UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa Interunidades de Pós-Graduação em Bioinformática

Túlio Morgan

GENÔMICA E PROTEÔMICA DE FUNGOS FILAMENTOSOS PARA APLICAÇÕES BIOTECNOLÓGICAS

Belo Horizonte

2021

Túlio Morgan

GENÔMICA E PROTEÔMICA DE FUNGOS FILAMENTOSOS PARA APLICAÇÕES BIOTECNOLÓGICAS

Tese apresentada ao Programa Interunidades de Pós-graduação em Bioinformática da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutor em Bioinformática.

Orientador: Tiago Antônio de Oliveira Mendes

Co-orientadora: Valéria Monteze Guimarães

Belo Horizonte

2021

043 Morgan, Túlio.
Genômica e proteômica de fungos filamentosos para aplicações biotecnológicas [manuscrito] / Túlio Morgan. – 2021.
134 f. : il. ; 29,5 cm.

Orientador: Tiago Antônio de Oliveira Mendes. Co-orientadora: Valéria Monteze Guimarães.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Biologia Computacional. 2. Fungos. 3. Biotecnologia. 4. Genômica. 5. Proteômica. I. Mendes, Tiago Antônio de Oliveira. II. Guimarães, Valéria Monteze. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 573:004

Ficha catalográfica elaborada pela bibliotecária Fabiane C M Reis – CRB 6 - 2680

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa Interunidades de Pós-Graduação em Bioinformática da UFMG

ATA DA DEFESA DE TESE

TULIO MORGAN

Às oito horas do dia **01 de março de 2021**, reuniu-se, no aplicativo Zoom, a Comissão Examinadora de Tese, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho de **Tulio Morgan** intitulado: **"Genômica e proteômica de fungos filamentosos para aplicações biotecnológicas"**, requisito para obtenção do grau de Doutor em **Bioinformática.** Abrindo a sessão, o Presidente da Comissão, Dr. Tiago Antonio de Oliveira Mendes, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa do candidato. Logo após, a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	**Instituição**	**Indicação**
Dr. Tiago Antonio de Oliveira Mendes - Orientador	UFV	**Aprovado**
Dra. Valéria Monteze Guimarães - Coorientadora	UFV	**Aprovado**
Dr. Aristóteles Góes Neto	UFMG	**Aprovado**
Dr. Daniel Luciano Falkoski	Novozymes Latin America	**Aprovado**
Dr. Murillo Peterlini Tavares	UFV	**Aprovado**
Dra. Vera Lúcia dos Santos	UFMG	**Aprovado**

Pelas indicações, o candidato foi considerado: **Aprovado**

O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 01 de março de 2021.

Documento assinado eletronicamente por **Vera Lucia dos Santos**, **Servidor(a)**, em 01/03/2021, às 13:02, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Daniel Luciano Falkoski**, **Usuário Externo**, em 01/03/2021, às 13:02, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Murillo Peterlini Tavares**, **Usuário Externo**, em 01/03/2021, às 13:02, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Tiago Antônio de Oliveira Mendes**, **Usuário Externo**, em 01/03/2021, às 14:33, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Valéria Monteze Guimarães**, **Usuário Externo**, em 01/03/2021, às 16:18, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Aristoteles Goes Neto**, **Professor do Magistério Superior**, em 02/03/2021, às 09:31, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0584825** e o código CRC **195E2AAA**.

Referência: Processo nº 23072.209276/2021-46 SEI nº 0584825

Agradecimentos

Aos meus pais, aos meus irmãos e à Angélica Gouveia pelo amor e amizade, por sempre me apoiarem, me incentivarem e acreditarem no meu potencial.

Ao prof. Dr. Tiago Antônio de Oliveira Mendes, meu principal mentor e quem abriu-me a porta de entrada à Bioinformática. Fico agradecido por todas as ideias e orientação neste trabalho, seus esforços em estabelecer parcerias e colaborações que tanto me aprimoram como pesquisador em Bioinformática. Sempre serei grato pela confiança depositada.

À minha co-orientadora, profa. Dra. Valéria Monteze Guimarães, por toda a ajuda, orientação e amizade.

Ao Dr. Daniel Falkoski pela grande parceria em experimentos de cultivo de fungos, atividades enzimáticas e sacarificação de biomassa vegetal. Forneceu-me conhecimentos cruciais para realização deste trabalho.

Ao prof. Dr. Olinto Liparini e ao Fábio Custódio, pelo auxílio em experimentos e análises de classificação taxonômica de fungos.

Ao Murillo Peterlini, pela amizade e auxílio em experimentos de espectrometria de massa.

Ao CNPEM e LNBio, pelo uso das instalações de espectrometria de massa, e as técnicas Bianca Alves e Romênia Ramos pelo suporte na realização e análise proteômica.

Ao Eduardo e João Guilherme, pela amizade e suporte nas análises de HPLC.

Aos professores e funcionários do programa de pós-graduação em Bioinformática, pelo profissionalismo, dedicação e atenção aos alunos. Não é à toa que o programa é excelente!

Aos membros da banca pela disponibilidade e interesse em avaliarem este trabalho.

À CAPES, CNPq e FAPEMIG, pelo apoio financeiro.

Resumo

Os fungos compõem um vasto e diverso grupo de eucariotos, sendo que estimativas apontam a existência de milhões de espécies de fungos, com alta diversidade genética. Contudo, apenas uma pequena fração dessas espécies já foram catalogadas e um número ainda menor foi objeto de estudos científicos aprofundados. Dada sua importância no campo da biotecnologia para geração de uma variedade de produtos ou em bioprocessos, faz-se relevante estudos visando avaliar o potencial biotecnológico desses organismos ou dos metabólitos produzidos por eles. Nesse sentido, análises genômicas, proteômica e transcriptômica podem ser importantes para caracterização de fungos, bem como os processos biológicos nos quais estão fazendo parte.

Inicialmente, reportamos uma montagem de genoma e análise funcional de uma nova espécie endofítica do gênero *Trichoderma.* Análises filogenéticas indicaram que esse isolado era uma nova espécie, proposta neste estudo como *Trichoderma orchidacearum* sp. nov. Genômica comparativa, análises evolutivas, proteômica e ensaios de atividade de enzimas ativas em carboidratos, forneceram importantes informações a respeito da evolução e estratégias de nutrição desse fungo, indicando uma simplificação do genoma e direcionamento para estilo de vida primariamente endofítico.

Posteriormente, também realizamos experimentos de triagem de fungos ascomicetos e basidiomicetos quanto à capacidade de produzir atividades enzimáticas para despolimerização de biomassa vegetal, e o potencial dessas enzimas para aumentar a eficiência da mistura comercial Cellic® CTec2/HTec2 sacarificação da palha da cana-de-açúcar pré-tratada. O sobrenadante de cultivo de *Penicillium ochrochloron* RLS11 apresentou promissor efeito de suplementação de Cellic® CTec2/HTec2, de forma que conduzimos o sequenciamento do genoma desse fungo, análises genômicas e proteômicas. Foram identificadas diversas celulases (famílias GH3, GH6, GH7) e outras enzimas possivelmente relacionadas com efeito de suplementação da mistura enzimática comercial.

Por fim, obtivemos uma montagem do genoma de *F. verticillioides* AZB, importante fungo patogênico de milho, sorgo e arroz, além de produtor de micotoxinas (especialmente fumonisinas). Análises de ortologia e dados de expressão gênica indicaram muitos genes compartilhados por diversas espécies de *Fusarium* que podem participar da colonização de plantas. Detectamos seleção positiva em 1826 genes ortólogos de *F. verticillioides* e muitos deles apresentaram similaridade de sequência suficiente com os fatores de virulência (banco de dados PHI, http://www.phi-base.org/). Dessa forma, identificamos diversos genes codificadores de proteínas que tem potencial para participar de processos de patogenicidade. Isso pode orientar estudos futuros para aumentar o conhecimento a respeito da virulência do fungo e elaborar melhores estratégias de controle.

Abstract

Fungi comprise a vast and diverse group of eukaryotes, and estimates point to the existence of millions of species of fungi, with high genetic diversity. However, only a small fraction of these species have already been identified and an even smaller number has been subjected to in-depth scientific studies. Given its importance in the field of biotechnology for the generation of a variety of products or bioprocesses, it is important to carry out studies to evaluate the biotechnological potential of these organisms or the metabolites produced by them. In this sense, genomic, proteomic, and transcriptomic analyzes can be important for the characterization of fungi, as well as the biological processes in which they are taking part.

Initially, we reported a genome assembly and functional analysis of a new endophytic species of the genus *Trichoderma*. Phylogenetic analyzes indicated that this isolate was a new species, proposed in this study as *Trichoderma orchidacearum* sp. nov. Comparative genomics, evolutionary analysis, proteomics, and enzyme activity assays, provided important information regarding the genome evolution and nutritional strategies of this fungus, indicating genome streamlining and evolution towards an endophytic lifestyle.

Subsequently, we also conducted a screening of ascomycetes and basidiomycetes for the ability to produce enzymatic activities for depolymerization of plant biomass, and the potential of these enzymes to increase the efficiency of the commercial mixture Cellic® CTec2/HTec2 for saccharification of pretreated sugarcane straw. The culture supernatant of *Penicillium ochrochloron* RLS11 showed a promising supplementation effect in Cellic® CTec2/HTec2, and we conducted the genome sequencing of this fungus, genomic and proteomic analyzes. Several cellulases (families GH3, GH6, GH7) and other enzymes possibly related to the supplementation effect of the commercial enzyme mixture have been identified.

Finally, we obtained a genome assembly of *Fusarium verticillioides* AZB, an important pathogen of corn, sorghum, and rice. Orthology analyzes and gene expression data indicated many genes shared by *Fusarium* species that may participate in plant colonization. We detected positive selection in 1,826 *F. verticillioides* orthologous genes and many of them showed sufficient sequence similarity to virulence effectors (PHI database, http://www.phi-base.org/). In this way, we identified several genes encoding proteins that have the potential to participate in pathogenic processes. This can guide future studies to increase knowledge about the virulence of the fungus and develop better control strategies.

Sumário

1. Introdução

O reino Fungi é conhecido por ser extremamente diverso, compreendendo espécies que são fundamentais para os ecossistemas onde atuam como decompositores, mutualistas e patógenos. Apesar dessa importância biológica, os fungos são relativamente pouco explorados, já que a maioria das espécies ainda não foi descrita ou não é caracterizada em detalhes (Schmit & Mueller, 2007; Wu et al., 2019).

Estimativas apontam que o número de espécies de fungos existentes é na faixa de milhões, e essa magnitude pode estar ligada a enorme diversidade genética e a capacidade de colonizar a maioria dos habitats do planeta (Tedersoo et al., 2014; Wu et al., 2019). Apesar de existirem 3 tipos principais de estilos de vida, saprotrófico, mutualista e patogênico, a diversidade de estratégias nutricionais empregadas pelos fungos é notável, sendo que um mesmo organismo pode ajustar seu metabolismo e mecanismos de adquirir nutrientes ao longo do ciclo de vida, visando maior adaptabilidade e proliferação no ambiente. A mudança na estratégia de nutrição pode ser drástica a ponto do estilo de vida do fungo ser alterado, por exemplo, *Fusarium verticillioides* pode passar de um estilo de vida endofítico, para um estilo de vida patogênico onde causa necrose dos tecidos da planta para fazer uso dos nutrientes liberados. Além disso, essa espécie também pode atuar como saprotrófico (Blacutt et al., 2018; Brown et al., 2014).

Independente do estilo de vida, os fungos de uma forma geral dependem de enzimas extracelulares para despolimerizar polímeros complexos e captar moléculas mais simples de modo a satisfazer suas necessidades nutricionais. Além disso, eles geralmente produzem pequenas moléculas denominadas metabólitos secundários, que geram vantagens adaptativas como por exemplo eliminar ou retardar o crescimento de organismos competidores (Richards & Talbot, 2013). Estando sujeitos a inúmeras forças evolutivas, os genomas e genes dos fungos são continuamente modelados para aprimorar sua adaptabilidade ao ambiente (Stajich, 2017). Tais eventos influenciam diretamente o repertório de metabólitos secundários e enzimas extracelulares que podem ser produzidas por esses organismos (Bills & Gloer, 2016; Zhao, Liu, Wang, & Xu, 2014).

Alguns fungos filamentosos são agentes fundamentais no processo de degradação de material vegetal no ambiente, uma característica altamente dependente da capacidade de produzir e secretar enzimas ativas em carboidratos (CAZymes). A eficiência dos fungos nesses processos resultou em grande interesse da comunidade acadêmica e industrial com relação às CAZymes, uma vez que podem ser aplicadas no processo de conversão de polímeros estruturais de plantas, a fonte de carbono mais abundante e renovável do planeta, em energia, alimentos ou uma diversidade de

insumos químicos (álcoois, ácidos orgânicos). Tais processos/produtos compõem uma economia circular, conferindo sustentabilidade à geração de energia e materiais.

Projeções recentes apontam para um aumento da população mundial assim como desenvolvimento econômico e industrial, o que tende a elevar a demanda por energia e materiais (EIA, 2019b, 2019a; OECD, 2018; UN, 2017) e estimular pesquisas e serviços relacionados à geração desses produtos. Apesar de ser a principal fonte de energia mundial, os combustíveis fósseis não estão disponíveis diretamente para todos os países. Portanto, os mesmos devem importar esses materiais, o que pode criar uma relevante fonte de débitos e impactos na economia local (Kilbane, 2016). Além disso, existe uma crescente preocupação com o uso de combustíveis fósseis devido aos seus efeitos prejudiciais sobre os ecossistemas e à saúde humana. Além de aumentar a concentração atmosférica de gases nocivos, os materiais a base de petróleo não são facilmente degradáveis e permanecem por muitas décadas no ambiente poluindo solo e águas (Rhodes, 2018; Vijayavenkataraman, Iniyan, & Goic, 2012).

Tais impactos econômicos e ambientais causados pelo uso de materiais fósseis poderiam em parte ser mitigados pelo emprego da biotecnologia. Este é um campo interdisciplinar da ciência, que pode ser definido como o uso de sistemas biológicos vivos ou seus metabólitos para modificar ou criar processos e produtos, e vem sendo rapidamente desenvolvido e aplicado para a geração sustentável de biocombustíveis e biomateriais (Fasciotti, 2017), por exemplo, com a produção de álcool combustível a partir de biomassas vegetais.

Embora abundante, as biomassas vegetais são resistentes à degradação por sistemas biológicos. Estes materiais são compostos basicamente de fibras de celulose dispostas em uma estrutura cristalina estável, e revestidas por hemiceluloses e lignina. Como a parede celular de plantas é altamente variável em estrutura e composição, é necessário um amplo e diversificado conjunto de CAZymes atuando em sinergia no processo de despolimerização enzimática.

A produção industrial e comercialização de CAZymes, principalmente celulases, estão aumentando anualmente e são dominadas por poucas empresas incluindo Novozymes e DuPont/Genencor (Jayasekara & Renuka, 2016). Essas misturas enzimáticas são constituídas de proteínas provenientes basicamente de *Trichoderma* spp., *Aspergillus* spp. e *Penicillium* spp., uma vez que são eficientes produtores de celulases e hemicelulases (Bischof, Ramoni, & Seiboth, 2016; Vaishnav et al., 2018). Embora já aplicadas em processos industriais, as misturas enzimáticas estão em constante aprimoramento (Sun et al., 2018), tanto por engenharia de proteínas quanto pela adição de enzimas de outras fontes fúngicas, contribuindo para a redução da carga enzimática e do custo do processo de sacarificação de biomassa.

Devido ao grande número e diversidade de espécies de fungos, é possível que muitas CAZymes ainda serão descobertas e essas enzimas podem possuir novas funcionalidades ou propriedades bioquímico-cinéticas superiores. Nesse sentido, o uso de genômica, transcriptômica, proteômica e ensaios funcionais incluindo atividades enzimáticas e de aplicação em sacarificação de biomassas são de grande valia para a prospecção e caracterização de fungos e seu conjunto de enzimas extracelulares. Essas técnicas aplicadas em combinação podem fornecer uma profunda caracterização das CAZymes, bem como o impacto dessas enzimas no processo de sacarificação de biomassa vegetal. A identificação de enzimas que promovem melhor desempenho de sacarificação pode levar à redução do requerimento dessas enzimas, o que reduz custos e favorece a implementação de bioprocessos em escala industrial.

De forma similar à geração de energia e materiais, atualmente a geração de alimentos se baseia em grande parte em práticas agrícolas que podem levar a impactos negativos no meio ambiente e na saúde humana, principalmente devido ao uso de fertilizantes químicos e agrotóxicos (F. P. Carvalho, 2006). Além disso, projeta-se um aumento da demanda per capita por alimentos nas próximas décadas impulsionada pelo aumento da população mundial, tornando necessária a expansão da produção de alimentos. Nesse sentido, o uso de fungos filamentosos é de grande interesse como parte de manejo agrícola sustentável. Isso inclui biocontrole de doenças de plantas (fungos fitopatogênicos, fitonematoides, insetos), biofertilizantes e promoção do crescimento vegetal que contribui para redução do tempo e aumento de produtividade vegetal.

O uso de produtos a base de fungos para a agricultura vem crescendo continuamente, assim como o número de produtos disponíveis no mercado (Kaewchai, Soytong, & Hyde, 2009). Dentre esses fungos, *Trichoderma* spp. são de grande interesse, sendo amplamente utilizados e eficazes para controle biológico de pragas e como biofertilizante (Topolovec-Pintarić, 2016). Os mecanismos de biocontrole e promoção do crescimento vegetal conduzidos pelos fungos são mediados por efetores extracelulares, como enzimas hidrolíticas (quitinases, β-1,3(6)-glucanases, proteases) e metabólitos secundários (toxinas, análogos de fitohormônios, ácidos orgânicos) (Ghorbanpour et al., 2018), de modo que esses efetores ou os próprios fungos podem ser usados em práticas agrícolas sustentáveis.

Enquanto algumas fungos atuam como agentes de biocontrole de pragas, inúmeras outros são fitopatógenos, causando prejuízos em importantes culturas agrícolas, comprometendo a produtividade e lucratividade dos produtores. Além disso, fungos fitopatogênicos tendem a produzir toxinas como parte do processo de patogenicidade (Schmidt & Panstruga, 2011), substâncias que contaminam alimentos e podem causar problemas à saúde de pessoas e animais (Palumbo et al.,

2020). Dessa forma, é de grande interesse o desenvolvimento e aprimoramento de estratégias de manejo visando mitigar o efeito negativo de fungos fitopatogênicos.

Nesse contexto, a patogenômica vem sendo usada como importante ferramenta para ganhar entendimento das bases genéticas da patogenicidade de fungos. A partir da identificação de genes de virulência e seus reguladores, bem como análises evolutivas e genômica comparativa, é possível indicar alvos para desenvolvimento de estratégias de manejo mais eficientes e maior entendimento da biologia do agente patogênico e sua interação com o hospedeiro (Rampersad, 2020).

Diante disso, a biotecnologia é um dos pilares para o desenvolvimento industrial e tecnológico sustentáveis, e sua importância cresce continuamente. Esse campo foi altamente desenvolvido nos últimos anos devido a progressos tecnológicos, como sequenciamento de DNA/RNA de alto desempenho, proteômica baseada em cromatografia líquida acoplada a espectrometria de massa e novas abordagens de biologia molecular e bioquímica para caracterização de biomoléculas e edição de genomas. Juntamente com os avanços computacionais e bioinformática, uma enorme quantidade de dados biológicos vem sendo gerada, o que amplia a compreensão de processos biológicos e de organismos como um todo (L. M. de Carvalho et al., 2019).

As ferramentas computacionais são fundamentais para desvendar as relações entre características genômicas e o fenótipo em escala global. A identificação de genes codificadores de proteínas e a genômica comparativa com relação a enzimas ativas a carboidratos, proteases e genes de biossíntese de metabólitos secundários, podem fornecer indícios de estratégias nutricionais de fungos, podendo revelar seu potencial biotecnológico. Além disso, estudos de estrutura e evolução dos genomas podem ser relevantes para caracterização profunda de espécies, formular hipóteses sobre o estilo de vida do organismo e orientar melhorias em aplicações biotecnológicas (Wilken et al., 2019).

Além da genômica, a proteômica baseada em espectrometria de massa é interessante para fornecer uma visão geral das proteínas produzidas por um organismo sob condições específicas de cultivo. A proteômica *bottom-up*, a estratégia mais usada atualmente, tem sido muito favorecida pelo grande volume de dados genômicos disponíveis, uma vez que facilita a correlação entre proteínas e os peptídeos detectados por espectrometria de massa (De Oliveira & De Graaff, 2011).

Com isso, a genômica e a proteômica associadas a ensaios biológicos são metodologias poderosas para caracterizar organismos e processos biológicos. Isso aprimora nosso conhecimento sobre esses sistemas, impulsionando melhorias na biotecnologia aplicada à agricultura e produção de energia sustentáveis.

2. Objetivos

Caracterização de genomas e proteomas, e analises evolutivas dos fungos *Trichoderma orchidacearum* COAD 3006, *Penicillium ochrochloron* RLS11 *Fusarium verticillioides* AZB de forma a ganhar entendimento sobre a biologia do fungo e suas potencialidades biotecnológicas. Foram conduzidos sequenciamento e análise de genomas, análises proteômicas, ensaios funcionais com enzimas ativas em carboidratos, bem como testes em casa de vegetação para avaliar o potencial agronômico (biocontrole de pragas e promoção de crescimento vegetal).

2.1 Objetivos específicos

2.1.1 Capítulo I

i. Avaliar a capacidade de *Trichoderma orchidacearum* COAD 3006 para atuar como fungo saprófito (cultivo em biomassa lignocelulósica);
ii. Obter uma montagem do genoma de *Trichoderma orchidacearum* COAD 3006;
iii. Ganhar entendimento sobre evolução de *Trichoderma orchidacearum* COAD 3006 e a relação com estilo de vida;
iv. Caracterização do exoproteoma utilizando espectrometria de massa (LC-MS/MS).

2.1.2 Capítulo II

i. Selecionar fungos filamentosos com maior potencial de produção de enzimas ativas em carboidratos e eficiência para suplementar a mistura enzimática comercial Cellic CTec2/HTec2 (Novozymes) na sacarificação de palha de cana-de-açúcar pré-tratada;
ii. Obter uma montagem do genoma de *Penicillium ochrochloron* RLS11;
iii. Caracterização do genoma, especialmente com relação a enzimas ativas em carboidratos;
iv. Caracterização do exoproteoma utilizando espectrometria de massa (LC-MS/MS), com foco em enzimas ativas em carboidratos.

2.1.3 Capítulo III

i. Obter uma montagem do genoma de *Fusarium verticillioides* AZB;
ii. Comparar genomas e conjuntos de genes codificadores de proteínas em diversas especies de *Fusarium*;

iii. Detectar genes possivelmente relacionados com patogenicidade, fazendo uso de bancos de dados de fatores de virulência e dados públicos de expressão de genes *in planta*;
iv. Caracterização de regiões repetitivas e relação com estilo de vida de *F. verticillioides* e evolução do genoma.

2.1.4 Anexo I

i. Avaliar a capacidade de *Trichoderma orchidacearum* COAD 3006 para promover crescimento vegetal e biocontrole de nematoide das galhas (*Meloidogyne javanica*) a partir de ensaios em casa de vegetação.

3. Delineamento da Tese

O presente trabalho está dividido em 3 capítulos, sendo apresentados na forma de artigos e já submetidos em periódicos indexados. A Tese ainda possui o Anexo I, que consistiu de ensaios para avaliar o potencial agronômico de *T. orchidacearum* COAD 3006:

- Capítulo I: Genome sequencing and evolutionary insights of a new endophytic *Trichoderma* species from a basal phylogenetic position

- Capítulo II: Carbohydrate-active enzymes from *Penicillium ochrochloron* RLS11 to improve commercial enzyme mixtures for plant biomass saccharification

- Capítulo III: Whole-genome sequence of *Fusarium verticillioides* AZB and insights into genomic evolutionary events and features associated to pathogenicity

- Anexo I: Ensaio em casa de vegetação para avaliar a capacidade de *Trichoderma orchidacearum* COAD 3006 para promover crescimento vegetal e biocontrole de nematoide das galhas (*Meloidogyne javanica*)

4. Capítulos

Capítulo I

Genome sequencing and evolutionary insights of a new endophytic *Trichoderma* species from a basal phylogenetic position

Túlio Morgan [a,b], Fábio Alex Custódio [c], Olinto Liparini Pereira [c], Valéria Monteze Guimarães [b], Tiago Antônio de Oliveira Mendes [b]

[a] Postgraduate program in Bioinformatics, Universidade Federal de Minas Gerais, Av. Presidente Antônio Carlos, 6627 - Pampulha, 31270-901, Belo Horizonte, Minas Gerais, Brazil
[b] Department of Biochemistry and Molecular Biology, Universidade Federal de Viçosa, Av. PH Rolfs, s/n, 36570-900, Viçosa, Minas Gerais, Brazil
[c] Department of Phytopathology, Universidade Federal de Viçosa, Av. PH Rolfs, s/n, 36570-900, Viçosa, Minas Gerais, Brazil

Genome sequencing and evolutionary insights of a new endophytic *Trichoderma* species from a basal phylogenetic position

Túlio Morgan [a,b], Fábio Alex Custódio [c], Olinto Liparini Pereira [c], Valéria Monteze Guimarães [b], Tiago Antônio de Oliveira Mendes [b,*]

[a] Postgraduate program in Bioinformatics, Universidade Federal de Minas Gerais, Av. Presidente Antônio Carlos, 6627 - Pampulha, 31270-901, Belo Horizonte, Minas Gerais, Brazil

[b] Department of Biochemistry and Molecular Biology, Universidade Federal de Viçosa, Av. PH Rolfs, s/n, 36570-900, Viçosa, Minas Gerais, Brazil

[c] Department of Phytopathology, Universidade Federal de Viçosa, Av. PH Rolfs, s/n, 36570-900, Viçosa, Minas Gerais, Brazil

* Corresponding author. Tel.: +55 (31) 3612-5107. Email address: tiagoaomendes@ufv.br (T.A.O. Mendes)

Abstract

The cosmopolitan fungal genus *Trichoderma* contains species commonly found in other fungi, on plant litter or living plants, showing parasitic, saprophytic, or symbiotic lifestyles. Such versatility of nutritional strategies and habitats is a consequence of its genomic repertoire, mainly regarding protein-coding genes, giving rise to great species diversity. Currently, there is a large number of accepted species names as *Trichoderma* and probably many remain to be discovered and described. Here we report the whole-genome sequence and functional analysis of a new endophytic *Trichoderma* species, and insights on genome evolution and nutritional strategies, especially regarding carbohydrate-active enzymes related to plant cell wall-degradation (pcwdCAZymes). Phylogenetic analysis indicated that this *Trichoderma* isolate was a new species, proposed in this study as *Trichoderma orchidacearum* sp. nov. Genome streamlining was the main event driving its evolution as indicated by the reduced repertoire of protein-coding genes (8,903 genes) and other comparative genomic analyses. The profiles of gene gains and losses showed massive gene losses related to gene expression regulation, signal transduction, oxidoreductase and hydrolytic activities, similar events that occurred for the common ancestor of the Longibrachiatum clade, and thus, probably related to nutritional strategies and ecological niches of *T. orchidacearum*. The exoproteome analysis combined with plant cell wall-degrading enzyme activities evidenced the ability of *T. orchidacearum* to feed on plant biomass, despite the large imbalance between cellulolytic and xylanolytic enzymes. This suggested a poor fit to saprophytism compared to other *Trichoderma* species, such as *T. reesei* and *T. longibrachiatum*, and further indicated the *T. orchidacearum* evolution towards an endophytic lifestyle.

Keywords: Evolution; Fungal endophyte; Hypocreales; Nutritional strategies; Orchidaceae; Whole-genome sequencing.

1. Introduction

Fungal species belonging to *Trichoderma* genus are ubiquitous in natural environments and wide climate zone, playing an important role in the equilibrium of ecosystems. Such wide distribution is accompanied by lifestyle diversity, being found mainly parasitizing other fungi (fungicolous), but also colonizing plants (as endophytes), and in the soil feeding on plant litter (saprotrophic). One of the main reasons for the great interest on this fungal genus is the ability of some *Trichoderma* species to antagonize plant pathogens, such as other fungi and nematodes, using competition for the substrate, antibiosis, and/or parasitism. These fungi are also reported as inducers of plant growth promotion and plant defensive system (Mukherjee et al. 2013). Moreover, a large proportion of the industrial production of pcwdCAZymes, as well as the understanding of expression mechanisms, is derived from *Trichoderma reesei* given its outstanding capability to feed on plant biomass (Lehmann et al. 2016). Thus, *Trichoderma* spp. are widely studied and employed in biotechnology, comprising species of agricultural and industrial interests.

Genome sequencing and comparative analysis have provided a wealth of data related to genome evolution and nutritional strategies of *Trichoderma* spp. (Druzhinina et al. 2018; Kubicek et al. 2019). Such data is crucial for understanding fungal biology, enabling strain improvement and more efficient usage for biotechnological purposes. Currently, genome-wide studies have mainly focused on *Trichoderma* species from the Harzinaum, Viride and Longibrachiatum clades, which contain the most commonly sampled species and with industrial/commercial relevance. However, since *Trichoderma* genus is highly diverse, investigating species from other taxonomic sections may be interesting to expand the understanding of genome evolution and nutritional strategies of these fungi. Also, it can contribute to detect new features not seen in species from most explored clades.

During a survey of the endophytic fungi associataed with orchids from the Brazilian Atlantic Forest of Minas Gerais state, a *Trichoderma* isolate (COAD 3006) was obtained from roots of

Anacheilium allemanoides. This fungus had hyaline conidia instead of green conidia found in more commonly sampled species, which is a prevalent morphological feature of lone lineages within this genus (Jaklitsch and Voglmayr 2015). Also, upon cultivation on lignocellulosic biomass, this fungus produced high titers of xylanolytic activity, while low cellulase activity was detected, even though the culture medium had a much higher cellulose content. Those morphological and nutritional features motivated a deeper characterization of *Trichoderma* COAD 3006 genome and protein secretome. Thus, we conducted genome sequencing, phylogenetic analysis and overall description of protein-coding genes related to nutritional strategies of *Trichoderma* COAD 3006, as well as comparative genomics and profile of gene gains and losses to study genome evolution of the taxon. Finally, the exoproteome data was correlated to the quantified enzyme activities to gain insights into one of its nutritional strategies (saprophytism), improving the understanding of species evolution and lifestyle.

2. Materials and methods

2.1. Culture conditions for plant cell wall-degrading enzymes production

The *Trichoderma* COAD 3006 was cultured in propagation medium (YPD) with the following composition: yeast extract 10 $g.L^{-1}$, peptone 20 $g.L^{-1}$, glucose 20 $g.L^{-1}$. Erlenmeyer flasks (250 mL) containing 100 mL of propagation medium were inoculated with 10 agar plugs (Ø = 5.0 mm) of PDA plates containing 7 days-old fungal mycelia and incubated on an orbital shaker for 7 days at 150 rpm and 28 ^{o}C. For enzyme production, the fungus was grown in liquid medium using 2.0 % (w/v) steam-exploded sugarcane straw as sole carbon source. The mineral solution for fungal cultivation consisted of NH_4NO_3, 2.0 $g.L^{-1}$; K_2HPO_4 , 2.0 $g.L^{-1}$; $MgSO_4$, 2.0 $g.L^{-1}$; NaCl, 1.0 $g.L^{-1}$; $CaCl_2$, 2.0 $g.L^{-1}$; citric acid, 2.8 $g.L^{-1}$; yeast extract, 3.0 $g.L^{-1}$; peptone, 2.0 $g.L^{-1}$ and 1.0 $mL.L^{-1}$ of

trace elements solution (EDTA, 50.0 $g.L^{-1}$; $MnCl_2$, 0.4 $g.L^{-1}$; $CoCl_2$, 0.16 $g.L^{-1}$; $CuSO_4$, 0.16 $g.L^{-1}$; H_3BO_3, 1.1 $g.L^{-1}$; $(NH_4)_6Mo_7O_{24}$, 0.13 $g.L^{-1}$, $FeSO_4$, 0.5 $g.L^{-1}$ e $ZnSO_4$, 2.2 $g.L^{-1}$). Fifth five milliliters of mineral solution and 1.2 grams of steam-exploded sugarcane straw (dry weight) were added in Erlenmeyer flasks (125 mL). The flasks were autoclaved at 120 ºC for 15 minutes and cooled to room temperature before adding 5 mL of propagation medium per flask (aprox. 2.0 x 10^6 spores.mL^{-1}). The flasks were kept at 28 ºC under agitation of 250 rpm. After 10 days of cultivation, the residual solids were filtered using a nylon filter followed by centrifugation at 15,000*g* for 10 minutes at 4 °C. The clarified supernatants were kept at 4 °C and phenylmethanesulfonyl fluoride (PMSF) was added to a final concentration of 1.0 $mmol.L^{-1}$. All experiments were performed in triplicate.

2.2. Protein quantification and plant cell wall-degrading enzyme activities

Endo-glucanase (CMCase) and endo-xylanase activities were determined by measuring the release of reducing sugars from carboxymethyl cellulose (CMC) and beechwood xylan, respectively, using dinitrosalicylic acid reagent (DNS) (Miller 1959). The enzymatic reactions were carried out in test tubes with 250 µL reaction volumes containing 200 µL of substrate suspension (1% w/v) in 50 $mmol.L^{-1}$ sodium acetate buffer, pH 5.0 and 50 µL of appropriately diluted enzyme preparation. The reactions were conducted for 15 min at 50 ºC and terminated by the addition of 250 µL of DNS reagent. Test tubes were boiled for 5 minutes and chilled at room temperature. The amount of reducing sugars released was determined by the correlation between absorbance at 540 nm and known concentrations of reducing sugar, with a calibration curve constructed with D-Glucose as standard.

Cellobiohydrolase, β-glucosidase, β-xylosidase and α-arabinofuranosidase activities were determined using the corresponding 4-nitrophenyl sugar derivative (β-D-cellobioside, β-D-

glucopyranoside, β-D-xylopyranoside and α-L-arabinofuranoside, respectively). The enzymatic reactions were carried out in 96-well plate with 100 µL reaction volume containing 50 µL of substrate suspension (0.5 $mmol.L^{-1}$) in 50 $mmol.L^{-1}$ sodium acetate buffer, pH 5.0 and 50 µL of appropriately diluted enzyme preparation. The reactions were conducted for 15 min at 50 ºC and terminated by the addition of 100 µL of 1 $mol.L^{-1}$ NaOH. The amount of 4-nitrophenolate released was determined by the correlation between absorbance at 410 nm and known concentrations of 4-nitrophenolate, with a calibration curve constructed with 4-nitrophenolate as standard (Sigma-Aldrich).

One unit of enzymatic activity was defined as the amount of enzyme that released 1 µmol of product per minute under assay conditions.

The quantification of soluble proteins in the crude enzymatic preparation was determined by the bicinchoninic acid method using bovine serum albumin as the standard (Smith et al. 1985).

2.3. Genomic DNA extraction of Trichoderma *COAD 3006*

The fungus was cultured on potato-dextrose agar (PDA) at 28 ºC for 10 days and the genomic DNA extraction was carried out from mycelia using a method described previously (Doyle and Doyle 1990).

2.4. Genome sequencing and assembly

The *Trichoderma* COAD 3006 genome was sequenced using Illumina Novaseq (Illumina Inc., San Diego, CA, USA). The fragments library was prepared using TruSeq Nano DNA prep kit (Illumina Inc., San Diego, CA, USA) with 150 bp pair-ends reads and 550 bp insertion length (inner distance, 250 bp). FASTQC v0.11.5 (Andrews 2010) was used to quality control assessment and

Trimmomatic v0.36 (Bolger et al. 2014) for raw data filtering with 4-mer sliding-window and mean *Q* value of 15. Filtered reads were used for *de novo* assembly using SPAdes v.3.9.0 (Bankevich et al. 2012) with k-mer sizes ranging from 27 to 123.

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAAOZV000000000. The version described in this paper is version JAAOZV010000000.

2.5. Phylogenetic analysis

The reference sequences of RNA polymerase II subunit 2 (*rpb2*), translation elongation factor 1-alpha (*tef1-α*) and internal transcribed spacer 1 – 5.8S – internal transcribed spacer 2 (*ITS*) of *Trichoderma* species were downloaded from GenBank (Table S1). Sequences were aligned with MAFFT v.7.305 (Katoh and Standley 2013) using the L-INS-i strategy. The three alignment files generated (one for each gene region) were combined with a custom Python script.

