima de 100x, resultou em 2282 contigs, tamanho do genoma de 40,8 Mb e um conteúdo GC de 47.7%, similar aos outros genomas de *Trichoderma*. Um total de 13932 genes foram anotados. Análise do *Core Eukariotic Genes Dataset* (CEGMA) sugere que o genoma está 100% completo e 97,9% das sequências de RNA-seq alinharam corretamente no genoma. A análise filogenética usando proteínas ortólogas com todas as espécies de *Trichoderma* sequenciadas no JGI, confirmam a divisão nas seções *Trichoderma* (*T. asperellum* e *T. atroviride*), *Longibrachiatum* (*T. reesei*, *T. citrinoviride* e *T. longibrachiatum*) e *Pachibasium* (*T. harzianum* e *T. virens*). Das proteínas ortólogas anotadas, 8242 compõem proteínas compartilhadas por todas as espécies, as proteínas espécie específicas variam de 262 (*T. reesei*) a 1803 (*T. longibrachiatum*). Os dois genomas de *T. harzianum* analisados sugerem uma alta similaridade entre eles, mesmo tendo sido isolados de locais e continentes distintos, um de solo de cerrado no Brasil e outro de solo de jardim na Inglaterra. Análises de genes envolvidos com o metabolismo secundário, CAZymes, transportadores, proteases e fatores de transcrição foram feitas. A seção *Pachibasium* expandiu virtualmente todas as categorias analisadas quando comparada com as outras seções. Análise CAFE mostrou uma correlação positiva entre estas expansões e o tamanho dos genomas. O subgrupo C1 das quitinases foi completamente perdido pela seção *Longibrachiatum*. Estes resultados sugerem que estas famílias proteicas tem um importante papel nos seus respectivos fenótipos. As abordagens transcritômicas mostraram que a interação entre *T. harzianum* e *S. sclerotiorum* é bem complexa e envolve a produção de metabólitos secundários e síntese de transportadores antes e durante o contato, com uma modulação principalmente de enzimas hidrolíticas após o contato. Dos genes encontrados diferencialmente expressos na condição de crescimento em parede celular, 86,8% foram também encontrados diferencialmente expressos na interação direta. Cerca de 25% de todo o genoma de *T. harzianum* foi modulado durante a interação com *S. sclerotiorum*.

ABSTRACT

Biological control is a complex process, which requires many mechanisms and a high diversity of biochemical pathways. *Trichoderma harzianum* species complex are well known for their biocontrol activity against many plant pathogens. To gain new insights into the biocontrol mechanism employed by *T. harzianum*, we sequenced genome of the isolate TR274 with its transcriptome during direct interaction with the fungal pathogen *Sclerotinia sclerotiorum* and its cell wall, using Illumina sequencing. Whole genome assembly was performed using AllPaths-LG, with a maximum coverage of 100x. The assembly resulted in 2282 contigs, with an estimated genome size of 40.8 Mb and GC content of 47.7%, similar to other *Trichoderma* genomes. Using the JGI Annotation Pipeline we predicted 13,932 genes, with high transcriptome support. *Core Eukariotic Genes Dataset* (CEGMA) tests suggested 100% genome completeness and 97.9% of RNA-SEQ reads mapped to the genome. The phylogenetic comparison using orthologous proteins with all *Trichoderma* genomes sequenced at JGI, corroborates the *Trichoderma* section division described previously (*T. asperellum* and *T. atroviride*), *Longibrachiatum* (*T. reesei*, *T. citrinoviride* and *T. longibrachiatum*) and *Pachibasium* (*T. harzianum* and *T. virens*). A Venn diagram was built with orthologs proteins, with 8242 composing the core protein group and species specific varying from 262 proteins (*T. reesei*) to 1803 (*T. longibrachiatum*). The comparison between two *Trichoderma harzianum* CBS 226.95 and TR274 isolates, suggests a high genome similarity. Analyses of the secondary metabolites, CAZymes, transporters, proteases and transcription factors were performed. The *Pachibasium* section expanded virtually all categories analyzed compared with the other sections. CAFE analysis showed positive correlation between these families and genome size. The chitinase subgroup C1 was completely absent in *Longibrachiatum* section members. These results suggest that these proteins families play an important role on their respective phenotypes. Transcriptome analysis suggests that the interaction between *T. harzianum* and *S. sclerotiorum* is complex, involving production of secondary metabolites and transporters before and during the contact, with a modulation of CAZymes after contact. A total of 86.8% of differentially expressed genes during growth on *Sclerotinia sclerotiorum* cell wall were found during direct interaction. Approximately the 25% of whole *T. harzianum* genome was seen to be modulated during the interaction with *S. sclerotiorum*.

INTRODUÇÃO GERAL

Da molécula de DNA aos genomas

Talvez um dos artigos científicos mais importantes da era moderna, nas ciências biológicas, seja aquele que descreveu a estrutura da molécula de DNA (Watson & Crick, 1953). Este modelo descrevia a maioria das características necessárias para o advento da Biologia Molecular que conhecemos atualmente. Não por acaso estes dois pesquisadores foram agraciados com o prêmio Nobel de medicina e fisiologia em 1962. Mas somente a descrição da estrutura da molécula não foi suficiente para elucidar todo o sistema responsável pela complexidade de um organismo vivo. Em 1958, Francis Crick postulou o “Dogma central da biologia molecular”, que descreve com acontece o fluxo de informação dentro da célula, desde o DNA até a construção das proteínas.



Figura 1. Esquema atual do Dogma Central da Biologia Molecular demonstrando o fluxo da informação genética.

A importância de todas estas descobertas é inquestionável, mas conseqüentemente geraram incontáveis perguntas biológicas. Como um alfabeto pequeno e unidimensional é responsável pela diversidade que conhecemos? Como existem organismos tão diferentes como bactérias e humanos sendo que todos contêm a mesma natureza informacional? A chave para responder boa parte destas perguntas foi o advento da técnica de sequenciamento do DNA.

Sequenciamento de DNA

Em 1977, Frederick Sanger e Alan Coulson publicaram dois trabalhos onde reportavam metodologias eficazes para a determinação da sequência de DNA de organismos, abrindo as portas para uma completa revolução na biologia com o desvendamento da sequência completa de genes e genomas (Sanger et al., 1977a; Sanger et al., 1977b). Em um desses trabalhos, o genoma completo do fago phi X174 era apresentado; tendo sido este o primeiro genoma de um organismo a ser completamente desvendado por seres humanos. Estes trabalhos renderam a Sanger seu segundo prêmio Nobel em química, entregue no ano de 1980.

Sua brilhante técnica de sequenciamento de DNA utilizando os chamados dideoxynucleotídeos imperou absoluta na ciência genômica ao longo dos 30 anos que se seguiram à sua publicação original. Sanger, portanto, abriu as portas para toda a análise informacional em ciências genômicas e mostrou que era definitivamente possível compreender genomas ao soletrá-los unidimensionalmente através da sequência de bases químicas do DNA: Adenina (A), Citosina (C), Timina (T) e Guanina (G).

Com a corrida pelo genoma humano no final da década de noventa e começo do novo século (Davies, 2001), novas estratégias começaram a ser pensadas para a produção de sequências genômicas de forma ainda mais massiva do que utilizando os sequenciadores de eletroforese capilar baseados em dideoxynucleotídeos fluorescentes (Boysen et al., 1997). Empresas biotecnológicas e mentes

inovadores rapidamente perceberam que surgimento de uma tecnologia de sequenciamento mais veloz certamente proporcionaria mais um salto ainda maior em nossa compreensão sobre os genomas de diferentes organismos.

Next Generation Sequencing (NGS)

Em 2005, os primeiros resultados dos esforços dos engenheiros em produzir máquinas ainda mais eficientes de sequenciamento de DNA, apareceram (Margulies et al., 2005). A empresa 454 Life Sciences anuncia então sua técnica de sequenciamento por síntese, onde cada base era lida à medida que fosse adicionada à cadeia recém-formada de DNA (Schuster, 2008). No método de Sanger, a base lida era verificada pela marcação fluorescente e pelo peso molecular da molécula contendo a sequência parcial do DNA através de uma análise eletroforética. O novo método que utilizava a inserção de cada base, parecia ser bastante confiável.

No trabalho original publicado na revista Nature, os autores descrevem um “sistema de sequenciamento paralelo e escalável com automatização significativamente maior do que os instrumentos padrão de eletroforese capilar” (Margulies et al., 2005). O equipamento da empresa 454 era capaz de produzir 25 milhões de pares de bases com precisão de 99% ou mais em apenas uma análise de quatro horas na máquina. Isso representava um aumento de 100 vezes na produção de sequências de DNA quando comparado à tecnologia mais moderna existente — um melhoramento grandioso da tecnologia originalmente descrita por Sanger realizado nos fins da década de 90 (Venter et al., 1996; Boysen et al., 1997). E para provar seus resultados e sua ideologia de “uma nova revolução genômica” está para acontecer, o trabalho original da 454 Life Sciences apresentava o resequenciamento da bactéria *Mycoplasma genitalium* com alta acurácia (99.96%) e realizado em uma única corrida dessas máquinas inovadoras (Margulies et al., 2005). Todavia este trabalho foi recebido

com cautela pela comunidade científica, que preferiu evidenciar as vantagens do sequenciamento Sanger sobre esta técnica do que abraçá-la sem hesitação.

De fato, a nova técnica ainda produzia sequências pequenas e de difícil análise (Wommack et al., 2008). Enquanto a técnica de Sanger produzia sequências de aproximadamente 750 pares de bases; a nova técnica gerava sequências contendo apenas cerca de 100 ou 200 letras de DNA encadeadas (Margulies et al., 2005). Já em 2006, entretanto, novos trabalhos publicados principalmente na área da metagenômica demonstravam a versatilidade e eficácia da nova técnica de sequenciamento ao amostrar a diversidade microbiana observada em uma mina (Edwards et al, 2006) ou nas profundezas do mar (Sogin et al., 2006), além de genomas virais observados em oceanos (Angly et al., 2006). Além disso, até 2014 aproximadamente, a tecnologia de pirosequenciamento evoluiu chegando ao tamanho médio de *reads* similares aos da técnica de Sanger. Mas devido ao alto custo, esta técnica foi descontinuada.

A máquina Genome Analyser (GA) da empresa Illumina, foi uma das novas máquinas criadas capazes de produzir dados genômicos em larga escala através de técnicas de sequenciamento por síntese. A cada execução da máquina com uma amostra diferente, milhões de sequências genômicas podem ser produzidas que ajudarão os pesquisadores na busca pela compreensão da biologia dos organismos — busca esta que invariavelmente traz benefícios para a sociedade, como na descoberta de alvos para fármacos ou enzimas de utilização biotecnológica.

Trabalhos subsequentes têm então demonstrado a extensa aplicabilidade de plataformas de sequenciamento de nova geração para a genômica em diversas áreas da pesquisa científica, como por exemplo: descobertas de rearranjos genômicos em tumores; genômica funcional (Morozova e Marra, 2008); análise do transcriptoma em larga escala (Asmann et al., 2008); testes de diagnósticos moleculares (Voelkerding et al, 2009); variação genética em diversos organismos (Imelfort et al., 2009), incluindo humanos (Guryev e Cuppen, 2009); descoberta de vacinas (Dhiman et al., 2009); melhoramento genético de plantas cultivadas (Varshney et al., 2009) e muitos outros.

Além disso, a produção e análise completa de genomas humanos, inaugura a era da genômica individual com a publicação do genoma diplóide de James Watson totalmente sequenciado por máquinas de nova geração (Wheller et al., 2008). O projeto 1000 genomas também apresenta o sequenciamento completo de diversos genomas individuais diplóides de seres humanos (<http://www.1000genomes.org>). Devido ao grande desafio de se comparar genomas de diferentes organismos, existem na comunidade científica outros projetos “1000 genomas”, como por exemplo focado em plantas (<http://phytozome.jgi.doe.gov/pz/portal.html>) e em fungos (<http://genome.jgi.doe.gov/programs/fungi/index.jsf>).

A análise dessa quantidade gigantesca de dados de sequenciamento de DNA, fez crescer uma área da ciência pouco conhecida pelo grande público: a bioinformática. Buscar auxílio em ferramentas computacionais para o estudo da biologia é um método colocado em prática por pesquisadores desde meados da década de 1970. Naquela época, programas eram desenvolvidos para comparar sequências de proteínas e aminoácidos, por exemplo. Foi apenas a partir de 1995, quando se publicou o primeiro genoma de uma bactéria, que houve uma explosão do uso das técnicas de informática para a realização de pesquisas biológicas. Chegará um momento em que a informática estará para a biologia assim como a matemática está para a física.

Este trabalho é composto por três capítulos. O primeiro descreve a análise comparativa dos genomas de sete espécies de fungos do gênero *Trichoderma* e seus desdobramentos na evolução deste gênero. O segundo capítulo trata do transcrito de *T. harzianum* TR274 durante o crescimento em parede celular do fitopatógeno *Sclerotinia sclerotiorum* comparado com crescimento em glicose. Este trabalho já foi publicado no periódico BMC Genomics em 2013 e será mostrado no formato da revista. O terceiro capítulo aborda a análise do transcrito de *T. harzianum* TR274 durante a interação em placas de *petri* com *S. sclerotiorum*, simulando de forma mais precisa uma interação *in vivo*. Esta tese faz parte de projetos cujos objetivos é descrever o genoma de um *T. harzianum* isolado de solo do

cerrado e o transcrito de este isolado na presença de fitopatógenos de importância econômica, neste caso, *S. sclerotiorum*.

REVISÃO BIBLIOGRÁFICA

Taxonomia do gênero *Trichoderma*

De acordo com o Index Fungorum (ver www.indexfungorum.org) o gênero *Trichoderma* pertence ao Domínio Eucaryota, — Grupo Fungi/Metazoa, Reino Fungi, Filo Ascomycota, Subfilo Pezizomycotina, Classe Sordariomycetes, Subclasse Hypocreomycetidae, Ordem Hypocreales, — Família Hypocreales mitospóricos (ou Família Hypocreaceae), Gênero *Trichoderma*. A taxonomia de *Trichoderma* está em plena evolução e pode ser estabelecida com base em critérios morfológicos, moleculares e, mais recentemente, por espectrometria de massa de células intactas. Por muitos anos, desde que o gênero *Trichoderma* foi descrito pela primeira vez por Persoon em 1754, o gênero foi frequentemente presumido como uma única espécie, *T. viride*. Rifai (1969) em sua classificação para o gênero, reconheceu nove espécies: *T. viridi* Persoon, *T. aureoviride* Rifai, *T. koningii* Lieckfeldt, *T. harzianum* Rifai, *T. longibrachiatum* Bisset, *T. polysporum* Link e *T. glaucum* Rifai. Ao estudar a discriminação das espécies delimitadas por Rifai, Bisset (1991a,b) propôs o termo “seções” para separar espécies similares dentro das nove espécies originalmente aceitas até então. De acordo com esse sistema taxonômico, os fungos do gênero *Trichoderma* estariam distribuídos em quatro seções: *Trichoderma*, *Longibrachiatum*, *Saturnisporum* e *Pachybasium*. Entretanto, seção *Saturnisporum* foi englobada como sendo parte da seção *Longibrachiatum* (Kuhls *et al.*, 1997). Outro fator importante é a descrição de espécies crípticas, que por definição, descreve duas ou mais espécies que se escondem atrás de um único nome. Existem alguns casos dentro do gênero *Trichoderma*, onde o mais estudado

é o complexo *T. harzianum* (Druzhinina *et al.*, 2010). Neste estudo se mostrou que existe um clado definido de *T. harzianum sensu stricto* e um conjunto de espécies não definidas denominada “matriz pseudoharzianum” mais relacionadas com o teleomorfo *Hypocrea lixii*. Neste trabalho também se concluiu que teleomorfos (*Hypocrea*) e anamorfos (*Trichoderma*) devem ser analisados como espécies diferentes.

Buscando entender o gênero *Trichoderma*, várias revisões foram realizadas baseadas em caracteres morfológicos e se mostraram insatisfatórias (Druzhinina *et al.*, 2006), pois não indicavam o suposto grau de divergência genética entre as espécies. Isso acontece devido à sobreposição de diversos caracteres distintivos (Chaverri & Samuels, 2003). Assim, a taxonomia baseada somente em observações morfológicas não é suficiente para identificar uma espécie de forma segura. Com o advento das técnicas moleculares esse cenário foi mudado, o que levou a um aumento no número de espécies de *Trichoderma* identificadas. Vários genes foram utilizados para esse fim. O primeiro estudado foi o espaçador interno da transcrição (ITS) (Kindermann *et al.*, 1998), seguido de vários outros como o fator de alongamento da transcrição (tef - exons e introns, sendo que esse último tem um alto poder de resolução dentro dos clados), endoquitinase (chit), Subunidade 2 da RNA polimerase (rpb2), calmodulina (cal) e actina (act) (Lopes *et al.*, 2012).

Em 2015 um trabalho na *International Mycological Association* descreveu 254 espécies de *Trichoderma* (Bisset *et al.*, 2015) juntamente com a descrição de todas as espécies e suas respectivas nomenclaturas. Outro parâmetro a ser seguido, decidido pelo ICTF (*International Commission on the Taxonomy of Fungi*) e *International Subcommittee on the Taxonomy Of Trichoderma and Hypocrea* (www.isth.info) é que se deve dar prioridade para o uso do nome *Trichoderma* ao termo *Hypocrea* para evitar confusões muito comuns quanto a classificação desses fungos.

O gênero *Trichoderma* e o micoparasitismo

O potencial das espécies do gênero *Trichoderma* como agentes de controle biológico foi reconhecido primeiramente por Weindling no começo da década de 1930, quando este autor descreveu a ação micoparasita de *Trichoderma* contra fungos do gênero *Rhizoctonia* e *Sclerotinia* e seus efeitos benéficos no controle das patologias causadas por estes fungos. Isso estimulou várias pesquisas nessa área e hoje algumas espécies vêm sendo utilizadas com sucesso, inclusive comercial, no controle biológico de fungos fitopatogênicos de interesse na agricultura. O gênero *Trichoderma* compreende um grupo de fungos saprófitas e micoparasitas componentes ativos da microbiota do solo, que são amplamente utilizados como agentes de controle biológico (Druzhinina et al, 2011). As espécies deste gênero caracterizam-se por utilizarem uma grande variedade de compostos como fonte de carbono e nitrogênio, serem resistentes a inibidores produzidos por outros microrganismos e tolerantes a diferentes tipos de fungicidas (Daryaei et al., 2016). Além disto, seu rápido crescimento, produção de clamidósporos e sua elevada capacidade de degradar carboidratos estruturais e não estruturais, os tornam de elevado interesse para fins de controle biológico, uma vez que estas características ecológicas os possibilitam competir, colonizar e proteger importantes pontos de entrada de patógenos nas plantas hospedeiras, como raízes, lesões e tecido em senescência (Seidl et al, 2009).

Os fungos da espécie *Trichoderma harzianum* são estudados como agentes no biocontrole de fitopatógenos de solo, folhas e frutos, sendo que pelo menos 50 formulados à base desta espécie já são comercializados e utilizados mundialmente no controle biológico de fungos fitopatogênicos (Woo et al, 2006). O mecanismo de biocontrole destes fungos é um processo complexo que pode ocorrer através de diferentes mecanismos diretos e indiretos ou por uma combinação destes, como antibiose,

competição por nutrientes, micoparasitismo ou indução de mecanismos de defesa nas plantas hospedeiras (Steindorff et al., 2012; Pereira et al., 2014).

O micoparasitismo é o mecanismo mais estudado para a espécie *T. harzianum* e ocorre em várias etapas que vão desde o reconhecimento até a morte do fungo hospedeiro. Os mecanismos de ação direta ocorrem a partir da interação física micoparasita-hospedeiro e são desencadeados eventos em cascata que culminam com o aumento na expressão de genes que codificam enzimas com ação na degradação da parede celular do fungo hospedeiro, bem como, de outras proteínas com papel ainda não elucidado neste processo (Druzhinina et al, 2011). A expressão de genes associados com o biocontrole parece ser regulada por uma cascata de transdução de sinal intracelular, ativada pela ligação a moléculas derivadas do hospedeiro e pode envolver a sinalização via proteína G, através de mensageiros secundários como o cAMP, bem como via proteína quinase mitógeno ativada (MAPK) (Zeilinger & Omann, 2007). Várias enzimas líticas, incluindo β -glucanases, quitinases e proteases, são produzidas durante ação micoparasita. Estas enzimas agem de forma sinérgica na degradação da parede celular do hospedeiro e apresentam um importante papel no mecanismo de biocontrole (Martin et al. 2007). Fungos do gênero *Trichoderma spp.* são grandes produtores destas enzimas que degradam parede celular dos hospedeiros, sendo bastante utilizados no campo no controle de pragas e indução de crescimento e defesa em plantas com resultados positivos (Carvalho, 2015). Um mecanismo comum à várias espécies micoparasitas é a produção de metabólitos secundários que inibem o crescimento de fungos antagonista (Brito et al., 2014). O termo “metabólito secundário” inclui uma ampla variedade de compostos que estão envolvidos em vias de sinalização, desenvolvimento e interação com outros organismos. Sabe-se que esta produção é variável entre as espécies do gênero e pode variar também dependendo da interação com diferentes fitopatógenos (Vinale et al., 2009).

Dentre outros metabólitos secundários produzidos por *Trichoderma spp.*, há também relatos e descrição de compostos como pironas, terpenóides, esteróides, gliotoxinas, gliovirinas (Patron et al.,

2007), e alguns peptídeos antibióticos, conhecidos como “peptaibols” (Brito et al., 2014). A descoberta de peptídeos antibióticos produzidos por fungos têm despertado interesse pelo fato de ser uma possível alternativa contra fitopatógenos. Alguns destes peptídeos produzidos por *Trichoderma harzianum* já foram descritos inibindo a atividade de β -glucanas sintetases de *Botrytis cinerea*, impedindo a reconstrução da parede celular do hospedeiro durante o antagonismo com *Trichoderma* (Lorito et al., 1996).

A análise comparativa dos genomas de *T. atroviride*, *T. virens* e *T. reesei* revelou que a capacidade micoparasita é inata ao gênero *Trichoderma* (Kubicek et al., 2011), mas as estratégias utilizadas por estas espécies durante a interação com o fitopatógeno *Rhizoctonia solani*, analisadas por transcritomas, foram diferentes (Figura 2) (Atanasova et al., 2013). A espécie *T. atroviride* não inibiu o crescimento da *R. solani* mas foi capaz de crescer por cima da colônia e consumi-la rapidamente, produto de sua maior produção de enzimas hidrolíticas. Já o *T. virens* foi mais eficaz na inibição do crescimento, produto de sua maior produção de metabólitos secundários. Neste experimento, o *T. reesei* não foi capaz de inibir o crescimento nem crescer por cima da colônia de *R. solani*, o que aconteceu foi o aparecimento de uma espécie de barreira impedindo o crescimento do antagonista, com fracos sinais de micoparasitismo (Atanasova et al., 2013).

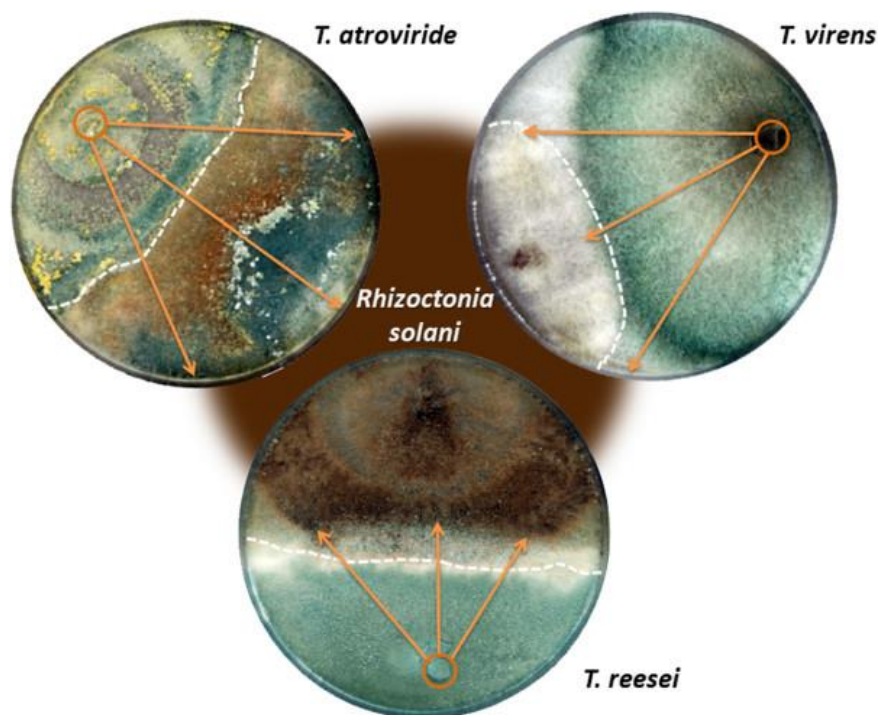


Figura 2. Ensaio de confronto direto entre *Trichoderma spp.* e *R. solani* após 10 dias de crescimento em meio BDA a 25°C. Setas laranjas indicam o quanto a colônia de *Trichoderma spp.* cresceu sobre a de *R. solani*. Figura retirada de Atanasova et al., 2013.

A espécie *T. harzianum*, já foi analisada em experimentos similares (Steindorff et al., 2012) na interação com *Fusarium solani* mostrando uma inibição efetiva no avanço do fitopatógeno, assim como o crescimento por cima da colônia. A resposta no ponto de vista de expressão foi mais complexa com a expressão de enzimas hidrolíticas, transportadores e metabólitos secundários (Steindorff et al., 2012). Um outro fator que traz mais complexidade à interação Trichoderma/fitopatógeno é a diversidade metabólica existente dentro de uma mesma espécie (Lopes et al., 2012). *T. harzianum* e *T. asperellum*, espécies mais prevalentes no Brasil, são bons exemplos desta complexidade. Neste trabalho sugere-se que a etapa inicial para escolha de um bom agente de biocontrole seria seu desempenho ambiental e capacidade micoparasítica, não direcionado à uma espécie ou outra.

O acúmulo de informações quanto à capacidade antagônica e habilidade de colonização e proliferação em diferentes habitats, associado a sistemas de produção maçal, formulação e aplicação

eficientes, tem levado ao desenvolvimento de produtos estáveis à base de *Trichoderma* (especialmente *T. harzianum*). Tais produtos, como RootShield™, T-22 Planter Box™ e Trichodex™, já se encontram em comercialização nos Estados Unidos, Europa e Israel. Estes produtos controlam eficientemente o crescimento de fungos fitopatogênicos do gênero *Fusarium*, *Phythium* e *Rhizoctonia* reduzindo a incidência de doenças fúngicas em culturas de interesse econômico. No entanto, o seu sucesso é maior quando aplicado em solos com características e pressão bióticas e abióticas semelhantes às do solo de origem. Desta forma, a seleção e avaliação da atividade antagônica de potenciais agentes de biocontrole é uma atividade constante em programas de controle biológico. E considerando-se o grande problema da produção agrícola no Brasil, ainda é necessária a busca por novos agentes efetivos às condições climáticas e pressões bióticas dos solos brasileiros (Carvalho et al, 2011).

***Sclerotinia sclerotiorum* e mecanismos de biocontrole**

O fungo *Sclerotinia sclerotiorum* pertence ao filo dos *Ascomycetos*, classe *Leotiomycetes*, ordem *Helotiales*, família *Sclerotiniaceae*. Este patógeno pertence ao grupo dos fungos mitospóricos, subgrupo Hifomicetos, apresenta distribuição mundial e é o agente causal da “podridão-de-esclerotinia”, ou “murcha-de-esclerotinia” ou “mofo-branco” em muitas espécies de plantas cultivadas. Boland e Hall (1994) consideram *S. sclerotiorum* como fungo necrotrófico e relataram ocorrência de 408 espécies de plantas suscetíveis, pertencentes a 278 gêneros e 75 famílias, dentre elas várias de importância econômica. Purdy (1979) descreveu este fungo como sendo o mais inespecífico dos fitopatógenos e com ampla distribuição geográfica. No Brasil, este fungo foi descrito pela primeira vez, em 1920, na cultura de batata, no estado de São Paulo. Ocorre também em várias espécies 20 cultivadas como alface, girassol, soja, feijão e algodão (Carvalho et al., 2015).

Este fungo sobrevive principalmente na forma de escleródios, que são estruturas compactas, melanizadas, podendo sobreviver por longos períodos de tempo sob condições adversas. Essas propriedades são essenciais para a sobrevivência, propagação e difusão do fungo, já que encontrando condições favoráveis como alta umidade e temperatura baixa os escleródios poderão dar início à germinação. O enorme potencial reprodutivo, juntamente com a capacidade de sobrevivência em longo prazo faz dos escleródios componentes centrais na epidemiologia de *S. sclerotiorum*. Escleródios podem germinar carpogenicamente ou miceliogenicamente, dependendo das condições ambientais, resultando em duas distintas categorias de doenças. A forma miceliogênica produz hifas que podem atacar diretamente os tecidos das plantas. A formacarpogênica produz apotécios e, posteriormente, ascósporos aéreos que infectam partes de plantas hospedeiras (Oliveira, 2011).

