UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MICROBIOLOGIA

Paula Luize Camargos Fonseca

Investigando o mitogenoma de *Trichoderma harzianum*: evolução, perfil transcricional e mecanismos regulatórios

Belo Horizonte
2021

Paula Luize Camargos Fonseca

Investigando o mitogenoma de *Trichoderma harzianum*: evolução, perfil transcricional e mecanismos regulatórios

Versão final

Tese apresentada ao Programa de Pós-Graduação em Microbiologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais como requisito para a obtenção do título de Doutora em Microbiologia.

Orientador :Prof. Dr. Aristóteles Góes Neto

Co-orientadores: Prof. Dr. Eric Roberto Guimarães Rocha Aguiar
Prof. Dra. Fernanda Badotti

Belo Horizonte

2021

043 Fonseca, Paula Luize Camargos.
Investigando o mitogenoma de Trichoderma harzianum: evolução, perfil transcricional e mecanismos regulatórios [manuscrito] / Paula Luize Camargos Fonseca. – 2021.

143 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Aristóteles Góes Neto. Co-orientadores: Prof. Dr. Eric Roberto Guimarães Rocha Aguiar; Profª. Dra. Fernanda Badotti.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Microbiologia.

1. Microbiologia. 2. Trichoderma. 3. Hypocreales. 4. Regulação. 5. Interferência de RNA. I. Goés Neto, Aristóteles. II. Aguiar, Eric Roberto Guimarães Rocha. III. Badotti, Fernanda. IV. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. V. Título.

CDU: 579

Ficha Catalográfica elaborada por Sônia M S Moraes CRB 6 1357

UNIVERSIDADE FEDERAL DE MINAS GERAIS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MICROBIOLOGIA

ATA DA DEFESA DE TESE DE PAULA LUIZE CAMARGOS FONSECA
Nº REGISTRO: 2017663691

Às 14:00 horas do dia 11 de março de 2021, reuniu-se, no Instituto de Ciências Biológicas da UFMG, a Comissão Examinadora composta pelos Drs. Francisco Pereira Lobo, João Trindade Marques, Eliane Ferreira Noronha, Leandro Lopes Loguercio e o Prof. Aristóteles Goés Neto – Orientador, para julgar o trabalho final "Investigando o Mitogenoma de *Trichoderma harzianum*: Evolução, perfil Transcricional e Mecanismos Regulatórios", da aluna **Paula Luize Camargos Fonseca**, requisito final para a obtenção do Grau de **DOUTOR EM CIÊNCIAS BIOLÓGICAS: MICROBIOLOGIA**. Abrindo a sessão, o Presidente da Comissão, Prof. Flávio Guimarães da Fonseca - Coordenador do Programa, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para a apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. A candidata foi considerada **APROVADA**. O resultado final foi comunicado publicamente à candidata 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, 11 de março de 2021. O candidato tem 60 (sessenta) dias, a partir desta data, para entregar a versão final da tese ao Programa de Pós-Graduação em Microbiologia da UFMG e requerer seu diploma.

Prof. Dr. Francisco Pereira Lobo

Prof. Dr. João Trindade Marques

Profa. Dra. Eliane Ferreira Noronha

Prof. Dra. Leandro Lopes Loguercio

Prof. Dr. Aristóteles Goés Neto (Orientador) Prof. Dr. Aristóteles Góes-Neto

Prof. Flávio Guimarães da Fonseca
Coordenador

AGRADECIMENTOS

Muitas pessoas participaram da realização deste trabalho e da minha formação como pesquisadora, seria impossível descrever a importância de cada um durante essa trajetória, contudo, todos conseguiram deixar a sua marca na minha história e coração.

Primeiramente, gostaria de agradecer a Deus que me deu a vida e que sempre esteve presente nos momentos mais difíceis e turbulentos. Muitas pessoas acreditam que ciência e religião não podem ser misturadas, mas o fato é que não consigo negar a presença de Deus ao meu redor, e sou muito grata pelo presente maravilhoso que é viver e a oportunidade de auxiliar no desenvolvimento do nosso país através da pesquisa.

Aos meus pais, Cristiane e Júnior por sempre acreditarem em mim, pelo amor incondicional e por terem me ensinado e me deixado a melhor herança que pais podem deixar, os valores de união, amizade, educação, respeito, disciplina, compreensão e honestidade. Ao meu irmão Pedro, pelo amor, compreensão, amizade e união. Seus exemplos de persistência, luta, respeito e humildade são responsáveis por grande parte da minha personalidade e escolhas de vida. Eu amo muito vocês e sou muito feliz por fazer parte dessa família.

Ao Charles, pelo amor, compreensão, respeito e paciência durante todo o doutorado. Seu suporte foi muito importante para mim. Aos meus sogros, Rose e Vanderley que me acolheram como filha e se tornaram minha segunda família.

Aos meus avós, tios, primos e agregados das famílias Camargos e Fonseca que fazem a minha vida mais alegre, simples e amável. Amo muito cada um de vocês, mas gostaria de fazer um agradecimento especial aos meus tios Iran, Eduardo, Odilon, Júlio, Cleyde, Hélia e Tânia por cuidarem de mim durante a minha infância. Sou eternamente grata a vocês. Aos meus avós pelos melhores cafés da tarde e almoços de domingo, melhores risadas e conversa despreocupada. Aos meus primos Luiz Gustavo, Cadu, Victor e João Paulo que sempre me "deram trabalho" durante o final de ano letivo. Vocês foram essenciais na minha formação como professora.

Aos meus amigos, pelo amor e amizade em todos os momentos de felicidade ou tristeza, todos foram muito valiosos para mim. Em especial a Nata do René, aos amigos do GAG do LabVírus, aos amigos da graduação por me auxiliarem e incentivarem na minha decisão de continuar na pesquisa. Ao meu time de paixão, Clube Atlético Mineiro, que apesar de fazer muita raiva nos seus fiéis torcedores, me permite presenciar uma das melhores sensações da vida, a mistura de cores, raças, sexos, idades e status social em um mesmo ambiente com respeito e amor.

A todos os brasileiros que contribuem com o pagamento dos impostos e permitem o financiamento da pesquisa no nosso país. Ao Governo brasileiro e as agências de fomento pela concessão de bolsas de pós graduação e investimento em pesquisas, pois somente assim poderemos desenvolver o país nos âmbitos da economia, educação, saúde, segurança e social.

Ao Prof. Aristóteles Góes-Neto, pela orientação, ensinamentos e alegria contagiante. Sempre serei grata por ter me dado a primeira oportunidade de recomeçar na pesquisa e a todas as oportunidades dadas a mim durante o período de trabalho. Ao Prof. Eric Aguiar, que mesmo sendo flamenguista, foi fundamental para a realização deste trabalho. Obrigada pela orientação, ensinamentos, paciência, oportunidades, puxões de orelha, amizade e também o jeito baiano de ser. A Profa. Fernanda Badotti, pela orientação, alegria, ensinamentos que foram imprescindíveis para a minha formação como pesquisadora. Foi um privilégio desenvolver a minha tese sob a orientação de três mentes brilhantes, e sempre serei grata a vocês por todos os momentos.

A Profa. Aline Vaz, pela amizade, ensinamentos e apoio. Sempre serei grata a você por ter me ensinado os princípios básicos de biologia molecular e estatística. Ao Prof. Colaborador Vasco Azevedo por ter aberto as portas do seu laboratório e pela oportunidade de integrar o seu grupo de pesquisa. Ao Prof. Colaborador Bertram Brenig pela disponibilidade e oportunidade de sequenciar as bibliotecas de DNA e RNA geradas nesta tese e pelo apoio e ensinamentos durante as nossas conversas virtuais. Ao Prof. Colaborador Luiz-Eduardo Del-Bem pelos ensinamentos e disponibilidade em auxiliar no desenvolvimento da minha tese. Ao Isaque Faria pela ajuda com os experimentos e análises dos dados de transcritoma. Todos vocês foram essenciais para o meu progresso científico.

Aos colegas e amigos do Laboratório de Biologia Molecular e Computacional de Fungos, pelos bons momentos e lutas nestes quatro anos de doutorado. Gostaria de agradecer em especial ao Marcelo, Thairine, Luz, Felipe, Gabriel, Daniel, Ruth e Mike pelas conversas, ensinamentos, amizade, parceria e alegria durante todos os dias. Em especial ao Marcelo que desde a época de Iniciação Científica está presente me ajudando e colaborando no meu projeto. Aos colegas e amigos do Laboratório de Bioinformática de Vírus pelas reuniões, ensinamentos e amizade. Vocês foram fundamentais para a elaboração dessa tese e para o meu crescimento profissional. Sempre lembrarei dos nossos almoços, lanches, churrasco dentro de ônibus e festividades.

A todos os professores que eu tive durante a minha jornada acadêmica. Vocês foram essenciais na minha escolha de profissão. A todos os colegas do programa de pós graduação em Microbiologia pelo apoio e convivência agradável.

Aos membros da banca examinadora por aceitarem o convite e pelas sugestões que serão essenciais para a conclusão da minha tese. A todos, o meu muito obrigada.

Resumo

A mitocôndria é uma organela citoplasmática encontrada na maioria dos organismos eucariotos e é responsável pela respiração e homeostase celular. O genoma mitocondrial (mitogenoma) vem sendo objeto de estudos em genética comparativa e evolução, uma vez que apresenta genes conservados entre todas as espécies, bem como variações em regiões intrônicas e repetitivas. A ordem Hypocreales é composta por espécies de fungos que desempenham diversas funções ecológicas, como sapróbios, patógenos e mutualistas. Uma das espécies de destaque é *Trichoderma harzianum*, que vem sendo muito empregado no controle biológico de outros fungos e nematódeos. O primeiro capítulo deste trabalho teve como objetivo investigar a variabilidade e evolução de 35 mitogenomas da ordem Hypocreales. Variações na estrutura e tamanho dos mitogenomas foram correlacionadas à presença de regiões não-codificantes. Análises evolutivas estimaram que mitogenomas que evoluem em uma taxa maior, apresentam uma região não-codificante maior. Apesar da variação no tamanho e conteúdo não-codificante, os mitogenomas avaliados apresentam um grupo de genes conservados, sendo o gene ribossomal *rns* um dos únicos contendo apenas um exon. A conservação gênica, o número de cópias por célula e a sua presença na maioria dos organismos eucariotos, tornam os mitogenomas promissores marcadores moleculares para a classificação taxonômica de espécies. No capítulo dois, avaliamos o potencial do gene *rns* para a identificação de fungos da ordem Hypocreales. O potencial do *rns* como marcador molecular foi comparada às de outros genes comumente empregados na identificação de fungos, incluindo o gênero *Trichoderma*, o qual possui inúmeras espécies crípticas, por isso sendo utilizado como estudo de caso. Os resultados encontrados sugerem que o gene *rns,* é capaz de separar todas as espécies da ordem Hypocreales testadas, e é suficiente para a identificação correta dos isolados de *Trichoderma*. Entretanto, a combinação dos genes *rns* e calmodulina se mostrou uma alternativa eficiente para a identificação de espécies do gênero. Mitocôndrias, apesar de possuírem o seu próprio genoma, necessitam de genes codificados no núcleo para o seu funcionamento. Grande parte do conhecimento sobre os genes envolvidos na regulação de mitogenomas é restrito a organismos modelo, como *Homo sapiens, Neurospora crassa* ou *Saccharomyces cerevisiae*. No entanto, alguns estudos já demonstraram em *N. crassa* a ausência de genes identificados em *H. sapiens* e *S. cerevisiae* que são fundamentais para o controle do genoma. Desta forma, com o objetivo de ampliar o conhecimento sobre os mecanismos de controle dos mitogenomas,

resultados do sequenciamento, montagem e anotação do genoma nuclear de *T. harzianum* e investigação da presença de genes com localização na organela mitocondrial são apresentados no capítulo três desta tese. Ao todo, foram identificados 159 genes, dos quais a maioria está envolvida no funcionamento da cadeia fosforilativa. Um número menor de genes está associado a regulação da replicação, transcrição e tradução do mitogenoma, enquanto 39 genes não tiveram sua função definida, sugerindo que os mecanismos de controle do mitogenoma são ainda pouco conhecidos em fungos filamentosos. Por último, no capítulo quatro, o perfil transcricional do mitogenoma de *T. harzianum* foi avaliado, estimando-se a capacidade do genoma em gerar RNA dupla fita (dsRNA), uma vez que o mtDNA é circular e apresenta características que sugerem existência de transcrição bidirecional. Através de amplificação fita específica foi possível confirmar a presença de dsRNA no gene *rns*, um dos genes com maior transcrição. Em outros grupos de organismos, como os animais, o dsRNA já havia sido reportado. Contudo, a detecção de dsRNA em mitocôndrias de fungos está sendo descrita pela primeira vez nesta tese. Em mamíferos, existe um complexo responsável pelo controle do dsRNA gerado, uma vez que pode alterar os processos celulares. No genoma nuclear de *T. harzianum* encontramos apenas um dos genes envolvidos no controle, sugerindo que em fungos filamentosos, ele também poderia ter a mesma função. Neste capítulo também foi possível estimar a geração de pequenos RNAs provenientes do processamento do dsRNA de origem mitocondrial no *T. harzianum*. Os principais genes envolvidos no funcionamento da via de RNA de interferência foram identificados, sugerindo que a via também possua algum papel no controle pós-transcricional do mitogenoma. Os resultados obtidos nesta tese demonstram que mitogenomas de fungos da ordem Hyporeales estaão sofrendo um processo de modificação estrutural gênica, provavelmente induzido por elementos móveis, e que a regulação pode ser realizada por genes ainda não descritos funcionalmente na mitocôndria.

Palavras-chave: Mitogenômica comparativa. Evolução. Hypocreales. Regulação. RNA de interferência. Transcrição. dsRNA.

Abstract

Mitochondria is a cytoplasmic organelle found in most eukaryotic cells and are responsible for respiration and cellular homeostasis. The mitochondrial genome (mitogenome) has been the subject of studies in comparative genetics and evolution, since it has conserved genes among all species, as well as variations in intronic and repetitive regions. The order Hypocreales is composed of fungal species that perform several ecological functions, which can be saprobes, pathogens or mutualistic. One of the prominent species is *Trichoderma harzianum*, which has been widely used in biological control of other fungi and nematodes. The first chapter of this thesis aimed to investigate the variability and evolution of 35 mitogenomes from the order Hypocreales. Variation in structure and size of mitogenomes were correlated with the presence of non-coding regions. Evolutionary analysis has estimated that fast-evolving mitogenomes have a larger non-coding region. Additionally, some intron sequences are conserved among species, suggesting horizontal transfer processes. Despite the variation in size and non-coding content, the mitogenomes evaluated have a group of conserved genes, with the ribosomal gene *rns* is being the only one containing one exon. The genetic conservation, the number of copies per cell and their presence in most eukaryotic organisms, make them a potential molecular marker for taxonomic classification of species. In chapter two, we evaluated whether the *rns* gene for the identification of fungal species from Hypocreales. The sequences of *rns* were compared to other genes commonly used to identify Hypocrealean fungi, including the genus *Trichoderma*, which has numerous cryptic species, as a case study. The results suggested that the *rns* gene, despite separating all the Hypocrealean species, is not sufficient for the correct identification of all the *Trichoderma* isolates. Nonetheless, the use of *rns* and calmodulin genes proved to be an efficient combination for the identification of species from the genus. Mitochondria, in spite of having their own genome, need genes encoded in the nucleus for their functioning. Much of the knowledge about the genes involved in their regulation is specific only to model organisms, such as *Homo sapiens*, *Neurospora crassa* or *Saccharomyces cerevisiae*. Nevertheless, some studies have already demonstrated in *N. crassa* the absence of genes identified in *H. sapiens* and *S. cerevisiae*, which are fundamental for mitogenome control. Thus, with the objective of expanding knowledge about mitogenome control mechanisms, the results of sequencing, assembly, and annotation of the *T. harzianum* nuclear genome, and the presence of genes with location in the mitochondria was investigated and

presented in chapter three of this thesis. Altogether, 159 genes were identified, of which the majority are involved in the functioning of the phosphorylation chain. A smaller number of genes have been identified in the replication, transcription, and translation regulation, while 39 genes have not had their function defined yet, suggesting that the mitogenome control mechanism is still poorly understood. Finally, in chapter four, the transcriptional profile of *T. harzianum* mitogenome was evaluated, estimating the capacity of the genome to generate double-stranded RNA (dsRNA), since it is circular and has a profile similar to bidirectional replication. Using specific-strand amplification, it was possible to confirm the presence of dsRNA in the *rns* gene, one of the genes with the highest transcription. In other groups of organisms, such as mammals, dsRNA had already been reported, however, the detection of dsRNA in fungal mitochondria is being described for the first time in this thesis. In mammals, there is a complex responsible for controlling the dsRNA generated, as it can alter cellular processes. In the nuclear genome of *T. harzianum* we found only one of the genes responsible for the control, suggesting that, in filamentous fungi, it could also have the same function. In this chapter it was also possible to estimate the generation of small RNA from *T. harzianum* mitogenome. The main genes of the RNA interference pathway were identified, indicating that this pathway also play a key role in mitogenome post-transcriptional control. The results obtained in this thesis demonstrate that the fungal mitogenomes have been still undergoing a process of structural genetic shaping, probably induced by mobile elements; and that regulation can be carried out by genes not yet functionally described in mitochondria.

Keywords: Comparative mitogenomics. evolution. Hypocreales. Regulation. Interference RNA. Transcription. dsRNA.

SUMÁRIO

RELEVÂNCIA E JUSTIFICATIVA

Mitocôndrias são organelas especializadas e localizadas no citoplasma de células eucarióticas, responsáveis pelo fornecimento de energia através da respiração aeróbica (via de fosforilação oxidativa), biossíntese de lipídios, apoptose, homeostase, dentre outros. O funcionamento da mitocôndria depende de proteínas traduzidas no citoplasma celular e posterior importação para a matriz mitocondrial. Apesar da dependência do genoma nuclear, a organela possui o seu próprio genoma (mitogenoma). Os mitogenomas são moléculas de DNA circulares ou lineares de tamanho variável que se replicam independentemente do ciclo celular nuclear. São constituídos por genes conservados essenciais, elementos transponíveis, *homing endonucleases* e regiões com potencial codificante não-conservados. Devido a diversidade de formas, tamanho e genes, o uso de mitogenomas fúngicos tem se mostrado uma estratégia promissora para estudos evolutivos e filogenéticos, principalmente na classificação correta de espécies, adaptabilidade em diferentes ambientes e investigação dos processos de modulação do mitogenoma na célula.

Apesar do importante papel desempenhado pela mitocôndria na célula, o conhecimento detalhado sobre a regulação gênica desta organela ainda é escasso, e pouco se sabe sobre o controle da expansão/contração de genomas mitocondriais em fungos. O sequenciamento de novos genomas mitocondriais de fungos, a caracterização genômica, assim como estudos de transcritoma, são de grande importância para auxiliar na expansão de conhecimento sobre o genoma dessa organela. A análise dos componentes gênicos mitocondriais e, principalmente, a investigação da função desempenhada por diferentes genes no controle do genoma mitocondrial de fungos são fundamentais para a compreensão da regulação mitocondrial e do controle da integridade do genoma. A investigação do funcionamento da organela mitocondrial e de seu genoma podem ainda fornecer novas estratégias para controle de fungos patógenos, assim como pode auxiliar na melhoria de produção de metabólitos secundários. Além disso, estes estudos ainda ajudam na investigação da influência da mitocôndria no funcionamento de processos celulares que são essenciais para o crescimento e desenvolvimento de fungos em diferentes ambientes.

Além disso, em mamíferos alguns estudos relataram que o DNA mitocondrial pode gerar RNA dupla-fita (dsRNA), o qual é degradado por um complexo de degradação

mitocondrial, contudo, ainda não há informações sobre a presença de dsRNA ou a presença deste complexo em fungos filamentosos. No citoplasma celular existem vias imunes que são responsáveis pelo controle pós-transcricional e que processam dsRNA de origem viral, e também podem ter função no processamento do dsRNA de origem mitocondrial.

Fungos da ordem Hypocreales apresentam diversas funções ecológicas, sendo essenciais em vários ecossistemas, e, portanto, constituem um excelente modelo para o estudo da evolução de genomas no reino Fungi. Além disso, apesar de a ordem Hypocreales apresentar grande interesse científico, este é o primeiro trabalho de análise de mitogenômica comparativa das espécies disponíveis em bancos de dados públicos.

Formatação da tese

Esta tese é composta por uma introdução geral, justificativa e relevância, objetivos, capítulos específicos, discussão integrativa, conclusões finais e perspectivas do projeto. O **capítulo um** refere-se ao estudo de mitogenômica comparativa da ordem Hypocreales e apresentação do mitogenoma de *Trichoderma harzianum*. Este trabalho foi aceito e publicado na revista *Frontiers in Microbiology*. O **capítulo dois** trata do potencial de uso do gene ribossomal *rns* para a identificação taxonômica de espécies da ordem Hypocreales. Este trabalho foi submetido para publicação na revista *FEMS Microbiology Letters*. O **capítulo três** fala sobre a montagem, anotação e identificação de genes de origem nuclear que são importados para a matriz mitocondrial em *T. harzianum*. O **capítulo quatro** é sobre o perfil transcricional, detecção de RNA dupla fita mitocondrial e caracterização da via de RNA de interferência em *T. harzianum*. A discussão integrada, as conclusões e as perspectivas do trabalho são encontradas após os quatro capítulos.

INTRODUÇÃO GERAL

Mitocôndrias são organelas presentes na maioria dos organismos eucariontes (aeróbios obrigatórios) e são responsáveis pelo fornecimento de energia (na forma de moléculas de ATP), utilizadas em diversos processos celulares (Friedman and Nunnari 2014). Outras funções também desempenhadas por essas organelas são biossíntese de lipídeos, homeostase e apoptose celular, além de auxiliar a célula na resposta a estresses ambientais (Spinelli and Haigis 2018). A origem da mitocôndria está associada a uma Alphaproteobactéria ancestral de vida livre, que se integrou na célula hospedeira, processo conhecido como endossimbiose (Sagan 1967; Gray 2017; Martijn et al. 2018). Durante a integração entre as duas células, ocorreu a aquisição da maquinaria de importação de RNA e proteína, modificação das membranas do endossimbionte, transferência de grande parte do genoma para o núcleo celular e formação da organela mitocondrial (Roger et al. 2017). Apesar da transferência de genes para o genoma nuclear, as mitocôndrias possuem o seu próprio genoma, no qual há a informação para o funcionamento da respiração aeróbica através do funcionamento da cadeia fosforilativa (Zardoya 2020).

Genomas mitocondriais são comumente denominados de mitogenomas ou mtDNA, podem ser circulares ou lineares e utilizar códigos genéticos diferentes daqueles utilizados pelo genoma nuclear (Zardoya 2020; Kulik et al. 2021). Em fungos, considerando os dois principais filos (Ascomycota e Basidiomycota), o conteúdo gênico mitocondrial é conservado em relação a presença de um *core* genome com 14 genes codificadores de proteínas (*nad1*-6, 4L, *cob*, *cox1*-3 e *atp6*, 8 e 9), dois genes ribossomais (*rns* e *rnl*) e um gene codificador de proteína ribossomal (*rps3*) envolvidos no processo de fosforilação oxidativa (Zardoya 2020). Entretanto, exceções podem ser encontradas. Adicionalmente, o tamanho do mitogenoma é bastante variável entre as espécies: por exemplo, a espécie *Spizellomyces punctatus* possui 1,136 pb, enquanto a espécie *Ophiocordyceps camponotifloridani* apresenta 272,497 pb. Essa variação é devida a presença de elementos acessórios como os íntrons, regiões intergênicas, ORF sem função definida (uORF) e genes de *homing endonucleases* (HEG) (Hausner 2003; Kulik et al. 2021).

Os íntrons são sequências nucleotídicas inseridas os genes e são removidas durante o processo de *splicing* alternativo. Em mitogenomas de fungos são encontrados íntrons ds grupos I e II, que diferem na sequência, estrutura e *splicing* (Craig et al. 2002;

Hausner 2003). Sequências de íntrons já foram identificadas em diferentes mitogenomas apresentando alta similaridade, e a disseminação dessas sequências através de eventos de transferência horizontal já foi observada em diferentes grupos de espécies (Hausner 2003; Fonseca et al. 2020; Kulik et al. 2021). Uma das hipóteses para essa transferência é a codificação de HEGs que são classificadas como sequências egoístas (*selfish* DNA) que possuem a capacidade de transposição e são responsáveis pela quebra da dupla fita de DNA em sítios específicos, que consequentemente ativa o processo de reparo do DNA por recombinação homóloga e geração de conversão gênica. As HEGs também podem ser encontradas livres, o que garante a sua disseminação e fixação no mitogenoma e pode inclusive, alterar a janela de leitura de genes codificadores de proteína (Hausner 2003; Barzel et al. 2011; Stoddard 2014). Atualmente existem seis famílias de HEGs, contudo, apenas duas famílias são encontradas nos mitogenomas fúngicos, LAGLIDADG e GIY-YIG (Hausner 2003). Adicionalmente a presença de íntrons e HEGs, mitogenomas de fungos também podem apresentar uORFs, cujo produto não possui similaridade com nenhum depósito disponível nos bancos de dados públicos. Assim como as HEGs, uORFs também podem estar associadas a regiões intrônicas e podem ser alteradas pela ação das HEGs (Sellem et al. 1996; Hausner 2003).

Por apresentarem alta variação estrutural em diferentes espécies biológicas, os mitogenomas de fungos vêm sendo muito estudados para investigar a origem, evolução e o processo de integração do endossimbionte (Kulik et al. 2021). Apesar da diversidade de tamanho e de genes, mitogenomas são uma ótima ferramenta para a elucidação de relações evolutivas entre espécies por apresentar uma baixa taxa de recombinação, herança maternal, diferença da taxa de mutação em relação ao genoma nuclear, o número de cópias no citoplasma celular e a facilidade para amplificação e sequenciamento de uma região ou o mitogenoma inteiro (Rubinoff and Holland 2005). Alguns trabalhos já foram desenvolvidos demonstrando a possibilidade do uso de genes mitocondriais para identificação de espécies fúngicas, entretanto é necessário o sequenciamento de mitogenomas de um número maior de espécies fúngicas para avaliar a presença de um marcador universal (Kulik et al. 2015; Smith 2016; Avin et al. 2017).

Até o mês de dezembro de 2020, 788 mitogenomas de fungos estavam depositados no banco de dados do *National Center of Biotechnology Information* (NCBI), sendo os filos Ascomycota e Basidiomycota, que compreendem o maior número de espécies descritas são os mais amostrados (601 e 129 mitogenomas respectivamente). Por outro lado, apenas oito (Ascomycota, Basidiomycota, Blastocladiomycota, Chytridiomycota,

Cryptomycota, Monoblepharidoycota, Mucoromycota e Zoopagomycota) dos 12 filos do reino Fungi estavam disponíveis, o que demonstra o viés amostral em dados públicos relacionados a sequenciamento de mitogenomas de fungos. Nos últimos anos, pode ser observado um número crescente de depósito de mitogenomas, sendo que a maioria dos estudos foca na descrição e análise comparativa dos mitogenomas (Kolondra et al. 2015), mas não extrapolam para o entendimento dos mecanismos envolvidos no controle e regulação dos genomas mitocondriais ou na interação deste com o genoma nuclear (Kolondra et al. 2015; Kulik et al. 2021).

O mitogenoma possui replicação independente do genoma nuclear e durante a transcrição, transcritos policistrônicos derivados de ambas as fitas (senso e anti-senso) são geradas, que são então processadas em mRNA, tRNA e rRNA. No entanto, existe a possibilidade de geração de RNA dupla fita (dsRNA) devido a transcrição simultânea das fitas do genoma, que pode ser acumulado na matriz mitocondrial (Dhir et al. 2018). A degradação dessas moléculas é realizada pelo degradossomo mitocondrial (mtEXO), responsável pela manutenção da integridade da mitocôndria, assim como a degradação de dsRNA na organela (Razew et al. 2018). O mtEXO é composto por dois genes codificadores de proteína: SUV3 DNA helicase e PNPase/PNPT1 no qual contém uma polimerase com atividade de exoribonuclease (Szczesny et al. 2010). Os dois genes são codificados no genoma nuclear e então importados para a mitocôndria. A proteína SUV3 tem capacidade de desenrolar o dsRNA, enquanto a PNPase é responsável pela degradação das moléculas. Em mamíferos, o nocaute desses dois genes é letal em embriões de camundongos (Wang et al. 2010; Chen et al. 2013), e em leveduras, a ausência dos genes também é letal, mas, como a respiração também pode ser realizada pela via fermentativa, o mitogenoma é perdido e as células ficam na forma de "*petite*" (células sem capacidade de crescimento em meios de cultivo com fonte não-fermentativa) (Stepien et al. 1992; Guo et al. 2011). A ausência dos genes do mtEXO favorece a acumulação de dsRNA na matriz mitocondrial, com consecutivo escape para o citoplasma celular (Dhir et al. 2018). No citoplasma, essas moléculas podem se tornar imunogênicas e ativar vias de imunidade do próprio hospedeiro (Dhir et al. 2018). Uma das vias capazes de degradar dsRNA no citoplasma é a via de RNA de interferência (RNAi) (Hannon 2002).

A via de RNAi é um mecanismo conservado em eucariotos, no qual pequenos RNAs são responsáveis pela regulação gênica em nível pós-transcricional (Hannon 2002; Lee et al. 2010; Nicolás and Ruiz-Vázquez 2013). Foi descrita pela primeira vez no

metazoário *Caenorhabditis elegans*, em um mecanismo mediado por dsRNA. Posteriormente, esse mecanismo foi descrito em outros animais, protozoários, plantas e fungos (Nakayashiki et al. 2006; Nicolás and Ruiz-Vázquez 2013).

Os principais componentes da maquinaria da via de RNAi, são as proteínas Dicer (DCL), Argonauta (AGO) e RNA polimerase dependente de RNA (RdRp). A proteína DCL pertence a família das RNases III, cuja função é processar longos RNAs de fita dupla em pequenas sequências de tamanho entre 21 a 26 nucleotídeos. As proteínas AGO são essenciais em complexos efetores multi-subunitários e contêm dois domínios conservados, o motivo PAZ (responsável pela ligação com ácidos nucléicos, com preferência por DNA e RNA de fita simples) e o domínio Piwi (responsável pela clivagem de uma das fitas do dsRNA) (Decker et al. 2015). A RdRp está envolvida na amplificação dos fragmentos de dsRNA reconhecidos e processados pela DCL (Nakayashiki et al. 2006; Nolan et al. 2008; Dang et al. 2011a). O produto final do funcionamento da via de RNAi são pequenos RNAs (sRNAs), gerados através do processamento de precursores de dsRNA pela DCL com consecutiva seleção de uma das fitas de dsRNA para serem usadas como molde no complexo se silenciamento induzido por RNA (RISC), no qual contém a enzima AGO e é responsável pelo reconhecimento da região alvo ao dsRNA e desestabilização do mRNA (Dang et al. 2011b).

Os sRNAs gerados são sequências curtas de RNA de tamanho geralmente entre 20-30 nt que podem ser classificados em três classes principais (microRNAs – miRNAs; *piwi-interacting* RNAs – piRNAs e pequenos RNAs de interferência – siRNAs) de acordo com a sua biogênese, modo de regulação e as vias biológicas reguladas (Hannon 2002; Ghildiyal and Zamore 2009; Holoch and Moazed 2015). Cada classe possui uma distribuição de tamanho e preferência de nucleotídeos diferentes. Além disso, essas classes são relacionadas entre si e cooperam para a regulação de genes e proteção do genoma contra modificações externas, como por exemplo, elementos transponíveis e vírus (Ghildiyal and Zamore 2009). A via de RNAi nunca foi identificada em leveduras *Saccharomyces cerevisiae*, contudo, nos fungos multicelulares já foi comprovada a presença do sistema de RNAi (Dang et al. 2011a). Em fungos filamentosos, dois mecanismos são conhecidos e denominados de *quelling* e de Silenciamento Meiótico Pareado por DNA (*Meiotic Silencing by Unpaired DNA* – MSUD) (Dang et al. 2011a; Billmyre et al. 2013).

O mecanismo de *quelling* foi descrito primeiramente em *N. crassa* ocorrendo no tecido somático (micélio), onde há o reconhecimento de sequências repetitivas e

homólogas que desencadeiam o silenciamento gênico em nível pós-transcricional de todos os genes homólogos do hospedeiro (Dang et al. 2011a; Billmyre et al. 2013). Já o mecanismo de MSUD ocorre apenas durante o ciclo sexual, e é ativado quando ocorre a fusão de dois núcleos haploides distintos para a formação de um núcleo diploide durante o processo de divisão celular. O MSUD age na prófase durante a primeira meiose na região entre os dois cromossomos parentais contendo DNA não-pareado. O DNA não-pareado pode ser causado por deleções ou inserções em uma fita parental. Durante essa etapa, ocorre o silenciamento de todas as sequências homólogas similares a estes elementos no genoma até o término do ciclo sexual. Acredita-se que o MSUD tenha importância na inibição da replicação de elementos móveis durante o ciclo sexual (Dang et al. 2011a).