The jmodeltest v.2.1.10 (Darriba et al. 2015) was used to infer the best substitution model under akaike information criterion. Maximum likelihood analysis was carried out in RaxML v.8.2.12 (Stamatakis 2014) using the GTR+I+G model with 3000 searches for the best-scoring tree. Maximum likelihood bootstrap proportions (MLBP) were calculated with 3000 replicates. Bayesian inference was performed with MrBayes v.3.2.6 (Ronquist et al. 2012) on CIPRES platform (Miller et al. 2010) using Markov Chain Monte Carlo (MCMC) algorithm. Six independent runs with 4 chains each were conducted for 2,000,000 generations, sampling every 100th generation. The initial 25 % trees were discarded as burn-in phase and the remaining trees were used for estimating bayesian inference posterior probability (BIPP) values. Trees were visualized in FigTree v.1.4.4 (Rambaut 2018) and statistic supports were evaluated by MLBP and BIPP.

2.6. Phylogenomic analysis

Genomes and putative proteins of *Trichoderma* spp. were downloaded from NCBI and JGI databases (Table S2). When not available, protein-coding genes were predicted as described in section 2.7.

To access the completeness of the genomes, we searched for 3,725 Sordariomycete universal single-copy orthologous genes with BUSCO v3 (Simão et al. 2015). Only the genomes with completeness equal or greater than 90.0 % were kept for the next analysis.

Protein orthogroups were identified with Orthofinder v.2.3.3 (Emms and Kelly 2015) with MCL inflation parameter of 1.5 and gene tree inference method using Mafft v.7.305 and FastTree v.2.1.9. The corresponding nucleotide sequences of proteins from each orthogroup of single-copy orthologs were aligned with Mafft v.7.305 using the L-INS-i strategy. The alignment files generated for each orthogroup were combined into a single alignment file with a custom Python script. Maximum likelihood analysis for the concatenated alignment file was performed with RaxML v.8.2.12 using the GTR+I+G model with 200 searches for the best-scoring tree. Maximum likelihood bootstrap proportions (MLBP) were calculated with 500 replicates. Tree visualization and annotations were performed with “ggtree” package of R software (Yu et al. 2017).

The information of species presence or absence (gene sequences) in each orthogroup were retrieved using a custom Python script. Estimation of gene losses and gains was carried out with dollop program from Phylip package v.3.696 (Felsenstein 2005). Dollop is based on the Dollo parsimony principle, which assumes that a character loss is irreversible and a character gain occurs only once in evolutionary history. We used the maximum likelihood phylogenomic tree as a guide for dollop program.

2.7. Gene prediction and annotation

Protein coding genes were predicted with a combination of *ab initio* and similarity-based methods. Augustus v3.2.2 (Stanke and Waack 2003) trained with *Trichoderma atroviride* gene parameters and GeneMark-ES (Ter-Hovhannisyan et al. 2008) using self-training mode were used for *ab initio* gene predictions. For similarity-based gene predictions, raw RNA-seq reads from *Trichoderma* spp. were downloaded from NCBI Sequence Read Archive database (Table S3). Raw data filtering was carried out with Trimmomatic v0.36 with 4-mer sliding-window and a mean *Q* value of 25. Filtered reads were assembled using Tophat v.2.1.1/Cufflinks v.2.2.0 pipeline (Trapnell et al. 2013) and the detection of coding regions within transcripts was performed with Transdecoder v.3.0.1 (Haas et al. 2013). Exonerate v.2.2.0 (Slater and Birney 2005) was used to predict genes by performing spliced-sequence alignments between assembled coding sequences and the target genome. The EvidenceModeler v.1.1.1 (Haas et al. 2008) was used to compute weighted consensus gene structure based on gene evidence from *ab initio* and similarity-based methods.

Protein functional annotation was carried out with InterProScan v.5.30-69.0 (Jones et al. 2014) using default mode and DIAMOND (Buchfink et al. 2014) searches against RefSeq fungi proteins downloaded from NCBI Protein database (downloaded in Fev. 2019). The dbCAN2 (Zhang et al. 2018) in default mode was used to retrieve annotations regarding carbohydrate-active enzymes families. For each protein, redundancy removal and annotation merging were performed with a custom Python script.

Predictions of protein subcellular localization were performed by three independent tools: CELLO v2.5, which is based on a two-level support vector machine system (Yu et al. 2006); Deeploc-1.0, which uses deep neural networks (Almagro Armenteros et al. 2017); and BUSCA, an integrative approach that combines various tools for prediction of short signal peptides and subcellular localization based on amino acid sequence (Savojardo et al. 2018). A consensus non-redundant result for each protein sequence was obtained by combining the outputs of the three tools.

2.8. Exoproteome analysis

The proteins in the crude enzymatic preparation were precipitated with pre-chilled ethanol (90% v/v) for 24 hours at -20 ºC. After centrifugation (9000*g* for 20 minutes at 4 °C), the protein pellet was kept at room temperature for 30 minutes to dry residual ethanol. The protein pellet was resuspended in 100 µL of resuspension solution containing urea 7.0 mol.L^{-1}, thiourea 2.0 mol.L^{-1} and CHAPS 4.0 % w/v. An electrophoresis was conducted with 30 µg of protein in 10% SDS-PAGE and the run was stopped when proteins migrated from stacking gel to resolving gel (*short run*). The unique protein band were excised, discolored with methanol 50% and acetic acid 5%, reduced with 50 mmol.L^{-1} dithiothreitol, alkylated with 100 mmol.L^{-1} iodoacetamide and subjected to in-gel trypsinization with 2 µg of trypsin (Trypsin Gold, Mass Spectrometry Grade - Promega™). The peptides derived from the trypsin-digested sample were desalted by stage-tips (Rappsilber et al. 2007), dried in a vacuum concentrator and reconstituted in 0.1% formic acid. The samples were run in triplicate. One microliter of peptide-containing solution were injected on PicoFrit Column (20 cm x ID75 µm, 5 µm particle size, New Objective) and analyzed on an ETD enabled Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) connected to the EASY-nLC system (Proxeon Biosystem, West Palm Beach, FL, USA). The identification of proteins was performed with MaxQuant v.1.6.3.3 (Cox and Mann 2008) against a combined set of protein sequences from *Trichoderma* COAD 3006 (this study), *T. reesei* QM6a (GCA_000167675.2), *T. atroviride* IMI 206040 (GCA_000171015.2), *T. harzianum* CBS 226.95 (GCA_003025095.1), *T. virens* Gv29-8 (GCA_000170995.2). Protein quantification was performed using the intensity-based absolute quantification (iBAQ) algorithm implemented in MaxQuant environment. False discovery rate (FDR) of 1% and a minimum of 2 unique peptides were used as confidence parameters.

2.9. Morphological study

Microscopic preparations were made by slide culture (Rosado et al., 2019). *Trichoderma* COAD 3006 was grown on Corn-meal Agar (CMD) plugs, placed directly onto a sterile microscope slide. Petri dishes with slides were incubated at 25 °C under a photoperiod of 12 h for 7 days. Microscopy slides were mounted in lactoglicerol. Photographs were taken with an Olympus BX53 microscope equipped with a digital camera, Olympus Q-Color5™. Measurements (*n*=30) of the relevant morphological characteristics (conidia, conidiophores, conidiogenous cells and chlamydospores) were made using the Olympus cellSens Dimension 1.9 software. For cultural characteristics, mycelium plugs of *Trichoderma* COAD 3006 were replicates for plates (one cm from the edge) containing PDA, CMD or Synthetic Nutrient Deficient Agar (SNA), the colonies were cultivated for 7 days at 25 ºC a photoperiod of 12 h.

3. Results

3.1. Trichoderma *COAD 3006 genome and putative protein-coding genes*

The genome of *Trichoderma* COAD 3006 was sequenced using whole-genome shotgun approach in an Illumina Novaseq platform, generating 57,263,146 reads and 8,646,735,046 base pairs, representing a mean coverage of 151x. After quality control and raw data filtering, 55,646,970 reads and 8,402,692,470 base pairs were kept in the data set, representing 97.2% of the total data generated. The assembled genome contained 165 contigs in 150 scaffolds (minimum lenght equal to 500 bp), with the largest scaffold of 3,158 Kbp. The assembly contiguity was good, with N50 of 1,411 Kbp and L50 of 9. The size of the assembled genome was 32,799,057 bp (Table 1), which is in the known range of *Trichoderma* spp. genome sizes (Kubicek et al. 2019).

Table 1. Summary of the *Trichoderma orchidacearum* COAD 3006 genome assembly results.

Total sequenced bases	8,646,735,046
Number of scaffolds	150
Largest scaffold (bp)	3,158,156
Total length (bp)	32,799,057
GC content (%)	49.15
N50	1,410,984
L50	9
N's per 100 kbp	3.73

A combination of an *ab initio* and similarity-based gene prediction was used to predict protein-coding genes in *Trichoderma* COAD 3006 genome. The final gene set comprised 8,903 protein-coding genes and a search for *Sordariomyceta* universal single-copy orthologs using BUSCO v3 (Simão et al. 2015) yielded 96.6 % of completeness, indicating good genome assembly. The gene density of *Trichoderma* COAD 3006 was 271 genes/Mbp, featuring at the lower end of *Trichoderma* spp. gene densities range (271 to 427 genes/Mbp). On the other hand, the average exon and intron lengths were similar to that of other species (Table S2). Taking together, these features indicated that the genome of *Trichoderma* COAD 3006 was composed of a slightly larger fraction of non-coding DNA.

3.2. Phylogenetic analysis

Initially, a BLASTn search against NCBI nucleotide collection (nr/nt) revealed that the sequences of *rpb2*, *tef1-α* and *ITS* from *Trichoderma* COAD 3006 had best sequence identities of 87.6 % with *T. alni* strain T24, 92.4 % with *T. neocrassum* strain 10617 and 91.7 % with *T. peltatum* strain CBS 127115, respectively. We used maximum likelihood (ML) and bayesian inference analysis of combined *rpb2*, *tef1-α* and *ITS* sequence alignments to perform phylogeny reconstruction and both approaches generated the same tree topology. The ML tree was presented

with MLBP and BIPP values at the nodes (Figure 1). *Trichoderma* COAD 3006 was clearly a new species, since it not grouped with any known species of *Trichoderma.*

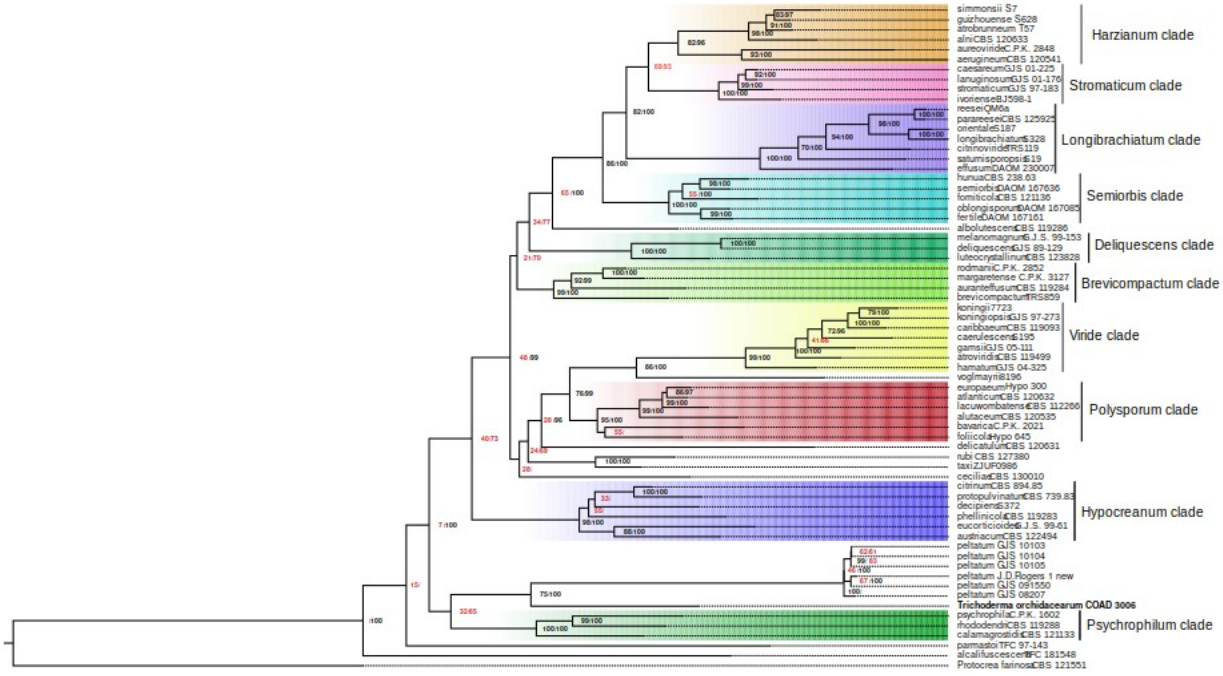


Figure 1. Phylogenetic tree based on the concatenated alignments of *rpb2* (RNA polymerase II subunit 2), *tef1-α* (translation elongation factor 1-alpha) and *ITS* (internal transcribed spacer 1 – 5.8S – internal transcribed spacer 2) genes. The tree is rooted to *Protocrea farinosa.* MLBP above 70% (left) and BIPP above 95% (right) are indicated in black at the nodes. The new species proposed are indicated in boldface. The Genbank accessions is given in Table S1.

Trichoderma COAD 3006 and *T. peltatum* formed an independent clade, and they were separated from each other with strong support values (MLBP/BIPP = 75%/1.00). The clade formed by these two *Trichoderma* grouped with the Psychrophilum clade, although with poor support values (MLBP/BIPP = 32%/0.65).

All other *Trichoderma* clades were in agreement with previous studies (Jaklitsch and Voglmayr 2015), including lone lineage species, which generally form weak associations with well-defined clades.

3.3. Taxonomy

Trichoderma orchidacearum. T. Morgan, F.A. Custódio, V.M, Guimarães, T.A.O. Mendes & O.L. Pereira, **sp. nov**. (Figure 2)

MycoBank: MB 834779

Etymology: In reference to the host family, Orchidaceae, from which the fungus was isolated as endophyte.

Description on CMD: Mycelium septate, branched, smooth, and hyaline, 2-5.5 µm diam hyphae. *Conidiophores* were usually straight or somewhat flexuous, arising as lateral branches of plagiotropous hyphae, cylindrical-oblong, hyaline, branched. *Phialides* were hyaline, lageniform to cylindrical-shaped, straight or curved, paired and rarely isolated, measuring 3.8-12.5 µm in length, 1.3-2.5 µm in width at the widest point, and 1.1–2.3 µm in width at the base. Phialides were formed from supporting cells measuring 4–10.6 x 1.9-4.3 µm; *Conidia* were subglobose to cylindrical, hyaline, smooth, non-septate, 2.75-4 x 1.5-2.35 µm. *Chlamydospores* were globose, hyaline, terminal or intercalated, 3–6.7 x 3.75–8 µm.

Culture characteristics: On PDA a radius of 55-56 mm after 72 h at 25 °C, mycelium abundant, dense, covering the plate after 4 days, showed concentric rings, with white-yellow surface and caramel on reverse, colourless to the periphery. On CMD a radius of 55-58 mm after 72 h at 25 °C, mycelium moderate, covering the plate after 4 days, showed concentric rings, with white surface and pale-caramel on reverse, colourless to the periphery. On SNA a radius of 32 mm after 72 h at 25 °C, mycelium scarce, covering the plate after 6 days with white surface and reverse. Material examined: Brazil, isolated as endophytic from roots of *Anacheilium allemanoides*, forest fragment between Canaã and Araponga, Brazil, 2008 O. L. Pereira (holotype VIC 47393, ex-type culture COAD 3006).

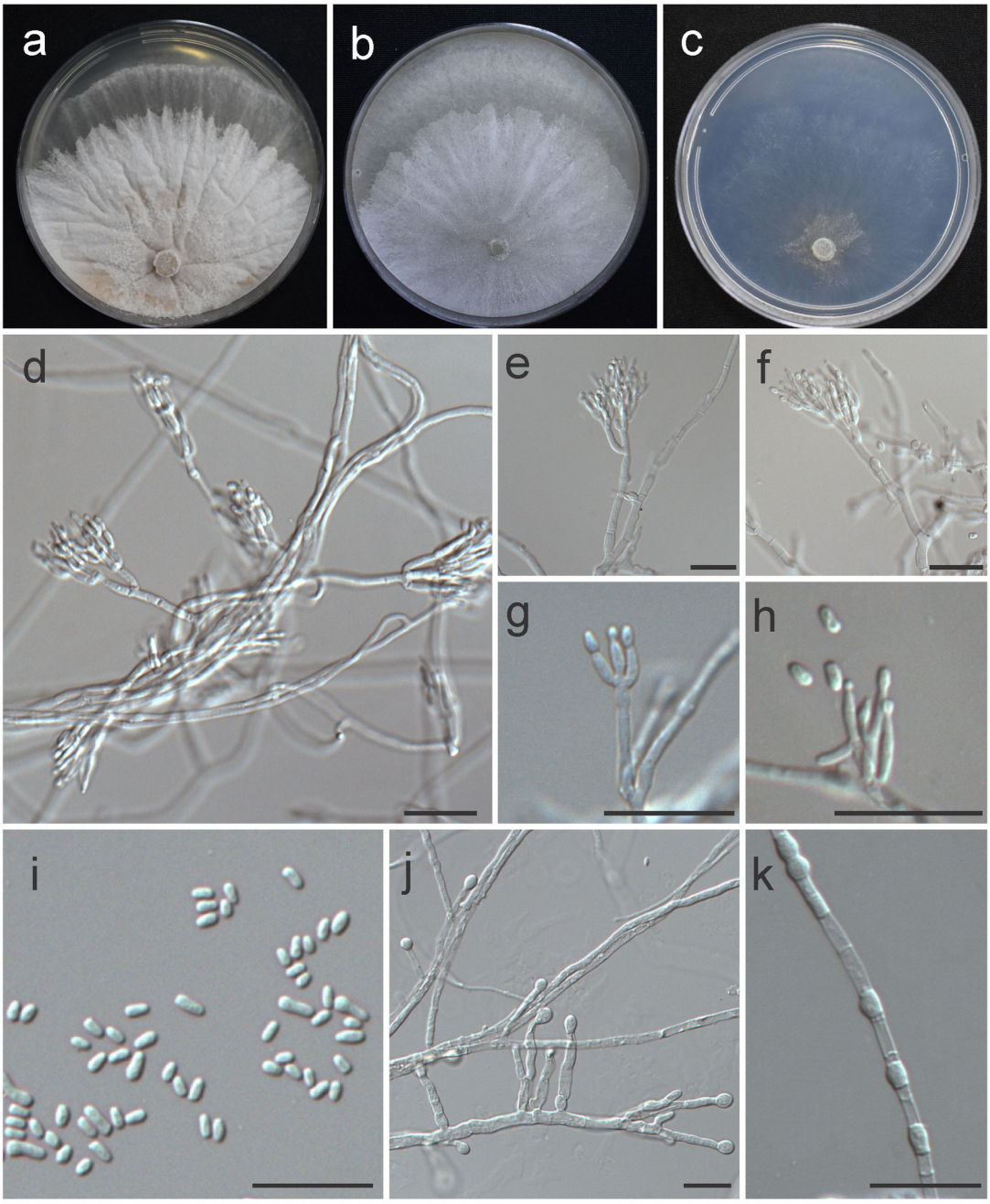


Figure 2. *Trichoderma orchidacearum* COAD 3006. a−c. Cultures after 7 d at 25 ºC (a. PDA; b. CMD; c. SNA). d–g. Conidiophores and phialides g. phialides. i. Conidia. j-k Chlamydospores (SNA, 14 d); Scale bars: d–k = 20 µm.

3.4. Phylogenomic analysis and Parsimonious Scenario of Gene Gain and Loss in the T. orchidacearum *and other* Trichoderma *species*

The phylogenomic analysis associated with profiles of gene gains and losses is a useful strategy to study the evolution of genomes and correlate with organism lifestyle. To investigate the

gene family evolutionary profiles in *T. orchidacearum*, we identified protein orthogroups between this species and other 29 *Trichoderma* spp. (17 different species) with genomes previously sequenced. We also included fungal genomes from many associated genera (*Escovopsis*, *Hypomyces*, *Beauveria* and *Metarhizium*) to obtain a more accurate estimation of the gene loss and gain events at the rise of the *Trichoderma*. All genomes used in this study had more than 92.0% of completeness (Table S2), as accessed by searches with BUSCO v3 using S*ordariomyceta* universal single-copy orthologs.

The data set for genomic analysis was composed of 440,558 proteins and 421,022 were assigned to 17,286 orthogroups, while 19,536 sequences showed no homology to any other in the data set (orphan genes). A total of 198 orthogroups had sequences from all *Trichoderma* species but *T. orchidacearum*, representing a putative contraction of less than 3.5 % of a putative "core genome" of the genus for this species (considering 5,783 orthogroups with all *Trichoderma* species present, including *T. orchidacearum* and sometimes other fungal genera). Furthermore, we detected exactly 50 orthogroups specifically composed of sequences from *Trichoderma* spp., which were less than 0.3 % off all orthogroups detected and evidenced the high diversity inside the genus (Table S4). Furthermore, other 12 orthogroups were specifically composed of *Trichoderma* spp. but *T. orchidacearum*, which was the highest number of missing "*Trichoderma*-specific orthologous genes" among *Trichoderma* species (Table S5). These missing orthologous genes in *T. orchidacearum* encoded mainly putative hydrolytic enzymes, which were related to carbon metabolism (glycosil-hydrolases) and proteolysis (peptidases).

Of the 17,286 orthogroups detected, 3,973 contained sequences from all fungi used in this study, and 1,709 consisted of entirely single-copy genes, which were used to build the phylogenomic tree. The evolutionary model obtained further showed that *T. orchidacearum* was considerably divergent from the other *Trichoderma* spp. with whole-genome data available (Figure 3). It is worthy mentioning that the tree topology was inconsistent regarding the position of *T.*

koningii JCM 1883 and *T. harzianum* Tr1 and thus, we performed a phylogenetic reconstruction using *rpb2*, *tef1-a* and *ITS* sequences and corrected the species names to *T. longibrachiatum* JCM 1883 and *T. pleuroticola* Tr1, respectively (Figure S1, Table S6). Furthermore, the *Hypomyces* isolates were not clustered in the same clade, but such topology was confirmed by phylogenetic reconstruction using *tef1-a* and *ITS* sequences, which indicated that *Hypomyces perniciosus* isolates were closest to *Trichoderma* genus than to other *Hypomyces* scpecies (Figure S2, Table S7)

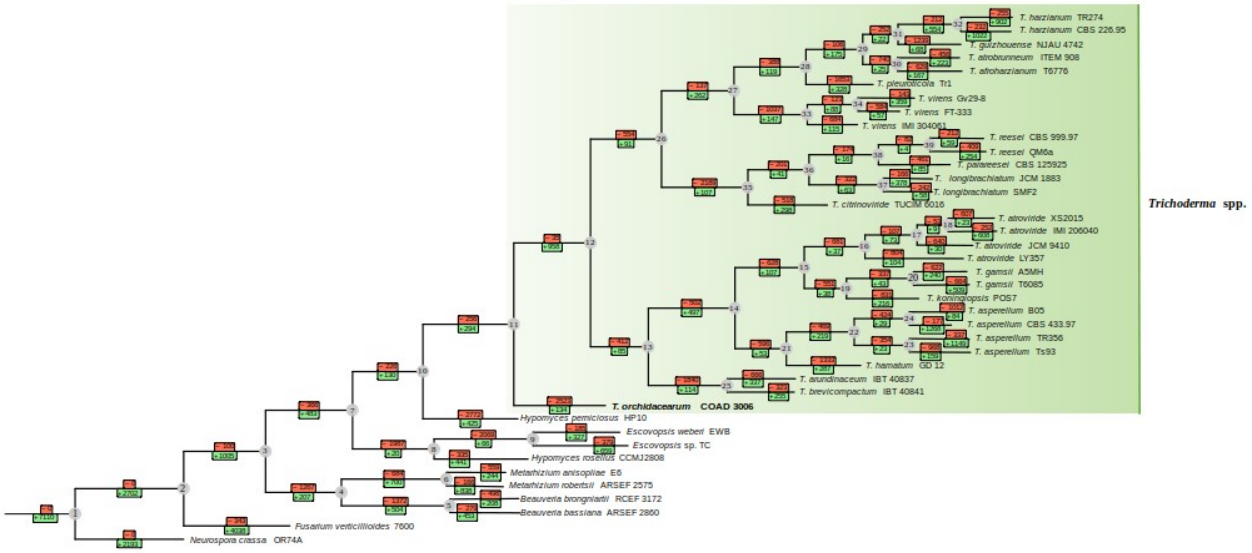


Figure 3. The best-scoring maximum-likelihood tree for *Trichoderma orchidacearum* COAD 3006 and other *Trichoderma* species. The tree was created based on 2,605,200 sites in 2,171 orthologous nucleotide sequences using the GTR+I+G substitution model and 500 maximum likelihood bootstrap replications. All bootstrap values were 100% and omitted from the phylogenomic tree. Numbers inside green boxes indicate genes gained and numbers inside red boxes indicate genes lost. The internal nodes were numbered from the root. The *Fusarium verticilliodes* 7600 and *Neurospora crassa* OR74A were used as outgroups.

To estimate gene gains and losses through the phylogenomic model, we used the information of species presence or absence in each protein cluster from the orthology analysis (36,822 clusters), and fitted these profiles to the phylogenomic tree (Figure 3). Regarding *Trichoderma*, a similar number of gene gains and losses was detected for the ancestral species that gave rise to the genus (Figure 3, node 11), followed by a dominance of gene gains in the ancestral of *Trichoderma* spp. after *T. orchidacearum* speciation (Figure 3, node 12). After that, prominent gene gains were detected for a few extant species only (*T. harzianum* CBS 226.95, *T. harzianum* TR274, *T. asperellum* CBS 433.97, *T. asperellum* TR356). Gene losses, on the other hand, were much more

common through *Trichoderma* genus, especially for *T. orchidacearum* (2523 genes lost and 134 genes gained) and the ancestors of the Longibrachiatum and Brevicompactum clades.

Regarding the genes acquired by *T. orchidacearum,* no homologs were detected in any fungal species analyzed (orphans genes) and we were able to infer functionality for only 10 of those genes using a combination of Pfam (database v.33.0) and PANNZER2 predictions (Table S8). These genes might be related to species-specific adaptations and may have arisen from sequence divergence or *de novo* emergence from non-coding regions (Tautz and Domazet-Lošo 2011). The set of genes lost by *T. orchidacearum* also encoded many hypothetical proteins (536 genes) but unlike the genes gained, we were able to assign function for most of them. They included mainly transcription factors (e.g., fungal specific transcription factors, Zn_2/Cys_6 zinc-finger and bZIP domains) and transmembrane transporters (e.g. MFS general substrate transporter, CorA-like Mg_2+ transporters), but prominent losses were also detected for ankyrin repeats, dehydrogenases (especially short chain dehydrogenases), heterokaryon incompatibility protein, methytransferases, kinases and domains related to protein-protein interactions such as F-box, WD40 and tetratricopeptide (Table S9).

Furthermore, we predicted the gene ontology (GO) terms of the genes lost in order to provide a comprehensive view of the major metabolic processes affected by the *T. orchidacearum* genome shrinkage. The molecular functions detected were mostly transferase activity (especially methyltransferase activity), transcription factor activity (DNA-binding transcription factor activity, RNA polymerase II-specific), hydrolase, transmembrane transporter and oxidoreductase activities (Figure S3, Table S9). The biological processes were especially enriched in oxidation-reduction process, transcription and transmembrane transport, while the major cellular components were integral component of membrane and nucleus.

Membrane transporters are important in osmotrophic organisms such as fungi, enabling specific acquisition of nutrients released by the action of extracellular enzymes and preventing the

internalization of harmful substances (Richards and Talbot 2013). The massive loss of genes that encoded membrane transporters in *T. orchidacearum* might be linked to its colonization niche (orchid roots, acting as an endophyte), an restricted environment with less competition for nutrients. Besides this, the common ancestor of all other *Trichoderma* spp. analyzed (Figure 3, node 12) showed a positive balance (more gains than losses) of genes coding for transmembrane transporters (Table S10). Since these genes may directly participate in the fungal nutritional strategy and development (e.g., for sugar and metal ion transport), such profiles of gene gains and losses indicated important shifts regarding nutrient acquisition of *T. orchidacearum* compared to other *Trichoderma* species.

Transcription factors (TF) gene families were also largely lost throughout *T. orchidacearum* evolution. These proteins regulates the amount of messenger RNA produced from a gene, having major impacts on organism development and fitness. The $Zn(II)_2Cys_6$ zinc-finger domain was the major TF family lost by *T. orchidacearum* (65 genes), followed by fungal specific transcription factor domain (63 genes) (Table S9). These TFs are abundant in fungi and required in numerous biological processes (Chang and Ehrlich 2013; Zhang et al. 2019), which suggested a lesser requirement of gene expression reprogramming in *T. orchidacearum*.

The set of genes lost by *T. orchidacearum* also included many genes related to oxidation-reduction processes, with the most prominent losses concerning the general NAD(P)/FAD-dependent oxidoreductases (64 genes), short-chain dehydrogenases (25 genes) and p450 monooxygenases (12 genes). Oxidation-reduction are extremely diverse processes and take place in virtually all metabolic pathways, from the most basal (e.g. glycolytic pathway) to secondary metabolite biosynthesis (e.g. nonribosomal peptides, polyketides, sesquiterpenes).

Other prominent gene losses comprised heterokaryon incompatibility (59 genes) and ankyrin-repeat coding genes (87 genes), which are putatively involved in enhanced ecological fitness and mycoparasitism of *Trichoderma* (Kubicek et al. 2019). In addition to the loss of

methyltransferases (53 genes) and protein kinases (40 genes), these events further indicated a simplification of signaling transduction and response to external stimuli in *T. orchidacearum*.

Regarding pcwdCAZymes, we detected the loss of 38 genes, most of them coding for pectin- and xylan-degrading enzymes (Figure 4, Table S11). Despite this, high titers of xylanase activity were detected in the culture supernatant of *T. orchidacearum* grown on plant biomass (see Table 2 in section 3.6), which indicated an important role of xylanases in *T. orchidacearum* nutritional strategy.

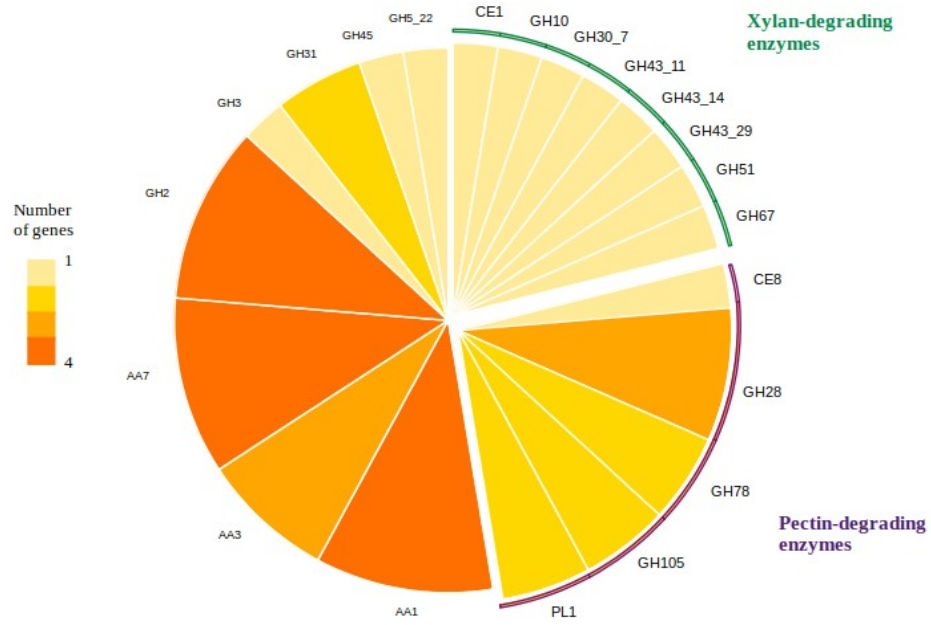


Figure 4. Pie chart showing the carbohydrate-active enzymes lost in the course of *Trichoderma orchidacearum* COAD 3006 evolution. The size of the external sector is proportional to the number of lost genes of each CAZy family, with the smallest sector corresponding to one gene and the largest sector corresponding to four genes.

In contrast, cellulase coding genes were much more resistant to loss (Figure 4). Only one gene coding for GH3 β-1,4-glucosidase and one gene of GH45 endo-β-1,4-glucanase were putatively lost by *T. orchidacearum*. The GH3 genes was the most abundant cellulase-coding genes in the *T. orchidacearum* genome, as well as for all other *Trichoderma* species, and the loss of such genes might be tolerated due to genetic redundancy. Genes coding for GH45 endo-β-1,4-glucanase were lost by many *Trichoderma* spp., and complete family loss were inferred for some isolates such

as *T. asperellum* Ts93 and *T. atroviride* LY357 (Figure 5, Table S12: clusters OG0005626 and OG0009676), which suggested a minor effect of these enzymes on the environmental fitness of some species. The origin of the GH45 genes was before the raise of the *Trichoderma* genus (had orthologs in *Neurospora*, *Fusarium* and other genera) and were subjected to sequence and function diversification, since we detected other pcwdCAZyme domains besides GH45 inside the OG0005626 and OG0009676 orthogroups.

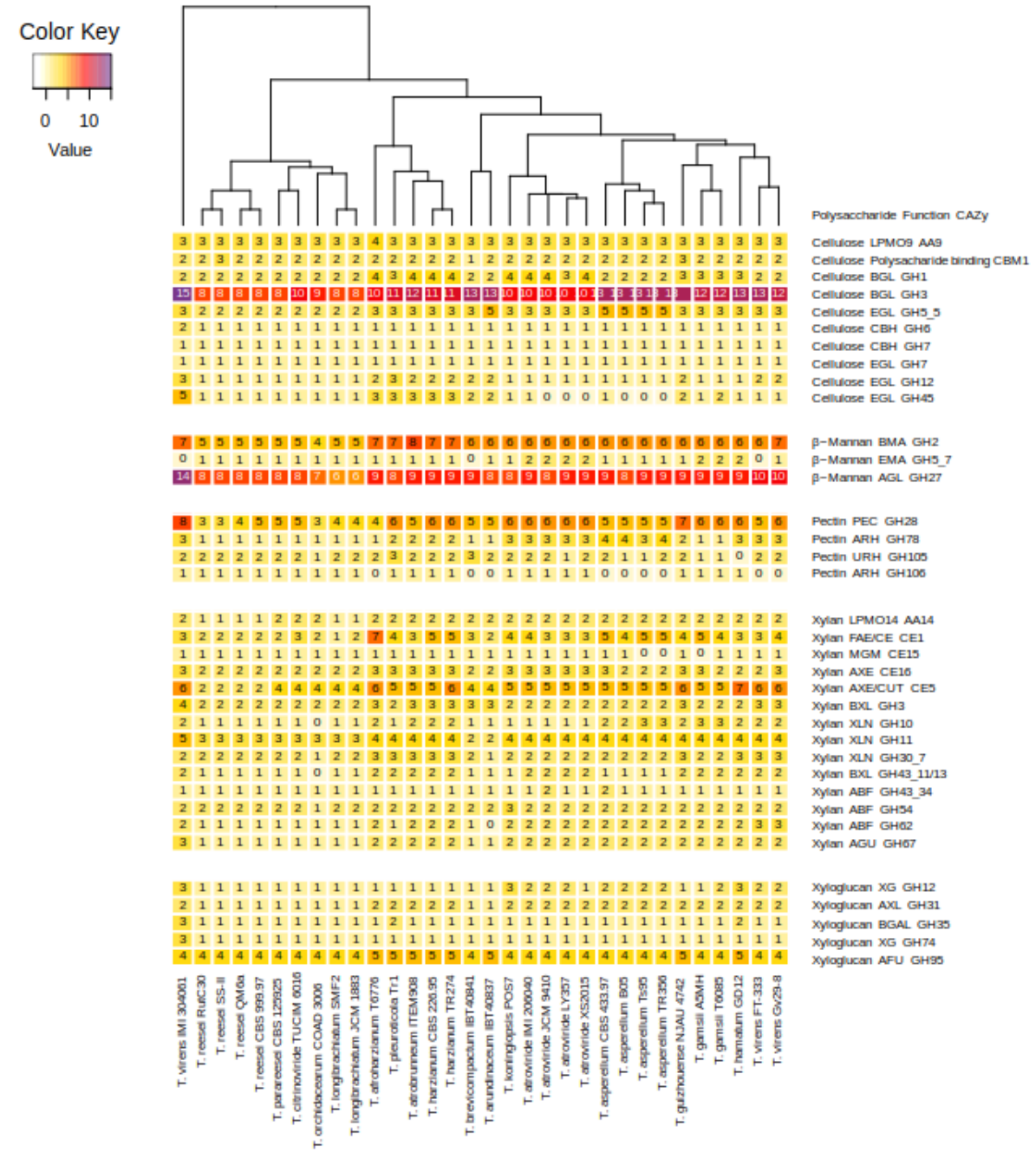


Figure 5. The inventory of carbohydrate-active enzymes (CAZymes) related to plant cell wall degradation in *Trichoderma*. CAZymes are grouped according to their polysaccharide substrate, which is given on the right side. The number of genes belonging to each CAZy family in the genome is indicated inside the frames. The cladogram on the top is based on the number of genes per each CAZy family. Abbreviations: ABF, α-arabinofuranosidase; AFU, α-fucosidase; AGL, α-galactosidase; AGU, α-glucuronidase; ARH, α-rhamnosidase; AXE, acetylesterase/acetyl xylan esterase; AXL, α-xylosidase; BGAL, β-galactosidase; BGL, β-glucosidase; BMA, β-mannosidase; BXL, β-xylosidase; CBH, cellobiohydrolase; CBM1 *, polysacharide binding (cellulose)/expansin-like; EGL, endo-β-1,4-glucanase; EMA, endo-β-1,4-mannanase; FAE, feruloyl esterase; LPMO9, AA9 lytic cellulose monooxygenase; LPMO14, AA14 lytic xylan monooxygenase; MGM, 4-O-methyl-glucuronoyl methylesterase; PEC, pectinases; URH, unsaturated rhamnogalacturonyl hydrolase; XG, xyloglucanase; XLN, endo-1,4-β-xylanase.
*: the recorded CBM1 domains were isolated, without a hydrolytic domain in the same protein sequence.