A principal estratégia de controle de fitopatógenos, mundialmente utilizada, ainda é a aplicação de fungicidas químicos. Esta medida de controle apesar de eficiente é um processo oneroso, leva à seleção de patógenos resistentes e ao surgimento de pragas secundárias (Pires et al, 2003; Carvalho et al., 2011). Além disto, os produtos químicos utilizados são tóxicos e poluentes, e desta forma impactam o meio ambiente e comprometem a saúde dos produtores agrícolas e consumidores (Punja & Utkhede, 2003). Outra desvantagem no uso de fungicidas químicos está relacionada à sua inespecificidade, podendo eliminar microrganismos já estabelecidos no solo e benéficos às plantas aumentando assim, sua suscetibilidade aos patógenos de solo (Pires et al, 2003). Nos últimos anos, devido à pressão imposta pelas leis de impacto ambiental, a crescente preocupação da população com o meio ambiente e saúde, bem como pela necessidade de desenvolvimento de métodos de controle mais eficientes, surgiu a demanda por novas estratégias neste controle (Carvalho et al., 2015).

Novas alternativas no combate a estes fitopatógenos já são utilizadas, como o controle biológico, que se baseia no uso de microrganismos antagonistas. Este controle pode ser direto quando os antagonistas são aplicados vivos diretamente nas culturas, ou indireto, através da aplicação de seus

metabólitos (Melo & Azevedo, 2000; Poletto et al, 2006). A maioria dos produtos já disponíveis comercialmente é à base do organismo selvagem, no entanto o desenvolvimento de linhagens mais eficientes por manipulação genética também é uma estratégia promissora na formulação de novos produtos aplicados ao controle de fungos fitopatogênicos em lavouras de relevância econômica. O desenvolvimento de cultivares resistentes por transgenia também é uma estratégia muito estudada e já utilizada para algumas culturas de importância econômica como a soja, o algodão e o milho. No entanto, a sua utilização depende do entendimento dos mecanismos de defesa da planta hospedeira, assim como da identificação de genes associados à defesa e resistência ao ataque de patógenos (Seidl et al, 2009; Druzhinina et al, 2011).

OBJETIVO GERAL

O objetivo geral deste trabalho foi sequenciar o genoma de *T. harzianum* TR274 e fazer uma análise comparativa de genomas de 7 espécies diferentes de *Trichoderma* e avaliar a evolução destes genomas. Buscamos também fazer uma análise do mecanismo de interação entre *Trichoderma harzianum* e o fungo fitopatogênico *S. sclerotiorum*, visando à descrição de novos genes, vias metabólicas e de transdução de sinais envolvidas nesta interação. Esta foi a primeira iniciativa de se utilizar NGS para análise do transcrito de *T. harzianum* em condições que simulam o antagonismo contra *S. sclerotiorum*, assim como interação direta entre eles. O conhecimento mais amplo e aprofundado do grupo de genes expressos durante esta interação será de grande importância no desenvolvimento de estratégias de biocontrole de fungos fitopatogênicos mais efetivas, como também contribuirão para um melhor entendimento de mecanismos básicos da interação entre estes

microrganismos amplamente distribuídos na natureza. Além de todo o entendimento deste mecanismo, este trabalho busca oferecer subsídios para futuros trabalhos.

Objetivos específicos

- Sequenciar o genoma do isolado TR274 de *T. harzianum* a fazer anotação estrutural e funcional dos genes preditos;
- Fazer uma análise comparativa dos genomas de *Trichoderma* disponíveis;
- Construir bibliotecas para RNA-seq utilizando RNA extraído de *T. harzianum* TR274 crescido em parece celular de *S. sclerotiorum* e regiões de interação em direta em placa de *Petri* entre os mesmos;
- Analisar *in silico* a expressão diferencial de genes gerados a partir do sequenciamento NGS;
- Validar por RT-qPCR os genes com expressão diferencial entre as condições de cultivo na ausência e presença do fitopatógeno.

CAPÍTULO 1: GENÔMICA COMPARATIVA DE FUNGOS

DO GÊNERO TRICHODERMA

INTRODUÇÃO

Em biologia, o genoma é definido por toda a informação hereditária de um organismo que está codificada em seu DNA (ou, em alguns vírus, no RNA). Isto inclui tanto os genes como as sequências não-codificadoras que são muito importantes para a regulação gênica, dentre outras funções. Tudo o que uma célula é capaz de fazer está, de alguma forma, inserido no genoma. Logo, este é um passo importante no entendimento do funcionamento de qualquer organismo.

Até o momento foram sequenciados os genomas de nove espécies de *Trichoderma*, mas somente cinco foram publicados: *Trichoderma reesei* QM6a (Martinez et al, 2008), grande produtor de celulases industriais; *T. virens* e *T. atroviride* (Kubicek et al. 2011), espécies descritas como micoparasitas largamente utilizadas no controle biológico de vários fitopatógenos como *B. cinerea* e *R. solani* (Chaverri et al, 2001; Dodd et al, 2003); *T. gamsii*, potencial agente de biocontrole contra *Fusarium spp.* (Baroncelli et al., 2016); *T. parareesei* (Yang et al, 2015) um ancestral do *T. reesei* e um isolado de *T. hamatum* (Studholme et al, 2013) também envolvido no biocontrole. O *Joint Genome Institute* (JGI), através do seu projeto *Mycocosm* (Grigoriev et al, 2014), sequenciou 7 destas espécies, sendo estas 7 espécies utilizadas para análise comparativa nesta tese.

A comparação de genomas de *Trichoderma* foi feita por Kubicek e colaboradores (Kubicek et al, 2011) onde foram analisados os genomas de *T. reesei*, *T. atroviride* e *T. virens*. Neste trabalho os principais resultados encontrados foram que a espécie *T. atroviride* é ancestral em relação as outras

duas, e que houve uma expansão de proteínas como enzimas hidrolíticas e genes envolvidos no metabolismo secundário nas espécies micoparasitas mais efetivas (*T. virens* e *T. atroviride*).

Um ponto interessante de se notar, é que os genomas dessas três espécies foram sequenciados utilizando a técnica de sequenciamento de Sanger. Sequenciamento de genomas utilizando esta técnica são bastante custosos e demorados. Com o advento de técnicas *Next Generation Sequencing* (NGS), sendo os principais representantes atualmente – Illumina e Pacific Biosystems (PacBio) – o sequenciamento de genomas se tornou mais barato e mais acessível. A contrapartida deste benefício é que a montagem destes genomas se tornou mais complexa e mais suscetíveis a erros (Wommack et al., 2008). Os demais genomas de *Trichoderma* sequenciados pelo JGI e outros grupos de pesquisa utilizam a técnica de sequenciamento Illumina. Com este aumento na quantidade de genomas de espécies do gênero *Trichoderma*, este tipo de análise comparativa pode trazer novas ideias de como esses genomas evoluíram, quais famílias foram expandidas ou perdidas ao longo do tempo evolutivo de separação destas linhagens.

METODOLOGIA

Manutenção dos isolados

A cepa *T. harzianum* TR274, isolado de solo do cerrado, assim como o fungo fitopatogênico *S. sclerotiorum* (Lib.) de Bary, da coleção de fungos de solo mantida no Laboratório de Enzimologia da UFG, foram utilizados nos experimentos de interação e análise de transcrito e genoma estrutural. O isolado TR274 foi escolhido como modelo deste trabalho, pois já foi caracterizado anteriormente como antagonista do fungo fitopatogênico *S. sclerotiorum* (Lopes et al. 2012). Estes isolados foram mantidos

com repiques periódicos em meio MYG (0,5% de extrato de malte, 0,25% de extrato de levedura, 1% de glicose e 2% de ágar) durante os experimentos e estocado em solução de glicerol (50%) em ultrafreezer -80 °C.

Condições de cultivo para isolamento de DNA

O isolado *T. harzianum* TR274, utilizado neste trabalho como modelo na análise do transcrito da interação com o fungo fitopatogênico *S. sclerotiorum*, foi também selecionado para sequenciamento do seu genoma estrutural. Neste caso, conídios do isolado TR274 (10^7 esporos ml^{-1}) foram inoculados em 100 ml de meio MYG líquido, a suspensão foi então incubada por 48 horas a 28 °C, 180 rpm. Após 48 horas de cultivo o meio de cultura foi filtrado e o micélio recuperado para isolamento do DNA genômico.

Sequenciamento, montagem e anotação do genoma estrutural

Para a extração do DNA genômico de *T. harzianum* TR274 foi utilizado o kit DNAeasy *plant and fungi* mini kit (Qiagen®). Este procedimento garante um DNA de alta qualidade, necessários para a etapa posterior de sequenciamento Illumina. O DNA extraído foi enviado para a empresa Eurofins (Alabama, EUA) para sequenciamento. A estratégia utilizada foi: sequenciamento *paired ends*, uma *lane* inteira de Hiseq 2000 - 2x100 pb com um tamanho médio de inserto de 250 pb e uma corrida de Miseq – 2x250 pb com um tamanho médio de inserto de 600 pb. Com estes dados temos cobertura suficiente para as análises. Os dados de sequenciamento obtidos foram submetidos à análise de qualidade utilizando o programa FastQC versão 0.10.1 (Babraham Bioinformatics). Seguido de um refinamento utilizando o programa Trimmomatic versão 0.22 (Lohse et al., 2012). Foram retiradas

regiões de sequências com qualidade menor que 20 (escala PHRED) da extremidade 3', como também 12 bases da região 5' suspeitas de conter frequências anômalas dos 4 nucleotídeos (hexamer primer bias e restos de adaptadores).

Para critério de comparação, a estratégia utilizada para o sequenciamento do isolado de *T. harzianum* CBS 226.95 foi: sequenciamento *paired ends* Hiseq 2000 – 2 x 100 pb com um tamanho médio de inserto de 270 pb e outra biblioteca 2 x 100 pb com um tamanho médio de inserto de 4 kb.

A montagem do genoma foi feita utilizando o programa AllPathsLG, assim como todos os outros genomas de *Trichoderma* sequenciados por Illumina (Butler et al., 2008), utilizando uma cobertura máxima de 100X. Para comparação de montagens, também foi utilizado o programa MaSurCA (Zimin et al., 2013) e as montagens visualizadas pelo programa QUAST v3.2 (Gurevich et al., 2013). A anotação funcional foi feita utilizando o *JGI Annotation Pipeline* versão 1.8 (Grigoriev et al., 2011), construído especificamente para anotação de fungos e desenvolvido pelo *Genome Annotation Team* (GAT) do *Joint Genome Institute* (JGI). Inicialmente se marcam sequências repetitivas e de transposons utilizando os programas RepeatMasker (<http://www.repeatmasker.org/>) e RepeatScout (Price et al., 2005), utilizando como banco de dados tanto o RepBase quanto predições *de novo* de regiões repetitivas. Com este genoma “mascarado” para regiões repetitivas, o *pipeline* começa a predição de modelos gênicos. Essa predição acontece utilizando três diferentes abordagens: *ab initio*, baseado em homologia e baseado no transcrito.

Os preditores *ab initio* Fgenesh (Salamov & Solovyev, 2000) e GeneMark (Ter-Hovhannisyan et al., 2008) foram treinados com o genoma de interesse para o início da predição. Os preditores baseados em homologia Fgenesh+ (Salamov & Solovyev, 2000) e Genewise (Birney et al., 2004) foram alimentados com alinhamentos oriundos do BLASTX contra o banco não-redundante de proteínas de fungos do NCBI. Preditores desenvolvidos pelo próprio grupo baseados em sequências oriundas do transcrito também foram utilizados. Após todas as predições, um algoritmo dá um peso para cada

predição para assim eleger o melhor modelo gênico para cada *locus*. Usualmente modelos com suporte de transcrito tem mais peso que suporte somente por homologia, que por sua vez tem mais peso que modelos preditos somente *ab initio*.

Todas as proteínas preditas foram então funcionalmente anotadas utilizando os programas SignalP (Petersen et al., 2011), TMHMM (Krogh et al., 2001), InterProScan (Hunter et al., 2009), domínios PFAM (Punta et al., 2012) e categorizadas de acordo com os bancos de dados GO (Blake et al., 2013), KEGG (Kanehisa et al., 2012) e KOG (Koonin et al., 2004). O portal contendo todos os dados de sequenciamento e anotação estão presentes no portal <http://genome.jgi.doe.gov/Trihar1/>.

Comparação dos genomas in silico

Os dados de todos os genomas utilizados nas análises comparativas foram retirados do portal Mycocosm (Grigoriev et al., 2013).

Identificação de genes ortólogos no gênero *Trichoderma*

Para a identificação de genes ortólogos, utilizamos os dados de proteínas preditas presentes no portal <http://jgi.doe.gov/fungi> para cada espécie dentro do gênero *Trichoderma*. Os arquivos de proteínas foram agrupados em um único arquivo e logo depois feito uma comparação “todos x todos” utilizando o algoritmo blastp. Logo depois foram aplicadas duas metodologias: Um script de perl customizado para encontrar os Best-Blast-Hit (BBH) de cada proteína com pelo menos 50% de cobertura e $e\text{-value} \leq 10^{-5}$, logo depois a construção dos clusters de presença e ausência e preenchimento do diagrama de Venn (Figura 5). Para checagem desta metodologia, o mesmo resultado do blastp “todos x todos” foi utilizado como entrada para o programa orthoMCL utilizando as

configurações padrão. Os dados obtidos com as duas metodologias foram bem similares e escolhemos a primeira metodologia para construção do diagrama de Venn.

O diagrama de Venn foi construído baseado na quantidade de combinações do Triângulo de Pascal para 7 conjuntos. Assim é possível representar bidimensionalmente todas as possibilidades de combinações.

Construção da filogenia de *Trichoderma*

Para construção da árvore filogenética, foram concatenadas 200 sequências de proteínas ortólogas e de cópia única presentes em todas as espécies analisadas. O alinhamento foi feito utilizando o programa MAFFT (Kato et al., 2002), utilizando as configurações padrão, gerando 80965 posições de aminoácidos. O programa GBlocks (Castresana, 2000) foi utilizado para retirar regiões com alinhamentos ruins resultando em 438 blocos totalizando 76493 posições de aminoácidos. Uma árvore de Máxima-Verossimilhança foi inferida utilizando o programa RAxML (Stamatakis, 2006) com o arquivo gerado pelo GBlocks utilizando o modelo PROTGAMMAWAG, 100 *bootstraps*, e dois grupos externos: *Fusarium graminearum* e *Chaetomium globosum*.

Categorização das proteínas quanto à função

Além das anotações já descritas, foram feitas anotações mais específicas para CAZymes, proteases, genes envolvidos no metabolismo secundário, fatores de transcrição e transportadores. Para cada categoria, exceto CAZyme e metabolismo secundário, foram construídos bancos de dados específicos juntamente com um script escrito em perl para fazer a comparação e montagem das tabelas. A anotação CAZyme foi feita comparando com o banco de dados do www.cazy.org e curagem manual para cada família.

A anotação dos genes de metabolismo secundário foi feita usando uma versão modificada do programa SMURF (Khaldi et al., 2008) para detecção de *poliketide syntases* (PKS), *non-ribosomal peptide syntase* (NRPS) e híbridos de PKS/NRPS juntamente com seus genes vizinhos possivelmente com funções relacionadas a produção de metabólitos secundários.

Para melhor entender a contração e expansão destas famílias gênicas dentro do gênero *Trichoderma*, nós usamos o software CAFE (de Bie et al. 2006) com um nível de significância de 0,05 e valor de λ calculado baseado nas distancias da árvore utilizada. Ao final, utilizamos o modelo de Viterbi para correção dos p-valores de cada ramo. Os arquivos de entrada foram: árvore filogenética descrita no item 4.7.2 no formato Newick e a tabela com as contagens de cada família gênica para cada isolado de *Trichoderma* e os respectivos grupos externos.

Análise de componentes principais (PCA) utilizando dados filogenéticos

O PCA filogenético foi implementado usando a função *phyl.pca* do pacote R *phytools* (www.phytools.org/). Como arquivos de entrada, utilizamos a árvore filogenética do gênero *Trichoderma* com suas respectivas distâncias e a tabela de contagem de todas as famílias gênicas preditas de acordo com a descrição do item acima. Este algoritmo (*phyl.pca*) corrige a matriz de correlação/covariância, levando em conta a correlação/covariância devido a relação puramente filogenética entre as espécies (Revell, 2009).

Análise comparativa das quitinases (GH18) de *Trichoderma*

Para a análise desta classe de enzimas utilizamos as anotações manuais específicas para a família de glicosil hidrolases família 18 (GH18) de todos os *Trichodermas*. As sequências de aminoácidos foram utilizadas para fazer a reconstrução filogenética utilizando o programa Muscle (Edgar, 2004)

para alinhar as sequências com a função *-refine* ativada. Este alinhamento foi então utilizado como entrada para construção da árvore filogenética utilizando o programa FastTree (Price et al., 2009) com o modelo WAG para evolução dos aminoácidos.

RESULTADOS E DISCUSSÃO

Sequenciamento, montagem e anotação do genoma estrutural do *T. harzianum*

TR274

O isolado TR274 de *T. harzianum* foi selecionado para sequenciamento do seu genoma estrutural, por apresentar atividade antagônica contra o fitopatógeno *S. sclerotiorum*, um dos organismos modelo deste trabalho, e também por dispormos de dados de RNAseq para este isolado.

Análises de resequenciamento geram somente dados de SNPs (*Single Nucleotide Polymorphisms*) e variantes estruturais mais complexas. Sendo assim, para uma melhor comparação e entendimento do genoma do isolado TR274, foi realizada a montagem de novo do genoma utilizando duas abordagens: o programa AllPathsLG, que utiliza sequencias pequenas para montagem e *jumping-libraries* para juntar contigs e o programa MaSuRCA assembler (Zimin et al., 2013), que combina a eficiência dos grafos de Brujin e o Overlap-Layout-Consensus para montagem dos *scaffolds*. A montagem do genoma do isolado TR274 utilizando o programa AllPathsLG ficou mais fragmentada, no entanto, gerou um número menor de *misassemblies* quando comparado com o genoma de referência de *T. harzianum* CBS 226.94 (Figura 3). Por isto, optou-se por utilizar a

montagem gerada pelo programa AllPathsLG. Outro fator que contribuiu para a escolha pela montagem do programa AllPathsLG é que este é o software recomendado dentro do JGI para montagem de genomas de fungos gerados por sequenciamento Illumina. Estes genomas são naturalmente menos complexos que genomas de plantas e animais, por exemplo, enquanto o MaSuRCA é direcionado para montagem de genomas de plantas, o que pode ter gerado *scaffolds* errôneos, como mostrado na figura 3.

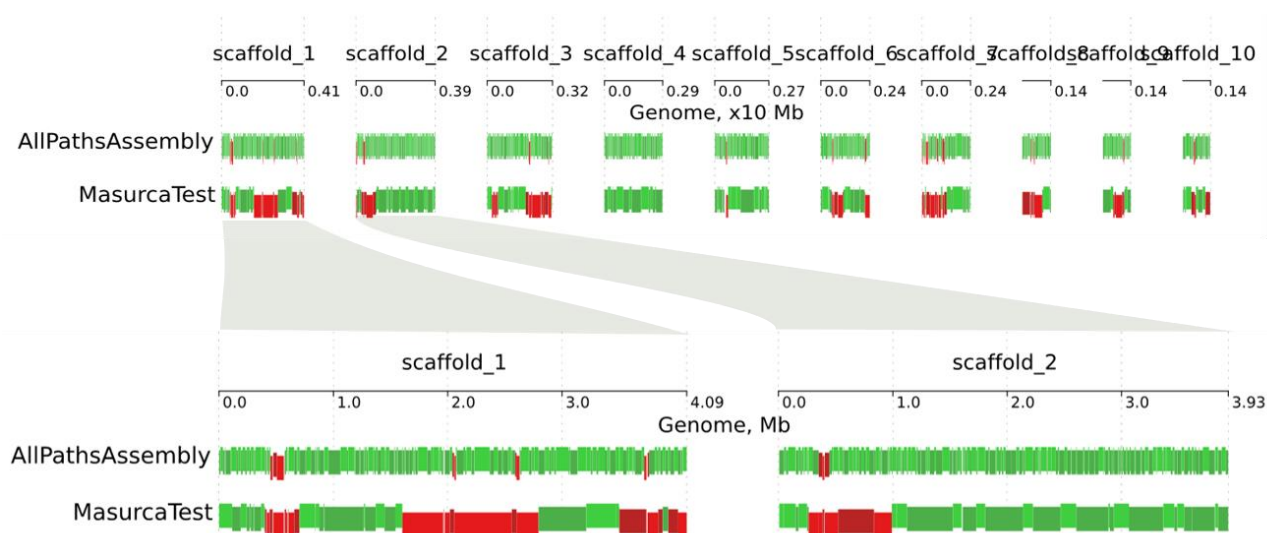


Figura 3. Comparação entre dois diferentes programas para montagem de genomas: AllPathsLG e MaSurCa. A comparação foi feita usando o genoma de referência de *T. harzianum* CBS 226.95. As linhas verdes são os alinhamentos com boa qualidade, mas com extremidades não correspondentes ao *scaffold* de referência e as linhas vermelhas são alinhamentos de baixa qualidade também extremidades não correspondentes. As diferenças de altura e tons representam diferentes *contigs*.

Para a anotação do genoma do isolado TR274, todos os passos descritos na metodologia foram seguidos para a anotação estrutural e funcional. Preditores que precisam de dados de transcrito foram alimentados de duas formas: dados crus oriundos de RNA-seq e transcrito montado *de novo*. Isto aumenta a confiabilidade das predições e também oferecem informações importantes como 5' e 3' UTR. É importante ressaltar que este pipeline não leva em consideração *splicing* alternativo de genes.

Com o *pipeline* finalizado, foi realizada a comparação das características estruturais e funcionais de todos os genomas de *Trichoderma* disponíveis para este tipo de análise (Tabela 1). Nesta tabela pode ser visto que não existem *gaps* na montagem do isolado TR274, enquanto nos outros existe sempre uma pequena fração de *gaps*, isso vem do fato que não foram utilizadas *jumping libraries* para montagem do isolado TR274. Este tipo de biblioteca ajuda na junção de diferentes *contigs* para geração dos *scaffolds*, mas muitas vezes faltam sequencias para preencher as regiões internas das *jumping libraries*, gerando essa ausência ou desconhecimento dos nucleotídeos em determinadas regiões do genoma, os *gaps*. Isso é refletido conseqüentemente na quantidade de *Scaffolds* presentes no isolado TR274, que é pelo menos quatro vezes maior que a dos outros isolados. Apesar dessa maior fragmentação, as variáveis como tamanho de genes e proteínas, tamanho de exons e introns e homologia com banco de dados, não foram afetadas significativamente, mostrando valores esperados para um genoma desta espécie.

Tabela 1. Comparação das características do genoma do isolado de *T. harzianum* TR274 com os genomas de outros isolados e espécies de *Trichoderma*.

Species	<i>T. asperellum</i>	<i>T. atroviride</i>	<i>T. harzianum</i>	<i>T. harzianum</i>	<i>T. virens</i>	<i>T. reesei</i>	<i>T. reesei</i>	<i>T. longibrachiatum</i>	<i>T. citrinoviride</i>
Strain	CBS 433.97	IMI 206040	CBS 226.95	TR274	Gv29-8	QM6a	RUTC-30	ATCC 18648	Trilo contaminant*
Designation	Trias	Triat	Triha	Trihar	Trivi	Trire	Triru	Trilo	Trici
Genome size	37,5	36,1	40,98	40,87	38,8	34,1	32,69	32,2	33,2
Coverage	120X	8.26x	120x	100x	8.05x	9x	47.6x	60	63,1
Assembly Gaps (Mbp)	0.2 (0.06%)	0.1 (0.16%)	0.04 (0.1%)	0	0.2 (0.4%)	0.05 (0.1%)	0.17 (0.5%)	1 (3.24%)	0.09 (0.2%)
Number of scaffolds	419	50	532	2282	135	89	182	130	533
Number of predicted genes	12586	11865	14095	13932	12518	9143	9852	10938	9737
Gene length (bp)	1651	1747,06	1628	1584	1710,05	1793,25	1696	1686	1702
Protein length (aa)	442	471,54	441	443	478,69	492,27	469	428	484
Exon per gene	2,68	2,93	2,7	2,8	2,98	3,06	2,88	2,71	3,1
Exon length (bp)	555	528,17	543	507	506,13	507,81	512	546	477
Intron length (bp)	98	104,2	96	94	104,95	119,64	119	121	107
Homology Support (NR)	10959 (87.07%)	10219 (92%)	12340 (87.55%)	12.308 (88.3%)	10915 (94%)	8409 (92%)	8779 (89.11%)	9056 (82.79%)	9388 (96.42%)
Homology Support (SwissProt)	8051 (63.97%)	8367 (75%)	8918 (63.27%)	8954 (64.27%)	8773 (75%)	6763 (74%)	7036 (71.42%)	6940 (63.45%)	7260 (74.56%)
PFAM domains	6991 (55.55%)	5883 (53%)	7818 (55.47%)	7847 (56.32%)	6267 (54%)	5096 (56%)	5976 (60.66%)	5771 (52.76%)	6084 (62.48%)

As características do genoma do isolado TR274 quando comparadas, especialmente com o genoma de *T. harzianum* CBS 226.95, pelo menos para os atributos analisados não apresentaram

diferenças notáveis, isso mostra que a qualidade da montagem e anotação estrutural e funcional em geral ficou comparável com os outros genomas e pronta para as próximas análises.

Genômica comparativa do gênero *Trichoderma*

Filogenia e conservação das proteínas

A árvore filogenética gerada pela comparação de 200 sequências de proteínas concatenadas e ortólogas e de cópia única presentes nos genomas de todas as espécies de *Trichoderma* analisadas apresentou valores de *bootstrap* máximos demonstrando confiabilidade da árvore (Figura 4A). Além disto, pode se ver a divisão entre as três seções *Trichoderma*, *Pachybasium* e *Longibrachiatum*. Como também o caráter ancestral da seção *Trichoderma*, já descrito por Kubicek e colaboradores em 2011. Segundo estes autores, o ancestral comum do gênero *Trichoderma* foi um micoparasita de basidiomicetos degradadores de madeira a adquiriu características saprofíticas ao seguir sua presa em seu substrato. Com esta mudança de habitat e conseqüente especialização a um ambiente pobre em nitrogênio (madeira em decomposição), foi descrita a transferência horizontal de uma enzima nitrato redutase (gene aparentemente perdido nas linhagens primitivas de *Sordariomicetos*) de *Basidiomicetos* para *T. reesei* (Slot & Hibbet 2007).

Baseados neste padrão de estilo de vida dentro do gênero *Trichoderma* (Kubicek et al., 2011), os ramos da árvore (Figura 4A) foram coloridos de forma diferente: vermelho para espécies com maior capacidade micoparasita e azul para espécies predominantemente saprófitas. Lembrando que nesta categorização não está implícito que organismos saprófitas não tem capacidade micoparasita e vice-

versa, eles somente têm uma maior predominância em um estilo de vida ou outro, mas podendo transitar entre eles. A linha tracejada horizontal na figura 4 está dividindo as duas estratégias. Existe uma discussão mais complexa na literatura a respeito de se dividir espécies baseadas em estilo de vida, especialmente espécies crípticas como o *T. harzianum*, onde existe uma variedade genética e metabólica grande (Druzhinina et al., 2010). Para uma melhor comparação e facilitar a análise fizemos esta divisão binária, mas não se pode generalizar estas análises para todos os isolados das espécies descritas neste trabalho.

A figura 4B mostra os tamanhos dos genomas para cada espécie e também a quantidade de sequências repetitivas. Estes genomas, como da maioria dos ascomicetos, não têm problema quanto às sequências repetitivas, o que atrapalha bastante a montagem e análise de genomas. Os genomas dos isolados da espécie *T. harzianum* apresentam uma maior quantidade de sequências repetitivas em números absolutos, devido a seu maior genoma. As porcentagens variaram de 2 a 5,3% do genoma. É importante ressaltar que os genomas de *T. reesei*, *T. atroviride* e *T. virens* foram sequenciados pela técnica de Sanger, portanto, a baixa porcentagem de sequências repetitivas não é resultado de artefatos da montagem dos genomas, observada quando se utiliza dados oriundos de sequenciamento Illumina.