Fungos estão entre os grupos de eucariotos mais abundantes e diversos do planeta. No entanto, grande parte da diversidade relações bióticas e habitats desses organismos ainda precisam ser descobertos e caracterizados (Hawksworth and Lücking 2017). A ordem Hypocreales (*Hypocreomycetidae*, *Sordariomycetes*, *Pezizomycotina*, *Ascomycota*) compõe um grupo monofilético com mais de 2700 espécies de fungos, distribuídas em 240 gêneros e classificadas em 14 famílias: *Bionectriaceae*, *Calcarisporiaceae*, *Clavicipitaceae*, *Cocoonihabitaceae*, *Cordycipitaceae*, *Flammocladiellaceae*, *Hypocreaceae*, *Myrotheciomycetaceae*, *Nectriaceae*, Niessliaceae, *Ophiocordycipitaceae*, Sarocladiaceae, *Stachybotryaceae* e *Tilachlidiaceae* (Wijayawardene 2020). Alguns gêneros, porém, ainda são classificados como *incertae sedis* (Varshney et al. 2016). Esses fungos são ubíquos e ecologicamente diversos, com funções que variam de sapróbios a biotróficos e patógenos de outras espécies fúngicas (incluindo líquens), insetos, plantas e animais. Espécies com importância econômica, nas áreas de farmacologia e medicina, controle biológico e biotecnologia também são encontradas (Rehner and Samuels 1995; Varshney et al. 2016), sendo o gênero *Trichoderma* (família Hypocreaceae) um dos exemplos de organismos com grande potencial biotecnológico.

As espécies desse gênero apresentam capacidade de adaptação em todos os ecossistemas, sendo encontradas em ambientes aquáticos (marinhos e continentais) e terrestres. Geralmente, a maioria das espécies é encontrada em solos, degradando matéria orgânica em materiais ricos em celulose como madeira em decomposição (como saprofíticos), ou no interior de tecidos de plantas, como endofíticos (Samuels 2006; Mukherjee et al. 2013). A adaptabilidade dessas espécies permitiu o seu estabelecimento

e colonização no interior de raízes, caules e folhas de plantas, sendo considerados como "*True endophytes*", no qual algumas são conhecidas por proteger a planta de doenças e estresse abiótico (Mukherjee et al. 2012). A sua adaptabilidade está relacionada com a capacidade de produzir uma variedade de metabólitos secundários, que podem auxiliar no crescimento e desenvolvimento de plantas, inibição de micro-organismos, resposta à estresses bióticos e abióticos, biorremediação e defesa contra patógenos (Hermosa et al. 2014). Além disso, espécies de *Trichoderma* possuem a habilidade de parasitar outros fungos (micoparasitismo) (Druzhinina et al. 2011; Gruber and Seidl-Seiboth 2012). Por esta razão, têm sido amplamente utilizados na Agricultura como biopesticidas, fertilizantes e agentes de controle biológico (Benítez et al. 2004; Mukherjee et al. 2013; Schmoll et al. 2016). A espécie *Trichoderma harzianum,* por exemplo, é muito utilizada no setor agrícola como agente de controle biológico, e, apesar de apresentar várias aplicações, entretanto, as características moleculares e genéticas dessa espécie ainda são pouco exploradas (Prasad et al. 2002; Srivastava and Shahid 2014a).

Nessa tese, o objetivo principal de estudo foi investigar o funcionamento e caracterizar os mitogenomas da ordem Hypocreales, utilizando a espécie *T. harzianum* como modelo. Para isso, análises de genômica comparativa, transcritoma de RNA total e de pequenos RNAs foram realizadas. Os resultados finais encontrados são de suma importância para a compreensão do funcionamento da organela nas células de fungos filamentosos.

OBJETIVOS

Objetivo Geral

Caracterizar os genomas mitocondriais de fungos da ordem Hypocreales elucidando o papel de íntrons e *homing endonucleases* na expansão/retração desses genomas e avaliar os genes envolvidos na regulação da replicação, transcrição e no controle desses mitogenomas.

Objetivos Específicos

- Identificar as prováveis razões pela diversidade genômica de mitocôndrias de espécies da ordem Hypocreales a partir da estrutura dos mitogenomas disponíveis em bancos de dados;
- Estimar o tempo de divergência e a influência das regiões não-codificantes nos genomas mitocondriais;
- Avaliar o uso de genes mitocondriais como possíveis marcadores moleculares para identificação de espécies da ordem Hypocreales;
- Identificar genes de origem nuclear com provável função de regulação mitocondrial a partir de sequenciamento, montagem e anotação do genoma nuclear da espécie *Trichoderma harzianum*;
- Avaliar a transcrição dos genes identificados na regulação mitocondrial;
- Analisar a transcrição do mitogenoma da espécie *T. harzianum;*
- Investigar a presença de dsRNA de origem mitocondrial no fungo *T. harzianum*;
- Investigar outras vias envolvidas no controle da produção de RNAs derivados da mitocôndria de *T. harzianum*.

CAPÍTULO 1

Exploring the Relationships Among Divergence Time and Coding and Non-coding Elements in the Shaping of Fungal Mitochondrial Genomes

Este capítulo foi publicado na revista *Frontiers in Microbiology* (2020)

DOI: 10.3389/fmicb.2020.00765

CAPÍTULO 2

The potential of the mitochondrial gene *rns* as a molecular marker for Hypocreales (Fungi, Ascomycota)

Este capítulo foi submetido para publicação na revista *FEMS Microbiology Letters*

The potential of the mitochondrial gene *rns* as a molecular marker for Hypocreales (Fungi, Ascomycota)

Paula L. C. Fonseca[1], Daniel S. Araújo[1], Ruth B. de-Paula[2], Luiz M. R. Tomé[1], Leonardo de Oliveira Barbosa[3], Thairine Mendes-Pereira[1], Fernanda Badotti[4], Ana C. F. Soares[3], Eric R. G. R. Aguiar[5,6,*] and Aristóteles Góes-Neto[1,5*]

[1]Molecular and Computational Biology of Fungi Laboratory, Institute of Biological Sciences, Department of Microbiology, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
[2] Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, United States
[3] Centro de Ciências Agrárias, Ambientais e Biológicas, Universidade Federal do Recôncavo da Bahia, Cruz das Almas 44380-000, Bahia, Brazil.
[4] Department of Chemistry, Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, Brazil
[5]Program of Bioinformatics, Institute of Biological Sciences, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
[6]Department of Biological Sciences, Center of Biotechnology and Genetics, Universidade Estadual de Santa Cruz, Ilhéus, Brazil

*** Correspondence:** Eric R. G. R. Aguiar and Aristóteles Góes Neto ericgdp@gmail.com; arigoesneto@icb.ufmg.br

Abstract

Hypocreales order is one of the most diverse fungal taxonomic groups concerning ecological functions. Despite its ecological and biotechnological importance, Hypocreales species identification is still problematic, once morphological characters are commonly insufficient for species-level identification. Thus, nuclear molecular markers, such as the ITS region, have been adopted as a standard DNA barcode for rapid fungal identification, although their use has been insufficient for some taxonomic groups. Mitochondrial genes are conserved and found in higher abundance than nuclear genes, then they may contribute or even be an alternative in the identification of fungal species. Previous studies revealed that the minor ribosome subunit gene (*rns*) does not present introns and is highly conserved among Hypocreales species. The aim of this study was to investigate the potential of the *rns* gene as a molecular marker for the Hypocreales. Our strategy was based on phylogenetic and barcoding gap analyses using ITS region, calmodulin (CAL) and transcription elongation factor (*tef*) along with the *rns* gene. Our results indicated that the *rns* gene is an efficient DNA barcode for the identification of Hypocreales representatives, which in association with CAL gene reached higher resolution than ITS alone or combined for species identification.

Keywords: Molecular identification, Mitogenomes, *rns*, Hypocreales.

Introduction

The kingdom Fungi is considered as one of the most diverse groups of eukaryotic organisms on the planet. Estimates suggest the presence of approximately five million species, but only 10% of its richness has already been identified (Hawksworth 1991; Blackwell 2011; Hibbett et al. 2011; Bass and Richards 2011; Hawksworth and Lücking 2017). Hypocreales is one of the most important orders of the phylum Ascomycota, with approximately 2,700 species which display different morphologies, lifestyles and ecological functions (Varshney et al. 2016; Wijayawardene et al. 2018). These fungi are adapted to different environments, suggesting that multiple events, such as interkingdom host jumping, occurred during their evolutionary history (Torres and White 2009). Several studies have investigated gene functions related to fungal development, pathogenicity, production of metabolites and enzymes as well as the study of the species fitness (Rehner and Samuels 1995; Varshney et al. 2016). Nevertheless, despite all research advances, and the variability of ecological functions presented by these species, their identification is still a problem, due to the existence of groups with similar and undistinguishable morphological characteristics and also the presence of cryptic species (O'Donnell et al. 2004; Starkey et al. 2007).

The use of information contained in specific DNA sequences has been one of the strategies most applied in the identification of fungal species (Kulik et al. 2020). In 2003, the DNA Barcoding Initiative began to search for an universal molecular marker for the identification of all organisms. A DNA barcode is a sequence of 400-800 bp universally present in a group of species, which displays enough variation for species distinction and can be rapidly amplified (Gao and Zhang 2013; Crous et al. 2016). In comparison with the methods based on morphological characters, the amplification of genic regions (molecular strategy) is advantageous because they are less laborious, faster, more reliable and highly sensitive (Kulik et al. 2020).

The first DNA barcode proposed was the mitochondrial cytochrome c oxidase subunit 1 (*cox1* - COI) for animals (Hebert et al. 2003). In the case of fungi, the *cox1* gene did not provide enough molecular resolution, which can be explained by its common fragmentation with introns that reduce the efficiency of amplification (Seifert et al. 2007; Seifert 2009). Currently, the Internal Transcribed Spacer (ITS) region is recommended as the primary DNA barcode for the identification of fungal taxa. This region comprises approximately 600 bp and was selected because it is easy to obtain (simplicity), highly conserved (universality), and displays high interspecific variation and very low

intraspecific variation (Schoch et al. 2012; Badotti et al. 2017). Despite of its advantages, the use of ITS for the identification of some groups of fungi has been insufficient due to its low power of discrimination of close and cryptic species, such as *Trichoderma*, *Fusarium*, and other genera (Seifert 2009; Chaverri et al. 2015).

Then, other nuclear genes have been proposed for the identification of fungal species, such as translation elongation factor (*tef*) (Kristensen et al. 2005; Rai et al. 2016) and the calmodulin gene (CAL) (Mulã et al. 2004). Despite exhibiting a better resolution than the ITS region, difficulties in amplifying these genes in the laboratory has limited their use (Schoch et al. 2012). Mitochondrial genes can be used in combination with nuclear genes to identify species, once the mitochondrial genome (mitogenome) has numerous copies inside the cell and its genic content tend to be conserved (Aguileta et al. 2014). Some studies showed that mitochondrial genes can be used to identify species of some groups of fungi (Vialle et al. 2009; Kulik et al. 2015; Avin et al. 2017). Our research group has previously estimated that the minor ribosome subunit gene (*rnS*) has a conserved size within the order Hypocreales and does not have introns (Fonseca et al. 2020), indicating that this gene could be a candidate molecular marker for identifying species of this order. In this study, we evaluated the potential of *rns* gene in discriminating species of the order Hypocreales by using phylogenetic analyses. The assessment of a set of isolates of the genus *Trichoderma*, widely recognized for containing cryptic species, revealed that *rns* gene has similar robustness in comparison to sequences from ITS, *tef* and CAL regions, and in some cases outperformed these other markers. Therefore, our data suggests that *rns* can be an important candidate for the molecular identification of cryptic species from the Hypocreales order.

Material and Methods

Hypocreales species identification and phylogeny-based analyses

A total of 35 reference mitogenomes of species from the Hypocreales were retrieved from NCBI Organelle Genome Resources database (https://www.ncbi.nlm.nih.gov/genome/organelle/). Identification of fungal species and accession numbers are provided on Supplementary Table S1. The mitochondrial genomes were selected and annotated following the steps published by (Fonseca et al. 2020). Briefly, we annotated all mitogenomes using the software MITOS2 (http://mitos2.bioinf.uni-leipzig.de/index.py), RNAweasel and MFannot (http://megasun.bch.umontreal.ca/RNAweasel/). Mitochondrial gene sequences were

extracted using an *in-house* python program developed for the purpose of this study (Supplementary text S1).

For mitochondrial-based phylogenies, the ribosomal gene *rns* was used and for nuclear-based phylogenies, the following biomarkers were used: Internal Transcribed Spacer (ITS), Translational Elongation Factor (*tef*), and Calmodulin (CAL).

For each phylogeny, sequence alignment was performed using the MAFFT program (Katoh et al. 2019). The selection of the best evolutionary models and the creation of maximum likelihood trees were done using the MEGA v.7 program, adopting 1000 replicates of bootstrap (Stecher et al. 2020). *Colletotrichum graminicola* was used as outgroup. The trees were edited and rooted using Geneious v. 2021.0.3 software (https://www.geneious.com).

DNA isolation, amplification and sequencing

Isolates from the species *Trichoderma harzianum*, *T. longibrachiatum*, *T. virens*, *T. asperellum*, *T. koningii*, *T. spirale*, *T. lentiforme* and *T. saturnisporum* were used for total DNA extraction, amplification and sequencing. For each isolate, the mycelium was collected and placed into 2.0 mL tube containing 1 mL of extraction buffer from FastDNA SPIN Kit (MP Biomedicals, CA, USA, 2020). The DNA extraction was carried out according to the manufacturer instructions. Each DNA sample was diluted to 1:100 for amplification. The ITS (˜650 bp), CAL (˜500 bp), and *tef* (˜550 bp) regions were amplified using the PCR protocol described by (Vaz et al. 2018). Primers were designed to amplify the *rns* gene (˜650 bp) using the program Primer3 from Geneious v 2021.0.3 software (https://www.geneious.com). The primer sequences are available in the Supplementary Table S2.

Successfully amplified PCR products were purified using an ethanol/ethylenediaminetetraacetic acid 125 mM precipitation protocol and the sequencing reactions were performed on an ABI 3130 automated sequencer (Applied Biosystems, Life Technologies Q7, CA, USA). Sequences were edited using Geneious (version 9.1.6) and then used to build new phylogenies using the same pipeline described in the section "*Hypocreales species identification and phylogeny-based analyses*". All sequences were deposited at NCBI database. The accession number are available in Supplementary Table S3.

Barcode Gap analyses

Barcode gap analyses were performed to evaluate the potential of the genes (ITS, *CAL, tef* and *rns*) to identify *Trichoderma* species according to (Badotti et al. 2017). Distance matrices were generated using *p* distance in the software Geneious Prime v 2021.0.3 (www.geneious.com). The intra and interspecific distances were separated in a matrix and jitter plots were generated using the package *ggplot2* in R program v 1.1.463 (Wickham 2016).

Results and Discussion

The use of *rns* for the molecular identification of Hypocreales species

The similarity of morphological characters among some groups of fungi makes their differentiation a challenging process (Xu 2016), then the use of molecular biomarkers has been fundamental and widely used for the correct identification of many species. In 2012, Schoch and colleagues proposed the ITS region as an universal molecular barcode for fungal identification (Schoch et al. 2012). Since then, ITS has been the most used molecular biomarker for fungal identification, however, this genomic region cannot be considered universal, because it does not have sufficient resolution for the separation of all fungal species, such as cryptic species (Badotti et al. 2018). Considering this limitation, other molecular markers have been proposed for the identification of fungal species. According to the review published by Badotti et al. (2018) data mining searches in specialized literature and sequence public databases have pointed out which nuclear genes are currently used as alternatives to the ITS region. It was found that the *tef*, CAL, and tubulin genes, among others, are generally used in association with the ITS for the identification and phylogenetic analysis of fungal species (Badotti et al. 2018). Based on these data, we selected the *tef*, CAL and ITS region to compare with *rns* gene (Chaverri et al. 2015; Badotti et al. 2018).

The order Hypocreales is composed by 14 families: Bionectriaceae, Calcarisporiaceae, Clavicipitaceae, Cocoonihabitaceae, Cordycipitaceae, Flammocladiellaceae, Hypocreaceae, Myrotheciomycetaceae, Nectriaceae, Niessliaceae, Ophiocordycipitaceae, Sarocladiaceae, Stachybotryaceae and Tilachlidiaceae (Wijayawardene 2020). The identification of species from this order is still an arduous task, once several different nuclear markers and morphological characters must be used for the classification at species level (Torbati et al. 2019; Sun et al. 2019). Mitochondrial genes have shown promising results to identify fungal species, with the advantage that a single cell can have several copies of the mitogenome, which enhances the chances of

successful gene amplification (Basse 2010; Kulik et al. 2020). In public databases, it is possible to find mitochondrial genomes of six out of 14 families belonging to Hypocreales (Fonseca et al. 2020). Recent study estimated that of all the mitochondrial genes present in species of the Hypocreales order the only one that presents a single exon (absence of introns) is the ribosomal gene *rns*. This gene has a size of approximately 1,500 bp, with a conserved portion among all the 35 mitogenomes from Hypocreales analyzed (Fonseca et al. 2020) . For this reason, we investigated whether the *rns* gene could be used for the taxonomic and phylogenetic classification of the species belonging to this fungal order. Figure 1a shows the *rns* phylogeny by the maximum likelihood method for the 35 species evaluated in the study published by Fonseca et al. (2020). We noticed that all the six sampled families were clustered with high statistical support (bootstrap value > 50%), suggesting that this is a promising region to be used for the identification of species from these families. Additionally, the genera *Trichoderma* (Hypocreaceae) and *Fusarium* (Nectriaceae) are recognized for containing cryptic species, whose identification is a hard task to be performed (Kulik et al. 2020). In the phylogeny of the *rns* gene, we observed that all the species tested for these two genera were well separated compared to the other regions tested, indicating that this gene could also be used to differentiate species at genera level.

In order to compare the results of the phylogenetic analysis for the *rns* gene, we searched public databases for sequences of other well-stablished molecular markers used to identify species of Hypocreales. The phylogenetic trees for the ITS, *tef*, and CAL genes are shown in Figure 1b-1d, respectively. Only the ITS phylogeny was able to separate the species among the corresponding families. In the phylogeny of *tef* gene, for example, the species *Trichoderma gamsii* (Hypocreaceae) and *Hirsutella rhossiliensis* (Ophiocordycipitaceae), *Lecanicillium muscarium* and *Metacordyceps chlamidosporia* (Clavicipitaceae) and *Beauveria pseudobassiana* (Cordycipitaceae) were grouped into a single cluster. In the phylogeny of the gene CAL, we observed that species of the family Hypocreaceae grouped with *Beauveria bassiana* (Cordycipitaceae), and species of the family Clavicipitaceae grouped with the species *Hypomyces auratius* (Hypocreaceae) and the species *M. chlamydosporia* (Clavicipitaceae) grouped with *C. graminicola* (outgroup). Despite the inconsistencies found, it was not possible to compare the phylogenetic relationships of the 35 species analyzed in the *rns* phylogeny due to the absence of sequences from the other regions used as molecular markers. It was possible to recover sequences from three regions only for 12 species. Overall, our data suggested

that the *rns* gene can be considered a biomarker for the identification of species of the order Hypocreales and its usage can be used together with ITS region.

In order to confirm that the *rns* gene could be used for other species within Hypocreales, we designed a pair of primers capable of amplifying a more conserved region of the gene for the identification of 18 *Trichoderma* isolates. These isolates were also identified by amplifying and sequencing the ITS, *tef* and CAL regions. Figure 2 displays the grouping of *Trichoderma* isolates by region. Using *rns* (Fig.2a), all the species were grouped into the correct clades, except for the *T. lentiforme* isolate (TCS1) that grouped with different isolates from the harzianum/virens clade, indicating that this region cannot separate the species from the clade. Figure 2b (ITS region), shows the separation of the harzianum/virens clade into two different clusters, and in Fig. 2c (*tef* gene) the species *T. virens* (harzianum/virens clade) grouped with the species *T. saturnisporum* (longibrachiatum clade) and *T. koningiopsis* (*Trichoderma* clade) (Kubicek et al. 2019). Moreover, in Fig. 2d, all the isolates of the same species or clade were grouped correctly, indicating that the CAL gene was the best gene for classifying the *Trichoderma* isolates tested in this study. Nevertheless, the *rns* gene and the ITS region showed better results than the *tef* gene and can also be used in *Trichoderma* identification studies.

Since CAL and *rns* presented the best results, we investigated whether the combination of the *rns* and CAL genes (Fig. 2e) or *rns* and ITS region (Fig. 2f) would increase the resolution for species classification, considering that any of them achieved optimal results. Phylogeny based on information from the *rns* and CAL genes succeeded in solving the clustering of *T. lentiforme* with *T. harzianum* isolates and grouped *T. koningiopsis* near *T. asperellum* and *T. saturnisporum*, indicating that this combination is effective in classifying *Trichoderma* species, unlike the *rns* and ITS combination, which did not result in the correct identification of all species. In parallel, we evaluated the combination of ITS and CAL for the classification of species (Supplementary Fig. 1a) and compared it with the other two combinations (*rns* and CAL, *rns* and ITS), which highlighted that *T. saturnisporum* was closer to the species from the harzianum clade than to the ones from asperellum and longibrachiatum clades and it is known that these species are more related among themselves than with the harzianum clade (Hassan et al. 2019). In addition, the phylogeny based on *rns*, CAL and ITS (Supplementary Fig. 1b) showed that the isolates TCS6 and TCS15 are closely related and that the species *T. saturnisporum* (R75 isolate) clustered together with asperellum and longibrachiatum

clades, indicating that the combination of the three regions is a good option for *Trichoderma* identification. While the combination of *rns* and CAL is the best option for the identification of species from the *Trichoderma* genera, the ITS region presented better clustering results for Hypocreales. For this reason, we suggest the use of *rns* and CAL to identify species of *Trichoderma,* and *rns* and ITS region to identify other species from the Hypocreales.

Effectiveness of *rns* and the other markers in the genera *Trichoderma*

The DNA barcode consists in the use of a short and standardized segment of DNA for fast and accurate identification of species, and the marker accepted as fungal barcode is the ITS region (Badotti et al. 2017). One of the methods used to evaluate the effectiveness of a marker is the barcode gap analysis, in which it is evaluated whether the genetic variation between the analyzed species exceeds the variation existing within these same species for the selected DNA segment (Hebert et al. 2003).

In our study, in order to assess whether the *rns* gene could be used as a good molecular marker, we compared the result of the barcode gap analyses from the molecular markers and the *rns* gene used to identify 20 *Trichoderma* isolates. All the four markers (ITS, *tef*, CAL and *rns*) showed a clear barcode gap (non-overlapping intraspecific divergence with interspecific divergence), indicating that all the analyzed markers can be used to separate different *Trichoderma* species (Fig. 3). Despite having a good resolution for species separation, phylogenetic analysis has shown that the *tef* and CAL genes are not sufficient for the taxonomic classification of some families of Hypocreales and the *Trichoderma* genus. So far, a universal marker for all the groups of fungi has not been found (Begerow et al. 2010). In this study, we suggest that the *rns* gene can indeed be used as a biomarker and can be used together with the ITS region and other molecular markers to identify species of the *Trichoderma* genus and the Hypocreales order.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Conceived and designed experiments: Paula L.C. Fonseca, Daniel S. Araújo, Ruth B. de-Paula, Luiz M. R. Tomé, Leonardo de Oliveira Barbosa, Thairine Mendes-Pereira and Fernanda Badotti. Analyzed the data: Paula L.C. Fonseca, Ana C. F. Soares, Eric R.G.R.

Aguiar and Aristóteles Goés-Neto. Wrote the manuscript: Paula L.C. Fonseca, Daniel S. Araújo, Ruth B. de-Paula, Luiz M. R. Tomé, Leonardo de Oliveira Barbosa,Thairine Mendes-Pereira, Fernanda Badotti, Ana C. F. Soares, Eric R.G.R. Aguiar and Aristóteles Goés-Neto. All authors read and approved the final manuscript.

Funding

This work was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The funders had no role in study design, data collection and analyses, decision to publish or preparation of the manuscript. Aristóteles Góes-Neto receives and Ana Cristina F. Soares a research grant for scientific productivity from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil (no. 310764/2016-5 for A. Góes-Neto and no. 308771/2017-6 for A.C.F.Soares).

Acknowledgments

The authors would like to thank the Graduate Programs of Microbiology (http://www.microbiologia.icb.ufmg.br/pos/) and Bioinformatics (http://www.pgbioinfo.icb.ufmg.br/) of the Universidade Federal de Minas Gerais (UFMG).

References

Aguileta G, de Vienne DM, Ross ON *et al.* High Variability of Mitochondrial Gene Order among Fungi. *Genome Biology and Evolution* 2014;6:451–65.

Avin FA, Subha B, Tan Y-S *et al.* Escaping introns in COI through cDNA barcoding of mushrooms: *Pleurotus* as a test case. *Ecol Evol* 2017;7:6972–80.

Badotti F, Fonseca PLC, Tomé LMR *et al.* ITS and secondary biomarkers in fungi: review on the evolution of their use based on scientific publications. *Braz J Bot* 2018;41:471–9.

Badotti F, de Oliveira FS, Garcia CF *et al.* Effectiveness of ITS and sub-regions as DNA barcode markers for the identification of Basidiomycota (Fungi). *BMC Microbiol* 2017;17:42.

Bass D, Richards TA. Three reasons to re-evaluate fungal diversity 'on Earth and in the ocean.' *Fungal Biology Reviews* 2011;25:159–64.

Basse CW. Mitochondrial inheritance in fungi. *Current Opinion in Microbiology* 2010;13:712–9.

Begerow D, Nilsson H, Unterseher M *et al.* Current state and perspectives of fungal DNA barcoding and rapid identification procedures. *Appl Microbiol Biotechnol* 2010;87:99–108.

Blackwell M. The Fungi: 1, 2, 3 … 5.1 million species? *American Journal of Botany* 2011;98:426–38.

Chaverri P, Branco-Rocha F, Jaklitsch W *et al.* Systematics of the *Trichoderma harzianum* species complex and the re-identification of commercial biocontrol strains. *Mycologia* 2015;107:558–90.

Crous PW, Groenewald JZ, Slippers B *et al.* Global food and fibre security threatened by current inefficiencies in fungal identification. *Phil Trans R Soc B* 2016;371:20160024.

Fonseca PLC, Badotti F, De-Paula RB *et al.* Exploring the Relationship Among Divergence Time and Coding and Non-coding Elements in the Shaping of Fungal Mitochondrial Genomes. *Front Microbiol* 2020;11:765.

Gao R, Zhang G. Potential of DNA Barcoding for Detecting Quarantine Fungi. *Phytopathology*® 2013;103:1103–7.

Hassan MM, Farid MA, Gaber A. Rapid identification of Trichoderma koningiopsis and Trichoderma longibrachiatum using sequence-characterized amplified region markers. *Egypt J Biol Pest Control* 2019;29:13.

Hawksworth DL. The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycological Research* 1991;95:641–55.

Hawksworth, Lücking. Fungal Diversity Revisited: 2.2 to 3.8 Million Species. *The Fungal Kingdom*. American Society of Microbiology, 2017, 79–95.

Hebert PDN, Cywinska A, Ball SL *et al.* Biological identifications through DNA barcodes. *Proc R Soc Lond B* 2003;270:313–21.

Hibbett DS, Ohman A, Glotzer D *et al.* Progress in molecular and morphological taxon discovery in Fungi and options for formal classification of environmental sequences. *Fungal Biology Reviews* 2011;25:38–47.

Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* 2019;20:1160–6.

Kristensen R, Torp M, Kosiak B *et al.* Phylogeny and toxigenic potential is correlated in Fusarium species as revealed by partial translation elongation factor 1 alpha gene sequences. *Mycol Res* 2005;109:173–86.

Kubicek CP, Steindorff AS, Chenthamara K *et al.* Evolution and comparative genomics of the most common Trichoderma species. *BMC Genomics* 2019;20:485.

Kulik T, Bilska K, Żelechowski M. Promising Perspectives for Detection, Identification, and Quantification of Plant Pathogenic Fungi and Oomycetes through Targeting Mitochondrial DNA. *IJMS* 2020;21:2645.

Kulik T, Ostrowska A, Buśko M *et al.* Development of an FgMito assay: A highly sensitive mitochondrial based qPCR assay for quantification of Fusarium graminearum sensu stricto. *International Journal of Food Microbiology* 2015;210:16–23.

Mulã G, Susca A, Stea G *et al.* Specific detection of the toxigenic species *Fusarium proliferatum* and *F. oxysporum* from asparagus plants using primers based on calmodulin gene sequences. *FEMS Microbiology Letters* 2004;230:235–40.

O'Donnell K, Ward TJ, Geiser DM *et al.* Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the Fusarium graminearum clade. *Fungal Genetics and Biology* 2004;41:600–23.

Rai S, Kashyap PL, Kumar S *et al.* Identification, characterization and phylogenetic analysis of antifungal Trichoderma from tomato rhizosphere. *SpringerPlus* 2016;5:1939.

Rehner SA, Samuels GJ. Molecular systematics of the Hypocreales: a teleomorph gene phylogeny and the status of their anamorphs. *Can J Bot* 1995;73:816–23.

Schoch CL, Seifert KA, Huhndorf S *et al.* Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences* 2012;109:6241–6.

Seifert KA. Progress towards DNA barcoding of fungi. *Molecular Ecology Resources* 2009;9:83–9.

Seifert KA, Samson RA, deWaard JR *et al.* Prospects for fungus identification using CO1 DNA barcodes, with Penicillium as a test case. *Proceedings of the National Academy of Sciences* 2007;104:3901–6.

Starkey DE, Ward TJ, Aoki T *et al.* Global molecular surveillance reveals novel Fusarium head blight species and trichothecene toxin diversity. *Fungal Genetics and Biology* 2007;44:1191–204.

Stecher G, Tamura K, Kumar S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Russo C (ed.). *Molecular Biology and Evolution* 2020;37:1237–9.

Sun J-Z, Ge Q-Y, Zhu Z-B *et al.* Three dominating hypocrealean fungi of the 'white mold spots' on acrylic varnish coatings of the murals in a Koguryo tomb in China. *Phytotaxa* 2019;397:225.

Torbati M, Arzanlou M, Sandoval-Denis M *et al.* Multigene phylogeny reveals new fungicolous species in the Fusarium tricinctum species complex and novel hosts in the genus Fusarium from Iran. *Mycol Progress* 2019;18:119–33.

Torres MS, White JF. Clavicipitaceae: Free-Living and Saprotrophs to Plant Endophytes. *Encyclopedia of Microbiology*. Elsevier, 2009, 422–30.

Varshney D, Jaiswar A, Adholeya A *et al.* Phylogenetic analyses reveal molecular signatures associated with functional divergence among Subtilisin like Serine Proteases are linked to lifestyle transitions in Hypocreales. *BMC Evol Biol* 2016;16:220.

Vaz ABM, Fonseca PLC, Badotti F *et al.* A multiscale study of fungal endophyte communities of the foliar endosphere of native rubber trees in Eastern Amazon. *Sci Rep* 2018;8:16151.

Vialle A, Feau N, Allaire M *et al.* Evaluation of mitochondrial genes as DNA barcode for Basidiomycota. *Molecular Ecology Resources* 2009;9:99–113.

Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. 2016. Cham: Springer International Publishing : Imprint: Springer, 2016.

Wijayawardene N. Outline of Fungi and fungus-like taxa. *Mycosphere* 2020;11:1060–456.

Wijayawardene NN, Hyde KD, Lumbsch HT *et al.* Outline of Ascomycota: 2017. *Fungal Diversity* 2018;88:167–263.

Xu J. Fungal DNA barcoding. Adamowicz S (ed.). *Genome* 2016;59:913–32.

Fig. 1. Phylogenetic analyzes of species from the Hypocreales order. **(a)** Phylogeny of the *rns* gene. **(b)** ITS region. **(c)** *tef* gene and **(d)** CAL gene. The specie *Colletotrichum graminicola* (Glomerellales order) was used as an outgroup.