Concerning the gene gains and losses for other *Trichoderma* species, the ancestor of the Longibrachiatum clade also showed a predominance of gene loss (Figure 3, internal node 35). These genes were mainly related to transcription factors (141 genes), transmembrane transporters (94 genes), general dehydrogenases (82 genes), ankyrin repeats (81 genes) and heterokaryon incompatibility protein (46 genes) (Table S13). Furthermore, the integral component of membrane was the most common cellular component (498 genes) followed by nucleus (204 genes), showing high similarity to the profile of gene losses of *T. orchidacearum.* Moreover, the ancestor of Brevicompactum species (*T. brevicompactum* and *T. arundinaceum*) also lost many genes (Figure 3, internal node 25) and the set of putative functions of those genes followed a similar trend as for *T. orchidacearum* and the ancestor of Longibrachiatum species (Table S14).

3.5. Comparative genomic analysis

The comparison of CAZyme-coding genes across genomes of closely related fungal species may provide information regarding major changes in nutritional strategies and evolution of polysaccharide-degrading enzyme machinery. In this sense, we identified the CAZyme-coding genes in *T. orchidacearum* genome and other 31 *Trichoderma* strains (17 different species). The *T. orchidacearum* genome harbored 309 putative CAZymes, the lowest number among all *Trichoderma* spp. analyzed (Table S15). The number glycosyl-hydrolases (GHs) resembled that of

species from the Longibrachiatum clade, while it had the lowest numbers of CAZymes of other families (CE, PL, AA, GT).

Regarding cellulases, the numbers of enzymes from GH6, GH7 and AA9 families were highly homogeneous among the *Trichoderma* species. On the other hand, we detected considerable variability concerning the endo-β-1,4-glucanases from GH3, GH5_5, GH12 and GH45 families (Figure 5). The number of cellulose-degrading enzymes of *T. orchidacearum* was slightly smaller than average (23 putative genes, mean = 27), but the difference does not exceed the standard deviation (s.d. = 4).

Regarding the xyloglucan-degrading enzymes, *T. orchidacearum* had a smaller repertoire than the most *Trichoderma* spp. analyzed (8 putative genes, mean = 10, s.d. = 2) and identical to that of species from Longibrachiatum clade. In addition, it presented the smallest repertoire of pectin- and β-mannan-degrading enzymes (12 putative genes, mean = 16, s.d. = 2). It is worth mentioning that the number of putative pectin-degrading enzymes in *Trichoderma* genomes were highly variable, especially for accessory pectinases (GH78, GH105 and GH106), which varied markedly even within a same species.

Furthermore, *T. orchidacearum* showed a considerable contraction in the set of genes coding for xylan-degrading enzymes (21 putative genes, mean = 31, s.d. = 6). This fungus was the only *Trichoderma* species with no glycosyl hydrolases from GH10 and GH43_11 families. The family GH10 comprises endo-β-1,4-xylanases with less specificity to substrates compared to the GH11 family (“true” xylanases) and GH43_11 comprises β-1,4-xylosidases, an important accessory enzyme for xylan degradation. In addition, *T. orchidacearum* had the smallest number of genes from the GH30_7 (endo-β-1,4-xylanases) and GH54 (α-arabinofuranosidases) families. These features clearly showed a simplification of the xylan-degrading machinery of *T. orchidacearum*.

Regarding CAZymes related to mycoparasitism, the most prominent contraction was detected for chitin-degrading enzymes families. *Trichoderma orchidacearum* had fewer genes

coding for GH18 and AA11 proteins compared to most *Trichoderma* species (Figure S4). In addition, we detected contractions in the GH5_15, GH64 and GH128 families, which comprises β-1,3/1,6-glucan-degrading enzymes. Taking together, the CAZyme profile of *T. orchidacearum* was more similar to those of species from Longibrachiatum clade, which are mainly saprotrophic.

In addition to carbohydrate-active enzymes, fungi secrete other enzymes for nutrient acquisition, such as proteases. Since these hydrolytic enzymes are also involved in intracellular metabolism and morphogenesis, we reported only the families that possessed more than 50% of members predicted as extracellular enzymes. Being putatively active in the extracellular environment, these enzymes can effectively contribute to the fungal nutritional strategy.

The genome of *T. orchidacearum* harbored one of the smallest sets of putative protease-coding genes compared to other *Trichoderma* species (Figure S5), especially regarding proteases from families S08, S09 and S10 (serine-proteases). On the other hand, the number of genes coding for S53 family (sedolisins) was more similar to that of species from Viride clade (*T. asperellum*, *T atroviride*) and considerable higher than the average of species from Longibrachiatum clade (mean = 5, s.d. = 1). It is worthy mentioning that the S53 family was more reduced in *Trichoderma* from Longibrachiatum clade, which might be due to specific adaptations of those species. Furthermore, we detected a minor contraction in the number of genes coding for aspartic-proteases (A01A family) e metallo-proteases (M14A, M20A, M28E, M36 families) in *T. orchidacearum* compared to the other species.

3.6. Enzyme activities secreted by Trichoderma orchidacearum COAD 3006 *cultivated on lignocellulosic biomass*

The set of extracellular pcwdCAZymes highly influences *Trichoderma* spp. lifestyle and is often shaped by evolution (Druzhinina et al. 2018). Furthermore, these biocatalysts are widely

exploited for biotechnological purposes, mainly by the food industry and for biomaterials production (Eggleston 2007).

To access the ability of *T. orchidacearum* COAD 3006 to produce pcwdCAZymes, we cultivated this fungus on steam-exploded sugarcane straw, recovered the secreted proteins from the culture supernatant and quantified the major xylanolytic and cellulolytic activities. Like other *Trichoderma* spp., the *T. orchidacearum* was able to produce these enzymes when cultivated on lignocellulosic biomass. The endo-xylanase activity was comparatively higher, reaching about 90x more activity than cellulases (CMCase, Cellobiohydrolase, β-glucosidase) (Table 2). Other quantified xylanolytic activities included β-xylosidase and α-arabinofuranosidase, which acts synergistically with endo-β-1,4-xylanases for arabinoxylan hydrolysis (Wilkens et al. 2017).

Table 2. Xylanolytic and cellulolytic activities detected in the culture supernatant of *Trichoderma orchidacearum* COAD 3006 grown on steam-exploded sugarcane straw.

Enzyme	Specific activity ($U.mg^{-1}$)
endo β-1,4-xylanase	96.0 ± 4.92
β-xylosidase	0.01 ± 0.00
α-arabinofuranosidase	0.13 ± 0.00
Cellobiohydrolase	0.01 ± 0.00
CMCase	1.09 ± 0.01
β-glicosidase	0.12 ± 0.00

The quantified enzymatic activities evidenced a large imbalance between xylanolytic and cellulolytic activities, suggesting that the growth conditions favored the expression of xylan-degrading enzymes. Interestingly, the steam-exploded sugarcane straw had a reduced xylan content compared to the cellulose content (Table S16), a feature observed for other steam-exploded sugarcane biomasses (Martín et al. 2008; da Silva et al. 2013).

3.7. Overview of exoproteome of T. orchidacearum *COAD 3006 cultivated on lignocellulosic biomass*

Mass spectrometry-based methods are an invaluable tools for deciphering the protein composition of fungal exoproteomes to better understand the nutritional strategies and biotechnological potential of these organisms.

To gain insight into the abundance of extracellular proteins produced by *T. orchidacearum* grown on cellulosic biomass, we performed quantitative proteomic analysis by LC-MS/MS. A total of 355 proteins were identified in the culture supernatant and 75 were classified as CAZymes, especially glycosyl hydrolases (Table S17).

Since a high xylanase activity was detected in the culture supernatant (Table 2), we expected a high relative abundance of xylan-degrading enzymes. Indeed, over 26 % of all detected proteins corresponded to an endo-β-1,4-xylanase from GH11 family (Figure 6), which was by far the most abundant endo-β-1,4-xylanase produced by *T. orchidacearum*. Another endo-β-1,4-xylanase (GH30_7 family) was detected, althought with much lower relative abundance (0.62%, Figure 7) and thus, probably had a minor contribution to the endo-xylanolytic activity of this fungus. This contrasts with a common feature of many fungi grown on plant biomass, which produces several endo-β-1,4-xylanases, especially from GH10 and GH11 families (Cologna et al. 2018; da Silva et al. 2019).

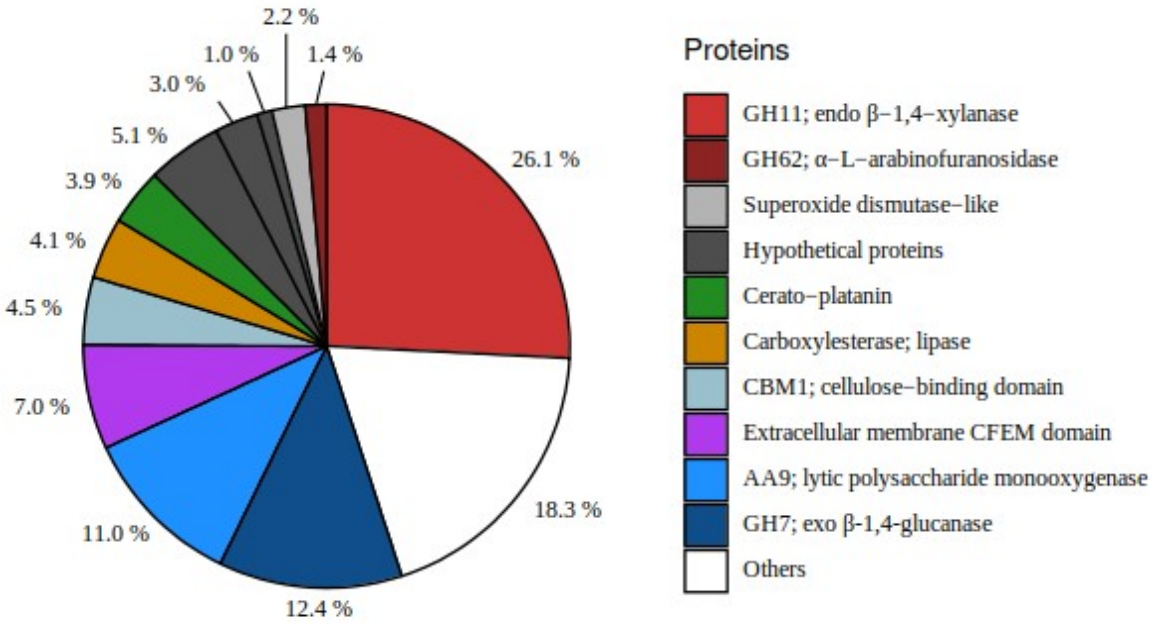


Figure 6. Composition of the exoproteome of *Trichoderma orchidacearum* COAD 3006 grown on steam-exploded sugarcane straw. Proteins with relative abundance greater than 1 % were depicted in the figure. The remaining proteins were grouped in the "Others" sector. Sequences that showed no significant BLASTp and InterProScan hits (E-value > 10^{-5}) were named hypothetical proteins.

Furthermore, the xylanolytic system secreted by *T. orchidacearum* was also composed of GH62 α-L-arabinofuranosidase (1.4 % of relative abundance), and small proportions of β-1,4-xylosidase (GH3), α-glucuronidase (GH67) and 4-O-methyl-glucuronoyl methylesterase (CE15, CIP2 protein) with relative abundances of 0.23%, 0.34% and 0.08% respectively (Figure 7).

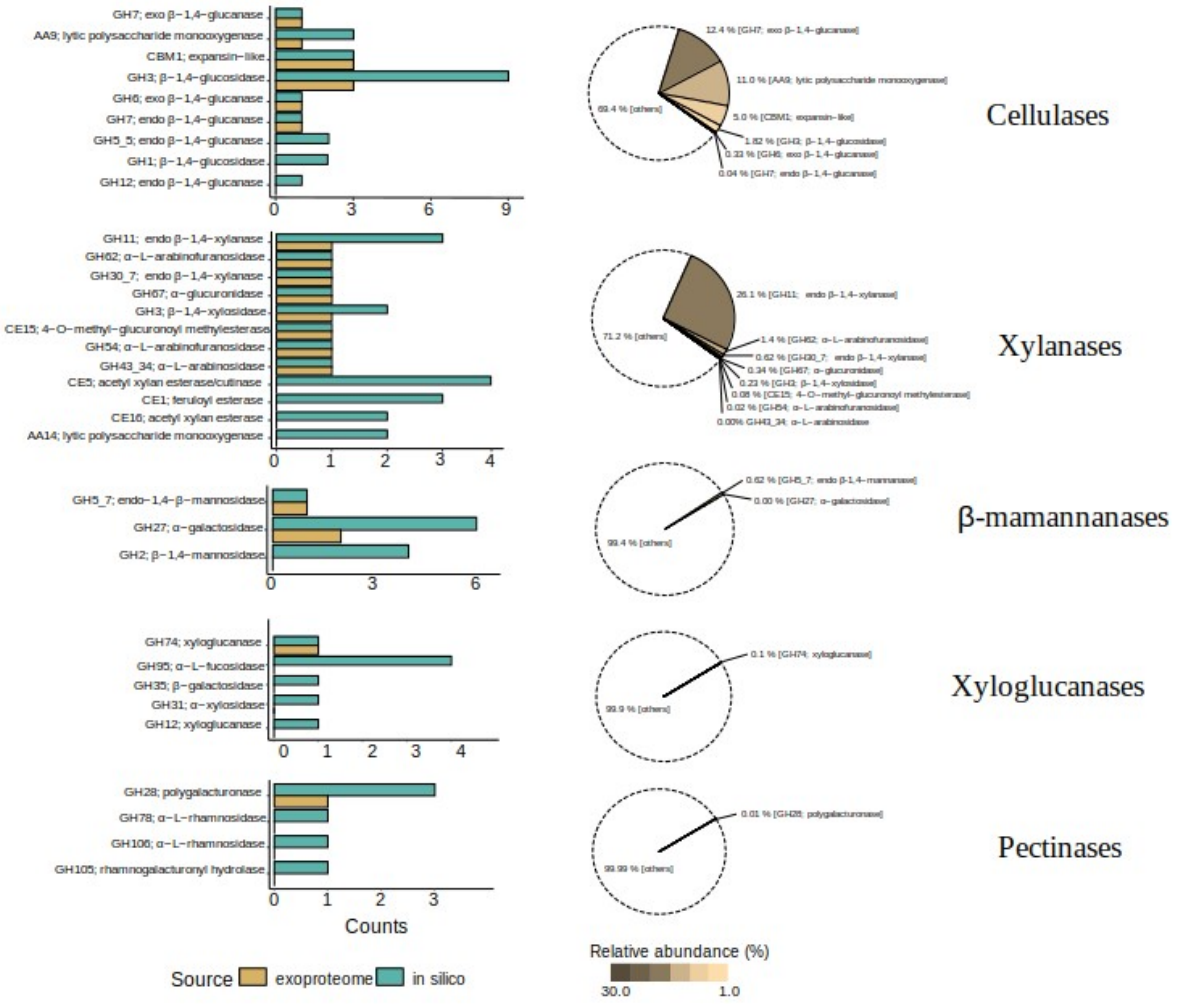


Figure 7. Quantity and abundance of plant cell wall-degrading enzymes detected in the exoproteome of *Trichoderma orchidacearum* COAD 3006 grown on steam-exploded sugarcane straw. Each pair of bar and pie charts refers to cellulases, xylanases, β-mannanases, xyloglucanases and pectinases. The horizontal bar plots indicate the number of proteins for each CAZy family detected in the exoproteome (brown bars) and the corresponding number of coding genes in the genome (blue bars). Pie charts indicate the proportion of exoproteome related to degradation of each plant cell wall polysaccharide (cellulose, xylan, β-mannan, xyloglucan and pectin) with a detailed relative abundance of each protein detected. We reported only the CAZymes predicted as extracellular proteins by at least 2 subcellular localization tools (CELLO v2.5, Deeploc-1.0 and BUSCA).

Regarding cellulolytic enzymes, we detected a GH7 exo-β-1,4-glucanase (cellobiohydrolase), AA9 lytic polysaccharide monooxygenase (LPMO) and a CBM1 domain (carbohydrate-biding module family 1) which was similar to *T. reesei* and *T. parareesei* CIP1 protein (81% and 75% of sequence identity, respectively). The CIP1 protein is an enhancing factor of lignocellulose degradation with unknown enzymatic activity (Lehmann et al. 2016), and is often found in *Trichoderma reesei* exoproteome under induction of pcwdCAZyme expression (Jacobson

et al. 2013). Surprisingly, we detected only one endo-β-1,4-glucanase in the *T. orchidacearum* exoproteome, which was an uncommon feature of plant biomass-degrading fungi. The detected enzyme belongs to the GH7 family and showed low relative abundance (less than 0.04 %), despite other putative genes coding for endo-β-1,4-glucanases (GH5 and GH12 families) were identified in the genome (Figure 5). Thus, the endo-β-1,4-glucanase activity of *T. orchidacearum* exoproteome might rely mainly on the AA9 LPMOs, which show oxidative activity on internal β-1,4-glycosidic bonds.

Furthermore, three β-1,4-glycosidases and 2 expansin-like proteins, which often have CBM1, were detected in the exproteome. All putative expansin-like proteins predicted in the genome were detected in the exoproteome (Figure 7), suggesting prominent importance for polysaccharide breakdown.

A minor abundance of other pcwdCAZymes detected included β-mannan-degrading enzymes (GH5_7 endo-β-1,4-mannanase, GH27 α-galactosidase), a pectin degrading enzyme (GH28 polygalacturonase) and a xyloglucan-degrading enzyme (GH74 xyloglucanase), where each set had relative abundances of 0.63 %, 0.01 % and 0.10 %, respectively. Such low relative abundances suggests a tight regulation of production of these enzymes by *T. orchidacearum*, since the cultivation medium did not contain their target polysaccharides.

Apart from pcwdCAZymes, a cerato-plantanin was detected at relatively high levels in the *T. orchidacearum* exoproteome (Figure 6). This protein belongs to the large group of small secreted cysteine-rich proteins (SSCPs), being reported as interaction effectors between the fungus and other organisms, as well as may exhibit an expansin-like activity on polysaccharides. In *Trichoderma* spp., these proteins were also reported as elicitors of plant resistance against pathogens, in addition to a possible role in the development of the fungus itself (Gaderer et al. 2014). Furthermore, proteins probably not addressed to the extracellular medium were present in the exoproteome with high relative abundance, such as membrane proteins (carboxylesterase and a CFEM domain

protein) and cytoplasmic proteins (superoxide dismutase-like superoxide and a hypothetical protein). Filamentous fungi generally have a strong cell wall, but intracellular or plasma membrane proteins often leak into the extracellular medium due to mechanical stress or cell death.

4. Discussion

This is the first report of a genome sequencing of a basal lineage *Trichoderma* species, in addition to describe a new species. Like other *Trichoderma* spp., *T. orchidacearum* has a versatile lifestyle, acting as endophytic (orchid roots) and possibly as saprotrophic fungus. Since it is a basal species and phylogenetically distant from other *Trichoderma* spp., our results regarding fungal cultivation on lignocellulose and CAZyme activities support the hypothesis that all species in this genus can efficiently feed on plant cell wall polysaccharides (Druzhinina et al. 2018).

The total length of the assembled genome of *T. orchidacearum* was at the lower end of the range of *Trichoderma* spp. genome sizes (https://genome.jgi.doe.gov/fungi/fungi.info.html) (Grigoriev et al. 2014) and more similar to that of species from Longibrachiatum clade (*T. reesei, T. parareesei, T. citrinoviride, T. longibrachiatum*), which primarily have saprophytic lifestyle instead of a more common mycoparasitic lifestyle showed by most *Trichoderma* species. Albeit not being the smallest one, the *T. orchidacearum* genome harbored the fewest number of putative protein-coding genes (8,903 genes), despite not belonging to the Longibrachiatum clade. Regarding this clade, the smaller genomes have a strong relationship with physiology and evolution of *Trichoderma* lifestyle (Xie et al. 2014), and similar events might also occurred during *T. orchidacearum* evolution. Thus, our study contributed to enhance the knowledge about the diversity and lifestyle shifts within *Trichoderma* genus.

Regarding the phylogenetic position of the *T. orchidacearum*, it grouped in the same clade of *T. peltatum*, but separated of this species by strong support values (MLBP/BIPP = 75 % / 1.00),

indicating that they were closely related species and possibly shared a recent evolutionary history. Since *T. peltatum* was classified as a lone lineage (Jaklitsch and Voglmayr 2015), we inferred that *T. orchidacearum* might be classified with the same taxonomic feature. Lone lineage species tend to form unstable phylogenetic relations with well-defined clades, which was evidenced by the poor support values (not significant) of clustering with the Psychrophilum clade.

The phylogenomic analysis further confirmed the high genetic variability of *T. orchidacearum* compared to other *Trichoderma* species, the most divergent species with whole-genome data available. The large evolutionary distance between *T. orchidacearum* and the other *Trichoderma* spp. was also evidenced by the absence of this species in many *Trichoderma*-specific orthogroups containing all other *Trichoderma* spp. analyzed (Table S5).

We used gene orthology information and the phylogenomic tree to estimate gene gains and losses throughout the evolutionary history of the *Trichoderma* genus to gain insights into the relationships between genome evolution and lifestyle. In this sense, we detected a massive reduction in the set of protein-coding genes throughout the evolution of *T. orchidacearum*, as opposed to the common ancestor of all other *Trichoderma* species analyzed (Figure 3). The putative function of these proteins were mainly related to gene expression regulation (methyltranferases, transcription factors, protein kinases), signal transduction (protein-protein interaction domains, protein kinases) general transmembrane transport and oxidation-reduction processes (Table S9). Such massive gene losses in *T. orchidacearum* genome indicated a reduction in organism metabolic complexity, which might be compatible with an endophytic lifestyle (less competition for nutrients, fewer pathogens, as well as abiotic stress). Furthermore, the gene losses might also be related to gene dispensability guaranteed by mutational robustness, i.e. occurrence of genetic redundancy or alternative metabolic pathways (Albalat and Cañestro 2016). Thus, we inferred a considerable simplification of *T. orchidacearum* metabolical processes, which might be a consequence of less versatility regarding ecological niches and/or nutritional strategies compared to other *Trichoderma*

species. The fungal adaptation to different lifestyles is tighly linked to changes in the genome content and organism metabolical needs and consequently, to the requirements of genes and gene products to conduct the biological processes (Stajich 2017).

Furthermore, the profiles of gene gains and losses across the *Trichoderma* genus revealed striking variations between isolates of the same species, such as *T. asperellum* strains. Therefore, there was no clear correlation between expansion or contraction in the set of protein-coding genes even at species level, indicating that these evolutionary events were somewhat frequent and explain, at least in part, the great diversity of lifestyles or nutritional modes observed in *Trichoderma* genus.

The horizontal gene acquisition encoding pcwdCAZymes were proposed as an important evolutionary event in *Trichoderma* genus, which provided the ability to feed on plant biomass (Druzhinina et al. 2018). We detected only pcwdCAZyme losses for *T. orchidacearum*, indicating a loss of ability to acquire novel pcwdCAZyme genes or the acquisition did not provided adaptive advantage and was readily lost. This species showed a reduced number of putative genes coding for cellulose-, β-mannan-, pectin- and xyloglucan-degrading enzymes compared to other *Trichoderma* spp., which further indicated a genome evolution towards a reduced set of genes, better fitted to less versatile nutritional strategies. This was more evident concerning xylan-degrading enzymes, since *T. orchidacearum* showed contractions in gene numbers for almost all glycosyl hydrolases families for xylan depolymerization and even complete family losses, namely GH10 and GH43_11 (endo-β-1,4,-xylanase and β-1,4,-xylosidase, respectively). Besides this, an endo-β-1,4-xylanase was detected at high levels in the exoproteome of *T. orchidacearum* grown on lignocellulosic biomass. Thus, the expansion/contraction of pcwdCAZymes families may not be directly related to the ability or preference to degrade a particular plant polysaccharide. Also, the high relative abundance of the GH11 endo-β-1,4-xylanase indicated a prominent role of this enzyme in fungal fitness and the fungus preference to consume the hemicellulosic fraction of plant cell wall.

The contraction of the pcwdCAZymes repertoire of *T. orchidacearum* clearly showed a simplification of enzymatic machinery for depolymerization of major plant cell wall polysaccharides. This can be advantageous, since the fungus may expend less metabolic resources to regulate gene expression, protein synthesis and secretion, which may provide faster growth and environment colonization. For example, GH11 xylanases tend to be more efficient than GH10 xylanases to depolymerize poorly soluble or complexed xylans (as in plant biomasses) (van Gool et al. 2013) and the combined action of xylanases from the two families may not show synergistic effect (Beaugrand et al. 2004; Yagi et al. 2019). Thus, the production of xylanases from both families might provide little or no adaptive advantage. Furthermore, the comparatively high losses of genes coding for pectin-degrading enzymes may be related to an adaptation to endophytic lifestyle, since such enzymes are related to virulence factors of plant pathogenic fungi (D'Ovidio et al. 2004). The loss of these genes may be an adaptation to minimize the elicitation of plant immune response, which might be advantageous to fungal establishment inside plant tissues.

The mycoparasitism is considered an ancient lifestyle of *Trichoderma* (Kubicek et al. 2011) and a high number of genes involved in the hydrolysis of chitin, β-1,3/1,6-glucan, α-mannans were observed in all *Trichoderma* spp. analyzed, probably because provides an advantage for environment colonization by killing or reducing the growth of other fungi. The contractions in the sets of genes coding for chitinases (families GH18 and AA11), β-1,3-glucanases (GH64) and serine-proteases (S08, S09 and S10) were detected for *T. orchidacearum* genome, which could be related to shifts regarding the ancestral lifestyle, evolutionary events similar to those that gave rise to the Longibrachiatum clade. For example, the shrinkage of the S08 and S10 families was a common feature of plant symbiont fungi since such proteases may contribute to a necrotrophic (mycoparasitic) phenotype (Muszewska et al. 2017).

Taking together, *T. orchidacearum* showed fewer gene copies related to environmental responses, such as polysaccharide-degrading enzymes and proteases, yet it presented approximately

97 % of the genes that form a possible core genome of *Trichoderma*, which fits well in the genome streamlining hypothesis (Giovannoni et al. 2014). For instance, the losses of many TF-coding genes may be a consequence of such streamlining effect, events that may have occurred after the loss of the genes they regulated. Also, duplicate genes (paralogous) tend to differ mainly regarding its transcriptional regulation (Wapinski et al. 2007) and thus, a relatively small genome with fewer paralogues might demand simpler gene expression regulation system. Thus, we inferred an overall simplification of the biological processes of *T. orchidacearum* possibly due to shifts to restricted ecological niche or nutritional strategies.

At first glance, there is no strong evidence that supports genetic drift as an important event for the *T. orchidacearum* genome evolution, since no expansion in intron number and size distribution, or dramatic reduction in gene density was observed for its genome (Table S2) (Kelkar and Ochman 2012). The rate of mutation fixation by genetic drift is higher when the effective population is reduced (Yi 2006), which is unlikely for *T. orchidacearum*. Although isolated as an endophytic, this fungus may show a free-living lifestyle, since it showed the ability to feed on major plant cell wall polysaccharides, an abundant and widely distributed carbon source.

Apart from genomic analysis, the exoproteome *T. orchidacearum* was analyzed using LC-MS/MS to obtain functional information about its saprophytic capabilities. In a broad sense, the set of secreted proteins was tailored for the depolymerization of polysaccharides available in its colonization environment, since 59.4 % of the total proteins detected in the exoproteome corresponded to pcwdCAZymes active on the main sugarcane straw polysaccharides (cellulose and xylans). Also, this enzyme set was relatively non-redundant, with little overlap between the cellulolytic or xylanolytic enzymes, regarding both the enzyme family and activity. This indicated a rationalized production of pcwdCAZymes, which might be a consequence of the genome streamlining.

The most abundant xylanolytic enzymes secreted by *T. orchidacearum* comprised endo-β-1,4-xylanase, α-L-arabinofuranosidase and β-1,4-xylosidase, an enzyme set well suited to depolymerize arabinoxylans. Only acetyl-xylan esterase and feruloyl esterase were not detected in the exoproteome, although putative genes were identified in the genome (Figure 5). Since acetyl groups and phenolic compounds are generally removed from steam-exploded pretreated biomass (Martín et al. 2008), such enzymes could provide little advantage for fungal growth, and therefore, may have been down-regulated. Although abundant in raw sugarcane biomasses, arabinoxylans are largely removed by steam-explosion pretreatment, and the residual fraction remains insoluble, linked to cellulose in the pretreated material (Berrin and Juge 2008; Wilkens et al. 2017). Even with a lower relative proportion of xylans in the growth medium, about a quarter of all proteins detected in the *T. orchidacearum* exoproteome corresponded to a GH11 endo-β-1,4-xylanase.

On the other hand, a small amount of endo-β-1,4-glucanase was detected in the exoproteome, despite the high abundance of cellulose in the growth medium and other cellulases comprising a considerable proportion of the exoproteome (30.6 %). This pattern of secreted pcwdCAZymes is not common in fungi with the ability to degrade plant biomass and suggested a less tightly expression regulation of GH11 and endo-β-1,4-glucanase genes in *T. orchidacearum*. The mechanisms of expression regulation of (hemi)cellulase-coding genes dependents on various external stimuli including light (Schuster et al. 2012), pH (Adav et al. 2011), temperature, nitrogen source (Guowei et al. 2011) and especially the balance between different carbon sources in the growth medium (Benocci et al. 2017). The signaling transduction cascades triggered by these stimuli usually involve protein kinases and transcription factors, which directly regulate gene expression (Schuster et al. 2012). Many genes coding for transcription factors and protein kinases were putatively lost during the evolution of *T. orchidacearum* and thus, this fungus might have a poor ability to tightly regulate the expression of particular genes (e.g. GH11 endo-β-1,4-xylanase and endo-β-1,4-glucanases) in some environmental conditions, including the one used in this study.

5. Conclusions

Trichoderma orchidacearum is a new species from a basal phylogenetic position within the *Trichoderma* genus. The genome sequencing of this species opens new paths to study and understand genome evolution of *Trichoderma* and the correlation with phenotypic diversity.

The gene gain and loss profiles of *T. orchidacearum* showed a genome evolution dominated by losses of protein-coding genes, a similar trend observed for the ancestors of Longibrachiatum and Brevicompatum clades. Thus, similar evolutionary processes might have occurred more than once in the evolutionary history of the *Trichoderma* genus, being associated with environmental fitness and thus, mainly guided by adaptive forces.

Like other *Trichoderma* species, the *T. orchidacearum* have the ability to feed on plant cell wall polysaccharides and the set of CAZymes secreted by this fungus is tailored for the depolymerization of polysaccharides available in its colonization environment. Nevertheless, the relative abundance of these enzymes has clearly been unbalanced, which directly impacts the efficiency of saprophytism as a nutritional strategy and might be related to its poorer ability to regulate gene expression and signaling transduction related to carbohydrate metabolism in some environmental growth conditions.