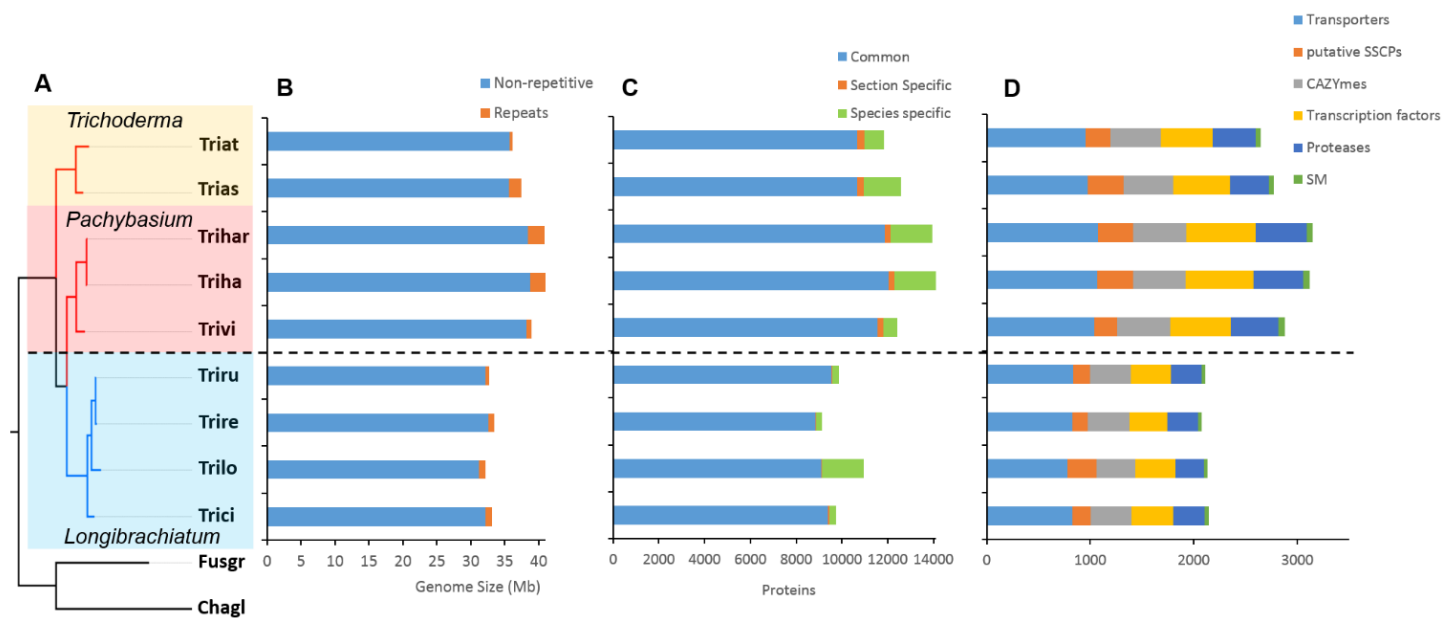


Figura 4. Filogenia e características dos genomas de espécies de *Trichoderma*. A – Árvore de máxima-verossimilhança de 9 isolados de *Trichoderma* (Triat: *T. atroviride*; Trias: *T. asperellum*; Trihar: *T. harzianum* TR274; Triha: *T. harzianum* CBS 226.95; Trivi: *T. virens*; Triru: *T. reesei* RUTC30; Trire: *T. reesei* QM6a; Trilo: *T. longibrachiatum*; Trici: *T. citrinoviride* e grupos externos: Fusgr: *Fusarium graminearum*; Chagl: *Chaetomium globosum*). Ramos vermelhos são espécies predominantemente micoparasitas e em azul, espécies predominantemente saprófitas. B – Proporção de regiões repetitivas nos genomas. C – Compartilhamento de proteínas entre as espécies de *Trichoderma*. D – Comparação das categorias de proteínas transportadoras (Transporters), proteínas pequenas e ricas em cisteínas (SSCP), enzimas envolvidas na degradação de polissacarídeos (CAZymes), fatores de transcrição, proteases e metabolismo secundário (SM).

A figura 4C mostra o quanto das proteínas preditas no genoma de cada um dos isolados é compartilhada entre as espécies de *Trichoderma* (azul), entre somente sua seção (laranja) e específicos daquele isolado (verde). Este dado mostra claramente que a maioria das proteínas presentes no gênero *Trichoderma* é compartilhada por todas as espécies. O número de proteínas específicas de cada espécie segue proporcionalmente o tamanho dos seus respectivos genomas, os maiores têm mais proteínas específicas e o contrário sendo observado para as espécies com genomas menores. Este padrão não é seguido somente quando se analisa o genoma de *T. longibrachiatum* (Trilo), que tem uma quantidade de 1803 genes. Isto pode ter sido causado devido as complicações associados à montagem deste

genoma. Como mostrado na tabela 1, os genomas de *T. longibrachiatum* e *T. citrinoviride* foram sequenciados como sendo de um isolado só, mas durante as análises se percebeu que na verdade se tratava de dois genomas cujo DNA foi isolado conjuntamente. Isso pode ter comprometido a montagem e anotação estrutural gerando genes falsos. Isso é corroborado pelo fato de que dentro dos 1803 genes exclusivos desta espécie, somente cinco tem domínio PFAM (Tabela 2) e, em geral, estas proteínas tem tamanhos entre 50 – 80 aminoácidos (o pipeline descarta modelos gênicos com menos de 150 pb) e ausência de introns. Logo temos duas hipóteses, realmente estas proteínas específicas são artefatos de anotação ou são proteínas com predição confiável, mas ainda não caracterizadas quanto à presença de domínios funcionais.

A figura 4D mostra a quantidade de proteínas classificadas como transportadores, *small secreted cystein proteins* (SSCP), CAZymes, fatores de transcrição, proteases e proteínas envolvidas com metabolismo secundário (SM) em cada um dos genomas analisados. Esta figura também mostra que as diferentes classes apresentam distribuição proporcionais. Isto pode ser explicado, como dito anteriormente, pelo fato de diferentes espécies de *Trichoderma* transitarem em diferentes estilos de vida, logo a composição gênica das grandes classes, não é o fator principal de diferenciação entre eles, mas outros fatores como quantidade de diferentes classes mais específicas de proteínas (serão abordados nos próximos tópicos) e regulação gênica.

Tabela 2. Total de proteínas ortólogas com anotação PFAM.

Total de proteínas com pelo menos um domínio PFAM							
	Trias1	Triat2	Triha1	Trivi2	Trici4	Trilo3	Trire2
Proteínas Comuns	6418 (77.8%)	6411 (77.7%)	6559 (79.6%)	6543 (79.4%)	6462 (78.4%)	6562 (79.6%)	6507 (78.9%)
Específica da espécie	27 (1.6%)	23 (2.7%)	47 (2.6%)	28 (4.6%)	5 (1.7%)	5 (0.3%)	5 (1.9%)
Específica da seção	36 (12.2%)	27 (9.1%)	51 (19.2%)	48 (18.2%)	5 (11.6%)	6 (13.9%)	4 (9.3%)
Comuns à <i>Trichoderma/Pachibasium</i>	91 (47.1%)	93 (48.2%)	84 (43.5%)	88 (45.6%)			

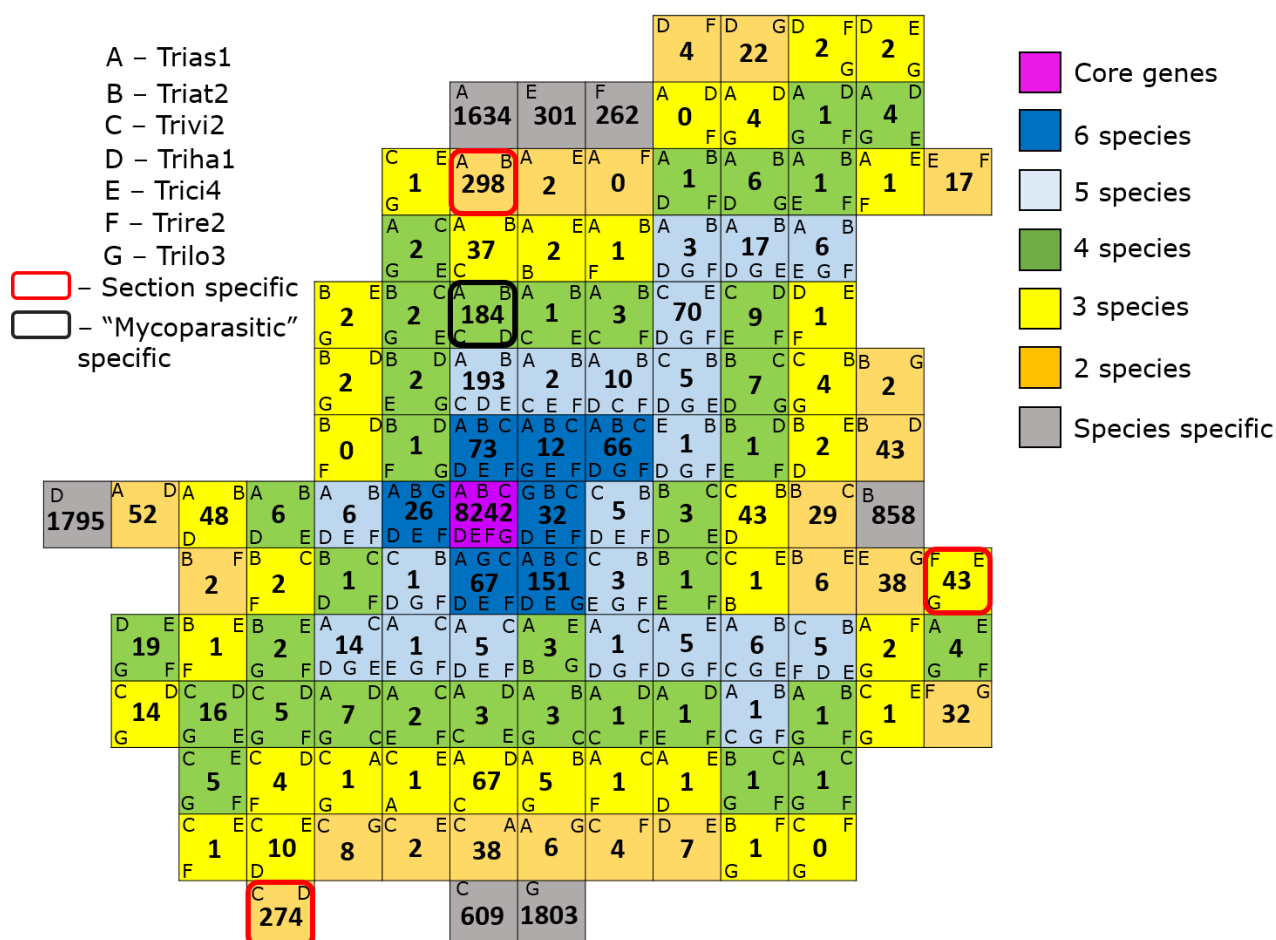


Figura 5. Diagrama de Venn com a distribuição de proteínas ortólogas de 7 diferentes espécies de *Trichoderma*. Nesta figura, não foram apresentadas isolados da mesma espécie. As cores representam o número de espécies comparadas e o número central a quantidade de proteínas compartilhadas entre as espécies de *Trichoderma*. Blocos circulado em vermelho correspondem a proteínas exclusivas das seções e em preto, proteínas compartilhadas entre seções *Trichoderma* e *Pachibasium*.

A figura 5 mostra um diagrama de Venn estruturado de forma que todas as intercessões são mostradas separadamente em blocos. Esta construção foi feita baseado na 7ª linha do triângulo de Pascal (não contando o primeiro 1 do topo) onde a combinação de 7 grupos é mostrada pela quantidade de blocos: 1 espécie: 7 blocos; 2 espécies: 21 blocos; 3 espécies: 35 blocos; 4 espécies: 35 blocos; 5 espécies: 21 blocos; 6 espécies: 7 blocos e 7 espécies: 1 bloco. Usando esta metodologia, podemos

construir diagramas de Venn com qualquer conjunto de dados, não restringindo a análise à no máximo 5 a 6 componentes que são utilizados visualmente.

A maioria das proteínas (8242 ou 46,3% do total de proteínas no diagrama) é comum a todos os isolados de *Trichoderma*, e um número bem menor foi detectado para as outras intercessões, de fato os blocos contendo de 2 a 6 espécies, o número de proteínas comuns ficou abaixo de 10. O que parte pode advir de a análise incluir combinações complexas de espécies de seções mais distantes, podendo, em alguns casos, ser artefatos do algoritmo *blast*. Um bom exemplo é a comparação entre uma espécie primitiva dentro do grupo, o *T. asperellum*, com uma derivada, o *T. reesei*. Neste caso, não foi encontrada nenhuma proteína compartilhada entre essas duas espécies que não esteja presente em outra. Desta forma é possível sugerir que espécies com estilo de vida similar, mas em seções diferentes compartilham mais proteínas, apesar da distância filogenética (Trias-Trivi: 38 proteínas; Trias-Triha: 52 proteínas; Triat-Trivi: 29 proteínas; Triat-Triha:43 proteínas). As proteínas únicas variaram de 262 em *T. reesei*, até 1803 em *T. longibrachiatum* (Figura 5).

A categorização destas proteínas quanto à quantidade de domínios PFAM mostra que à medida que se afasta do centro do diagrama de Venn, a presença desses domínios vai diminuindo. Nas proteínas centrais, a proporção varia de aproximadamente 78% de proteínas com domínio PFAM, enquanto na espécie específica, uma proporção máxima de 4,6% no caso de *T. virens* (Tabela 2). Este padrão é esperado, pois proteínas comuns a vários organismos são as mais conhecidas e estudadas, já as específicas são exatamente aquelas onde existe uma grande falta de informação, mas que podem revelar novas funções e direcionar novos estudos. Neste caso, é necessário passar estas sequências por um filtro para se certificar que não são artefatos de anotação, se realmente são expressas, se tem um produto proteico, dentre outros mecanismos.

Função e evolução de classes protéicas

Com relação à análise de famílias gênicas nos isolados de *Trichoderma* e verificação da expansão ou contração de famílias gênicas, pode-se afirmar que na seção *Pachibasium* ocorreram várias expansões de famílias gênicas desde a separação com a seção *Longibrachiatum* em todas as diferentes classes de proteínas. Destacando-se a classe de proteases S33 e fator de transcrição com motivo de dedo de zinco Zn_clus, que tiveram expansões em praticamente todos os ramos da seção *Pachibasium*. Nos ramos da seção *Longibrachiatum* se vê o contrário acontecendo, somente contrações de famílias gênicas exceto a família de fungal_trans em *T. reesei*. Esta família de fatores de transcrição é constituída por um fator de transcrição específico de fungos que possivelmente está envolvido com seu estilo de vida mais saprófita degradador de biomassa.

Na figura 6B, no entanto, não se observa uma quantidade significativa de contrações na seção *Longibrachiatum* quanto às enzimas hidrolíticas, somente expansões de quitinases (GH18) e seus respectivos domínios de ligação (CBM50 e CBM18) na seção *Pachibasium*. Este resultado também está relacionado com o estilo de vida mais micoparasita desta seção comparado às outras (Hartl et al., 2012). A expansão de famílias de genes do metabolismo secundário NRPS e PKS também correlaciona com um estilo de vida mais micoparasita, especialmente a espécie *T. virens*, que tem um dos maiores repertórios de metabolismo secundário dentro do reino *Fungi* (Kubicek et al., 2011). Um resultado interessante na figura 6B é que a seção *Pachibasium* e a espécie *T. harzianum* tiveram mais expansões que qualquer outra espécie. Sugerindo que a expansão do tamanho do genoma está relacionada com a expansão de famílias gênicas, logo, algum mecanismo de duplicação gênica aconteceu nessa linhagem (Magadum et al., 2013). Mais investigações a respeito dos mecanismos envolvidos na expansão destes genomas devem ser feitas.

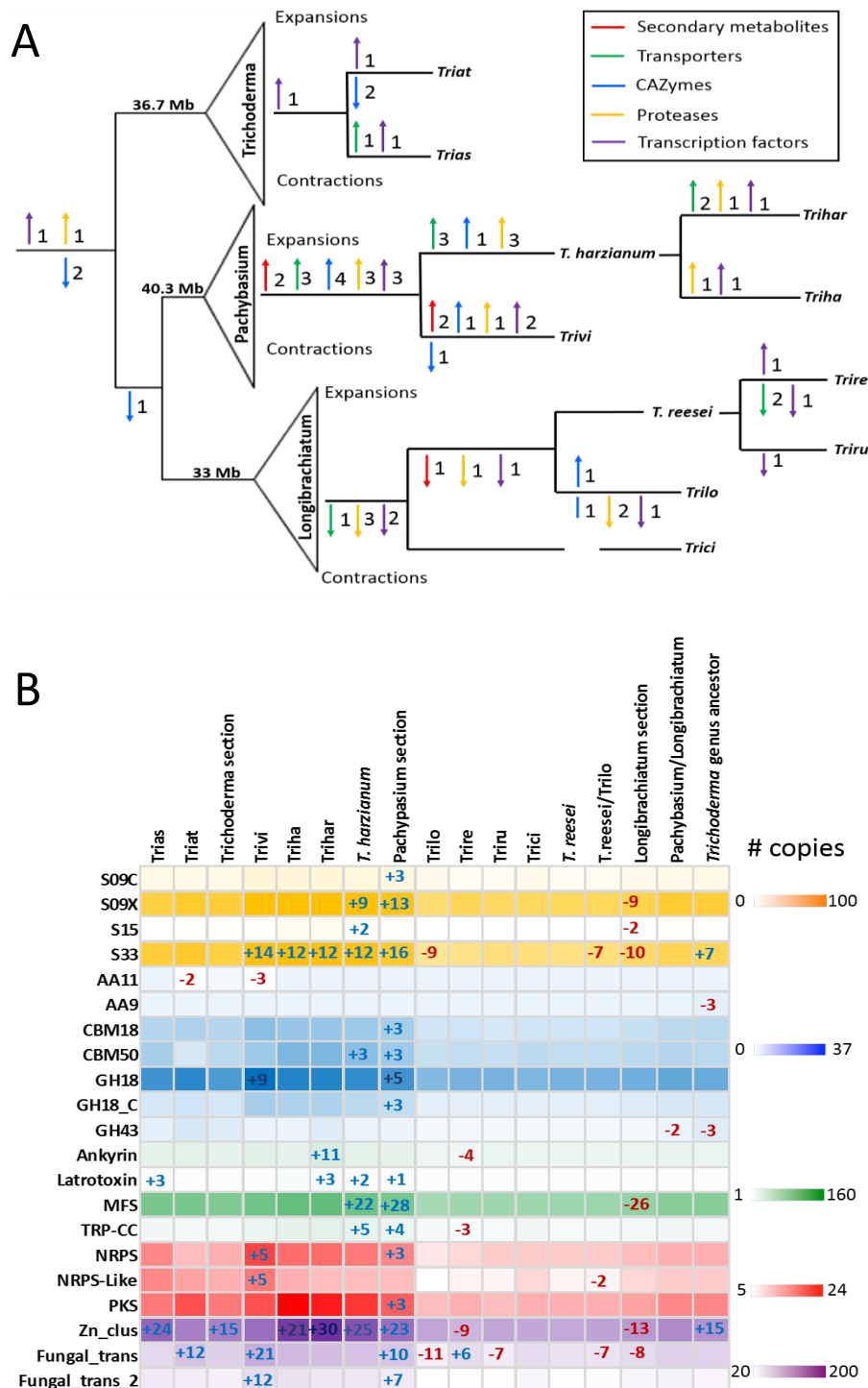


Figura 6. Evolução das famílias genicas em *Trichoderma*. A – Árvore filogenética juntamente com a quantidade de famílias gênicas que sofreram expansão (seta para cima) e contração (seta para baixo) em cada nó relativo a seu ancestral. Cada cor nas setas representam uma classe de proteínas. Os valores

representados nos nós basais de cada seção representam a média do tamanho dos genomas dos isolados constituintes. B – Famílias que sofreram expansão e contração durante a evolução do gênero *Trichoderma*. Cada linha representa uma família gênica e cada coluna representa um nó da árvore filogenética. O nível da cor de cada célula demonstra a quantidade de cópias gênicas predita para cada ancestral de acordo com o programa CAFE. Os números dentro das células mostram a quantidade de cópias que foram significativamente expandidas (azul) e contraídas (vermelho) com p-valor $\leq 0,05$.

A figura 7 mostra um PCA filogenético com a mesma contagem de famílias gênicas utilizadas na figura 6. Existe uma divisão entre estilos de vida saprófita e micoparásita nas duas componentes principais, especialmente na componente 1, onde está a maior porcentagem de variação. Esta figura mostra que apesar de uma distribuição proporcional das classes de proteínas nos genomas (Figura 5D), quando analisamos a quantidade de proteínas de famílias mais específicas, nota-se a sua importância na divisão dos estilos de vida. As seções *Pachibasium* (Triha, Trihar e Trivi) são mais próximas filogeneticamente da seção *Longibrachiatum* (Trici, Trilo, Trire, Triru), mas quando olhamos no ponto de vista da composição das famílias gênicas, estas seções estão, na verdade, se afastando (Figura 7). Este dado sugere que ao compararmos diferentes estilos de vida – neste caso poderíamos extrapolar para o fenótipo - a composição gênica no nível de famílias, é mais importante que análises filogenéticas, que olham somente a composição de nucleotídeos/aminoácidos e sua relação entre organismos.

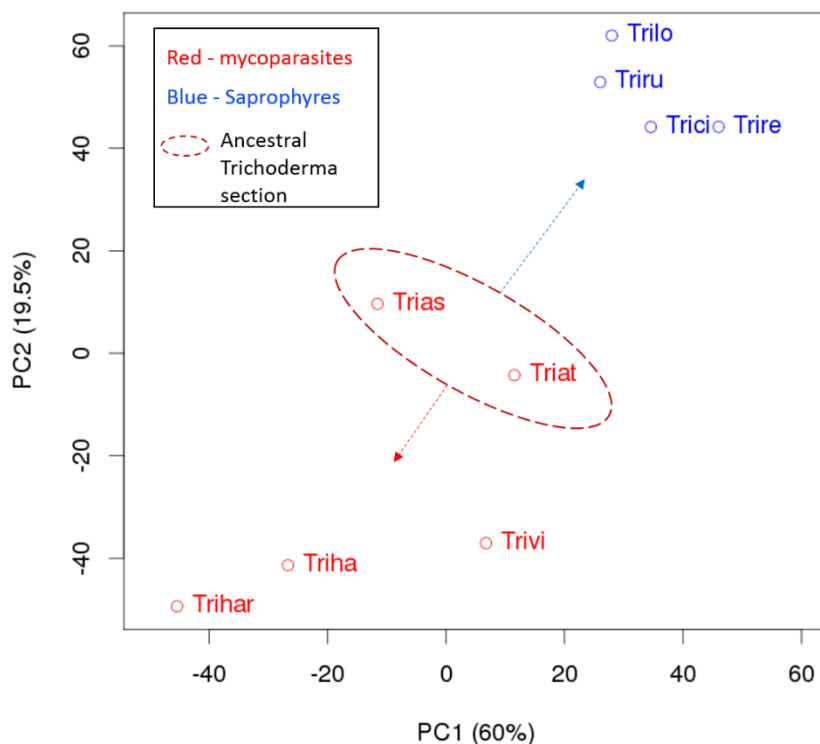


Figura 7. Análise de componentes principais (PCA) filogenético construídos utilizando a tabela de contagem de famílias gênicas e a árvore filogenética do gênero *Trichoderma*. Espécies predominantemente micoparasitas são mostradas em vermelho e predominantemente saprófitas em azul. O círculo vermelho mostra a seção ancestral do gênero *Trichoderma*.

Perda das GH18 subgrupo C1 pela seção *Longibrachiatum*

A família de enzimas GH18 que inclui as quitinases é conhecida como um dos mais importantes efetores envolvidos no micoparasitismo (Hartl et al., 2012). As quitinases são divididas em subgrupos de acordo com os domínios presentes na sua estrutura (figura 8). O subgrupo A inclui as quitinases com massa molecular entre 40 e 50 kDa e que não apresentam domínio de ligação a carboidrato, sendo o subgrupo de maior ocorrência em fungos filamentosos e mais bem estudado (Seidl, 2008). O subgrupo B é bem variável à massa molecular, 30 a 90 kDa, e domínios de ligação a carboidratos na região carboxi-terminal. Algumas quitinases, como a CHIA de *A. nidulans* tem um domínio de ancoragem-GPI, sugerindo uma função de remodelamento de parede celular (Takaya et al. 1998). No

gênero *Trichoderma* é bastante comum a presença de domínios de ligação a celulose nesta classe de enzimas, como o CBM1. Este domínio não é exclusivo de ligação à celulose, mas também a outros polissacarídeos como a quitina (Sedl et al. 2008). O subgrupo C é constituído por enzimas de massa molecular entre 140 e 170 kDa com uma maior distinção dos outros subgrupos contendo domínios de ligação à quitina (CBM18) e domínios LysM. Estes domínios LysM são envolvidos com a ligação à peptideoglicanos de parede celular. Algumas quinases de plantas contem este domínio e sugere-se que esteja relacionado ao reconhecimento de bactérias simbiotes e resposta a fungos patogênicos (Buist et al. 2008).

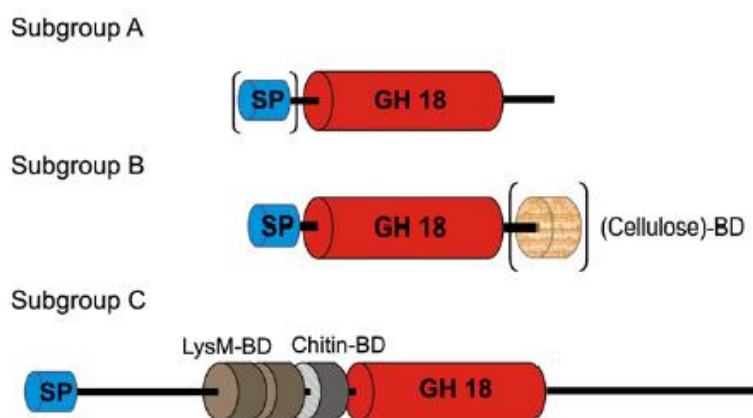


Figura 8. Organização dos domínios de estruturais e funcioanis dos subgrupos de quitinases. SP – Peptídeo Sinal. Domínio conservado típico da família de glicosil hidrolase GH18. BD – Domínio de Ligação. Retirado de Seidl, 2008.

Para analisar a distribuição das quitinases no gênero *Trichoderma*, utilizamos todas as sequencias de proteínas preditas com a presença do domínio catalítico GH18, incluindo uma proteína de cada subgrupo de *N. crassa*. A figura 9 mostra esta árvore. Percebe-se que os subgrupos foram separados corretamente. Em trabalhos anteriores, as árvores eram construídas separadamente para cada subgrupo, neste caso escolhemos reconstruí-la utilizando todas as sequencias juntas para se analisar melhor a relação entre os grupos. De fato, os subgrupos C1 e C2 não são tão próximos como se imaginava, apesar de ambas terem o domínio de ligação CBM18. O subgrupo C1 está mais próximo

filogeneticamente do subgrupo A, onde estão localizadas quitinases clássicas relacionadas ao micoparasitismo, como a chit33, chit 37 e chit 42 (Steindorff et al., 2012; Ramada et al., 2015), enquanto o subgrupo C2 está mais próximo do subgrupo B. Uma característica interessante deste subgrupo é a presença de um domínio conhecido por se ligar à celulose (CBM1), mas que já foi descrita sua capacidade de se ligar a uma diversidade grande de polissacarídeos, incluindo quitina. Algumas enzimas do subgrupo B também tem a capacidade de se ligar a membrana plasmática, através de um sinal de ancoragem-GPI (Seidl, 2008), sugerindo que este subgrupo tem funções mais complexas que simplesmente a degradação de quitina extracelular.

Na figura 9, nota-se um ramo acima do subgrupo B5 que não pertence a nenhuma categoria descrita até aqui. Estas são as proteínas com atividade de glicoproteína mannosil endo-N-acetil- β -D-glucosaminidase (ENGases). Estas proteínas são classificadas na família das GH18 e participam de processos como deglicosilação de proteínas intracelulares e hidrolise e oligossacarídeos oriundos de glicoproteínas externas (Dubey et al., 2012).