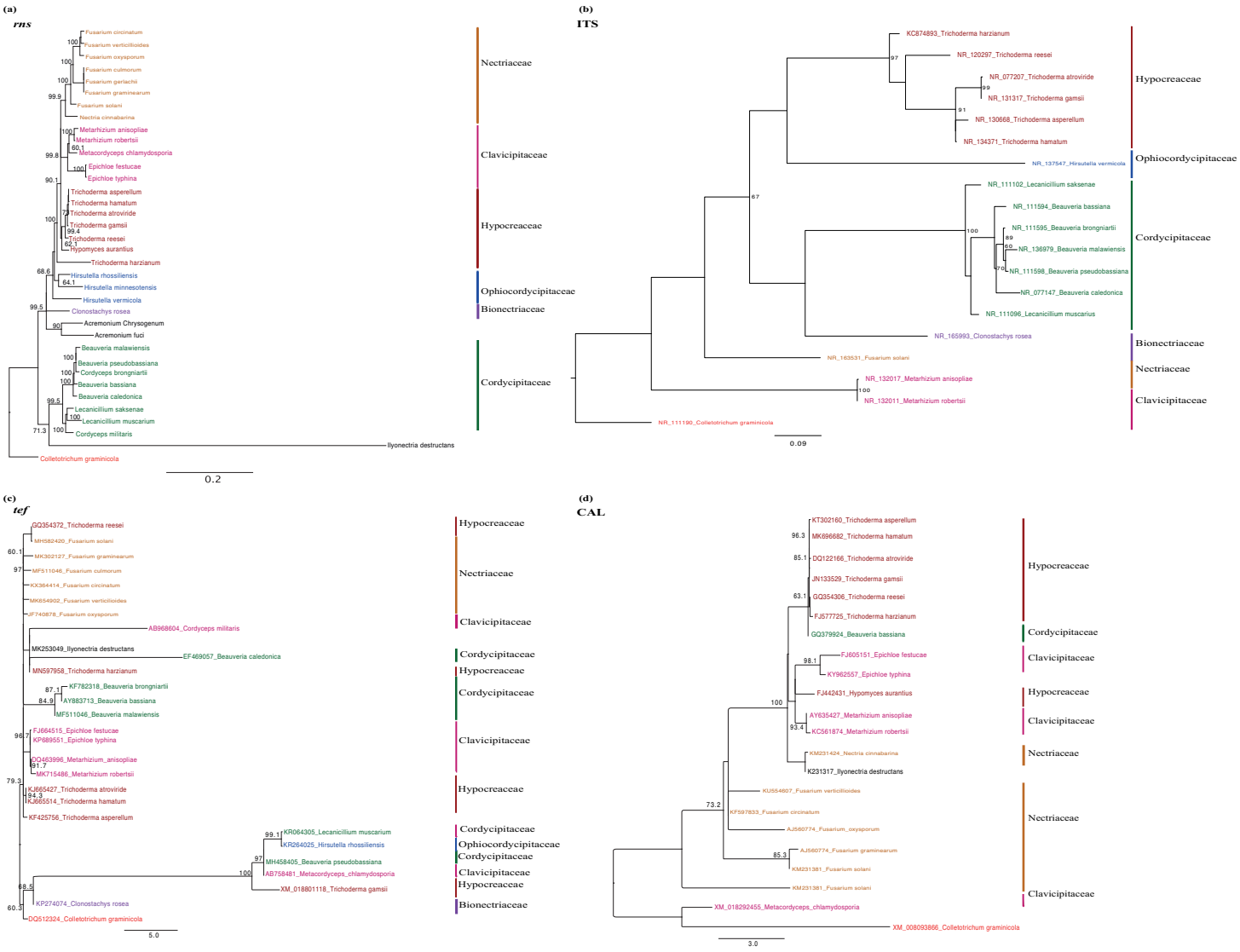


Fig. 2. Phylogenetic analyzes of species from the genus *Trichoderma*. **(a)** *rns* gene **(b)** ITS region **(c)** *tef* gene **(d)** CAL gene. **(e)** *rns* and CAL genes **(f)** *rns* gene and ITS region.

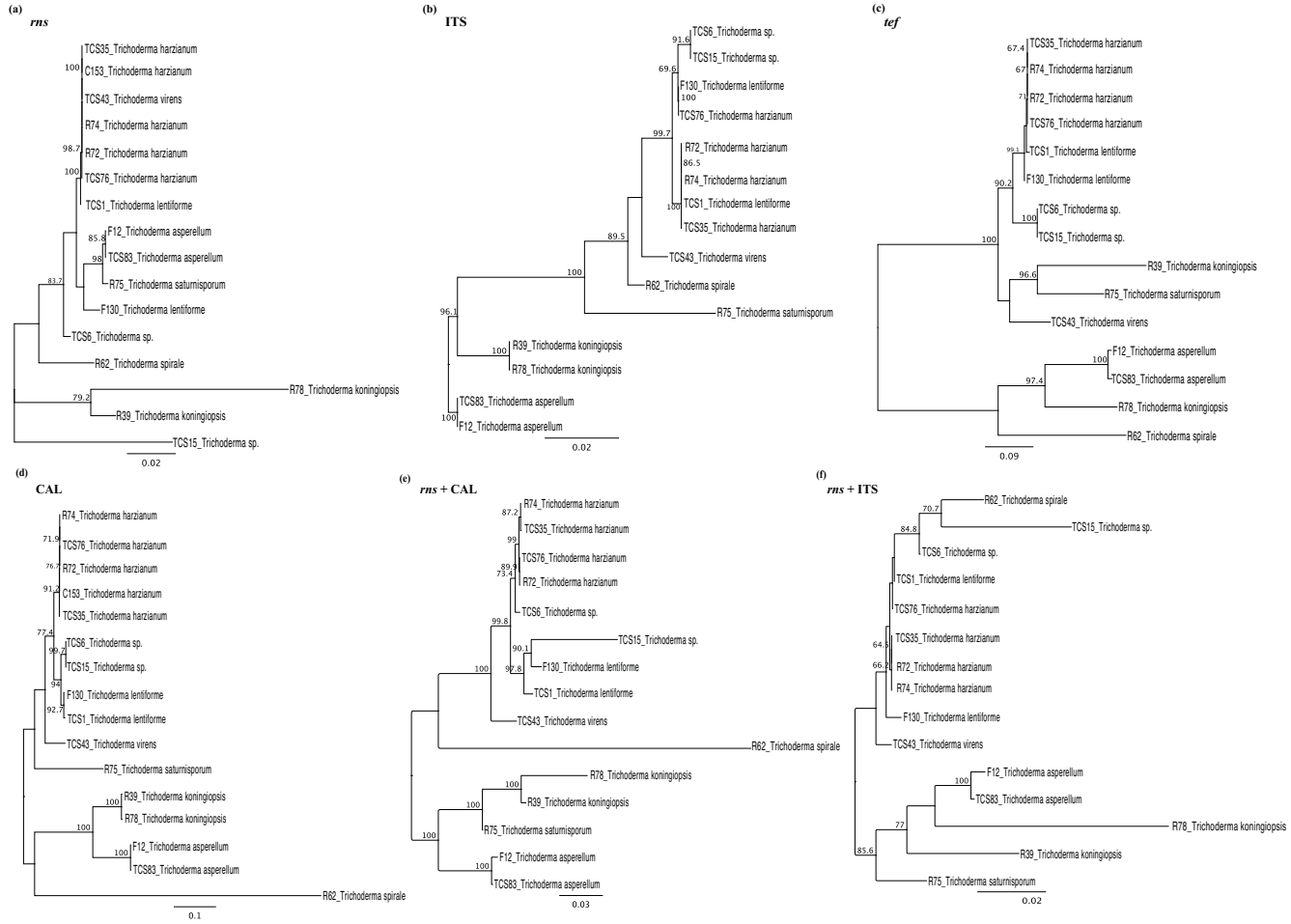
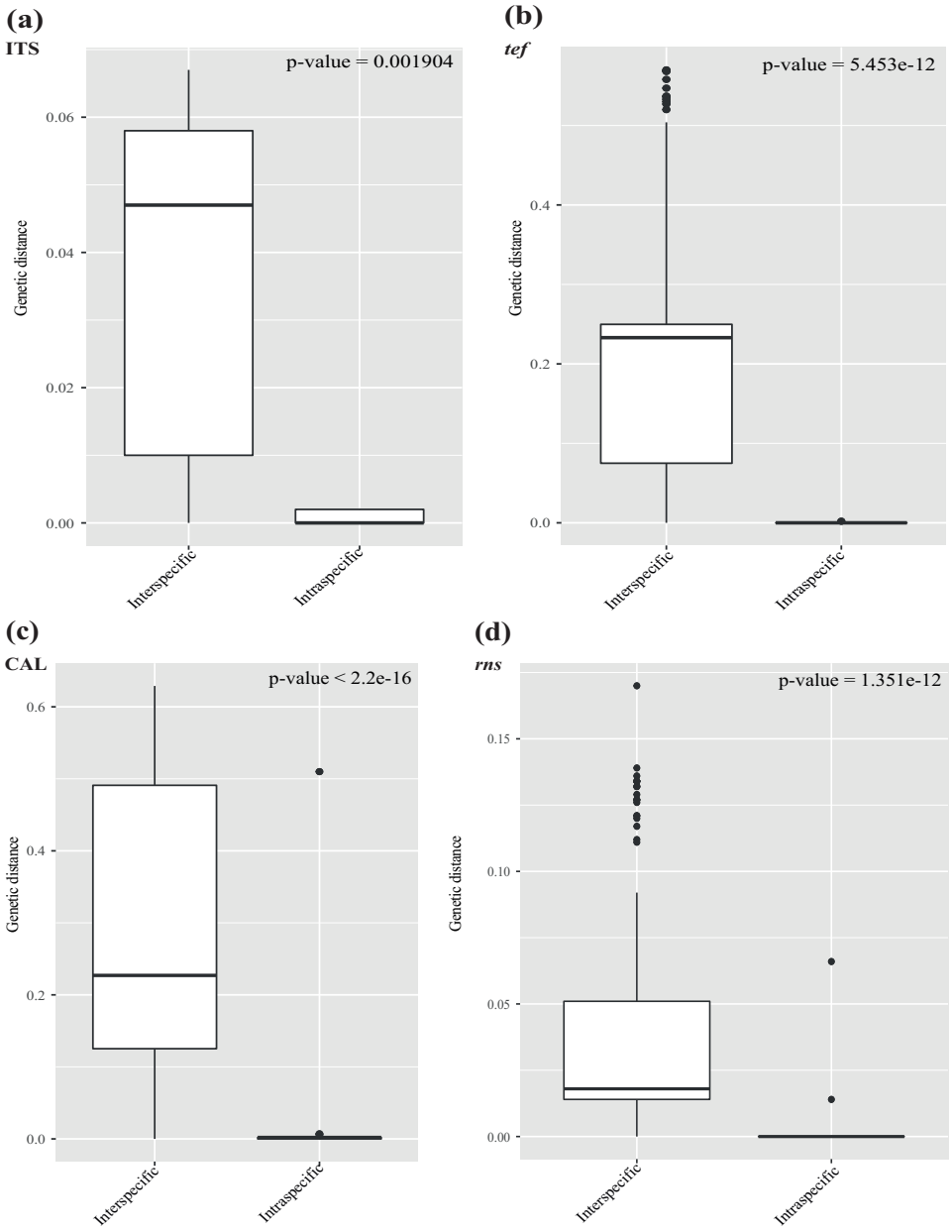


Fig. 3. Barcode gap analysis based on intraspecific and interspecific distances for isolates of the genera *Trichoderma*. (a) ITS region. (b) *tef* gene. (c) Analysis of CAL gene. (d) *rns* gene.



Supplementary Text 1. Script used in this work.

```
#This script is part of supplementary documents of " The potential of the mitochondrial gene rns as a molecular marker for Hypocreales (Fungi, Ascomycota)"

#This script uses a gff and fasta files of a target species to get the sequences of genes of interest (GOI)
#The gff file provide the start and end positions of each GOI.
#The output file is in fasta format, with the names (preceded by '>') and the sequence of genes of interest (GOI)

#*********************************************************************************#
#                    Run the code in Python 3+                    #
#*********************************************************************************#



import sys
import os.path
from os import path


def checkInputFiles():
   #Check if all the necessary files names are passed as arguments
   if (len(sys.argv)!=3 or sys.argv[1].find(".gff")==-1 or sys.argv[2].find(".fasta")==-1):
      print ("\nUsage:\npython getGeneSeqGff.py [file_path_name.gff] [file_path_name.fasta]\n\n")
      sys.exit(0)

   gff_file_name=sys.argv[1]
   fasta_file_name=sys.argv[2]


   #Check if path/files exists
   if (not (path.exists(gff_file_name) or path.exists(fasta_file_name))):
      print("\nOne or more files not found! Check the path and file names.\n")
      exit(0)

   return gff_file_name, fasta_file_name

#Reads whole sequence from input fasta file
def readFasta(fasta_file):
   #Position 0 of whole_genome will not be used
   whole_genome=" "
   for line in fasta_file:
      if (line.find(">")==-1):
         whole_genome=whole_genome+line.strip()
   fasta_file.close()
   return whole_genome
```

```
#This function check if the gene_name is present in the whitelist array, case true, it
saves the name and sequence of the gene
def search_genes(gene_name,gene_sequence, output_file, GOI):
    for GOI in GOI:
        if (gene_name.startswith(GOI)):
            output_file.write(">"+gene_name+"\n"+gene_sequence+"\n\n")
            break

#Read the gff file to extract data.
#The gff file contains 1 gene per row with several values ordered by 'tab'. Its straight
forward to get the name and positions of a single gene
#and retrieve th sequence from the whole_genome
def readGffSelGenes(GOI, gff_file, output_file, whole_genome):
    for line in gff_file:
        values=line.split("\t")
        #Start position is at index 3
        start_gene_position=values[3]
        #End position is at index 4
        end_gene_position=values[4]
        #Name is at index 8
        gene_name=values[8][values[8].find("Name=")+5:].strip()
        #Get gene sequence

gene_sequence=whole_genome[int(start_gene_position):int(end_gene_position)+1]
        #Verify if it is on gene_whitelist to save in output file
        search_genes(gene_name,gene_sequence,output_file,GOI)

def main():


GOI={"rrnL","rps3","nad2","nad3","atp9","cox2","nad4l","nad5","cob","cox1","nad1",
"nad4","atp8","atp6","rrnS","cox3","nad6"}

    gff_file_name, fasta_file_name=checkInputFiles()

    gff_file=open(gff_file_name,'r')
    fasta_file=open(fasta_file_name,'r')

    whole_genome=readFasta(fasta_file)

    #Get ID specie from gff file name
    #The strip will remove '.\' that appear on console in Windows 10 before path\filename
    if (os.name=="nt"):
        gff_file_name=gff_file_name.strip(".\\")

    output_file_name=gff_file_name[0:gff_file_name.find(".")]+"_GOI.fasta"
    #Open output file with '_GOI.fasta' extension
    output_file=open(output_file_name,'w')
```

```
    readGffSelGenes(GOI,gff_file,output_file, whole_genome)

    gff_file.close()
    output_file.close()

print("\n\n_____________________________________________________________________")
    print("\nResults saved in: "+output_file_name)

print("_____________________________________________________________________\n\n\n"
)

if __name__ == '__main__':
    main()
```

Supplementary Table 1. Accession number for the mitochondrial genomes and other molecular regions used in this study.

Species	Mitochondrial access number	ITS	TEF	CAL
Acremonium chrysogenum	KF757229			
Acremonium fuci	NC_029851			
Beauveria bassiana	NC_010652	NR_11594	AY883713	GQ379924
Beauveria caledonica	NC_030636	NR_077147	EF469057	
Beauveria malawiensis	NC_030635	NR_136979	DQ376246	
Beauveria pseudobassiana	NC_022708	NR_111598	MH458405	
Clonostachys rosea	NC_036667	NR_165993	KP274074	
Cordyceps brongniartii	NC_011194	NR_111595	KF782318	
Cordyceps militaris	NC_022834		AB968604	
Epichloe festucae	NC_032064		FJ664515	FJ605151
Epichloe typhina	NC_032063		KP689551	KY962557
Fusarium circinatum	NC_022681	NR_120263	KX364414	KF597833
Fusarium culmorum	NC_026993		MF511046	LR215910
Fusarium gerlachii	KM486533			
Fusarium graminearum	KP966551		MK302127	HQ412343
Fusarium oxysporum	AY945289		JF740878	AJ560774
Fusarium solani	NC_016680	NR_163531	MH582420	KM231381
Gibberella moniliformis	NC_016687		MK654902	KU554607
Hirsutella minnesotensis	NC_027660			
Hirsutella rhossiliensis	NC_030164		KR264025	
Hirsutella vermicola	NC_036610			NR_137547
Hypomyces aurantius	NC_030206		FN868743	FJ442431
Ilyonectria destructans	NC_030340		MK253049	KM231317
Lecanicillium muscarium	NC_004514	NR_111096	KR064305	
Lecanicillium saksenae	NC_028330	NR_111102		
Metacordyceps chlamydosporia	NC_022835		AB758481	XM_018292455
Metarhizium anisopliae	NC_008068	NR_132017	DQ463996	AY635427
Metarhizium robertsii	JELW01000367	NR_132011	MK715486	KC561874
Nectria cinnabarina	NC_030252			KM231424

Trichoderma asperellum	KR952346	NR_130668	KF425756	KT302160
Trichoderma atroviride	JGI1185329	NR_077207	KJ665427	DQ122166
Trichoderma gamsii	NC_030218	NR_131317	XM_018801118	JN133529
Trichoderma hamatum	NC_036144	NR_134371	KJ665514	MK696682
Trichoderma harzianum	MT263519	KC874893	MN597958	FJ577725
Trichoderma reesei	NC_003388	NR_120297	GQ354372	GQ354306
Colletotrichum graminicola	NW_007361658	NR_111190	DQ512324	XM_008093866

Supplementary Table S2: Primer sequences used in this study.

Molecular region	Primer Forward (5`-3`)	Primer Reverse (5`-3`)
ITS	ITS5 - GGAAGTAAAAGTCGTAACAAGG	ITS4 - TCCTCCGCTTATTGATATGC
tef	EF700F - TCTACCAGTGCGGTGGTA	TEF1R - GCCATCCTTCGAACCAGC
CAL	CAL228F - GAGTTCAAGGAGGCCTTCTCCC	CAL737R - CATCTTTCTGGCCATCATGG
rns	rnsF - GAGAAGTGGCAAGTCCTATT	rnsR - AACCACAGTTTCACAACATT

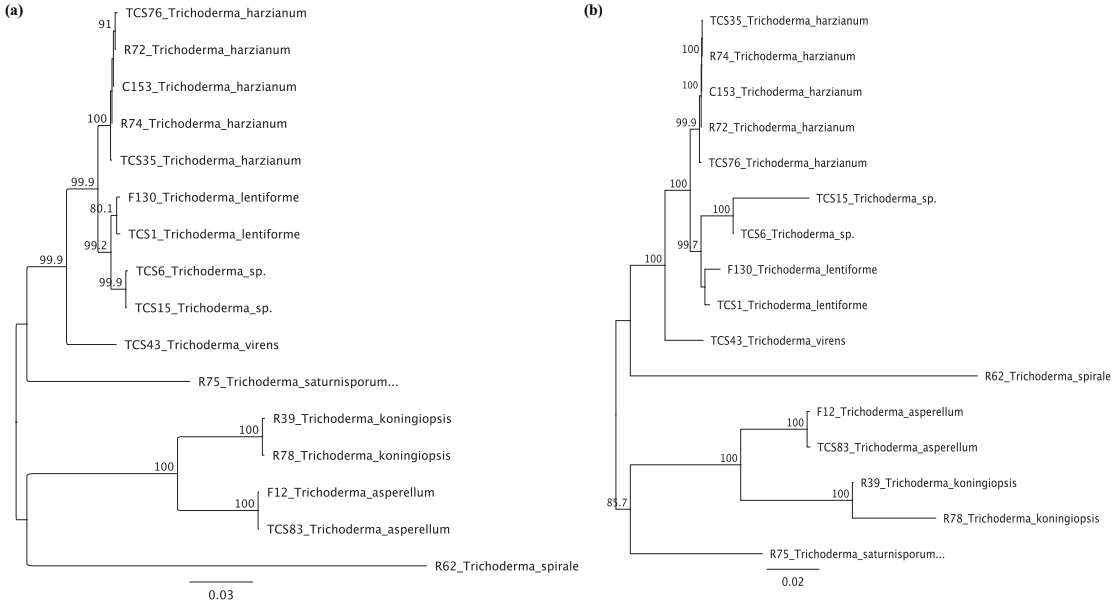
Suppelementary Table S3: Isolates identification and accession number

Gene/Region	Contig	Specie	BLAST result			Closest Access number
			Identidade	Cover	*e-value*	
TC83 Isolate						
ITS	Yes	*Trichoderma asperellum*	100	100	0.0	KU059966
Calmodulina	Yes	*Trichoderma asperellum*	99.11	94	0.0	FJ442414
TEF	Yes	*Trichoderma asperellum*	99.5	100	0.0	EF185998
R78 Isolate						
ITS	Yes	*Trichoderma koningiopsis*	100	100	0.0	MK396898
Calmodulina	Yes	*Trichoderma koningiopsis*	100	99.61	0.0	FJ467647
TEF	Yes	*Trichoderma koningiopsis*	100	100	0.0	DQ367692
R62 Isolate						
ITS	Yes	*Trichoderma spirale*	100	99	0.0	NR_077177
Calmodulina	Yes	*Trichoderma spirale*	100	92	0.0	FJ436153
TEF	Reverse	*Trichoderma spirale*	97.56	100	0.0	FJ442389
TCS15 Isolate						
ITS	Yes	*Trichoderma auroviride*	100	0	100	AF194013
Calmodulina	Yes	*Trichoderma lentiforme*	96.36	0	98	FJ442306
TEF	Yes	*Trichoderma camerunense*	99.07	0	100	AF348107
TCS76 Isolate						
ITS	Yes	*Trichoderma harzianum*	100	0	99	MF782822
Calmodulina	Yes	*Trichoderma harzianum*	99.78	0	96	FJ442338
TEF	Yes	*Trichoderma harzianum*	99.62	0	100	AY605774
TCS43 Isolate						
ITS	Yes	*Trichoderma virens*	100	0	100	MH624149
Calmodulina	Yes	*Trichoderma virens*	97.57	0	96	FJ442405
TEF	Yes	*Trichoderma virens*	99.81	0	97	EU280060
TCS35 Isolate						
ITS	Yes	*Trichoderma harzianum*	100	0	100	KP009228

Calmodulina	Yes	*Trichoderma harzianum*	99.78	0	97	FJ442338
TEF	Yes	*Trichoderma harzianum*	98.99	0	100	AY605768
TCS6 Isolate						
ITS	Yes	*Trichoderma lixii*	99.84	0	99	HQ608121
Calmodulina	Yes	*Trichoderma lentiforme*	96.37	0	99	FJ442306
TEF	Yes	*Trichoderma camerunense*	99.05	0	100	AF348107
TCS 1 Isolate						
ITS	Yes	*Trichoderma lentiforme*	99.67	0	100	AF443913
Calmodulina	Yes	*Trichoderma lentiforme*	99.78	0	100	FJ442287
TEF	Yes	*Trichoderma lentiforme*	99.35	0	100	FJ463309
R75 Isolate						
ITS	Yes	*Trichoderma saturnisporum*	100	0	100	NR_103704
Calmodulina	Yes	*Trichoderma saturnisporum*	98.50	0	40	JN388898
TEF	Yes	*Trichoderma saturnisporum*	97.35	0	99	KJ713203
R74 Isolate						
ITS	Yes	*Trichoderma harzianum*	100	0	100	MG490820
Calmodulina	Yes	*Trichoderma harzianum*	99.78	0	98	FJ442338
TEF	Yes	*Trichoderma harzianum*	99.57	0	99	AY605775
R72 Isolate						
ITS	Yes	*Trichoderma harzianum*	100	0	100	MG490811
Calmodulina	Yes	*Trichoderma harzianum*	99.78	0	97	FJ442338
TEF	Yes	*Trichoderma harzianum*	98.62	0	100	AY605768
R39 Isolate						
ITS	Yes	*Trichoderma koningiopsis*	100	0	100	MH624141
Calmodulina	Yes	*Trichoderma koningiopsis*	99.79	0	100	DQ367669
TEF	Yes	*Trichoderma koningiopsis*	99.46	0	100	FJ467647
F130 Isolate						
ITS	Yes	*Trichoderma lentiforme*	100	0	100	AF443913

Calmodulina	Yes	*Trichoderma lentiforme*	99.07	0	100	FJ463356
TEF	Yes	*Trichoderma lentiforme*	99.78	0	97	FJ442306
F12 Isolate						
ITS	Yes	*Trichoderma asperellum*	100	0	100	KU059966
Calmodulina	Yes	*Trichoderma asperellum*	99.76	0	100	EU248632
TEF	Reverse	*Trichoderma asperellum*	99.12	0	92	DQ122170
C153 Isolate						
ITS	Yes	*Trichoderma harzianum*	100	0	100	KP009228
Calmodulina	Yes	*Trichoderma harzianum*	99.52	0	100	KP757745
TEF	Yes	*Trichoderma harzianum*	99.78	0	100	FJ442338

Supplementary Fig.1. Phylogenetic analyzes from the genus *Trichoderma* using the ITS region and CAL gene concatenated **(a)** and using the combination *rns*, CAL and ITS **(b)**.



CAPÍTULO 3

Core components for the control of the replication and transcription of mitochondrial genome in *Trichoderma harzianum* (Hypocreales) mitogenome

Core components for the control of the replication and transcription *Trichoderma harzianum* (Hypocreales) mitogenome

Paula L. C. Fonseca[1], Luiz M. R. Tomé[1], Vasco A. Azevedo[2], Bertram Brenig[3], Eric R. G. R. Aguiar[2,4,*] and Aristóteles Góes-Neto[1,2*]

[1] Molecular and Computational Biology of Fungi Laboratory, Instituto de Ciências Biológicas, Department of Microbiology, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
[2.] Program of Bioinformatics, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
[3] Institute of Veterinary Medicine, Burckhardtweg, University of Gottingen, Germany
[4] Department of Biological Science, Center of Biotechnology and Genetics, Universidade Estadual de Santa Cruz, Ilhéus, Brazil

*** Correspondence:**

Eric R. G. R. Aguiar and Aristóteles Góes-Neto

ericgdp@gmail.com; arigoesneto@icb.ufmg.br

Abstract

Fungal mitochondrial genomes encode 14 coding genes and two ribosomal genes that are essential for the functioning of the oxidative phosphorylation process. Nevertheless, the maintenance and control of expression of mitogenomes are dependent on the import of several nuclear-encoded proteins which controls this process in the mitochondria. This regulation is a complex mechanism that still poorly characterized in filamentous Fungi. *Trichoderma harzianum* is one of the main species of the genus and is recognized for its biotechnological potential. Herein, we report a new draft genome from the fungus *T.harzianum* HB324. The genome annotation identified the presence of 16,858 putative genes, and amongst them, 159 nuclear-encoded genes related to the mitochondrion. These genes were grouped in five classes according to its role: replication, transcription, translation, energy production and control of molecules to the internal matrix and cytoplasm. Altogether, our results provide new pieces of information regarding genes involved in the control and maintenance of the organelle and can also create new perspectives to the study and understand of mitochondrial control in filamentous fungi.

Keywords: genomics, functional annotation, mitochondria, maintenance and replication.

Introduction

The *Trichoderma* genus (Ascomycota - Hypocreaceae), is one of the most diverse taxon in the kingdom Fungi, with 300 species currently identified (Bissett et al. 2015). These fungi have the characteristic of being ubiquitous and adapted to different lifestyles and functions in the environment, and thus, they can be sapro-, simbio- or pathotrophs (Kubicek et al. 2011; Vaz et al. 2018; Sood et al. 2020). The species *Trichoderma harzianum*, as well as other species in the genus, have already been described as biocontrol agents against other fungi and nematodes, producing secondary metabolites with antimicrobial action, among others (Guo et al. 2018, 2019; Sood et al. 2020).

In recent years, many studies of comparative genomics have been published on the *Trichoderma*, exploring the biotechnological potential and presence of genes encoding lignocellulolytic enzymes in this group of fungi (Kubicek et al. 2011, 2019; Venice et al. 2020). Nevertheless, basal and essential processes for the survival and adaptation of these organisms in different environments are poorly understood (Selbmann et al. 2013). The fundamental source for maintaining the life of eukaryotes is the energy production, which is produced primarly by aerobic respiration through the functioning of the mitochondria (Zardoya 2020).

Mitochondria is an organelle found in eukaryotic cells whose main function is to generate energy through aerobic respiration, as well as to be involved in other cellular processes, such as fatty acids and cofactor generation (Malina et al. 2018; Zardoya 2020). This organelle descends from a likely alphaproteobacterial ancestor that was integrated into the ancestor of eukaryotic cells, a process known as endosymbiosis, which led to the retainment of part of the ancestral genetic material (Carvalho et al. 2015; Gustafsson et al. 2016; Zardoya 2020). Mitogenomes (mtDNAs) are double-stranded DNA molecules that can be circular or linear and contain a group of core and accessory genes that are replicated, transcribed, and translated independently from the nuclear genome and based on its own genetic code (Kulik et al. 2021). The study of fungal mitogenomes can also provide information on the functioning of the organelle and the evolution of this group of organisms, as well as present new genomic features that may serve as biotechnological tools (Kulik et al. 2021).

The knowledge about mitochondrial functioning is of great importance, once the energy produced by mitochondria is related to adaptation of fungi in different environments and resistance to environmental stresses, in the production of mycotoxins and pathogenicity of the fungus (Grahl et al. 2012; Tang et al. 2018). Mitochondria have

also been described as essential in drug tolerance, since one of the main factors responsible for this resistance are mutations in nuclear-encoding genes exported to the mitochondrial matrix (Mosbach et al. 2017). Nevertheless, exploratory studies on maintenance and control of the mitochondria and mitogenome in the kingdom Fungi are generally focused on the yeast species *Saccharomyces cerevisiae* (Dirick et al. 2014; Nguyen et al. 2020; Kulik et al. 2021). This species has several differences compared to filamentous fungi. Yeasts have a high tolerance to mutations that can inactivate the oxidative phosphorylation process, in addition, *S. cerevisiae* may not have the mitogenome and may not be lethal for the cell, once it still uses fermentation for energy production (Malina et al. 2018). Comparatively, filamentous fungi require the mitogenome for energy production as they are obligate aerobes (Kulik et al. 2021).

Herein, we assembled a new nuclear genome of the endophytic fungus *T. harzianum* HB324 isolated from leaves of *Hevea brasiliensis* (rubber tree). The final genome size was 38.8kbp distributed in 216 contigs. Through the annotation performed, we investigated the presence of nuclear-coding genes with mitochondrial location. In total, 159 genes were identified, in which, the majority are related to the production and export of energy molecules (ATP and GTP). Furthermore, genes involved in the maintenance and transcriptional control of the mitogenome were also identified. It was possible to estimate the transcription for all genes identified in this study. The results provided a perspective for a better understanding of the functioning and regulation of the mitogenome in fungal cells. Altogether, we present a new genomic source and improve the knowledge of mitogenome regulation carried out by the nuclear genome of *T. harzianum*.

Material and Methods

DNA extraction and genome sequencing

The isolation and sequencing of the *Trichoderma harzianum* HB324 isolate was carried out as described by Fonseca et al., 2020 (Fonseca et al. 2020). Briefly, the isolate was cultured for five days in 2% Malt Extract agar (MEA) and the mycelium was collected from the agar surface. The total DNA was extracted with the FastDNA Kit according to the manufacturer`s instructions (MP Biomedicals, CA, United States), and the extracted DNA was quantified using the Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, Waltham, MA, United States). The sequencing library was prepared using the NEBNext Fast DNA Fragmentation and Library Preparation Kit (New England Biolabs, Ipswich,

NE, United States) and subsequently sequenced on the HiSeq 2500 sequencer (Illumina, San Diego, CA, United States).

RNA extraction, transcriptomics sequencing and assembly

The isolate *T. harzianum* HB324 was cultivated for five days in MEA medium. RNA extraction was performed using the mycelium collected by the surface of the plate and then transferred to a microtube containing TRIzol® reagent (Thermo Fisher Scientific, CA, United States) and macerated with liquid nitrogen. The integrity was checked by eluting the RNA on an agarose gel with sodium hypochlorite (Aranda et al. 2012) and visualizing the presence of three bands corresponding to the ribosomal RNA (28S, 18S and 5.8S). In addition to the agarose gel, the quality of the total RNA was also verified by the Bioanalyzer equipment with the Agilent RNA 6000 Nano Kit (Agilent Technologies, Waldbronn, DE) through the microfluidic electrophoresis methodology (On-Chip electrophoresis) provided. The sequencing library was prepared according to suppliers' protocols of the kits NEBNext Ultra II RNA Library Prep Kit and NEBNext Poly (A) mRNA Magnetic Isolation Module (Ipswich, MA, USA) for enrichment of poly-A transcripts and then sequenced on the HiSeq 2500 sequencer (Illumina, San Diego, CA, United States). The transcripts were assembled using SPAdes v3.11.1 (Prjibelski et al. 2020).

Genome assembly and functional annotation and comparison of genome assembly quality

Quality of the library`s reads was verified using the FastQC v.0.11.5 program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapters and low-quality reads (Phred score < 20) were removed using the BBduk software from the BBtools package (https://sourceforge.net/projects/bbmap/). In order to calculate the coverage of the genome, it was considered 38 Mb as the expected size of the assembled genome (Srivastava and Shahid 2014b) and the number of reads after the adapter`s trimming. The remaining reads were used to assemble the genome. Two assemblers were tested, SPAdes v.3.11.1 (Prjibelski et al. 2020) and MASURCA v.3.3.1 (Zimin et al. 2013). Metrics from the assembly were accesses using the QUAST v.2 (Gurevich et al. 2013) and BUSCO v.3 (Waterhouse et al. 2018) programs.

Functional annotation of the genome was performed using the MAKER program (Cantarel et al. 2008) with four rounds of annotation. The generated RNA-seq were used

to improve the quality of annotation and to predict gene regions. In the first round, an internal MAKER algorithm was used to predict the genes and the RepeatModeler program (Flynn et al. 2020) was used for repeat masking and identification of DNA transposons and retroelements. In the second round of annotation, the SNAP software (Korf 2004) was trained to predict genes, using the *Fusarium graminearum* as a model for prediction. In the third round, the gene prediction was done with the program GeneMark (Besemer 2001) and in the fourth round, we used the SNAP and Augustus again. After the four rounds of annotation, a final file was generated with all the predicted genes, which was used to perform the protein characterization according to the Uniprot database using the *makeblastdb* application and *blastp* algorithm (The UniProt Consortium 2019). Infernal software was used to predict non-coding RNAs (Nawrocki and Eddy 2013). Gene ontology analysis was carried out in the GoFeat (Gene Ontology Functional Enrichment Annotation Tool) (Araujo et al. 2018) platform to assess the predicted proteome with e-value of 1e-10 for similarity analysis.