References

Adav SS, Ravindran A, Chao LT, et al (2011) Proteomic analysis of pH and strains dependent protein secretion of Trichoderma reesei. J Proteome Res 10:4579–4596. doi: 10.1021/pr200416t

Albalat R, Cañestro C (2016) Evolution by gene loss. Nat Rev Genet 17:379–391. doi: 10.1038/nrg.2016.39

Almagro Armenteros JJ, Sønderby CK, Sønderby SK, et al (2017) DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics 33:3387–3395. doi: 10.1093/bioinformatics/btx431

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Bankevich A, Nurk S, Antipov D, et al (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. doi: 10.1089/cmb.2012.0021

Beaugrand J, Chambat G, Wong VWK, et al (2004) Impact and efficiency of GH10 and GH11 thermostable endoxylanases on wheat bran and alkali-extractable arabinoxylans. Carbohydr Res 339:2529–2540. doi: 10.1016/j.carres.2004.08.012

Benocci T, Aguilar-Pontes MV, Zhou M, et al (2017) Regulators of plant biomass degradation in ascomycetous fungi. Biotechnol Biofuels 10:1–25. doi: 10.1186/s13068-017-0841-x

Berrin JG, Juge N (2008) Factors affecting xylanase functionality in the degradation of arabinoxylans. Biotechnol Lett 30:1139–1150. doi: 10.1007/s10529-008-9669-6

Blin K, Shaw S, Steinke K, et al (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. doi: 10.1093/nar/gkz310

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. doi: 10.1093/bioinformatics/btu170

Buchfink B, Xie C, Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. doi: 10.1038/nmeth.3176

Chang PK, Ehrlich KC (2013) Genome-wide analysis of the Zn(II)2Cys6 zinc cluster-encoding gene family in Aspergillus flavus. Appl Microbiol Biotechnol 97:4289–4300. https://doi.org/10.1007/s00253-013-4865-2

Chen LH, Yang SL, Chung KR (2014) Resistance to oxidative stress via regulating siderophore-mediated iron acquisition by the citrus fungal pathogen Alternaria alternata. Microbiol (United Kingdom) 160:970–979. doi: 10.1099/mic.0.076182-0

Chitwood DJ (2002) Phytochemical based strategies for nematode control. Annu Rev Phytopathol 40:221–249. doi: 10.1146/annurev.phyto.40.032602.130045

Cologna N de M di, Gómez-Mendoza DP, Zanoelo FF, et al (2018) Exploring Trichoderma and Aspergillus secretomes: Proteomics approaches for the identification of enzymes of biotechnological interest. Enzyme Microb Technol 109:1–10. doi: 10.1016/j.enzmictec.2017.08.007

Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372. doi: 10.1038/nbt.1511

D'Ovidio R, Mattei B, Roberti S, Bellincampi D (2004) Polygalacturonases, polygalacturonase-inhibiting proteins and pectic oligomers in plant-pathogen interactions. Biochim Biophys Acta - Proteins Proteomics 1696:237–244. doi: 10.1016/j.bbapap.2003.08.012

da Silva AS, Teixeira RSS, Moutta RO, et al (2013) Sugarcane and Woody Biomass Pretreatments for Ethanol Production. Intech 13. doi: 10.5772/53378

da Silva DS, Dantzger M, Assis MA, et al (2019) Lignocellulolytic characterization and comparative secretome analysis of a Trichoderma erinaceum strain isolated from decaying sugarcane straw. Fungal Biol 123:330–340. doi: 10.1016/j.funbio.2019.01.007

Darriba D, Taboada GL, Doallo R, Posada D (2015) jModelTest 2: more models, new heuristics and high-performance computing Europe PMC Funders Group. Nat Methods 9:772. doi: 10.1038/nmeth.2109.jModelTest

De Bruyne L, Van Poucke C, Di Mavungu DJ, et al (2016) Comparative chemical screening and genetic analysis reveal tentoxin as a new virulence factor in Cochliobolus miyabeanus, the causal agent of brown spot disease on rice. Mol Plant Pathol 17:805–817. doi: 10.1111/mpp.12329

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus (Madison) 12:13–15

Druzhinina IS, Chenthamara K, Zhang J, et al (2018) Massive lateral transfer of genes encoding plant cell wall-degrading enzymes to the mycoparasitic fungus Trichoderma from its plant-associated hosts

Eggleston G (2007) Advances in the industrial application of enzymes on carbohydrate-based materials. ACS Symp Ser 972:1–16. doi: 10.1021/bk-2007-0972.ch001

Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16:1–14. doi: 10.1186/s13059-015-0721-2

Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle

Gaderer R, Bonazza K, Seidl-Seiboth V (2014) Cerato-platanins: A fungal protein family with intriguing properties and application potential. Appl Microbiol Biotechnol 98:4795–4803. doi: 10.1007/s00253-014-5690-y

Giovannoni SJ, Cameron Thrash J, Temperton B (2014) Implications of streamlining theory for microbial ecology. ISME J 8:1553–1565. doi: 10.1038/ismej.2014.60

Grigoriev I V., Nikitin R, Haridas S, et al (2014) MycoCosm portal: Gearing up for 1000 fungal genomes. Nucleic Acids Res 42:699–704. doi: 10.1093/nar/gkt1183

Guowei S, Man H, Wang S, He C (2011) Effect of some factors on production of cellulase by Trichoderma reesei HY07. Procedia Environ Sci 8:357–361. doi: 10.1016/j.proenv.2011.10.056

Haas BJ, Papanicolaou A, Yassour M, et al (2013) De novo transcript sequence recostruction from RNA-Seq: reference generation and analysis with Trinity

Haas BJ, Salzberg SL, Zhu W, et al (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol 9:1–22. doi: 10.1186/gb-2008-9-1-r7

Jacobson F, Karkehabadi S, Hansson H, et al (2013) The Crystal Structure of the Core Domain of a Cellulose Induced Protein (Cip1) from Hypocrea jecorina, at 1.5 Å Resolution. PLoS One 8:. doi: 10.1371/journal.pone.0070562

Jaklitsch WM, Voglmayr H (2015) Biodiversity of Trichoderma (Hypocreaceae) in Southern Europe and Macaronesia. Stud Mycol 80:1–87. doi: 10.1016/j.simyco.2014.11.001

Jones P, Binns D, Chang HY, et al (2014) InterProScan 5: Genome-scale protein function classification. Bioinformatics 30:1236–1240. doi: 10.1093/bioinformatics/btu031

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 30:772–780. doi: 10.1093/molbev/mst010

Kelkar YD, Ochman H (2012) Causes and consequences of genome expansion in fungi. Genome Biol Evol 4:13–23. doi: 10.1093/gbe/evr124

Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, et al (2011) Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of Trichoderma. Genome Biol 12:. doi: 10.1186/gb-2011-12-4-r40

Kubicek CP, Steindorff AS, Chenthamara K, et al (2019) Evolution and comparative genomics of the most common Trichoderma species. BMC Genomics 20:1–24. doi: 10.1186/s12864-019-5680-7

Lehmann L, Rønnest NP, Jørgensen CI, et al (2016) Linking hydrolysis performance to Trichoderma reesei cellulolytic enzyme profile. Biotechnol Bioeng 113:1001–1010. doi: 10.1002/bit.25871

Martín C, Marcet M, Thomsen AB (2008) Comparison between wet oxidation and steam explosion as pretreatment methods for enzymatic hydrolysis of sugarcane bagasse. BioResources 3:670–683. doi: 10.15376/biores.3.3.670-683

Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 31:426–428

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees

Mukherjee PK, Horwitz BA, Herrera-Estrella A, et al (2013) Trichoderma Research in the Genome Era . Annu Rev Phytopathol 51:105–129. doi: 10.1146/annurev-phyto-082712-102353

Muszewska A, Stepniewska-Dziubinska MM, Steczkiewicz K, et al (2017) Fungal lifestyle reflected in serine protease repertoire. Sci Rep 7:1–12. doi: 10.1038/s41598-017-09644-w

Rambaut A (2018) FigTree v1.4.4, a graphical viewer of phylogenetic trees. Available at https://github.com/rambaut/figtree/releases

Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2:1896–1906. doi: 10.1038/nprot.2007.261

Richards TA, Talbot NJ (2013) Horizontal gene transfer in osmotrophs: Playing with public goods. Nat Rev Microbiol 11:720–727. doi: 10.1038/nrmicro3108

Ronquist F, Teslenko M, Van Der Mark P, et al (2012) Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. doi: 10.1093/sysbio/sys029

Rosado AWC, Custódio FA, Pinho DB, et al (2019) *Cladosporium* species associated with disease symptoms on *Passiflora edulis* and other crops in Brazil, with descriptions of two new species. Phytotaxa 409: 239-260. doi: 10.11646/phytotaxa.409.5.1

Savojardo C, Martelli PL, Fariselli P, et al (2018) BUSCA: An integrative web server to predict subcellular localization of proteins. Nucleic Acids Res 46:W459–W466. doi: 10.1093/nar/gky320

Schuster A, Tisch D, Seidl-Seiboth V, et al (2012) Roles of protein kinase A and adenylate cyclase in light-modulated cellulase regulation in Trichoderma reesei. Appl Environ Microbiol 78:2168–2178. doi: 10.1128/AEM.06959-11

Simão FA, Waterhouse RM, Ioannidis P, et al (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. doi: 10.1093/bioinformatics/btv351

Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:1–11. doi: 10.1186/1471-2105-6-31

Smith PK, Krohn RI, Hermanson GT, et al (1985) Measurement of protein using bicinchoninic acid. Anal Biochem 150:76–85. doi: 10.1016/0003-2697(85)90442-7

Stajich JE (2017) Fungal genomes and insights into the evolution of the kingdom. Microbiol Spectr 5:1–25. doi: 10.1016/j.physbeh.2017.03.040

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. doi: 10.1093/bioinformatics/btu033

Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19:215–225. doi: 10.1093/bioinformatics/btg1080

Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. Nat Rev Genet 12:692–702. doi: 10.1038/nrg3053

Ter-Hovhannisyan 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. doi: 10.1101/gr.081612.108

Trapnell C, Roberts A, Goff L, et al (2013) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578. doi: 10.1038/nprot.2012.016.Differential

van Gool MP, van Muiswinkel GCJ, Hinz SWA, et al (2013) Two novel GH11 endo-xylanases from Myceliophthora thermophila C1 act differently toward soluble and insoluble xylans. Enzyme Microb Technol 53:25–32. doi: 10.1016/j.enzmictec.2013.03.019

Wapinski I, Pfeffer A, Friedman N, Regev A (2007) Natural history and evolutionary principles of gene duplication in fungi. Nature 449:54–61. doi: 10.1038/nature06107

Wilkens C, Andersen S, Dumon C, et al (2017) GH62 arabinofuranosidases: Structure, function and applications. Biotechnol Adv 35:792–804. doi: 10.1016/j.biotechadv.2017.06.005

Xie B Bin, Qin QL, Shi M, et al (2014) Comparative genomics provide insights into evolution of Trichoderma nutrition style. Genome Biol Evol 6:379–390. doi: 10.1093/gbe/evu018

Yagi H, Takehara R, Tamaki A, et al (2019) Functional Characterization of the GH10 and GH11 Xylanases from *Streptomyces olivaceoviridis* E-86 Provide Insights into the Advantage of GH11 Xylanase in Catalyzing Biomass Degradation. J Appl Glycosci 66:29–35. doi: 10.5458/jag.jag.jag-2018_0008

Yi S V (2006) Non-adaptive evolution of genome complexity. BioEssays 28:979–982. doi: 10.1002/bies.20478

Yu C-S, Chen Y-C, Lu C-H, Hwang J-K (2006) Prediction of Protein Subcellular Localization. PROTEINS Struct Funct Bioinforma 64:643–651. doi: 10.1002/prot

Yu G, Smith DK, Zhu H, et al (2017) GGTREE: an R Package for Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other Associated Data. Methods Ecol Evol 8:28–36. doi: 10.1111/2041-210X.12628

Zhang C, Huang H, Deng W, Li T (2019) Genome-wide analysis of the Zn(II) 2 Cys 6 Zinc cluster-encoding gene family in Tolypocladium guangdongense and its light-induced expression. Genes (Basel) 10:. https://doi.org/10.3390/genes10030179

Zhang H, Yohe T, Huang L, et al (2018) DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. doi: 10.1093/nar/gky418

Formatting of funding sources

This work was supported by the Conselho Nacional de desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). The authors thank the “Diretoria de

Tecnologia de Informação” (DTI) at “Universidade Federal de Viçosa” for availability of the computational cluster and software used in this work, the “Núcleo de Pesquisa e Conservação de Orquídeas” (NPCO) at “Universidade Federal de Viçosa” for maintenance of studied plants, and the Mass Spectrometry Facility at Brazilian Biosciences National Laboratory (LNBio), CNPEM, Campinas, Brazil for their support on mass spectrometry analysis.

Capítulo II

Carbohydrate-active enzymes from *Penicillium ochrochloron* RLS11 to improve commercial enzyme mixtures for plant biomass saccharification

Túlio Morgan [a, b], Daniel Luciano Falkoski [c], Murillo Peterlini Tavares [b], Mariana Bicalho Oliveira [b], Valéria Monteze Guimarães [b], Tiago Antônio de Oliveira Mendes [b]

[a] Postgraduate Program in Bioinformatics, Federal University of Minas Gerais, Av. Presidente Antônio Carlos, 6627 - Pampulha, 31270-901, Belo Horizonte, Minas Gerais, Brazil
[b] Departament of Biochemistry and Molecular Biology, Federal University of Viçosa, Av. PH Rolfs, s/n, 36570-900 Viçosa, MG, Brazil
[c] Novozymes Latin America - R. Prof. Francisco Ribeiro, 683 - Barigui, Araucária, PR - 83707-660, Brazil.

Carbohydrate-active enzymes from *Penicillium ochrochloron* RLS11 improve commercial enzyme mixtures for plant biomass saccharification

Túlio Morgan [a, b], Daniel Luciano Falkoski [c], Murillo Peterlini Tavares [b], Mariana Bicalho Oliveira [b], Valéria Monteze Guimarães [b], Tiago Antônio de Oliveira Mendes [b, *]

[a] Postgraduate Program in Bioinformatics, Federal University of Minas Gerais, Av. Presidente Antônio Carlos, 6627 - Pampulha, 31270-901, Belo Horizonte, Minas Gerais, Brazil

[b] Departament of Biochemistry and Molecular Biology, Federal University of Viçosa, Av. PH Rolfs, s/n, 36570-900 Viçosa, MG, Brazil

[c] Novozymes Latin America - R. Prof. Francisco Ribeiro, 683 - Barigui, Araucária, PR - 83707-660, Brazil.

* Corresponding author. Tel.: +55 (31) 3612-5107. Email address: tiagoaomendes@ufv.br (T.A.O. Mendes)

Abstract

Filamentous fungi are prolific producers of carbohydrate-active enzymes (CAZymes) and important agents that carry out plant cell wall degradation in nature. The number of fungal species is frequently reported being in the millions range with a huge species diversity and genetic variability, reflecting on a wide repertoire of CAZymes that can be produced by these organisms. In this study, we evaluated the ability of previously selected ascomycete and basidiomycete fungi to produce plant cell wall-degrading enzyme (PCWDE) activities and the potential of the culture supernatants to increase the efficiency of the Cellic® CTec2/HTec2 mixture for steam-exploded sugarcane straw saccharification. In general, the ascomycetes produced higher (hemi)cellulolytic activity, while lignin-degrading activity was detected only for basidiomycetes. In saccharification assays, the culture supernatant of *Penicillium ochrochloron* RLS11 showed a promising supplementation effect on Cellic® CTec2/HTec2, and we conducted the whole-genome sequencing of this fungus. The size of the assembled genome was 38.06 Mbp and a total of 12,015 protein-coding genes were identified. The repertoire of PCWDE-coding genes was comparatively high among *Penicillium* spp. and showed an expansion in important cellulases and xylanases families, such as GH3, GH6, GH7 and GH11. A proteomic analysis indicated cellulases of *P. ochrochloron* RLS11 that probably enhanced the biomass saccharification performance of the Cellic® CTec2/HTec2, which included enzymes from GH3, GH6 and GH7 families. This enzymes are promising targets for biochemical characterization and inclusion in modern commercial CAZyme cocktails.

Keywords: Plant biomass saccharification; Cellic® CTec2/HTec2; *Penicillium*; whole-genome sequencing; proteomics.

Significance

Fungi comprise a vast and diverse group of eukaryotes, but only a small portion of species have already been identified and an even smaller number has been subjected to in-depth scientific studies. Thus, it is interesting to evaluate the biotechnological potential of these organisms and their metabolites, such as carbohydrate-active enzymes for biorefinery purposes. In this work, we reported the ability of poorly explored ascomycetes and basidiomycetes to produce carbohydrate-active enzymes, and the potential of their exoproteomes to increase the efficiency of the Cellic® CTec2/HTec2 (Novozymes A/S) for saccharification of pretreated sugarcane straw. The *Penicillium ochrochloron* RLS11 was identified as a good source of proteins to improve saccharification yields of the commercial enzyme cocktail, and genomic and proteomic analyzes indicated several cellulases (families GH3, GH6, GH7) and other enzymes that were probably related to the good supplementation effect on the Cellic® CTec2/HTec2. These enzymes may improve biomass saccharification processes and contribute to a better understanding of the structure-function of CAZymes. Besides this, the genome sequence of *P. ochrochloron* may contribute to other genomic and evolutionary studies regarding this important fungal genus.

1. Introduction

Fungi are a diverse and large group of eukaryotes, comprising species adapted to almost all terrestrial and aquatic ecosystems [1]. The widespread and multitude of lifestyles of fungi entail a huge genetic variability, even inside the same genus [2]. This diversity is accompanied by biological importance, such as its prominent role in soil carbon cycling, pathogenesis and plant symbiosis. These features may be highly dependent on the fungal ability to produce and secrete carbohydrate-active enzymes (CAZymes), that must act on a highly heterogeneous substrates, and thus, a huge diversity regarding sequence and function are frequently observed for these enzymes [3].

Filamentous fungi significantly diverge in the way that they decompose lignocellulose, which is a consequence of their genetic background driving the production and secretion of plant cell wall-degrading enzymes (PCWDEs). Soft-rot fungi include ascomycetes and basidiomycetes which mainly employ hydrolases to depolymerize polysaccharides from plant cell wall. Brown-rot species are basidiomycetes with ability to produce non-enzymatic oxidative systems, while white-rot fungi are prolific producers of lignin-degrading oxidoreductases [4]. Necrotrophic phytopathogenic fungi generally have a enriched repertoire of PCWDE-coding genes to rapidly and effectively decompose plant polysaccharides, while symbionts and biotrophic fungi generally posses fewer CAZyme-coding genes [5]. Although this classification of fungi provide a glimpse of plant cell wall-degrading capabilities, a closer look into the repertoire of such genes in fungal genomes may reveal considerable variations, even considering close related species. This knowledge was enriched by a large number of whole-genome sequences available and the comprehensive characterization of exoproteomes by mass spectrometry [6-8].

Despite being an industrially attractive process, the use of enzymes for plant biomass saccharification represents a significant proportion of the capital cost to generate value-added products from lignocellulose [9]. In this sense, there is a continuous effort to improve

(hemi)cellulolytic cocktails, which is driven by protein engineering and enzyme screening from wild organisms, aiming at the production or discovery of more efficient biocatalysts.

The production and commercialization of (hemi)cellulase mixtures for industrial processes are dominated by companies such as Novozymes (Cellic® enzyme cocktails) and DuPont (Accellerase® and Spezyme® enzyme cocktails). Although the composition of commercial enzyme mixtures are not known in detail, they are largely composed of *Trichoderma reesei* enzymes that were possibly improved by protein engineering techniques. Probably, the enzyme mixtures also contain β-glucosidase from *Aspergillus* sp., as well as other proteins that enhance lignocellulose conversion obtained from a variety of organisms, such as swollenins and LPMOs from AA9 family [10].

In this study, we evaluated the ability of filamentous fungi to produce PCWDEs when cultivated on pretreated sugarcane straw, as well as the application of these enzymes in the saccharification of the same biomass. We used a relatively small set of fungi in this work, which comprised fungal isolates identified as good CAZyme-producers in previous screening studies (data not published). The saccharification assays were conducted using the Cellic® CTec2/HTec2 (Novozymes A/S) supplemented with crude fungal enzyme preparations to access the ability of fungi to secrete proteins to improve the saccharification performance of the commercial enzyme mixture. Our analysis indicated that the culture supernatant of *Penicillium ochrochloron* strain RLS11 showed a promising supplementation effect and we conducted whole-genome sequencing of this fungus to uncover its PCWDEs coding-genes repertoire. Also, LC-MS/MS analysis was carried out to comprehensively characterize its exoproteome and infer the proteins that provided the good ability to supplement the Cellic® CTec2/HTec2 for sugarcane straw saccharification. The cellulases and xilanases of *P. ochrochloron* RLS11 identified here are promising targets for future experiments to improve modern commercial CAZyme preparations.

Materials and methods

Fungal strains and culture conditions for PCWDEs production

The ascomycete fungi (*Trichoderma* sp. J7, *Penicillium* sp. RLS11, and *Fusarium verticilliodes* AZB) were obtained from the mycologycal culture collection of the Biochemical Analysis Laboratory, Bioagro, Universidade Federal de Viçosa (Brazil). The basidiomycete fungi (*Pleurotus cornucopiae* PLOCOR, *Pleurotus floridanus* PLO09, *Pleurotus ostreatoroseus* PLO13, *Pholiota adiposa* KGM, *Pholiota nameko* PH01, *Hericium erinaceus* HE) were kindly provided by the Laboratory of Mycorrhizal Association, from the same University. These fungal isolates were chosen based on previous screening studies evaluating CAZyme production and activity. Each fungus was cultivated on potato-dextrose-agar (PDA) plates for 7 days at 28 ^{o}C in the dark. Ten agar plugs (Ø = 5.0 mm) of PDA-containing mycelia were added to sterile Erlenmeyer flasks (250 mL) containing 100 mL of propagation medium with the following composition: yeast extract 10 $g.L^{-1}$, peptone 20 $g.L^{-1}$, glucose 20 $g.L^{-1}$. The flasks were kept under agitation (150 rpm) for 7 days at 28 ^{o}C in the dark.

For the production of PCWDEs, 5 mL of propagation medium from each fungal culture were added to sterile Erlenmeyer flasks (125 mL) containing 55 mL of mineral solution and 1.2 grams of steam-exploded sugarcane straw (dry weight) as sole carbon source. The mineral solution consisted of NH_4NO_3, 2.0 $g.L^{-1}$; K_2HPO_4 , 2.0 $g.L^{-1}$; $MgSO_4$, 2.0 $g.L^{-1}$; NaCl, 1.0 $g.L^{-1}$; $CaCl_2$, 2.0 $g.L^{-1}$; citric acid, 2.8 $g.L^{-1}$; yeast extract, 3.0 $g.L^{-1}$; peptone, 2.0 $g.L^{-1}$ and 1.0 $mL.L^{-1}$ of trace elements solution (EDTA, 50.0 $g.L^{-1}$; $MnCl_2$, 0.4 $g.L^{-1}$; $CoCl_2$, 0.16 $g.L^{-1}$; $CuSO_4$, 0.16 $g.L^{-1}$; H_3BO_3, 1.1 $g.L^{-1}$; $(NH_4)_6Mo_7O_{24}$, 0.13 $g.L^{-1}$, $FeSO_4$, 0.5 $g.L^{-1}$ e $ZnSO_4$, 2.2 $g.L^{-1}$). The flasks were kept at 28 ^{o}C under agitation of 250 rpm in the dark. The ascomycetes were grown for 10 days and the basidiomycetes fungi for 15 days, and aliquots were periodically taken from the culture medium during these

periods. For extracellular enzyme recovery, the residual solids were filtered using a nylon filter followed by centrifugation at 15,000*g* for 10 minutes at 4 °C. The clarified supernatants were kept at 4 °C and phenylmethanesulfonyl fluoride (PMSF) was immediately added to a final concentration of 1.0 mmol.L^{-1}. All experiments were performed in triplicate.

Plant cell wall-degrading enzyme activities

Endo-β-1,4-glucanase (CMCase) and β-1,4-xylanase activities were measured with carboxymethyl cellulose (CMC) and beechwood xylan as substrates, respectively. The enzymatic reactions were conducted in test tubes with substrate suspension (1.0 % w/v) in sodium acetate buffer (pH 5.0) and appropriately diluted enzyme preparation. The tubes were incubated at 50 ºC for 15 minutes. The released reducing sugars were quantified using dinitrosalicylic acid reagent [11]. Enzymatic activity units (U) were determined by the correlation between absorbance at 540 nm and known concentrations of analyte.

β-glucosidase, β-xylosidase and cellobiohydrolase activities were quantified using 4-nitrophenyl sugar derivatives as substrates (β-D-glucopyranoside, β-D-xylopyranoside and β-D-cellobioside, respectively). The enzymatic reactions were carried out on substrate suspension (0.5 mmol.L^{-1}) in sodium acetate buffer (pH 5.0) and appropriately diluted enzyme preparation. After 15 minutes at 50 ºC, the reactions were terminated by the addition of NaOH (1 mol.L^{-1}). The 4-nitrophenolate was quantified by the correlation between absorbance at 410 nm and known concentrations of analyte.

Laccase activity was measured by monitoring the oxidation of the ABTS (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)). The enzymatic reactions contained 300 µL of the appropriately diluted enzyme suspension, 600 µL of sodium acetate buffer (0.05 mol.L^{-1} pH 5.0) and 100 µL of 0.01 mol.L^{-1} ABTS. This mixture was incubated for 5 min at 50 ºC. After that, the

absorbance was taken at 420 nm. Laccase activity was calculated using the Lambert-Beer law, using a molar extinction coefficient of $3.6 \times 10^4\ M^{-1}cm^{-1}$.

The manganase peroxidase activity (MnP) was carried out using 100 µL of 0.25 $mol.L^{-1}$ sodium lactate, 50 µL of 0.002 $mol.L^{-1}$ $MnSO_4$, 20 µL of 0.5 % (m/v) bovine serum albumin, 100 µL of 0.1 % (m/v) phenol red, 500 µL appropriately diluted enzyme suspension and 50 µL of 0.002 $mol.L^{-1}$ H_2O_2 in sodium succinate buffer (0.025 $mol.L^{-1}$ and pH 4.5). The mixture was incubated for 5 minutes at 30 ºC and the reaction was ended by the addition of 40 µL of 2 $mol.L^{-1}$ NaOH. Control assays were conducted in the absence of Mn^{2+} by omitting the addition of $MnSO_4$ in the reaction mixture. The absorbance was taken at 610 nm, subtracting the value of the control reaction. MnP activity was calculated using the Lambert-Beer law, using a molar extinction coefficient of the oxidized phenol red (0.022 $M^{-1}cm^{-1}$).

One unit of enzymatic activity was defined as the amount of enzyme that released 1 µmol of product per minute under assay conditions.

Steam-exploded sugarcane straw saccharification by Cellic® CTec2/HTec2 supplementend with crude fungal enzyme preparations

The Cellic® CTec2/HTec2 enzyme mixture (75 % of CTec2 and 25 % of HTec2) was applied alone and in combination with fungal culture supernatants for steam-exploded sugarcane straw saccharification. Reactions were conducted in 25 mL Erlenmeyer flasks with 5 mL of reaction medium consisting of pretreated plant biomass 5.0 % (w/v), 0.1 mol.L^{-1} sodium citrate buffer pH 5.0 and sodium azide 0.001 mol.L^{-1}. The reactions contained 5.00 mg of protein per gram of dry matter (referred as “Mix”), 5.50 mg of protein per gram of dry matter (referred as “Mix + 10 %”) or 6.25 mg of protein per gram of dry matter (referred as “Mix + 25 %”). The protein supplementations (Mix +10 % and Mix + 25 %) were conducted using the commercial mixture itself Cellic® CTec2/HTec2 (control) or the fungal culture supernatants (treatments).

The flasks were incubated at 250 rpm and 50 ^{o}C for 72 hours. Aliquots were taken and immediately boiled to protein denaturation and centrifuged at 16000*g* for 15 minutes. Glucose were quantified in a high-performance liquid chromatography system equipped with Aminex HPX-87H column (300 x 7.8 mm) and refractive index detector.

The experiment was performed in triplicate for each condition and the statistical analyses were conducted in the R environment (version 3.5.3) [12]. The T-test was used for comparison of means. The null hypothesis was defined as the control mean higher than the treatment mean and the alternative hypothesis consisted of equal means or control mean less than treatment mean. The significance level was 5 % for all conclusions.

DNA extraction, genome sequencing and assembly

For genomic DNA extraction, *Penicillium* sp. RLS11 was cultured on PDA plates and the mycelium was collected with a sterile scalpel and immediately frozen with liquid nitrogen. DNA isolation was carried out according to a method described previously [13].

The genome was sequenced using Illumina Novaseq platform (Illumina Inc., San Diego, CA, USA). The library was prepared using TruSeq Nano DNA prep kit (Illumina Inc., San Diego, CA, USA) with 150 bp pair-ends reads and 550 bp fragment length. Raw data filtering was performed with Trimmomatic v0.36 [14] using 4-mer sliding-window and mean *Q* equal to 15. Filtered reads were used for *de novo* genome assembly using SPAdes v.3.9.0 [15] using k-mer sizes ranging from 27 to 123. The genome completeness was accessed with BUSCO v3 [16], using the *Ascomycota* and *Eurotiomycetes* single-copy orthologous genes.

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession PRJNA614650. The version described in this paper is version PRJNA614650.

Phylogenetic analysis

Initially, *ITS* regions (internal transcribed spacer 1 – 5.8S – internal transcribed spacer 2) were searched in the assembled genome of *Penicillium* sp. RLS11 using BLASTn searches with sequences from Fungi RefSeq ITS project (NCBI BioProject: **PRJNA177353**). After identifying the fungal species that provided the best-hit (based on sequence identity), its RNA polymerase II subunit 2 (*RPB2*) and β-tubulin (*TUB*) sequences were downloaded from Genbank and used in new BLASTn searches to identify these genes in the *Penicillium* sp. RLS11 assembly.

For phylogenetic analysis, reference sequences of *RPB2*, *TUB* and *ITS* regions were downloaded from GenBank (Table S1). These sequences were separately aligned with MAFFT v.7.305 [17] using the L-INS-i strategy. The three multiple sequence alignment files generated were combined with a custom Python script.

The jmodeltest v.2.1.10 [18] was used to infer the best substitution model under akaike information criterion. Phylogenetic reconstruction using the maximum likelihood method was carried out in RaxML v.8.2.12 [19] using the GTR+I+G model with 2,000 searches for the best-

scoring tree. Maximum likelihood bootstrap proportions (MLBP) were calculated with 5,000 replicates. Bayesian inference was carried out with MrBayes v.3.2.7 [20] using Markov Chain Monte Carlo (MCMC) algorithm. Six independent runs with four chains each were conducted for 2,000,000 generations, sampling every 100th generation. The initial 25 % trees were discarded as burn-in phase and the remaining trees were used for estimating Bayesian inference posterior probability (BIPP) values. Trees were visualized in FigTree v.1.4.4 [21].

Gene prediction and annotation

A combination of *ab initio* and similarity-based methods were applied to predict protein-coding genes in the *Penicillium* sp. RLS11 genome assembly. Augustus v3.2.2 [22] trained with *Penicillium chrysogenum* NRRL 1951 gene parameters and GeneMark-ES v4.38 [23] using self-training mode were used for *ab initio* gene predictions. The similarity-based gene predictions were based on RNA-seq data from *Penicillium* species downloaded from NCBI Sequence Read Archive (Table S2). Transcriptome assembly was conducted with Tophat/Cufflinks pipeline [24] and the coding sequences were used as input for Exonerate v.2.2.0 [25], which predicted its most likely position in the *Penicillium* sp. RLS11 assembly using spliced-sequence alignments. The EvidenceModeler v.1.1.1 [26] was used to merge the genes predicted by *ab initio* and similarity-based methods applying a weighted consensus metric.

Carbohydrate-active enzymes annotation was carried out with dbCAN2 [27] using HMMER, DIAMOND and Hotpep predictions. Only CAZy families predicted by at least 2 tools were kept in the data set. The dbCAN2 predictions were refined with BLASTp (cutoff values: identity >=60 % and coverage >= 80 %) and InterProScan best-hits (e-value <= 1e-30) aiming to annotate enzymes at activity level. This prediction was also applied to other *Penicillium* genomes used in this work (Table S3).

Mass spectrometry-based proteomics of culture supernatants of *Penicillium* sp. RLS11

Proteins of the 5-day and 10-day culture supernatants were precipitated with ethanol (90 % v/v) for 24 hours at -20 ºC and centrifugated at 9000*g* for 20 minutes and 4 °C. The protein pellet was resuspended in 100 µL of a solution containing 7.0 mol.L^{-1} urea, 2.0 mol.L^{-1} thiourea and 4.0 % (w/v) CHAPS. A SDS-PAGE was conducted with 30 µg of proteins and the run was stopped when proteins migrated from stacking gel to resolving gel. The unique protein bands were excised, discolored with methanol 50 % and acetic acid 5.0 %, treated with 0.05 mol.L^{-1} dithiothreitol and 0.10 mol.L^{-1} iodoacetamide, and subjected to in-gel trypsinization with 2 µg of trypsin (Trypsin Gold, Promega™). The peptides were desalted using stage-tips [28], dried in a vacuum and reconstituted in 0.10 % formic acid. One microliter of peptide-containing solution were injected on PicoFrit Column (20 cm x ID75 µm, 5 µm particle size, New Objective) and analyzed on an Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) connected to the EASY-nLC system (Proxeon Biosystem, West Palm Beach, FL, USA). Protein identification was performed with MaxQuant v.1.6.3.3 [29] using a database of predicted proteins sequences from *Penicillium* sp. RLS11, *P. oxalicum* 114-2, *P. brasilianum* and *P. rubens* Wisconsin 54-1255. Quantification was performed using the label-free quantification (LFQ) method. False discovery rate (FDR) of 1 % and a minimum of 2 unique peptides were used as identification confidence parameters.

The experiment was conducted in triplicate for each condition (5-day and 10-day supernatants) and statistical and exploratory data analysis were conducted in the R environment (version 3.5.3) [11]. The T-test was used to compare the means (protein relative abundance). The null hypothesis was defined as equal means and the alternative hypothesis as different means. The significance level was 5 % for all conclusions.

Results

Plant cell wall-degrading enzyme activities in fungal culture supernatants

Although enzymatic cocktails are already used in industrial processes to generate plant biomass-derived products, they are subjected to continuous improvements aiming at better saccharification efficiency. In this sense, a deep inspection of high-performance fungal enzyme systems could provide important information regarding new or more efficient biocatalysts to supplement existing enzyme cocktails.

In this study, we evaluated the ability of 9 pre-selected fungal isolates (3 ascomycetes and 6 basidiomycetes) to produce enzymes for depolymerization of major cell wall components of steam-exploded sugarcane straw, and its potential to improve the saccharification yields conducted by the Cellic® CTec2/HTec2 enzyme cocktail.

Generally speaking, the culture supernatants of basidiomycetes showed less cellulase and xylanase activities (Table 1). All basidiomycetes produced laccase activity but only *Pleurotus floridanus* PLO09 was able to produce manganese peroxidase in detectable amounts. This fungus and the *Pleurotus ostreatoroseus* PLO13 showed higher PCWDEs activities than other basidiomycetes and thus, might be better suited for saccharification of pretreated plant biomasses.

Table 1. Plant cell wall-degrading enzyme activities detected in the culture supernatant of ascomicetes and basidiomicetes grown on steam-exploded sugarcane straw.

	Enzymatic activity ($U.mg^{-1}$)						
	CMCase	β-glucosidase	Cellobiohydrolase	Endo-Xylanase	β-xylosidase	Laccase	Manganese peroxidase
Fusarium verticillioides AZB	1.3 ± 0.1	0.1 ± 0.0	< 0.1	7.0 ± 0.1	< 0.1	nd	nd
Trichoderma sp. J7	1.1 ± 0.0	0.1 ± 0.0	< 0.1	96 ± 4.9	< 0.1	nd	nd
Penicillium sp. RLS11	4.0 ± 0.3	0.3 ± 0.0	0.1 ± 0.0	40 ± 2.2	< 0.1	nd	nd
Pleurotus floridanus PLO09	2.2 ± 0.1	0.1 ± 0.0	< 0.1	11 ± 0.7	nd	24 ± 1.8	< 0.1
Pleurotus ostreatoroseus PLO13	2.4 ± 0.1	0.2 ± 0.0	< 0.1	10 ± 0.4	< 0.01	0.1 ± 0.0	nd
Pleurotus cornucopiae PLOCOR	0.2 ± 0.0	< 0.1	< 0.1	2.1 ± 0.1	< 0.01	< 0.01	nd
Pholiota adiposa PKGM	0.3 ± 0.0	< 0.1	< 0.1	3.9 ± 0.4	< 0.01	1.2 ± 0.0	nd
Hericium erinaceus HE	0.2 ± 0.0	< 0.1	< 0.1	2.0 ± 0.3	< 0.01	1.1 ± 0.1	nd
Pholiota nameko PH01	1.4 ± 0.1	< 0.1	< 0.1	8.1 ± 0.6	< 0.01	0.8 ± 0.0	nd

nd: not detected
The values depicted represent the detected specific enzyme activity ± standard deviation
Values below 0.1 $U.mg^{-1}$ were represented as < 0.1

Regarding the ascomycetes, the *Trichoderma* sp. J7 produced a high xylanolytic activity, several orders of magnitude higher than the cellulase activities, while *A. zeae* AZB provided a more balanced set of cellulase and xylanase activities. In turn, the *Penicillium* sp. RLS11 was by far the best producer of cellulases specific activities (CMCase, β-glucosidase and cellobiohydrolase) as well as high xylanolytic specific activity. Hence, this fungus showed the best ability to produce proteins that are active enzymes or produced enzymes with higher efficiency to catalyze reactions on the model substrates. In fact, *Penicillium* species are proving to be promising sources of enzymes for biomass saccharification [30], producing a complete and balanced set of enzymes to catalyze polysaccharide conversion into fermentable sugars [31-33]. Therefore, the *Penicillium* sp. RLS11 showed the greatest potential among the fungi evaluated for plant biomass saccharification.

Steam-exploded sugarcane straw saccharification by Cellic® CTec2/HTec2 supplemented with fungal culture supernatants

The sugarcane straw was used as a model substrate to evaluate the ability of enzyme mixtures to convert polysaccharides into fermentable sugars, as it is abundant in Brazil and easily recovered from commercial plantations. For saccharification, the Cellic® CTec2/HTec2 (3:1) was supplemented with each fungal culture supernatant to access its effect on the glucose release by the commercial enzymatic mixture. The steam-exploded sugarcane straw had a minor amount of hemicelluloses (Table S4) leading to low xylose release in the enzymatic saccharification. Thus, the xylose concentration was not reported.

Overall, the culture supernatants of basidiomycetes promoted a significant decrease in glucose release from steam-exploded sugarcane straw, especially of *Pholiota nameko* PH01 and *Pholiota adiposa* PKGM (Figure 1A).