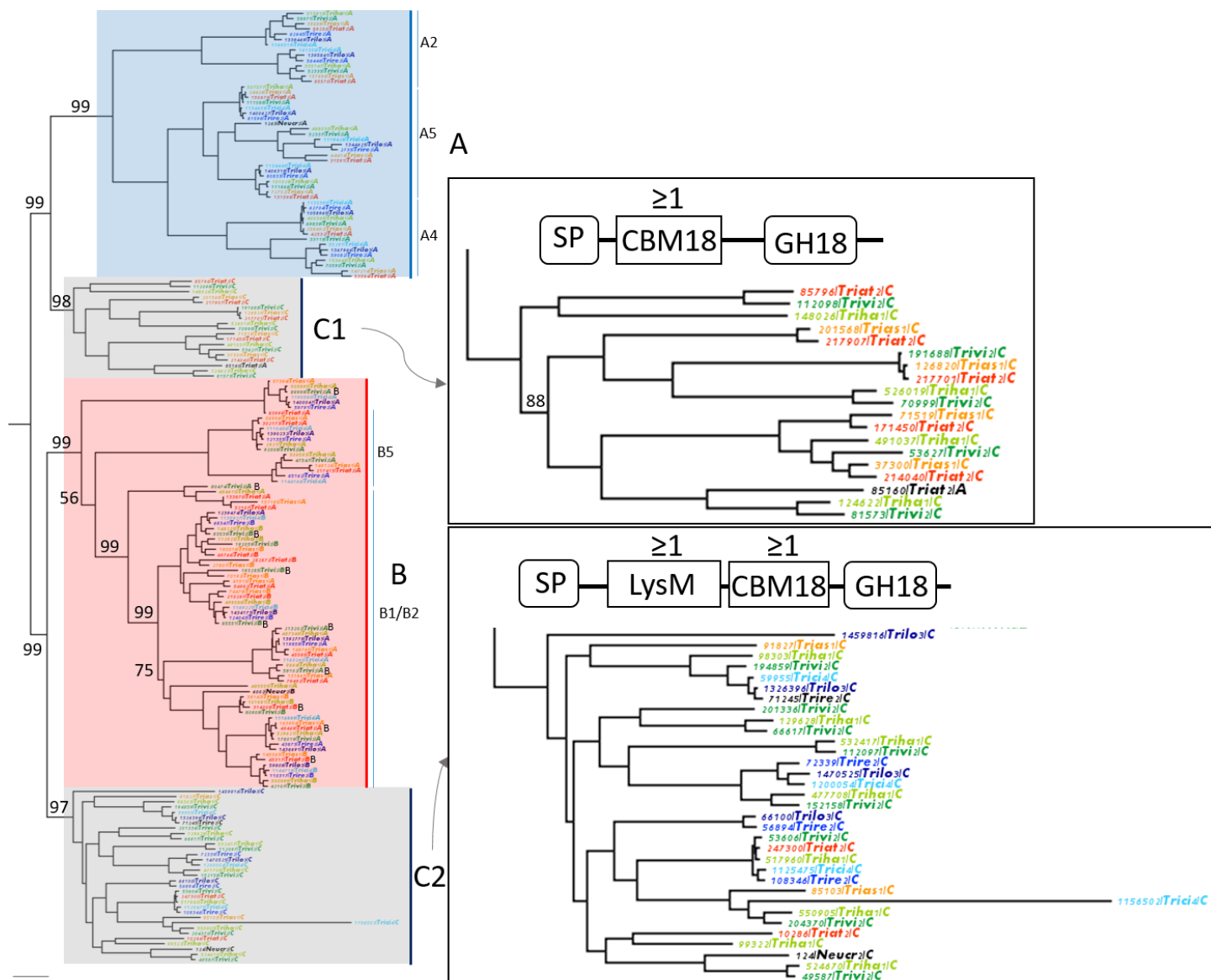


Figura 9. Árvore filogenética de máxima verossimilhança de quitinases do gênero *Trichoderma*. Subgrupos C1 e C2 enfatizados juntamente com sua estrutura de domínios. Cada cor nos nomes representa uma espécie diferente. Nós onde os bootstraps não são mostrados apresentam valores maiores que 90.

O dado mais interessante dessa árvore é a ausência de espécies da seção *Longibrachiatum* subgrupo C1 de quitinases. Na análise comparativa dos genomas de *T. atroviride*, *T. virens* e *T. reesei* (Kubicek et al. 2011) notou-se esta ausência, mas como o número de espécies era pequeno, não havia suporte para sustentar uma perda deste subgrupo na seção *Longibrachiatum*. Nesta análise foi

confirmado que realmente este subgrupo foi perdido nas três espécies analisadas. Este dado sugere que este subgrupo especificamente, pode ter um papel mais importante do que se imagina na preferência pelo estilo de vida micoparasita, diminuindo a “eficiência micoparasita” de fungos desta seção. A relação filogenética mais próxima entre os subgrupos C1 e A também dão suporte à esta hipótese. Nota-se também que todas as espécies têm quitinases do subgrupo C2, que contem domínios de ligação CBM18 e LysM. Isso sugere que estas são enzimas que fazem parte do arcabouço micoparasita deste gênero. Podendo assim ter uma função mais de proteção contra o ataque de outros microrganismos, conferindo maior vantagem competitiva durante o consumo de matéria orgânica, do que ataque direto à antagônicas.

CONCLUSÕES

Neste estudo vimos que o genoma de *T. harzianum* TR274 não apresentou diferenças significativas no seu genoma quando comparado com a cepa CBS 226.95 já sequenciada. Quando comparadas todos os genomas disponíveis de *Trichoderma*, vemos uma expansão no tamanho do genoma nas espécies da seção *Pachibasium* e uma contração nas espécies da seção *Longibrachiatum*. Este mesmo padrão é visto na quantidade de genes/proteínas destes isolados. Estes dados sugerem uma correlação da contração dos genomas da espécie *Longibrachiarum* com a diminuição da sua efetividade, ou simplesmente preferência, pelo estilo de vida micoparasita. A baixa proporção de domínios PFAM nas proteínas espécie específica sugerem que podemos ter a geração de artefatos durante a predição, ou que nos falta conhecimento sobre as funções de várias proteínas, especialmente de baixa massa molecular.

A análise da evolução das famílias genicas corrobora estas expansões e contrações ao nível de famílias gênicas. Famílias de todas as classes analisadas (transportadores, proteases, Fatores de transcrição, CAZymes e metabolismo secundário), foram expandidas na seção *Pachibasium* e contraídas na seção *Longibrachiatum*, com destaque para as quitinases. Foi demonstrado que o subgrupo C1 desta família foi perdido provavelmente no ancestral comum da seção *Longibrachiatum* e expandidas na seção *Pachibasium*.

Estas análises dão suporte para futuros trabalhos de isolamentos de *Trichoderma* em campo, onde fungos do gênero *Trichoderma* pertencentes à seção *Pachibasium* seriam os mais interessantes no ponto de vista do potencial de biocontrole ou produção de moléculas bioativas.

CAPÍTULO 2: ANÁLISE DO TRANSCRITOMA **DE *T. harzianum* CRESCIDO NA PRESENÇA DE** **PAREDE CELULAR DE *S. sclerotiorum* (artigo** **publicado BMC Genomics)**

INTRODUÇÃO

Sclerotinia sclerotiorum (Lib.) de Bary é um dos mais devastadores e cosmopolitas patógenos de plantas. Este fungo infecta mais de 400 espécies no mundo, várias delas com importância econômica (Bolland e Hall, 1994). *S. sclerotiorum* causa doenças em culturas como girassol, soja, feijão, lentilha, amendoim, cebola e tulipas (Attanayake *et al.*, 2013), além de ser capaz de infectar flores, folhas, frutas e caules (Attanayake *et al.*, 2013). Seu ciclo de vida se inicia pela germinação do escleródio e formação de propágulos infecciosos. Durante o crescimento da lavoura, que depende de vários fatores ambientais, o escleródio germina e forma o micélio, que pode infectar as plantas diretamente ou produzir ascósporos. Logo após, estes ascósporos se desenvolvem e formam os apotécios (Sun e Yang, 2000). Os ascósporos são os propágulos infectivos primários de *S. sclerotiorum* em várias culturas, podendo funcionar como mecanismo de dispersão para outras culturas da mesma região (Bolland e Hall, 1994).

O uso de pesticidas químicos é a principal estratégia usada para controlar doenças causadas por fungos. Iniciativas para desenvolvimento de alternativas sustentáveis para o controle do mofo branco e que não causem impactos econômicos e ambientais na produção, como observados no uso de fungicidas químicos, foram isoladas cepas de *Trichoderma* com potencial capacidade de biocontrole de *S. sclerotiorum*. Este trabalho inclui avaliação da capacidade antagonista, produção de enzimas

degradadoras de parede celular e produção de metabólitos secundários (Lopes *et al.*, 2012). Nossos resultados identificaram o *T. harzianum strito sensu* TR274 como um promissor agente de controle biológico contra *S.sclerotiorum in vitro* e sob condições de campo (Geraldine *et al.*, 2013).

O controle biológico é um processo complexo que requer o reconhecimento do hospedeiro pelo agente de biocontrole, seguido pela produção de enzimas hidrolíticas e produção de antibióticos que é ativada pelo contato da hifa do micoparasita com o fungo hospedeiro. Este contato é mediado por lecitinas, ativando vias de sinalização compostas por proteínas G e MAPK, modulando os padrões de expressão de genes de *Trichoderma* (Lorito *et al.*, 2010). No entanto, o mecanismo molecular detalhado envolvido neste processo ainda não foi elucidado. Este complexo mecanismo é influenciado pelo patógeno e também a cepa de *Trichoderma* avaliados (Druzhinina *et al.*, 2011). Neste contexto, estudos conduzidos com diferentes isolados são necessários para o entendimento dos mecanismos envolvidos no biocontrole.

O sequenciamento de bibliotecas de *expressed-sequence-tag* (EST) em diferentes isolados de *Trichoderma* cultivados em presença de fungos fitopatógenos tem contribuído significativamente para a identificação em larga escala de genes envolvidos no micoparasitismo (Vizcaíno *et al.*, 2007; Seidl *et al.*, 2009; Steindorff *et al.*, 2012). Microarranjos de DNA tem sido utilizado para avaliar a interação de *Trichoderma* com plantas (Rubio *et al.*, 2012), somente dois estudos utilizaram abordagens de alto desempenho para investigar os mecanismos de micoparasitismo de *Trichoderma* (Reithner *et al.*, 2011; Samolski *et al.*, 2009). O sequenciamento de RNA (RNA-Seq), uma Tecnologia de alto desempenho usado para sequenciar cDNA, tem sido bastante utilizada como uma ferramenta revolucionária em transcritomas (Ozsolak e Milos, 2011). O genoma completo e disponível ao público de *T. harzianum* CBS 226.95 (Grigoriev *et al.*, 2012), que foi liberado recentemente pelo *Joint Genome Institute* (JGI) (<http://genome.jgi.doe.gov/Triha1/>), nos permitiu fazer análise de mapeamento de sequencias de transcrito neste genoma de referência. Isso nos permitiu identificar

genes envolvidos no micoparasitismo, assim como entender melhor os mecanismos moleculares envolvidos nesta interação.

RESEARCH ARTICLE

Open Access

Identification of mycoparasitism-related genes against the phytopathogen *Sclerotinia sclerotiorum* through transcriptome and expression profile analysis in *Trichoderma harzianum*

Andrei Stecca Steindorff¹, Marcelo Henrique Soller Ramada², Alexandre Siqueira Guedes Coelho³, Robert Neil Gerard Miller¹, Georgios Joannis Pappas Júnior¹, Cirano José Ulhoa^{4*} and Eliane Ferreira Noronha¹

Abstract

Background: The species of *T. harzianum* are well known for their biocontrol activity against plant pathogens. However, few studies have been conducted to further our understanding of its role as a biological control agent against *S. sclerotiorum*, a pathogen involved in several crop diseases around the world. In this study, we have used RNA-seq and quantitative real-time PCR (RT-qPCR) techniques in order to explore changes in *T. harzianum* gene expression during growth on cell wall of *S. sclerotiorum* (SSCW) or glucose. RT-qPCR was also used to examine genes potentially involved in biocontrol, during confrontation between *T. harzianum* and *S. sclerotiorum*.

Results: Data obtained from six RNA-seq libraries were aligned onto the *T. harzianum* CBS 226.95 reference genome and compared after annotation using the Blast2GO suite. A total of 297 differentially expressed genes were found in mycelia grown for 12, 24 and 36 h under the two different conditions: supplemented with glucose or SSCW. Functional annotation of these genes identified diverse biological processes and molecular functions required during *T. harzianum* growth on SSCW or glucose. We identified various genes of biotechnological value encoding proteins with functions such as transporters, hydrolytic activity, adherence, appressorium development and pathogenesis. To validate the expression profile, RT-qPCR was performed using 20 randomly chosen genes. RT-qPCR expression profiles were in complete agreement with the RNA-Seq data for 17 of the genes evaluated. The other three showed differences at one or two growth times. During the confrontation assay, some genes were up-regulated during and after contact, as shown in the presence of SSCW which is commonly used as a model to mimic this interaction.

Conclusions: The present study is the first initiative to use RNA-seq for identification of differentially expressed genes in *T. harzianum* strain TR274, in response to the phytopathogenic fungus *S. sclerotiorum*. It provides insights into the mechanisms of gene expression involved in mycoparasitism of *T. harzianum* against *S. sclerotiorum*. The RNA-seq data presented will facilitate improvement of the annotation of gene models in the draft *T. harzianum* genome and provide important information regarding the transcriptome during this interaction.

Keywords: *T. harzianum*, *S. sclerotiorum*, RNA-seq, Gene expression, Mycoparasitism

* Correspondence: ulhoa@icb.ufg.br

⁴Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Goiás, Campus Samambaia, Instituto de Ciências Biológicas, CEP 74.090-900 Goiânia, GO, Brazil

Full list of author information is available at the end of the article

Background

Sclerotinia sclerotiorum (Lib.) de Bary is one of the most devastating and cosmopolitan plant pathogens. This fungus infects over 400 species of plants worldwide including important crops and numerous weeds [1]. *S. sclerotiorum* poses a threat to crops such as sunflower, soybean, oilseed rape, edible dry bean, chickpea, peanut, dry pea, lentils, onion and tulip [1]. It is capable of infecting flowers, leaves, fruits or stems [2] and its life cycle initiates by the germination of sclerotia and formation of infectious propagules. During the crop growing season, which depends on a set of environmental factors, fungal sclerotia germinate to form mycelia, which can directly infect host plants, or produce ascospores. Further, ascospores develop forming apothecia [3]. Ascospores are the primary infective propagules of *S. sclerotiorum* on many crops, and also can act in disease scattering since they may be transported to neighboring fields as well as over longer distances [1].

Chemical treatment is today the main strategy employed worldwide for fungal disease control. In order to develop alternative and sustainable methods for control of white mold, which do not cause negative environmental or economic impacts during crop production, as observed with the routine use of fungicides, our research group has isolated strains of a number of *Trichoderma* species from diverse agro-ecosystems in Brazil and assessed their potential for biocontrol of *S. sclerotiorum*. This analysis has included evaluation of antagonistic capacity, production of cell wall-degrading enzymes and production of volatile antibiotics [4]. Our results have identified *T. harzianum* Rifai (anamorph) strain TR274 as a promising biocontrol agent against *S. sclerotiorum* *in vitro* and under field conditions [5,6].

Biological control is a complex process which requires the host to be recognised by the biocontrol agent, followed by hydrolytic enzyme and antibiotic production which is triggered by the direct attachment of the mycoparasite to the host fungus. This contact is mediated by lectins and proteins harboring cellulose binding modules from hyphae of the host and mycoparasitic fungus, respectively, thereby eliciting a signaling cascade comprising G-proteins and MAPKs that can modulate *Trichoderma*'s protein expression pattern [7,8]. However, the detailed molecular mechanisms involved in this process remain unknown. This complex mechanism is influenced by the pathogen and *Trichoderma* isolates evaluated [9]. In this context, studies conducted on different strains are necessary for increased understanding of the biocontrol mechanism.

The sequencing of expressed-sequence-tag (EST) libraries for different *Trichoderma* strains cultivated in the presence of host fungi has contributed significantly to the large-scale identification of mycoparasitism-related genes [10-12]. Our research group has described gene mapping using EST and suppression subtractive hybridization (SSH) approaches

[12,13] during the interaction of *T. harzianum* with *Fusarium solani* and proteomic approaches for *T. harzianum* grown in liquid containing *Rhizoctonia solani*, *Macrophomina phaseolina* and *Fusarium sp* cell walls [14]. Whilst DNA microarrays have been used to study the interaction among *Trichoderma* and host plants [15], only two studies have employed high-throughput transcriptomic approaches to investigate mycoparasitism mechanisms of *Trichoderma* [16,17]. RNA sequencing (RNA-Seq), a high-throughput technology used to sequence complementary DNA, has been widely thought of as a revolutionary tool for transcriptomics [18-21]. The publically available whole genome sequence for *T. harzianum* CBS 226.95 [22], which was recently released by the Joint Genome Institute (JGI) (<http://genome.jgi.doe.gov/Triha1/Triha1.home.html>), now allows for use of RNA-Seq approaches and mapping of data to the reference sequence, which will likely contribute to identification of mycoparasite-related genes, as well as the molecular mechanisms by which this fungus is able to inhibit phytopathogen fungal growth.

The present study is the first initiative to use RNA-seq for identification of differentially expressed genes in *T. harzianum* strain TR274, in response to the phytopathogenic fungus *S. sclerotiorum*. *T. harzianum* TR274 was cultivated on liquid medium containing *S. sclerotiorum* cell wall (SSCW) to mimic fungal host presence. Quantitative real-time PCR (RT-qPCR) supported *in silico*-based evidence for differential gene expression of candidate genes involved in mycoparasitism.

Results and discussion

Illumina sequencing and mapping onto the *T. harzianum* reference genome

In this present study, an RNA-seq approach was used to map genes differentially expressed during *T. harzianum* growth on SSCW. The experimental design enabled comparison of gene expression in the presence of host cell wall or glucose as sole carbon source. Samples of mRNA from *T. harzianum* following three growth periods in the presence of SSCW (12, 24 and 36 h) were used to construct six Illumina libraries. A total of 171,442,148 sequence reads were obtained after quality trimming, varying from 25 to 100 bp in length. Each library was represented by at least 16,845,349 reads, representing a coverage of 66X when compared with the full transcriptome, a density regarded as adequate to perform gene expression analysis [23]. Complete mapping information can be accessed in Additional file 1: Table S8.

Sequence reads were aligned onto the *T. harzianum* CBS 226.95 reference genome [22]. This strain was isolated from garden soil in the UK, while strain TR274 was isolated from Brazilian cerrado soil. Despite the genetic differences among *T. harzianum* isolates described in the literature [24,4] and the fact that the reference genome

published is only a first draft, $82.6\% \pm 10.08\%$ of the obtained reads were mapped onto the reference genome using the default settings of the Bowtie2 aligner. The high mapping percentage suggests a high similarity between these isolates, at least at the transcriptome level. Only $0.8\% \pm 0.15\%$ of reads for each library was mapped in more than one region on the reference genome and these reads were not used in the expression analysis.

Gene expression analysis using “in silico” approach

The present work presents a first draft version of *T. harzianum* CBS 226.95 [22] using the RNA-seq approach, and provided a total of 14095 predicted genes. From these, the cuffdiff software analysis detected a total of 297 differentially expressed genes in the presence of SSCW in comparison to glucose as carbon source, as showed by the Venn diagram (Figure 1A). Remarkably, differences in gene distribution pattern were detected when the three induction times were compared. Data suggest a similar expressed gene set distribution between 24 and 36 hours, with most changes detected between 12 hours and 24 hours after *Trichoderma* growth in the presence of SSCW. This same expression pattern was confirmed by gene regulatory network analysis (GNR) (Figure 1B). The main modulation pattern was shared between 12 and 24 hours and a clear inversion between 12 and 36 h. The main genes with different modulation after 12 and 36 h belong to CAZymes and transporters (Additional file 2: Table S1 and Additional file 3: Table S4). This suggests that some proteins required during initial phases of cell wall degradation are not necessary in late times such as 36 hours.

The most notable increase in the gene expression was observed between 12 and 24 hours with a noteworthy change in expression pattern and protein functions detected between 12 and 36 hours. The main genes up or downregulated after 12 and 36 h of growth encoded CAZymes and transporters (Additional file 2: Table S1 and Additional file 3: Table S4). This suggests that some proteins required during early growth phases which are involved in cell wall degradation and sugar transport are no longer necessary after 36 hours of growth.

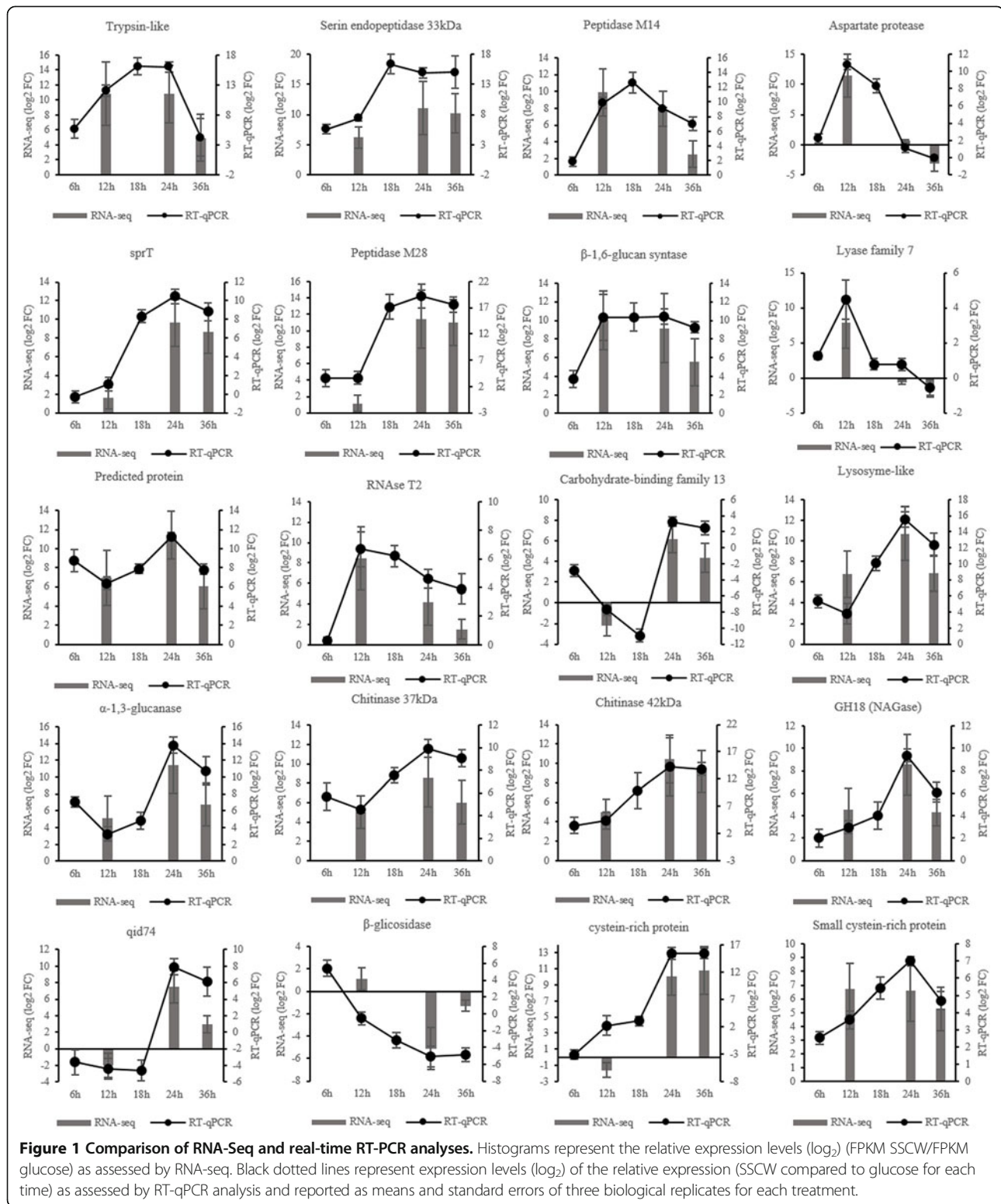
In order to further evaluate the time course expression profile, the top 10 differentially expressed genes after 12, 24 and 36 hours were identified (Table 1). The top 10 up-regulated genes at 12 h included a chitinase *chi18-17* and a GH25, four proteases, an isotrichodermin c-15 hydroxylase and three proteins of unknown function (Table 1). The top 10 up-regulated genes at 24 h included two peptidases, one Carbohydrate Esterase Family 5 (cutinase), two GH (P1 and α -1,3-glucanase), one transcription factor (srcap-like), two conidiation related and one unknown protein (Table 1). The top 10 up-regulated genes at 36 h included three proteases, a GH76, a MFS transport protein, two proteins involved in cell adhesion (hydrophobin 1 and

fasciclin domain protein), an alcohol oxidase, a protease inhibitor kazal and a CBM3 protein. It is interesting to note that there are more proteases/peptidases across the three growth times than GHs in the top 10 up-regulated genes. The combination of proteases and GHs seems to be preferential in mycoparasitism-related conditions [12-14]. Other genes also within the top 10 genes are involved in accessory functions like cell adhesion, antibiotic biosynthesis, conidiation and transport, complementing the main degradation activity.

To broadly compare gene expression patterns between growth periods, functional categories were assigned to the differentially expressed genes according to Gene Ontology (GO) guidelines [25] using Blast2GO [26]. Interproscan tool was used to improve the Gene Ontology annotations (Additional file 4: Table S2). To enrich the category analysis for up and down regulated genes at each growth time point, an exact Fisher test ($p < 0.05$) was performed (Figure 2). Data showed a clear up-regulation of transcripts categorized as involved in primary metabolic processes and presenting hydrolase activity, such as chitinases, β -1,3-glucanases and proteases [10,12,13]. This sort of expression pattern is expected, since the host fungus *S. sclerotiorum* cell wall is composed basically of proteins, chitin and β -1,3/1,6/ α -1,3glucans [27] and presents the first barrier to be overcome by the mycoparasite to achieve host cell invasion. On the other hand, *Trichoderma* can also use the fungal cell wall as carbon and nitrogen sources, and therefore has to increase expression level of hydrolytic enzymes (chitinases, mutanases, β -1,3-glucanases and proteases) and primary metabolism encoding genes to degrade and metabolize the host cell wall.

The down-regulated transcripts for all the stages of growth were categorized into oxidoreductase activity, oxidation-reduction process and some “binding” child categories, such as small molecule binding and nucleotide binding proteins. A hypothesis for repression could be the nature of basal metabolism of the genes belonging to these categories. This fact is consistent with the extensive metabolic activity expected for a filamentous fungus growing on a rich medium (glucose 2% medium) with an easily assimilable substrate [17]. Under this culture condition up-regulation of a specific subset of oxidoreductases and nucleotide binding proteins encoding genes related to primary metabolism is commonly observed for *Trichoderma* species, which are down-regulated in the presence of complex substrates [13]. Vieira *et al.* in 2013 showed the same pattern of repressed categories when *T. harzianum* was grown on *Fusarium solani* cell wall, suggesting that this result is not pathogen-dependent.

Finally, the expression profiles of the differentially expressed genes were determined by cluster analysis based on the SOTA method using Pearson’s correlation distance. These genes were divided into five groups based on their



expression modulation over time (Figure 3). Cluster 1 contains genes up-regulated after 12 h growth and down-regulated after 24 and 36 h, cluster 2 contains genes

up-regulated after 12 and 24 h and down-regulated after 36 h, cluster 3 contains genes up-regulated during the whole time course, cluster 4 contains genes down-regulated

Table 1 log₂ Fold Change (FC) of the top 10 differentially expressed genes after 12, 24 and 36 h

JGI ID	Putative function	Top 10	12 h	24 h	36 h
86893	Aspartic protease		11.4208	0.62564	-3.18003
91534	Acid proteinase protein		11.1366	0.04225	-7.00264
526221	Trypsin-like protease		10.8711	10.817	4.95966
154554	Alkaline serine protease		10.7861	0.062284	-8.35379
500888	Chitinase chi18-17		10.7537	-0.26535	-3.95398
521588	Isotrichodermin c-15 hydroxylase	12 h up	10.6739	6.96978	-1.43531
100207	Hypothetical protein TRIVIDRAFT_62551		10.488	1.96719	1.64245
524327	N,O-diacetyl muramidase(GH25)		10.0928	6.71248	3.93055
43497	Hypothetical protein TRIVIDRAFT_130513		10.0535	6.23974	1.18343
476485	Uncharacterized serine-rich protein		10.0402	2.01486	-0.79684
518894	Peptidase tripeptidyl-peptidase		-0.221944	16.3412	21.0713
485240	P1 β-1,6-glucanase		5.2448	11.7243	7.33577
11443	Transcriptional activator srcap-like protein		2.8351	11.5825	8.72928
502174	Conidiation-specific protein (con-13) protein		6.34144	11.5327	4.66579
11439	Hypothetical protein FOC4_g10000877		2.74235	11.4602	8.54921
96797	wsc domain-containing protein	24 h up	6.99977	11.4475	6.05176
525334	α-1,3-glucanase		5.13055	11.4057	6.72423
501003	Peptidase family m28		1.1649	11.4017	10.9878
128023	Cutinase		0.32008	11.3122	7.58831
503197	Related to spore coat protein sp96 precursor		3.26222	11.1371	9.11624
518894	Peptidase tripeptidyl-peptidase		-0.221944	16.3412	21.0713
491972	Fasciclin domain containing protein		0.17976	2.80756	11.4693
501003	Peptidase family m28		1.1649	11.4017	10.9878
96734	Glycoside hydrolase family 76		0.570337	9.93525	10.8109
511478	Proteinase inhibitor kazal		-1.56863	10.0734	10.7725
533861	Hydrophobin 1	36 h up	-0.624081	9.8541	10.5943
482878	Endonuclease/exonuclease/phosphatase family protein		0.314306	11.0782	10.372
78602	MFS carboxylic acid transport protein		-0.582243	10.0355	10.2595
110777	Alkaline proteinase/serin endopeptidase		6.23477	11.0992	10.226
10644	Alcohol oxidase		6.7904	8.15543	10.0216

after 12 and up-regulated after 24 and 36 h, and cluster 5 contains genes down-regulated after 12 h and 24 h up-regulated after 36 h. Figure 3A summarizes the clustering analysis as a matrix in which clusters 1 and 2 are presented as mirror images of clusters 4 and 5.