Identification of core components to the maintenance of mitochondrial replication and transcription and their quantification

Through the annotation of the genome performed by Maker program and the results found by GoFeat platform, it was possible to identify genes of probable mitochondrial function by onthology. The sequences were extracted using the Bedtools program (Quinlan and Hall 2010) and then manually cured for domain prediction. Seven transcriptome libraries were used to evaluate the transcriptional profile of the genes that were correctly identified with Salmon v. 1.0.4 (Patro et al. 2017) considering the value of Transcripts per million (TPM).

Identification of Repeat Point Mutation pathway in *T. harzianum* HB324

In order to investigate the presence and expression of the gene *rid1* involved in the Repeat Point Mutation (RIP) process, we create a database containing amino acid sequences from the proteins identified and characterized in *Neurospora crassa* (XP_011392925.1) and *T. atroviride* (AEM66210.1) available in the NCBI database (Geer et al. 2010) and compared with the assembled genome using the BLASTX tool considering an E-value of $1e^{-5}$ and coverage of 85% (Altschul et al. 1990). The sequence obtained was translated and had the domains predicted with HMMER program (Finn et al. 2015). Transcriptional profile of the *rid1* gene was assessed in the poly-A RNA sequencing using the Bowtie2

mapper (Langmead and Salzberg 2012, p. 2), and the value was normalized by RPKM (Reads *per kilo* base *per* million mapped reads) and compared with the normalized value of the actin constitutive gene identified in *T. harzianum* HB324 (Malinich et al. 2019).

Results

Trichoderma harzianum HB324 Assembly metrics and quality

In this work, we carried out the assembly of the *T. harzianum* HB324 nuclear genome. After removing the adapters and low-quality reads, 20,463,861,980 bases remained, which corresponds to a coverage of 538X. We tried to assemble the genome using two, SPAdes and MASURCA assemblers, and we also evaluated whether removing contigs smaller than 100 bp would improve the assembly. Afterwards, the assemblage quality was evaluated using the BUSCO program using a *Fusarium graminearum* genome as reference. We observed that the resulting genome assembled by MASURCA had a greater number of fragmented and absent genes (Supplementary Figure 1). Comparatively, the assembly by the SPAdes showed a smaller number of fragmented and absent genes (three and 18 respectively) (Supplementary Figure 1). Additionally, the elimination of contigs smaller than 100bp using the SPAdes assembler did not affect the number of identified genes. Therefore, we selected SPAdes assembled genome as reference for *T. harzianum*. The final genome size was 38,836,727 bp distributed in 216 contigs, in which the largest one presented 1,388,115bp. The values of N50 and N90 were 450,962 and 271,392 respectively. The GC content was 48.89%. The size and GC content are similar to the nuclear genomes of other *Trichoderma* species already sequenced and published (Kubicek et al. 2019).

Number of predicted genes in *Trichoderma harzianum* genome

The composition of *T. harzinaum* HB324 genome was estimated by gene prediction. The genome comprises 16,858 genes (35,015,413 nt), of which 5,705 genes could not have their function assigned by sequence similarity strategies. The number of non-coding elements identified was 254, distributed in 185 tRNAs (15,421 nt) and 69 rRNAs (32,835 nt). The intergenic region represented 9,92% (3,860,035 nt) of total genome length. The total number of single repeats were 8,501 (357,970 nt) and 1,599 (78,981 nt) low complexity repeats. No retroelements or DNA transposons were found; however, we identified 16 (1,446 nt) of rolling circles type and 1392 (456,423 nt) of uncharacterized repetitions. In addition, nine snoRNAs (1,125 nt) were identified.

Biological processes, molecular function and cellular components

Among the total predicted genes in *T. harzianum*, we were able to assign gene onthology (GO) term to 66.16% (11,153/16,858 genes). Of these, 22.58% of the genes could have biological processes designated (events related to cellular functioning), 28.52% genes presented information about cellular components (intra and extracellular location) and for 58.90% we found information regarding their molecular functions (product of genes related to elementary activities at the molecular level) (Figure 1A).

The top 20 GO terms for each of the three function groups are detailed in Figure 1B, the main groups of biological processes are related to transcription (32%), carbohydrate metabolic process (11%) and translation (9%). Furthermore, cellular components are mainly related to integral component of membrane (47%), nucleus (22%) and host cell nucleus (11%). The organelle with more GO terms is the mitochondrion (2%). In relation to molecular functions, the ATP binding (17%), zinc ion binding (16%), and DNA binding (12%) are those with the highest number of hits.

Identification of nuclear encoded genes related to the mitochondrial genome replication and transcription

Nuclear-coding genes with mitochondrial location are extremely important to keep the organelle functioning and thus, the energy supply to the cell (Gustafsson et al. 2016). Analyzing the *T. harzianum* HB324 genome, it was possible to identify 159 genes related to the membrane and internal matrix of mitochondria. Among them, 39 genes have not yet been characterized, but the term GO refers to the mitochondrial location and to processes of ions and amino acids transport by membrane proteins and ribosome formation, suggesting that there are unknown genes involved in the mitochondria functioning process. The other identified genes were classified into five functional groups: replication, transcription, translation, energy production and control of transit molecules to the internal matrix. The group energy production has more characterized proteins, accounting with 28 nuclear-coding genes. The identification of the genes found is available in Supplementary Table 1.

Regarding the maintenance of replication, transcription and mitochondria control of molecules to the internal matrix, the main genes were represented in Figure 2. Two families of genes have been identified that are responsible for controlling mitogenome replication. The first, DNA-directed DNA polymerase, is responsible for the extent of replication. This enzyme requires other subunits for DNA sequence recognition and base

repair (epsilon, kappa, and iota subunits respectively). The second family identified is the mitochondrial genome maintenance protein (MGM1) gene, responsible for the maintenance and repair of the mitogenome. The third group identified is the mitochondrial escape protein, responsible for maintaining the mitochondrial genome and controlling the escape of mtDNA into the cell cytoplasm.

In the transcription group, the gene encoding the mitochondrial RNA polymerase (POLRMT) enzyme, responsible for the initiation of mtDNA transcription was also identified. After transcription, the gene encoding protein 10 in the formation of the mitochondrial spliceosome and in the integrity of mtDNA. After transcription, the translation of mitochondrial proteins is mediated by genes responsible for the proteasome, such as ubiquitin protein ligase and the proteasome subunit alpha gene, besides the presence of ribosomal genes responsible for ribosome formation and elongation factors.

The fourth group of energy production exhibits genes linked to the functioning of the tricarboxylic acid cycle, pyruvate metabolic process, ribose phosphate metabolic process, ketone body catabolic process and interaction with generated energy molecules (ATP and GTP). In the last group, control of molecules to the internal matrix, genes involved in protein processing and targeting, integral proteins to enable the transfer of proteins and other molecules between the internal environment (mitochondrial matrix) and the external environment (cell cytoplasm), such as mitocarriers, a gene responsible for mitochondrial autophagy and mitofusin, responsible for the approximation and fusion of mitochondrial membranes were also identified.

Moreover, we found four genes encoding the LCL3 endonuclease, a gene encoding a LAGLIDADG homing endonuclease that is common to be identified in mitochondrial genomes of fungi and was previously identified in the mitogenome of *T. harzianum* HB324 (Fonseca et al. 2020) as well as a homologous endonuclease III gene that has the repair of excision sites as a putative function.

Furthermore, all genes identified in the nuclear genome involved in mitochondrial processes presented transcripts. Figure 3 shows the transcription of the genes classified in the mitogenome replication and transcription processes. The genes protein 10 and RNA polymerase (transcription process) and MGM and the épsilon subunit of the DNA polymerase showed the highest transcription.

RIP mechanism identification and transcription

In ascomycetes, the Repeat Point Mutation (RIP) pathway is responsible for controlling retrotransposons elements (Gladyshev 2017). In many fungal species it has also been suggested that this pathway is associated with the absence of transposable elements in the nuclear genome (Gladyshev 2017). As we did not identify any DNA or retrotransposons, we investigated the presence of the *rid1* gene, an essential gene identified in the RIP pathway (John Clutterbuck 2011). In the Supplementary figure 2, we can check the gene identified in *T. harzianum* and compare it with the *N. crassa* gene, the model organism of filamentous fungi. The protein predicted in *T. harzianum* is smaller (473 aa compared to 845 aa of *N. crassa*), but the DNA methylase domain has the same size (291 aa) in both sequences. Furthermore, we checked the gene expression in the transcriptome used to improve the annotation of the genome, and we found an expression of 1.1791 RPKM (normalized value), close to the transcription of the constitutive actin gene (1.1700 RPKM).

Discussion

Trichoderma harzianum is a fungus with great biotechnological potential and can be used in several industrial processes (Harman et al. 2004). In this work, we carried out the assembly and annotation of the nuclear genome of *T. harzianum*. The genome characterization revealed the size of 38kbp distributed into 216 scaffolds. The genome assembled, displayed 538X of coverage, however, it was not possible to assemble the genome in a smaller number of scaffolds or chromosomes. This may be due to the type of sequencing that relies on fragmentation and sequencing of these fragments, resulting in the failure to assemble important genetic regions to fulfill the gaps in the genome (Dominguez Del Angel et al. 2018). In spite of this feature inherent to the sequencing type, the result obtained in this work was more informative than that presented in the study by Kubiceck and colleagues (2019), in which two *T. harzianum* nuclear genomes were analyzed, the CBS 226.95 isolate with a size of 40 kbp and 841 scaffolds (GCA_003025095) and the TR274 isolate with 2,281 scaffolds (GCA_002838845) (Kubicek et al. 2019).

In order to improve the genome annotation, we sequenced seven RNA-seq libraries poly-A enriched. These libraries were grouped and used to identify exons and genes. Among all the identified genes, more than 30% had no similarity with GO terms, indicating that many genes have not yet been characterized and do not have predicted domains. Furthermore, no transposable element was identified. In the Hypocreales order,

the absence or reduced number of these elements in the genome has already been described in many species. For instance, in the species *T. virens*, *T. reesei* and *T. atroviride*, few copies of LINE, gypsy, and copia elements have been identified (Kubicek et al. 2011) and in *F. graminearum* only a gypsy retrotransposon has been found (King et al. 2015). The absence of transposable elements in the nuclear genome of *T. harzianum* may be due to the mechanism known as RIP, a sequence silencing pathway that is believed to be responsible for controlling the dispersion of these elements. This mechanism was first described in *N. crassa* (Margolin et al. 1998). It is still not well known how the pathway works, however, in *N. crassa* the functioning is dependent on the *rid1* gene (RIP-defective) (John Clutterbuck 2011). In order to corroborate the possible presence of this pathway in *T. harzianum*, we performed the search for similarity of the *rid1* gene sequences and then estimated its expression. Based on the transcriptome libraries, the expression levels identified were low and may be due to the absence of transposable elements to silence.

Usually, genomics studies focus in the investigation of biotechnological interest genes, such as the production of enzymes, secondary metabolites or pathogenicity factors (Wei et al. 2002). In this study, we attempted to assess the presence of nuclear-coding genes involved in the mitochondrial organelle and genome processes. This assessment is of great importance since the production of energy by the fungal cell is based on the functioning of the mitogenome. Nevertheless, there are few studies in mycology that attempt to reveal the control of replication, transcription, translation and energy production carried out in the mitochondria (Lipinski et al. 2010). The lack of information about the probable function of these genes may be due to the quality of the genome, the number of proteins identified, and the time and work spent to characterize *in silico* and *in vitro* a single gene in filamentous fungi. Therefore, we made an outline of how the mitochondria work from the genes identified in the genome annotation. As the number of scaffolds in our genome is lower than the other published genomes, the use of the genetic information from this genome is an opportunity for the investigation of these protein-coding genes.

The endosymbiotic theory proposes that mitochondria is the product of the symbiosis between a eukaryote and an Alphaproteobacteria ancestor. During evolutionary time, the genetic material of the symbiont was transferred to the eukaryote nuclear genome, and much of the information had been lost. Many genes involved in the production of energy in the mitochondrial organelle are encoded in the nuclear genome

and exported to mitochondria. This process is recognized in several eukaryote organisms, as well as in fungi (Grahl et al. 2012; Gustafsson et al. 2016).

Unlike other organisms, fungi exhibit a huge diversity of size and composition of mitogenomes, being able to code more than 100 genes and size greater than 200kbp (Liu et al. 2020). Despite this diversity, it is not yet known how the control of replication and transcription is performed in these mitogenomes. Here, 159 nuclear genes were identified whose GO terms refer to the mitochondrial location. The largest number of genes identified was from the energy production group, which involves proteins and the functioning of the oxidative phosphorylation cycle. Genes involved in the maintenance and control of replication, transcription, translation and control of entry and exit of molecules to and from the mitochondrion have also been identified. In yeasts, the presence of more than 400 nuclear genes encoding mitochondrial proteins has been reported (Karlberg et al. 2000), and the mitochondrial proteome of *S. cerevisiae*, involves more than 1,000 proteins (Bolender et al. 2008). In *N. crassa*, proteome studies have also been carried out, but differently, only 438 mitochondrial proteins have been identified (Ambrosio et al. 2013).

The correct functioning of the mitochondria depends on a complex crosstalk process between the nuclear and mitochondrial genomes. Damage generated in either the mitochondrial or nuclear genomes may be responsible for the cell`s fitness and response to stimuli. The replication and transcription of the mitogenome is dependent on the nuclear genome transcription and, without the presence of the protein-coding genes involved in both processes, proteins are not generated and, thus, oxidative phosphorylation does not work. In mammals, different transcription profiles of nuclear genes related to mitochondrial replication have been described. This is due to variations in the mitogenome, suggesting that mtDNA provides feedback to the nucleus, and the nucleus respond accordingly to the mitochondrial differences (Lechuga-Vieco et al. 2020).

Many studies have already been carried out and have identified the presence of Nuclear Mitochondrial DNA (NUMT) in fungi, including in the *T. harzianum* nuclear genome (Brankovics et al. 2018; Fonseca et al. 2020). The mechanism involved in the escape of DNA to the cytosol and, consecutively, to the nucleus are still not well known. In *S. cerevisiae*, transfer of mtDNA to the nuclear genome has already been observed when the cell had mutations in specific nuclear genes, depending on the structure of the mitogenome and availability of sugars in the medium (Shafer et al. 1999). In this study,

the gene coding for the mitochondrial escape protein was identified, whose GO refers to the control of mtDNA escape to the cytosol, and this this gene may be involved in the process of controlling NUMTs to the nucleus. In *S. cerevisiae*, this gene has already been characterized in the control of mtDNA (Hanekamp and Thorsness 1996). Additionally, other studies suggest that the transfer control of these molecules to the nucleus is also dependent on the organelle`s morphology. In the study carried out by Dimmer and colleagues (2005) in *S. cerevisiae*, it was described that the deletion of the genes encoding membrane proteins Mdm31 and Mdm32 alter the morphology of the mitochondria, making them spherical and the mtDNA unstable, impairing the inheritance process of mtDNA (Dimmer et al. 2005). mtDNA instability and high mutation rates influence the transfer of mitochondrial genes to the nucleus (Berg and Kurland 2000). In our study, we have not identified any sequence similar to the genes described in *S. cerevisiae*, however, three genes encoding membrane proteins have been identified with mitochondrial inheritance, which can then be related to this process. Moreover, we have identified a gene encoding mitofusion, a transmembrane protein of the GTPase family that helps in the fusion of the outer membrane of mitochondria. This process is extremely important for maintaining the organelle`s morphology. Deficiency of this gene in *N. crassa* has been described as capable of altering the life span of the fungus. Thus, the presence of the mitofusin gene, allows the organism to grow for much longer than organisms that lack this gene (Kurashima et al. 2013).

In *T. harzianum* annotation, a gene encoding a homing endonuclease (HEG) – LAGLIDADG, was also identified, which is frequently found in fungal mitogenomes (Fonseca et al. 2020). HEGs are associated with the process of intron mobility or self-splicing introns, in which a region is recognized for invasion, followed by transposition in different sites (Hausner 2003). The presence of HEGs has already been identified in the nuclear genome of other fungi. In some species, it was suggested that these genes remained active performing the process of intron homing (Haugen et al. 2004), whereas in others, it was shown that the start codon of the gene was lost, not being functional, however, part of the domain is transcribed with part of another gene, and thus enabling to alter the putative function of that gene (Dutheil et al. 2020).

In conclusion, the identification of nuclear-coding genes related to the mitochondria can be useful to the development of new studies in applied areas in Mycology, such as development of fungicides, and communication between organelles

and other mechanisms found in the cell. Additionally, other nuclear-coding genes can be identified helping to control the mitogenomes and its expression.

REFERENCES

Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. Journal of Molecular Biology 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ambrosio AB, do Nascimento LC, Oliveira BV, et al (2013) Global analyses of Ceratocystis cacaofunesta mitochondria: from genome to proteome. BMC Genomics 14:91. https://doi.org/10.1186/1471-2164-14-91

Aranda PS, LaJoie DM, Jorcyk CL (2012) Bleach gel: a simple agarose gel for analyzing RNA quality. Electrophoresis 33:366–369. https://doi.org/10.1002/elps.201100335

Araujo FA, Barh D, Silva A, et al (2018) GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. Sci Rep 8:1794. https://doi.org/10.1038/s41598-018-20211-9

Berg OG, Kurland CG (2000) Why Mitochondrial Genes are Most Often Found in Nuclei. Molecular Biology and Evolution 17:951–961. https://doi.org/10.1093/oxfordjournals.molbev.a026376

Besemer J (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Research 29:2607–2618. https://doi.org/10.1093/nar/29.12.2607

Bissett J, Gams W, Jaklitsch W, Samuels GJ (2015) Accepted Trichoderma names in the year 2015. IMA Fungus 6:263–295. https://doi.org/10.5598/imafungus.2015.06.02.02

Bolender N, Sickmann A, Wagner R, et al (2008) Multiple pathways for sorting mitochondrial precursor proteins. EMBO Rep 9:42–49. https://doi.org/10.1038/sj.embor.7401126

Brankovics B, Kulik T, Sawicki J, et al (2018) First steps towards mitochondrial pan-genomics: detailed analysis of *Fusarium graminearum* mitogenomes. PeerJ 6:e5963. https://doi.org/10.7717/peerj.5963

Cantarel BL, Korf I, Robb SMC, et al (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18:188–196. https://doi.org/10.1101/gr.6743907

Carvalho DS, Andrade RFS, Pinho STR, et al (2015) What are the Evolutionary Origins of Mitochondria? A Complex Network Approach. PLoS One 10:e0134988. https://doi.org/10.1371/journal.pone.0134988

Dimmer KS, Jakobs S, Vogel F, et al (2005) Mdm31 and Mdm32 are inner membrane proteins required for maintenance of mitochondrial shape and stability of mitochondrial DNA nucleoids in yeast. Journal of Cell Biology 168:103–115. https://doi.org/10.1083/jcb.200410030

Dirick L, Bendris W, Loubiere V, et al (2014) Metabolic and Environmental Conditions Determine Nuclear Genomic Instability in Budding Yeast Lacking Mitochondrial DNA. G3 4:411–423. https://doi.org/10.1534/g3.113.010108

Dominguez Del Angel V, Hjerde E, Sterck L, et al (2018) Ten steps to get started in Genome Assembly and Annotation. F1000Res 7:148. https://doi.org/10.12688/f1000research.13598.1

Dutheil JY, Münch K, Schotanus K, et al (2020) The insertion of a mitochondrial selfish element into the nuclear genome and its consequences. Ecol Evol 10:11117–11132. https://doi.org/10.1002/ece3.6749

Finn RD, Clements J, Arndt W, et al (2015) HMMER web server: 2015 update. Nucleic Acids Res 43:W30–W38. https://doi.org/10.1093/nar/gkv397

Flynn JM, Hubley R, Goubert C, et al (2020) RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA 117:9451–9457. https://doi.org/10.1073/pnas.1921046117

Fonseca PLC, Badotti F, De-Paula RB, et al (2020) Exploring the Relationship Among Divergence Time and Coding and Non-coding Elements in the Shaping of Fungal Mitochondrial Genomes. Front Microbiol 11:765. https://doi.org/10.3389/fmicb.2020.00765

Geer LY, Marchler-Bauer A, Geer RC, et al (2010) The NCBI BioSystems database. Nucleic Acids Research 38:D492–D496. https://doi.org/10.1093/nar/gkp858

Gladyshev E (2017) Repeat-Induced Point Mutation and Other Genome Defense Mechanisms in Fungi. Microbiol Spectr 5:. https://doi.org/10.1128/microbiolspec.FUNK-0042-2017

Grahl N, Dinamarco TM, Willger SD, et al (2012) Aspergillus fumigatus mitochondrial electron transport chain mediates oxidative stress homeostasis, hypoxia responses and fungal pathogenesis. Mol Microbiol 84:383–399. https://doi.org/10.1111/j.1365-2958.2012.08034.x

Guo H, Wang X-D, Lee D-J (2018) Proteomic researches for lignocellulose-degrading enzymes: A mini-review. Bioresource Technology 265:532–541. https://doi.org/10.1016/j.biortech.2018.05.101

Guo Y, Ghirardo A, Weber B, et al (2019) Trichoderma Species Differ in Their Volatile Profiles and in Antagonism Toward Ectomycorrhiza Laccaria bicolor. Front Microbiol 10:891. https://doi.org/10.3389/fmicb.2019.00891

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

Gustafsson CM, Falkenberg M, Larsson N-G (2016) Maintenance and Expression of Mammalian Mitochondrial DNA. Annu Rev Biochem 85:133–160. https://doi.org/10.1146/annurev-biochem-060815-014402

Hanekamp T, Thorsness PE (1996) Inactivation of YME2/RNA12, which encodes an integral inner mitochondrial membrane protein, causes increased escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae. Mol Cell Biol 16:2764–2771. https://doi.org/10.1128/MCB.16.6.2764

Harman GE, Howell CR, Viterbo A, et al (2004) Trichoderma species — opportunistic, avirulent plant symbionts. Nat Rev Microbiol 2:43–56. https://doi.org/10.1038/nrmicro797

Haugen P, Reeb V, Lutzoni F, Bhattacharya D (2004) The Evolution of Homing Endonuclease Genes and Group I Introns in Nuclear rDNA. Molecular Biology and Evolution 21:129–140. https://doi.org/10.1093/molbev/msh005

Hausner G (2003) Fungal Mitochondrial Genomes, Plasmids and Introns. In: Applied Mycology and Biotechnology. Elsevier, pp 101–131

John Clutterbuck A (2011) Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. Fungal Genetics and Biology 48:306–326. https://doi.org/10.1016/j.fgb.2010.09.002

Karlberg O, Canbäck B, Kurland CG, Andersson SG (2000) The dual origin of the yeast mitochondrial proteome. Yeast 17:170–187. https://doi.org/10.1002/1097-0061(20000930)17:3<170::AID-YEA25>3.0.CO;2-V

King R, Urban M, Hammond-Kosack MCU, et al (2015) The completed genome sequence of the pathogenic ascomycete fungus Fusarium graminearum. BMC Genomics 16:544. https://doi.org/10.1186/s12864-015-1756-1

Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics 5:59. https://doi.org/10.1186/1471-2105-5-59

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:R40. https://doi.org/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 20:485. https://doi.org/10.1186/s12864-019-5680-7

Kulik T, Van Diepeningen AD, Hausner G (2021) Editorial: The Significance of Mitogenomics in Mycology. Front Microbiol 11:628579. https://doi.org/10.3389/fmicb.2020.628579

Kurashima K, Chae M, Inoue H, et al (2013) A *uvs-5* Strain Is Deficient for a Mitofusin Gene Homologue, *fzo1* , Involved in Maintenance of Long Life Span in Neurospora crassa. Eukaryotic Cell 12:233–243. https://doi.org/10.1128/EC.00226-12

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923

Lechuga-Vieco AV, Justo-Méndez R, Enríquez JA (2020) Not all MITOCHONDRIAL DNAs are made equal and the nucleus knows it. IUBMB Life iub.2434. https://doi.org/10.1002/iub.2434

Lipinski KA, Kaniak-Golik A, Golik P (2010) Maintenance and expression of the S. cerevisiae mitochondrial genome—From genetics to evolution and systems biology. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1797:1086–1098. https://doi.org/10.1016/j.bbabio.2009.12.019

Liu W, Cai Y, Zhang Q, et al (2020) The mitochondrial genome of Morchella importuna (272.2 kb) is the largest among fungi and contains numerous introns, mitochondrial non-conserved open reading frames and repetitive sequences. International Journal of Biological Macromolecules 143:373–381. https://doi.org/10.1016/j.ijbiomac.2019.12.056

Malina C, Larsson C, Nielsen J (2018) Yeast mitochondria: an overview of mitochondrial biology and the potential of mitochondrial systems biology. FEMS Yeast Research 18:. https://doi.org/10.1093/femsyr/foy040

Malinich EA, Wang K, Mukherjee PK, et al (2019) Differential expression analysis of Trichoderma virens RNA reveals a dynamic transcriptome during colonization of Zea mays roots. BMC Genomics 20:280. https://doi.org/10.1186/s12864-019-5651-z

Margolin BS, Garrett-Engele PW, Stevens JN, et al (1998) A methylated Neurospora 5S rRNA pseudogene contains a transposable element inactivated by repeat-induced point mutation. Genetics 149:1787–1797

Mosbach A, Edel D, Farmer AD, et al (2017) Anilinopyrimidine Resistance in Botrytis cinerea Is Linked to Mitochondrial Function. Front Microbiol 8:2361. https://doi.org/10.3389/fmicb.2017.02361

Nawrocki EP, Eddy SR (2013) Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29:2933–2935. https://doi.org/10.1093/bioinformatics/btt509

Nguyen THM, Sondhi S, Ziesel A, et al (2020) Mitochondrial-nuclear coadaptation revealed through mtDNA replacements in Saccharomyces cerevisiae. In Review

Patro R, Duggal G, Love MI, et al (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14:417–419. https://doi.org/10.1038/nmeth.4197

Prjibelski A, Antipov D, Meleshko D, et al (2020) Using SPAdes De Novo Assembler. Current Protocols in Bioinformatics 70:. https://doi.org/10.1002/cpbi.102

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. https://doi.org/10.1093/bioinformatics/btq033

Selbmann L, Egidi E, Isola D, et al (2013) Biodiversity, evolution and adaptation of fungi in extreme environments. Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology 147:237–246. https://doi.org/10.1080/11263504.2012.753134

Shafer KS, Hanekamp T, White KH, Thorsness PE (1999) Mechanisms of mitochondrial DNA escape to the nucleus in the yeast Saccharomyces cerevisiae. Current Genetics 36:183–194. https://doi.org/10.1007/s002940050489

Sood M, Kapoor D, Kumar V, et al (2020) Trichoderma: The “Secrets” of a Multitalented Biocontrol Agent. Plants 9:762. https://doi.org/10.3390/plants9060762

Srivastava M, Shahid M (2014) Trichoderma Genome to Genomics: A Review. J Data Mining Genomics Proteomics 05: https://doi.org/10.4172/2153-0602.1000162

Tang G, Zhang C, Ju Z, et al (2018) The mitochondrial membrane protein FgLetm1 regulates mitochondrial integrity, production of endogenous reactive oxygen species and mycotoxin biosynthesis in Fusarium graminearum. Mol Plant Pathol 19:1595–1611. https://doi.org/10.1111/mpp.12633

The UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research 47:D506–D515. https://doi.org/10.1093/nar/gky1049

Vaz ABM, Fonseca PLC, Badotti F, et al (2018) A multiscale study of fungal endophyte communities of the foliar endosphere of native rubber trees in Eastern Amazon. Sci Rep 8:16151. https://doi.org/10.1038/s41598-018-34619-w

Venice F, Davolos D, Spina F, et al (2020) Genome Sequence of Trichoderma lixii MUT3171, A Promising Strain for Mycoremediation of PAH-Contaminated Sites. Microorganisms 8:1258. https://doi.org/10.3390/microorganisms8091258

Waterhouse RM, Seppey M, Simão FA, et al (2018) BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Molecular Biology and Evolution 35:543–548. https://doi.org/10.1093/molbev/msx319

Wei L, Liu Y, Dubchak I, et al (2002) Comparative genomics approaches to study organism similarities and differences. Journal of Biomedical Informatics 35:142–150. https://doi.org/10.1016/S1532-0464(02)00506-3

Zardoya R (2020) Recent advances in understanding mitochondrial genome diversity. F1000Res 9:270. https://doi.org/10.12688/f1000research.21490.1

Zimin AV, Marçais G, Puiu D, et al (2013) The MaSuRCA genome assembler. Bioinformatics 29:2669–2677. https://doi.org/10.1093/bioinformatics/btt476

Figure 1. GO-terms functional annotation of genes in the *Trichoderma harzianum* HB324 genome. (**A**) The total number of sequences designated in each of the three groups: Biological process, Molecular function and Cellular component. (**B**) The top 20 groups of sequences classified into each of the groups.

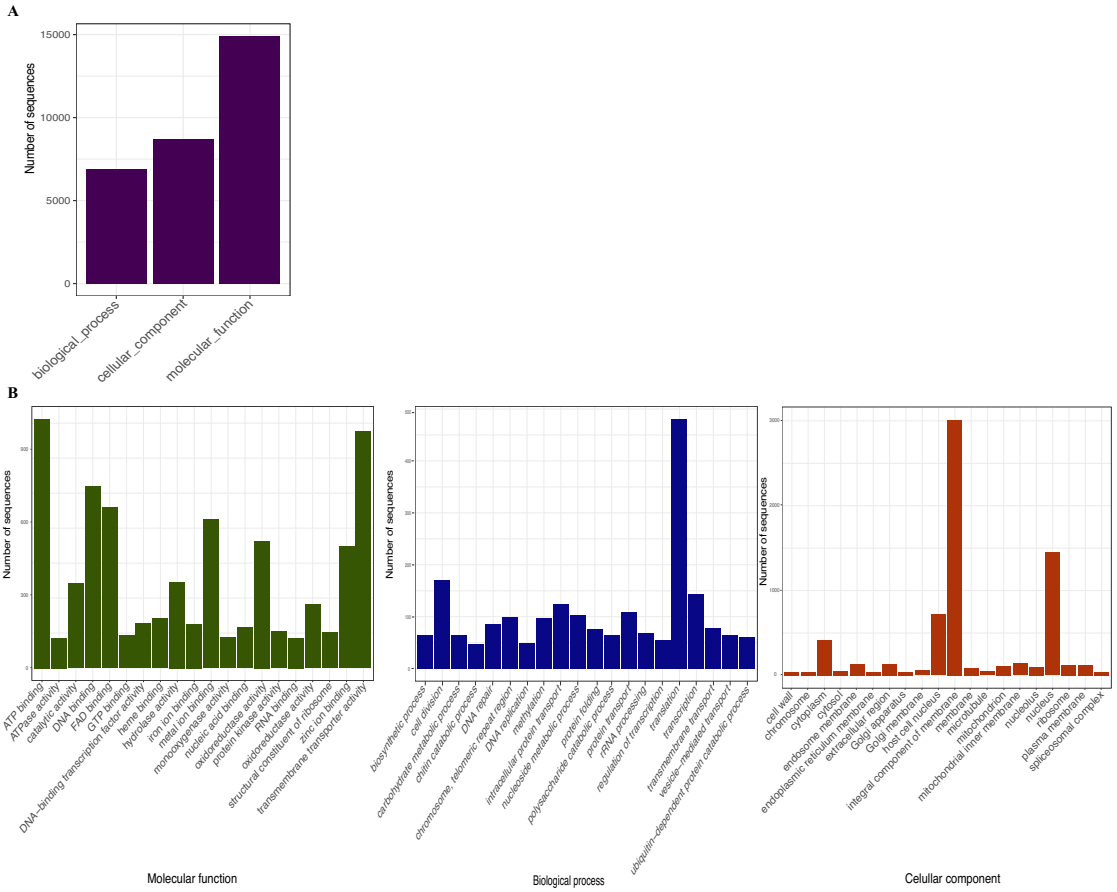


Figure 2. The functioning of mitochondrial organelles. Proteins are translated from the information from two different genomes, the mitochondrial and the nuclear. The mitochondrial genome (mtDNA) encodes 14 protein-coding, two ribosome and one ribosomal protein gene (see Fonseca et al., 2020). The nuclear genome (dDNA) encodes 158 genes involved in the maintenance, replication, transcription and translation of mitochondrial organelles. These genes are synthesized in the cytosol and transported to the mitochondria. The integral proteins of membrane allow the change of information between the cytosol and the mitochondrial matrix.