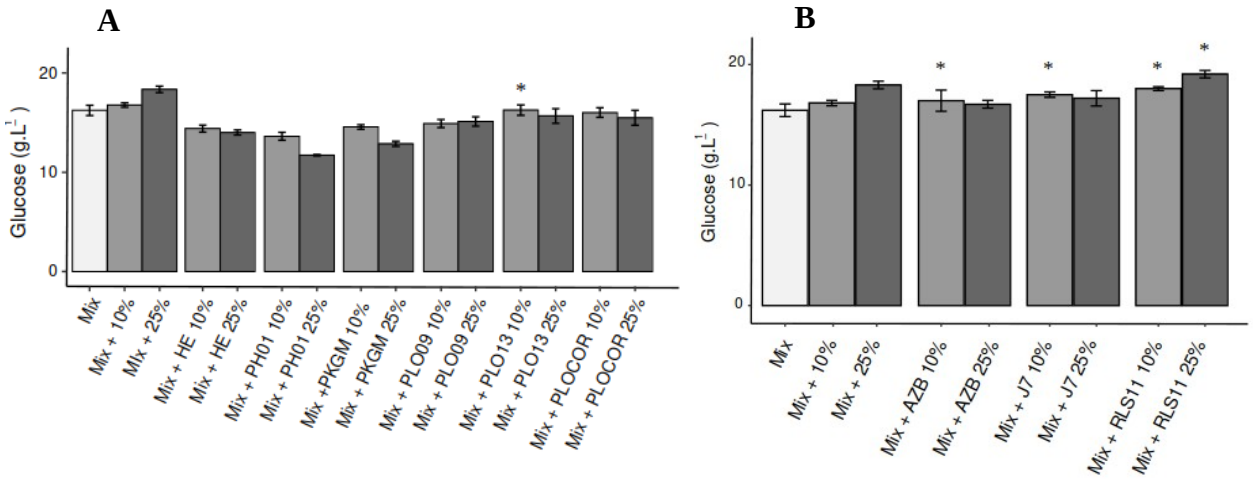


Figure 1. Glucose concentration ($g.L^{-1}$) in the reaction medium of steam-exploded sugarcane straw saccharification. The reaction was conducted with 5 % (w/v) dry matter loading for 72 hours at 50 ºC under orbital agitation of 250 rpm. The "**Mix**" refers to saccharification conducted by the Cellic® CTec2/HTec2 enzymatic mixture in a 3:1 ratio with a final concentration of 5.0 milligrams of protein per gram of dry matter (5 mg/g DM). The "Mix" was supplemented base on protein loading with the Cellic® CTec2/HTec2 itsef or with the culture supernatants of basidiomycetes (15 days of cultivation) and ascomycetes (10 days of cultivation). **Mix + 10 %**: Cellic® CTec2/HTec2 + Cellic® CTec2/HTec2 (0.5 mg/g DM); **Mix + 25 %**: Cellic® CTec2/HTec2 + Cellic® CTec2/HTec2 (1.25 mg/g DM).
(A) **Mix + HE 10 %** : Cellic® Ctec2/HTec2 + *Hericium erinaceus* HE (0.5 mg/g DM); **Mix + HE 25 %** : Cellic® CTec2/HTec2 + *Hericium erinaceus* HE (1.25 mg/g DM); **Mix + PH01 10 %** : Cellic® CTec2/HTec2 + *Pholiota nameko* PH01 (0.5 mg/g DM); **Mix + PH01 25 %** : Cellic® CTec2/HTec2 + *Pholiota nameko* PH01 (1.25 mg/g DM); **Mix + PKGM 10 %** : Cellic® CTec2/HTec2 + *Pholiota adiposa* KGM (0.5 mg/g DM); **Mix + PKGM 25 %** : Cellic® CTec2/HTec2 + *Pholiota adiposa* KGM (1.25 mg/g DM); **Mix + PLO09 10 %** : Cellic® CTec2/HTec2 + *Pleurotus floridanus* PLO09 (0.5 mg/g DM); **Mix + PLO09 25 %** : Cellic® CTec2/HTec2 + *Pleurotus floridanus* PLO09 (1.25 mg/g DM); **Mix + PLO13 10 %** : Cellic® Ctec2/HTec2 + *Pleurotus ostreatoroseus* PLO13 (0.5 mg/g DM); **Mix + PLO13 25 %** : Cellic® CTec2/HTec2 + *Pleurotus ostreatoroseus* PLO13 (1.25 mg/g DM); **Mix + PLOCOR 10 %** : Cellic® CTec2/HTec2 + *Pleurotus cornucopiae* PLOCOR (0.5 mg/g DM); **Mix + PLOCOR 25 %** : Cellic® CTec2/HTec2 + *Pleurotus cornucopiae* PLOCOR (1.25 mg/g DM).

(B) **Mix + 10% AZB**: Cellic® CTec2/HTec2 + *Fusarium verticillioides* AZB (0.5 mg/g DM); **Mix + 25% AZB**: Cellic® CTec2/HTec2 + *Fusarium verticillioides* AZB (1.25 mg/g DM); **Mix + 10% J7:** Cellic® CTec2/HTec2 + *Trichoderma* sp. J7 (0.5 mg/g DM); **Mix + 25% J7**: Cellic® CTec2/HTec2 + *Trichoderma* sp. J7 (1.25 mg/g DM); **Mix + 10% RLS11**: Cellic® CTec2/HTec2 + *Penicillium* sp. RLS11 (0.5 mg/g DM); **Mix + 25% RLS11**: Cellic® CTec2/HTec2 + *Penicillium* sp. RLS11 (1.25 mg/g DM).
* There is evidence that the treatment mean is equal to or greater than the control mean with 95 % of probability.

Increasing the protein loads from these fungi in the saccharification medium had a greater detrimental effect on Cellic® CTec2/HTec2 enzymes, since was detected lower glucose concentration relative to the control with less protein loading (Mix). The culture supernatant of *Pleurotus ostreatoroseus* PLO13 was the only from basidiomycetes that showed a promising result at 10% supplementation, providing a glucose release equal to or greater than the control supplemented with the commercial enzyme mixture itself at same protein loading (Mix + 10%). Nevertheless, the same effect was not observed for 25% supplementation (Mix + PLO13 25%). Thus, these results indicated that all evaluated basidiomycetes have poor ability to improve the commercial enzymatic mixture for saccharification.

Similarly, the culture supernatants of *Trichoderma* sp. J7 and *Fusarium verticilliodes* AZB provided significant results at 10% protein supplementation, but a lower glucose release when protein load was increased (25% protein supplementation) (Figure 1B). At first glance, these crude enzymatic preparations have no components that enhance the efficiency of Cellic® CTec2/HTec2 for sugarcane straw saccharification.

On the other hand, the secretome of *Penicillium* sp. RLS11 showed a promising result. The addition of this fungal culture supernatant to the Cellic® CTec2/HTec2 mixture led to positive supplementation effects on both protein loading levels (10% and 25%), that is, glucose release compatible to the commercial mixture alone at same protein loading. During growth, fungi probably produce and secrete proteins that do not participate in the biomass saccharification or even show a negative effect, such as proteases. Thus, the above results provided strong evidences that *Penicillium* sp. RLS11 produced more efficient enzymes or enzymatic activities that were not present in the Cellic® CTec2/HTec2. Some *Penicillium* species are efficient producers of PCWDEs

and have been gaining prominence on research related to enzymes for biofuel production [34]. Several reports indicated that cellulases produced by *Penicillium* species exhibit better catalytic performance than widely used cellulases from *Trichoderma reesei* [35-37], showing a potential to improve the conversion of plant biomasses into fermentable sugars.

Steam-exploded sugarcane straw saccharification by Cellic® CTec2/HTec2 supplemented with *Penicillium* sp. RLS11 culture supernatants

In order to confirm the potential of Penicillium sp. RLS11 to improve the Cellic® CTec2/HTec2 mixture, we conducted enzymatic saccharification assays supplementing the commercial mixture with culture supernatants of *Penicillium* sp. RLS11 obtained after 10 days of cultivation, as in previous saccharification assay, and culture supernatants obtained in other cultivation periods (3, 5 and 7 days). The 10-day culture supernatant yielded the same result as the previous assay, but the other culture supernatants did not showed such positive supplementation effect (Figure 2).

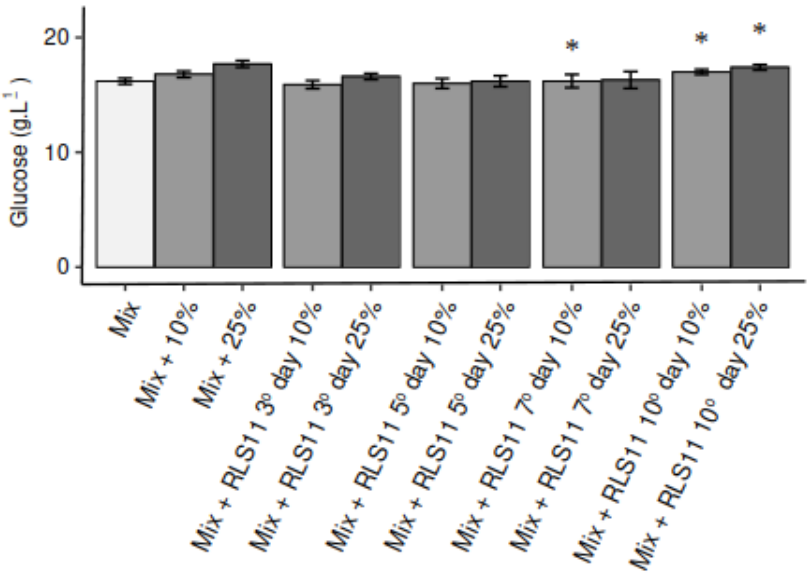


Figure 2. Glucose concentration ($g.L^{-1}$) in the reaction medium of steam-exploded sugarcane straw saccharification. The reaction was conducted with 5% (w / v) solids loading for 72 hours at 50 ºC under orbital agitation of 250 rpm. The "**Mix**" refers to saccharification conducted by the Cellic® CTec2/HTec2 enzymatic mixture in a 3:1 ratio with a final

concentration of 5.0 milligrams protein per gram of dry matter (5 mg/g DM). The "**Mix**" was supplemented, based on protein loading, with the Cellic® CTec2/HTec2 itsef or with the culture supernatants of *Penicillium* sp. RLS11 (3-day, 5-day, 7-day and 10-day culture supernatants). **Mix + 10 %**: Cellic® CTec2/HTec2 + Cellic® CTec2/HTec2 (0.5 mg/g DM); **Mix + 25 %**: Cellic® CTec2/HTec2 + Cellic® CTec2/HTec2 (1.25 mg/g DM). **Mix +RLS11 3º day 10 %** : Cellic® CTec2/HTec2 + 3-day culture supernatant of *Penicillium* sp. RLS11 (0.5 mg/g DM); **Mix +RLS11 3º day 25 %** : Cellic® CTec2/HTec2 + 3-day culture supernatant of *Penicillium* sp. RLS11 (1.25 mg/g DM); **Mix +RLS11 5º day 10 %** : Cellic® CTec2/HTec2 + 5-day culture supernatant of *Penicillium* sp. RLS11 (0.5 mg/g DM); **Mix +RLS11 5º day 25 %** : Cellic® CTec2/HTec2 + 5-day culture supernatant of *Penicillium* sp. RLS11 (1.25 mg/g DM); **Mix + RLS11 7º day 10 %** : Cellic® CTec2/HTec2 + 7-day culture supernatant of *Penicillium* sp. RLS11 (0.5 mg/g DM); **Mix + RLS11 7º day 25 %** : Cellic® CTec2/HTec2 + 7-day culture supernatant of *Penicillium* sp. RLS11 (1.25 mg/g DM); **Mix +RLS11 10º day 10 %** : Cellic® CTec2/HTec2 + 10-day culture supernatant of *Penicillium* sp. RLS11 (0.5 mg/g DM); **Mix +RLS11 10º day 25 %** : Cellic® CTec2/HTec2 + 10-day culture supernatant of *Penicillium* sp. RLS11 (1.25 mg/g DM).
* There is evidence that the treatment mean is equal to or greater than the control mean with 95 % of probability.

This suggested that proteins produced in the late stage of cultivation had the positive effect on the supplementation of Cellic® CTec2/HTec2 or these proteins accumulated in the medium during fungal colonization. Thus, we conducted proteomic analysis of the 10-day and 5-day culture supernatants to gain insights into which proteins are promoting the good supplementation effect.

Also, the genome sequencing and assembly of this fungus was important to gain knowledge on the genetic basis of its promising ability to supplement commercial enzyme mixtures and predict the sequence of genes to aid protein identification by proteomic analysis.

Penicillium sp. RLS11 genome and protein-coding genes prediction

The *Penicillium* sp. RLS11 genome sequencing generated 71,026,620 reads with 151 base pairs (bp) each, totalizing 10,725,019,620 bp. After quality control and raw data filtering, 69,872,948 reads and 10,550,815,148 bp were kept in the data set, representing 98.4 % of the total data generated. The assembled genome contained 991 contigs in 354 scaffolds (minimum length equal to 500 bp), with the largest scaffold of 2,094 Kbp. The N50 and L50 parameters were 832,608 bp and 16, indicating a good genome assembly (Table 2). The total size of the assembled genome was approximately 38 Mbp, featuring in the upper range of known genome sizes of *Penicillium* species (https://mycocosm.jgi.doe.gov) [38].

Table 2. Summary of the *Penicillium* sp. RLS11 genome assembly results.

Total sequenced bases	10,725,019,620
Number of scaffolds	354
Largest scaffold (bp)	2,093,928
Total length (bp)	38,056,224
GC content (%)	49.03
N50	832,608
L50	16
N's per 100 kbp	7.45
Completeness (*Ascomycota*) *	99.2 %
Completeness (*Eurotiomycetes*) *	97.7 %
# protein-coding genes	12,015
Average exons/gene	3.0
Average introns/gene	2.0

* The genome completeness was calculated using the *Ascomycota* and *Eurotiomycetes* universal single-copy orthologs with BUSCO v3 (Simão et al., 2015)

We employed *ab initio* and similarity-based approaches to predict protein-coding genes in the *Penicillium* sp. RLS11 assembly. The final consensus gene set comprised 12,015 protein-coding genes (supplementary file 1: protein sequences in fasta format) and a search for *Ascomycota* and *Eurotiomycetes* universal single-copy orthologs with BUSCO v3 yielded 99.2 % and 97.7 % of completeness, respectively, indicating good genome assembly and gene prediction.

Phylogenetic analysis

Preliminary searches with sequences from Fungi RefSeq ITS bioproject using BLASTn provided best-hits between *Penicillium* sp. RLS11 and sequences from *Penicillium ochrochloron* strains, with 100% sequence identity. The combined multiple sequence alignment of *RPB2*, *TUB* and *ITS* sequences was used for phylogenetic reconstruction and classification of *Penicillium* sp. RLS11.

The maximum likelihood (ML) and Bayesian inference analysis generated the same tree topology, and MLBP/BIPP values were depicted on the ML tree (Figure 3). All sections obtained

were in accordance with previous studies and the currently accepted *Penicillium* sections [39]. Regarding the *Penicillium* sp. RLS11, it grouped in a clade containing *Penicillium ochrochloron* strains with reliable bootstrap proportions and posterior probabilities, and thus, we classified *Penicillium* sp. RLS11 as *Penicillium ochrochloron* strain RLS11.

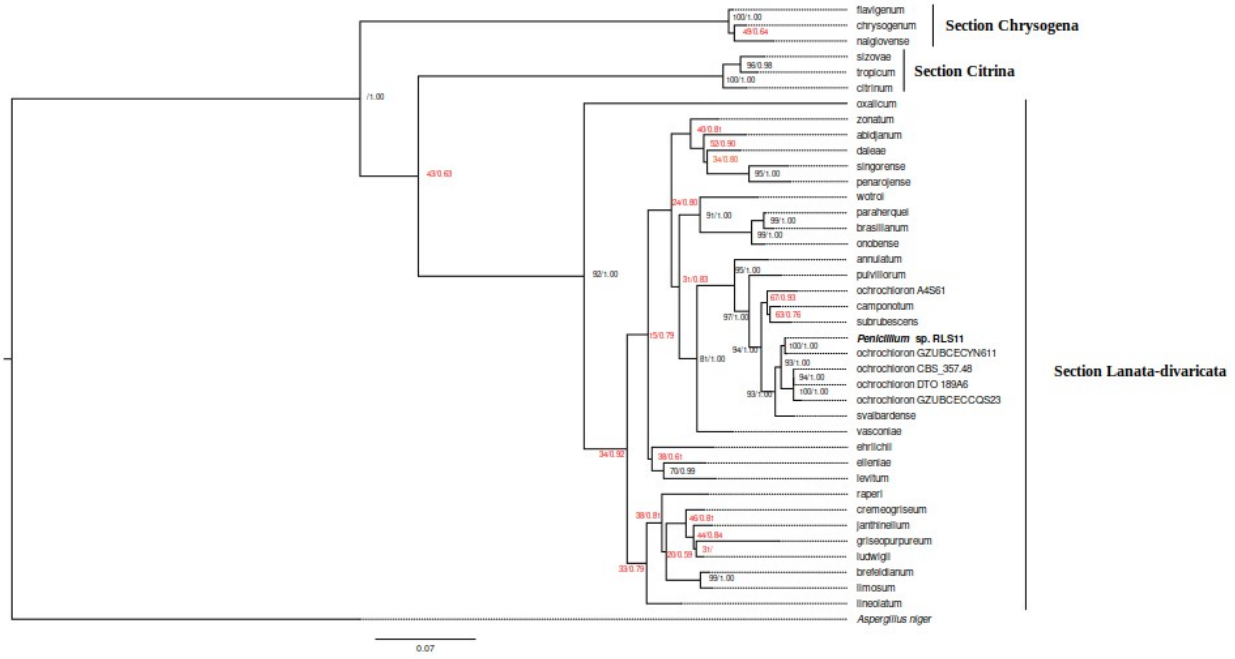


Figure 3. Best-scoring Maximum Likelihood tree of concatened sequence alignments of *RPB2* (RNA polymerase II subunit 2), *β-tubulin* and *ITS* (internal transcribed spacer 1 – 5.8S – internal transcribed spacer 2) from species belonging to *Penicillium* sections Chrysogena, Citrina and Lanata-divaricata. The tree is rooted to *Aspergillus niger*. MLBP above 70% (left) and BIPP above 95% (right) are indicated in black at the nodes. The *Penicillium* strain in boldface was used in this study. The Genbank accessions in given in Table S1.

Comparative analysis of CAZyme repertoire of *P. ochrochloron* RLS11 and other *Penicillium* spp.

The number of genes that encode CAZymes in *P. ochrochloron* RLS11 genome were compared to the other *Penicillium* species. The *Penicillium* genus contains species with good ability to depolymerize plant biomasses, being promising sources of enzymes [40] and already used in commercial preparations [34].

Among the 12,015 protein-coding genes of *P. ochrochloron* RLS11, 737 were predicted as carbohydrate-active enzymes. Glycosyl-hydrolases was by far the most abundant family with 391 predicted genes, followed by glycosyl-transferases (121 genes), carbohydrate esterases (105 genes), auxiliary activity enzymes (90 genes) and polysaccharide-lyases (12 genes) (Table S5).

Overall, *P. ochrochloron* RLS11 showed a higher number genes related to depolymerization of cellulose, xylans and pectins compared to other *Penicillium* species (Figure 4). This PCWDE-coding genes repertoire was more similar to that of *Penicillium subrubescens*, which is close related to *P. ochrochloron* (Figure 3) and is a good source of enzymes for plant biomass saccharification [30].

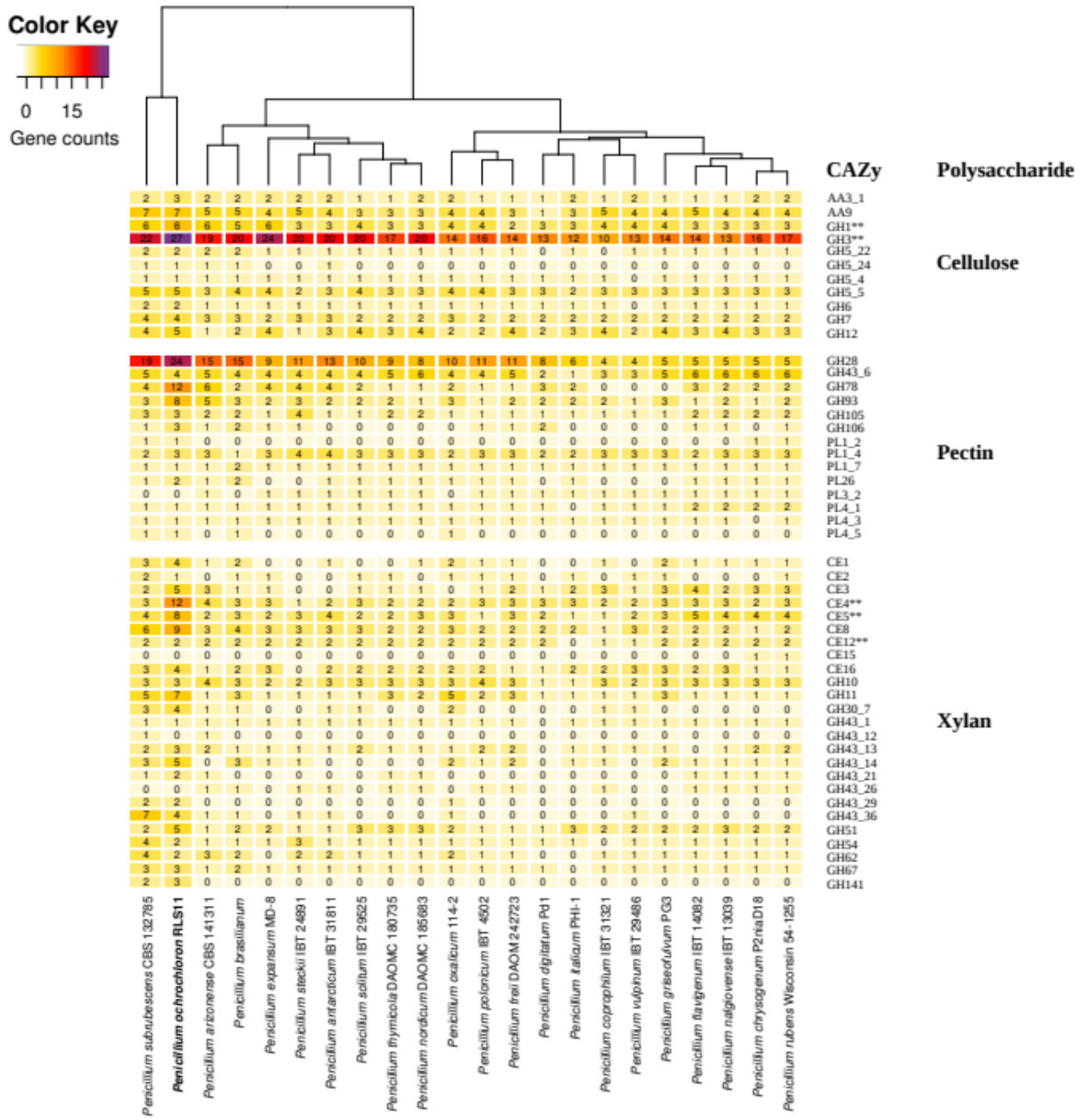


Figure 4. The putative inventory of carbohydrate-active enzymes (CAZymes) in selected *Penicillium* spp. genomes. CAZyme families were grouped according to their major polysaccharide substrate, which is given on the right side (cellulose, pectin, xylan). The number of genes belonging to CAZy families in each genome is indicated inside the frames. The cladogram on the top is based on the number of genes per CAZy family. CAZy families marked with ** contain members acting on other polysaccharides in addition to that depicted (see http://www.cazy.org/)

Cellulose is the main source of fermentable sugars in most plant biomasses, and thus, cellulases are crucial components of enzyme cocktails for biorefinary purposes. In this sense, *P. ochrochloron* RLS11 showed the largest number of protein-coding genes related to its depolymerization (65 genes), outperforming all other *Penicillium* species analyzed (min = 27 genes, max = 56 genes, s.d. = 7). This included oxidative (AA9 and AA3_1 families) and hydrolytic

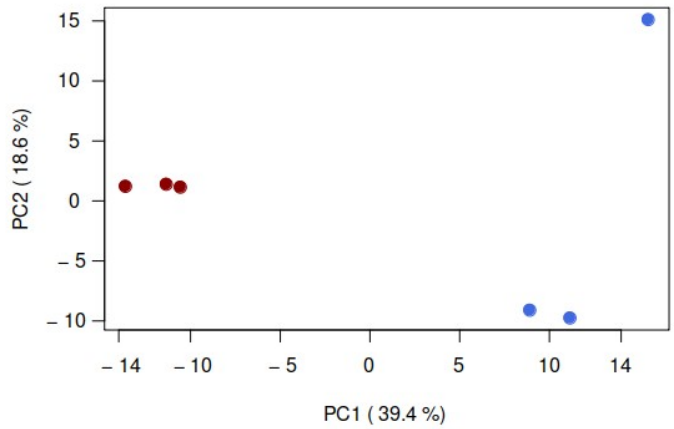
enzymes (GH3, GH5_5, GH6 and GH7 families). Furthermore, *P. ochrochloron* RLS11 genome harbors the largest number of endo-β-1,4-xylanases (GH10, GH11, GH30_7 families), α-arabinofuranosidases (GH51), β-xylosidases/α-arabinofuranosidases (GH43_14) and acetylxylan/feruloyl esterases (CE1, CE3, CE4, CE5). Taking together, such genetic features suggested that *P. ochrochloron* RLS11 might be very effective for saccharification of graminaceous feedstock, such as sugarcane straw/bagasse and corn stover, biomasses with high potential to support biorefineries [41].

Proteomic analysis of *P. ochrochloron* RLS11 culture supernatants

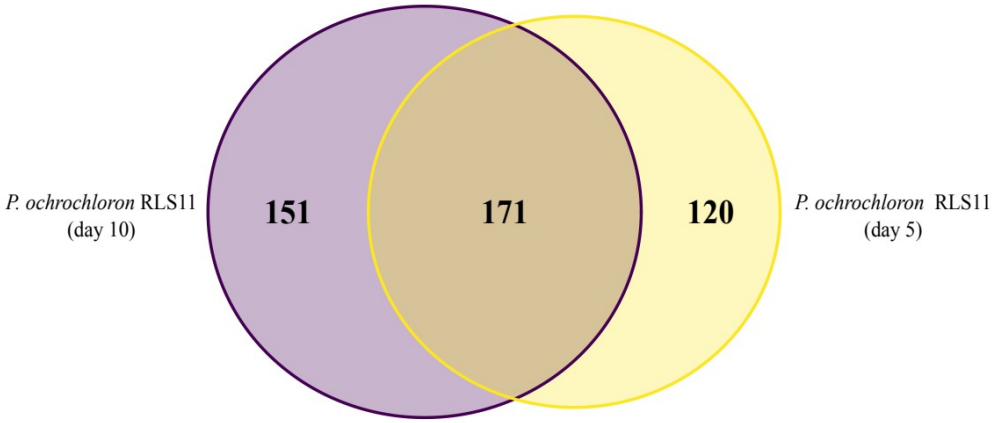
We conducted LC-MS/MS analysis of the *P. ochrochloron* RLS11 culture supernatants obtained after 5 and 10 days of growth to address the differences on protein composition between these samples (Tables S6 and S7). This could indicate proteins that led to the promising supplementation effect observed for the 10-day culture supernatant on Cellic® CTec2/HTec2 mixture.

To detect quantitative changes in the abundance of PCWDEs between the samples, we first accessed the similarity of replicates using principal component analysis (Figure 5A). Overall, the replicates were close in the bidimensional space using the two principal components that mostly encompasses the data variance (58 % of the total variance). Thus, the similarity of samples agrees with the experimental design.

A


15
10
5
0
– 5
– 10
PC2 (18.6 %)
– 14
– 10
– 5
0
5
10
14
PC1 (39.4 %)

B


P. ochrochloron RLS11
(day 10)
151
171
120
P. ochrochloron RLS11
(day 5)

C

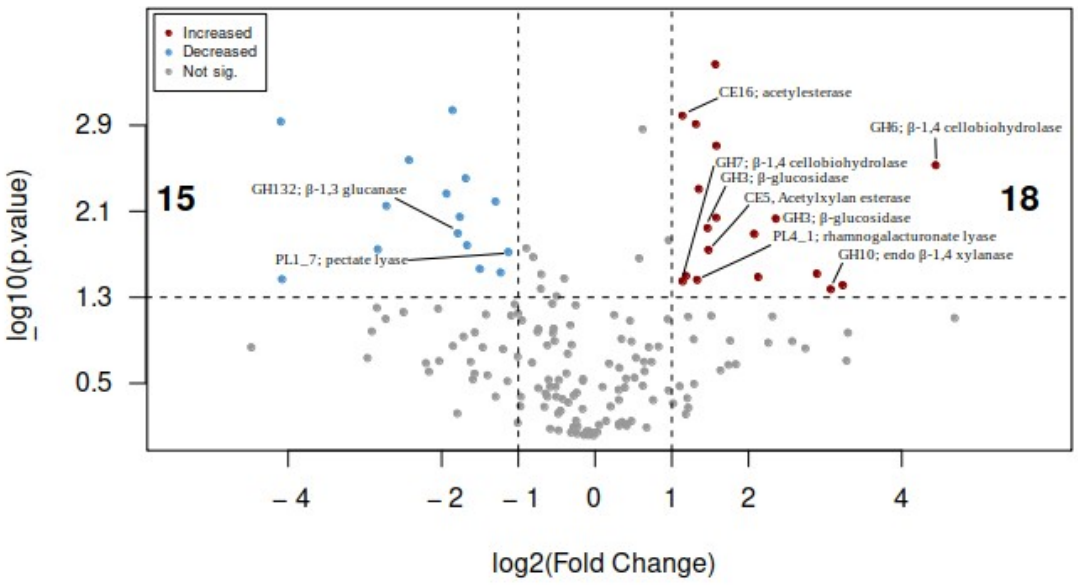

Increased
Decreased
Not sig.
15
18
CE16; acetylesterase
GH6; β-1,4 cellobiohydrolase
GH7; β-1,4 cellobiohydrolase
GH3; β-glucosidase
CE5, Acetylxylan esterase
GH3; β-glucosidase
PL4_1; rhamnogalacturonate lyase
GH10; endo β-1,4 xylanase
GH132; β-1,3 glucanase
PL1_7; pectate lyase
log10(p.value)
2.9
2.1
1.3
0.5
– 4
– 2
– 1
0
1
2
4
log2(Fold Change)

Figure 5. Exoproteome analysis by LC/MS-MS of 5-day and 10-day culture supernatants of *P. ochrochloron* RLS11. (A) Principal component analysis of 5-day culture supernatant (red dots) and 10-day culture supernatant (blue dots). PC1 explains 39.4 % of variance, while PC2 explains 18.6 % of variance. (B) Venn diagrams showing the number of proteins detected exclusively in each culture supernatant and the proteins found in both samples. (C) Volcano plot showing the significance (-log p-value > 1.30) and log2 fold-change of proteins with the 10-day culture supernatant proteome as reference. The red dots represent proteins with significantive higher relative abundance compared with the 5-day culture supernatant relative abundances, while the blue dots represent proteins with significantive lower relative abundance.

An overview of the number of proteins detected in each culture supernatant was given in the Venn diagram (Figure 5B). A total of 151 and 120 unique proteins were detected in 10-day and 5-day culture supernatants, respectively. Although a large number of exclusive proteins were detected, only a small portion of these corresponded to PCWDEs (Table 3). The 10-day culture supernatant had a higher number of unique cellulases, comprising β-glucosidases, exo- and endo-β-1,4-glucanases, which might contributed to the improvements of Cellic® CTec2/HTec2. A volcano plot was built to visualize the proteins that had a significant fold change between samples (Figure 5C). A $\log_2$(fold change) cutoff associated with statistical analysis (T test) indicated 18 proteins with higher abundance in the 10-day culture supernatant, which comprised GH6 and GH7 exo-β-1,4-glucanases (cellobiohydrolases) and GH3 β-1,4-glucosidases (T test, p-value < 0.05 and $\log_2$(fold change) > 1). These proteins are promising targets for biochemical characterization and supplementation of modern commercial enzyme cocktails to increase plant biomass saccharification efficiency.

Table 3. Unique carbohydrate-active enzymes of each *P. ochrochloron* RLS11 culture supernatant (day 5 and day 10).

***P. ochrochloron* RLS11 culture supernatant after 5 days of cultivation**			
CAZy family	Putative function*	Polysaccharide	Relative abundance** (%)
AA7	FAD linked oxidase	-	0.10
AA7	FAD linked oxidase	-	0.01
CBM63	Cellulose-binding	Cellulose	0.01
GH35	β-galactosidase	β-mannans	0.01
GH54	α-arabinofuranosidase	Xylan	0.02
***P. ochrochloron* RLS11 culture supernatant after 10 days of cultivation**			
CAZy family	Putative function*	Polysaccharide	Relative abundance** (%)
GH2	β-galactosidase	β-mannans	0.01
GH3	β-glucosidase	Cellulose	0.01
GH5_4	Endo β-1,4-glucanase	Cellulose	0.01
GH5_22	Endo β-1,4-glucanase	Cellulose	0.02
GH6	Exo β-1,4-glucanase	Cellulose	0.06
GH7	Endo β-1,4-glucanase	Cellulose	0.01
GH12	Endo β-1,4-glucanase	Cellulose	0.01

* The functional assignments were manually inspected using BLASTp best-hits (identity >=60% and coverage >= 80%) and Interpro domains (e-value <=1e-30) due to various enzymatic activities in the same CAZy family.

** The relative abundance was calculated taking into account all proteins detected in the respective culture supernatant.

Although a qualitative analysis indicated a high similarity between the two culture supernatants (Figures 6A and 6B), especially regarding endo-β-1,4-xylanases (GH10 and GH11) and β-1,4-glucanases (GH5, GH6 and GH7), the 10-day sample was more enriched with total cellulases than 5-day sample (10.9 % vs 4.1 % of total relative abundance). On the other hand, the relative abundances of the xylan-degrading enzymes were more similar (23.5 % vs 22.4 %, respectively). Such enzyme profiles may be a consequence of higher proportion of cellulose in steam-exploded sugarcane straw (Table S4) and accumulation of cellulose over xylan during the cultivation period, leading to more expression of cellulases than hemicellulases genes.

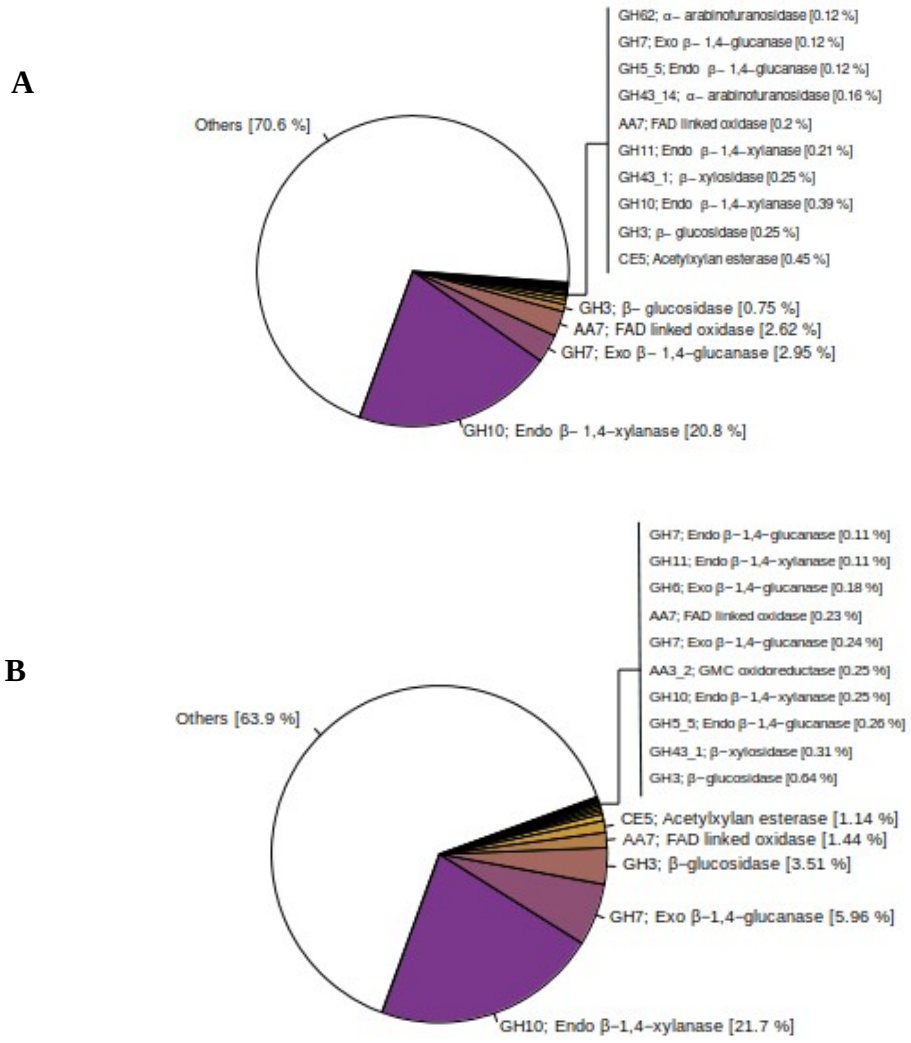


Figure 6. Exoproteome analysis of culture supernatants of *P. ochrochloron* RLS11 by LC/MS-MS. Relative abundances of plant cell wall-degrading enzymes detected in the (A) 5-day culture supernatant and (B) 10-day culture supernatant of *Penicillium ochrochloron* RLS11 grown on steam-exploded sugarcane straw. The “Others” sector comprised proteins that were annotated as CAZymes by dbCAN2 tool.