The functional category distribution frequency for each cluster was then calculated to identify differences in the distribution of genes among the three *Trichoderma* growth periods (Figure 3D). High percentages of hydrolases, peptidases and transporter activities were observed in clusters 1, 2 and 3, which included most of the genes up-regulated for all growth times, mainly after 12 h. Clusters 4 and 5 are represented by the lowest number of genes; however they presented a high diversity of functional categories and the smallest percentage of the hydrolase activity category.

Cluster 4 is mainly represented by specific transporters, oxidoreductase and peroxidase activities, which are absent in the other clusters. Cluster 5 was less diverse, but contains a high percentage of peptidases and the appearance of the lyase activity category. Clusters 4 and 5 contain genes induced after 24 and 36 hours. This pattern of categorization suggests a late increasing in the expression levels of genes encoding substrate transporters and other CAZymes such as lyases, instead of the classical cell wall hydrolases (chitinases, β-1,3-glucanases and proteases). Recently, we have described the production of polysaccharide lyases by *T. harzianum* during growth in the presence of different phytopathogen cell walls using a proteomic approach, which is in agreement with the presented data in this work (unpublished data). These results

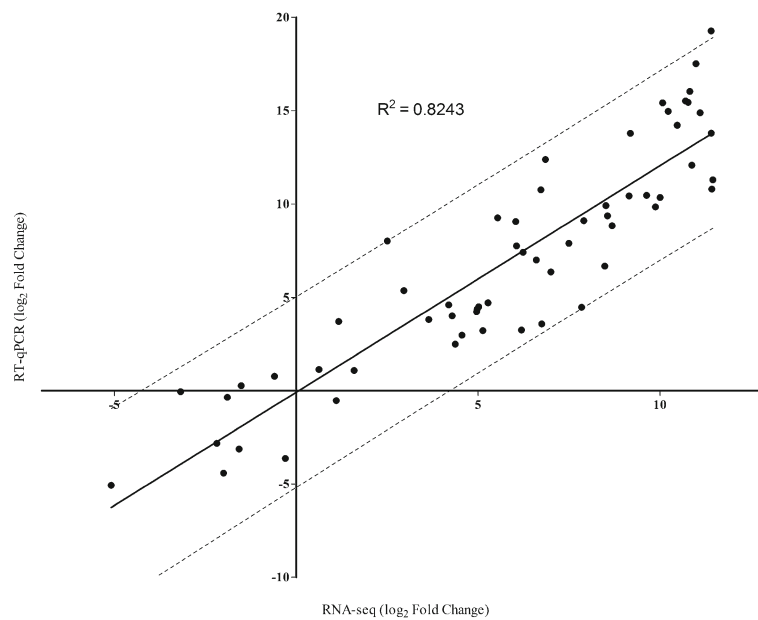


Figure 2 Pearson correlation between the RNA-seq and RT-qPCR data. All expression data were normalized in \log_2 — Represents the 95% confidence interval.

suggest a potential role of these enzymes, especially the lyase families 7 and 8, in mycoparasitism by *Trichoderma* independent of the host pathogen.

Validation of RNA-seq gene expression

To validate the expression profile obtained by “*in silico*” analysis of RNA-Seq data, RT-qPCR was performed using 20 genes randomly chosen among up or down-regulated differentially expressed genes. For a better understanding of expression kinetics, two additional growth times were

added to the analysis (6 and 18 hours). RT-qPCR expression profiles were in complete agreement with the RNA-Seq data for 17 of the genes evaluated. The other three (Lyase family 7, β -glucosidase, cysteine-rich protein) showed differences at one or two growth times, but the modulation pattern of expression was maintained as observed through the “*in silico*” analysis (Figure 4).

One of the primary goals of transcriptome sequencing is to compare gene expression levels in different samples. In the present work, RNA-seq analysis was carried out for a

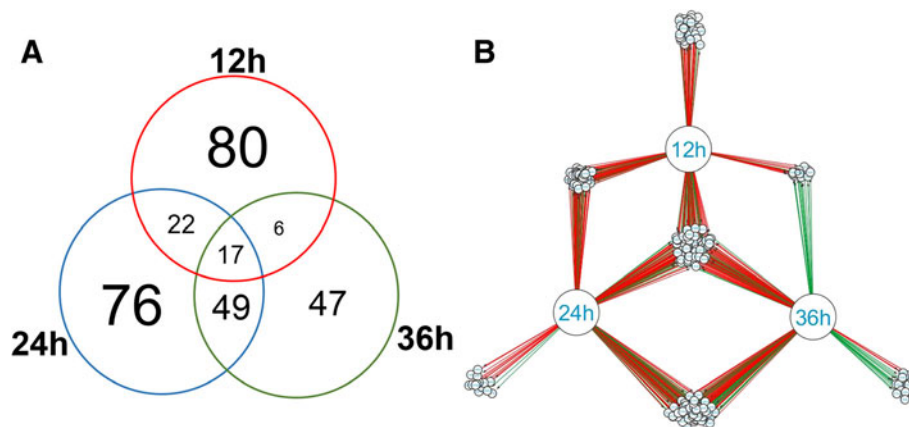


Figure 3 Venn diagram and network gene interaction of the 297 differentially expressed genes. (A) Venn diagram showing the distribution of differentially expressed genes detected by cuffdiff ($p < 0.05$) for *T. harzianum* grown on SSCW for 12, 24 and 36 hours when compared with glucose growth. (B) Network gene interaction showing differentially expressed genes up- and down-regulated (Fold Change > 2), for different times (12, 24 and 36 hours) and growth conditions, totaling 297 genes. Genes are represented as nodes (shown as small circles), and interactions are represented as edges (shown as lines, i.e. red interactions up-regulated and down-regulated interactions green), that connect the nodes: 592 interactions.

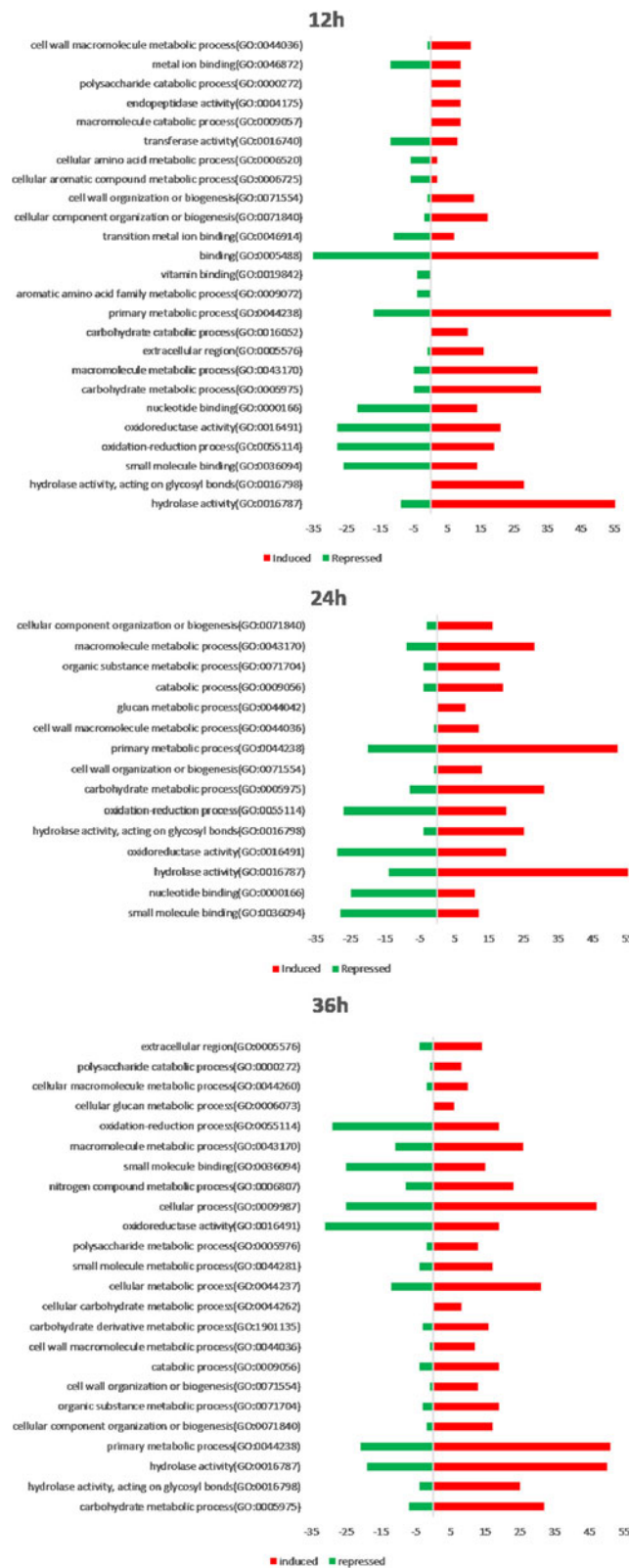


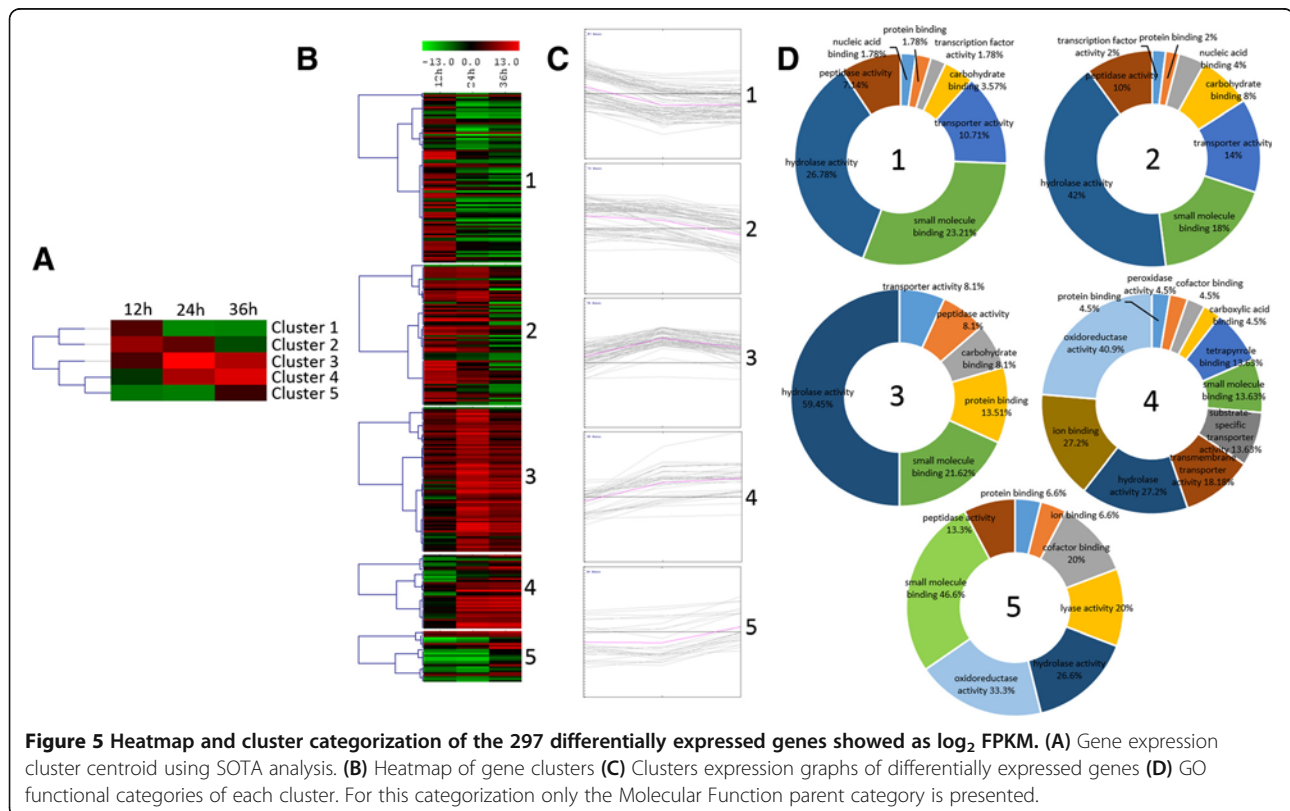
Figure 4 Annotation of differentially expressed genes. Distribution of differentially expressed genes using the Blast2GO function prediction tool. All GO functional categories in the differentially expressed genes were compared in the induced ($\log_2\text{FPKM} > 0$) (red) and repressed ($\log_2\text{FPKM} < 0$) (green) genes according to Fisher exact test ($p < 0.05$).

biological sample of pooled mycelia from three different growth cultures (biological replicates). Validation experiments using qPCR were subsequently carried out using the three biological replicates, and revealed a high Pearson's correlation coefficient between RNA-seq and qPCR expression data ($R^2 = 0.8243$) (Figure 5), this correlation enables us to use this data in differential expression analysis.

As a set of peptidases were up-regulated in the presence of SSCW, two serine peptidases genes (trypsin-like peptidase and serine endopeptidase 33 kDa), one aspartyl protease gene, one aminopeptidase gene (Peptidase M28) and one carboxypeptidase gene (Peptidase M14) were chosen to perform qPCR experiments. The expression data over the five growth times revealed an upward trend along the time course with the highest expression values at 18 hours followed by a decrease after this time or 24 hours. Their diversity and uniform time course production indicated that these enzymes may form a synergistic proteolytic system related to mycoparasitism in *T. harzianum*. However, their exact role in mycoparasitism has not been clearly elucidated yet. The main accepted hypothesis presents these enzymes as contributing to the breakdown of the fungal host cell wall, constituted by chitin and glucan polymers embedded in, and covalently linked to, a protein matrix [28], and also as acting as proteolytic inactivators of pathogen enzymes which are typically involved in the plant infection process [29].

The *Trichoderma* species, *T. reesei*, *T. atroviride*, and *T. virens*, may have one of the largest sets of proteases among fungi, of which the total number of predicted proteases are 383 (4.2% of all predicted protein coding genes), 445 (3.75%), and 479 (3.85%), respectively [30]. These authors described that the dominant groups are classified as aspartyl proteases, serine proteases, subtilisin-like proteases, and dipeptidyl and tripeptidyl peptidases. These enzymes have been described as performing a central role in the mycoparasitic activity of *Trichoderma* species, as they have been consistently identified during interaction against different phytopathogenic fungi using transcriptomic and proteomic approaches [12-14]. Their diversity and uniform time course production provide evidence that these enzymes may form a synergistic proteolytic system related to mycoparasitism in this *Trichoderma* species. Indeed, in our study, a number of genes encoding a serine peptidase, an aspartyl protease, a subtilisin-like, a trypsin-like beyond metallopeptidases (M28, M14) were also differentially expressed in the presence of SSCW. These results strongly suggest a role of these enzymes in *T. harzianum* mycoparasitism against *S. sclerotiorum*, and support a putative common action mechanism of mycoparasite fungi within the *Trichoderma* genus.

About 29% of the differentially expressed genes encode CAZy enzymes (Additional file 2: Table S1), suggesting a central role for them in *Trichoderma* mycoparasitism,



probably during host fungal cell wall degradation. The kinetics expression for CAZy category enzymes presented as expression average showed a decrease trend along the growth timecourse, with a maximal expression at 24 h. An exception were the enzymes categorized as auxiliary activities and polysaccharide lyases (Additional file 5: Table S3). Glycosyl hydrolase family 18 and other enzymes which act mainly as fungal cell wall degrading enzymes (CWDE) have also been described as presenting a central role in mycoparasitism in *Trichoderma atroviride* [31]. Genes encoding this family were also up-regulated in the present work based on “*in silico*” RNA-seq data analysis. As a consequence, eight genes encoding three GH18 (Chitinase 37 kDa, Chitinase 42 kDa and a not well characterized Chitinase), a α -1,3-glucanase, a β -glucosidase, a lyzosome-like, a polysaccharide lyase Family 7 and a carbohydrate binding Family 13 protein were selected for expression validation by RT-qPCR. The expression levels of all CWDE genes followed the same trend, except for the β -glucosidase gene that was repressed from 12 hours onwards. The common expression profile was an increase in transcripts after 12 hours until 24 hours and a similar level or decrease of transcripts at 36 hours. This kinetic suggests use of small sugars during the first growth time and an expression of CWDE after 12 hours, indicating a role in degradation of the cell wall as carbon source to allow continued growth. KEGG pathway analysis (Additional file 6: Table S7) shows the mapped genes in starch and sucrose metabolism. All genes mapped are in the final stages of the pathway, mainly in the formation of small sugars such as D-glucose and D-xylose.

Alginate lyases are enzymes classified as belonging to the polysaccharide lyase Family 7 and are usually involved in the deconstruction of complex polysaccharides, such as polyguluronate and polymannuronate [32]. Their expression observed in our study suggests a possible role in mycoparasitism, complementing the classical GH activity which is known to be involved. The carbohydrate binding module Family 13 gene encodes a protein with a domain related to lectins, which, in *Rhizoctonia solani*, is implicated in fungal insecticidal activity [33]. Our expression data suggest that this gene is related to the lectins and may play a role mediating the physical contact with the host and elicitation of the signaling cascade comprising G-proteins and MAPKs.

Small secreted cysteine-rich proteins (SSCPs) have been described as up-regulated in *Trichoderma* species during mycoparasitism against phytopathogenic fungi [31]. The present work is the first to report their probable role in mycoparasitism of *T. harzianum* against *S. sclerotiorum*.

In this work we also identified genes encoding two predicted cistein-rich proteins and *qid74*, all up-regulated in the presence of SSCW. These genes were highly expressed in the later induction time periods (24 and 36 h) as showed by “*in silico*” analysis of RNA-seq and are

in agreement with the results of validation by RT-qPCR. Hydrophobins I and II were also identified by the “*in silico*” analysis of RNA-seq data, as shown in Additional file 4: Table S2 with a high expression at 24 and 36 hours.

Small secreted cysteine-rich proteins (SSCPs) are one of the largest groups of proteins secreted by *Trichoderma*. Hydrophobins, probably the most widely known SSCP, are characterized by the presence of eight positionally conserved cysteine residues of which four occur in doubles. They are found on the outer surfaces of cell walls of hyphae and conidia, where they mediate interactions between the fungus and the environment and also between the fungus and host plant roots [34]. Class II hydrophobins represent the predominant class described for *Trichoderma* species [35]. *T. atroviride* and *T. virens* have also class I-like hydrophobins, however they present differences in several aspects when compared to other fungi and form a separate clade in phylogenetic analysis within the Ascomycetes [36]. As well as other cysteine-rich proteins, *T. harzianum* the *qid74* gene encodes a cell wall protein which has an important role in adherence to hydrophobic surfaces and cellular protection [34]. The gene RNaseT2, which has been described as a stress related protein and involved in permeability and stability of the plasmatic membrane in *Saccharomyces cerevisiae* [37], was also up-regulated after 12 hours growth of *T.harzianum* on SSCW, decreasing over time.

Among the 297 differentially expressed genes, 30 encode transporter proteins (Additional file 3: Table S4). MFS (Major Facilitator Superfamily) permeases are the most abundant proteins among transporter proteins over the three growth times. These proteins enable the transport of essential nutrients and ions, in addition to the excretion of end products of metabolism and cell-environment communication [38]. Their expression levels vary according to the time of growth and culture growth condition (presence or absence of SSCW and glucose).

In summary, our results demonstrated a time course dependent expression of *T. harzianum* genes during growth on media with *S. sclerotiorum* cell wall as sole carbon and nitrogen source. The majority of the genes described in this work have also been reported in the literature during growth of *Trichoderma* spp. on cell-wall of phytopathogenic fungi [10,12,13], as well as during confrontation against *R. solani* [31].

RT-qPCR for dual cultures

The direct confrontation assay is a powerful tool for studying mycoparasitism by *Trichoderma* [12,13,30] under laboratory conditions. To validate RNA-seq data using a growth condition which closely mimics the interaction *Trichoderma*/host fungus, RT-qPCR was also performed using total RNA from dual cultures of *T. harzianum* and *S. sclerotiorum* over three different interaction stages:

before physical contact (BC) between the mycoparasite and the host, during the contact (C) and after the contact (AC). As a control, a confrontation assay was conducted in which *T. harzianum* was challenged against itself. The same genes chosen for the RNA-seq “*in silico*” analysis validation were used in this analysis, with three genes presenting expression patterns in agreement with RNA-seq data (Figure 6). The peptidases serine endopeptidase, M14 peptidase (carboxypeptidase) and aspartate protease were up-regulated during and after contact between *T. harzianum* and *S. sclerotiorum*, when compared with the control bioassay, providing further evidence for these enzymes as important factors in the mycoparasitism (Figure 6A).

The two small cystein-rich proteins were up-regulated during the interaction between *T. harzianum* and *S. sclerotiorum* although they showed differing expression patterns over the time period (Figure 6B). The gene cystein rich (511478) was highly expressed after contact, whilst the gene small cystein rich (518220) before contact. These

data suggests that this group of proteins (SSCP) may present a synergistic time course dependent activity during the interaction. The RNase T2 gene was also up-regulated during the contact stage, confirming its role in this interaction, possibly through conferring membrane stability during contact with the phytopathogen hyphae (Figure 6B).

Among the eight genes encoding putative glycoside hydrolases, three were up-regulated in the interaction: polysaccharidelyase Family 7, β -glucosidase and the CB module Family 13. This data suggests a role of these proteins in the mycoparasitism, complementing the classical GH activity in the interaction. The carbohydrate binding module Family 13 gene was induced during and after contact with *S. sclerotiorum*. The other genes identified as up-regulated by RNA-seq “*in silico*” analysis and validated by qRT-PCR using SCCW were not up-regulated in this condition. This “not-complete” agreement between cell wall induction with SCCW and direct confrontation with

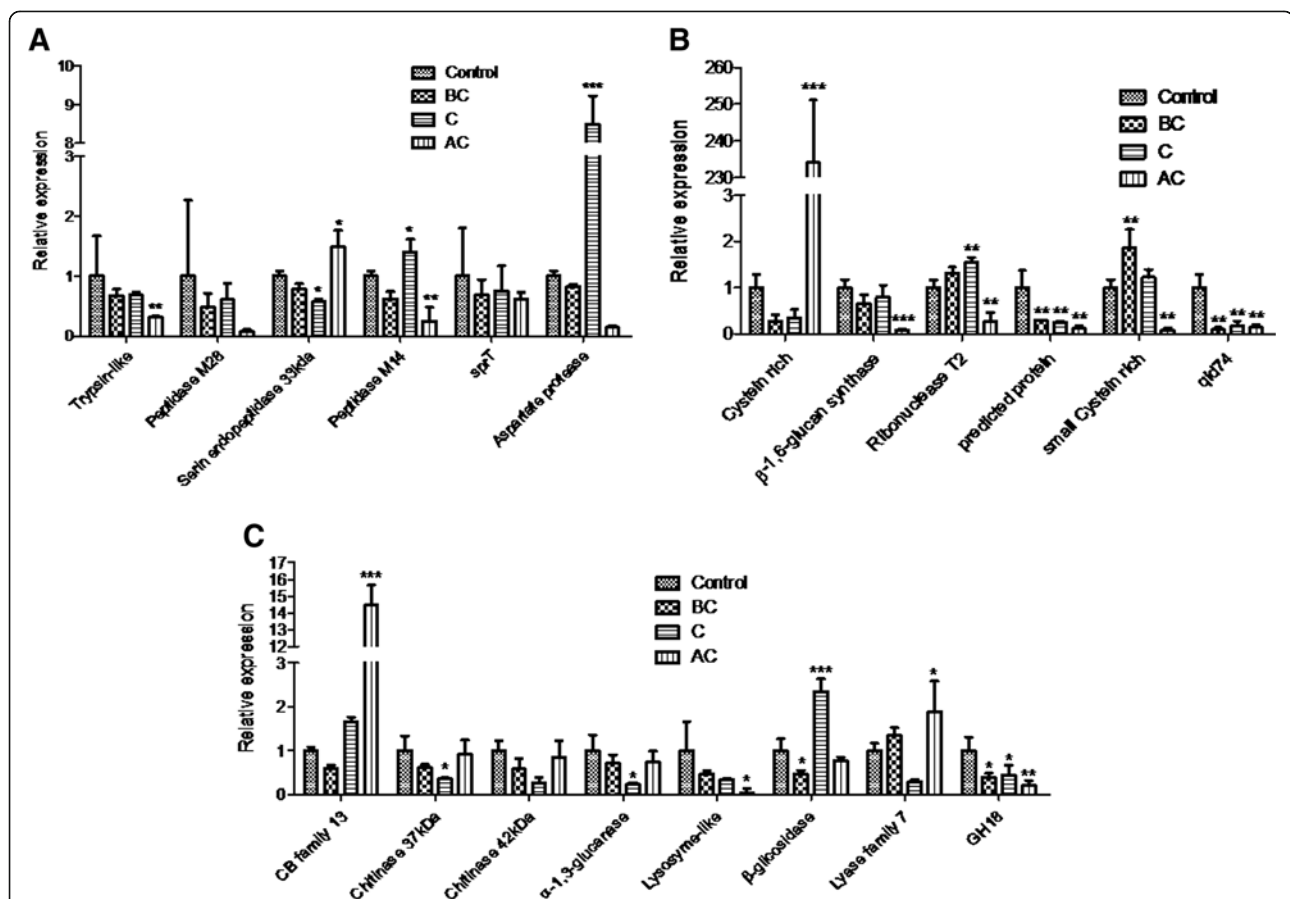


Figure 6 Differential expression analysis and quantification of transcript levels of biocontrol-related genes expressed by *T. harzianum* under mycoparasitic conditions against *S. sclerotiorum*. Control, BC – Before contact, C – Contact, AC – After Contact. (A) Expression analysis of Trypsin-like protein, Peptidase M28, serine endopeptidase 33 kDa, Peptidase M14, sprT, Aspartate protease. (B) Expression analysis of cystein rich protein, β -1,6-glucan synthase, ribonuclease T2, predicted protein, small cysteine rich protein, qid74. (C) Expression analysis of carbohydrate binding module family 13 protein, chitinase 37 kDa, Chitinase 42 kDa, α -1,3-glucanase, lysozyme-like protein, β -glucosidase, lyase family 7 protein, Glycoside Hydrolase family 18 protein. The data were presented as fold change using the $2^{-\Delta\Delta Ct}$ method. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

S. sclerotiorum is expected due to differences in growth conditions for *T. harzianum*. Investigation of mycoparasitism using inactivated cell walls is a useful approach to evaluate the broad arsenal of induced genes in *Trichoderma* spp. and to identify candidate mycoparasitism related genes that can be further evaluated for expression during mycoparasitism through dual culture assays. The genes which were observed to be up-regulated in both interaction bioassays are certainly promising candidates for future biotechnological applications as well as further detailed investigations to unveil their precise function in *T. harzianum* mycoparasitism.

Conclusions

The RNA-seq data presented will not only facilitate improvement of the annotation of gene models in the draft *T. harzianum* genome, but also provide important information regarding the transcriptome during growth on SSCW and during “*in vivo*” interactions with *S. sclerotiorum*. Our data represent an important step towards understanding the mycoparasitic process of *T. harzianum* during its interaction with *S. sclerotiorum*. Further studies for functional characterization of candidate genes reported here are necessary in order to better define the exact pathways involved in mycoparasitism in *T. harzianum*.