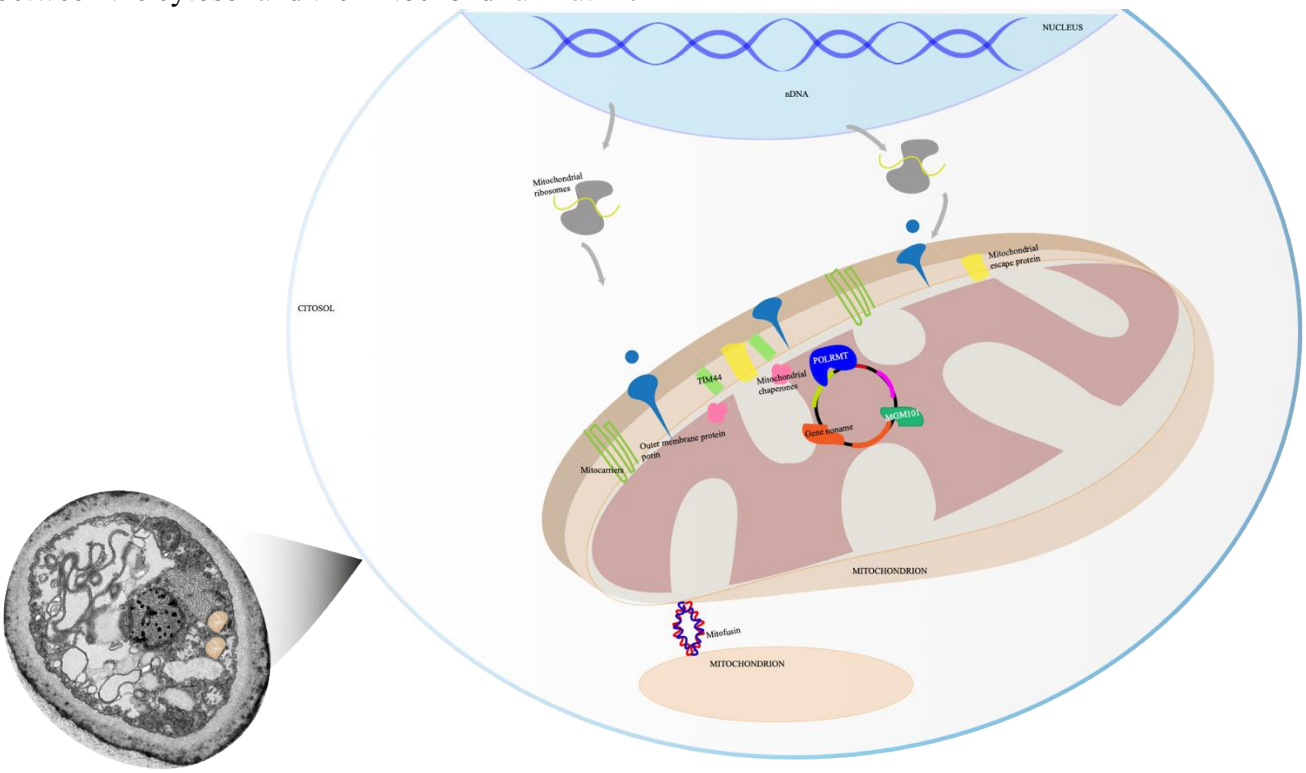
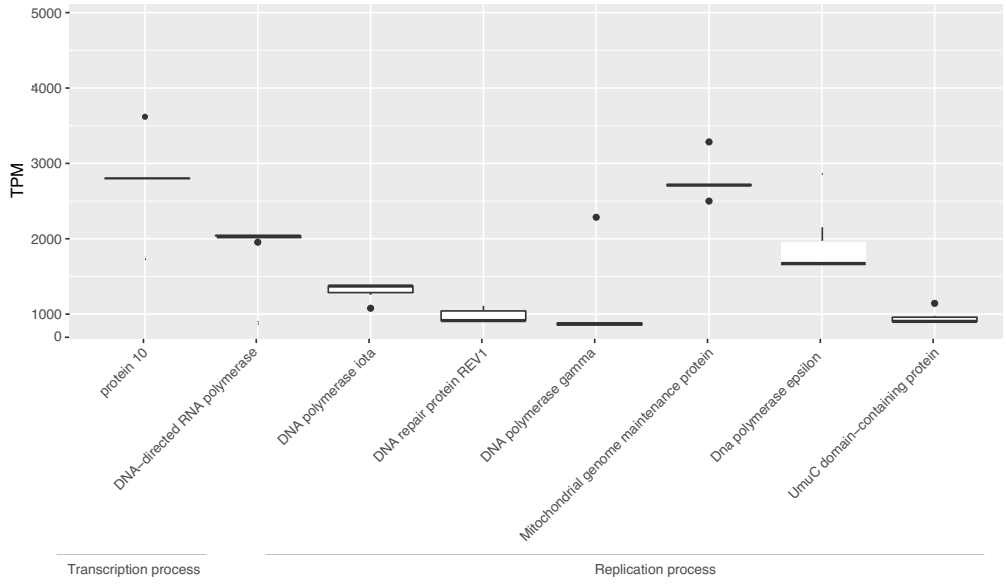


Figure 3. Transcription of the genes identified in the replication and transcription process of mitogenome. The values were normalized by TPM.



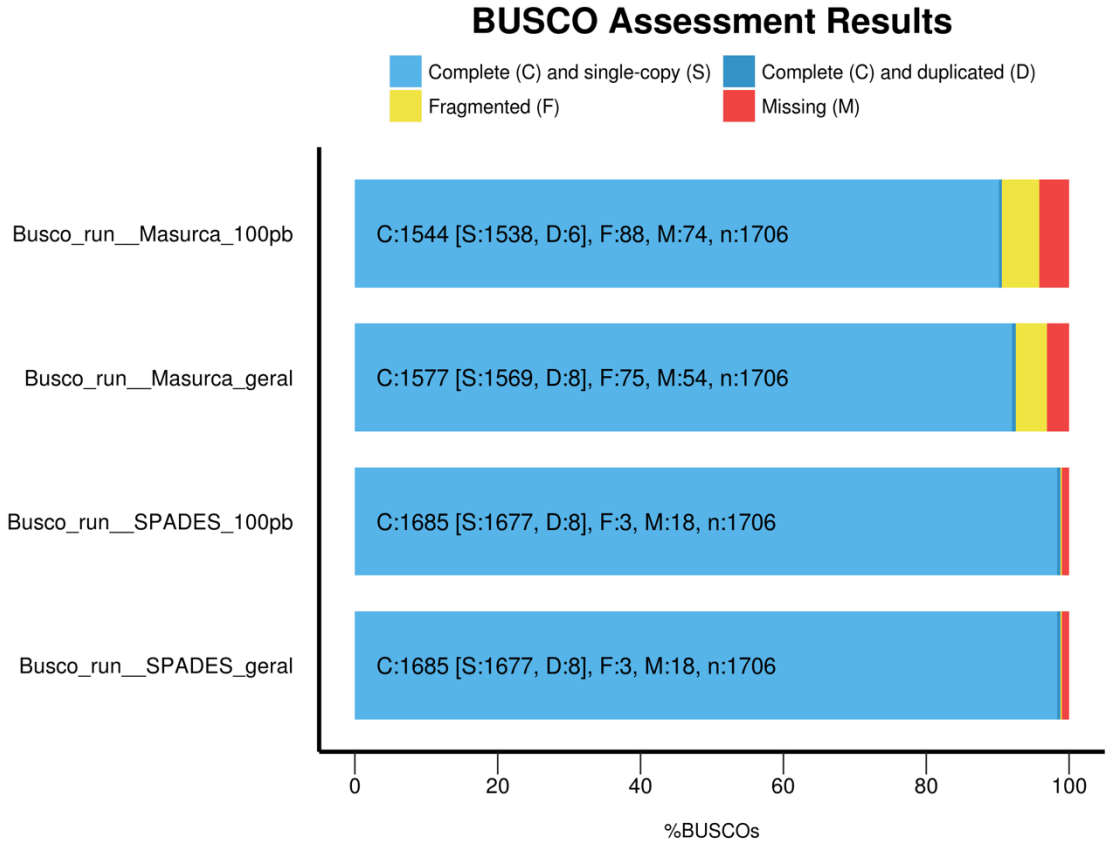
Supplementary Table 1. GO-based functional annotation of nuclear-encoded genes involved in the function of mitochondrial organelles.

Group 1 - Replication		
Gene	**GO term**	**Function**
DNA-directed DNA polymerase	GO:0000166	DNA binding
Mitochondrial genome maintenance protein	GO:0000002	DNA repair
DNA polymerase epsilon catalytic subunit	GO:0000166	premeiotic DNA replication
DNA polymerase kappa	GO:0003684	damaged DNA binding
DNA polymerase iota subunit	GO:0003684	damaged DNA binding
DNA repair protein REV1	GO:0003684	damaged DNA binding
UmuC domain-containing protein	GO:0003684	damaged DNA binding
Group 2 - Transcription		
Gene	**GO term**	**Function**
Mitochondrial RNA polymerase	GO:0003677	transcription DNA
protein 10	GO:0000002	mRNA splicing via spliceosome
N-acetyltransferase	GO:0003723	RNA processing
Group 3 - Translation		
Gene	**GO term**	**Function**
ubiquitin-protein ligase	GO:0000151	protein polyubiquitination
proteasome subunit alpha	GO:0004298	proteasome core complex alpha
54S ribosomal protein L51 mitochondrial	GO:0003735	mitochondrial translation
Elongation factor G mitochondrial	GO:0003746	mitochondrial translational elongation
MRP-L46 domain-containing protein	GO:0003735	structural constituent of ribosome
MRP-S28 domain-containing protein	GO:0003735	structural constituent of ribosome
Mitochondrial ribosomal protein L43	GO:0003735	structural constituent of ribosome
L51_S25_CI-B8 domain-containing protein	GO:0003735	structural constituent of ribosome
Letm1 RBD domain-containing protein	GO:0005739	ribosome binding
Alanine--tRNA ligase	GO:0000049	tRNA ligase activity
Elongator complex protein 3	GO:0000049	tRNA binding
Ubiquinone biosynthesis protein	GO:0005739	ubiquinone biosynthetic process
Group 4 - Energy production		
Gene	**GO term**	**Function**
Aconitate hydratase mitochondrial	GO:0003994	tricarboxylic acid cycle
Acetylglutamate kinase	GO:0003991	arginine biosynthetic process
Actin-related protein 2	GO:0000001	ATP binding
Aminomethyltransferase	GO:0004047	glycine catabolic process
ATP-dependent DNA helicase PIF1	GO:0000723	ATP binding
Biotin synthase	GO:0004076	biotin biosynthetic process
CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	GO:0005524	ATP binding

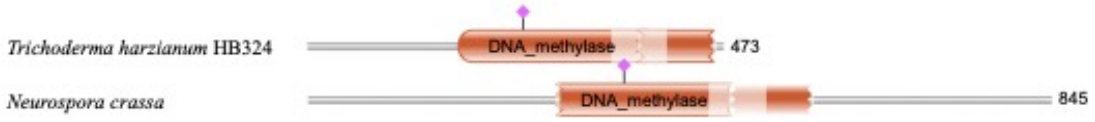
Cyclic pyranopterin monophosphate synthase	GO:0005525	GTP binding
Dynamin-related GTPase	GO:0000001	GTP binding
Homoaconitase mitochondrial	GO:0004409	lysine biosynthetic process via aminoadipic acid
Ferroxidase	GO:0004322	heme biosynthetic process
Glycine cleavage system H protein	GO:0005739	glycine decarboxylation via glycine cleavage system
Glycine cleavage system P protein	GO:0004375	glycine catabolic process
Homoaconitase mitochondrial	GO:0004409	lysine biosynthetic process via aminoadipic acid
Isocitrate dehydrogenase [NAD] subunit mitochondrial	GO:0000287	tricarboxylic acid cycle
Malic enzyme	GO:0004471	pyruvate metabolic process
Mitochondrial Rho GTPase	GO:0003924	GTP binding
NADPH-dependent diflavin oxidoreductase 1	GO:0003958	hemoprotein reductase activity
NADPH:adrenodoxin oxidoreductase mitochondrial	GO:0005739	adrenodoxin reductase activity
NUDIX domain-containing protein	GO:0005739	ribose phosphate metabolic process
Ornithine transcarbamylase	GO:0004585	ornithine carbamoyltransferase activity
Protein arginine methyltransferase NDUFAF7	GO:0005739	N symmetric methyltransferase activity
RNA helicase	GO:0003724	ATP binding
Succinate--CoA ligase [ADP-forming] subunit alpha mitochondrial	GO:0000166	tricarboxylic acid cycle
Succinyl-CoA:3-ketoacid-coenzyme A transferase	GO:0005739	ketone body catabolic process
Glutamyl-tRNA(Gln) amidotransferase subunit A mitochondrial	GO:0004040	ATP binding
Glutamyl-tRNA(Gln) amidotransferase subunit B mitochondrial	GO:0005524	ATP binding
Flavin prenyltransferase PAD1 mitochondrial	GO:0005739	flavin prenyltransferase activity
Group 5 - Control		
Gene	**GO term**	**Function**
Mitochondrial escape protein	GO:0000002	mitochondrial genome maintenance
ATP12 chaperone	GO:0005739	transporting ATP synthase complex assembly
Mitochondrial substrate carrier	GO:0005743	magnesium ion export from mitochondrion
Mitochondrial inner membrane protease subunit 2	GO:0006465	protein processing involved in protein targeting to mitochondrion
Mitochondrial intermediate peptidase	GO:0004222	protein processing involved in protein targeting to mitochondrion
Mitochondrial glycine transporter	GO:0005743	glycine import into mitochondrion

Cytochrome c oxidase subunit 3	GO:0004129	aerobic electron transport chain
Complex I-ESSS	GO:0005739	integral component of membrane
ATP synthase regulation protein NCA2	GO:0005739	integral component of membrane
NADH-ubiquinone oxidoreductase chain 3	GO:0005739	integral component of membrane
Autophagy-related protein 11	GO:0000045	autophagosome
Altered inheritance of mitochondria protein 24 mitochondrial	GO:0005739	mitochondria inheritance

Supplementary Figure 1: BUSCO assessments results of genome assembly using SPAdes and MASURCA assemblers.



Supplementary Figure 2: Characterization of *rid1* gene in *Trichoderma harzianum*.



CAPITULO 4

Transcriptional profile of *Trichoderma harzianum* mitogenome

Transcriptional profile of *Trichoderma harzianum* mitogenome

Paula L. C. Fonseca[1], Isaque João da Silva de Faria[2], Vasco A. Azevedo[3], Bertram Brenig[4], Aristóteles Góes-Neto[1,3*] and Eric R. G. R. Aguiar[3,5,*]

[1] Department of Microbiology, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
[2.] Department of Biochemistry, Institutlo de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
[3]Program of Bioinformatics, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
[4]Institute of Veterinary Medicine, Burckhardtweg, University of Gottingen, Germany
[5]Department of Biological Science, Center of Biotechnology and Genetics, Universidade Estadual de Santa Cruz, Ilhéus, Brazil

*** Correspondence:**

Eric R. G. R. Aguiar and Aristóteles Góes Neto

ericgdp@gmail.com; arigoesneto@icb.ufmg.br

Abstract

Mitochondria can be found in most eukaryotic organisms and are mainly responsible for energy production. Despite having a conserved gene content, fungal mitogenomes display a high length variation due to the presence of non-coding regions. The study of transcription and its control can provide clues about the role and how coding and non-coding regions are maintained in the genome. In mammals, mitogenome replication can generate double-stranded RNA (dsRNA) during transcription that can escape from the organelle and alter cell function. RNA interference (RNAi) pathway is able to degrade dsRNA and is responsible for post-transcriptional control in the cell. Moreover, it has been shown that other surveillance pathways and proteins can also play a role on degradation of aberrant dsRNA. Herein, we evaluated the transcriptional profile of *Trichoderma harzianum* mitogenome by sequencing small and long RNAs and investigated possible mechanisms that prevent abnormal molecules produced during the transcription to damage the cell. We detected transcription for all core genes, in which ribosomal genes showed the highest abundance. Interesting, we note that some non-coding elements also presented transcription, even higher than genes known as essential to the mitochondria. For many features, we observed sense and antisense transcription, indicating the existence of mtDNA-derived dsRNA. Indeed, we performed a strand-specific qPCR that detected sense and antisense transcripts in the *rns* ribosomal gene, confirming that dsRNA is produced during transcription. It has been shown that mitochondria-derived dsRNA molecules must be controlled to avoid misfunction of cellular processes. Therefore, we investigated the existence of different mechanisms described as targeting dsRNAs from the mitochondria. We were able to identify and detect transcriptional activity for an orthologue of SUV3, dsRNA-degrading protein described in mammals, and for the core genes of RNA interference (RNAi) pathway suggesting they are functional. Nevertheless, small RNA deep sequencing revealed small RNAs entire mitochondrial genome. Molecular characteristics of mtDNA-derived small RNAs resembles miRNA-like molecules derived from Fungi RNAi pathways, suggesting that its functional and play a role in the silencing of aberrant mitochondrial dsRNA. Altogether, our results indicate for the first time the production of mitochondria-derived dsRNA in Fungi, which is targeted by dsRNA-degradation machinery, and possibly works as a mechanism of control, such as those observed in mammals.

Keywords: Mitogenome, transcriptional profile, RNAi pathway, SUV3, dsRNA, Hypocreales

Introduction

Mitochondria are essential in most eukaryotic cells for producing energy that is used in cellular processes. Its functioning is dependent on the expression of nuclear genes that are imported to the mitochondrial matrix. In spite of this dependence, mitochondria also has its own genome (Kotiadis et al. 2014; Mendoza et al. 2020). The fungal mitochondrial genome (mitogenome or mtDNA) is of great importance for cellular respiration since it encodes essential genes for the functioning of the phosphorylative chain (Malina et al. 2018; Zardoya 2020). The proper functioning of mitochondria is dependent on coordinated regulation mechanisms for the maintenance of cellular processes (Mendoza et al. 2020). Moreover, different studies have been shown that the characterization, transcription and regulation of mitogenomes is one of the areas in Mycology that can provide essential knowledge for the investigation of species adaptation in different environments, pathogenicity, biofilm formation, virulence and oxidative stress (Grahl et al. 2012; Tang et al. 2018).

The fungal mitogenome transcription is initiated by the Mitochondrial RNA polymerase (POLRMT) enzyme encoded by the nuclear genome and imported into the mitochondria (Mendoza et al. 2020). In mammals, mitogenomes are transcribed in long polycistronic precursor transcripts derived from the two Heavy (H) and Light (L) strands, in which 22 tRNAs, rRNAs and mRNAs are excised according to the punctuation tRNA model (Ojala et al. 1981). The mRNA and rRNA sequences are then polyadenylated at the 3` end and translated for use in the electron transport chain (Nagaike et al. 2005).

In yeasts, transcription of polycistronic transcripts derived from both strands have also been described. Their transcription is polycistronic, and exhibits differences in the transcription levels of some genes, suggesting that there is post-transcriptional regulation (Kolondra et al. 2015). In filamentous fungi, however, the transcriptional profile of mitogenomes is still poorly characterized. Indeed, there is a lack of information about the profile and abundance of transcripts, the presence of non-coding RNAs, and how post-transcriptional control is performed, since the available information is almost totally restricted to genomic characterization data. (Kulik et al. 2021).

Bidirectional transcription of the mitogenomes can generate double-stranded RNA, as demonstrated in mammals (Dhir et al. 2018). According to the aforementioned authors, the mitogenome transcription is controlled by an RNA decay process called the degradosome mitochondrial RNA (mtEXO). In the absence of this degradosome complex, there is an accumulation of dsRNA in the mitochondrial matrix with

consecutive escape to the cytoplasm (Dhir et al. 2018). Furthermore, dsRNA of viral and cellular origin are recognized and processed by immune processes, altering cellular metabolism (Wiatrek et al. 2019).

One of the mechanisms capable of silencing dsRNA in the cytoplasm is the RNA interference (RNAi) pathway, responsible for the regulation of gene expression, at the pos-transcriptional level (Li et al. 2010; Dang et al. 2011b). dsRNA molecules are cleaved into small RNAs (sRNAs) between 21-30 nt by the Dicer, an RNAse III enzyme. The sRNAs are then loaded into the RNA-induced silencing complex (RISC) that contains an Argonaute protein. This complex will be responsible for recognizing the target RNA by base complementarity, and consecutive modulation of gene transcription (Kim et al. 2009; Czech and Hannon 2011; Huntzinger and Izaurralde 2011). This mechanism is conserved among eukaryotes, and in fungi, two RNAi pathways have been described: Quelling and Miotic Silencing by Unpaired DNA (MSUD). Quelling pathway is active during the asexual phase, while MSUD is active during the sexual phase (Li et al. 2010; Dang et al. 2011b). Nevertheless, the genes involved in each of the two pathways have already been found in both sexual and asexual phases (Carreras-Villaseñor et al. 2013; Campo et al. 2016). RNA silencing in fungal species are associated to pathogenicity (Weiberg et al. 2013), development (Carreras-Villaseñor et al. 2013) and antiviral defense (Segers et al. 2007; Campo et al. 2016).

Herein, we provide a comprehensive map of the *Trichoderma harzianum* mitogenome transcriptome that contains a circular genome of 32,277 bp in length, 14 protein-coding genes, 28 tRNAs, two ribosomal genes and a ribosomal protein coding gene (Fonseca et al. 2020). We performed different deep sequencing methods: mRNA strand specific sequencing with rRNA depletion, poly-A enriched mRNA, and small RNAs (sRNAs) to conduct the study. Our analysis show transcription derived from both strands and different levels for each feature. Sequences from Introns, homing endonucleases and ORFs without known function (uORF) display transcription level similar to those observed in genes considered essential for the functioning of mitochondria. The mitochondrial-derived sRNA showed profile consistent with the activation of the RNAi-related genes, suggesting that the pathway is probably functional. Moreover, the presence of sense and antisense transcription in several features of the mitogenome were detected. In order to confirm the generation of mt-derived dsRNA, we performed stranded-specific amplification of the *rns* gene, and we detected RNAs derived from both strands, detected even with low concentration. Additionally, we characterized

the protein-coding gene SUV3, the only mtEXO gene found in *T. harzianum*. Taken together, our findings are of great importance for understanding the post-transcriptional control in fungal mitogenomes.

Material and Methods

RNA deep sequencing of *T. harzianum* HB324

Pure culture of *T. harzianum* HB324 isolate was grown on 2% malt extract agar (MEA) for five days. The mycelium was collected and ground with liquid nitrogen and 500 mg aliquots were used for total RNA extraction using TRIzol Reagent (Thermo Fisher Scientific, Carlsbad, CA, USA) following the manufacture`s recommendations. RNA quality and quantity were evaluated using an automated electrophoresis system (2100 Bioanalyzer, Agilent RNA 6000 Nano Kit, Agilent Technologies, Waldbronn, DE).

For small RNA (sRNA) deep sequencing, RNA samples were stored with 30 µL of RNA protection reagent (OMEGA bio-tek, Norcross, GA, USA). The samples were prepared for sequencing with the NEXTflex Small RNA – Seq Kit V3, and sRNAs were selected by size (15–35nt) in denaturing SDS-PAGE electrophoresis (Bio Scientific Corp, Austin, TX, USA), and sequenced using Illumina HiSeq (Illumina, San Diego, CA, USA).

Two types of library preparation were carried out for total RNA deep sequencing. The first one using the kits NEBNext Ultra II RNA Library Prep Kit and NEBNext Poly (A) mRNA Magnetic Isolation Module (Ipswich, MA, USA) for enrichment of poly-A transcripts, and the second method for strand-specific with rRNA depletion sequencing using the kits QIAseq Stranded Total RNA Library Kit together with the QIAseq FastSelect RNA Removal Kit for depletion of rRNA and globin mRNA (Germantown, MD, USA). Small RNA deep sequencing was prepared using the kit NEXTflex Small RNA-seq Kit V3 (Bio Scientific Corp, Austin, TX, USA). The sRNAs were selected by size (15-35 nt) in denaturing SDS-PAGE electrophoresis, and the libraries were sequenced on the HiSeq 2500 sequencer (Illumina, San Diego, CA, United States).

Bioinformatic analyses

Pre-processing of raw RNA reads comprised adapter trimming and quality filter, in which reads with Phred quality <20 or less than 15 nt were removed using the BBduk package from BBmap program (https://sourceforge.net/projects/bbmap/). The remaining sequences from each library were mapped against the nuclear genome of the *T. harzianum* HB324 using Bowtie (sRNA library) or Bowtie2 (total RNA libraries) programs

(Langmead 2010), allowing one mismatch. Only pre-processed reads that did not map into the nuclear genome, were used for further analyses of the mtDNA of *T. harzinaum* (MT263519).

The sRNA size profile was calculated as the frequency of each sRNA read size on the reference mitogenome, considering each polarity separately. The density of sRNAs was calculated as the number of times sRNA sequences covered each nucleotide position on the reference mitogenome. The sRNA size profile and density were plotted using the R program with the ggplot2 (Wickham 2009, p. 2).

In order to estimate the number of reads that aligned on each feature annotated in mtDNA, we used the Bedtools program (Quinlan and Hall 2010). In the case of sRNA libraries, only sequences with size between 20-23 nt were considered. For sRNAs library, feature counts were normalized by reads per million (RPM value), and for total RNA libraries, we normalized the counts by Reads per kilo base per million mapped reads (RPKM value). We also analyzed public fungal sRNAs deep sequencing libraries available at NCBI database. The information for each library is available in Table 1.

dsRNA detection in mitochondrion genes in *T. harzianum* HB324

In order to corroborate the data on the presence of dsRNA generated during the transcription of mtDNA, we performed strand-specific RNA detection using a pair of specific primers for amplification of the *rns* gene designed by our group (Table 2). Firstly, total RNA was extracted with TRIzol reagent (Thermo Fisher Scientific, Carlsbad, CA, USA) and then digested with DNAse I (New England BioLabs, MA, USA) followed by cDNA synthesis, in which, five reactions were performed: (i) using random primers, (ii) without reverse transcriptase enzyme, (iii) without the presence of primers and (iv) with the designed forward and (v) reverse primers. The first three reactions were considered control and the other two are specific for detection, using the forward primer to transcribe the cDNA from negative (antisense) strand, and the reverse primer to transcribe the cDNA from positive (sense) strand. For all the reactions, the mixture between primers (500 nM) and RNA (200 ng) were incubated at 90°C for five minutes and then placed on ice for another five minutes. cDNA was synthesized using the Superscript I enzyme (Thermo Fisher Scientific, Carlsbad, CA, USA) at 50°C, for 30 minutes, and inactivated for 15 minutes at a temperature of 95°C. After synthesis, exonuclease I digestion (New England BioLabs, MA, USA) was performed to inactivate primer sequences that remain in the

reaction. Then, the cDNA was diluted 1:10 to be used for quantitative amplification (qPCR).

The qPCR reaction was performed using the SsoAdvanced Universal SYBR Green Supermix kit (Bio-Rad, CA, USA) with the appropriate primer containing a tag at 5` end (See Table 2). Each reaction contained primers in the concentration of 10 nM and 2uL of 1:10 diluted cDNA and a final volume of 10 uL. The cycling conditions were 95°C for 10`, followed by 40 cycles of 95°C for 15", 60°C for 30", 72°C for 30" and a final extension cycle of 72°C for 10`. In order to confirm the amplification, the qPCR product was visualized in 1% agarose gel and sequenced by Sanger methodology.

Characterization of genes encoding proteins from RNA interference and SUV3

In order to perform the identification of genes encoding proteins involved in the RNAi pathway, a database containing amino acids of each of the genes Argonaute (AGO), Dicer-like (DCL) and RNA polymerase RNA-dependent (RdRp) sequences were created according to (Carreras-Villaseñor et al. 2013). For the identification of mitochondrial exoribonuclease complex (mtEXO), a database of amino acids sequences was created according to (Dhir et al. 2018). The databases were used as reference to perform sequence similarity searches using Blast package (Altschul et al. 1990) with the *tblastn* parameter, comparing amino acid sequences (databases) with similar nucleotide sequences in the genome of *T. harzianum*. The best results were filtered, translated, and submitted to analysis on the PHMMER platform (Finn et al. 2015), which returned information on the presence of protein domains in each of the sequences.

Sequences of each of the identified genes were further classified phylogenetically. The sequences were aligned by the MAFFT program (Katoh et al. 2019) and ProtTest program was used to verify the best evolutionary model according to the Akaike Information Criterion (AIC) value (Akaike 1973). Phylogenetic trees were produced using maximum likelihood method with 1,000 bootstrap replicates, which were edited using the Mega software 6.0 and Geneious v. 2021.0.3.

Results

Overview of the mitochondrial transcription in *Trichoderma harzianum*

The long RNAs (lRNA) libraries enriched for poly-A generated an average of 46,793,666 reads, of which 0.05% mapped into mtDNA. On the other hand, lRNAs libraries depleted for ribossomal genes generated an average of 37,143,603 reads, of which 2.7% mapped

into mtDNA. The difference in the number of sequences mapped into mitogenome suggests that the transcripts derived from the mitogenome are poorly or not polyadenylated. Moreover, it was possible to detect reads derived from all the features annotated on the mitogenome in both types of libraries (polyadenylated and strand-specific). Nevertheless, ribosomal genes (*rrnL*, *rrnS* and *rps3*) were responsible for the majority of the transcription (Figure 1). Interesting, we note that some features, considered as accessory elements, exhibited higher counts than some protein-coding genes as, for instance, orf312 and orf448 showed higher abundance than the *nad5*, *nad4l* and *nad3* genes. Sense and antisense transcription were also identified strand-specific rRNA depleted libraries in all the features, mainly in ribosomal genes, for example, the gene *rrnS,* which had a transcription of 209.1294 and 174,121 RPKM for sense and antisense strands, respectively.

Detection of antisense transcription derived from the mitochondria

In order to confirm the presence of dsRNA derived from the mitochondria in *T. harzianum*, we performed a ssqPCR method. These assays, employing tagged primers, has been well-demonstrated to detect viral RNA of a specific polarity (Gu et al. 2007; Plaskon et al. 2009). According to the results generated by the sequencing of total RNAs, the genes with the highest antisense transcription was of ribosomal origin. We selected the *rns* gene for the design of specific primers with a tag to ensure that only cDNA that contains the tag is detected and quantified. The reaction for detection of the sense strand was amplified at CT 17, while the amplification of the negative strand was at CT 35 (late amplification). An electrophoresis was performed to confirm the amplification of the two reactions and then subsequently, sequenced. The obtained sequences confirmed the presence of the tag in the 3` end, indicating the detection of sense and antisense transcription on thr *rns* gene (Figure 2).

Identification of genes involved on mitochondrial degradosome

In mammals, two proteins have already been described interacting in a mitochondrial dsRNA degradation complex (Dhir et al. 2018). The proteins involved are Polyribonucleotide Nucleotidyltransferase 1 (PNPT1), ATP-dependent RNA helicase (SUV3), and 37S mitochondrial RNA-binding ribosomal protein (MRP1). The PNPT1 protein is found in the mitochondria membrane and is involved in the import of RNA into mitochondria. The MRP1 protein is a component of the mitochondrial ribosome found in

the inner mitochondrial membrane. The SUV3 protein is a helicase located in the mitochondria and is involved in the RNA degradation complex (Dhir et al. 2018).

Only the gene encoding the SUV3 protein was found *in T. harzianum* genome as well as in other species from Hypocreales order (Figure 3A). As in other eukaryotes, the amino acid sequence identified also has the same domains: two helicases of RNA and SUV3 (Figure 3B). We also detected transcriptional activity for SUV3 orthologue, with estimated abundance of 24 RPKM (Figure 4). However, more experiments are necessary to assess the possible role of this orthologue of SUV3 protein of the degradation of mtDNA-derived dsRNA.

Characterization of the core proteins of the RNA interference pathway in *Trichoderma harzianum*

In order to characterize the RNAi pathway in *T. harzianum*, we searched for the core proteins in the assembled genome. Using sequence search strategy, we were able to identify three AGO, two DCL and three RdRp gene sequences. These sequences were translated for domain analysis in the PHMMER program. AGO proteins showed the PAZ domain responsible for recognition and binding on the 3` end of siRNA and miRNA, and the Piwi domain involved in target cleavage (Figure 5). DCL proteins contain the RNA helicase domains for processing long substrates, dsRNA binding and two RNAse III catalytic domains. The DCL sequence, classified as quelling, also features a DEAD box domain to help the helicase in unwinding the RNA strand (Figure 6). All the three RdRp protein sequences contain the RdRp domain (Figure 7).

The proteins identified in *T. harzianum* were grouped with proteins from other species of the Ascomycota phylum. Using phylogenetic analysis, it was possible to classify each of the proteins in the respective classes of the RNAi pathway: Quelling and MSUD. Two AGO (Figure 5), one DCL (Figure 6), and one RdRp (Figure 7) proteins were classified in Quelling group, while one AGO, DCL, and RdRp were classified in MSUD group. Additionally, one RdRp sequence protein was classified as unknown (Figure 7). Quelling is responsible for maintaining the genome integrity, acting on the response against viral sequences and mobile elements, growth control, and mycelial development during the asexual phase. In contrast, MSUD is responsible for protection against mobile elements during meiosis (sexual phase) (Carreras-Villaseñor et al. 2013; Torres-Martínez and Ruiz-Vázquez 2017).