Discussion

Recent estimates place the number of fungal species in the millions range [42], and such magnitude can be linked to huge species diversity, genetic variability and the widespread across the globe [43]. This led to a variety of fungal lifestyles, which directly influences the repertoire of extracellular enzymes related to their nutritional strategies, such as CAZymes [5]. Therefore, the search for enzymes with superior biochemical-kinetic properties or even unknown enzyme activities

is justified, since there is a high number of unexplored fungal species, as well as CAZymes sequence variability [44].

In this work, we first accessed the ability of 9 fungal species from *Ascomycota* and *Basidiomycota* phyla to produce PCWDE activities when grown on steam-exploded sugarcane straw. Most of basidiomycetes showed poor ability to produce enzyme activities compared to ascomycetes (Table 1). This could be a consequence of insufficient cultivation time or not appropriated growth conditions, where these fungi may not effectively colonized the substrate and thus, not produced high amounts of extracellular CAZymes.

On the other hand, the ascomycetes showed higher potential to produce cellulases and xylanases, but no ability to produce lignin-degrading activities (e.g., laccase activity). Taking into account only the PCWDE activities detected in the culture supernatants, the enzyme mixtures produced by the ascomycetes were more suitable for saccharification of grass biomass, since this material generally have high proportion cellulose, especially after the pretreatment. The biomass pretreatment generally removes hemicelluloses and removes/redistributes the lignin on the surface of the lignocellulosic material. Therefore, cellulases are the main set of enzymes that drive the production of fermentable sugars from such biomasses.

Commercial enzyme mixtures are specifically tailored for biomass saccharification, containing protein stabilizers, a great proportion of proteins that are active (hemi)cellulases, and the absence of proteases. Thus, it is unfair to directly compare such enzyme cocktails with crude fungal enzymatic preparations. However, in the saccharification of steam-exploded sugarcane straw, the Cellic® CTec2/HTec2 supplemented with the 10-day culture supernatant of *P. ochrochloron* RLS11 provided a comparable glucose release to the commercial mixture alone at the same protein loading. This suggested that *P. ochrochloron* RLS11 may produced more efficient enzymes or enzyme activities absent from the Cellic® CTec2/HTec2 mixture, and thus, have potential to increase the efficiency of the commercial enzyme mixture for biomass saccharification.

The (hemi)cellulolytic enzyme system of *P. ochrochloron* RLS11 induced by steam-exploded sugarcane straw was composed mostly of hydrolases (families GH3, GH5, GH6, GH7, GH10, GH11) and a lower proportion of oxidative enzymes from AA3 and AA9 families (Figure 6A and 6B). Such enzyme profile seems to be common among *Penicillium* species since a similar trend was observed in other studies [33,45]. Surprisingly, an endo-β-1,4-xylanase was the most abundant enzyme in the *P. ochrochloron* RLS11 exoproteome, although steam-exploded sugarcane straw was composed primarily of cellulose and lignin with small amounts of xylan. This could be due to transcription factors that induce xylanase-coding genes expression in cellulose medium, such as POX02484 of *P. oxalicum* HP7-1 [46] (homolog of evm_model_NODE_31_72 of *P. ochrochloron* RLS11)

Furthermore, the xylanase activity was the highest for all fungi analyzed in this work. This suggested a similar induction and/or regulation of xylanase-coding genes, probably at the transcriptional level [47]. Transcription factors are poorly characterized in basidiomycetes compared to ascomycetes, but genomic studies indicate that transcription factor sets vary considerably in these two phyla [48]. Thus, the underlying mechanisms of regulation may vary, although the overall regulatory effect might be similar.

Although highly similar qualitatively, the relative abundance of PCWDEs between the culture supernatants (5-day and 10-day) of *P. ochrochloron* RLS11 varied substantially, which may be a consequence of variations of the carbon source composition over the cultivation period. It is well known that the secretion of PCWDEs by fungi is tightly regulated and largely dependent on the carbon source available. As the fungus grows and modifies the surrounding environment, a change in the requirement of enzymes may take place and the fungus adapts its metabolism for better growth performance. This can be evidenced by the greater abundance of cellulases in 10-day culture supernatant compared to the 5-day (more than twice), while a similar trend was not observed for total xylan-degrading enzymes.

The 10-day culture supernatant possessed many unique cellulases (Table 3) and other enzymes with higher relative abundance (Figure 5C, Figures 6A and 6B). The exo-β-1,4-glucanases (cellobiohydrolases - CBH) from GH6 and GH7 families are fundamental hydrolases for plant biomass saccharification and previous studies indicated that such enzymes from a *Penicillium* species possessed a better hydrolysis potential than CBH from *Trichoderma reesei* [36]. Thus, CBHs are very interesting targets for next experiments to pinpoint if these proteins may increment saccharification yields by Cellic® CTec2/HTec2.

Also, β-1,4-glucosidases are essential cellulases to alleviate the inhibitory effects of cellobiose on CBHs and endo-β-1,4-glucanases, being important enzymes to improve enzymatic saccharification yields [49]. A β-1,4-glucosidase from GH3 family was 6x more abundant in the 10-day culture supernatant *P. ochrochloron* RLS11 and thus, might be involved in the positive supplementation effect observed for this sample.

These results are the first steps towards improvements of modern commercial enzyme preparations for plant biomass depolymerization and have shown the potential of poorly explored fungi to produce efficient PCWDEs. Also, biochemical and kinetics characterization of the cellulases identified here may provide deeper knowledge on the structure-function of enzymes with superior catalytic properties. Taking together, the results presented here may contribute to the development of biotechnology regarding the production of plant biomass-derived biofuels and biomaterials.

Conclusions

Poorly explored fungi are powerful sources of carbohydrate-enzymes and have potential to improve modern commercial enzyme mixtures, such as Cellic® CTec2/HTec2 for biomass saccharification.

Genomic, proteomic and functional assays evidenced the ability of *Penicillium ochrochloron* RLS11 to produce CAZymes for plant biomass saccharification, as well as provided a wealth of “omics” data that may be used for other biotechnological applications involving this fungus.

The *P. ochrochloron* RLS11 cellulases from GH3, GH6 and GH7 families are promising targets for next experiments aiming at the improvement of modern commercial enzyme mixtures for biomass saccharification.

References

[1] Blackwell, M. The fungi: 1, 2, 3 ... 5.1 million species? Am. J. Bot. 2011; 98:426–438. https://doi.org/10.3732/ajb.1000298

[2] Galagan, J.E., Henn, M.R., Ma, L.J., Cuomo, C.A., Birren, B. Genomics of the fungal kingdom: Insights into eukaryotic biology. Genome Res. 2005; 15:1620–1631. https://doi.org/10.1101/gr.3767105

[3] Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 2009; 37:D233-238. https://doi.org/10.1093/nar/gkn663

[4] Kameshwar, S.A.K., Qin, W. Comparative study of genome-wide plant biomass-degrading CAZymes in white rot, brown rot and soft rot fungi. Mycology 2017; 9:93–105. https://doi.org/10.1080/21501203.2017.1419296

[5] Zhao, Z., Liu, H., Wang, C., Xu, J.R. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics 2013; 14(274):1-12. https://doi.org/10.1186/1471-2164-15-6

[6] Couturier, M., Navarro, D., Favel, A., Haon, M., Lechat, C., Lesage-Meessen, L., Chevret, D., Lombard, V., Henrissat, B., Berrin, J.G. Fungal secretomics of ascomycete fungi for biotechnological applications. Mycosphere 2016; 7:1546–1553. https://doi.org/10.5943/mycosphere/si/3b/6

[7] Presley, G.N., Panisko, E., Purvine, S.O., Schilling, J.S. Coupling secretomics with enzyme activities to compare the temporal processes of wood metabolism among white and brown rot fungi. Appl. Environ. Microbiol. 2018; 84:1–12. https://doi.org/10.1128/AEM.000159-18

[8] Rytioja, J., Hildén, K., Yuzon, J., Hatakka, A., de Vries, R.P., Mäkelä, M.R. Plant-Polysaccharide-Degrading Enzymes from Basidiomycetes. Microbiol. Mol. Biol. Rev. 2014 ; 78:614–649. https://doi.org/10.1128/mmbr.00035-14

[9] Humbird, D., Davis, R., Tao, L., Kinchin, C., Hsu, D., Aden, A., Schoen, P., Lukas, J., Olthof, B., Worley, M., Sexton, D., Dudgeon, D. Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol: Dilute-Acid Pretreatment and Enzymatic Hydrolysis of Corn Stover. Natl. Renew. Energy Lab. 2011; NREL/TP-5100-47764:1-147. https://doi.org/10.2172/1107470

[10] Gusakov, A. V. Alternatives to Trichoderma reesei in biofuel production. Trends Biotechnol. 2011; 29:419–425. https://doi.org/10.1016/j.tibtech.2011.04.004

[11] Miller, G.L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal. Chem. 1959; 31:426–428.

[12] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. URL http://www.R-project.org/.

[13] Doyle, J.J., Doyle, J.L. Isolation of plant DNA from fresh tissue. Focus (Madison). 1990; 12:13–15.

[14] Bolger, A.M., Lohse, M., Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

[15] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 2012; 19:455–477. https://doi.org/10.1089/cmb.2012.0021

[16] Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 2015; 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351

[17] Katoh, K., Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 2013; 30:772–780. https://doi.org/10.1093/molbev/mst010

[18] Darriba, D., Taboada, G.L., Doallo, R., Posada, D.,. jModelTest 2: more models, new heuristics and high-performance computing. Nat. Methods 2015; 9(8)-772. https://doi.org/10.1038/nmeth.2109.jModelTest

[19] Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014; 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033

[20] Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P. Mrbayes 3.2: Efficient bayesian phylogenetic

inference and model choice across a large model space. Syst. Biol. 2012; 61:539–542. https://doi.org/10.1093/sysbio/sys029

[21] Rambaut, A. FigTree v1.4.4, a graphical viewer of phylogenetic trees. 2018. Available at https://github.com/rambaut/figtree/releases.

[22] Stanke, M., Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 2003; 19:215–225. https://doi.org/10.1093/bioinformatics/btg1080

[23] Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O., Borodovsky, M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 2008; 18:1979–1990. https://doi.org/10.1101/gr.081612.108

[24] Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Patcher, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 2013; 7:562–578. https://doi.org/10.1038/nprot.2012.016.

[25] Slater, G.S.C., Birney, E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 2005; 6:1–11. https://doi.org/10.1186/1471-2105-6-31

[26] Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Robin, C.R., Wortman, J.R. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008; 9:1–22. https://doi.org/10.1186/gb-2008-9-1-r7

[27] Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P.K., Xu, Y., Yin, Y. DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018; 46:W95–W101. https://doi.org/10.1093/nar/gky418

[28] Rappsilber, J., Mann, M., Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. 2007; 2:1896–1906. https://doi.org/10.1038/nprot.2007.261

[29] Cox, J., Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 2008; 26:1367–1372. https://doi.org/10.1038/nbt.1511

[30] Vaishnav, N., Singh, A., Adsul, M., Dixit, P., Sandhu, S.K., Mathur, A., Puri, S.K., Singhania, R.R. Penicillium : The next emerging champion for cellulase production. Bioresour. Technol. Reports 2018; 2:131–140. https://doi.org/10.1016/j.biteb.2018.04.003

[31] Mäkelä, M.R., Mansouri, S., Wiebenga, A., Rytioja, J., de Vries, R.P., Hildén, K.S. Penicillium subrubescens is a promising alternative for Aspergillus niger in enzymatic plant biomass saccharification. N. Biotechnol. 2016; 33:834–841. https://doi.org/10.1016/j.nbt.2016.07.014

[32] Schneider, W.D.H., Gonçalves, T.A., Uchima, C.A., Reis, L. dos, Fontana, R.C., Squina, F.M., Dillon, A.J.P., Camassola, M. Comparison of the production of enzymes to cell wall hydrolysis using different carbon sources by Penicillium echinulatum strains and its hydrolysis potential for lignocelullosic biomass. Process Biochem. 2018; 66:162–170. https://doi.org/10.1016/j.procbio.2017.11.004

[33] Song, W., Han, X., Qian, Y., Liu, G., Yao, G., Zhong, Y., Qu, Y. Proteomic analysis of the biomass hydrolytic potentials of Penicillium oxalicum lignocellulolytic enzyme system. Biotechnol. Biofuels 2016; 9:1–15. https://doi.org/10.1186/s13068-016-0477-2

[34] Gusakov, A. V, Sinitsyn, A.P. Cellulases from Penicillium species for producing fuels from biomass. Biofuels 2012; 3:463–477. https://doi.org/10.4155/bfs.12.41

[35] Chekushina, A. V., Dotsenko, G.S., Sinitsyn, A.P. Comparing the efficiency of plant material bioconversion processes using biocatalysts based on Trichoderma and Penicillium verruculosum enzyme preparations. Catal. Ind. 2013; 5:98–104. https://doi.org/10.1134/S2070050413010042

[36] Taylor, L.E., Knott, B.C., Baker, J.O., Alahuhta, P.M., Hobdey, S.E., Linger, J.G., Lunin, V. V., Amore, A., Subramanian, V., Podkaminer, K., Xu, Q., Vanderwall, T.A., Schuster, L.A., Chaudhari, Y.B., Adney, W.S., Crowley, M.F., Himmel, M.E., Decker, S.R., Beckham, G.T. Engineering enhanced cellobiohydrolase activity. Nat. Commun. 2018; 9:1–10. https://doi.org/10.1038/s41467-018-03501-8

[37] Thygesen, A., Thomsen, A.B., Schmidt, A.S., Jørgensen, H., Ahring, B.K., Olsson, L. Production of cellulose and hemicellulose-degrading enzymes by filamentous fungi cultivated on wet-oxidised wheat straw. Enzyme Microb. Technol. 2003; 32:606–615. https://doi.org/10.1016/S0141-0229(03)00018-8

[38] Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I. MycoCosm portal: Gearing up for 1000 fungal genomes. Nucleic Acids Res. 2014; 42:699–704. https://doi.org/10.1093/nar/gkt1183

[39] Houbraken, J., Samson, R.A. Phylogeny of Penicillium and the segregation of Trichocomaceae into three families. Stud. Mycol. 2011; 70:1–51. https://doi.org/10.3114/sim.2011.70.01

[40] Saini, R., Saini, J.K., Adsul, M., Patel, A.K., Mathur, A., Tuli, D., Singhania, R.R. Enhanced cellulase production by Penicillium oxalicum for bio-ethanol application. Bioresour. Technol. 2015; 188:240–246. https://doi.org/10.1016/j.biortech.2015.01.048

[41] van der Weijde, T., Alvim Kamei, C.L., Torres, A.F., Vermerris, W., Dolstra, O., Visser, R.G.F., Trindade, L.M. The potential of C4 grasses for cellulosic biofuel production. Front. Plant Sci. 2013; 4:1–18. https://doi.org/10.3389/fpls.2013.00107

[42] Wu, B., Hussain, M., Zhang, W., Stadler, M., Liu, X., Xiang, M. Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. Mycology 2019; 10:127–140. https://doi.org/10.1080/21501203.2019.1614106

[43] Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Ruiz, L.V., Vasco-Palacios, A.M., Thu, P.Q., Suija, A., Smith, M.E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Rat, D., Abarenkov, K. Global diversity and geography of soil fungi. Science 2014; 346:1052–1053. https://doi.org/10.1126/science.aaa1185

[44] Helbert, W., Poulet, L., Drouillard, S., Mathieu, S., Loiodice, M., Couturier, M., Lombard, V., Terrapon, N., Turchetto, J., Vincentelli, R., Henrissat, B. Discovery of novel carbohydrate-active enzymes through the rational exploration of the protein sequences space. Proc. Natl. Acad. Sci. U. S. A. 2019; 116:6063–6068. https://doi.org/10.1073/pnas.1815791116

[45] Schneider, W.D.H., Gonçalves, T.A., Uchima, C.A., Couger, M.B., Prade, R., Squina, F.M., Dillon, A.J.P., Camassola, M. Penicillium echinulatum secretome analysis reveals the fungi potential for degradation of lignocellulosic biomass. Biotechnol. Biofuels 2016; 9:1–26. https://doi.org/10.1186/s13068-016-0476-3

[46] Zhao S, Yan YS, He QP, Yang L, Yin X, Li CX, et al. Comparative genomic, transcriptomic and secretomic profiling of Penicillium oxalicum HP7-1 and its cellulase and xylanase hyper-producing mutant EU2106, and identification of two novel regulatory genes of cellulase and xylanase gene expression. Biotechnol Biofuels 2016;9:1–17. https://doi.org/10.1186/s13068-016-0616-9.

[47] Amore, A., Giacobbe, S., Faraco, V. Regulation of Cellulase and Hemicellulase Gene Expression in Fungi. Curr. Genomics 2013; 14:230–249. https://doi.org/10.2174/1389202911314040002

[48] Todd, R.B., Zhou, M., Ohm, R.A., Leeggangers, H.A.C.F., Visser, L., de Vries, R.P. Prevalence of transcription factors in ascomycete and basidiomycete fungi. BMC Genomics 2014; 15:1–12. https://doi.org/10.1186/1471-2164-15-214

[49] de Andrade, L.G.A., Maitan-Alfenas, G.P., Morgan, T., Gomes, K.S., Falkoski, D.L., Alfenas, R.F., Guimarães, V.M. Sugarcane bagasse saccharification by purified β-glucosidases from Chrysoporthe cubensis. Biocatal. Agric. Biotechnol. 2017; 12:199–205. https://doi.org/10.1016/j.bcab.2017.10.007

Formatting of funding sources

This work was supported by the Conselho Nacional de desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). The authors thank the “Diretoria de Tecnologia de Informação” (DTI) at “Universidade Federal de Viçosa” for availability of the computational cluster and software used in this work, and the Mass Spectrometry Facility at Brazilian Biosciences National Laboratory (LNBio), CNPEM, Campinas, Brazil for their support on mass spectrometry analysis.

Capítulo III

Whole-genome sequence of *Fusarium verticillioides* AZB and insights into genomic evolutionary events and features associated to pathogenicity

Túlio Morgan [a, b], Murillo Peterlini Tavares [b], Valéria Monteze Guimarães [b], Tiago Antônio de Oliveira Mendes [b]

[a] Postgraduate Program in Bioinformatics, Federal University of Minas Gerais, Av. Presidente Antônio Carlos, 6627 - Pampulha, 31270-901, Belo Horizonte, Minas Gerais, Brazil
[b] Departament of Biochemistry and Molecular Biology, Federal University of Viçosa, Av. PH Rolfs, s/n, 36570-900 Viçosa, MG, Brazil

Whole genome sequence of *Fusarium verticillioides* AZB and insights into genomic evolutionary events and features associated to pathogenicity

Túlio Morgan [a, b], Murillo Peterlini Tavares [b], Valéria Monteze Guimarães [b], Tiago Antônio de Oliveira Mendes [b, *]

[a] Postgraduate Program in Bioinformatics, Federal University of Minas Gerais, Av. Presidente Antônio Carlos, 6627 - Pampulha, 31270-901, Belo Horizonte, Minas Gerais, Brazil

[b] Departament of Biochemistry and Molecular Biology, Federal University of Viçosa, Av. PH Rolfs, s/n, 36570-900 Viçosa, MG, Brazil

* Corresponding author. Tel.: +55 (31) 3612-5107. Email address: tiagoaomendes@ufv.br (T.A.O. Mendes)

Abstract

Fusarium verticillioides is a fungal plant-pathogen spread across the globe, being one of the major agents of ear rot in maize and grain contamination with mycotoxins. This species is part of the *F. fujikuroi* species complex (FFSC), comprising evolutionary related species but with varied lifestyle and plant hosts. The phenotypic diversity between close related species or even strains are due to their genetic background and are shaped by various evolutionary events. Here, we presented the genome sequence of *F. verticillioides* AZB, expanding the availability of genomes of this important cereal pathogen. Orthology analysis and gene expression data indicated many genes shared by diverse *Fusarium* species that might participate in plant colonization, as well as strain- and species-specific genes that may confer uniqueness for *F. verticillioides*. Many of the most variable genomic regions between *F. verticillioides* strains were within protein-coding genes that may be related to pathogenicity such as proteases, chitinases, LysM domain and VIT domain. Furthermore, diversifying selection was detected among 1826 orthologous genes of *F. verticillioides* and many of these showed sufficient sequence similarity to characterized virulence factors (PHI database, http://www.phi-base.org/). There was no correlation between the genomic positions of these genes and transposable elements (TEs) in *F. verticillioides* AZB genome, indicating that repetitive DNA might have little impact on genome evolution and fungal fitness. In fact, the TEs identified in *F. verticillioides* AZB were highly mutated/fragmented and together with genomic hallmarks of RIP, provided evidences of strong depletion of repetitive DNA in its genome. Since *F. verticillioides* populations tends to reproduce sexually, their evolution may be guided mainly by meiotic recombination.

1. Introduction

The genus *Fusarium* includes a large and diverse set of species, many of which are plant pathogens and produce an array of mycotoxins that contaminate food and feed-stocks. The ability to cause disease in a wide range of plant species associated with their huge genetic diversity encouraged many studies towards a better understanding of the *Fusarium* spp. biology and mechanisms of interaction with their hosts. In this context, the genomes of many *Fusarium* species have been sequenced and subjected to functional genomic studies and evolutionary analysis [1].

Many mycotoxins are reported to promote harmful effects on the health of humans and animals, and various important cereals are frequently reported to contain these toxic compounds, such as maize contaminated with fumonisins [2-4]. These highly toxic group of secondary metabolites are primarily produced by *Fusarium verticillioides*, which is also the causal agent of ear rot in maize [5] that reduces grain yield production. Therefore, *F. verticillioides* impair food production and quality, revealing the importance of deeper studies to characterize genomic features and evolutionary events that impact its pathogenicity and ability to produce toxins.

Intimate pathogen-host associations, such as *F. verticillioides*-maize, had impacts in the genome evolution of both pathogen and host populations [6] being mainly directed towards protein-coding genes that participate in the interaction. The plant resistance against *Fusarium* ear rot is a complex trait, as well as the pathogenic mechanisms employed by *Fusarium* spp., which include a diverse set of effectors, enzymes and secondary metabolites. For example, the deletion of effector genes in *F. oxysporum* (SIX1, SXI3, SIX5 e SIX6) decreased but not eliminated the pathogenicity against tomato, indicating that the pathogenic mechanisms were dependent on other virulence factors [7]. Therefore, employing comprehensive functional genomics and analyzing evolutionary events may be useful to unveil the genetic basis of pathogenicity and help developing better control strategies.

Fungal virulence genes may be identified using *in planta* gene expression or predicted from fungal genomes using evolutionary evidence such as diversifying selection and its proximity to transposable elements and other hotspots of genome evolution [8].

Transposable elements may cause an increase in the rate of DNA mutation, promote large insertions and deletions in the genome, as well as gene duplication events. Thus, the identification and classification of TEs are essential to understand their evolutionary effects on the genome. Furthermore, horizontal gene/chromosome transfer between Fusarium strains could provide the ability to cause plant disease [9], improving fungal fitness. On the other hand, some *Fusarium* genomes not harbor supernumerary chromosomes (*F. graminearum*) or are enriched in repetitive regions (*F. verticillioides*) [7], suggesting that sexual recombination or other genomic features are the main drives of genome evolution.

In this work, we sequenced the whole genome of *Fusarium verticillioides* AZB isolated in Minas Gerais – Brazil, increasing the set of *F. verticillioides* genomes available. Comparative genomics and evolutionary analysis were conducted, providing knowledge about genomic features affecting fungal virulence. Using orthology analysis and gene expression data we identified orthologous genes shared by *Fusarium* spp. from *F. fujikuroi* species complex (FFSC) that may participate in pathogenicity, as well as *F. verticillioides* unique genes that may confer species-specific virulence features. In addition to the more widely recognized processes of diversifying selection, the genome evolutionary analysis indicated that gene gains and losses were important mechanisms of adaptation, occurring even at strain level. Although present in *F. verticillioides* genome, many TE-coding genes were putatively lost after *F. verticillioides* speciation and most of the remaining ones were fragmented or highly mutated. Furthermore, the presence of AT-rich blocks and other hallmarks of repeat-induced point mutations (RIP) indicated strong mechanisms to suppress TE proliferation and thus, repetitive DNA possibly has little impact on *F. verticillioides* genome evolution. Collectively, this study increased the availability of *Fusarium* spp. genomic resources

and identified genomic evolutionary events and features of *F. verticillioides*, especially regarding virulence factors.

2. Materials and methods

2.1. Microorganism isolation and storage

The *Fusarium verticillioides* AZB was isolated from maize grains and belongs to the mycological collection of Laboratory of Biochemical Analysis at Federal University of Viçosa, MG. The fungus was stored in glycerol 20 % (v/v), glycerol 50 % (v/v) and NaCl 0.9 % (m/v) at -80 ºC, and was routinely propagated on potato-dextrose-agar plates at 28 ºC.

2.2. DNA extraction, genome sequencing and assembly

Fusarium verticillioides AZB was cultured on potato-dextrose-agar (PDA) at 28 ºC for 5 days, the mycelium was collected and immediately frozen with liquid nitrogen. Genomic DNA extraction was performed by a method described elsewhere [10].

For genome sequencing, fragments library was prepared using TruSeq Nano DNA prep kit (Illumina Inc., San Diego, CA, USA) with 150 bp pair-ends reads and 550 bp fragment length and sequenced using Illumina Novaseq platform (Illumina Inc., San Diego, CA, USA). Raw data filtering was carried out with Trimmomatic v0.36 [11] using 4-mer sliding-window and mean *Q* equal to 15. SPAdes v.3.9.0 [12] was used for *de novo* genome assembly using k-mer sizes ranging from 27 to 123. Assembly metrics for quality assessment were retrieved using QUAST v.4.3 [13]. Searches for *Ascomycota* and *Eurotiomycetes* universal single-copy orthologous genes was performed with BUSCO v3 [14] to access the genome assembly completeness.

2.3. Gene prediction and annotation

The prediction of protein-coding genes in *F. verticillioides* AZB assembly was performed using *ab initio* and similarity-based methods. Augustus v3.2.2 [15] trained with *Fusarium graminearum* gene parameters and GeneMark-ES v4.57 [16] using self-training mode were used for *ab initio* gene finding. The similarity-based gene predictions were conducted with protein sequences downloaded from NCBI Protein database in February 2020 (search key = fusarium; source databases = RefSeq and UniProtKB/Swiss-Prot; 226,826 sequences retrieved). Spliced-sequence alignments were performed with Exonerate v.2.2.0 [17], which provided the most likely position of query protein sequences in the *F. verticillioides* AZB assembly. The genomic coordinates of genes predicted by the different gene finding approaches were submitted to EvidenceModeler v.1.1.1 [18] using equal weights to generate consensus gene structures.

2.4. Comparative genomics and phylogenetic analysis

Genomes and the predicted protein-coding genes of *Fusarium* spp. were downloaded from NCBI database (Table S1). When not available, gene prediction was performed as described in section 2.3.

Universal single-copy orthologous genes of Sordariomycetes (3,725 genes) were searched in each genome using BUSCO v3 [14]. Only the genomes that yielded completeness equal to or greater than 90.0 % were kept for the next analysis.

Protein orthogroups were inferred with Orthofinder v.2.3.3 [19] using default parameters and gene tree inference method (Mafft v.7.305 [20] and FastTree v.2.1.9 [21]). The corresponding nucleotide sequences of each protein were retrieved using a custom Python script. Nucleotide

sequences of single-copy orthologs were aligned using Mafft v.7.305 (parameters: --localpair --maxiterate 1500 --op 1.5 --ep 0.1) and the alignments generated were combined into a single sequence for each species using a custom Python script. Phylogenetic analysis was conducted using maximum likelihood method by RAxML v.8.2.12 [22] using the GTR+I+G model with 300 searches for the best-scoring tree. Maximum likelihood bootstrap proportions (MLBP) were calculated with 1000 replicates. Tree visualization and annotations were performed with “ggtree” package of R software [23].

To estimate gene gains and losses through the evolutionary model obtained, a boolean matrix representing the presence and absence of species in each orthogroup was fitted to the phylogenetic tree using dollop program from Phylip package v.3.696 [24].
Genome alignments were conducted using NUCmer algorithm of the MUMmer v.3.23 (http://mummer.sourceforge.net/), which is useful for large-scale global alignments such as a set of contigs to a finished genome sequence.

2.5. RNA-seq data collection and identification of up-regulated genes during maize infection

We downloaded publicly available transcriptome data of *Fusarium* spp. from *F. fujukuroi* species complex from the NCBI GEO (GSE77595). The RNA-seq data consisted of gene expression of five *Fusarium* strains while infecting maize roots and a control condition (CM agar *in vitro*). Fungal genes with a minimum of fourfold increase in expression [$\log_2$(FPKM+1) >= 2] during plant infection compared to the control condition were considered as up-regulated [25].

2.6. In silico *detection of repetitive DNA – tandem repeats and transposable elements*

The genomic coordinates of tandem repeats (TRs) and transposable elements (TEs) were inferred by *de novo* and similarity-based methods. For tandem repeats identification, a combination of seven tools was used: XSTREAM [26], T-Reks [27], Phobos [28], Tantan [29], TRF [30], etandem [31] and mreps [32]. The outputs from each tool were combined in a single file and the genomic coordinates were merged using bedtools merge algorithm [33] to remove redundancies (parameters: -c 4 -o count_distinct -d -10). Only the TRs identified by at least 5 tools were taken. For *de novo* detection of transposable elements, IFR [34] (parameters: 2 3 5 80 10 20 50000 10000 -a3 -t4 1000 -t5 5000 -d -h), RED [35] and RepeatModeler [36] were used. The output of each tool was analyzed with CD-HIT [37] to remove overlapping sequences (parameters: -c 1 -l 50 -aS 0.99 -d 1000). To classify the identified TEs, we build HMMs for each repeat class based on the RepBase-RepeatMaskerEdition-20181026 database. First, the database was split into multiple fasta files based on repeat classes. Sequences from each class (fasta file) were subjected to clustering using MCL v.14.137 [38] (parameters: --abc -I 1.2) based on the best length-normalized bitscore for each sequence pair (bitscore/(length sequence A * length sequence B)) obtained from pairwise alignments using BLASTn v.2.6.0 [39]. Next, sequences from each cluster were aligned with Mafft v.7.305 using the L-INS-i strategy and HMMs were build using HMMER v.3.2.1 [40]. Sequences of the same repeat class that were not clustered by MCL were merged in a single cluster and also subjected to HMM build. The nhmmer algorithm (HMMER v.3.2.1) was used to search each HMM query against a nucleotide sequence database (repetitive sequences identified by IRF, RED and RepeatModeler). Only the alignments with E-value <= 1e-02 and HMM coverage >= 50 % were used to classify the repeats, otherwise, the “unknown” tag was added.

The *F. verticillioides* specific repetitive sequences (TE-classified sequences and tandem repeats) were merged with RepBase-RepeatMaskerEdition-20181026 and the combined database was used as a library for RepeatMasker v.4.1.0 [41] (additional parameters: -gccalc -no_is).

2.7. Di-nucleotide frequency and AT-rich blocks detection

RIPCAL v.2.0 [42] was used to calculate di-nucleotide frequencies for each repeat family and non-repetitive sequences (control). We analyzed the frequencies of CpN and TpN di-nucleotides between each repeat class and control sequences, as well as the GpN to ApN transitions to consider the features in the reverse complement strand. Furthermore, we inferred the presence of AT-rich blocks in the genome using OcculterCut v.1.1 [43]. This tool calculates the %GC in genome segments and report frequencies within 1 % intervals (from 0 % to 100 %), providing a way to detect bimodal distribution of GC-content.

2.8. Testing for adaptive evolution in orthologous genes

First, we used the one-ratio model of codeml [44] (M0, model = 0, NSsites = 0, fix_omega = 0, omega = 0.2) to estimate the average dN/dS for whole sequences of single-copy orthologous genes of all *Fusarium* species from the African clade (all branches of this clade).

To test for selection acting on *F. verticillioides* protein-coding genes, we used codeml branch-site model A (model = 2, NSsites = 2), which allowed dN/dS to vary among codon sites and across branches of the phylogeny. By setting the branch leading to *Fusarium verticillioides* strains as the foreground branch, we compared a selection model that allowed codons on that branch to have dN/dS > 1 (Model A, fix_omega = 0) to a null model that constrained codons to have dN/dS <= 1 (Model A1, fix_omega = 1, omega = 1). The likelihood of model A was computed 3 times using 3 different initial omega values (0.5, 1.5 and 2.5) and the highest value was used in the likelihood ratio test (LRT) to check if the difference between the model A and model A1 was significant. A false discovery rate (FDR) of 1 % was applied and the adjusted p-values were reported as $-\log_2$(adjusted p-value).

3. Results

3.1. Genome assembly and protein-coding genes prediction

The genome of *Fusarium verticillioides* AZB was sequenced using whole-genome shotgun approach in an Illumina Novaseq platform, generating 55,603,914 reads and 8,396,191,014 base pairs. After quality control and raw data filtering, 54,382,388 reads and 8,211,740,588 base pairs were kept in the data set, representing 97.8 % of the total data generated. The assembled genome contained 29 scaffolds (minimum length equal to 500 bp), with the largest scaffold of 6,255,750 bp (Table 1). The assembly length was 43.14 Mbp, providing a sequencing depth of 190 fold (number of reads * read length / assembly size). The assembly quality metrics were good, with N50 of 3.86 Mbp and L50 of 5.

Table 1. Summary of the *Fusarium verticillioides* AZB genome assembly results.

Total sequenced bases	8,211,740,588
Number of scaffolds	29
Largest scaffold (bp)	6,255,750
Total length (bp)	43,144,266
GC content (%)	47.94
N50	3,864,563
L50	5
N's per 100 kbp	0.42
Completeness (*Ascomycetes*) *	99.2 %
Completeness (*Sordariomyceta*) *	98.9 %
# protein-coding genes	14,497

* The genome completeness was calculated using the *Ascomycota* and *Sordariomyceta* universal single-copy orthologs with BUSCO v3 (Simão et al., 2015)

We identified 6 scaffolds of *F. verticillioides* AZB flanked by telomeric repeats (start: 5'-TAACCC-3', end: 5'-TTAGGG-3'). Four of these were possibly complete or near-complete since had similar lengths to the chromosomes of the finished genome of *Fusarium verticillioides* 7600 (reference strain, NCBI Assembly: GCA_000149555.1), as well as provided high sequence coverage and identity in sequence alignments to the reference (97.13 % and 99.00 % of minimum coverage and identity, respectively)(Table S2).

To find protein-coding genes in the *Fusarium verticillioides* AZB assembly, we employed a combination of similarity-based and *ab initio* methods. The consensus protein-coding genes set contained 14,497 sequences, corresponding to a gene density of 336 protein-coding genes/Mbp. The average exon length and exons/gene were 391 bp and 2.8, respectively, while the average intron length and introns/gene were 70 bp and 1.8. The overall quality of *F. verticillioides* AZB gene set was good since 99.2 % and 98.9 % of the *Ascomycota* and *Sordariomycetes* universal single-copy orthologs were detected using BUSCO v3.

3.2. Comparative genomics between Fusarium *spp. and* Fusarium verticillioides *strains*

To perform comparative genomics between *Fusarium* spp. we identified orthologous genes between 36 genomes, comprising 19 different species. Five of those genomes belonged to *Fusarium verticillioides*, including the newly generated assembly of *F. verticillioides* AZB.
The data set was composed of 564,044 sequences, with 551,673 assigned to 21,871 orthogroups and 12,731 sequences classified as orphan genes (no homology was found to any sequence in the data set).

The core genome of *Fusarium* spp. analyzed was composed of 5510 genes (had orthologs in all 36 *Fusarium* genomes) and 2320 of these consisted entirely of single-copy genes, which were used to build the phylogenetic tree. The tree topology was in agreement with previous studies

[45,46], showing a major section formed by *F. fujikuroi* species complex (FFSC) subdivided in independent clades (Asian, African, American) (Figure 1).