Methods

Organism culture conditions

T. harzianum TR274 was isolated from soil samples from the Brazilian Cerrado region and identified to species level based upon ribosomal RNA ITS 1 and 2 sequence identities (Genbank number KC993076). This strain is deposited in the ICB Enzymology Group culture collection at the Universidade Federal de Goias. A strain of the phytopathogenic fungus *S. sclerotiorum* belonging to the EMBRAPA-CNPAF culture collection was originally isolated from *Phaseolus vulgaris*. These fungi were grown on MYG medium (w/v: 0.5% malt extract, 0.25% yeast extract, 1% glucose and 2% agar) for 2 days at 28°C. *T. harzianum* spores were collected from cultures by washing with sterile water and centrifugation at 2000 rpm. Following two rounds of washes, spore suspensions at a concentration of 10^7 spores mL⁻¹ were used to inoculate flasks containing 50 mL of TLE medium [14]. Cultures were incubated at 28°C with constant shaking at 120 rpm. After 24 hours growth, mycelia were collected and transferred to flasks containing 50 mL of minimal medium (TM) (w/v: 0.2% KH₂PO₄, 1.4% (NH₄)₂SO₄, 0.03% MgSO₄·7H₂O) supplemented with 2% (w/v) of glucose or 0.5% (w/v) of inactivated SSCW (previously autoclaved at 120°C for 20 min). Cultures were incubated with constant shaking at 120 rpm at 28°C. After 6, 12, 18, 24 and 36 hours of growth, mycelium was harvested and immediately flash frozen in liquid nitrogen and stored at -80°C until RNA

isolation. This experiment was carried out in triplicate for each growth period, with mycelia subsequently pooled to constitute composed samples.

RNA isolation, cDNA library preparation and sequencing

RNA was extracted from macerated frozen mycelia using TRI-Reagent (Sigma-Aldrich), according to manufacturer's instructions. Integrity and quantity of isolated RNA were assessed using a RNA Pico chip on an Agilent Bioanalyzer 2100 (Agilent Technologies) (Additional file 7: Table S5). The RNA samples were stored at -80°C until library construction for sequencing and RT-qPCR. Normalized starting quantities of total RNA extracted from mycelia grown for 12, 24 and 36 h under the two different conditions (supplemented with glucose or SSCW), were then used to prepare six separate barcoded RNA-seq libraries using the TruSeq™ RNA sample preparation kit (Illumina, CA, USA). Three biological replicates were pooled in preparation of each of the six final samples. All library preparation and sequencing was carried out by Eurofins MWG Operon (AL, USA). Barcoded libraries were prepared according to the manufacturer's instructions, then sequenced on a single lane on an Illumina HiSeq2000 sequencer, to generate 100 bp paired-end reads.

Mapping of sequenced reads and assessment of gene expression

FastQC files were used to visualize the libraries quality before and after trimming. For quality trimming and sequence filtering, the software Trimmomatic (version 0.30) [39] was employed to remove sequencing adapters, k-mers, and bases with a Phred quality score lower than 20 from the read ends. All reads shorter than 25 nucleotides were then discarded. Filtered reads were mapped onto the *Trichoderma harzianum* v1.0 genome sequence (<http://genome.jgi.doe.gov/Triha1/Triha1.home.html>) using TopHat 2.0.8 release with default settings [40]. Gene expression values were determined using Cufflinks 2.1.1 release [41], and the FPKM (Fragments per kilobase mapped) values were calculated for each transcript. The gene expression levels in the *T. harzianum* genome were obtained using the Cuffdiff software within Cufflinks. Fungal transcript levels were calculated using uniquely mapped reads onto the genome. The expression profiles of differentially expressed genes were determined by SOTA (Self Organizing Tree Algorithm), cluster analysis was carried out using the MeV 4.9 software, and gene ontology assignment conducted according to Gene Ontology (GO) guidelines [25] using the Blast2GO platform [26].

Each *T. harzianum* sample grown in SSCW had a corresponding control sample of culture growth on glucose, enabling normalization using the relation FPKM SSCW/FPKM glucose. Positive values were considered as up-regulated transcripts in the presence of cell wall and

negative values were considered as down-regulated transcripts. Expression level differences were judged to be significant and the expression level estimate status acceptable when a gene was identified as differentially expressed with an FDR of the Benjamini-Hochberg multiple tests of 5% ($P < 0.05$).

Gene regulatory network of *T. harzianum*

A regulatory network was generated using Cytoscape 3.0.2 software and a table of data containing the differentially expressed genes detected using the cuffdiff program, presenting Fold Change > 2 , the interaction type (up- or down-regulated) and the target gene (i.e., the Protein ID of each gene affected). This analysis was carried out in order to reconstruct a *T. harzianum* time course network representation for all 297 identified genes (Figure 1B) [42].

RT-qPCR

Twenty genes were randomly selected between the differentially expressed genes *in silico* for the expression analysis by RT-qPCR. Real-time qPCR (Additional file 8: Table S6) primers were designed using the PerlPrimer v1.1.20 software. Total RNA from the above described preparations was digested with DNase I (Invitrogen) and a total of 5 μ g from each pooled sample was reverse transcribed into cDNA using the Maxima™ First Strand cDNA synthesis kit for RT-qPCR in a final volume of 20 μ L (Fermentas). The synthesized cDNA was diluted with 80 μ L of water and used as a template for real-time PCR reactions using the instrument iQ5 real-time PCR system (Bio-Rad). Each reaction (20 μ L) contained 10 μ L of MAXIMA® SYBR-green PCR Master Mix (Fermentas), forward and reverse primers (500 nM each), cDNA template, and nuclease free water. PCR cycling conditions were: 10 min at 95°C (1 cycle), 15 s at 95°C followed by 1 min at 60°C (40 cycles), and a melting curve ramping from 60°C to 95°C with an increasing temperature of 0.2°C for 10 s (1 cycle) and continuous data collection to test for primer dimers and nonspecific amplification. Determination of the PCR efficiency was performed using triplicate reactions from a dilution series of cDNA (1, 0.1, 10^{-2} , and 10^{-3}). Amplification efficiency was then calculated from the given slopes in the IQ5 Optical System Software v2.0 (Additional file 8: Table S6). The α -tubulin and β -actin transcript were used as internal references to normalize the amount of total RNA present in each reaction [12]. Gene expression levels were calculated from the threshold cycle according to the $2^{-\Delta\Delta CT}$ method [43]. All samples were analyzed in at least two independent experiments with three technical replicates in each run.

Analysis of expression of biocontrol-related genes

RT-qPCR was used to evaluate *T. harzianum* gene expression during confrontation against the fungal pathogen *S.*

sclerotiorum. Confrontation bioassays were conducted as described in Steindorff *et al.* [12]. *T. harzianum* TR274 and *S. sclerotiorum* circular plaques of 5 mm diameter were cut from 7-day-old cultures on MYG Plates. *T. harzianum* TR274 plaques were inoculated onto plates containing solid minimal medium supplemented with 0.2% of glucose and overlaid with cellophane at a distance of 7 cm from *S. sclerotiorum* plaque mycelia. A control confrontation assay was conducted following the same setup described above, except that *T. harzianum* was challenged against itself. Confrontation plates were incubated in the dark at 28°C and the mycelia were harvested at different growth stages: prior to fungal contact, at contact, and after contact (overlapping mycelia). The confrontation assays were conducted in triplicate with RNA extracted for all treatments and replicates. The RNA samples were used for RT-qPCR reactions as described above, with results compared by ANOVA coupled with the Dunnett's test ($\alpha = 5\%$) using GraphPad Prism 5 software, to allow analysis of differences between confrontation assay gene expression patterns.

Availability of supporting data

Sequences have been deposited at the Sequence Read Archive (SRA) of the National Center for Biotechnology under BioProject number PRJNA216008. Raw sequence reads can be found in <http://www.ncbi.nlm.nih.gov/sra/?term=PRJNA216008>.

Additional files

Additional file 1: Table S8. KEGG pathway analysis showing enzymes mapped in Starch and Sucrose Metabolism. Arrows in red, green and blue (12, 24 and 36 h respectively) show the up-regulated condition of all genes.

Additional file 2: Table S1. Primers used in qPCR experiments.

Additional file 3: Table S4. CAZY enzymes differentially expressed after 12, 24 and 36 h.

Additional file 4: Table S2. RNA-Seq sequencing and read mapping.

Additional file 5: Table S3. Functional annotation of the 297 differentially expressed genes.

Additional file 6: Table S7. \log_2 Fold change expression of CAZY classes. The data shown is the average and standard deviation on each time condition.

Additional file 7: Table S5. Transporters differentially expressed in 12, 24 and 36 hours.

Additional file 8: Table S6. Bioanalyser profile of the six samples used to construct RNA-seq libraries.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ASS, GJP, ASGC, RNGM, EFN performed the experimental design, RNA isolation, quality control and designed the bioinformatics analysis. ASS and MHSR performed the RT-PCR analyses and evaluation of the data. ASS drafted the manuscript. EFN and RNGM were responsible for revision of the manuscript. All authors approved the final version of the paper.

Acknowledgements

These sequence data (reference genome) were produced by the US Department of Energy Joint Genome Institute <http://www.jgi.doe.gov/> in collaboration with the user community. The current study was funded by the National Council for Scientific and Technological Development (CNPq) (Process 559680/2009-0). EFN and CJU were supported by a biotechnology research grant (FAPEGO and CNPq). ASS and MHSR were supported by CNPq PhD scholarships.

Author details

¹Departamento de Biologia Celular, Universidade de Brasília, Campus Universitário Darcy Ribeiro, Instituto de Ciências Biológicas, CEP 70.910-900 Brasília, DF, Brazil. ²EMBRAPA Recursos Genéticos e Biotecnologia, Parque Estação Biológica, CP 02372, CEP 70.770-900 Brasília, DF, Brazil. ³Escola de Agronomia e Engenharia de Alimentos, Universidade Federal de Goiás, Campus Samambaia, P.O. Box 131CEP 74001-970 Goiânia, GO, Brasil. ⁴Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Goiás, Campus Samambaia, Instituto de Ciências Biológicas, CEP 74.090-900 Goiânia, GO, Brazil.

Received: 28 November 2013 Accepted: 6 March 2014
Published: 18 March 2014

References

- Boland GJ, Hall R: **Index of plant hosts of *Sclerotinia sclerotiorum***. *Can J Pl Pathol* 1994, **16**:93–108.
- Attanayake RN, Carter PA, Jiang D, Del Río-Mendoza L, Chen W: ***Sclerotinia sclerotiorum* populations infecting canola from China and the United States are genetically and phenotypically distinct**. *Phytopathology* 2013, **103**(7):750–761.
- Sun P, Yang XB: **Light, temperature, and moisture effects on Apothecium production of *Sclerotinia sclerotiorum***. *Plant Dis* 2000, **84**:1287–1293.
- Lopes FA, Steindorff AS, Geraldine AM, Brandão RS, Monteiro VN, Lobo M Jr, Coelho AS, Ulhoa CJ, Silva RN: **Biochemical and metabolic profiles of *Trichoderma* strains isolated from common bean crops in the Brazilian Cerrado, and potential antagonism against *Sclerotinia sclerotiorum***. *Fungal Biol* 2012, **116**(7):815–824.
- Qualhato TF, Lopes FA, Steindorff AS, Brandão RS, Jesuino RS, Ulhoa CJ: **Mycoparasitism studies of *Trichoderma* species against three phytopathogenic fungi: evaluation of antagonism and hydrolytic enzyme production**. *Biotechnol Lett* 2013, **35**(9):1461–1468.
- Geraldine AM, Lopes FAC, Carvalho DDC, Barbosa ET, Rodrigues AR, Brandão RS, Ulhoa CJ, Lobo-Junior M: **Cell wall-degrading enzymes and parasitism of sclerotia are key factors on field biocontrol of white mold by *Trichoderma* spp.** *Biol Control* 2013, **67**(3):308–316.
- Lorito M, Woo SL, Harman GE, Monte E: **Translational research on *Trichoderma*: from Omics to the field**. *Ann Rev Phytopathol* 2010, **48**:395–417.
- Hermosa R, Viterbo A, Chet I, Monte E: **Plant-beneficial effects of *Trichoderma* and of its genes**. *Microbiology* 2012, **158**:17–25.
- Druzhinina IS, Seidl-Seiboth V, Herrera-Estrella A, Horwitz BA, Kenerley CM, Monte E, Mukherjee PK, Zeilinger S, Grigoriev IV, Kubicek CP: ***Trichoderma*: the genomics of opportunistic success**. *Nat Rev Microbiol* 2011, **16**(9):749–759.
- Vizcaino JA, Redondo J, Suarez MB, Cardoza RE, Hermosa R, Gonzales FJ, Rey M, Monte E: **Generation, annotation and analysis of ESTs from four different *Trichoderma* strains grown under conditions related to biocontrol**. *Appl Microbiol Biotechnol* 2007, **75**:853–862.
- Seidl V, Song L, Lindquist E, Gruber S, Koptchinskiy A, Zeilinger S, Schmol M, Martínez M, Sun J, Grigoriev I, Herrera-Estrella A, Baker SE, Kubicek CP: **Transcriptomic response of the mycoparasitic fungus *Trichoderma atroviride* to the presence of a fungal prey**. *BMC Genomics* 2009, **10**:567.
- Steindorff AS, Silva RN, Coelho ASG, Noronha EF, Ulhoa CJ: ***Trichoderma harzianum* expressed sequence tags for identification of genes with putative roles in mycoparasitism against *F. solani***. *Biol Control* 2012, **61**(2):134–140.
- Vieira PM, Coelho AS, Steindorff AS, De Siqueira SJ, Silva Rdo N, Ulhoa CJ: **Identification of differentially expressed genes from *Trichoderma harzianum* during growth on cell wall of *Fusarium solani* as a tool for biotechnological application**. *BMC Genomics* 2013, **14**:177.
- Monteiro VN, Silva RN, Steindorff AS, Costa FT, Noronha EF, Ricart CAO, Sousa MV, Vainstein MH, Ulhoa CJ: **New insights in *Trichoderma harzianum* antagonism of fungal plant pathogens by secreted protein analysis**. *Curr Microb* 2010, **61**:298–305.
- Rubio MB, Domínguez S, Monte E, Hermosa R: **Comparative study of *Trichoderma* gene expression in interactions with tomato plants using high-density oligonucleotide microarrays**. *Microbiology* 2012, **158**(Pt 1):119–128.
- Reithner B, Ibarra-Laclette E, Mach RL, Herrera-Estrella A: **Identification of mycoparasitism-related genes in *Trichoderma atroviride***. *Appl Environ Microbiol* 2011, **77**(13):4361–4370.
- Samolski I, de Luis A, Vizcaino JA, Monte E, Suárez MB: **Gene expression analysis of the biocontrol fungus *Trichoderma harzianum* in the presence of tomato plants, chitin, or glucose using a high-density oligonucleotide microarray**. *BMC Microbiol* 2009, **9**:217.
- Wang WC, Lin FM, Chang WC, Lin KY, Huang HD, Lin NS: **miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression**. *BMC Bioinform* 2009, **10**:328.
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y: **RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays**. *Genome Res* 2008, **18**(9):1509–1517.
- Ozsolak F, Milos PM: **RNA sequencing: advances, challenges and opportunities**. *Nat Rev Genet* 2011, **12**(2):87–98.
- Martin JA, Wang Z: **Next-generation transcriptome assembly**. *Nat Rev Genet* 2011, **12**(10):671–682.
- Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D, Kuo A, Minovitsky S, Nikitin R, Ohm RA, Otilar R, Poliakov A, Ratnere I, Riley R, Smirnova T, Rokhsar D, Dubchak I: **The genome portal of the Department of energy joint genome Institute**. *Nucleic Acids Res* 2012, **40**:26–32.
- Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA: **Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells**. *Genome Res* 2008, **18**:610–621.
- Druzhinina IS, Kubicek CP, Komon-Zelazowska M, Mulaw TB, Bissett J: **The *Trichoderma harzianum* demon: complex speciation history resulting in coexistence of hypothetical biological species, recent agamospecies and numerous relict lineages**. *BMC Evol Biol* 2010, **10**:94.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology: the gene ontology consortium**. *Nat Genet* 2000, **25**(1):25–29.
- Conesa A, Götz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research**. *Bioinformatics* 2005, **21**:3674–3676.
- Free SJ: **Fungal cell wall organization and biosynthesis**. *Adv Gene* 2013, **81**:33–82.
- Kapteyn JC, Montijn RC, Vink E, de la Cruz J, Llobell A, Douwes JE, Shimoi H, Lipke PN, Klis FM: **Retention of *Saccharomyces cerevisiae* cell wall proteins through a phosphodiester-linked beta-1,3-/beta-1,6-glucan heteropolymer**. *Glycobiology* 1996, **6**(3):337–345.
- Elad Y, Kapat A: **The role of *Trichoderma harzianum* protease in the biocontrol of *Botrytis cinerea***. *Eur J Plant Pathol* 1999, **105**:177–189.
- Druzhinina IS, Shelest E, Kubicek CP: **Novel traits of *Trichoderma* predicted through the analysis of its secretome**. *FEMS Microbiol Lett* 2012, **337**(1):1–9.
- Atanasova L, Crom SL, Gruber S, Couplier F, Seidl-Seiboth V, Kubicek CP, Druzhinina IS: **Comparative transcriptomics reveals different strategies of *Trichoderma* mycoparasitism**. *BMC Genomics* 2013, **14**:121.
- Wang Y, Guo EW, Yu WG, Han F: **Purification and characterization of a new alginate lyase from a marine bacterium *Vibrio* sp.** *Biotech Let* 2013, **35**(5):703–708.
- Skamnaki VT, Peumans WJ, Kantsadi AL, Cubeta MA, Plas K, Pakala S, Zographos SE, Smagghe G, Nierman WC, Van Damme EJ, Leonidas DD: **Structural analysis of the *Rhizoctonia solani* agglutinin reveals a domain-swapping dimeric assembly**. *FEBS J* 2013, **280**(8):1750–1763.
- Samolski I, Rinco AM, Pinzo LM, Viterbo A, Monte E: **The qid74 gene from *Trichoderma harzianum* has a role in root architecture and plant biofertilization**. *Microbiology* 2012, **158**:129–138.
- Kubicek CP, Baker S, Gamauf C, Kenerley CM, Druzhinina IS: **Purifying selection and birth-and-death evolution in the class II hydrophobin gene families of the ascomycete *Trichoderma/Hypocrea***. *BMC Evol Biol* 2008, **8**:4.

36. Seidl-Seiboth V, Gruber S, Sezerman U, Schwecke T, Albayrak A, Neuhoef T, Von Öhren H, Baker SE, Kubicek CP: **Novel hydrophobins from *Trichoderma* define a new hydrophobin subclass: protein properties, evolution, regulation, and processing.** *J Mol Evol* 2011, **72**:339–351.
37. MacIntosh GC, Bariola PA, Newbiggin E, Green PJ: **Characterization of Rny1, the *Saccharomyces cerevisiae* member of the T2 RNase family of RNases: unexpected functions for ancient enzymes?** *Proc Natl Acad Sci USA* 2001, **98**(3):1018–1023.
38. Pao SS, Paulsen IT, Saier MH: **Major facilitator superfamily.** *Microbiol Mol Biol R* 1998, **62**:1–34.
39. Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B: **RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics.** *Nucleic Acids Res* 2012, **40**:W622–W627.
40. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL: **TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.** *Genome Biol* 2013, **14**:R36.
41. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13**:2498–2504.
42. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.** *Nat Biotechnol* 2010, **28**:511–515.
43. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_t}$ method.** *Methods* 2001, **25**(4):402–408.

doi:10.1186/1471-2164-15-204

Cite this article as: Steindorff *et al.*: Identification of mycoparasitism-related genes against the phytopathogen *Sclerotinia sclerotiorum* through transcriptome and expression profile analysis in *Trichoderma harzianum*. *BMC Genomics* 2014 **15**:204.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



CONCLUSÃO

Os dados de RNA-seq apresentados irão não somente facilitar e melhorar a anotação de modelos gênicos no genoma de *T. harzianum* CBS 226.95, mas também irá prover importantes informações a respeito do transcrito durante o crescimento em parede celular e na interação *in vivo* com *S. sclerotiorum*. Estes dados representam um importante passo no sentido do entendimento do micoparasitismo de *T. harzianum* na interação com *S. sclerotiorum*. Estudos adicionais de caracterização funcional de genes candidatos reportados aqui são necessários para um melhor entendimento das vias envolvidas no micoparasitismo em *T. harzianum*.

CAPÍTULO 3: ANÁLISE DO TRANSCRITOMA DA **INTERAÇÃO IN VIVO ENTRE *T. harzianum* E *S.*** ***sclerotiorum***

INTRODUÇÃO

Como o processo pelo qual espécies de *Trichoderma* atuam como micoparasitas envolve alteração no padrão de expressão de seus genes na presença e no íntimo contato com o hospedeiro (Mendoza et al., 2015), as análises de transcrito e proteoma são importantes ferramentas no mapeamento de genes e proteínas relacionados a este antagonismo. A análise de transcritos (EST's), transcrito, representa uma eficiente forma de caracterizar o conjunto de genes expressos por um organismo, principalmente quando existe um genoma anotado disponível para o organismo estudado. A análise da interação *Trichoderma* e os fungos fitopatogênicos *Rhizoctonia solani* e *Botrytis cinerea* utilizando a construção e sequenciamento pelo método de Sanger de bibliotecas de EST's já foram publicados (Liu e Yang, 2005; Vizcaíno et al, 2006; Vizcaíno et al, 2007; Seidl et al, 2009). Somada a esta técnica novas estratégias de sequenciamento de menor custo e com maior capacidade de geração de dados já foram descritas, trazendo novas perspectivas para o estudo do transcrito da interação de *Trichoderma* frente à fitopatógenos.

Estas novas tecnologias de sequenciamento (*Next Generation Sequencing* - NGS) já estão sendo aplicadas com sucesso na análise do transcrito de eucariotos, permitindo a análise simultânea do nível de expressão, da variação estrutural e de sequência em um determinado *locus* (Wang et al, 2009). Estas técnicas possibilitam a obtenção de milhões de sequências e por isto, reduzem os custos destas análises e aumentam a possibilidade de detecção de transcritos raros.

Até o momento somente um trabalho utilizando *Next Generation Sequencing* – NGS no estudo da interação *Trichoderma*/fungo fitopatogênico foi publicado (Reithner et al, 2011). Neste trabalho foi utilizada a técnica de piro sequenciamento (Roche®) para a análise do transcrito da interação *T. atroviride/R. solani*, no entanto esta não é a ferramenta mais indicada para este tipo de análise em função de problemas inerentes à técnica relacionados à avaliação da expressão relativa com a quantidade de *reads*. Análises de sequências curtas utilizando sequenciamento de Illumina® se mostram mais eficazes neste sentido. A base molecular do controle biológico por espécies de *Trichoderma* permanece em boa parte desconhecida, mas acredita-se que sua atividade no controle biológico ocorra devido à produção de antibióticos, competição por nutrientes, produção de enzimas envolvidas na hidrólise da parede celular do hospedeiro ou por uma combinação destas atividades (Steindorff et al., 2012; Ramada et al., 2015).

Existe uma diferença nos genes e em seus padrões de expressão relacionados a diferentes fungos fitopatogênicos e linhagens de *Trichoderma* relacionados ao biocontrole (Vizcaíno et al, 2007, Atanasova et al., 2013). O mapeamento dos genes envolvidos no micoparasitismo vem para contribuir no entendimento deste mecanismo, bem como direcionar futuros trabalhos que envolva a obtenção de isolados mais efetivos no biocontrole e plantas resistentes à fitopatogenos.

METODOLOGIA

Manutenção dos isolados

O isolado *T. harzianum* TR274 e o fungo fitopatogênico *S. sclerotiorum* da coleção de fungos de solo mantida no Laboratório de Enzimologia da UFG, foram utilizados nos experimentos de interação

e análise de transcrito e genoma estrutural. O isolado TR274 foi escolhido como modelo deste trabalho, pois já foi caracterizado anteriormente como antagonista do fungo fitopatogênico *S. sclerotiorum*. Estes isolados foram mantidos com repiques periódicos em meio MYG (0,5% de extrato de malte, 0,25% de extrato de levedura, 1% de glicose e 2% de ágar) durante os experimentos e estocado em solução de glicerol (50%) em ultrafreezer -80 °C.

Condições de cultivo para isolamento de RNA

Foram utilizadas duas estratégias para obtenção dos micélios para os experimentos posteriores de análise da expressão gênica do isolado *T. harzianum* TR274 cultivado em meio líquido contendo a parede celular do fungo, condição que simula o micoparasitismo e o bioensaio de confronto direto entre *T. harzianum* e *S. sclerotiorum* em placas de Petri. Conídios do isolado *T. harzianum* TR274 (1×10^7 conídios.ml⁻¹) foram inoculados em frascos de 250 ml contendo 50 ml de meio TLE [Bactopetona 1,0 g l⁻¹, Uréia 0,3 g l⁻¹, KH₂PO₄ 2,0 g l⁻¹, (NH₄)₂SO₄ 1,4 g l⁻¹, MgSO₄.7H₂O 0,3 g l⁻¹, CaCl₂.2H₂O 0,2 g l⁻¹, 2% (v/v) de solução de elementos traços, pH 5,0] com 2% (p/v) de glicose e crescidos por 36 h. Após esse crescimento, os micélios foram filtrados, lavados com solução salina 0,9% e transferidos para frascos de 250 ml contendo 50 ml de Meio Mínimo [KH₂PO₄ 2,0 g l⁻¹, MgSO₄.7H₂O 0,3 g l⁻¹, CaCl₂.2H₂O 0,2 g l⁻¹, 2% (v/v) de solução de elementos traços, pH 5,0] com 0,5% (p/v) de parede celular de *S. sclerotiorum* ou 2% de glicose (controles). A parede celular de *S. sclerotiorum* foi macerada em grau e pistilo, liofilizada e lavada para retirada de proteínas intracelulares, logo depois liofilizada novamente para uso. Os frascos foram incubados em agitador rotatório à 28 °C e 180 rpm, após 12, 24 e 36 horas de crescimento os micélios foram coletados e utilizados para posterior extração de RNA total.

Para o experimento de interação in vivo entre *T. harzianum* TR274 e *S. sclerotiorum*. As condições analisadas foram: 1 cm antes do contato, no momento do contato, 48 h após o contato e um controle de *T. harzianum* crescido contra ele mesmo em meio BDA suplementado com 0,2% de glicose. Este experimento foi feito em triplicata biológica. Como controle para todas as amostras, utilizamos o crescimento do *T. harzianum* sozinho na placa de *petri* com o mesmo meio de cultura. Cada réplica biológica contém 3 placas de interação. Estas amostras foram utilizadas para os experimentos de RNA-seq e qPCR.

Preparação das amostras de RNA para sequenciamento

Os micélios obtidos conforme acima descrito foram macerados em gral e pistilo usando nitrogênio líquido. Este macerado foi utilizado para isolamento do RNA total utilizando o reagente TRIreagent (Sigma), de acordo com as recomendações do fabricante. Foi analisada a qualidade das preparações de RNA utilizando NanoDrop (Thermo scientific) e o perfil do BioAnalyser (Agilent Technologies) com um RIN > 8. Após certificada a qualidade, as amostras de RNA foram enviadas para a empresa Eurofins (EUA) que procedeu à construção das bibliotecas e seu sequenciamento no equipamento Hiseq2000 (Illumina). Para o sequenciamento das amostras de RNA dos micélios obtidos por cultivo em meio líquido contendo parede celular de *S. sclerotiorum* foram construídas seis bibliotecas, enquanto que para o bioensaio de confronto quatro bibliotecas foram construídas.

Sequenciamento e análise dos dados de RNA-seq

O processamento inicial de todos os dados de RNA-seq foi feito pela poda das regiões com baixa qualidade (Phred < 20) e adaptadores utilizando o programa Trimmomatic versão 0.30. Sequências

menores que 25 pb foram descartadas. O transcrito de fungo *T. harzianum* crescido em parede celular de *S. sclerotiorum* foi mapeado no genoma disponível naquele momento: *Trichoderma harzianum* v1.0 (<http://genome.jgi.doe.gov/Triha1>) utilizando o programa TopHat 2.0.8 (Trapnell et al., 2009) nas configurações padrão. As sequências da interação *in vivo* passaram pelo mesmo processamento, exceto que o mapeamento foi feito no próprio genoma do isolado (<http://genome.jgi.doe.gov/Trihar1>). Os valores de expressão gênica foram determinados utilizando o programa Cufflinks 2.1.1 e os valores de FPKM (Fragments per kilobase mapped) foram calculados para cada gene. Os níveis de expressão comparativa foram obtidos utilizando o programa CuffDiff somente com sequências mapeadas uma única vez no genoma. Os perfis de expressão diferencial e agrupamentos foram analisados pelo SOTA (*Self Organizing Tree Algorithm*) utilizando o programa MeV 4.9, e anotação dos agrupamentos foi feita utilizando a plataforma Blast2GO.

Cada amostra de *T. harzianum* crescida em parede celular de *S. sclerotiorum* (SSCW) foi normalizada com seu respectivo controle crescido em glicose. Valores positivos de Log_2FC (fold change) foram considerados *up-regulated* e valores negativos *down-regulated*. Níveis de expressão diferencial foram considerados significantes utilizando FDR (False Discovery Rate) do teste de Benjamini-Hochberg ($P < 0.05$).