In order to confirm that the protein-coding genes were being transcribed, we estimated the transcription of each identified gene. Genes classified as from Quelling pathway exhibited a greater transcription, for instance, the Dicer and Rdrp genes from MSUD pathway had almost no transcription (less than 2 RPKM). Moreover, the gene with the highest transcription was the RdRp of unknown function (45 RPKM) (Figure 4).

Small RNA profile in mitogenomes

The RNAi pathway in fungi has the main function to maintain the integrity of the nuclear genome against selfish elements and strike viral infections (Dang et al. 2011b). Nevertheless, in despite of the involvement in the control of viruses, there is no information about the role of sRNAs in silencing elements from the mitogenomes. Thus, we sequenced sRNAs to investigate the role of RNAi pathway in controlling the integrity of the mtDNA.

The sRNA library generated a total of 6,536,053 reads. These sequences were compared with the mitogenome of *T. harzianum* HB324, in which 1,67% reads were mapped. The sRNA profile showed a peak size between 20-22 nt (Figure 8A), typical of RNAi pathway activation in fungi, especially miRNA-like (Schumann et al. 2010). Small RNAs were mapped throughout the mitogenome of *T. harzianum* HB324 (Figure 8B), and the sRNAs with size between 20-22 nt were considered to estimate the transcription of each feature of the annotated mtDNA (Figure 9). Ribosomal genes showed a higher number of sRNAs, probably due its high transcription previously identified in lRNAs libraries. Furthermore, other features such as, orf324, homing endonucleases (LAGLIDADG and GIY) and Intron IB also showed a considerable number of sequences when compared to other genes that are essential for the functioning of mitochondria, such as *atp8*, *atp9*, *nad1* and *nad4l*. In order to ensure that the mapped reads are of mitochondrial origin, we also perform the mapping against the nuclear genome and, therefore, confirmed that the result are were actually from the mitogenome (Supplementary figure 1).

In order to compare the results found in sRNA library of *T. harzianum* HB324, we evaluated other public sRNAs libraries. In Figure 10 we can see the same profile of size and mapping identified in our library in the species *T. atroviride* (Figure 10A) and *Fusarium graminearum* (Figure 10B). RNAi pathway has three groups of genes essential for its functioning. The previously studied *F. graminearum* species had knockout sRNA libraries for each of the gene groups, enabling to assess the role of each protein on the

production of small RNAs from mitogenome. Figure 11 shows the small RNA abundance in libraries with knockout for the genes DCL1, DCL2, DCL1 and 2, AGO1, AGO2, AGO1 and 2. Knocked-out libraries displayed a difference in the number of sRNAs 20-23 nt mapped in mtDNA. The AGO (Quelling pathway) knockout exhibited a higher number of sRNAs than the DCL (Quelling pathway) knockout. The AGO and DCL knockout classified as MSUD pathway did not show a variation in the number of generated sRNAs. This difference is due to the sexual phase of the species *F. graminearum* (asexual phase), and the genes of the Quelling pathway have a greater influence on the sRNA generation. Altogether, these indicates that the RNAi pathway plays a role in the transcriptional regulation of mtDNA (Figure 12).

Discussion

Mitochondria are essential for most of the eukaryotic cells as it provides energy to cells in the form of ATP molecules (Vartak et al. 2013). Changes in the functioning of mitochondria can result in serious consequences. In mammals, for instance, numerous diseases caused by mitochondrial disorders have already been characterized, such as Parkinson`s disease, cardiomyopathy and others (Li et al. 2019). Thus, studies that aim to evaluate the mitogenome transcription are extremely important for understanding the mitochondrial functioning. Nevertheless, in the kingdom Fungi, mitogenome transcription studies are restricted to yeast species (Kolondra et al. 2015) and is assumed that the knowledge obtained is universal for all fungi.

Herein, we evaluated the transcriptional profile of the *T. harzianum* HB324 mitogenome (Vaz et al. 2018) by performing the sequencing of the transcriptome enriched for mRNAs and strand-specific rRNA depleted libraries. In both methods it was possible to identify sense and antisense reads in most of the features of mitogenome. A higher number of reads mapping the mitogenomes were found in the strand-specific libraries, suggesting that fungal mitochondrial transcripts are not polyadenylated, such as the mitochondrial mRNAs from yeasts (Chang and Tong 2012). Nonetheless, in the study by Gowda and colleagues (2010), transcriptome enriched for mRNAs was performed in the species *Magnaporthe oryzae,* and only 43 reads mapped to the mitogenome, suggesting that the transcripts are not polyadenylated (Gowda et al. 2010).

The genes that showed the highest transcription in the evaluated libraries were genes of ribosomal origin. Other features considered non-coding elements such as orf324, showed a greater transcription than other protein-coding genes, indicating that these

regions may have a probable function. In mice, it has been reported that 15% of the mitochondrial transcriptome is represented by long non-coding RNA (mt-lncRNA), found in both ribosomal and protein coding genes (Villegas et al. 2000; Rackham et al. 2011). Thus, we suggest that some of the genes classified as non-coding may have the function of lncRNAs; however, further studies are needed for confirming this hypothesis.

Antisense transcription was identified in many features in all the evaluated libraries. Circular mitogenomes are bi-directionally transcribed, and are being able to generate overlapping transcripts that may result in double-stranded RNA structures (Dhir et al. 2018). In mammals, it is known that the accumulation of dsRNA in the cytoplasm of cells can be toxic due to the inhibition of protein synthesis (Cooper et al. 1979). Therefore, eukaryotic cells may have a mechanism to control dsRNA generated by mitochondria. In mammals, two proteins have already been described interacting in a mitochondrial dsRNA degradation complex (mtEXO) (Dhir et al. 2018). The proteins involved are PNPT1 and SUV3. The PNPT1 protein is found in the mitochondria membrane and is involved in the import of RNA into the organelle, stability of mRNAs generated in mitochondria and degradation of non-coding RNA. Silencing studies have shown deleterious effects on RNA import and translation of the mitogenome (Vedrenne et al. 2012).

The SUV3 protein is a helicase located in the mitochondria and is involved in the RNA degradation complex mtEXO. Only the gene SUV3 was found in the nuclear genome of *T. harzianum*. This protein has already been identified in other species of filamentous fungi and yeasts. In the latter case, functional studies have already been carried out and it has been demonstrated that the loss of SUV3 was responsible for the accumulation of intronic RNA sequences and subsequent changes on the stability of mitochondrial transcripts, indicating that this gene is essential for the respiratory chain (Golik et al. 2004), and suggesting that, in filamentous fungi, this gene has the same function.

In order to confirm the presence of RNAi pathway in *T. harzianum*, we characterized the genes encoding AGO, DCL, and RdRp genes in the nuclear genome. Two sequences of the DCL gene and three sequences were found for each of the AGO and RdRp genes. Besides, two genes from AGO family were classified in the Quelling pathway (asexual phase). The presence of two AGO proteins in the RNAi pathway has already been reported in the fungal specie *Magnaporthe oryzae* by Nguyen and colleagues (2018) and has been related to a mechanism for suppressing the activity of the other two

AGO genes. This mechanism would be important to maintain a baseline level of transposition of mobile elements during the asexual phase, allowing some genetic alteration to occur (Nguyen et al. 2018). Each DCL gene was classified into a sexual phase. As for RdRp, one of the genes was classified in a group considered unknown. A previous study by Zhang and colleagues (2014) identified this third group of RdRps as a non-functional pseudogene; however, the protein has the complete domain and, therefore, studies should be developed to investigate their possible functions (Zhang et al. 2014). In this study, we verified that the RdRp gene classified as non-functional displayed the highest transcription level when compared to all the other evaluated genes, indicating that the gene is transcribed and that it may have an uncharacterized function. Additionally, the classified genes belonging to the Quelling pathways showed a higher expression than the genes classified as MSUD. This was already expected since the genus *Trichoderma* is considered the asexual phase (anamorph) of the genus *Hypocrea* (teleomorph or sexual phase). Similarly, in the study conducted by Campo and colleagues (2016) the expression of each AGO, DCL, and RdRp protein in the specie *C. higginsianum* were reported and the proteins classified in the Quelling pathway had a higher expression than the proteins classified in MSUD, which is consistent with the sexual phase (Campo et al. 2016).

The data obtained from sRNA sequencing suggests that RNAi pathway is active, due to the presence of a peak between 20-22 nt. Furthermore, it was possible to map the entire extension of the mitogenome in both strands (sense and antisense), indicating that sRNAs are generated in mitochondria (Lax et al. 2020). Nonetheless, no other study had shown evidence on this pathway activity in controlling the mitochondrial genome.

In order to corroborate the presence of sRNAs throughout the fungal mitogenome, we evaluated public sRNA libraries from two other species, *T. atroviride* and *F. graminearum*. In both, it was possible to identify the same size profile and mitogenome mapping. In the case of *F. graminearum*, it was still possible to evaluate the impact of AGO and DCL genes on the generation of mitochondrial sRNAs. Usually, the knockout libraries for the DCL genes had a reduction in the number of sRNAs when compared to the control library and the knockout libraries for the AGO genes. This difference has already been observed in *Colletotrichum higginsianum* (Campo et al. 2016), and it is already expected since the knockout of the DCL prevents the dsRNA from being cut into smaller pieces, thus reducing the number of sRNAs. In the case of AGO knockout, an accumulation of sRNAs will occur, since they are cleaved by the DCL but not loaded into

the RISC complex (Campo et al. 2016). Taken together, these results indicate that the RNAi pathway has a role in controlling mitogenomes.

In conclusion, the data found in this chapter can help in the transcription and its control knowledge in filamentous fungi Current knowledge about mitochondrial functioning is based on model and easily growing organisms, such as yeasts, however, it is known that each group of organisms has its idiosyncrasies and, thus, differences in functioning. For instance, many yeast species do not have genes involved in the RNAi pathway (Drinnenberg et al. 2009), unlike filamentous fungi, and our findings suggest a possible role of this pathway in the post-transcriptional control of mitogenomes in filamentous fungal species.

References

Akaike (1973) Information Theory and an Extension of the Maximum Likelihood Principle. In: Proceedings of the 2nd International Symposium on Information Theory. B. N. Petrov, & F. Csaki, Budapest: Akademiai Kiado

Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. Journal of Molecular Biology 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Campo S, Gilbert KB, Carrington JC (2016) Small RNA-Based Antiviral Defense in the Phytopathogenic Fungus Colletotrichum higginsianum. PLoS Pathog 12:e1005640. https://doi.org/10.1371/journal.ppat.1005640

Carreras-Villaseñor N, Esquivel-Naranjo EU, Villalobos-Escobedo JM, et al (2013) The RNAi machinery regulates growth and development in the filamentous fungus *Trichoderma atroviride*: Function of sRNAs in *Trichoderma*. Molecular Microbiology 89:96–112. https://doi.org/10.1111/mmi.12261

Chang JH, Tong L (2012) Mitochondrial poly(A) polymerase and polyadenylation. Biochim Biophys Acta 1819:992–997. https://doi.org/10.1016/j.bbagrm.2011.10.012

Cooper JA, Morser J, Colby MS, Burke DC (1979) The Toxic Effect of Double-stranded RNA for Interferon-treated Cells: Evidence for a Heterogeneous Cellular Response and the Role of the Cell Nucleus. Journal of General Virology 43:553–561. https://doi.org/10.1099/0022-1317-43-3-553

Czech B, Hannon GJ (2011) Small RNA sorting: matchmaking for Argonautes. Nat Rev Genet 12:19–31. https://doi.org/10.1038/nrg2916

Dang Y, Yang Q, Xue Z, Liu Y (2011) RNA interference in fungi: pathways, functions, and applications. Eukaryot Cell 10:1148–1155. https://doi.org/10.1128/EC.05109-11

Dhir A, Dhir S, Borowski LS, et al (2018) Mitochondrial double-stranded RNA triggers antiviral signalling in humans. Nature 560:238–242. https://doi.org/10.1038/s41586-018-0363-0

Drinnenberg IA, Weinberg DE, Xie KT, et al (2009) RNAi in budding yeast. Science 326:544–550. https://doi.org/10.1126/science.1176945

Finn RD, Clements J, Arndt W, et al (2015) HMMER web server: 2015 update. Nucleic Acids Res 43:W30–W38. https://doi.org/10.1093/nar/gkv397

Fonseca PLC, Badotti F, De-Paula RB, et al (2020) Exploring the Relationship Among Divergence Time and Coding and Non-coding Elements in the Shaping of Fungal Mitochondrial Genomes. Front Microbiol 11:765. https://doi.org/10.3389/fmicb.2020.00765

Golik P, Zwolinska U, Stepien P, Lazowska J (2004) The gene from is a functional equivalent of its orthologue and is essential for respiratory growth. FEMS Yeast Research 4:477–485. https://doi.org/10.1016/S1567-1356(03)00160-0

Gowda M, Nunes CC, Sailsbery J, et al (2010) Genome-wide characterization of methylguanosine-capped and polyadenylated small RNAs in the rice blast fungus Magnaporthe oryzae. Nucleic Acids Research 38:7558–7569. https://doi.org/10.1093/nar/gkq583

Grahl N, Dinamarco TM, Willger SD, et al (2012) Aspergillus fumigatus mitochondrial electron transport chain mediates oxidative stress homeostasis, hypoxia responses and fungal pathogenesis. Mol Microbiol 84:383–399. https://doi.org/10.1111/j.1365-2958.2012.08034.x

Gu C, Zheng C, Shi L, et al (2007) Plus- and minus-stranded foot-and-mouth disease virus RNA quantified simultaneously using a novel real-time RT-PCR. Virus Genes 34:289–298. https://doi.org/10.1007/s11262-006-0019-2

Huntzinger E, Izaurralde E (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet 12:99–110. https://doi.org/10.1038/nrg2936

Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics 20:1160–1166. https://doi.org/10.1093/bib/bbx108

Kim VN, Han J, Siomi MC (2009) Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol 10:126–139. https://doi.org/10.1038/nrm2632

Kolondra A, Labedzka-Dmoch K, Wenda JM, et al (2015) The transcriptome of Candida albicans mitochondria and the evolution of organellar transcription units in yeasts. BMC Genomics 16:827. https://doi.org/10.1186/s12864-015-2078-z

Kotiadis VN, Duchen MR, Osellame LD (2014) Mitochondrial quality control and communications with the nucleus are important in maintaining mitochondrial function and cell health. Biochim Biophys Acta 1840:1254–1265. https://doi.org/10.1016/j.bbagen.2013.10.041

Kulik T, Van Diepeningen AD, Hausner G (2021) Editorial: The Significance of Mitogenomics in Mycology. Front Microbiol 11:628579. https://doi.org/10.3389/fmicb.2020.628579

Langmead B (2010) Aligning Short Sequencing Reads with Bowtie. Curr Protoc Bioinform 32:. https://doi.org/10.1002/0471250953.bi1107s32

Lax C, Tahiri G, Patiño-Medina JA, et al (2020) The Evolutionary Significance of RNAi in the Fungal Kingdom. IJMS 21:9348. https://doi.org/10.3390/ijms21249348

Li H, Slone J, Fei L, Huang T (2019) Mitochondrial DNA Variants and Common Diseases: A Mathematical Model for the Diversity of Age-Related mtDNA Mutations. Cells 8:608. https://doi.org/10.3390/cells8060608

Li L, Chang S, Liu Y (2010) RNA interference pathways in filamentous fungi. Cell Mol Life Sci 67:3849–3863. https://doi.org/10.1007/s00018-010-0471-y

Malina C, Larsson C, Nielsen J (2018) Yeast mitochondria: an overview of mitochondrial biology and the potential of mitochondrial systems biology. FEMS Yeast Research 18:. https://doi.org/10.1093/femsyr/foy040

Mendoza H, Perlin MH, Schirawski J (2020) Mitochondrial Inheritance in Phytopathogenic Fungi-Everything Is Known, or Is It? Int J Mol Sci 21:. https://doi.org/10.3390/ijms21113883

Nagaike T, Suzuki T, Katoh T, Ueda T (2005) Human mitochondrial mRNAs are stabilized with polyadenylation regulated by mitochondria-specific poly(A) polymerase and polynucleotide phosphorylase. J Biol Chem 280:19721–19727. https://doi.org/10.1074/jbc.M500804200

Nguyen Q, Iritani A, Ohkita S, et al (2018) A fungal Argonaute interferes with RNA interference. Nucleic Acids Research 46:2495–2508. https://doi.org/10.1093/nar/gkx1301

Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–474. https://doi.org/10.1038/290470a0

Plaskon NE, Adelman ZN, Myles KM (2009) Accurate Strand-Specific Quantification of Viral RNA. PLoS ONE 4:e7468. https://doi.org/10.1371/journal.pone.0007468

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. https://doi.org/10.1093/bioinformatics/btq033

Rackham O, Shearwood A-MJ, Mercer TR, et al (2011) Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins. RNA 17:2085–2093. https://doi.org/10.1261/rna.029405.111

Schumann U, Ayliffe M, Kazan K, Wang M-B (2010) RNA silencing in fungi. Front Biol 5:478–494. https://doi.org/10.1007/s11515-010-0550-3

Segers GC, Zhang X, Deng F, et al (2007) Evidence that RNA silencing functions as an antiviral defense mechanism in fungi. Proc Natl Acad Sci U S A 104:12902–12906. https://doi.org/10.1073/pnas.0702500104

Tang G, Zhang C, Ju Z, et al (2018) The mitochondrial membrane protein FgLetm1 regulates mitochondrial integrity, production of endogenous reactive oxygen species and mycotoxin biosynthesis in Fusarium graminearum. Mol Plant Pathol 19:1595–1611. https://doi.org/10.1111/mpp.12633

Torres-Martínez S, Ruiz-Vázquez RM (2017) The RNAi Universe in Fungi: A Varied Landscape of Small RNAs and Biological Functions. Annu Rev Microbiol 71:371–391. https://doi.org/10.1146/annurev-micro-090816-093352

Vartak R, Porras CA-M, Bai Y (2013) Respiratory supercomplexes: structure, function and assembly. Protein Cell 4:582–590. https://doi.org/10.1007/s13238-013-3032-y

Vaz ABM, Fonseca PLC, Badotti F, et al (2018) A multiscale study of fungal endophyte communities of the foliar endosphere of native rubber trees in Eastern Amazon. Sci Rep 8:16151. https://doi.org/10.1038/s41598-018-34619-w

Vedrenne V, Gowher A, De Lonlay P, et al (2012) Mutation in PNPT1, which encodes a polyribonucleotide nucleotidyltransferase, impairs RNA import into mitochondria and causes respiratory-chain deficiency. Am J Hum Genet 91:912–918. https://doi.org/10.1016/j.ajhg.2012.09.001

Villegas J, Zárraga AM, Muller I, et al (2000) A novel chimeric mitochondrial RNA localized in the nucleus of mouse sperm. DNA Cell Biol 19:579–588. https://doi.org/10.1089/104454900439809

Weiberg A, Wang M, Lin F-M, et al (2013) Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. Science 342:118–123. https://doi.org/10.1126/science.1239705

Wiatrek DM, Candela ME, Sedmík J, et al (2019) Activation of innate immunity by mitochondrial dsRNA in mouse cells lacking p53 protein. RNA 25:713–726. https://doi.org/10.1261/rna.069625.118

Wickham H (2009) Ggplot2: elegant graphics for data analysis. Springer, New York

Zardoya R (2020) Recent advances in understanding mitochondrial genome diversity. F1000Res 9:270. https://doi.org/10.12688/f1000research.21490.1

Zhang D-X, Spiering MJ, Nuss DL (2014) Characterizing the roles of Cryphonectria parasitica RNA-dependent RNA polymerase-like genes in antiviral defense, viral recombination and transposon transcript accumulation. PLoS One 9:e108653. https://doi.org/10.1371/journal.pone.0108653

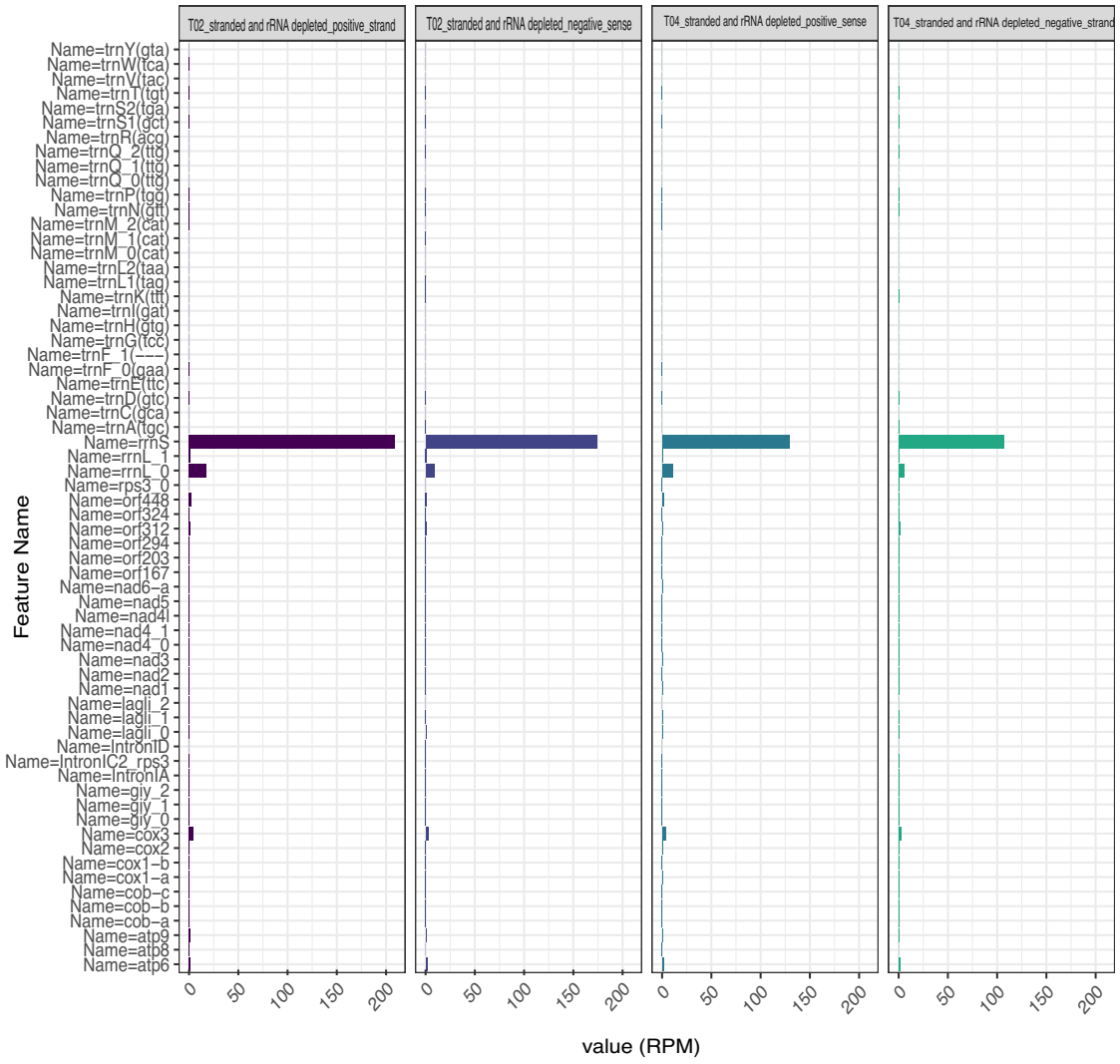
Table 1. Public sRNAs libraries used in this study.

Condition	Species	SRA ID
Five days of cultive	*Trichoderma atroviride*	SRR8280363
Five days of cultive	*Fusarium graminearum*	SRR4415737
DCL1 Knockout	*Fusarium graminearum*	SRR4415738
DCL2 Knockout	*Fusarium graminearum*	SRR4415739
DCL1 and 2 Knockout	*Fusarium graminearum*	SRR4415740
AGO1 Knockout	*Fusarium graminearum*	SRR4415741
AGO2 Knockout	*Fusarium graminearum*	SRR4415742
AGO1 and 2 Knockout	*Fusarium graminearum*	SRR4415743

Table 2. Sequences of primers for quantitative PCR

Purpose	Primer name	Step	Primer sequence
Positive Strand detection	rns_R_tag	RT-PCR	5` **GCGTCATGGTGGCGAATAA**ACAGCCATGCAACACCTGTA 3`
	rns_F_tag	qPCR	5` **GCGTCATGGTGGCGAATAA**TGGCAACGCTGGAACTGTAA 3`
	Tag	qPCR	GCGTCATGGTGGCGAATAA
Negative strand detection	rns_F_tag	RT-PCR	5` **GCGTCATGGTGGCGAATAA**TGGCAACGCTGGAACTGTAA 3`
	rns_R_tag	qPCR	5` **GCGTCATGGTGGCGAATAA**ACAGCCATGCAACACCTGTA 3`
	Tag	qPCR	GCGTCATGGTGGCGAATAA

Figure 1. Analysis of the number of lRNAs reads mapped in the *Trichoderma harzianum* mitogenome in poly(A) libraries and in strand-specific rRNA depleted libraries. The values are normalized by RPKM.



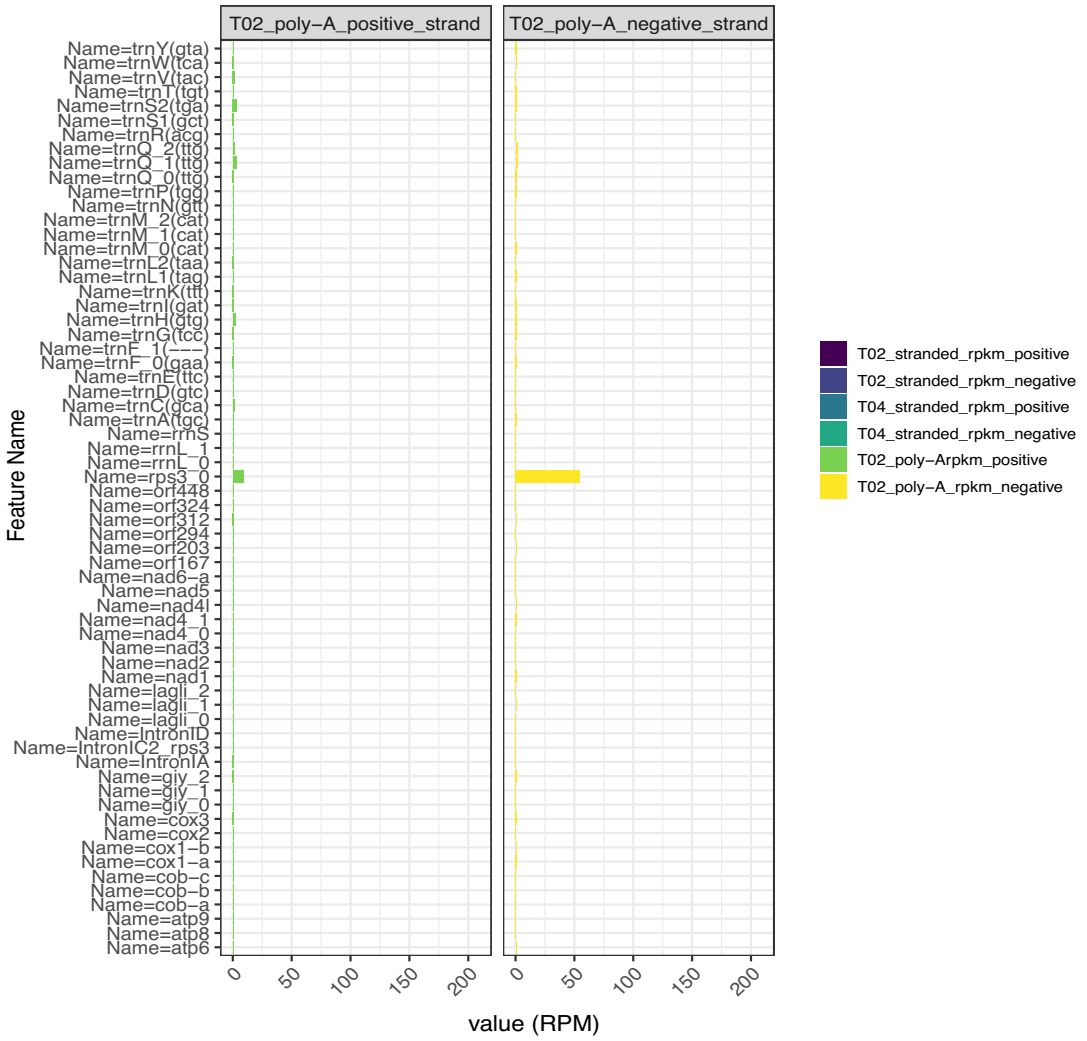

T02_poly−A_positive_strand
T02_poly−A_negative_strand
Feature Name
Name=trnY(gta)
Name=trnW(tca)
Name=trnV(tac)
Name=trnT(tgt)
Name=trnS2(tga)
Name=trnS1(gct)
Name=trnR(acg)
Name=trnQ_2(ttg)
Name=trnQ_1(ttg)
Name=trnQ_0(ttg)
Name=trnP(tgg)
Name=trnN(gtt)
Name=trnM_2(cat)
Name=trnM_1(cat)
Name=trnM_0(cat)
Name=trnL2(taa)
Name=trnL1(tag)
Name=trnK(ttt)
Name=trnI(gat)
Name=trnH(gtg)
Name=trnG(tcc)
Name=trnF_1(---)
Name=trnF_0(gaa)
Name=trnE(ttc)
Name=trnD(gtc)
Name=trnC(gca)
Name=trnA(tgc)
Name=rrnS
Name=rrnL_1
Name=rrnL_0
Name=rps3_0
Name=orf448
Name=orf324
Name=orf312
Name=orf294
Name=orf203
Name=orf167
Name=nad6−a
Name=nad5
Name=nad4l
Name=nad4_1
Name=nad4_0
Name=nad3
Name=nad2
Name=nad1
Name=lagli_2
Name=lagli_1
Name=lagli_0
Name=IntronID
Name=IntronIC2_rps3
Name=IntronIA
Name=giy_2
Name=giy_1
Name=giy_0
Name=cox3
Name=cox2
Name=cox1−b
Name=cox1−a
Name=cob−c
Name=cob−b
Name=cob−a
Name=atp9
Name=atp8
Name=atp6
0
50
100
150
200
value (RPM)
T02_stranded_rpkm_positive
T02_stranded_rpkm_negative
T04_stranded_rpkm_positive
T04_stranded_rpkm_negative
T02_poly−Arpkm_positive
T02_poly−A_rpkm_negative

Figure 2. The gene *rns* sequenced with primers with specific tag for detecting (A) sense and (B) antisense transcription.

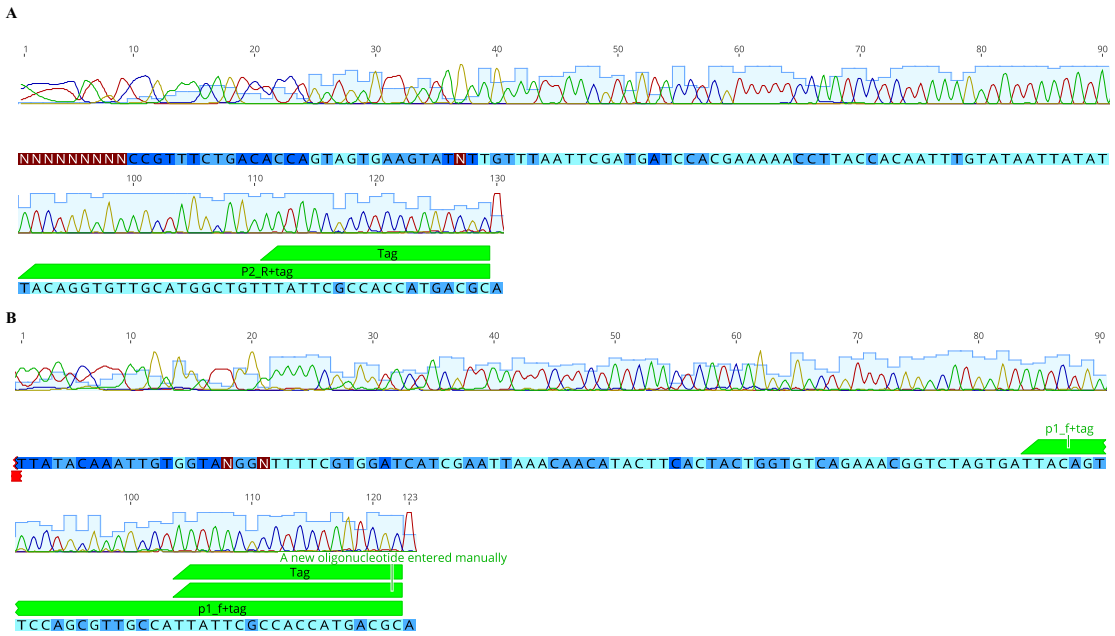


Figure 3. Characterization of the protein-coding gene SUV3. (A) Phylogenetic analysis using the Maximum Likelihood method with 1000 bootstrap replicates. (B) Domains found in the SUV3 protein.