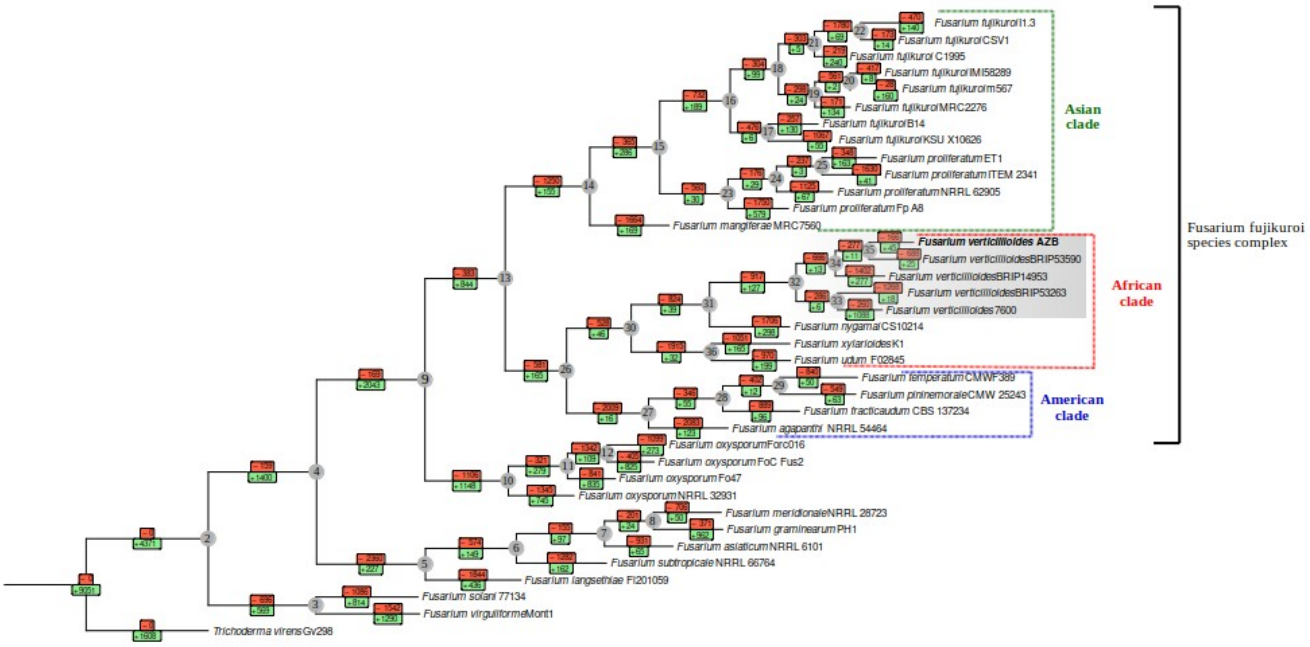


Figure 1. The best-scoring maximum-likelihood tree for *Fusarium verticillioides* AZB and other *Fusarium* species. The tree was created based on 2,320 orthologous genes (nucleotide sequences). The RAxML v.8.2.12 performed 200 searches for the best-scoring tree using the GTR+I+G substitution model, followed by 500 maximum likelihood bootstrap replications. All bootstrap values were 100% and omitted from the phylogenomic tree. Numbers inside green boxes indicate genes gained and numbers inside red boxes indicate genes lost based on the profile of sequences from each orthogroup. The internal nodes were numbered from the root. *Trichoderma virens* Gv29-8 were used as outgroup.

Furthermore, we detected 8598 orthogroups containing all *Fusarium* species from FFSC, which represented approximately 40 % of the total number of orthogroups detected containing 2 or more species (21771 orthogroups). The core genome of FFSC contained 2026 genes with significant blast-hits (e-value: 1e-05, identity >= 20 % and subject coverage >= 30%) to proteins from PHI-database [47], most of them comprising “reduced virulence” genes, i.e., somewhat related to pathogenicity (Table S3). To better understand the biological significance of these orthologous genes, we analyzed previously published RNA-seq data (NCBI GEO: GSE77595) to check if they were expressed during plant infection [25]. We identified 11537 fungal genes up-regulated while infecting maize seedling roots, considering all the *Fusarium* strains from FFSC used in the RNA-

seq analysis (*F. fujikuroi* IMI58289, *F. proliferatum* NRRL 62905, *F. proliferatum* ET1, *F. mangiferae* MRC7560, *F. verticillioides* 7600). These deferentially expressed genes were distributed across 6398 orthogroups (Table S4) and 3603 were part of the FFSC “core genome” identified in this work (composed of 8598 genes). Furthermore, we detected 307 orthogroups of up-regulated genes with representatives from all five *Fusarium* strains, which might take part in plant colonization mechanisms shared by these species. Regarding these orthogroups, the OG000061 contained the highest number of up-regulated genes during plant infection (13 genes) and were related to GDSL lipase acylhydrolase. Other orthogroups that contained a high number of up-regulated genes included mainly MFS transporters (e.g., alpha glucoside transporter, MFS monosaccharide transporter) and carbohydrate-active enzymes (e.g., extracellular endoglucanase, acetylxylan esterase, alpha-galactosidase, pectate lyase d, endo-beta-xylanase b).

The orthology analysis further indicated 22 orthogroups specifically composed of all *F. verticillioides* strains. Four of these (OG0016990, OG0016994, OG0016996, OG0016999) contained genes of *F. verticillioides* 7600 up-regulated during plant infection (Table S4), while three provided significant alignments to PHI-database sequences (OG0016991, OG0016993, OG0016994) and four were composed of putative effectors according to EffectorP v.2.0 [48] (OG0016992, OG0016993, OG0016996, OG0016999). Protein domain searches carried out using InterProScan v.5.44-79.0 [49] and PANNZER2 [50] indicated that these genes encoded mainly hypothetical proteins (Table S5). Besides this, other *F. verticillioides* specific orthologous genes encoded proteins related to gene expression (transcription factors, histone methyltransferases, endoribonuclease L-PSP), protein-protein interaction (ankyrin repeats), transmembrane transporter (amino acid permease, major facilitator superfamily), secondary metabolite biosynthesis and hydrolytic activity. Owing to such putative functions, these genes might also participate in plant infection mechanisms, although this was not detected in gene expression data (GSE77595).

Regarding strain-specific genes, we detected 45 unique genes of *F. verticillioides* AZB. Most of these genes encoded short proteins (40 proteins had less than 200 amino acids, and 20 of them had less than 100 amino acids) with no functional domains detected by InterProScan or significant alignments to sequences from PHI-database. However, 7 of these genes of *F. verticillioides* AZB were predicted as effectors by EffectorP v.2.0. Since this tool is not dependent on sequence similarity, it is better suited to detect highly divergent virulence-related genes.
Furthermore, the number of unique genes of *F. verticillioides* AZB corresponded to 0.3 % of the total genes of this isolate, which was lower than the average found in other *Fusarium* strains analyzed (mean = 2.08, min = 0.1, max = 8.6, sd = 1.88) (Table S6). Considering only *F. verticillioides* strains, the number o unique genes varied considerably, ranging from 0.1 % (*F. verticillioides* BRIP53263) to 5.3 % (*F. verticillioides* 7600). Similar to *F. verticillioides* AZB, the proteins encoded by these strain-specific genes were mainly classified as hypothetical since no significant similarities were found to sequences with known function.

3.3. Whole-genome sequence alignments between F. verticillioides *strains*

In order to identify the genomic regions with the greatest variability among *F. verticillioides* strains, we conducted pairwise alignments between the reference chromosomes of *F. verticillioides* 7600 and contigs of the other isolates (AZB, BRIP53263, BRIP53590, BRIP14953). In general, the reference genome was well covered (low percentage of gaps) with high sequence identity (very few regions with identity less than 85%), indicating good agreement between the draft assemblies and the reference genome (Figure 2).

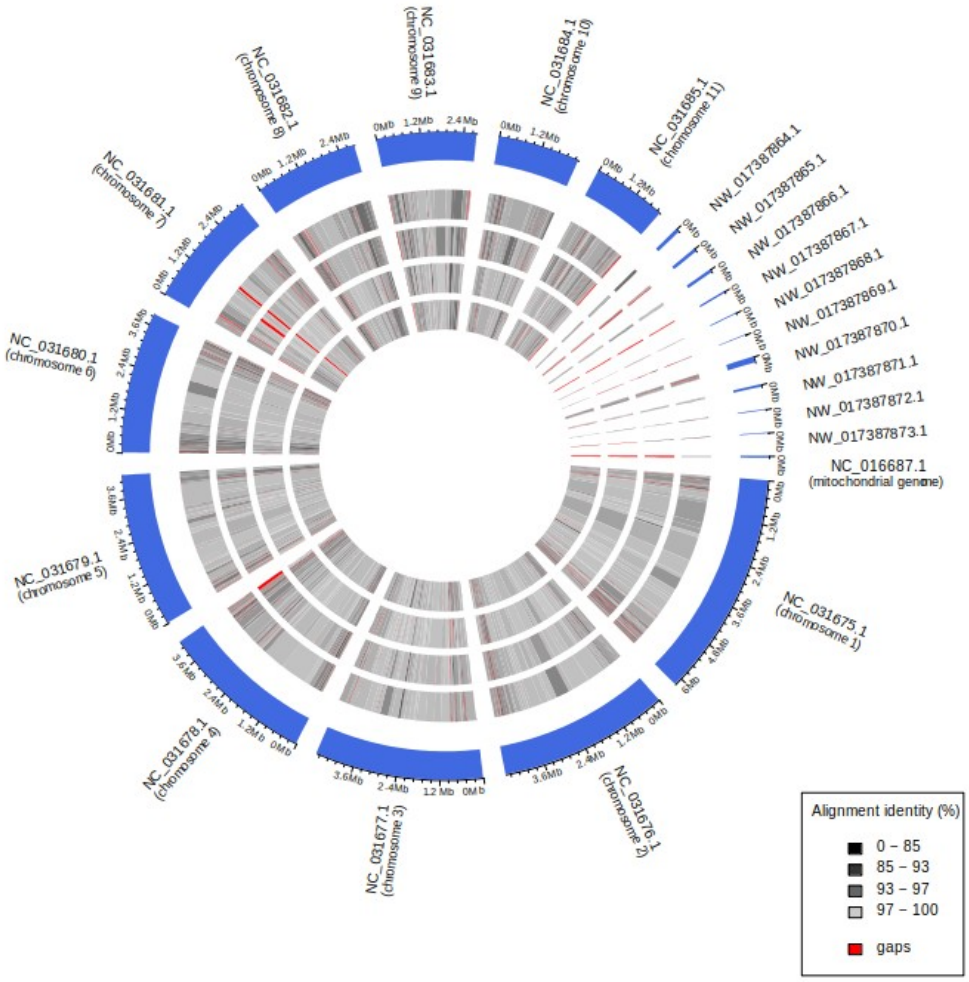


Figure 2. Contig alignments between *Fusarium verticillioides* strains. Each section of the outer circle of the circos plot (blue) represent chromosomes of *F. verticillioides* 7600. The grey circles represent the contigs of other *F. verticillioides* strains (outer to inner: AZB, BRIP14953, BRIP53263, BRIP53590) where each was aligned to the reference (*F. verticillioides* 7600) using the NUCmer algorithm (MUMmer v.3.23). The identities of the sequence alignments were represented on greyscale and gaps were colored in red.

Despite being highly similar, regions of variability were detected between the reference and the other assemblies, indicating sites of rapid evolution in *F. verticillioides*. In general, the most variable regions occurred in the same genomic positions in each strain, being found mainly at the ends of the chromosomes/contigs. However, these regions were somewhat more dispersed on chromosomes 9, 10 and 11 of the reference isolate.

To gain knowledge about genomic features in regions of greater variability among strains, we conducted a deeper analyses in the main sites of variability between *F. verticillioides* 7600 and

F. verticillioides AZB, considering an identity threshold of 85 %. We found 182 genomic regions under the identity threshold, and 49 of those were located in coding sequences (CDS) in both strains (Table S7). Most of these CDS were of hypothetical proteins (not showed significant alignments in InterProScan and PANNZER2 searches), although few were putatively involved in hydrolytic activity (Peptidase S8, Chitinase) and possibly in pathogen-plant interaction process (LysM domain, VIT domain, NB- ARC / Tetratricopeptide, NACHT / Ankyrin repeat), in addition to proteins that can directly affect genome variability (Heterokaryon incompatibility protein) [51,52].

3.4. Phylogenomic analysis and Parsimonious Scenario of Gene Gain and Loss in Fusarium verticillioides *strains*

The profile of gene family gains and losses is interesting to study the dynamics of genome evolution through a phylogenetic model and correlate with fungal lifestyles. The intimate association between *F. verticillioides* and host plant produces evolutionary pressures towards enhanced pathogen fitness, which may lead to the acquisition or loss of protein-coding genes. The ancestor of *F. verticillioides* putatively lost 917 genes while acquired only 127 (Figure 1, node 32) and in both cases, most of the genes encoded hypothetical proteins(Table S8 and S9). Despite this, we detected considerable gene losses regarding transposable elements (30 genes) including HTH CENPB domain, transposases, reverse transcriptases, Pogo transposable element, and others. We also detected considerable losses of transcription factors (25 genes) and transmembrane transporters (21 genes). Furthermore, pairwise sequence alignments against PHI-database indicated that several lost genes (133 genes) were related to pathogenicity such as pisatin demethylase, isotrichodermin C-15 hydroxylase (trichothecene modifying enzyme), short-chain dehydrogenases (e.g. abscisic acid production) and FAD-binding protein (e.g. maackiain detoxification) (Table S10).

On the hand, only 7 genes acquired by the ancestor of *F. verticillioides* provided sufficient similarity to sequences from the PHI-database (Table S11). In general, the acquired genes encoded hypothetical proteins, reflecting the need for deeper characterization of the *F. verticillioides* proteome. Many genes may be related to pathogen-host interaction processes given the pathogenic lifestyle of this fungus.

Furthermore, the profile of gene losses and gains among *F. verticillioides* strains indicated that these events were heterogeneous and relatively frequent, and may be a consequence of changes in the size of the genomes or sequence diversification, which might affect fungal pathogenicity. For example, we found that *F. verticillioides* BRIP14953 and *F. verticillioides* AZB putatively lost genes that were orthologous to FVEG_13321 of *F. verticillioides* 7600 (fungal specific transcription factor). This gene composed subnetwork modules associated with pathogenicity against maize [53]. Also, we detected 161 unique genes of *F. verticillioides* 7600 up-regulated during maize infection (NCBI GEO: GSE77595, Table S4, Table S12) which were inferred as acquired exclusively by this strain (were classified as orphan genes) (Figure 1, Table S13).

3.5. Adaptive Selection

Fungal genes related to niche specialization or host–pathogen interactions are frequently under diversifying selection, playing crucial role for organisms environmental fitness [54,55].

We used codeml algorithm of PAML/4.9h package to calculate the ratio of non-synonymous to synonymous substitutions (dN/dS) and infer the mode of natural selection acting on 7538 single copy orthologous genes of *Fusarium* spp. from African clade. As expected, the M0 model of codeml indicated a high number of genes with $dN/dS < 1$, indicating that most experienced purifying selection (Figure S1). Only 16 genes had average $dN/dS > 1$ considering the whole

sequence in all branches analyzed, and therefore were under strong diversifying selection (Table S14).

Furthermore, we used the branch-site model MA of codeml, since the most likely scenario is that positive selection takes place only in specific sites of genes and in some branches of the phylogeny. We set up the clade of *F. verticillioides* strains as the foreground branch while the *F. nygamai*, *F. xylarioides* and *F. udum* composed the background branch (Figure 1). Under this approach, 1182 orthologous genes undergone positive selection ($-\log_2$(adjusted p-values) > 6.644) among the 7538 single-copy orthologous genes tested (Figure 3), and many of these encoded extracellular proteins, which might be effectors that participate in interacting processes with the host plant. Furthermore, 195 sequences had sufficient similarity to proteins from PHI-database, genes that might be important to evolutionary "arms race" between host plant and *F. verticillioides* (Table S15).

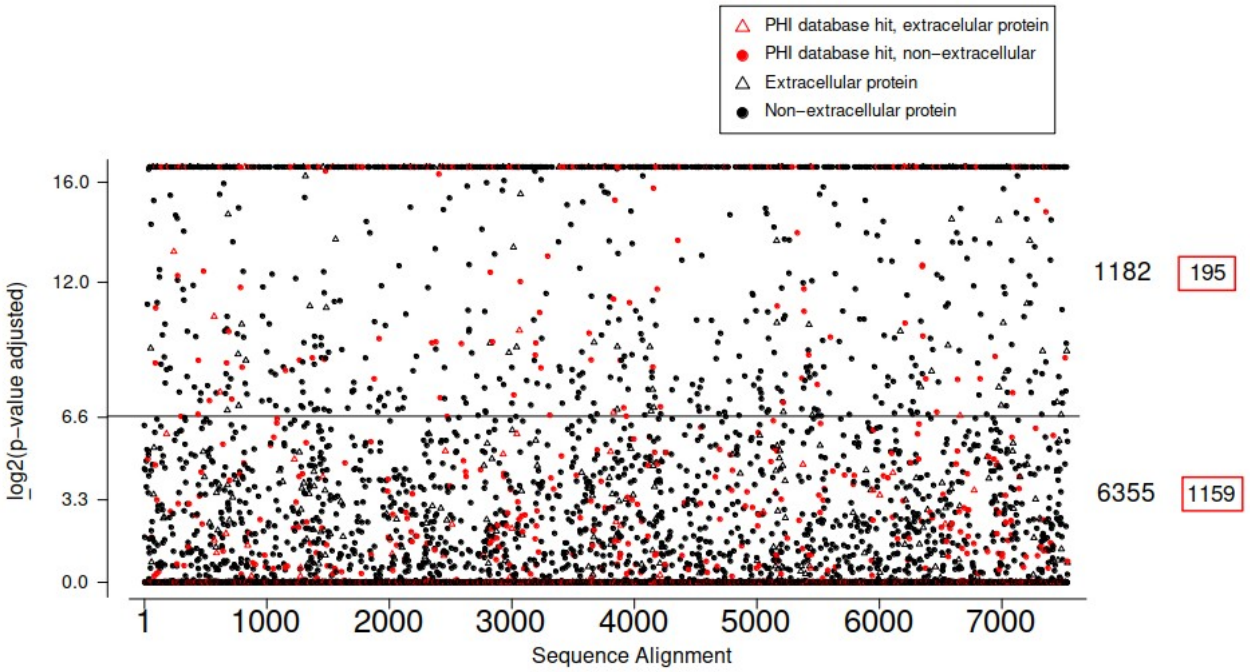


Figure 3. The significance level [-log2 (adjusted p-value)] of the likelihood ratio test (LRT) applied on the likelihoods of model A1 and model A (codeml). The analysis encompassed 7538 single-copy orthologous genes. The clade composed by F. verticillioides strains were set up as the foreground branch and other species from African clade composed the background branch. The null model (H0) constrained codons to have dN / dS <= 1 (Model A1, fix_omega = 1, omega = 1) and the alternative model (H1) allowed codons to have dN / dS> 1 (Model A, fix_omega = 0). The p-values were adjusted using FDR 1% to reduce the false-positives, and those equal to or greater than 6.644 were considered significant (evidence of diversifying selection on the foreground branch). Orthogroups composed of genes

with sufficient similariry to PHI-database sequences were colored in red, otherwise were colored in black. Orthogroups composed of genes coding for extracellular proteins were depicted as unfilled triangules and non-extracellular proteins were depicted as filled circles. the numbers inside the red boxes on the right indicate the number of orthogroups comprising genes that gave significant alignments with PHI-database sequences.

A comprehensive annotation using PANNZER2 revealed that most prevalent functions for the genes under diversifying selection were transcription factor (including many related to mycotoxins production), substrate transporter, s-adenosyl methyltransferase and extracellular glycoside hydrolases. Besides, few genes directly related to mycotoxin production were also under diversifying selection such as nonribosomal peptide synthases and satratoxin biosynthesis sc1 cluster protein 4 (FVEG_16152, FVEG_07488 and FVEG_05749 of *F. verticillioides* 7600 which have orthologues in other *F. verticillioides* strains), suggesting that they have prominent role for *F. verticillioides* pathogenic mechanisms.

In addition, the genes under diversifying selection were evenly distributed across *F. verticillioides* 7600 chromosomes (Table S16), indicating no bias on the content of rapidly evolving genes between *F. verticillioides* 7600 chromosomes.

3.6. Transposable elements in F. verticillioides *AZB genome*

Transposable elements are mobile repetitive DNA sequences that proliferates across the host genome and during this process might create or disrupt protein-coding genes. Thus, the identification of these repetitive elements are of interest due to its impact on genome architecture and evolution [56].

Using a combination of *ab initio* and similarity-based approaches, we detected 2097 retroelements (class I transposable elements), 864 DNA transposons (class II transposable elements) and 1974 unclassified elements in *F. verticillioides* AZB, spanning 1.55 Mbp or 3.60 % of the genome (Table S17). These numbers were higher than the reported in other studies [5], which may

be due to species-specific repetitive sequences identified in this work and used in RepeatMasker searches.

In many fungal plant pathogens, genes related to virulence are often found within or close to transposable elements [57]. For *F. verticillioides* AZB, on the other hand, we found that most of the genes within or close to transposable elements were not related to virulence (Figure 4).

Furthermore, it is worth to mention that most TE sequences of *F. verticillioides* AZB not showed Pfam or CDD domains related to transposable elements [58] and thus, were fragmented or highly mutated sequences. Therefore, most of the TEs of *F. verticillioides* AZB was probably inactive, which may be a consequence of strong mechanisms to suppress its proliferation such as repeat-induced point mutations (RIP).

Figure 4. Heatmap showing the number of protein-coding genes with intersection (0 bp) or close (500 bp to 5000 bp) to transposable elements in *F. verticillioides* AZB genome. The total number of genes that fall into those features were depicted in the first column and the other 3 columns indicated the number of genes putatively related to virulence. We used 3 different approaches to identify these genes: BLASTp against PHI database, predictions with EffectorP and

orthology to F. verticillioides 7600 genes up-regulated during plant infection. Each gene could be counted more than once if it was identified by more than one approach.

3.7. Evidences of repeat-induced point mutations (*RIP) in* F. verticillioides *AZB genome*

The RIP mechanism induces transitions between pyrimidines or purines (C ↔ T or G ↔ A) in repeated sequences, generally with 400 bp or more, and operates in the haploid nuclei of fungi during meiosis in the preparation for sexual reproduction [59].

The RIP process identified in *N. crassa* is dependent on two genes that encode the 5-cytosine methyltransferases RID (RIP deficient) and DIM-2 (defective in methylation), as well as the cofactors HP-1 (heterochromatic protein), and DIM-3, -5, -7, -8 and -9. This set of fungal RIP effectors was previously identified in *F. verticillioides* 7600 genome [60] and we also detected they in *F. verticillioides* AZB and other *F. verticillioides* strains (BRIP53263, BRIP53590, BRIP14953). As shown by pairwise global sequence alignments, the DIM-9 and RID sequences of *F. verticillioides* were the most divergent to the *N. crassa* OR74A sequences, while the DIM-3 provided the highest identity and coverage (Table S18).

Besides harboring putative RIP effectors, *F. verticillioides* AZB genome had a bimodal %GC distribution, showing a major peak at 49.0 % and a minor peak at 18.9 % (Figura 5). This revealed the existence of AT-rich regions, a hallmark of RIP-affect genome.

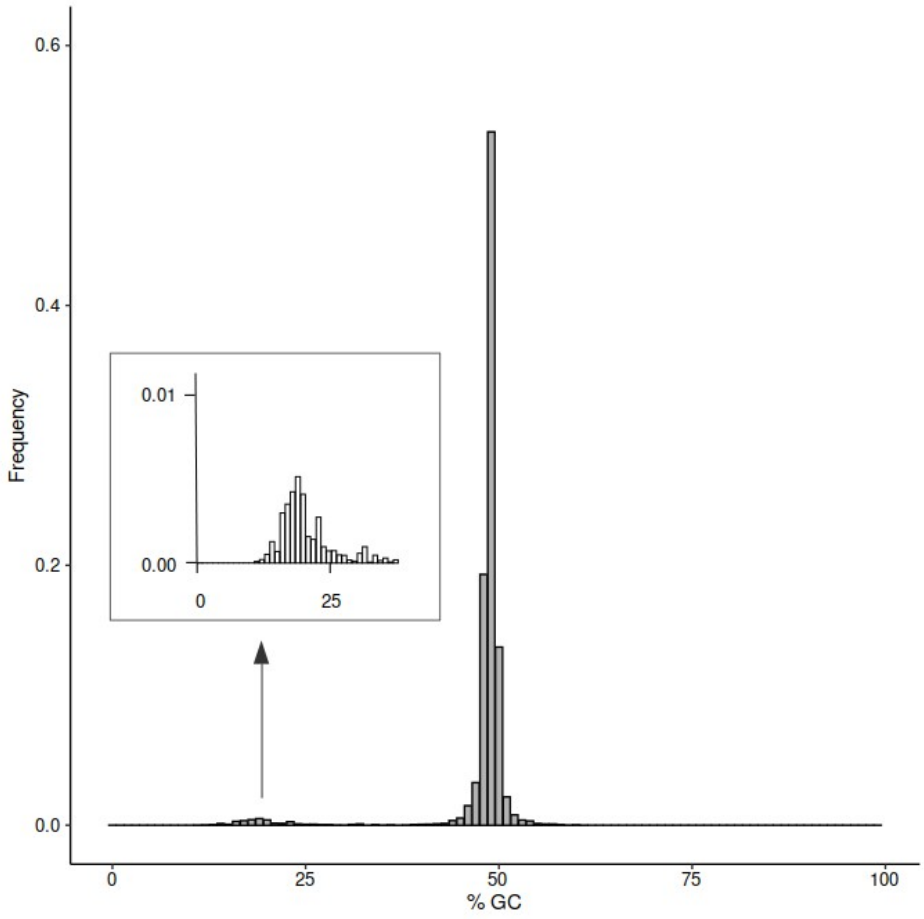


Figure 5. The frequency distribution of the GC content of *F. verticillioides* AZB. The x-axis indicates percentage GC content and the y-axis illustrates the frequencies of each %GC across genome. The secondary peak of GC content (18.9 %) was enlarged for better visualization.

Furthermore, the analysis of di-nucleotide frequencies in repetitive regions greater than 400 bp revealed that simple repeats and many transposable elements families (LTR/Gypsy, DNA/CMC-EnSpm, DNA/hAT, DNA/hAT-Blackjack and Unknown) presented a depletion of CpA di-nucleotides and an increased frequency of TpA (Table S19), suggesting that this type of mutation was predominant in the *F. verticillioides* RIP mechanism.

Albeit RIP mutates transposable elements efficiently, it may also operate in protein-coding genes causing few highly similar gene sequences in the genome [61]. Such RIP side-effect was also detected for *F. verticillioides* AZB, where most of its protein sequences provided less than 45 % identity to each other (Figure S2).

4. Discussion

The genomes of fungal plant pathogens are under constant evolutionary pressure due to the intimate association between pathogen and host. Identifying genomic traits related to adaptive diversification of fungal species or strains could enhance our knowledge of the genetic basis underlying pathogenicity, which may be useful to develop better control strategies. In this study, we reported the whole-genome sequence of *F. verticillioides* AZB, the first sequenced strain obtained in Brazil. Using the generated assembly and other *Fusarium* genomes, we focused our work on genome evolutionary events and features, such as gene orthology, gene gains and losses, transposable elements content, hallmarks of RIP and diversifying selection on protein-coding genes. To study the interplay between genome evolution and virulence-related genes in *F. verticillioides*, we used 36 genome sequences from diverse *Fusarium* species, especially from *F. fujikuroi* species complex (FFSC). The selected genomes harbor high genetic diversity, and the results therefore provided insights into adaptive diversification of species analyzed, which is linked to enhanced fitness to particular hosts and pathogenic mechanisms. Furthermore, we used previously published RNA-seq data (NCBI GEO: GSE77595) in an attempt to deepen our knowledge about the biological significance of the evolutionary features inferred.

The integration of genomic and transcriptomic data revealed that most of the FFSC genomes were composed of orthologous genes shared by all species (8598 orthogroups, the “core genome”) and many of these genes were up-regulated during maize infection (7071 genes distributed across 3603 orthogroups of the core genome). Thus, although have a broad host range, *Fusarium* spp. from the FFSC share many genes related to pathogenicity, which may be interesting to study aiming at universal control strategies. For example, the OG0000061 contained sequences from almost all *Fusarium* species used in this study (included all species from FFSC) as well as genes up-regulated by different *Fusarium* species during maize infection (Table S4). This orthogroup was composed

mostly of extracellular GDSL acyl-hydrolases, which may be involved in the fatty acid release from various lipids, such as membrane phospholipids and cutin [62] and thus, might participate in fungus-host interaction. Furthermore, BLASTp searches indicated similar GDSL acyl-hydrolases sequences in other plant pathogens such as *Colletotrichum* spp. and *Verticillium* spp., and there is also evidence of up-regulation of these genes during plant infection [63].

The orthology analyses further indicated many orthogroups composed only by genes from *F. verticillioides* strains, signals of adaptive divergence occurred after speciation. These genes encoded mainly hypothetical proteins, but expression data (GSE77595) indicated that some were up-regulated by *F. verticillioides* 7600 during plant infection and thus, were possibly related to species-specific pathogenic mechanisms. The absence of these orthologous genes in closely related species (*F. xylarioides* and *F. udum*) suggested that them might contribute to unique traits of *F. verticillioides* and help to better understand the species biology and specificities.

Furthermore, a marked variability in the number of strain-specific genes (orphan genes) between *F. verticillioides* genomes were detected, and most of these genes encoded small proteins (less than 200 amino acids). Although these data should be viewed with caution due to possible errors of assembly and gene prediction in draft genomes, it suggested that all strains had genomic traits that could provide enhanced fitness under specific ecological scenarios, which may include mycotoxin production, host genotype, fungicide resistance or climate conditions. Owing to its sequence divergence, we were not able to assign function for most of the strain-specific genes using sequence similarity approaches such as BLASTp and InterProScan. However, a machine learning-based tool (EffectorP) indicated strain-specific genes of *F. verticillioides* AZB as candidates of virulence effectors, suggesting the importance of highly specific genes of this species for pathogenesis.

Since strain-specific genes are product of fast-evolving genomic regions we analyzed if these genes of *F. verticillioides* AZB had intersection or were close to transposable elements in the

genome, but such evolutionary pattern was not prevalent. Furthermore, we detected very few highly divergent regions (sequence identity < 85 %) in genome alignments between *F. verticillioides* strains, and these divergent regions were located especially on chromosome ends. These genomic features associated with the high potential for sexual reproduction of *F. verticillioides* [64,65], suggested that sexual recombination is the major event that generates genetic variability in this species, which directly impacts the evolution of virulence-related genes. In many plant pathogenic fungi, on the other hand, the virulence factors are typically located near and within transposable elements or in specific chromosomes [66], and these genomic features are linked to genetic diversification in fungi which mainly shown asexual reproduction such as *Verticillium* spp., *Colletotrichum* spp. and *F. oxysporum* [9, 67-69].

Since rapidly-evolving genes of plant pathogenic fungi are frequently associated to virulence [54,55], we calculated the dN/dS ratios for each orthogroup comprising single-copy orthologous genes of *Fusarium* spp. from African clade to check the variations in the selection pressure among these genes. A simpler model (model M0 of codeml) indicated only 0.2% of the total genes analyzed were under adaptive selection. This approach provided an overview of orthologous genes that underwent strong adaptive selection since they presented average dN/dS > 1 considering the whole sequence and all *Fusarium* species from the African clade. These genes encode mostly hypothetical proteins (table S14) and are interesting targets for further studies to assess their function or impact on the lifestyle of *Fusarium* species investigated. Also, we used the more realistic branch-site model to estimate dN/dS ratios among single-copy orthologous genes (model A *vs* A1 of codeml). Using this model, we detected 1182 genes of *F. verticillioides* under positive selection and many encoded putative proteins related to pathogenicity (195 genes with similarity to PHI-database sequences, including 8 related to mycotoxin production, and many extracellular glycosidases and peptidases)(Table S15). Therefore, this approach was useful to identify diversifying selection in protein-coding genes of *F. verticillioides* using sequences of closely related

species as genetic background. These genes are likely to be involved in host–pathogen interactions and are targets for future experimental verification that might contribute to develop or improve strategies to control *F. verticillioides* as well as diminish mycotoxin contamination.

Another important genome evolutionary event consists of gene gains and losses, and analyzing these genomic changes could provide important insights into fungal adaptations due to associations with the host. The profiles of gene gains and losses through *Fusarium* evolutionary model revealed striking variations between species and even strains, showing that these events are frequent and thus, important for fungal adaptive fitness. The speciation that originated *F. verticillioides* (Figure 1, internal node 32) was characterized by the considerable loss of genes related to pathogenicity (195 genes), such as pisatin demethylase and maackiain detoxification protein, which can be associated with host specialization. Besides this, the major group of genes lost by *F. verticillioides* was related to transposable elements, which could directly impacts the evolution of their genomes. Although *F. verticillioides* AZB genome harbored a relatively large amount of transposable elements (3.6 % of the genome), these repetitive elements were fragmented or highly mutated and thus, were probably inactive. Previous studies, as well as ours, indicated hallmarks of RIP in *F. verticillioides*, and since this species has a strong tendency to perform sexual reproduction, RIP may have operated on its genome profoundly and earlier in evolutionary history. This suggests that repetitive regions, especially transposable elements, have a little impact in the evolution of *F. verticillioides* towards better fit to pathogenicity.

A detailed inspection of the profile of gene gains and losses in each *F. verticillioides* strain revealed that genes related to pathogenicity were lost by some strains but not others, such as the orthologues of FVEG_13321 lost by *F. verticillioides* BRIP14953 and *F. verticillioides* AZB. Besides this, using gene expression data available (GSE77595) we detected that genes gained only by *F. verticillioides* 7600 were up-regulated during plant infection, indicating that these evolutionary events might contribute to variations in the pathogenicity level or mechanisms

between *F. verticillioides* strains. Since there might be considerable variation in the degree of virulence between *F. verticillioides* strains [70], the profile of gene gains and losses could provide evolutionary clues relating the genetic background and phenotype.

Finally, most of the genes acquired by the ancestor of *F. verticillioides* strains, as well as by each strain, encoded proteins with no known function (hypothetical proteins). This reflects the long way through functional genomics to better characterize this species and its relationship with plants. Based on the results of orthology, adaptive selection and gene gains and losses, we provided knowledge about genomic evolutionary events and features of *F. verticillioides*, as well as a set of genes that probably participate in plant pathogenic processes. Thus, the data generated here might be a starting point for future functional studies to unveil the biological importance of genes putatively involved in pathogen-host associations.

5. Conclusions

In this study, we sequenced and analyzed the genome of *F. verticillioides* AZB as well as performed comparative genomics and evolutionary analysis, especially regarding virulence-related genes. Gene expression data associated with genome-scale comparative analysis evidenced genomic features of *F. verticillioides* and other species from FFSC that may be related to fungal fitness, which could serve as a substrate for functional studies to further understand pathogenicity mechanisms and mycotoxin production.

The *F. verticillioides* AZB genome harbored a high number of fragmented or highly mutated transposable element-coding genes, AT-rich blocks and other hallmarks of RIP, indicating that mobile DNA exerts little effect on *F. verticillioides* genome evolution. Since this species had a tendency to reproduce sexually, the meiotic recombination is probably the major source of genetic variability. We found a high number of protein-coding genes under positive selection and many

strain-specific genes, both considered as rapidly-evolving genomic features. This may guide futures studies to increase the knowledge of the evolutionary dynamics of the pathogen and elaborate better control strategies.