Todos os dados oriundos do RNA-seq foram utilizados para montagem do transcrito total utilizando o programa Trinity (Grabher et al., 2011), nas configurações padrão. Estes dados serviram de entrada para a anotação estrutural do genoma.

RESULTADOS E DISCUSSÃO

Neste estudo, foi utilizada a técnica de RNA-seq para avaliar os genes diferencialmente expressos durante a interação entre *T. harzianum* TR274 e *S. sclerotiorum* em três diferentes situações: antes do contato (BC), durante o contato (C) e após o contato (AC). Nesta etapa, como o genoma deste isolado já estava montado e anotado, ele foi utilizado para o mapeamento. Com isso foi obtido $95,3\% \pm 1,33\%$ de *reads* mapeadas unicamente no genoma. A critério de comparação, o artigo publicado na seção anterior com o mapeamento no genoma do isolado CBS 226.95 obteve $82,6\% \pm 10,8\%$ de *reads* mapeadas.

Dentre os 13932 modelos gênicos preditos para este genoma, 3396 foram diferencialmente expressos em pelo menos uma das condições analisadas. Mais precisamente 392 genes diferencialmente expressos antes do contato, 256 durante o contato e 2748 depois do contato (Figura 10).

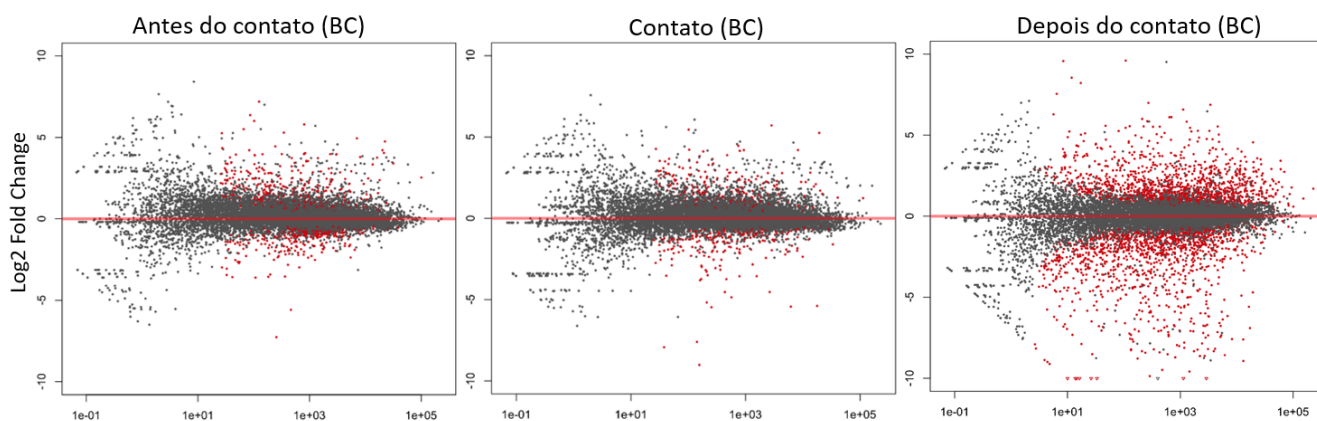


Figura 10. Comparação dos perfis de expressão genoma completo de *T. harzianum* TR274 na interação com *S. sclerotiorum* antes do contato (BC) com 392 genes diferencialmente expressos, durante o contato (C) com 256 genes diferencialmente expressos e após o contato (AC) com 2748 genes diferencialmente expressos mostrados em vermelho ($P < 0.05$).

A figura 11 é composta por dois diagramas de Venn mostrando a quantidade de genes reprimidos (*down-regulated*) e induzidos (*up-regulated*). Nota-se que a quantidade de genes

exclusivos de uma única condição é maior que as intercessões, sugerindo que cada condição tem suas particularidades e requerem diferentes genes para a ação do fungo *T. harzianum*. Outro fator a se notar é que a quantidade de genes induzidos e reprimidos, no total (1719 e 1677 genes respectivamente) não é tão variável. Isso mostra que a ativação e inativação da transcrição gênica são importantes durante a interação, diferente das induções utilizando parece celular, onde existe um padrão maior de ativação da expressão gênica.

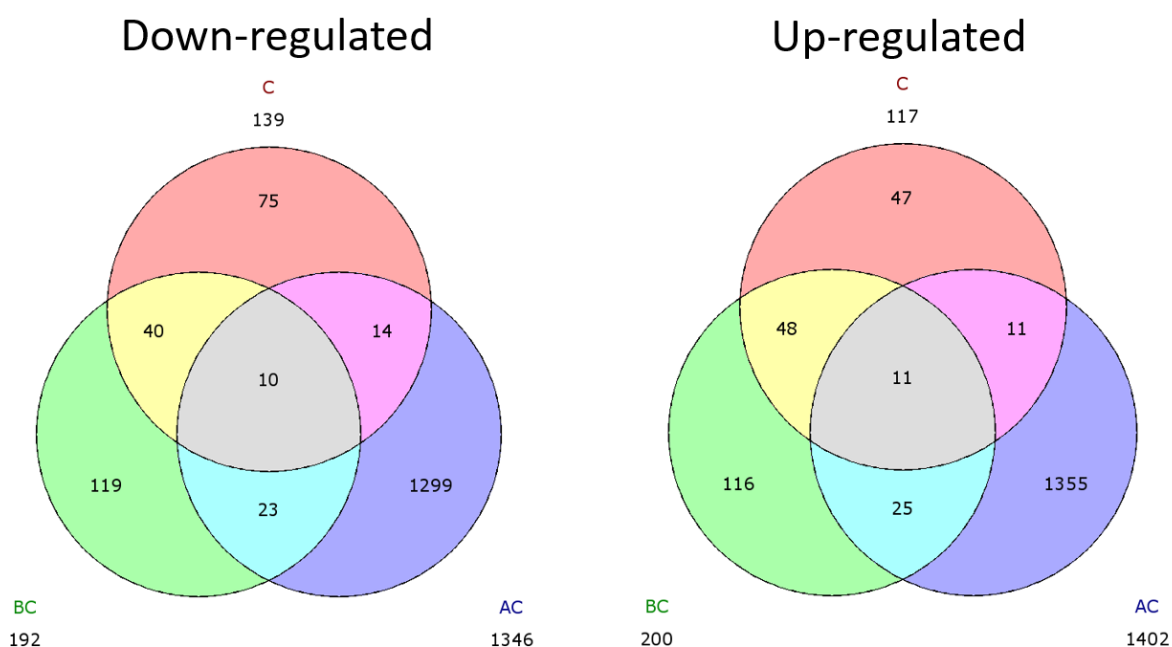


Figura 11. Diagrama de Venn dos 3396 genes diferencialmente expressos, divididos em reprimidos (down-regulated) e induzidos (up-regulated) durante as três condições de interação: Antes do contato (BC), durante o contato (C) e após o contato (AC).

Para categorização das proteínas quanto à sua função em experimentos de transcrito, são utilizadas ferramentas como *Gene Ontology* (Ashburner et al., 2000), KOG (Tatusov et al., 2003) e KEGG (Kanehisa et al., 2015). São ferramentas muito interessantes e eficazes em genomas bem anotados, o que não é o caso de fungos, onde cerca de 60% das proteínas preditas não tem qualquer anotação. Para evitar a geração de anotações genéricas, foram utilizadas as anotações já feitas para o

genoma, de fatores de transcrição, transportadores, CAZymes, proteases, SSCPs e genes do metabolismo secundário. A figura 12 mostra a distribuição dessas anotações nas condições analisadas. Nota-se que existe uma quantidade maior dessas classes de proteínas depois do contato. Antes do contato existe uma quantidade levemente maior de genes induzidos quando comparado com o contato. Mais precisamente de fatores de transcrição e transportadores, sugerindo que o fungo *T. harzianum* tem uma certa sensibilidade a presença do outro fungo.

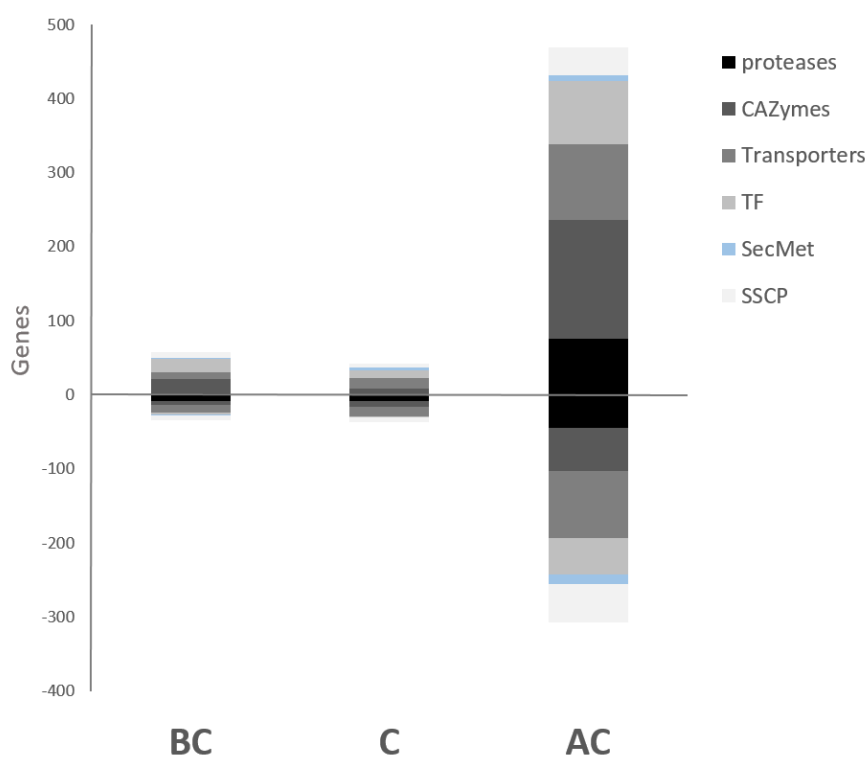


Figura 12. Quantidade de genes diferencialmente expressos categorizados como proteases, CAZymes, transportadores, fatores de transcrição (TF), metabolismo secundário (SecMet) e pequenas proteínas secretadas e ricas em cisteínas (SSCP). As barras positivas mostram a quantidade de genes induzidos e as barras negativas, reprimidos. As barras representam as condições: antes do contato (BC), durante o contato (C) e após o contato (AC).

Para melhor avaliar a distribuição das funções ao longo da interação, foi feita uma análise de agrupamento dos genes de acordo com seus respectivos padrões de expressão. Foram obtidos cinco

clusters a partir desta análise (Figura 13). O que se vê é que as condições BC e C são bem similares quanto aos seus padrões de expressão, enquanto a condição AC é bem diferenciada. Isso sugere que antes e no momento do contato, o fungo ainda se encontra em uma fase preparatória para uma interação efetiva. O intervalo de 48 horas entre o contato e após o contato é suficiente para o fungo *T. harzianum* ativar seu arsenal gênico de competição contra seu antagonista.

Apesar da divisão em 5 grupos, percebe-se somente dois padrões de expressão: Indução nas condições BC e C com repressão na AC (*clusters* 1, 2 e 3) e a ativação na AC com repressão nas condições BC e C (*clusters* 4 e 5).

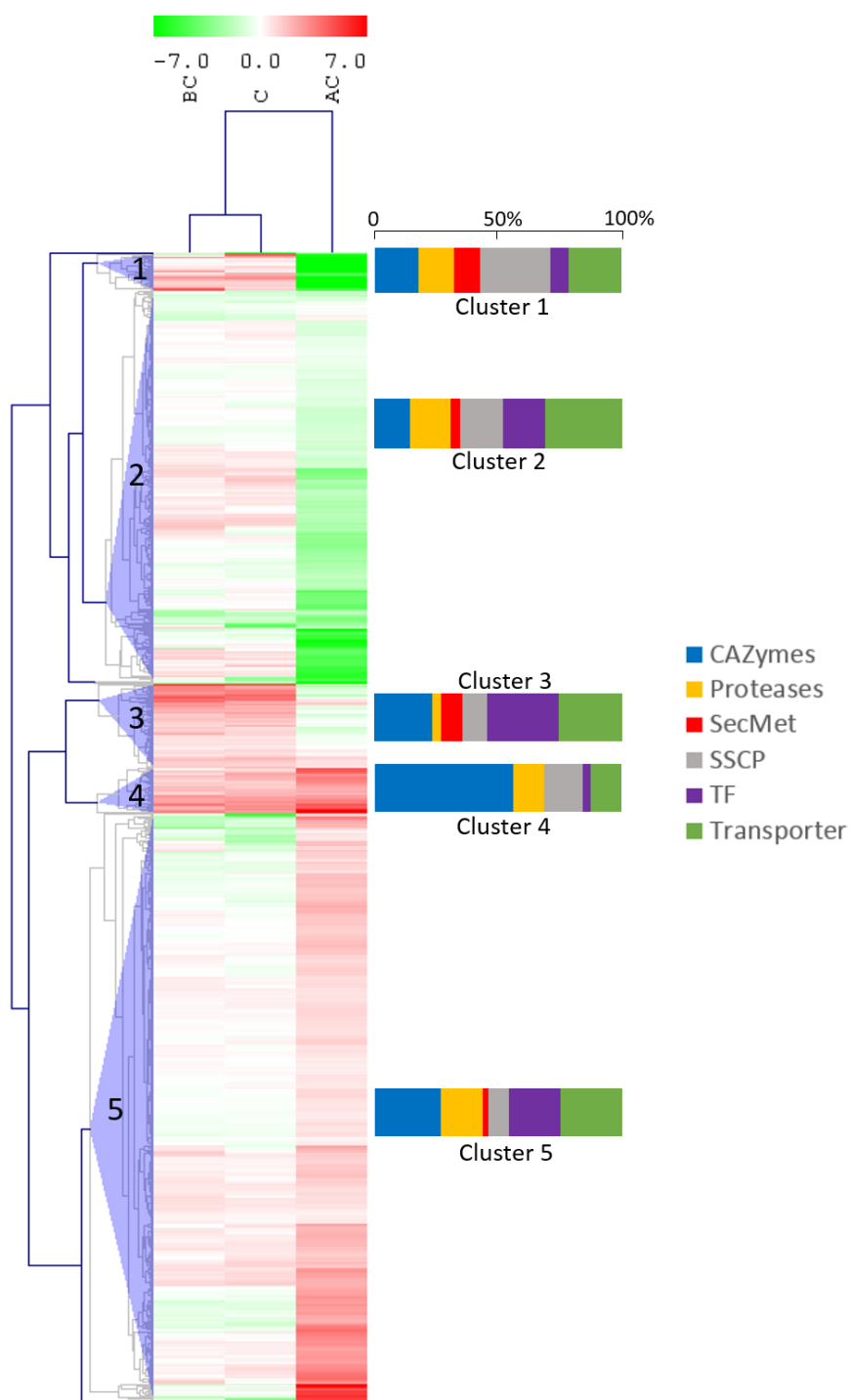


Figura 13. *Heatmap*, agrupamento hierárquico (distância euclidiana) e categorização dos agrupamentos dos genes diferencialmente expressos. A escala dos valores de expressão é mostrada em \log_2 fold change. As categorias mostradas nas barras horizontais mostram a proporção de cada

classe de CAZymes, proteases, metabolismo secundário (SecMet), pequenas proteínas secretadas ricas em cisteínas (SSCP), Fatores de transcrição (TF) e transportadores.

As funções relacionadas aos grupamentos mostram que os grupos 4 e 5, com expressão aumentada depois do contato, mostram um aumento na quantidade de enzimas do CAZy, mais perceptível no *cluster* 4. As categorias de genes onde seus produtos têm ação mais indireta, tem uma expressão mais alta antes e durante o contato, como transportadores (*cluster*2), fatores de transcrição (*cluster*3), SSCP (*cluster* 1) e metabolismo secundário (*cluster* 1 e 3). A quantidade de proteases não variou tanto ao longo dos *clusters*, exceto no *cluster* 3, onde a quantidade foi bem menor comparada com os outros. Isso mostra que as proteases têm papel importante em todas as etapas da interação.

Uma outra forma de avaliar quais categorias são significativas e diferencialmente expressas, é o teste exato de Fisher, frequentemente utilizado com as categorizações do *Gene Ontology* (Ashburner et al., 2000). Como o propósito era evitar o uso desse sistema de categorização, foi feita a construção de um banco dados de anotações de domínios PFAM de todo o genoma e depois avaliamos quantos desses domínios foram diferencialmente expressos em relação ao total de domínios PFAM do genoma (Tabelas 3, 4 e 5).

Na condição antes do contato, os domínios PFAM significativamente induzidos são aqueles envolvidos com metabolismo secundário, mais especificamente domínios frequentes em PKS e NRPS (Condensation, PP-binding, AMP-binding, FAD/NAD-binding e KR) e genes que também podem estar presentes em *clusters* de metabolismo secundário, como transportadores (C2) e fatores de transcrição (Fungaltrans2). Outros domínios presentes são Chitinsynth2 e Glycohydro3 que neste caso, podem estar envolvidos com remodelamento de parede celular.

Tabela 3. Domínios PFAM diferencialmente expressos pelo teste exato de Fisher ($P < 0,05$) antes do contato.

Reprimido Antes do Contato			Induzido Antes do Contato		
PFAM	Frequencia	P-valor	PFAM	Frequencia	P-valor
CofilinADF	2/6 (33.33%)	2,73E-03	Condensation	18/90 (20.00%)	4,75E-16
EcKinase	2/7 (28.57%)	3,79E-03	PP-binding	18/116 (15.52%)	5,09E-14
tRNAanti-codon	2/9 (22.22%)	6,38E-03	AMP-binding	16/116 (13.79%)	8,79E-12
tRNA-synt2	2/11 (18.18%)	9,58E-03	C2	4/16 (25.00%)	6,55E-05
Aminotran3	2/12 (16.67%)	1,14E-02	FADbinding8	3/16 (18.75%)	1,42E-03
BBE	2/17 (11.76%)	2,24E-02	Dak1	2/5 (40.00%)	1,99E-03
FADbinding4	3/46 (6.52%)	2,54E-02	Dak2	2/5 (40.00%)	1,99E-03
Pectatelyase3	2/20 (10.00%)	3,05E-02	Fungaltrans2	6/110 (5.45%)	5,06E-03
APH	2/25 (8.00%)	4,61E-02	Cyt-b5	3/27 (11.11%)	6,61E-03
			Chitinsynth2	2/9 (22.22%)	6,91E-03
			AMP-bindingC	3/30 (10.00%)	8,90E-03
			Glycotranf23	2/11 (18.18%)	1,04E-02
			Glycotrans23	2/11 (18.18%)	1,04E-02
			Ank2	10/297 (3.37%)	1,08E-02
			adhshort	9/253 (3.56%)	1,09E-02
			Cellulase	2/12 (16.67%)	1,23E-02
			Ferricroduct	2/14 (14.29%)	1,67E-02
			Fn3-like	2/15 (13.33%)	1,90E-02
			Rhodanese	2/15 (13.33%)	1,90E-02
			Glycohydro3C	2/16 (12.50%)	2,16E-02
			Glycohydro3	2/17 (11.76%)	2,42E-02
			KR	7/201 (3.48%)	2,61E-02
			NaCaex	2/18 (11.11%)	2,70E-02
			NADbinding6	2/19 (10.53%)	2,99E-02
			adhshortC2	5/126 (3.97%)	3,52E-02

Durante o contato os domínios diferencialmente expressos e induzidos, não mudam tanto quando comparados com antes do contato, o que muda é a adição de mais alguns domínios com funções semelhantes, como fatores de transcrição (Fungaltrans, Znclus), transportadores (Sugartr) e domínios também presentes em genes do metabolismo secundário (Ketoacyl-synt, Acyltransf1, ketoacyl-synt, Methyltransf).

Tabela 4. Domínios PFAM diferencialmente expressos pelo teste exato de Fisher ($P < 0,05$) durante o contato.

Reprimido Contato			Induzido Contato		
PFAM	Frequencia	p-valor	PFAM	Frequencia	p-valor
Cupin1	4/9 (44.44%)	1,15E-06	PP-binding	8/116 (6.90%)	5,81E-06
Cupin2	4/32 (12.50%)	2,75E-04	Condensation	7/90 (7.78%)	1,03E-05
Asp	3/25 (12.00%)	1,90E-03	Ketoacyl-syntC	4/30 (13.33%)	1,09E-04
DUF1776	2/8 (25.00%)	2,66E-03	Acyltransf1	4/32 (12.50%)	1,42E-04
Glytransfsug	2/11 (18.18%)	5,12E-03	Dak1	2/5 (40.00%)	6,88E-04
NADbinding10	5/143 (3.50%)	1,43E-02	Dak2	2/5 (40.00%)	6,88E-04
ADHN	4/99 (4.04%)	1,72E-02	DUF3433	2/7 (28.57%)	1,43E-03
p450	4/123 (3.25%)	3,47E-02	Fungaltrans	8/269 (2.97%)	1,94E-03
FADbinding3	3/81 (3.70%)	4,72E-02	ketoacyl-synt	3/34 (8.82%)	2,86E-03
Grp1Fun34YaaH	1/5 (20.00%)	4,89E-02	Methyltransf12	3/46 (6.52%)	6,74E-03
STAS	1/5 (20.00%)	4,89E-02	Rhodanese	2/15 (13.33%)	6,84E-03
Sulfatetransp	1/5 (20.00%)	4,89E-02	Znclus	9/412 (2.18%)	7,88E-03
			NaCaex	2/18 (11.11%)	9,80E-03
			Sugartr	5/171 (2.92%)	1,46E-02
			PS-DH	2/23 (8.70%)	1,58E-02
			AMP-binding	4/116 (3.45%)	1,65E-02
			Methyltransf11	3/66 (4.55%)	1,80E-02
			Methyltransf23	3/71 (4.23%)	2,18E-02
			Catalase-rel	1/5 (20.00%)	4,13E-02
			DUF4419	1/5 (20.00%)	4,13E-02
			Glycohydro28	1/5 (20.00%)	4,13E-02
			Grp1Fun34YaaH	1/5 (20.00%)	4,13E-02
			LactamaseB3	1/5 (20.00%)	4,13E-02
			PA14	1/5 (20.00%)	4,13E-02
			SAM2	1/5 (20.00%)	4,13E-02
			TransketolaseC	1/5 (20.00%)	4,13E-02
			Glycohydro30	1/6 (16.67%)	4,94E-02
			Glycohydro43	1/6 (16.67%)	4,94E-02
			Transketpyr	1/6 (16.67%)	4,94E-02
			Yphosphatase	1/6 (16.67%)	4,94E-02

Tabela 5. Domínios PFAM diferencialmente expressos pelo teste exato de Fisher ($P < 0,05$) após o contato.

Reprimido Depois do Contato			Induzido Depois do Contato		
PFAM	Frequencia	P-valor	PFAM	Frequencia	P-valor
PP-binding	47/116 (40.52%)	9,56E-19	WSC	18/27 (66.67%)	1,95E-12
Condensation	40/90 (44.44%)	7,23E-18	Pectatelyase3	12/20 (60.00%)	6,02E-08
AMP-binding	43/116 (37.07%)	1,38E-15	Septin	6/7 (85.71%)	6,58E-06
PUF	10/17 (58.82%)	7,07E-07	DUF3328	14/37 (37.84%)	6,72E-06
MORN	6/6 (100.00%)	8,05E-07	Glycohydrocc	5/5 (100.00%)	1,03E-05
ABCtran	26/111 (23.42%)	1,53E-05	SNase	5/6 (83.33%)	5,64E-05
PAN1	6/8 (75.00%)	1,90E-05	Glycohydro71	5/7 (71.43%)	1,81E-04
LysM	13/37 (35.14%)	2,34E-05	Condensation	21/90 (23.33%)	1,85E-04
DUF3659	5/6 (83.33%)	4,61E-05	Fungaltrans2	24/110 (21.82%)	2,01E-04
VCBS	6/9 (66.67%)	5,23E-05	Aldoketred	14/49 (28.57%)	2,34E-04
ABCmembrane	17/65 (26.15%)	1,05E-04	DUF1034	5/8 (62.50%)	4,42E-04
PAN4	6/13 (46.15%)	7,59E-04	Glycohydro75	4/5 (80.00%)	4,70E-04
Ank2	46/297 (15.49%)	8,73E-04	CFEM	8/21 (38.10%)	6,32E-04
HxxPfrpt	5/10 (50.00%)	1,39E-03	3BetaHSD	15/60 (25.00%)	6,90E-04
AMP-bindingC	9/30 (30.00%)	1,56E-03	LysM	11/37 (29.73%)	7,45E-04
NACHT	20/102 (19.61%)	1,62E-03	Sel1	11/37 (29.73%)	7,45E-04
Cuamineoxid	4/7 (57.14%)	2,39E-03	Fringe	4/6 (66.67%)	1,30E-03
Hydrophobin2	4/7 (57.14%)	2,39E-03	FADbinding4	12/46 (26.09%)	1,56E-03
Acetyltransf1	13/57 (22.81%)	2,59E-03	NADbinding10	26/143 (18.18%)	2,06E-03
PNPUDP1	12/53 (22.64%)	3,97E-03	Glycohydro18	9/30 (30.00%)	2,08E-03
Cholinekinase	4/8 (50.00%)	4,41E-03	Abhydrolase6	27/153 (17.65%)	2,66E-03
ABC2membrane3	3/5 (60.00%)	7,75E-03	DUF3433	4/7 (57.14%)	2,78E-03
DUF4238	3/5 (60.00%)	7,75E-03	Homeobox	5/11 (45.45%)	2,82E-03
Mo-codimer	3/5 (60.00%)	7,75E-03	Abhydrolase5	21/111 (18.92%)	3,26E-03
PeptidaseC14	3/5 (60.00%)	7,75E-03	AMP-binding	21/116 (18.10%)	5,56E-03
Ank	49/364 (13.46%)	1,06E-02	GFA	8/29 (27.59%)	6,43E-03
Methyltransf2	5/15 (33.33%)	1,10E-02	NaCaex	6/18 (33.33%)	6,58E-03
Cu-oxidase2	4/10 (40.00%)	1,13E-02	Epimerase	24/142 (16.90%)	7,72E-03
ADHzincN2	10/47 (21.28%)	1,28E-02	Glycohydro28	3/5 (60.00%)	8,70E-03
ketoacyl-synt	8/34 (23.53%)	1,38E-02	Glycohydro72	3/5 (60.00%)	8,70E-03
DSBA	3/6 (50.00%)	1,44E-02	ADHN	18/99 (18.18%)	9,37E-03
Thioesterase	3/6 (50.00%)	1,44E-02	Asp	7/25 (28.00%)	9,73E-03
Amidase	9/41 (21.95%)	1,46E-02	ADHzincN	18/101 (17.82%)	1,15E-02
ADHzincN	17/101 (16.83%)	1,63E-02	peroxidase	4/10 (40.00%)	1,30E-02
DynaminM	3/7 (42.86%)	2,34E-02	DUF202	3/6 (50.00%)	1,61E-02
Acetyltransf3	5/18 (27.78%)	2,46E-02	HomeoboxKN	3/6 (50.00%)	1,61E-02
ADHN	16/99 (16.16%)	2,75E-02	COesterase	10/47 (21.28%)	1,68E-02
NADbinding4	10/53 (18.87%)	2,86E-02	MFS1	53/394 (13.45%)	1,73E-02
Acyltransf1	7/32 (21.88%)	3,03E-02	FAAhydrolase	5/16 (31.25%)	1,74E-02
Methyltransf31	10/54 (18.52%)	3,22E-02	Glycohydro16	5/16 (31.25%)	1,74E-02
Polysaccsynt2	6/26 (23.08%)	3,43E-02	Glycohydro2	4/11 (36.36%)	1,89E-02
Catalase	3/8 (37.50%)	3,48E-02	SnoaL2	4/11 (36.36%)	1,89E-02
Cu-oxidase	3/8 (37.50%)	3,48E-02	Sugartr	26/171 (15.20%)	2,13E-02
MIP	3/8 (37.50%)	3,48E-02	PeptidaseS8	10/49 (20.41%)	2,22E-02
Methyltransf18	10/55 (18.18%)	3,60E-02	PP-binding	19/116 (16.38%)	2,24E-02
PQ-loop	4/14 (28.57%)	3,95E-02	Prenyltrans	5/17 (29.41%)	2,26E-02
Cu-oxidase3	3/9 (33.33%)	4,85E-02	ATP-syntC	3/7 (42.86%)	2,61E-02
LipaseGDSL2	3/9 (33.33%)	4,85E-02	EcKinase	3/7 (42.86%)	2,61E-02
PCMT	3/9 (33.33%)	4,85E-02	MFS2	3/7 (42.86%)	2,61E-02
Transferase	4/15 (26.67%)	4,98E-02	SnoaL4	3/7 (42.86%)	2,61E-02
Methyltransf11	11/66 (16.67%)	5,03E-02	Aminotran3	4/12 (33.33%)	2,61E-02

Na condição após o contato, devido a quantidade maior de genes, nota-se consequentemente uma variedade muito maior de domínios diferencialmente expressos. As enzimas do CAZy, com

função de degradação de polissacarídeos, aparecem com maior frequência, como GH18, LysM (quitinases), GH75 (chitosanases) com quase todos os domínios do genoma sendo expressos (4 de 5), PectateLyase3, WSC (domínio de ligação à carboidratos), GH71 (α -1,3-glucanase), GH28 (poligalacturonase), GH72 (β -1,3-glucanosiltransglicosilase), GH16 (endo-1,3- β -glucanase) e GH2 (β -manosidase). Proteases, por ser uma classe bem diversa, existem domínios tanto induzidos (Peptidase S8) como reprimidos (PeptidaseC14). Transportadores da do tipo MFS, os mais prevalentes em células eucarióticas, também foram induzidos. Nesta condição aparecem uma grande diversidade também de domínios reprimidos, como aqueles envolvidos no metabolismo secundário (Condensation, PP-binding, AMP-binding, Acyltransf e Methiltransf), transporte (principalmente transportadores ABC), e proteases.