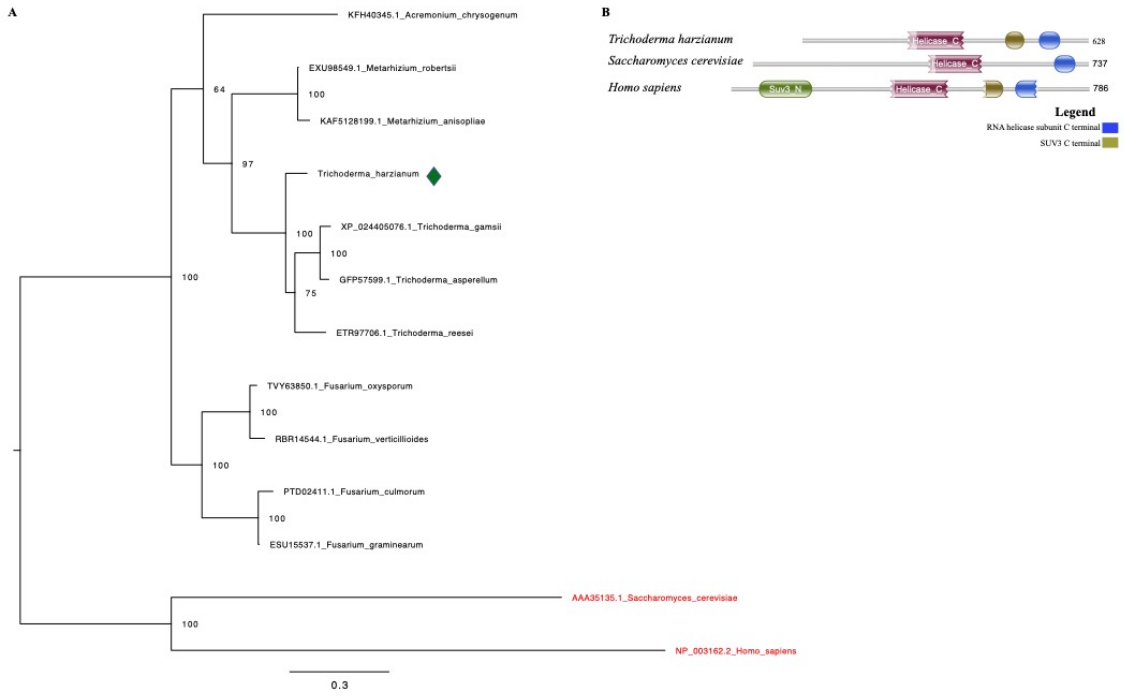


Figure 4. Transcriptional profile of the protein-coding genes involved in the mtEXO and RNAi pathway identified in *Trichoderma harzianum*. The values are normalized by RPKM.

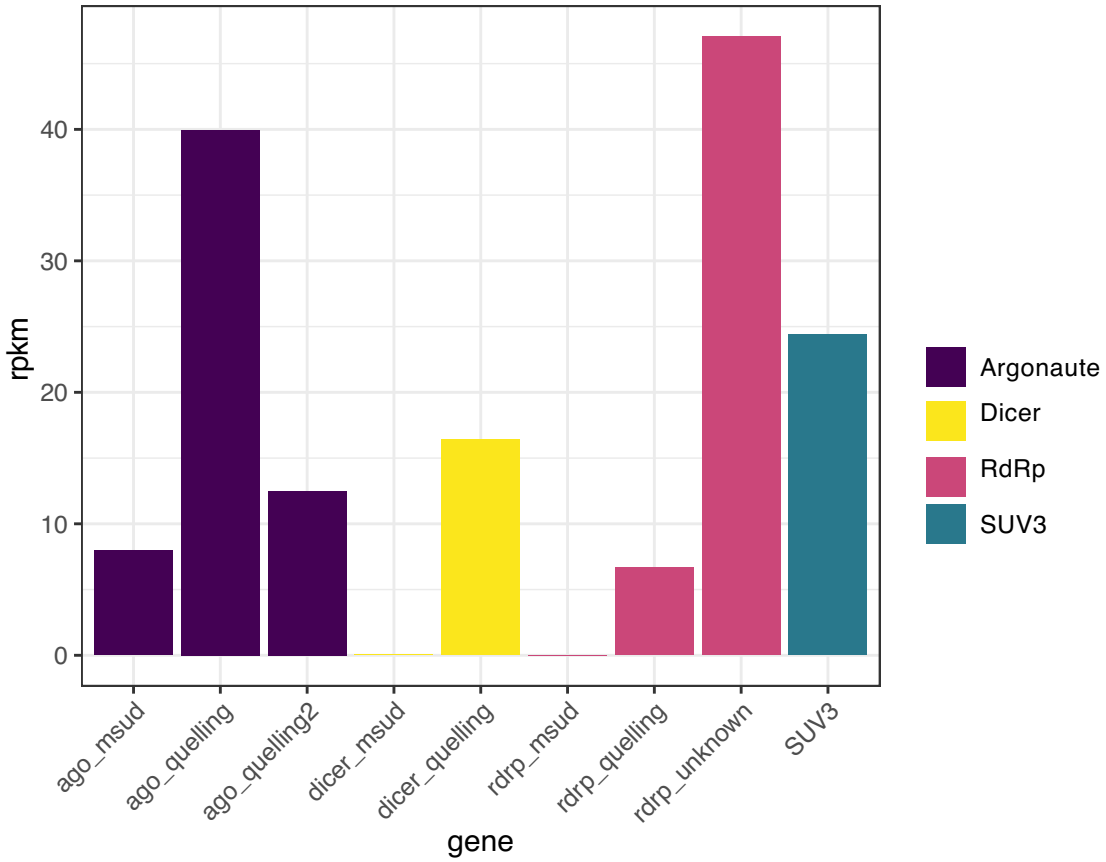


Figure 5. Phylogenetic and domain analyses of Argonaute proteins using maximum likelihood method with 1000 bootstrap replicates. The green diamond represents the proteins identified in *Trichoderma harzianum*. The blue and yellow clades represent MSUD and Quelling respectively.

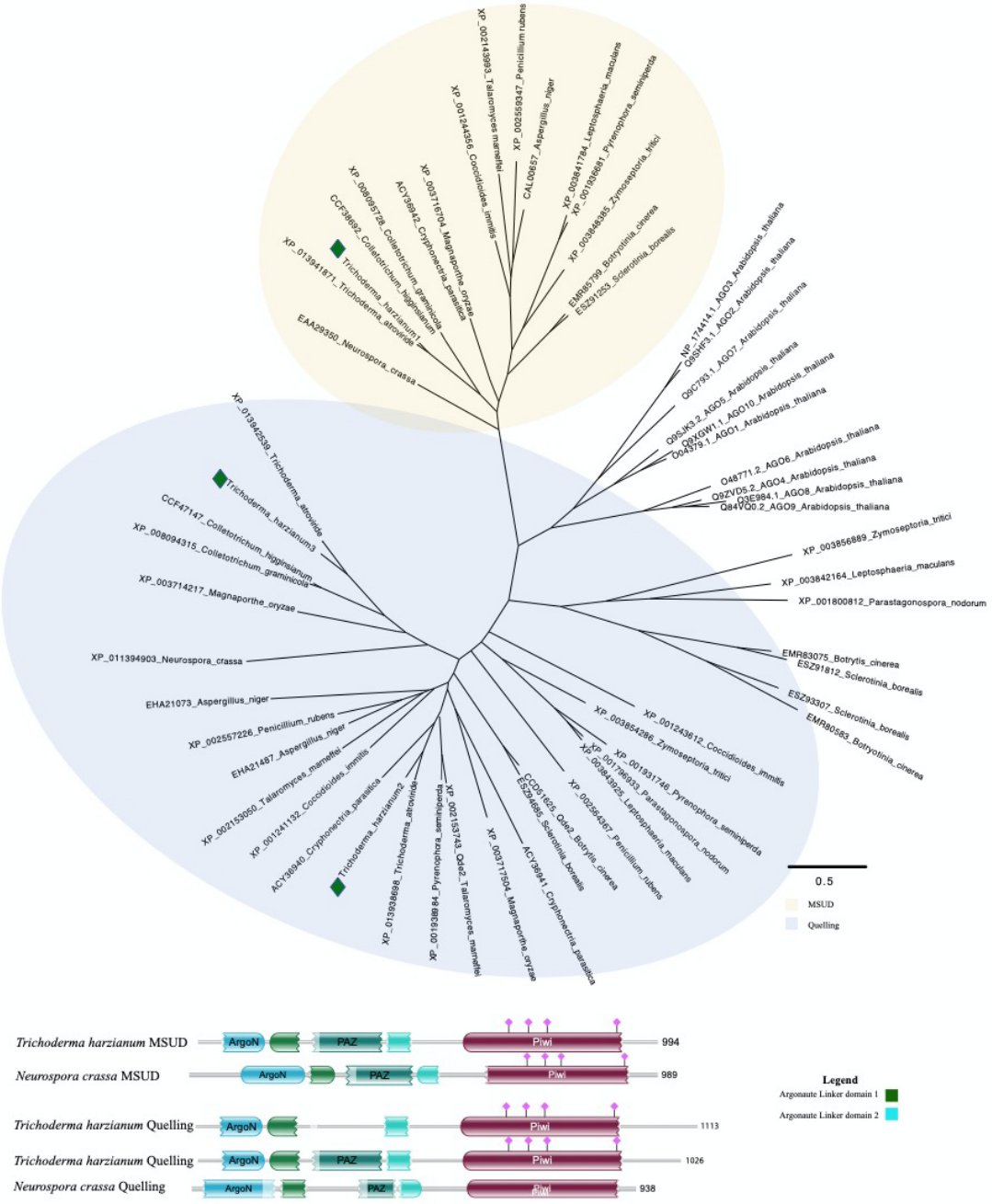


Figure 6. Phylogenetic and domain analyses of Dicer proteins using maximum likelihood method with 1000 bootstrap replicates. The green diamond represents the proteins identified in *Trichoderma harzianum*. The blue and yellow clades represent MSUD and Quelling respectively.

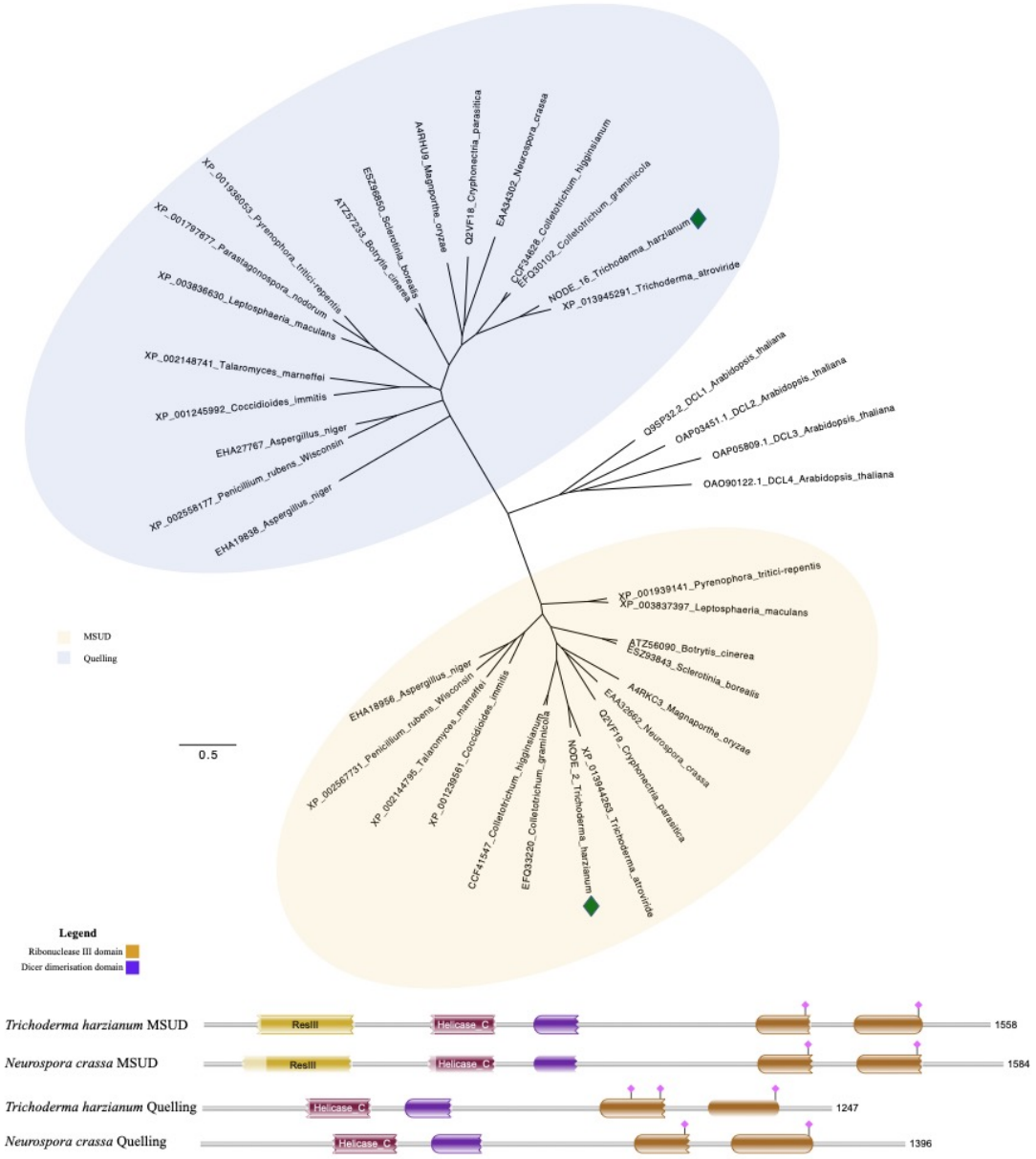
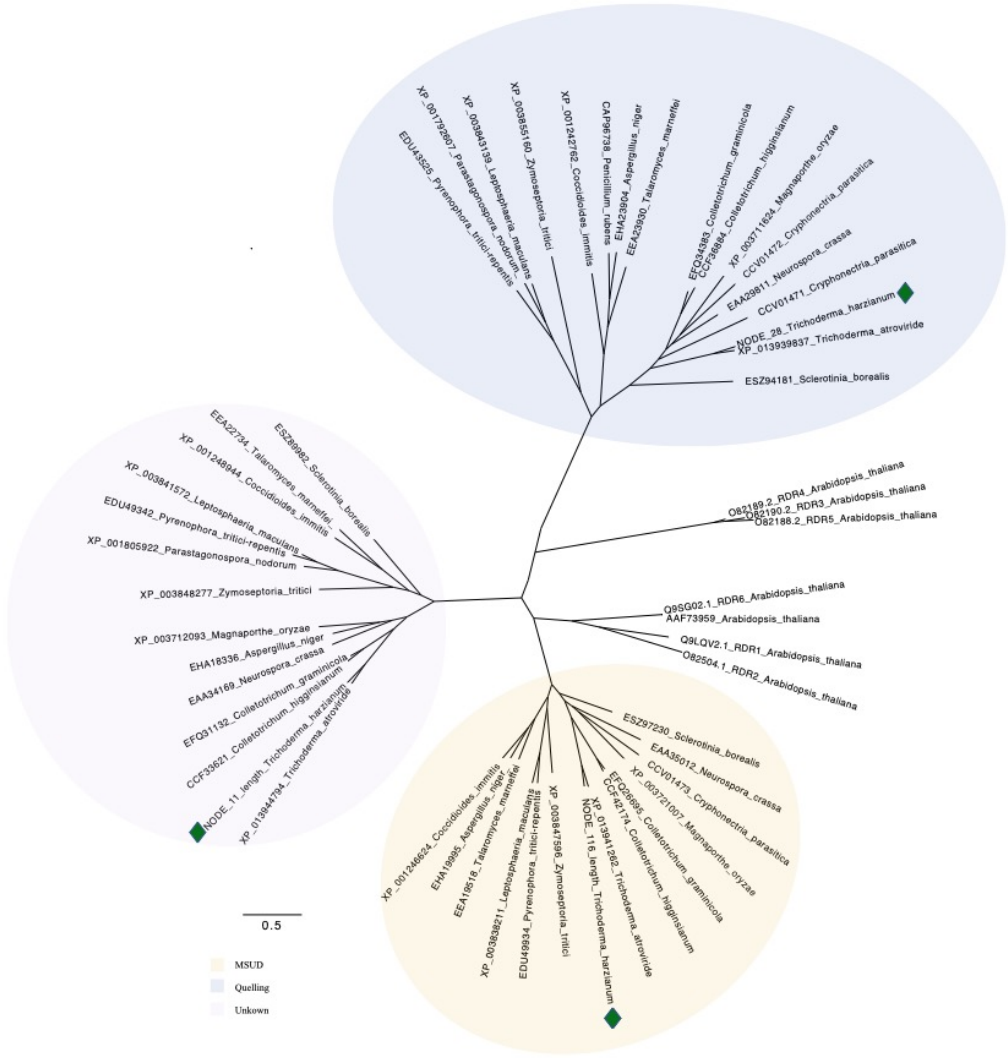


Figure 7. Phylogenetic and domain analyses of RNA-dependent RNA polymerase proteins using maximum likelihood method with 1000 bootstrap replicates. The green diamond represents the proteins identified in *Trichoderma harzianum*. The blue, yellow and pink clades represent MSUD, Quelling and unknown class respectively.



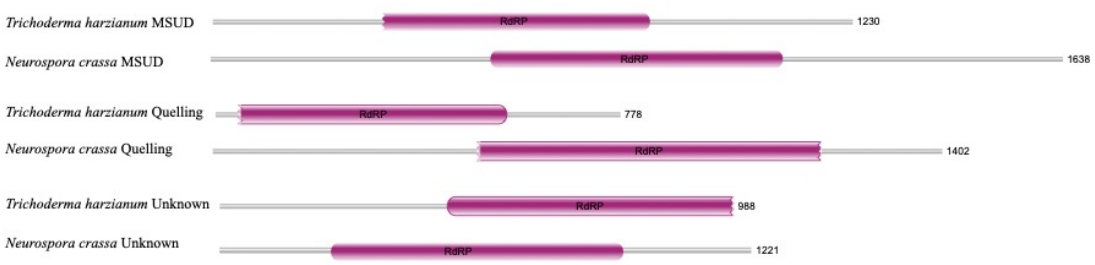


Figure 8. Analysis of sequencing of sRNAs mapped in the mitogenome of *Trichoderma harzianum* HB324. (A) Profile of sRNAs mapped in the mitogenome and presence of a peak between 20-22 nt. The colors represent the preference of bases at the 5` end of the small RNAs. (B) Sequences of sRNAs mapped in the mitogenome. Sense sequences (represented in blue) above the y axis and antisense sequences (represented in brown) below the y axis.

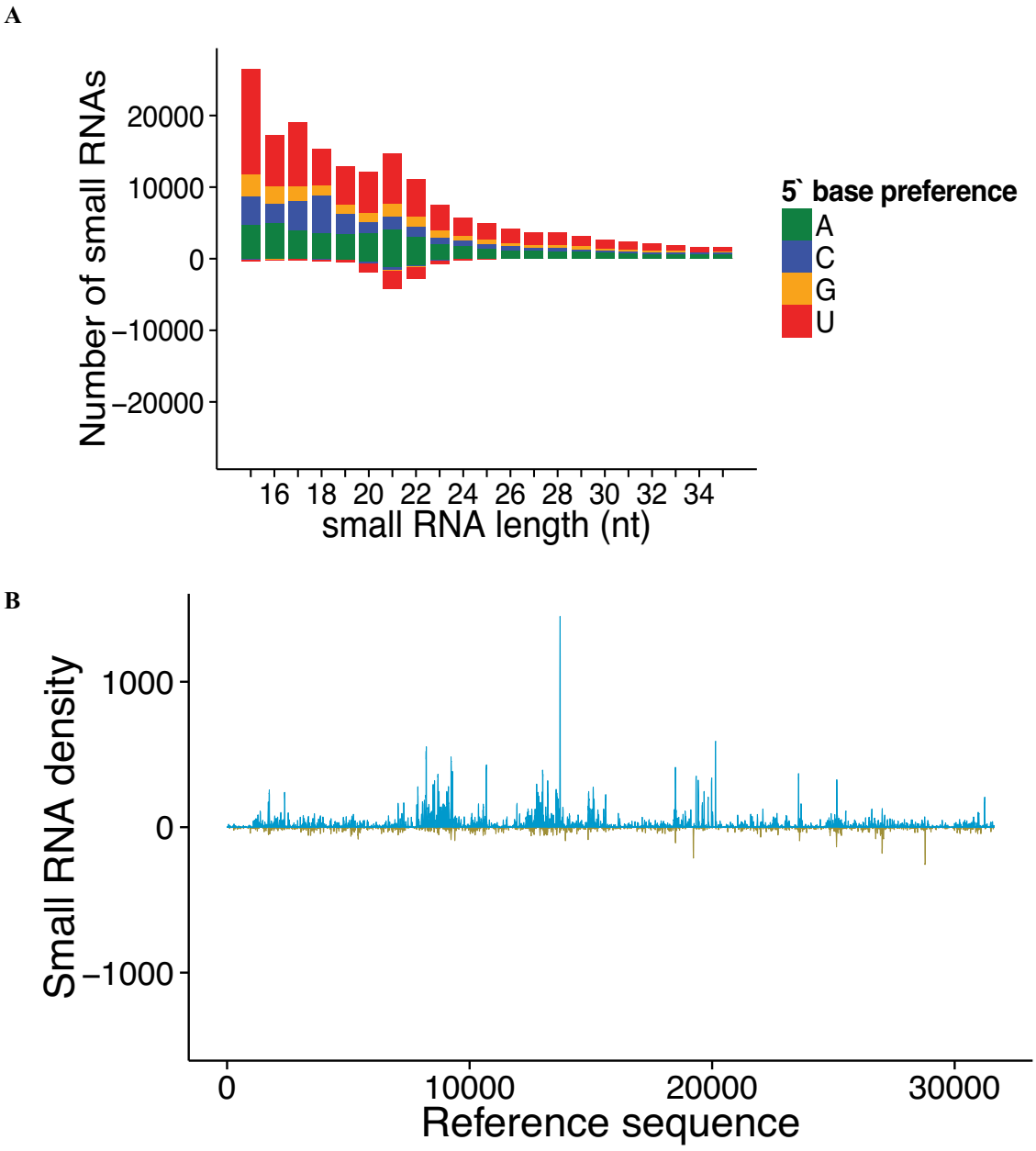


Figure 9. Analysis of the number of sRNAs reads mapped in each feature of the *Trichoderma harzianum* mitogenome. The values are normalized by RPM.

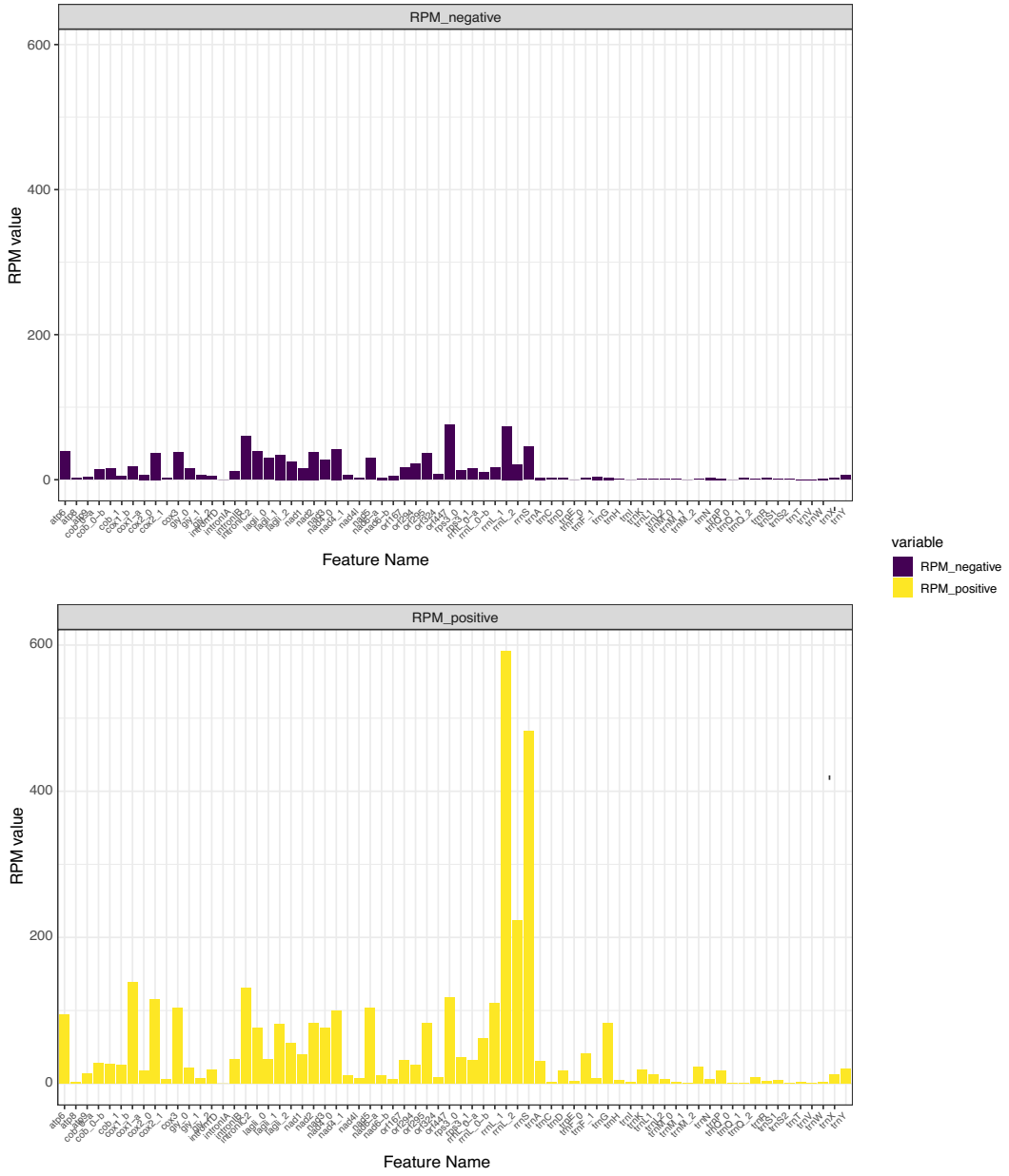


Figure 10. Profile of sRNAs mapped in the (A) *Trichoderma atroviride* and (B) *Fusarium graminearum* mitogenomes. The colors represent the preference of bases at the 5` end of the small RNAs. Sense sequences (represented in blue) above the y axis and antisense sequences (represented in brown) below the y axis.

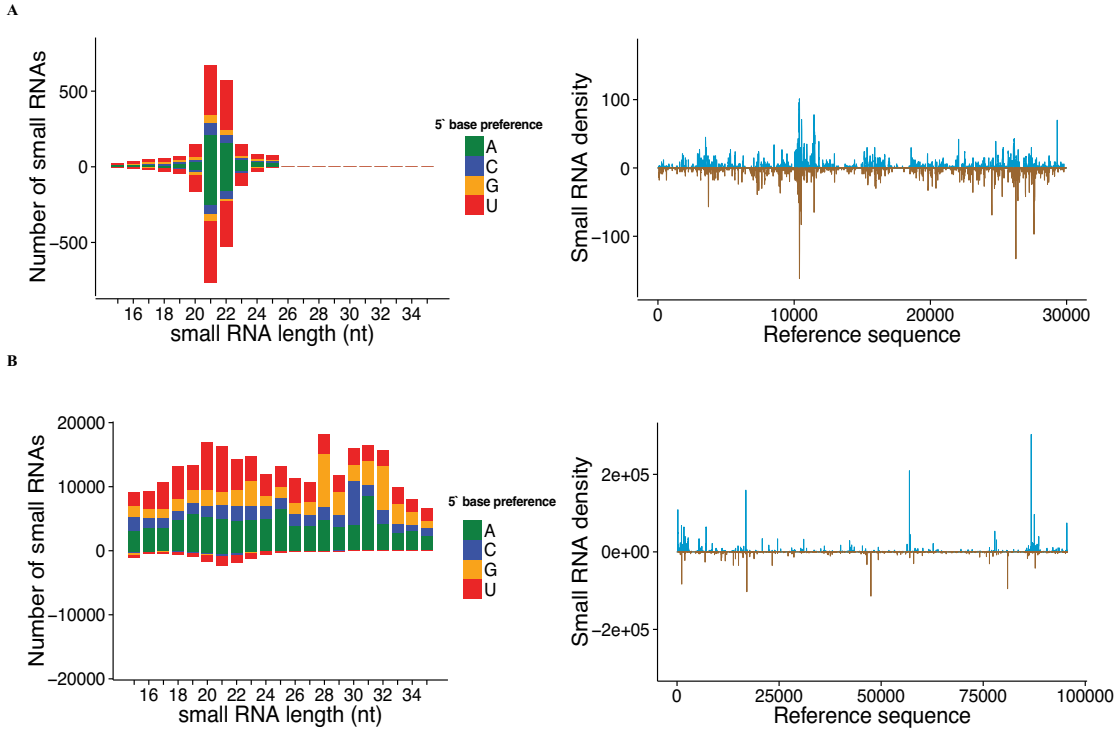
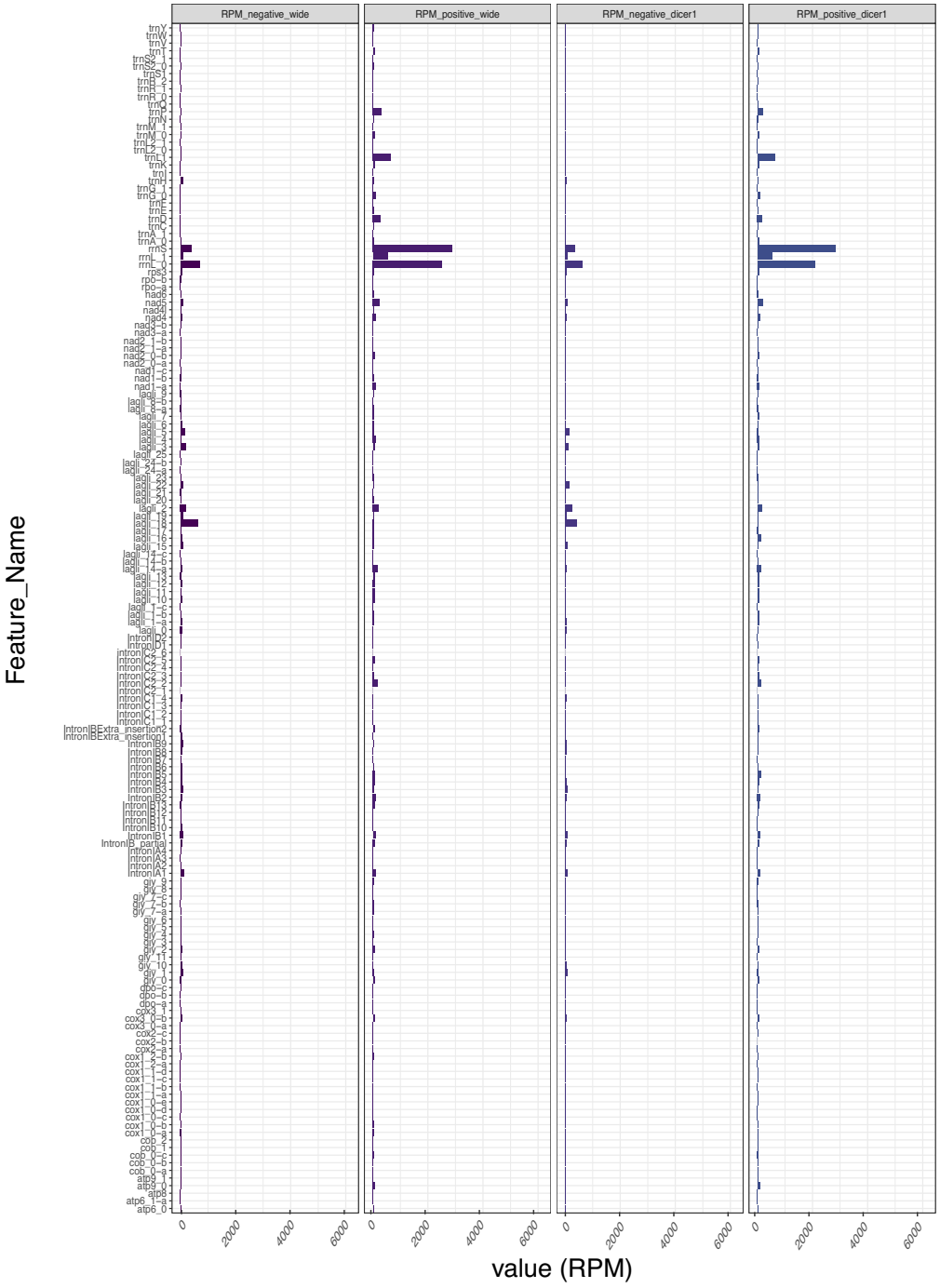
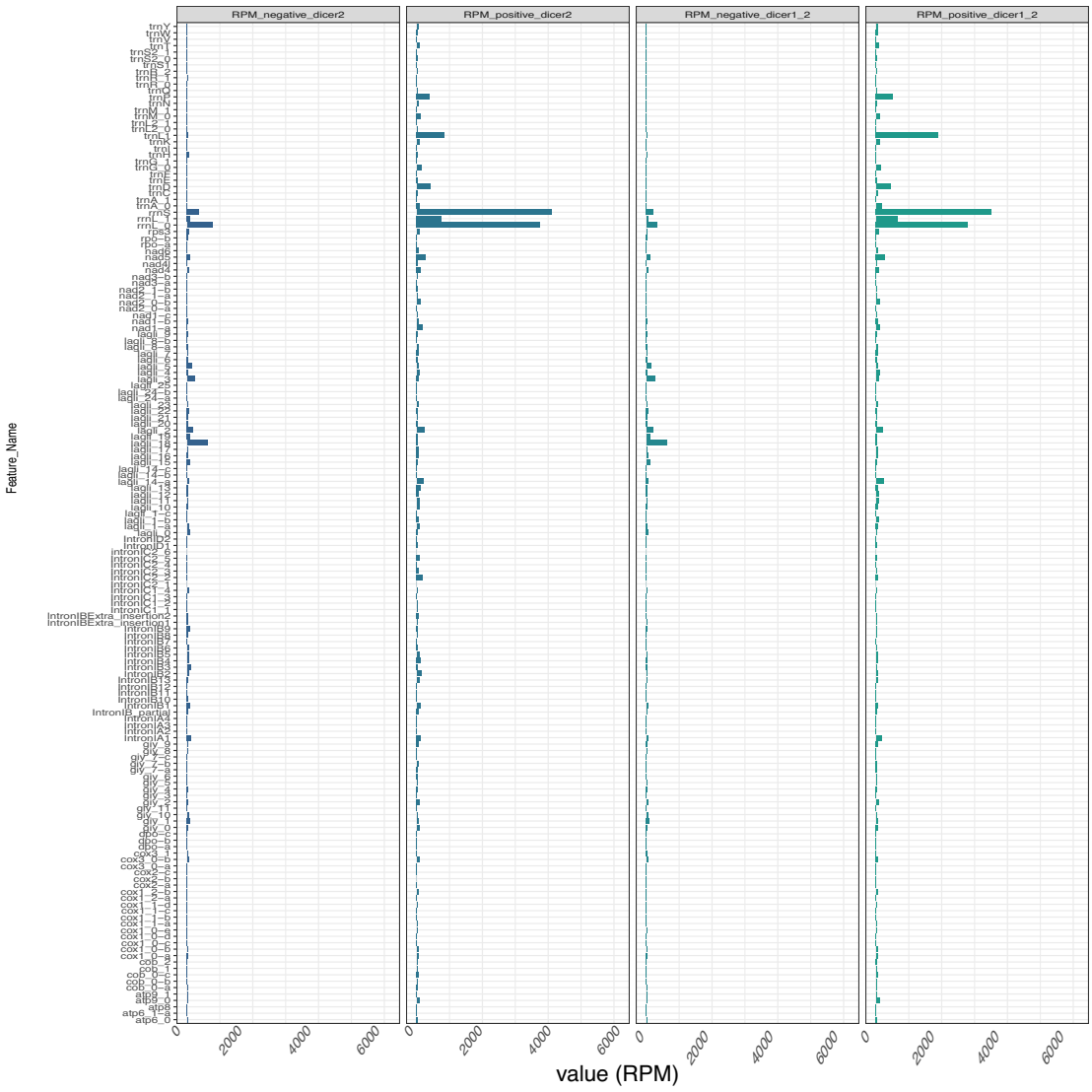
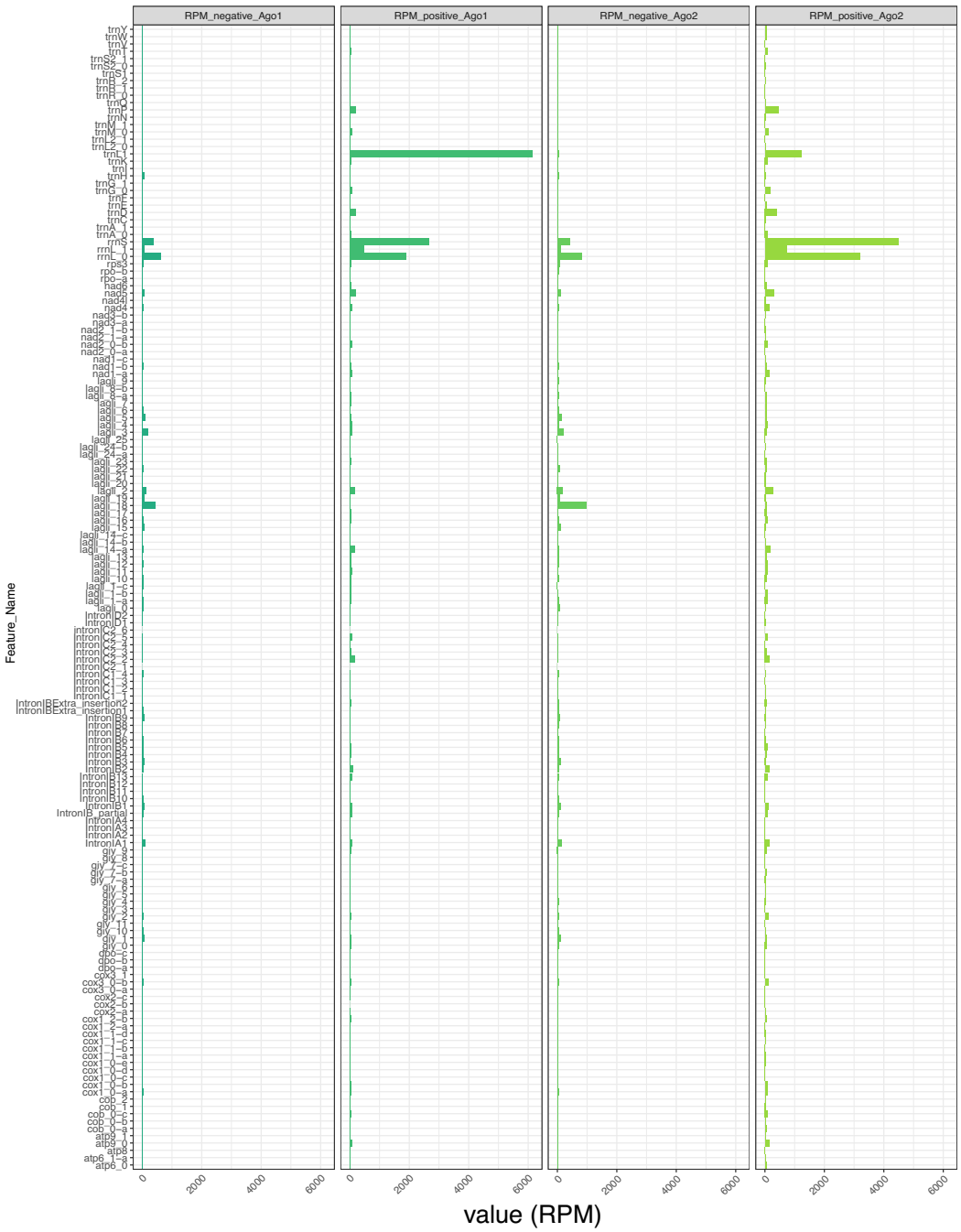