References

[1] Ma, L.J., Geiser, D.M., Proctor, R.H., Rooney, A.P., O'Donnell, K., Trail, F., Gardiner, D.M., Manners, J.M., Kazan, K., 2013. Fusarium pathogenomics. Annu. Rev. Microbiol. 67, 399–416. https://doi.org/10.1146/annurev-micro-092412-15565

[2] Krout-Greenberg, N.D., Puschner, B., Davidson, M.G., DePeters, E.J., 2013. Preliminary study to assess mycotoxin concentrations in whole corn in the California feed supply. J. Dairy Sci. 96, 2705–2712. https://doi.org/10.3168/jds.2012-5957

[3] Palumbo, R., Crisci, A., Venâncio, A., Abrahantes, J.C., Dorne, J. Lou, Battilani, P., Toscano, P., 2020. Occurrence and co-occurrence of mycotoxins in cereal-based feed and food. Microorganisms 8. https://doi.org/10.3390/microorganisms8010074

[4] Queiroz, V.A.V., De Oliveira Alves, G.L., Da Conceição, R.R.P., Guimarães, L.J.M., Mendes, S.M., De Aquino Ribeiro, P.E., Da Costa, R.V., 2012. Occurrence of fumonisins and zearalenone in maize stored in family farm in Minas Gerais, Brazil. Food Control 28, 83–86. https://doi.org/10.1016/j.foodcont.2012.04.039

[5] Blacutt, A.A., Gold, S.E., Voss, K.A., Gao, M., Glenn, A.E., 2018. Fusarium verticillioides: Advancements in understanding the toxicity, virulence, and niche adaptations of a model mycotoxigenic pathogen of maize. Phytopathology 108, 312–326. https://doi.org/10.1094/PHYTO-06-17-0203-RVW

[6] Zila, C.T., Ogut, F., Romay, M.C., Gardner, C.A., Buckler, E.S., Holland, J.B., 2014. Genome-wide association study of Fusarium ear rot disease in the U.S.A. maize inbred line collection. BMC Plant Biol. 14, 1–15. https://doi.org/10.1186/s12870-014-0372-6

[7] Zhang, Y., Ma, L.J., 2017. Deciphering Pathogenicity of Fusarium oxysporum From a Phylogenomics Perspective, 1st ed, Advances in Genetics. Elsevier Inc. https://doi.org/10.1016/bs.adgen.2017.09.010

[8] Croll, D., Lendenmann, M.H., Stewart, E., McDonald, B.A., 2015. The impact of recombination hotspots on genome evolution of a fungal plant pathogen. Genetics 201, 1213–1228. https://doi.org/10.1534/genetics.115.180968

[9] Ma, L.J., Van Der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P.M., Kang, S., Shim, W.B., Woloshuk, C., Xie, X., Xu, J.R., Antoniw, J., Baker, S.E., Bluhm, B.H., Breakspear, A., Brown, D.W., Butchko, R.A.E., Chapman, S., Coulson, R., Coutinho, P.M., Danchin, E.G.J.,

Diener, A., Gale, L.R., Gardiner, D.M., Goff, S., Hammond-Kosack, K.E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C.D., Koehrsen, M., Kumar, L., Lee, Y.H., Li, L., Manners, J.M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S.Y., Proctor, R.H., Regev, A., Ruiz-Roldan, M.C., Sain, D., Sakthikumar, S., Sykes, S., Schwartz, D.C., Turgeon, B.G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C.A., Kistler, H.C., Rep, M., 2010. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature 464, 367–373. https://doi.org/10.1038/nature08850

[10] Doyle, J.J., Doyle, J.L., 1990. Isolation of plant DNA from fresh tissue. Focus (Madison). 12, 13–15.

[11] Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

[12] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

[13] Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

[14] Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E.M., 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

[15] Stanke, M., Waack, S., 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19, 215–225. https://doi.org/10.1093/bioinformatics/btg1080

[16] Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O., Borodovsky, M., 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 18, 1979–1990. https://doi.org/10.1101/gr.081612.108

[17] Slater, G.S.C., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6, 1–11. https://doi.org/10.1186/1471-2105-6-31

[18] Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Robin, C.R., Wortman, J.R., 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9, 1–22. https://doi.org/10.1186/gb-2008-9-1-r7

[19] Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 1–14. https://doi.org/10.1186/s13059-015-0721-2

[20] Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 30:772–780. doi: 10.1093/molbev/mst010

[21] Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 -- Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE, 5(3):e9490. doi:10.1371/journal.pone.0009490

[22] Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

[23] Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.Y., 2017. GGTREE: an R Package for Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other Associated Data. Methods Ecol. Evol. 8, 28–36. https://doi.org/10.1111/2041-210X.12628

[24] Felsenstein, J., 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

[25] Niehaus, E.M., Münsterkötter, M., Proctor, R.H., Brown, D.W., Sharon, A., Idan, Y., Oren-Young, L., Sieber, C.M., Novák, O., Pěnčík, A., Tarkowská, D., Hromadová, K., Freeman, S., Maymon, M., Elazar, M., Youssef, S.A., El-Shabrawy, E.S.M., Shalaby, A.B.A., Houterman, P., Brock, N.L., Burkhardt, I., Tsavkelova, E.A., Dickschat, J.S., Galuszka, P., Güldener, U., Tudzynski, B., 2016. Comparative “omics” of the fusarium fujikuroi species complex highlights differences in genetic potential and metabolite synthesis. Genome Biol. Evol. 8, 3574–3599. https://doi.org/10.1093/gbe/evw259

[26] Newman, A.M., Cooper, J.B., 2007. XSTREAM: A practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. BMC Bioinformatics 8, 1–19. https://doi.org/10.1186/1471-2105-8-382

[27] Jorda, J., Kajava, A. V., 2009. T-REKS: Identification of Tandem REpeats in sequences with a K-meanS based algorithm. Bioinformatics 25, 2632–2638. https://doi.org/10.1093/bioinformatics/btp482

[28] Mayer, C., 2010. Phobos - a tandem repeat search tool for complete genomes.

[29] Frith, M.C., 2011. A new repeat-masking method enables specific detection of homologous sequences. Nucleic Acids Res. 39. https://doi.org/10.1093/nar/gkq1212

[30] Benson, G., 1999. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580. https://doi.org/10.1093/nar/27.2.573

[31] Rice, P., Longden, L., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277. https://doi.org/10.1016/S0168-9525(00)02024-2

[32] Kolpakov, R., Bana, G., Kucherov, G., 2003. mreps: Efficient and flexible detection of tandem repeats in DNA. Nucleic Acids Res. 31, 3672–3678. https://doi.org/10.1093/nar/gkg617

[33] Quinlan, A.R., Hall, I.M., 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

[34] Warburton, P.E., Giordano, J., Cheung, F., Gelfand, Y., Benson, G., 2004. Inverted repeat structure of the human genome: The X-chromosome contains a preponderance of large, highly homologous inverted repeated that contain testes genes. Genome Res. 14, 1861–1869. https://doi.org/10.1101/gr.2542904

[35] Girgis, H.Z., 2015. Red: An intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. BMC Bioinformatics 16, 1–19. https://doi.org/10.1186/s12859-015-0654-5

[36] Smit, A., Hubley, R., 2015. RepeatModeler Open-1.0 http://www.repeatmasker.org.

[37] Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

[38] Enright, A.J., Van Dongen, S., Ouzounis, C.A., 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30, 1575–1584. https://doi.org/10.1093/nar/30.7.1575

[39] Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

[40] Eddy, S.R., 2011. Accelerated profile HMM searches. PLoS Comput. Biol. 7. https://doi.org/10.1371/journal.pcbi.1002195

[41] Smit, A., Hubley, R., Green, P., 2015. RepeatMasker Open-4.0. http://www.repeatmasker.org.

[42] Hane, J.K., Oliver, R.P., 2008. RIPCAL: A tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. BMC Bioinformatics 9, 1–12. https://doi.org/10.1186/1471-2105-9-478

[43] Testa, A.C., Oliver, R.P., Hane, J.K., 2016. OcculterCut: A comprehensive survey of at-rich regions in fungal genomes. Genome Biol. Evol. 8, 2044–2064. https://doi.org/10.1093/gbe/evw121

[44] Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591. https://doi.org/10.1093/molbev/msm088

[45] Herron, D.A., Wingfield, M.J., Wingfield, B.D., Rodas, C.A., Marincowitz, S., Steenkamp, E.T., 2015. Novel taxa in the Fusarium fujikuroi species complex from Pinus spp. Stud. Mycol. 80, 131–150. https://doi.org/10.1016/j.simyco.2014.12.001

[46] Villani, A., Proctor, R.H., Kim, H.S., Brown, D.W., Logrieco, A.F., Amatulli, M.T., Moretti, A., Susca, A., 2019. Variation in secondary metabolite production potential in the Fusarium incarnatum-equiseti species complex revealed by comparative analysis of 13 genomes. BMC Genomics 20, 1–22. https://doi.org/10.1186/s12864-019-5567-7

[47] Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S.Y., De Silva, N., Martinez, M.C., Pedro, H., Yates, A.D., Hassani-Pak, K., Hammond-Kosack, K.E., 2020. PHI-base: The pathogen-host interactions database. Nucleic Acids Res. 48, D613–D620. https://doi.org/10.1093/nar/gkz904

[48] Sperschneider J., Dodds P.N., Gardiner D.M., Singh K.B., Taylor J.M., 2018. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Molecular Plant Pathology Sep;19(9):2094-2110. doi: 10.1111/mpp.12682

[49] Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics (Oxford, England), 30(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

[50] Toronen P., Medlar A., Holm L., 2018. PANNZER2: A rapid functional annotation webserver. Nucl. Acids Res.46, W84-W88 doi: 10.1093/nar/gky350

[51] Daskalov, A., Heller, J., Herzog, S., Fleiβner, A., Glass, N.L., 2017. Molecular Mechanisms Regulating Cell Fusion and Heterokaryon Formation in Filamentous Fungi. The Fungal Kingdom 215–229. https://doi.org/10.1128/9781555819583.ch10

[52] Schoustra, S.E., Debets, A.J.M., Slakhorst, M., Hoekstra, R.F., 2007. Mitotic recombination accelerates adaptation in the fungus Aspergillus nidulans. PLoS Genet. 3, 0648–0653. https://doi.org/10.1371/journal.pgen.0030068

[53] Kim, M.S., Zhang, H., Yan, H., Yoon, B.J., Shim, W.B., 2018. Characterizing co-expression networks underpinning maize stalk rot virulence in Fusarium verticillioides through computational subnetwork module analyses. Sci. Rep. 8, 1–13. https://doi.org/10.1038/s41598-018-26505-2

[54] Grandaubert, J., Dutheil, J.Y., Stukenbrock, E.H., 2019. The genomic determinants of adaptive evolution in a fungal pathogen. Evol. Lett. 3, 299–312. https://doi.org/10.1002/evl3.117

[55] Sperschneider, J., Gardiner, D.M., Thatcher, L.F., Lyons, R., Singh, K.B., Manners, J.M., Taylor, J.M., 2015. Genome-wide analysis in three Fusarium pathogens identifies rapidly evolving chromosomes and genes associated with pathogenicity. Genome Biol. Evol. 7, 1613–1627. https://doi.org/10.1093/gbe/evv092

[56] Werren, J.H., 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. Proc. Natl. Acad. Sci. U. S. A. 108, 10863–10870. https://doi.org/10.1073/pnas.1102343108

[57] Dong, S., Raffaele, S., Kamoun, S., 2015. The two-speed genomes of filamentous pathogens: Waltz with plants. Curr. Opin. Genet. Dev. 35, 57–65. https://doi.org/10.1016/j.gde.2015.09.001

[58] Muszewska, A., Steczkiewicz, K., Stepniewska-Dziubinska, M., Ginalski, K., 2019. Transposable elements contribute to fungal genes and impact fungal lifestyle. Sci. Rep. 9, 1–10. https://doi.org/10.1038/s41598-019-40965-0

[59] Gladyshev, E., 2017. Repeat-Induced Point Mutation and Other Genome Defense Mechanisms in Fungi. The Fungal Kingdom 5, 687–699. https://doi.org/10.1128/9781555819583.ch33

[60] van Wyk, S., Wingfield, B.D., De Vos, L., van der Merwe, N.A., Santana, Q.C., Steenkamp, E.T., 2019. Repeat-induced point mutations drive divergence between Fusarium circinatum and its close relatives. Pathogens 8. https://doi.org/10.3390/pathogens8040298

[61] Galagan, J.E., Selker, E.U., 2004. RIP: The evolutionary cost of genome defense. Trends Genet. 20, 417–423. https://doi.org/10.1016/j.tig.2004.07.007

[62] Grienenberger, E., Geoffroy, P., Mutterer, J., Legrand, M., Heitz, T., 2010. The interplay of lipid acyl hydrolases in inducible plant defense. Plant Signal. Behav. 5, 1181–1186. https://doi.org/10.4161/psb.5.10.12800

[63] Bhadauria, V., Vijayan, P., Wei, Y., Banniza, S., 2017. Transcriptome analysis reveals a complex interplay between resistance and effector genes during the compatible lentil-Colletotrichum lentis interaction. Sci. Rep. 7, 1–13. https://doi.org/10.1038/srep42338

[64] Danielsen, S., Meyer, U.M., Funck Jensen, D., 1998. Genetic characteristics of Fusarium verticillioides isolates from maize in Costa Rica. Plant Pathol. 47, 615–622. https://doi.org/10.1046/j.1365-3059.1998.00277.x

[65] Gomes, A.A.M., de Melo, M.P., Tessmann, D.J., Lima, C.S., 2020. Sexual reproduction parameters in Fusarium verticillioides populations from maize in Brazil. Eur. J. Plant Pathol. 156, 317–323. https://doi.org/10.1007/s10658-019-01881-1

[66] Möller, M., Stukenbrock, E.H., 2017. Evolution and genome architecture in fungal plant pathogens. Nat. Rev. Microbiol. 15, 756–771. https://doi.org/10.1038/nrmicro.2017.76

[67] Amyotte, S.G., Tan, X., Pennerman, K., del Mar Jimenez-Gasco, M., Klosterman, S.J., Ma, L.J., Dobinson, K.F., Veronese, P., 2012. Transposable elements in phytopathogenic Verticillium spp.: insights into genome evolution and inter- and intra-specific diversification. BMC Genomics 13. https://doi.org/10.1186/1471-2164-13-314

[68] Gordon, T., Martyn, R., 1997. The evolutionary biology of Fusarium oxysporum. Annu. Rev. Phytopathol. 35, 111–128. https://doi.org/10.2307/2419438

[69] Tsushima, A., Gan, P., Kumakura, N., Narusaka, M., Takano, Y., Narusaka, Y., Shirasu, K., 2019. Genomic plasticity mediated by transposable elements in the plant pathogenic fungus Colletotrichum higginsianum. Genome Biol. Evol. 11, 1487–1500. https://doi.org/10.1093/gbe/evz087

[70] Olowe, O.M., Sobowale, A.A., Olawuyi, O.J., Odebode, A.C., 2018. Variation in pathogenicity of Fusarium verticillioides and resistance of maize genotypes to Fusarium ear rot. Arch. Phytopathol. Plant Prot. 51, 939–950. https://doi.org/10.1080/03235408.2018.1535812

Anexo I

Ensaio em casa de vegetação para avaliar a capacidade de *Trichoderma orchidacearum* COAD 3006 para promover crescimento vegetal e biocontrole de nematoide das galhas (*Meloidogyne javanica*)

Introdução

Os nematoides parasitas de plantas são responsáveis por limitar a produção agrícola das principais culturas agrícolas em todo o mundo, causando enormes perdas econômicas e ameaçando a segurança alimentar global. O manejo desses parasitas depende em grande parte do uso de nematicidas químicos, que geralmente são prejudiciais ao meio ambiente e à saúde humana. Nesse sentido, a biotecnologia pode fornecer meios eficazes para o manejo desses patógenos de forma ecologicamente correta, como por exemplo, com o uso de fungos filamentosos como agentes de biocontrole.

Um vez que *Trichoderma orchidacearum* se trata de uma espécie ainda não descrita e pertencente a uma seção do gênero *Trichoderma* pouco explorada, é interessante avaliar o potencial que possui como agente de biocontrole de fitonematoides. O estilo de vida micoparasita é considerado comum a todas as espécies do genêro *Trichoderma* (Druzhinina et al., 2011), porém com variações no nível de agressividade. Além disso, existem espécies de *Trichoderma* que também promovem crescimento vegetal, sendo atrativas para desenvolvimento de práticas agrícolas sustentáveis (Stewart & Hill, 2014)

Os experimentos descritos abaixo foram realizados no Laboratório de Biocontrole de Fitonematoides do Departamento de Fitopatologia da Universidade Federal de Viçosa. Os experimentos foram conduzidos sob coordenação do prof. Leandro Grassi, que possuiu ampla experiência no assunto e foi o principal responsável pelo desenvolvimento do produto RizoTec® a base de fungo de biocontrole *Pochonia chlamydosporia* (https://www.stoller.com.br/solucoes/rizotec/).

Metodologia

O fungo *Trichoderma orchidacearum* COAD 3006 foi cultivado em meio sólido constituído de 150 g de arroz e umedecido com 40 mL de água. O material foi autoclavado por 20 minutos a 120 ºC e após esse processo foi acrescentado discos de BDA contendo o fungo. Seu crescimento foi mantido por 21 dias a 27 º C para obtenção dos conídios. Após esse intervalo de tempo, essas estruturas foram recuperadas com extração aquosa e filtradas com auxílio de gazes.

O inóculo de *M. javanica*, obtido de população pura, foi multiplicado em tomateiros do grupo Santa Clara, em casa de vegetação. Os ovos foram extraídos das raízes utilizando-se a técnica descrita por Boneti e Ferraz (1981) e, em câmara de Peters, foi feita a calibração da suspensão de ovos, com auxílio do microscópio óptico.

Mudas de tomate Santa Clara com 21 dias foram transplantadas em recipientes de 2 L e em seguida o solo esterilizado foi infestado com *M. javanica* (3000 ovos por vaso) e recebeu os devidos tratamentos: água esterilizada (testemunha), *T. orchidacearum* nas concentrações de 2500; 5000; 7500; 10000 conídios por grama de solo, com 8 repetições por tratamento. Para testes de promoção de crescimento vegetal, foi apenas omitida a adição de ovos de *M. javanica*. As plantas permaneceram em casa de vegetação a 28ºC. Após 40 dias, foi avaliado o número de ovos por raiz (no experimento que continha *M. javanica*), altura e massa fresca da parte aérea e massa da raiz. ANOVA foi conduzida no pacote R (versão 3.5.3), com teste F de variâncias a 5,0 % de probabilidade.

Resultados

A partir das métricas avaliadas concluímos que *T. orchidacearum* COAD 3006 não apresenta capacidade de biocontrole de *Meloidogyne javanica*. O número de ovos por raiz não apresentou variação significativa entre os diferentes tratamentos (Figura 1), assim como o peso e altura da parte aérea (Figuras 2 e 3) e peso da raiz (Figura 4). De forma similar, os experimentos indicaram que o fungo não tem capacidade de promover crescimento vegetal de tomateiros, uma vez que peso e altura da parte aérea (Figuras 5 e 6) e peso da raiz (Figura 7) permanecem constantes, independente da quantidade de fungo inoculada no solo.

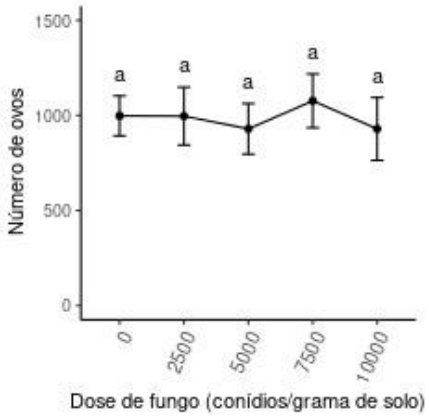


Figura 1: Número médio de ovos por raiz. Todos os tratamentos receberam 3000 ovos de *M. javanica* por vaso. As barras de erros representam o desvio-padrão amostral. Todas médias foram iguais, de acordo com teste F a 5,0 % de significância.

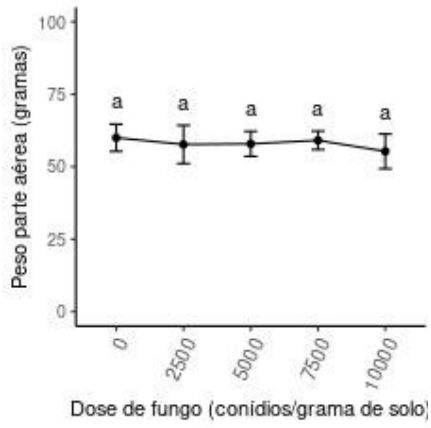


Figura 2: Peso em gramas da parte aérea (caule, folhas). Todos os tratamentos receberam 3000 ovos de *M. javanica* por vaso. As barras de erros representam o desvio-padrão amostral. Todas médias foram iguais, de acordo com teste F a 5,0 % de significância.

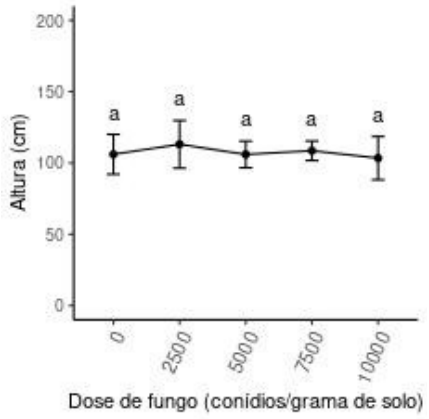


Figura 3: Altura da parte aérea (base do caule até o topo da planta). Todos os tratamentos receberam 3000 ovos de *M. javanica* por vaso. As barras de erros representam o desvio-padrão amostral. Todas médias foram iguais, de acordo com teste F a 5,0 % de significância.

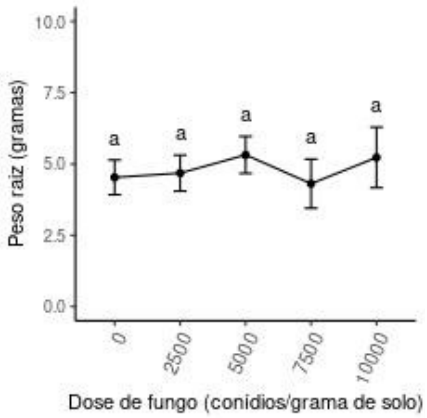


Figura 4: Peso da raiz em gramas. Todos os tratamentos receberam 3000 ovos de *M. javanica* por vaso. As barras de erros representam o desvio-padrão amostral. Todas médias foram iguais, de acordo com teste F a 5,0 % de significância.

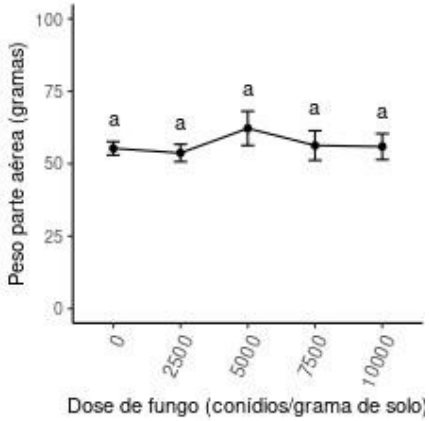


Figura 5: Peso em gramas da parte aérea (caule, folhas). As barras de erros representam o desvio-padrão amostral. Todas médias foram iguais, de acordo com teste F a 5,0 % de significância.

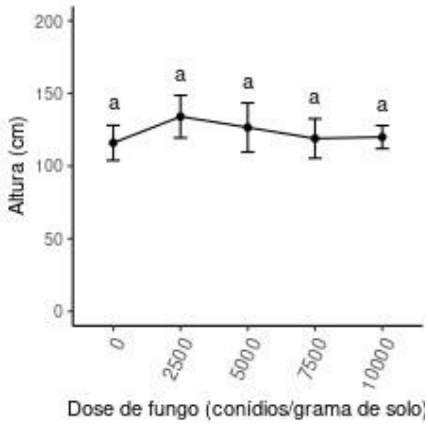


Figura 6: Altura da parte aérea (base do caule até o topo da planta). As barras de erros representam o desvio-padrão amostral. Todas médias foram iguais, de acordo com teste F a 5,0 % de significância.

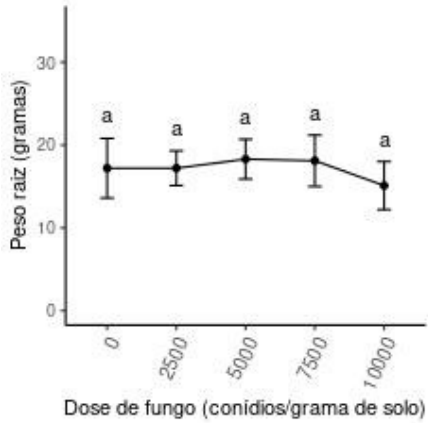


Figura 7: Peso da raiz em gramas. As barras de erros representam o desvio-padrão amostral. Todas médias foram iguais, de acordo com teste F a 5,0 % de significância.

Conclusão

Trichoderma orchidacearum COAD 3006 não tem capacidade de biocontrole de nematoide das galhas (*M. javanica*) em solo, assim como não foi capaz de promover crescimento vegetal de tomateiros nas condições experimentais utilizadas.

Apesar de não produzir os efeitos pretendidos, *T. orchidacearum* é um fungo endofítico pertencente a um gênero de fungos com potencial agronômico elevado, e portanto, pode de alguma forma ser benéfico para plantas. Novos testes, em menor escala, podem ser conduzidos com outras variedades vegetais e avaliando se o fungo de fato é capaz de colonizar os tecidos das plantas. Dessa froma, conclusões mais assertivas poderão ser tomadas quanto ao potencial agronômico de *Trichoderma orchidacearum* COAD 3006.

Referências

Boneti, J.I.S., Ferraz, S. 1981. Modificação do método de Hussey & Barker para extração de ovos de *Meloidogyne exigua* em raízes de cafeeiro. Fitopatologia Brasileira ,553.

Druzhinina, I. S., Seidl-Seiboth, V., Herrera-Estrella, A., Horwitz, B. A., Kenerley, C. M., Monte, E., Kubicek, C. P. (2011). Trichoderma: The genomics of opportunistic success. *Nature Reviews Microbiology*, *9*(10), 749–759. https://doi.org/10.1038/nrmicro2637

Stewart, A., & Hill, R. (2014). Applications of Trichoderma in Plant Growth Promotion. In *Biotechnology and Biology of Trichoderma*. https://doi.org/10.1016/B978-0-444-59576-8.00031-X

5. Conclusão geral

Dentre diversas espécies de *Penicillium* com genoma sequenciado, *P. ochrochloron* RLS11 possui comparativamente maior quantidade de genes codificantes para diversas famílias de CAZymes e essas enzimas foram produzidas pelo fungo quando cultivado em subtrato celulósico. Ensaios funcionais comprovaram que as enzimas produzidas por *P. ochrochloron* RLS11 são efetivas para despolimerização de biomassa vegetal, com potencial de aprimorar misturas enzimáticas comerciais para sacarificação de biomassa vegetal (Cellic® CTec2/HTec2). Além disso, análises genômicas e proteômica de *Penicillium ochrochloron* RLS11 forneceram uma riqueza de dados que podem ser usados para outras aplicações biotecnológicas envolvendo este fungo.

Por meio de análises de ortologia gênica e filogenia sustentamos a hipótese da descoberta de uma nova espécie do gênero *Trichoderma*, nomeada *T. orchidacearum* COAD 3006, sendo uma espécie evolutivamente distante das demais atualmente conhecidas e caracterizadas. Análises proteômica e ensaios funcionais indicaram uma grande capacidade desse fungo em produzir xilanases, enzimas fundamentais para despolimerizar a fração hemicelulósica de materiais vegetais. Além disso, avaliamos o potencial de *T. orchidacearum* COAD 3006 em manejo sustentável de pragas agrícolas (fitonematoides) e promoção de crescimento vegetal. Entretanto, os ensaios conduzidos em casa de vegetação indicaram que o fungo não apresentou os efeitos pretendidos.

Com o sequenciamento do genoma de *F. verticillioides* AZB aumentamos o número de genomas disponíveis desse importante patógeno de cereais, e as análises genômicas, transcriptômicas e evolutivas forneceram indícios de genes possivelmente relacionados à virulência. Além disso, a identificação e análise de regiões repetitivas (repetições em tandem e elementos transponíveis) contribuíram para maior entendimento da evolução de *F. verticillioides* e caracterização de genomas dessa espécie.

Com isso, no presente trabalho foram conduzidas análises funcionais, genômica e proteômica para acessar o potencial de produção de proteínas relacionadas com aplicabilidade biotecnológica, bem como caracterização/classificação de organismos de importância agronômica ou industrial. Em próximos estudos, poderão ser aprofundados os conhecimentos nesses tópicos a

partir dos dados já obtidos e gerar resultados de impacto no âmbito de utilização biotecnológica de fungos, tais como produção de CAZymes ou manejo de fungos fitopatogênicos (*Fusarium* spp.).

6. Artigos publicados em colaboração e outras contribuições acadêmicas

Somando-se ao trabalho apresentado, durante o período do doutorado realizei colaborações com diversos grupos de pesquisa da UFV, trabalhando principalmente em genômica funcional e evolução de microrganismos.

Fui co-autor de 2 artigos publicados recentemente nas revistas *Journal of Proteomics* (*Secretomic insight into the biomass hydrolysis potential of the phytopathogenic fungus Chrysoporthe cubensis*) e *Applied Biochemistry and Biotechnology* (*Genome-scale characterization of fungal phytases and a comparative study between beta-propeller phytases and histidine acid phosphatases*) em colaboração com o grupo da profa. Valéria Monteze Guimarães.

Além de contribuições em pesquisa, fui ministrante de mini-cursos em duas edições da jornada de inverno de Bioquímica e Biologia Molecular da UFV (2018 e 2019) em assuntos relacionados a genômica de fungos (*Avaliação do potencial de fungos fitopatogênicos para a produção de enzimas de interesse industrial*) e bactérias (*Bioinformática - da genômica a desenho de fármacos*).

7. Referências

Bills, G. F., & Gloer, J. B. (2016). Biologically Active Secondary Metabolites from the Fungi. *Microbiol Spectrum*, *4*(6), 1–32. https://doi.org/10.1016/0734-9750(93)90042-L

Bischof, R. H., Ramoni, J., & Seiboth, B. (2016). Cellulases and beyond: The first 70 years of the enzyme producer Trichoderma reesei. *Microbial Cell Factories*, *15*(1), 1–13. https://doi.org/10.1186/s12934-016-0507-6

Blacutt, A. A., Gold, S. E., Voss, K. A., Gao, M., & Glenn, A. E. (2018). Fusarium verticillioides: Advancements in understanding the toxicity, virulence, and niche adaptations of a model mycotoxigenic pathogen of maize. Phytopathology, 108(3), 312–326. https://doi.org/10.1094/PHYTO-06-17-0203-RVW

Brown, D. W., Busman, M., & Proctor, R. H. (2014). Fusarium verticillioides SGE1 is required for full virulence and regulates expression of protein effector and secondary metabolite biosynthetic genes. Molecular Plant-Microbe Interactions, 27(8), 809–823. https://doi.org/10.1094/MPMI-09-13-0281-R

Carvalho, F. P. (2006). Agriculture, pesticides, food security and food safety. *Environmental Science and Policy*, *9*(7–8), 685–692. https://doi.org/10.1016/j.envsci.2006.08.002

de Carvalho, L. M., Borelli, G., Camargo, A. P., de Assis, M. A., de Ferraz, S. M. F., Fiamenghi, M. B., … Carazzolle, M. F. (2019). Bioinformatics applied to biotechnology: A review towards bioenergy research. *Biomass and Bioenergy*, *123*(March), 195–224. https://doi.org/10.1016/j.biombioe.2019.02.016

De Oliveira, J. M. P. F., & De Graaff, L. H. (2011). Proteomics of industrial fungi: Trends and insights for biotechnology. *Applied Microbiology and Biotechnology*, *89*(2), 225–237. https://doi.org/10.1007/s00253-010-2900-0

EIA, U. S. E. I. A. (2019a). EIA projects nearly 50% increase in world energy usage by 2050, led by growth in Asia. Retrieved from https://www.eia.gov/todayinenergy/detail.php?id=41433

EIA, U. S. E. I. A. (2019b). *International Energy Outlook 2019 - with projections to 2050* (p. 85). p. 85. https://doi.org/10.5860/choice.44-3624

Fasciotti, M. (2017). Perspectives for the use of biotechnology in green chemistry applied to biopolymers, fuels and organic synthesis: from concepts to a critical point of view. *Sustainable Chemistry and Pharmacy*, *6*(March), 82–89. https://doi.org/10.1016/j.scp.2017.09.002

Ghorbanpour, M., Omidvari, M., Abbaszadeh-Dahaji, P., Omidvar, R., Kariman, K., & Kariman, K. (2018). Mechanisms underlying the protective effects of beneficial fungi against plant diseases. *Biological Control*, *117*, 147–157. https://doi.org/10.1016/j.biocontrol.2017.11.006

Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., & Foust, T. D. (2007). Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science*, *315*(5813), 804–807. https://doi.org/10.1126/science.1137016

Jayasekara, S., & Renuka, R. (2016). Microbial Cellulases: An Overview and Applications. *IntechOpen*, 21.

Kaewchai, S., Soytong, K., & Hyde, K. D. (2009). Mycofungicides and fungal biofertilizers. *Fungal Diversity*, *38*, 25–50.

Kilbane, J. J. (2016). Future applications of biotechnology to the energy industry. *Frontiers in Microbiology*, *7*(FEB), 4–7. https://doi.org/10.3389/fmicb.2016.00086

Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., & Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels*, *6*(1), 1–14. https://doi.org/10.1186/1754-6834-6-41

OECD. (2018). Global Material Resources Outlook to 2060 - Economic drivers and environmental consequences. *Global Material Resources Outlook to 2060*, p. 24. https://doi.org/10.1787/9789264307452-en

Palumbo R, Crisci A, Venâncio A, Abrahantes JC, Dorne J Lou, Battilani P, et al. Occurrence and co-occurrence of mycotoxins in cereal-based feed and food. Microorganisms 2020;8. https://doi.org/10.3390/microorganisms8010074.

Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., … Tschaplinski, T. (2006). The path forward for biofuels and biomaterials. *Science*, *311*(5760), 484–489. https://doi.org/10.1126/science.1114736

Rampersad SN. Pathogenomics and management of fusarium diseases in plants. Pathogens 2020;9. https://doi.org/10.3390/pathogens9050340.

Rhodes, C. J. (2018). Plastic pollution and potential solutions. *Science Progress*, *101*(3), 207–260. https://doi.org/10.3184/003685018X15294876706211

Richards, T. A., & Talbot, N. J. (2013). Horizontal gene transfer in osmotrophs: Playing with public goods. *Nature Reviews Microbiology*, *11*(10), 720–727. https://doi.org/10.1038/nrmicro3108

Schmidt, S.M. and Panstruga, R. Pathogenomics of fungal plant parasites: what have we learnt about pathogenesis? *Current Opinion in Plant Biology*, (14), 392–399. http://dx.doi.org/10.1016/j.pbi.2011.03.006

Schmit, J. P., & Mueller, G. M. (2007). An estimate of the lower limit of global fungal diversity. *Biodiversity and Conservation*, *16*(1), 99–111. https://doi.org/10.1007/s10531-006-9129-3

Stajich, J. E. (2017). Fungal genomes and insights into the evolution of the kingdom. *Microbiol Spectr.*, *5*(4), 1–25. https://doi.org/10.1016/j.physbeh.2017.03.040

Sun, F., Marie Rose Mukasekuru, D. C., Wei, Y., Han, L., Lin, X., & Fang, X. (2018). Determination of Cellulase Activities and Model for Lignocellulose Saccharification. In X. Fang & Q. Yinbo (Eds.), *Fungal Cellulolytic Enzymes Microbial Production and Application* (pp. 223–238). Springer.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., … Abarenkov, K. (2014). Global diversity and geography of soil fungi. *Science (New York, N.Y.)*, *346*(6213), 1052–1053. https://doi.org/10.1126/science.aaa1185

Topolovec-Pintarić, S. (2016). Trichoderma: Invisible Partner for Visible Impact on Agriculture. *IntechOpen*, 13. https://doi.org/http://dx.doi.org/10.5772/57353

UN. (2017). World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100. Retrieved from https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html

Vaishnav, N., Singh, A., Adsul, M., Dixit, P., Sandhu, S. K., Mathur, A., … Singhania, R. R. (2018). Penicillium : The next emerging champion for cellulase production. *Bioresource Technology Reports*, *2*(2017), 131–140. https://doi.org/10.1016/j.biteb.2018.04.003

Vijayavenkataraman, S., Iniyan, S., & Goic, R. (2012). A review of climate change, mitigation and adaptation. *Renewable and Sustainable Energy Reviews*, *16*(1), 878–897. https://doi.org/10.1016/j.rser.2011.09.009

Wilken, S. E., Swift, C. L., Podolsky, I. A., Lankiewicz, T. S., Seppälä, S., & O'Malley, M. A. (2019). Linking 'omics' to function unlocks the biotech potential of non-model fungi. *Current Opinion in Systems Biology*, *14*(February), 9–17. https://doi.org/10.1016/j.coisb.2019.02.001

Wu, B., Hussain, M., Zhang, W., Stadler, M., Liu, X., & Xiang, M. (2019). Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. *Mycology*, *10*(3), 127–140. https://doi.org/10.1080/21501203.2019.1614106

Zhao, Z., Liu, H., Wang, C., & Xu, J. R. (2014). Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*, *15*(1). https://doi.org/10.1186/1471-2164-15-6