Estes dados corroboram os dados encontrados na análise de agrupamento, onde temos uma predominância de genes do metabolismo secundário antes e durante o contato e enzimas hidrolíticas após o contato. Estas são duas estratégias bastante utilizadas no micoparasitismo para competição mais efetiva entre fungos. O isolado *T. harzianum* TR274 se mostrou capaz de modular sua expressão para ativar e inativar diferentes vias e conseqüentemente utilizar diferentes estratégias dependendo do estágio de sua interação com um fungo antagonista.

A critério de comparação entre as duas abordagens, parede celular e interação direta com *S. sclerotiorum*, foi feita uma comparação dois genes encontrados nas duas condições. Como a análise do transcrito crescendo em parede celular foi mapeado no genoma do isolado CBS 226.95 e da interação direta com o TR274 uma tabela de ortólogos foi construída entre os dois isolados. Dos 287 genes diferencialmente expressos na abordagem de parede celular, 249 (86,8%) genes também foram encontrados na interação direta, sua maioria na condição depois do contato. Isso mostra que a indução com parede celular simula somente uma situação de necrotrofia, onde o *Trichoderma* produz seu aparato proteico para consumo da parede celular disponível. A interação direta, como mostrado nas

análises, é um processo bem mais complexo que envolve a modulação de pelo menos 25% dos genes presentes no genoma de *T. harzianum* TR274.

CONCLUSÕES

As abordagens transcritômicas mostraram que a interação entre *T. harzianum* e *S. sclerotiorum* é bem complexa e envolve a produção de metabólitos secundários e síntese de transportadores antes e durante o contato, com uma modulação principalmente de enzimas hidrolíticas após o contato. Dos genes encontrados diferencialmente expressos na condição de crescimento em parede celular, 86,8% foram também encontrados diferencialmente expressos na interação direta. Cerca de 25% de todo o genoma de *T. harzianum* foi modulado durante a interação com *S. sclerotiorum*.

Este trabalho nos dá algumas direções de como aconteceu a evolução do gênero *Trichoderma* no aspecto genômico e transcritômico. Uma perspectiva é aumentar da quantidade de espécies sequenciadas para uma análise mais robusta dentro do gênero e uma comparação mais acurada com gêneros próximos filogeneticamente e um melhor entendimento da origem dessa dualidade micoparasita/saprófito. Assim como entender melhor a interação de fungos do gênero *Trichoderma* não só com fungos, mas com outros organismos, como por exemplo nematóides, que causam grandes perdas na agricultura do Brasil e do mundo, como também as próprias plantas hospedeiras em um sistema *Trichoderma*/Patógeno/Planta.

REFERÊNCIAS BIBLIOGRÁFICAS

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle CA, Rohwer F. (2006) The marine viromes of four oceanic regions. *PLoS Biol.* 4(11):368.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology: the gene ontology consortium. *Nat Genet.* 25(1):25–29.

Asmann YW, Wallace MB, Thompson EA. (2008) Transcriptome profiling using next-generation sequencing. *Gastroenterology.* 135(5):1466-8.

Atanasova L, Le Crom S, Gruber S, Couplier F, Seidl-Seiboth V, Kubicek CP, Druzhinina IS (2013) Comparative transcriptomics reveals different strategies of *Trichoderma* mycoparasitism. *BMC Genomics.* 22:114-121.

Attanayake RN, Carter PA, Jiang D, Del Río-Mendoza L, Chen W (2013) *Sclerotinia sclerotiorum* populations infecting canola from China and the United States are genetically and phenotypically distinct. *Phytopathology.* 103(7):750-761.

Baroncelli R, Zapparata A, Piaggeschi G, Sarrocco S, Vannacci G. (2016) Draft Whole-Genome Sequence of *Trichoderma gamsii* T6085, a Promising Biocontrol Agent of *Fusarium* Head Blight on Wheat. *Genome Announc.* 4(1):e01747-15.

Birney E, Clamp M, Durbin R. (2004) GeneWise and genomewise. *Genome Res.* 14:988-995.

Bisset J. (1991a) A revision of the genus *Trichoderma* II. Infrageneric classification. *Canadian Journal of Botany* 68:2357-2372.

Bisset J. (1991b) A revision of the genus *Trichoderma* III. Section *Pachybasium* sect. nov. *Canadian Journal of Botany* 69:2373-2417.

Bissett J, Gams W, Jaklitsch W, Samuels GJ (2015) Accepted *Trichoderma* names in the year 2015. *IMA Fungus.* 6(2):263–295.

Blake JA, Dolan M, Drabkin H, Hill DP, Li N, Sitnikov D, Bridges S, Burgess S, Buza T, McCarthy F, *et al.* (2013) Gene ontology annotations and resources. *Nucleic Acids Res.* 41:530-535.

Boland GJ, Hall R (1994) Index of plant hosts of *Sclerotinia sclerotiorum*. *Canadian Journal Plant Pathology*, 16(2):93-108.

Boysen C, Simon MI, Hood L. (1997) Fluorescence-based sequencing directly from bacterial and P1-derived artificial chromosomes. *Biotechniques*. 23(6):978-82.

Brito JP, Ramada MH, de Magalhães MT, Silva LP, Ulhoa CJ (2014) Peptaibols from *Trichoderma asperellum* TR356 strain isolated from Brazilian soil. *Springerplus*. 13(3):600.

Buist G, Steen A, Kok J, Kuipers OP. (2008) LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol Microbiol*. 68(4):838-47.

Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB. 2008. ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Research* 18: 810–20.

Carvalho DDC, Geraldine AM, Lobo Junior M, Mello SCM (2015) Biological control of white mold by *Trichoderma harzianum* in common bean under field conditions. *Pesq. agropec. bras.* 50(12):1220-1224.

Carvalho, D.D.C., Mello, S.C.M., Lobo Júnior, M., Silva, M.C. (2011) Control of *Fusarium oxysporum* f.sp.phaseoli in vitro and on seeds and growth promotion of common bean in early stages by *Trichoderma harzianum*. *Tropical plant pathology*, 36.

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17(4):540–552.

Chaverri, P., Samuels, G.J. & Stewart, E.L. (2001) *Hypocrea virens* sp. nov., the teleomorph of *Trichoderma virens*. *Mycologia* 93:1113-1124.

Chaverri P, Samuels GJ. (2003) *Hypocrea/Trichoderma* (*Ascomycota*, *Hypocreales*, *Hypocreaceae*): species with green ascospores. *Studies in Mycology*. 48:1–116.

Daryaei A, Jones EE, Ghazalibiglar H, Glare TR, Falloon RE (2016) Effects of temperature, light and incubation period on production, germination and bioactivity of *Trichoderma atroviride*. *J Appl Microbiol*.

Davies K. (2001) Decifrando o genoma: a corrida para desvendar o DNA humano. COMPANHIA DAS LETRAS.

De Bie T, Cristianini N, Demuth JP, Hahn MW (2006) CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* 22(10):1269–1271.

Dhiman N, Smith DI, Poland GA. (2009) Next-generation sequencing: a transformative tool for vaccinology. *Expert Rev Vaccines*. 8(8):963-7.

Dodd SL, Lieckfeldt E, Samuels GJ (2003) *Hypocrea atroviridis* sp. nov., the teleomorph of *Trichoderma atroviride*. *Mycologia* 95:27–40.

Druzhinina IS, Kopchinskiy AG. (2006) TrichOKEY v. 2 - a DNA oligonucleotide BarCode program for the identification of multiple sequences of *Hypocrea* and *Trichoderma*. Proceedings of the 8th International Mycological Congress, Cairns, Australia.

Druzhinina IS, Kubicek CP, Komoń-Zelazowska M, Mulaw TB, Bissett J (2010) The *Trichoderma harzianum* demon: complex speciation history resulting in coexistence of hypothetical biological species, recent agamospecies and numerous relict lineages. *BMC Evol Biol.* 1;10:94.

Druzhinina, I.S., Seidl-Seiboth, V., Herrera-Estrella, A., Horwitz, B.A., Kenerley, C.M., Monte, E., Mukherjee, P.K., Zeilinger, S., Grigoriev, I.V., Kubicek, C.P. (2011) *Trichoderma*: the genomics of opportunistic success. *Nature Reviews.* 9:749-759.

Dubey MK, Ubhayasekera W, Sandgren M, Jensen DF, Karlsson M. (2012) Disruption of the Eng18B ENGase gene in the fungal biocontrol agent *Trichoderma atroviride* affects growth, conidiation and antagonistic ability. *PLoS One.* 7(5):36152.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792-1797.

Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, Saar MO, Alexander S, Alexander EC Jr, Rohwer F. (2006) Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics.* PMC1483832.

Geraldine AM, Lopes FAC, Carvalho DDC, Barbosa ET, Rodrigues AR, Brandão RS, Ulhoa CJ, Lobo-Junior M (2013) Cell wall-degrading enzymes and parasitism of sclerotia are key factors on field biocontrol of white mold by *Trichoderma* spp. *Biol Control.* 67(3):308-316.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 15;29(7):644-52.

Grigoriev IV, Cullen D, Goodwin SB, Hibbett D, Jeffries TW, Kubicek CP, Kuske C, Magnuson JK, Martin F, Spatafora JW, Tsang A, Baker SE. (2011) Fueling the future with fungal genomics. *Mycology.* 2(3):192-209.

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42(1):699-704.

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*, 40:26–32.

Gurevich A, Saveliev V, Vyahhi N and Tesler G. (2013) QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29 (8): 1072-1075.

Guryev V, Cuppen E. (2009) Next-generation sequencing approaches in genetic rodent model systems to study functional effects of human genetic variation. *FEBS Lett.* 5;583(11):1668-73.

Hartl L, Zach S, Seidl-Seiboth V (2012) Fungal chitinases: diversity, mechanistic properties and biotechnological potential. *Appl Microbiol Biotechnol.* 93(2):533–543.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, et al. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.* 33:116-120.

Imelfort M, Duran C, Batley J, Edwards D. (2009) Discovering genetic polymorphisms in next-generation sequencing data. *Plant Biotechnol J.* 7(4):312-7.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40:109-114.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. (2015) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, Vol. 44, 457–462.

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30(14):3059–3066.

Khalidi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND (2010) SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genet Biol.* 47(9):736–741.

Kindermann J, El-Ayouti Y, Samuels GJ, Kubicek CP. (1998) Phylogeny of the genus *Trichoderma* based on sequence analysis of the internal transcribed spacer region 1 of the rDNA cluster. *Fungal Genetics and Biology.* 24:298–309.

Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, et al. (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* 5:7.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305:567-580.

Kubicek C.P., Komon-Zelazowska M., Sándor E., Druzhinina, I.S. (2007) Facts and challenges in the understanding of the biosynthesis of peptaibols by *Trichoderma*. *Chemistry Biodiversity*. 4 (6).1068-1082.

Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, Mukherjee M, Kredics L, Alcaraz LD, Aerts A, Antal Z, Atanasova L, Cervantes-Badillo MG, Challacombe J, Chertkov O, McCluskey K, Coulpier F, Deshpande N, von Döhren H, Ebbole DJ, Esquivel-Naranjo EU, Fekete E, Flipphi M, Glaser F, Gómez-Rodríguez EY, Gruber S, Han C, Henrissat B, Hermosa R, Hernández-Oñate M, Karaffa L, Kosti I, Le Crom S, Lindquist E, Lucas S, Lübeck M, Lübeck PS, Margeot A, Metz B, Misra M, Nevalainen H, Omann M, Packer N, Perrone G, Uresti-Rivera EE, Salamov A, Schmoll M, Seiboth B, Shapiro H, Sukno S, Tamayo-Ramos JA, Tisch D, Wiest A, Wilkinson HH, Zhang M, Coutinho PM, Kenerley CM, Monte E, Baker SE, Grigoriev IV. (2011) Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. *Genome Biology*, 12(4):40.

Kuhls K, Lieckfeldt E, Samuels GJ, Meyer W, Kubicek CP, Börner T. (1997) Revision of *Trichoderma* sect. *Longibrachiatum* Including Related Teleomorphs Based on Analysis of Ribosomal DNA Internal Transcribed Spacer Sequences. *Mycologia*. 89(3):442-460.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL (2004) Versatile and open software for comparing large genomes. *Genome Biology*, 5(2):12

Langmead B, Salzberg S. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 9:357-359.

Liu, P.G., Yang, Q. (2005) Identification of genes with a biocontrol function in *Trichoderma harzianum* mycelium using the expressed sequence tag approach. *Research Microbiology*, 156(3):416-423.

Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B. (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 40:622-7.

Lopes FA, Steindorff AS, Geraldine AM, Brandão RS, Monteiro VN, Lobo M Jr, Coelho AS, Ulhoa CJ, Silva RN (2012) Biochemical and metabolic profiles of *Trichoderma* strains isolated from

common bean crops in the Brazilian Cerrado, and potential antagonism against *Sclerotinia sclerotiorum*. Fungal Biol. 116(7):815–824.

Lorito M, Farkas V, Rebuffat S, Bodo B, Kubicek CP. (1996) Cell Wall Synthesis Is a Major Target of Mycoparasitic Antagonism by *Trichoderma harzianum*. J Bacteriol, 178(21):6382–6385.

Lorito M, Woo SL, Harman GE, Monte E (2010) Translational research on Trichoderma: from Omics to the field. Ann Rev Phytopathol. 48:395-417.

Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. (2013) Gene duplication as a major force in evolution. J. Genet. 92,155–161.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature. 15;437(7057):376-80.

Martin K., Mcdougall B.M., Mccilroy S., Jayus, Chen J. E., Seviour R.J (2007). Biochemistry and molecular biology of exocellular fungal β -(1,3)- and β -(1,6)-glucanases. FEMS Microbiology Reviews. 31:168–192.

Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EG, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barabote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*) Nat Biotechnol. 26:553–560.

Melo IS, & Azevedo JL. (2000). Controle biológico, Jaguariúna: EMBRAPA, Meio Ambiente. p. 388.

Mendoza JL, Pérez MI, Prieto JM, Velásquez JD, Olivares JG, Langarica HR (2015) Antibiosis of *Trichoderma* spp strains native to northeastern Mexico against the pathogenic fungus *Macrophomina phaseolina*. Braz J Microbiol. 46(4):1093-101.

Morozova O, Marra MA. (2008) From cytogenetics to next-generation sequencing technologies: advances in the detection of genome rearrangements in tumors. *Biochem Cell Biol.* 86(2):81-91.

Oliveira MB (2011) Caracterização da expressão de inibidores de poligalacturonases (PGIPs) em resposta aos estresses biótico e abiótico em plantas de feijão comum (*Phaseolus vulgaris* L.) Dissertação de Mestrado - UFG.

Ozsolak F, Milos PM. (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet.* 12 (2): 87-98.

Patron NJ, Waller RS, Cozijnsen AJ, Straney DC, Gardiner DM, Nierman WC, Howlett BJ. (2007) Origin and distribution of epipolythiodioxopiperazine (ETP) gene clusters in filamentous ascomycetes. *BMC Evol Biol.* 7:174.

Pereira JL, Queiroz RM, Charneau SO, Felix CR, Ricart CA, da Silva FL, Steindorff AS, Ulhoa CJ, Noronha EF (2014) Analysis of *Phaseolus vulgaris* response to its association with *Trichoderma harzianum* (ALL-42) in the presence or absence of the phytopathogenic fungi *Rhizoctonia solani* and *Fusarium solani*. *PLoS One.* 30;9(5).

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods.* 8:785-786.

Pires, C.S.S., Fontes, E.M.G., Sujii, E.R. (2003). Impacto ecológico de plantas geneticamente modificadas: O algodão como estudo de caso. Embrapa Recursos Genéticos e Biotecnologia.

Poletto I, Muniz, MFB. Seconi DE, Santin D, Weber MND, Blume E. (2006) Zoneamento e identificação de *Fusarium* spp. causadores de podridão de raízes em plantios de erva-mate (*Ilexparaguariensis* a. st.-hil.) na região do vale do Taquarí, RS. *Ciência Florestal, Santa Maria.* 16(1): 1-10.

Price A.L., Jones N.C. and Pevzner P.A. (2005) De novo identification of repeat families in large genomes. To appear in Proceedings of the 13 Annual International conference on Intelligent Systems for Molecular Biology (ISMB-05).

Price, M.N., Dehal, P.S., and Arkin, A.P. (2009) FastTree: Computing Large Minimum-Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution* 26:1641-1650.

Punja, Z.K. & Utkhede, R.S. (2003). Using fungi and yeasts to manage vegetable crop diseases. *Trends Biotechnology* 21(9):400-7.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, et al. (2012) The Pfam protein families database. *Nucleic Acids Res.* 40:290-301.

Purdy, L.H. (1979) *Sclerotinia sclerotiorum*: history, diseases and symptomatology, host range, geographic distribution, and impact. *Phytopathology* 69:875-880.

Ramada MH, Steindorff AS, Bloch C Jr, Ulhoa CJ (2015) Secretome analysis of the mycoparasitic fungus *Trichoderma harzianum* ALL 42 cultivated in different media supplemented with *Fusarium solani* cell wall or glucose. *Proteomics*. 16(3):477-490.

Reithner, B., Ibarra-Laclette, E., Mach, R.L., Herrera-Estrella, A. (2011) Identification of mycoparasitism-related genes in *Trichoderma atroviride*. *Applied and Environmental Microbiology*. 77(13):4361-70.

Revell LJ (2009) Size-correction and principal components for interspecific comparative studies. *Evolution*. 63(12):3258–3268.

Rifai MA (1969). A revision of the genus *Trichoderma*. *Mycological Papers*.

Rubio MB, Domínguez S, Monte E, Hermosa R (2012) Comparative study of *Trichoderma* gene expression in interactions with tomato plants using high-density oligonucleotide microarrays. *Microbiology*. 158: 19-128.

Salamov AA, Solovyev VV. Ab initio gene finding in *Drosophila* genomic DNA. (2000) *Genome Res*. 10:516-522.

Samolski I, de Luis A, Vizcaíno JA, Monte E, Suárez MB. (2009) Gene expression analysis of the biocontrol fungus *Trichoderma harzianum* in the presence of tomato plants, chitin, or glucose using a high-density oligonucleotide microarray. *BMC Microbiol*. 9: 217

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*. 1977(a) 74(12):5463-7.

Schuster SC. (2008) Next-generation sequencing transforms today's biology. *Nature Methods* 5:16-18.

Seidl V. (2008) Chitinases of filamentous fungi: a large group of diverse proteins with multiple physiological functions. *Fungal Biology Reviews*. 22:36-42.

Seidl V, Song L, Lindquist E, Gruber S, Koptchinskiy A, Zeilinger S, Schmoll M, Martinez P, Sun J, Grigoriev I, Herrera-Estrella A, Baker ES, Kubicek CP. (2009) Transcriptomic response of the mycoparasitic fungus *Trichoderma atroviride* to the presence of a fungal prey. *BMC Genomics* 10:567.

Slocombe PM, Smith M. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*.

Slot JC, Hibbett DS (2007) Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study. *PLoS One*, 2:1097.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc Natl Acad Sci USA*. 8;103(32):12115-20.

Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.

Steindorff AS, Silva RN, Coelho ASG, Nagata T, Noronha EF, Ulhoa CJ (2012) *Trichoderma harzianum* expressed sequence tags for identification of genes with putative roles in mycoparasitism against *Fusarium solani*. *Biological Control* 61(2):134-140.

Studholme DJ1, Harris B, Le Cocq K, Winsbury R, Perera V, Ryder L, Ward JL, Beale MH, Thornton CR, Grant M. Investigating the beneficial traits of *Trichoderma hamatum* GD12 for sustainable agriculture-insights from genomics. (2013) *Front Plant Sci*. 30(4):258.

Sun P, Yang XB (2000) Light, temperature, and moisture effects on Apothecium production of *Sclerotinia sclerotiorum*. *Plant Dis*. 84: 1287-1293.

Takaya N, Yamazaki D, Horiuchi H, Ohta A, Takagi M. (1998). Cloning and characterization of a chitinase-encoding gene (*chiA*) from *Aspergillus nidulans*, disruption of which decreases germination frequency and hyphal growth. *Bioscience, Biotechnology and Biochemistry* 62:60-65.

Ter-Hovhannisyanyan V, Lomsadze A, Chernoff YO, Borodovsky M. (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res*. 18:1979-1990.

Trapnell C, Pachter L, Salzberg SL. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105-1111.

Varshney RK, Nayak SN, May GD, Jackson SA. (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.*;27(9):522-30.

Venter JC, Smith HO, Hood L. (1996) A new strategy for genome sequencing. *Nature*. 30;381(6581):364-6.

Vinale F, Ghisalberti EL, Sivasithamparam K, Marra R, Ritieni A, Ferracane R, Woo S, Lorito M. (2009) Factors affecting the production of *Trichoderma harzianum* secondary metabolites during the interaction with different plant pathogens. *Lett Appl Microbiol*. 48:705–711.

Vizcaíno JA, Gonzáles FJ, Suárez MB, Redondo J, Heinrich J, Delgado-Jarana J, Hermosa R, Gutiérrez R, Monte E, Llobell A, Rey M. (2006) Generation, annotation and analysis of ESTs from *Trichoderma harzianum* CECT 2413. *BMC Genomics*, 7:193.

Vizcaíno JA, Redondo J, Suárez MB, Cardoza RE, Hermosa R, González FJ, Rey M, Monte E. (2007) Generation, annotation, and analysis of ESTs from four different *Trichoderma* strains grown under conditions related to biocontrol. *Applied Microbiology Biotechnology*, 75:853–862.

Voelkerding KV, Dames SA, Durtschi JD. (2009) Next-generation sequencing: from basic research to diagnostics. *Clin Chem*. 55(4):641-58.

Wang WC, Lin FM, Chang WC, Lin KY, Huang HD, Lin NS (2009) miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinform*. 10:328.

Watson JH, Crick FH. (1953) The structure of DNA. *Nature*. 18:123-31.

Weidling, R. (1934) Studies on lethal principle effective in the parasitic action of *Trichoderma lignorum* on *Rhizoctonia solani* and other soil fungi, *Phytopathology*. 24.

Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM. (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 17;452(7189):872-6.

Wommack KE, Bhavsar J, Ravel J. (2008) Metagenomics: read length matters. *Appl Environ Microbiol*. 74(5):1453-63.

Woo, S.L., Scala, F., Ruocco, M., Lorito, M. (2006) The molecular biology of the interactions between *Trichoderma* spp., phytopathogenic fungi, and plants. *Phytopathology*. 96:181–185.

Yang D, Pomraning K, Kopchinskiy A, Karimi Aghcheh R, Atanasova L, Chenthamara K, Baker SE, Zhang R, Shen Q, Freitag M, Kubicek CP, Druzhinina IS. (2015) Genome Sequence and Annotation of *Trichoderma parareesei*, the Ancestor of the Cellulase Producer *Trichoderma reesei*. *Genome Announc*. 13;3(4).

Zeilinger, S. Oman, F. (2005) Signal transduction by Tga3, a novel G protein alpha subunit of *Trichoderma atroviride*. *Applied Environment Microbiology*. 71:1591-1597.

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA (2013) The MaSuRCA genome assembler. *Bioinformatics*. 1;29(21):2669-77.

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*. 11(4):41.

ARTIGOS PUBLICADOS DURANTE O DOUTORADO

1. Lopes, Fabyano Alvares Cardoso; Steindorff, Andrei Stecca; Geraldine, Alaerson Maia; Brandão, Renata Silva; Monteiro, Valdirene Neves; Júnior, Murillo Lobo; Coelho, Alexandre Siqueira Guedes; Ulhoa, Cirano José; Silva, Roberto Nascimento. Biochemical and metabolic profiles of *Trichoderma* strains isolated from common bean crops in the Brazilian Cerrado, and potential antagonism against *Sclerotinia sclerotiorum*. *FUNGAL BIOL-UK*, v. 116, p. 815-824, 2012.
2. Steindorff, Andrei Stecca; Silva, Roberto do Nascimento; Coelho, Alexandre Siqueira Guedes; Nagata, Tatsuya; Noronha, Eliane Ferreira; Ulhoa, Cirano José. *Trichoderma harzianum* expressed sequence tags for identification of genes with putative roles in mycoparasitism against *Fusarium solani*. *Biological Control (Print)*, v. 61, p. 134-140, 2012.
3. Qualhato, Thiago Fernandes; Lopes, Fabyano Alvares Cardoso; Steindorff, Andrei Stecca; Brandão, Renata Silva; Jesuino, Rosália Santos Amorim; Ulhoa, Cirano José. Mycoparasitism studies of *Trichoderma* species against three phytopathogenic fungi: evaluation of antagonism and hydrolytic enzyme production. *Biotechnology Letters*, p. 1461-1468, 2013.
4. Vieira, Pabline Marinho; Coelho, Alexandre Siqueira; Steindorff, Andrei Stecca; De Siqueira, Saulo José; Silva, Roberto Do; Ulhoa, Cirano José. Identification of differentially expressed genes from *Trichoderma harzianum* during growth on cell wall of *Fusarium solani* as a tool for biotechnological application. *BMC Genomics*, v. 14, p. 177, 2013.
5. Steindorff, Andrei Stecca; Ramada, Marcelo Henrique; Coelho, Alexandre Siqueira; Miller, Robert Neil; Pappas, Georgios Joannis; Ulhoa, Cirano José; Noronha, Eliane Ferreira. Identification of mycoparasitism-related genes against the phytopathogen *Sclerotinia sclerotiorum* through transcriptome and expression profile analysis in *Trichoderma harzianum*. *BMC Genomics*, v. 15, p. 204, 2014.

6. Troian, Rogério Fraga; Steindorff, Andrei Stecca; Ramada, Marcelo Henrique Soller; Arruda, Walquiria; Ulhoa, Cirano José. Mycoparasitism studies of *Trichoderma harzianum* against *Sclerotinia sclerotiorum*: evaluation of antagonism and expression of cell wall-degrading enzymes genes. *Biotechnology Letters*, p. 2095-2101, 2014.
7. Pereira, Jackeline L.; Queiroz, Rayner M. L.; Charneau, Sébastien O.; Felix, Carlos R.; Ricart, Carlos A. O.; Da Silva, Francilene Lopes; Steindorff, Andrei Stecca; Ulhoa, Cirano J.; Noronha, Eliane F. Analysis of *Phaseolus vulgaris* Response to Its Association with *Trichoderma harzianum* (ALL-42) in the Presence or Absence of the Phytopathogenic Fungi *Rhizoctonia solani* and *Fusarium solani*. *Plos One*, v. 9, p. e98234, 2014.
8. Dos Santos Castro, Lilian; Pedersoli, Wellington Ramos; Antoniêto, Amanda Cristina; Steindorff, Andrei Stecca; Silva-Rocha, Rafael; Martinez-Rossi, Nilce M; Rossi, Antonio; Brown, Neil Andrew; Goldman, Gustavo H; Faça, Vitor M; Persinoti, Gabriela F; Silva, Roberto Nascimento. Comparative metabolism of cellulose, sophorose and glucose in *Trichoderma reesei* using high-throughput genomic and proteomic analyses. *Biotechnology for Biofuels*, v. 7, p. 41, 2014.
9. Freitas, Rachel Silveira; Steindorff, Andrei Stecca; Ramada, Marcelo Henrique Soller; Siqueira, Saulo José Linhares; Noronha, Eliane Ferreira; Ulhoa, Cirano José. Cloning and characterization of a protein elicitor Sm1 gene from *Trichoderma harzianum*. *Biotechnology Letters*, v. 36, p. 783-788, 2014.
10. Ramada, Marcelo Henrique Soller ; Steindorff, Andrei Stecca ; Bloch, Carlos ; Ulhoa, Cirano José. Secretome analysis of the mycoparasitic fungus *Trichoderma harzianum* ALL 42 cultivated in different media supplemented with *Fusarium solani* cell wall or glucose. *Proteomics (Weinheim. Print)*, v. 16, p. 477-490, 2016.