Figure 11. Analysis of the number of sRNAs reads mapped in each feature of the *Fusarium graminearum* mitogenome in knockout libraries for the Argonaute and Dicer genes. The values are normalized by RPM.




RPM_negative_dicer2
RPM_positive_dicer2
RPM_negative_dicer1_2
RPM_positive_dicer1_2
Feature_Name
value (RPM)
0
2000
4000
6000


RPM_negative_Ago1
RPM_positive_Ago1
RPM_negative_Ago2
RPM_positive_Ago2
Feature_Name
value (RPM)
0
2000
4000
6000

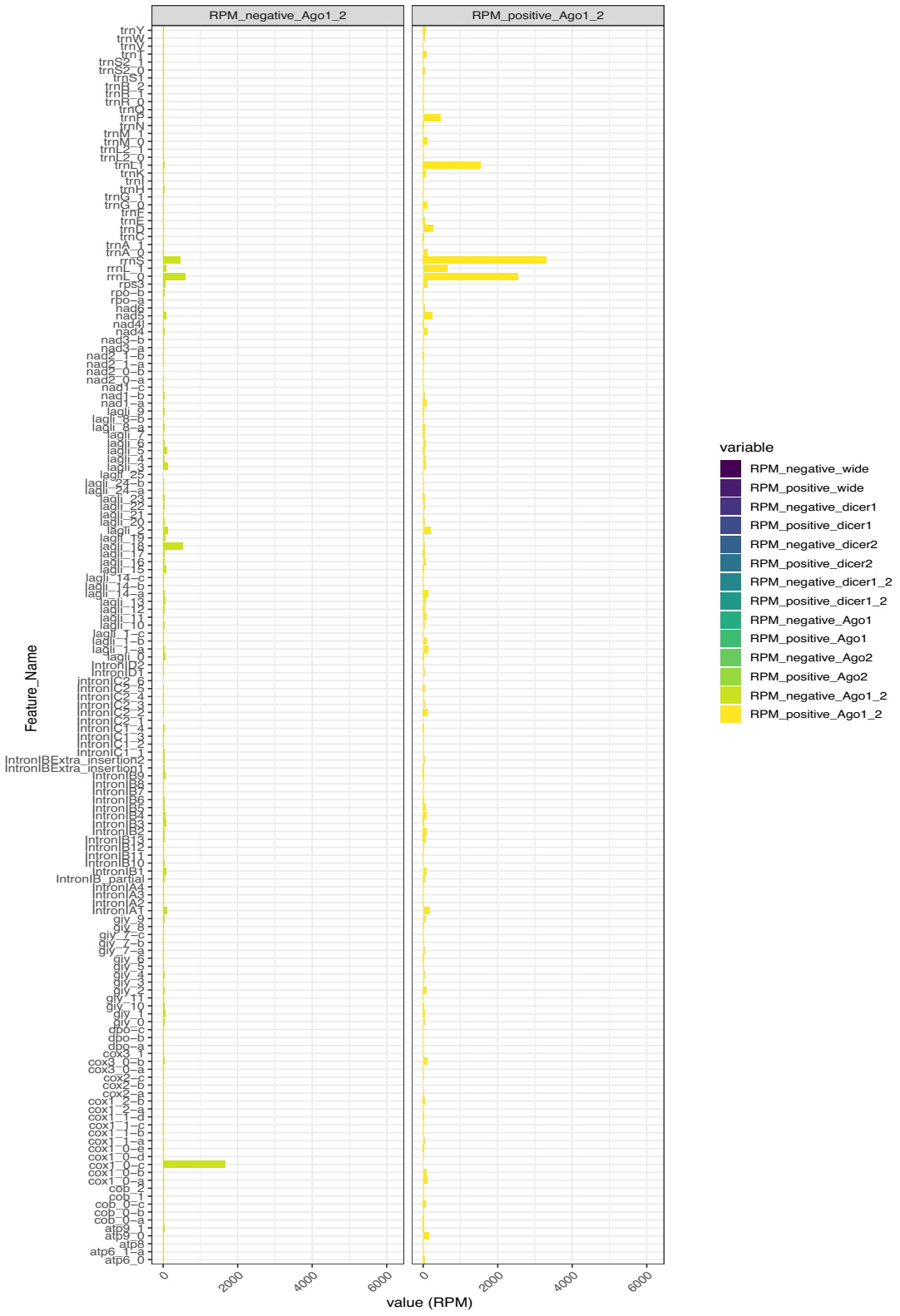
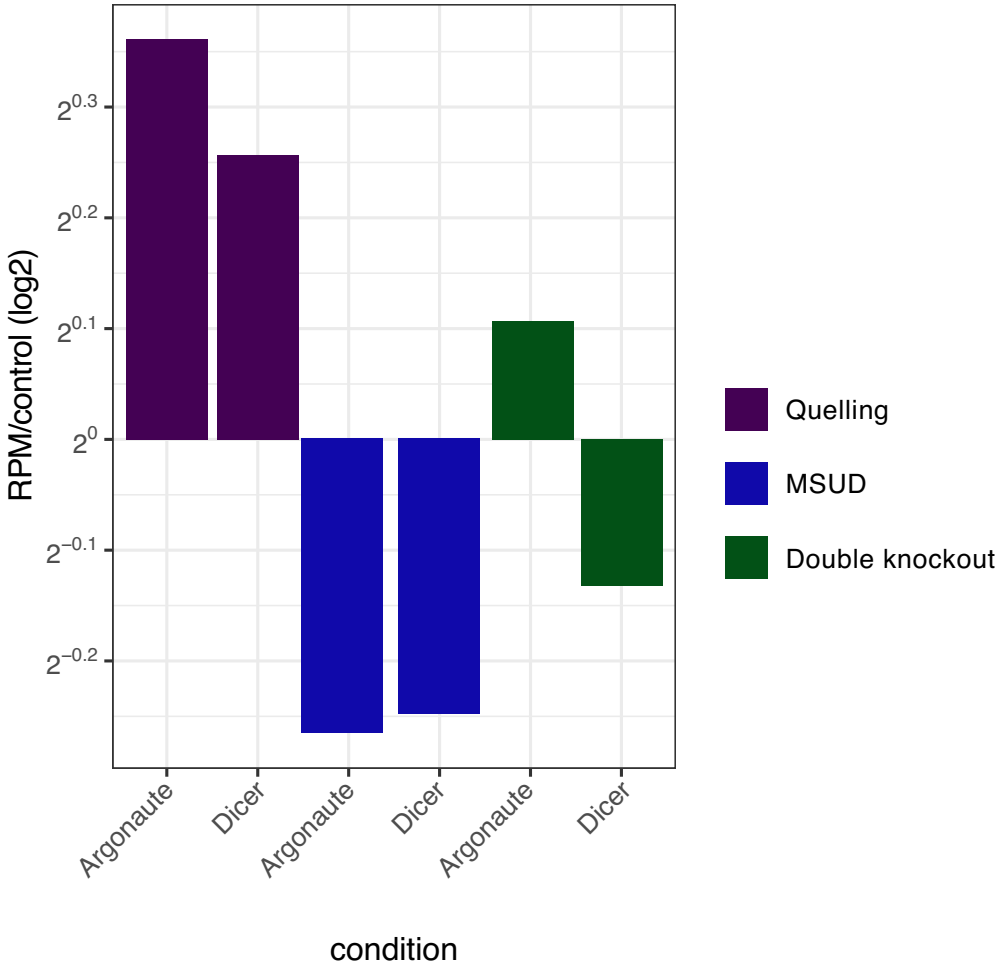
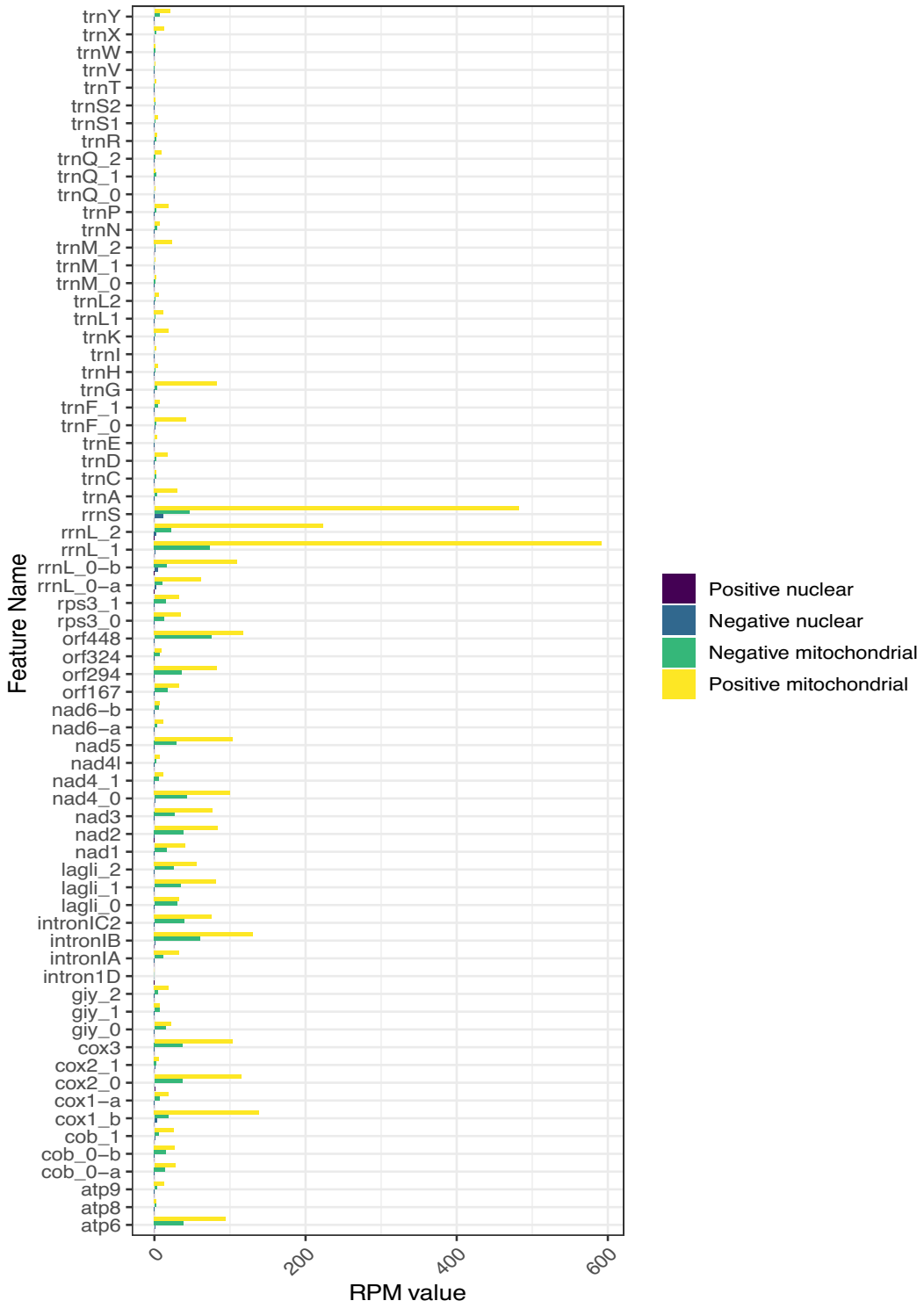

RPM_negative_Ago1_2
RPM_positive_Ago1_2
Feature_Name
value (RPM)
0
2000
4000
6000
variable
RPM_negative_wide
RPM_positive_wide
RPM_negative_dicer1
RPM_positive_dicer1
RPM_negative_dicer2
RPM_positive_dicer2
RPM_negative_dicer1_2
RPM_positive_dicer1_2
RPM_negative_Ago1
RPM_positive_Ago1
RPM_negative_Ago2
RPM_positive_Ago2
RPM_negative_Ago1_2
RPM_positive_Ago1_2

Figure 12. Reads mapped into *T. harzianum* mitogenome from small RNA libraries. The values were normalized by RPM divided by control library and put in log2. The Quelling, MSUD and double knockout are classified by different colors.



Supplementary Figure 1. Analysis of the number of sRNAs reads mapped in each feature of the *Trichoderma harzianum* mitogenome.



DISCUSSÃO INTEGRADA

Genomas mitocondriais de fungos (mitogenomas ou mtDNA) são altamente variáveis em tamanho e conteúdo gênico devido a presença de região intergênica, introns, ORFs sem função definida (uORFs) e eventos de recombinação (Hausner 2003; Pogoda et al. 2019; Kulik et al. 2021). Adicionalmente, íntrons mitocondriais podem ser *self-splicing*, podendo ser excisados e inseridos em outra região do mtDNA através de enzimas com função de *homing endonuclease* ou por fatores do próprio genoma hospedeiro que ainda não são bem conhecidos (Hausner 2003; Kulik et al. 2021). Geralmente, mitogenomas codificam um grupo de 14 genes codificadores de proteínas envolvidas na forsforilação oxidativa, dois genes ribossomais (subunidade menor e maior) e um gene codificador de proteína ribossomal (Hausner 2003; Kulik et al. 2021).

No capítulo um dessa tese, foi investigado o perfil de variação de tamanho e conteúdo de mitogenomas na ordem Hypocreales. Essa ordem é conhecida por ser uma das mais diversas e ricas do reino Fungi, desempenhando inúmeras funções ecológicas e com capacidade de adaptação rápida em diferentes ambientes. Nesse capítulo, foram avaliados 35 mitogenomas, sendo o mtDNA de *T. harzianum* apresentado pela primeira vez. Nas análises genômicas comparativas, foi possível identificar que a maioria das espécies apresentam o conjunto dos 17 genes *core* (genes codificadores de proteínas e ribossomais). Entretanto, a variação de tamanho dos mitogenomas foi correlacionado positivamente com o tamanho das regiões não-codificantes. Análises evolutivas indicaram que as espécies cujo mitogenomas possuem uma maior taxa de mutações sinônimas (evolução mais rápida) tendem a acumular um maior número e tamanho das regiões não-codificadoras. Muitos estudos demonstraram que a aquisição de regiões não-codificantes é devido a eventos de transferência horizontal entre espécies (Wu and Hao 2014; Kanzi et al. 2016; Kolesnikova et al. 2019), entretanto, neste capítulo demonstramos que grande parte da região codificante é correlacionada com a taxa de evolução dos mitogenomas. Neste capítulo também foram identificadas sequências de genes mitocondriais duplicados no genoma nuclear (NUMT). Além disso, a maioria dos genes *core* foi encontrada em, pelo menos, um genoma nuclear, com exceção dos genes *atp8*, *atp9* e *cox3*.

Mitogenomas são conservados e podem apresentar mais de uma cópia por célula. Por essa razão, alguns trabalhos vêm sendo desenvolvidos utilizando-se genes mitocondriais para a identificação molecular de espécies (Vialle et al. 2009; Kulik et al.

2015, 2020). Em mamíferos, por exemplo, o gene *cox1* é considerado o gene de escolha para uso como DNA *barcode* (Hebert et al. 2003). Em fungos, alguns estudos já sugeriram o uso do gene *cox1* e *cob* para alguns gêneros, apresentando bons resultados na separação e diferenciação de espécies. No entanto, esses dois genes apresentaram o maior número de íntrons na ordem Hypocreales de acordo com a análise realizada no capítulo um. Em contrapartida, o gene da subunidade menor do ribossomo, *rns*, foi o único gene que apresentou um único exon em todas as espécies analisadas, sugerindo que este gene poderia ser um candidato a biomarcador para a identificação de espécies da ordem (Fonseca et al. 2020). Por essa razão, no capítulo dois, testou-se o uso do gene *rns* poderia para a identificação de espécies e utilizando o gênero *Trichoderma* como modelo de estudo. Os resultados encontrados foram comparados com a identificação por outras três regiões, ITS, calmodulina e fator de elongação. Conforme os dados obtidos, a identificação específica realizada pela amplificação por apenas uma das regiões não é suficiente, mas a combinação das regiões *rns*, calmodulina e ITS apresenta uma ótima estratégia para a identificação correta das espécies.

Devido a variação na composição gênica, e a presença de sequências de NUMTs no genoma nuclear, o controle, manutenção, replicação e transcrição de mitogenomas deve ser realizado por uma maquinaria coordenada e eficiente para o funcionamento correto e, consequentemente produção de energia (Mosbach et al. 2017; Tang et al. 2018). Contudo, grande parte do conhecimento sobre os genes e maquinaria envolvidos no controle da organela e genoma são baseados no conhecimento desenvolvido em leveduras e poucos na espécie-modelo de fungos filamentosos *Neurospora crassa* (Kulik et al. 2021). Por essa razão, no capítulo três, foi realizado a identificação de genes possivelmente envolvidos na replicação, transcrição, tradução, produção de energia e manutenção da morfologia e transferência de moléculas entre o meio interno (matriz intracelular da mitocôndria) e externo (matriz extracelular - citoplasma). Para confirmar os dados encontrados, é necessário ainda um estudo aprofundado e análises de proteoma para identificar se os genes são traduzidos e se possuem sítio de ação no interior da organela mitocondrial. Essa etapa ainda será desenvolvida posteriormente.

No capítulo quatro, é apresentado o perfil transcricional do mitogenoma de *T. harzianum*. Inicialmente, foi realizado o sequenciamento de pequenos RNAs, e a partir dos resultados encontrados foi possível identificar sequências de RNA proveniente das duas fitas (senso e anti-senso). Em mamíferos, já foi descrito que o mtDNA está sujeito a transcrição bidirecional, sendo capaz de gerar RNA dupla fita (dsRNA) (Dhir et al.

2018). Para confirmar a geração desse dsRNA, foi realizado o sequenciamento de RNA total utilizando a metodologia de enriquecimento para sequências poliadeniladas e fita específica. Os resultados encontrados sugerem que o RNA gerado na transcrição não é poliadenilado, sugerindo novamente a presença de dsRNA. A amplificação fita específica do gene *rns* confirmou a presença de dsRNA proveniente do mitogenoma, sabe-se que o acúmulo de dsRNA mitocondrial pode provocar alterações no funcionamento da mitocôndria. Em mamíferos, já foi descrito a presença do complexo de degradossomo mitocondrial que possui a finalidade de cortar o dsRNA em fragmentos menores. A ausência desse complexo resulta no acúmulo de dsRNA no interior da matriz mitocondrial com consequente evasão para o citoplasma (Rongvaux et al. 2014; Dhir et al. 2018).

Neste último capítulo, também foi realizada a identificação e caracterização dos principais genes envolvidos na via de RNA de interferência (RNAi) e o gene SUV3, único gene do degradossomo mitocondrial identificado em *T. harzianum*. Todos os genes foram quantificados de acordo com a sua expressão nas bibliotecas sequenciadas de transcritoma. De forma geral, sugerimos que a via de RNAi e o gene SUV3 sejam responsáveis pelo controle transcricional e de dsRNA nos mitogenomas. Todavia, é necessário a realização de experimentos de silenciamento ou nocaute para confirmação dessa hipótese.

Em conclusão, esta tese contribui para o conhecimento da diversidade de mitogenomas fúngicos e o seu controle regulatório nas células. Mitogenomas são essenciais para a maioria das espécies de fungos na geração de energia e, sem o controle adequado, pode ocasionar a não sobrevivência das espécies no ambiente. Por essa razão, é de extrema importância os estudos de caracterização, regulação e funcionamento desses genomas. A presente tese, portanto, forneceu novos conhecimentos além dos já descritos na literatura sobre esse tema.

CONCLUSÃO

O estudo de caracterização e transcrição do genoma mitocondrial de fungos ainda é escasso. Além disso, mitogenomas fúngicos apresentam uma riqueza em relação ao tamanho, composição e elementos core. A partir dos resultados encontrados neste trabalho, foi possível alcançar as seguintes conclusões:

- A mitogenômica comparativa de fungos da ordem Hypocreales mostrou uma grande variedade de tamanho e composição, principalmente de elementos não-codificantes. Muitos desses elementos, são compartilhados entre as espécies e podem apresentar a mesma localização de inserção. Apesar da variação do conteúdo, o ordenamento dos genes *core* é conservado entre as espécies;
- A análise filogenética baseada no tempo de divergência revelou que a ordem evoluiu há aproximadamente 137.39 Mya. Além disso, a correlação positiva entre o tamanho total de elementos não codificantes (introns, HEGs e uORFs) com a taxa de mutações sinônimas indicaram que os mitogenomas que evoluem mais rápido tendem a acumular modificações na região não codificante. Em contra-partida, não há correlação entre a taxa de mutações sinônimas e o tamanho total de genes codificadores de proteínas, indicando que esses genes estão sobre influência de seleção negativa;
- O gene da subunidade menor ribossomal *rns* pode ser utilizado como um biomarcador associado ao gene da calmodulina e o da região ITS para a identificação à nível de espécies do gênero *Trichoderma*;
- Cinco grupos de genes codificadores de proteínas mitocondriais identificados através da montagem e anotação do genoma nuclear de *Trichoderma harzianum* HB324 foram encontrados e descritos neste trabalho. A identificação desses genes auxiliou na compreensão do controle da replicação e transcrição de mitogenomas. Além disso, 39 genes sem similaridade foram identificados, sugerindo que muitos genes ainda não tiveram a sua função descrita;
- A análise de pequenos RNAs indicou que a via de RNA de interferência tenha papel fundamental no controle transcricional dos mitogenomas e detectou-se a presença de geração de RNA dupla fita (dsRNA) proveniente de genes mitocondriais. A comprovação da geração de dsRNA foi feita através de amplificação fita-específica no gene *rns*.

- Transcritos provenientes da mitocôndria não são poliadenilados, sugerindo que devem sofrer um processamento durante a tradução;
- Os dados obtidos neste trabalho são promissores na compreensão do funcionamento de mitogenomas e pode auxiliar no desenvolvimento de pesquisas futuras na área de biotecnologia e investigação de novas drogas antifúngicas.

PERSPECTIVAS

- Diante dos resultados descritos neste trabalho, iremos expandir a análise de caracterização de mitogenomas de fungos. Esta etapa já está em andamento e os resultados dessa comparação serão publicados no jornal Frontiers in Microbiology;
- Para melhor entender e descrever os genes envolvidos no controle do genoma e da morfologia da mitocôndria, iremos realizar um estudo de proteoma para a identificação de proteínas e sua localização na organela;
- A partir dos resultados obtidos pela análise das bibliotecas de pequenos e longos RNAs, iremos realizar o estudo da função da proteína SUV3 para verificar a relação da proteína com a degradação de RNA dupla fita. Além disso, iremos certificar se a via de RNA de interferência está atuando no controle do genoma mitocondrial de fungos e avaliar a geração de dsRNA mitocondrial durante o crescimento do fungo.

REFERÊNCIAS

Avin FA, Subha B, Tan Y-S, et al (2017) Escaping introns in COI through cDNA barcoding of mushrooms: *Pleurotus* as a test case. Ecol Evol 7:6972–6980. https://doi.org/10.1002/ece3.3049

Barzel A, Obolski U, Gogarten J, et al (2011) Home and away- the evolutionary dynamics of homing endonucleases. BMC Evol Biol 11:324. https://doi.org/10.1186/1471-2148-11-324

Benítez T, Rincón AM, Limón MC, Codón AC (2004) Biocontrol mechanisms of Trichoderma strains. Int Microbiol 7:249–260

Billmyre RB, Calo S, Feretzaki M, et al (2013) RNAi function, diversity, and loss in the fungal kingdom. Chromosome Res 21:561–572. https://doi.org/10.1007/s10577-013-9388-2

Chen P-L, Chen C-F, Chen Y, et al (2013) Mitochondrial genome instability resulting from SUV3 haploinsufficiency leads to tumorigenesis and shortened lifespan. Oncogene 32:1193–1201. https://doi.org/10.1038/onc.2012.120

Craig, Craigie, Gellert, Lambowitz (eds) (2002) Mobile Introns: Pathways and Proteins. In: Mobile DNA II. American Society of Microbiology, pp 761–783

Dang Y, Yang Q, Xue Z, Liu Y (2011a) RNA Interference in Fungi: Pathways, Functions, and Applications. Eukaryotic Cell 10:1148–1155. https://doi.org/10.1128/EC.05109-11

Dang Y, Yang Q, Xue Z, Liu Y (2011b) RNA interference in fungi: pathways, functions, and applications. Eukaryot Cell 10:1148–1155. https://doi.org/10.1128/EC.05109-11

Decker LM, Boone EC, Xiao H, et al (2015) Complex formation of RNA silencing proteins in the perinuclear region of Neurospora crassa. Genetics 199:1017–1021. https://doi.org/10.1534/genetics.115.174623

Dhir A, Dhir S, Borowski LS, et al (2018) Mitochondrial double-stranded RNA triggers antiviral signalling in humans. Nature 560:238–242. https://doi.org/10.1038/s41586-018-0363-0

Druzhinina IS, Seidl-Seiboth V, Herrera-Estrella A, et al (2011) Trichoderma: the genomics of opportunistic success. Nat Rev Microbiol 9:749–759. https://doi.org/10.1038/nrmicro2637

Fonseca PLC, Badotti F, De-Paula RB, et al (2020) Exploring the Relationship Among Divergence Time and Coding and Non-coding Elements in the Shaping of Fungal Mitochondrial Genomes. Front Microbiol 11:765. https://doi.org/10.3389/fmicb.2020.00765

Friedman JR, Nunnari J (2014) Mitochondrial form and function. Nature 505:335–343. https://doi.org/10.1038/nature12985

Ghildiyal M, Zamore PD (2009) Small silencing RNAs: an expanding universe. Nat Rev Genet 10:94–108. https://doi.org/10.1038/nrg2504

Gray MW (2017) Lynn Margulis and the endosymbiont hypothesis: 50 years later. Mol Biol Cell 28:1285–1287. https://doi.org/10.1091/mbc.E16-07-0509

Gruber S, Seidl-Seiboth V (2012) Self versus non-self: fungal cell wall degradation in Trichoderma. Microbiology (Reading) 158:26–34. https://doi.org/10.1099/mic.0.052613-0

Guo XE, Chen C-F, Wang DD-H, et al (2011) Uncoupling the roles of the SUV3 helicase in maintenance of mitochondrial genome stability and RNA degradation. J Biol Chem 286:38783–38794. https://doi.org/10.1074/jbc.M111.257956

Hannon GJ (2002) RNA interference. Nature 418:244–251. https://doi.org/10.1038/418244a

Hausner G (2003) Fungal Mitochondrial Genomes, Plasmids and Introns. In: Applied Mycology and Biotechnology. Elsevier, pp 101–131

Hawksworth, Lücking (2017) Fungal Diversity Revisited: 2.2 to 3.8 Million Species. In: The Fungal Kingdom. American Society of Microbiology, pp 79–95

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B 270:313–321. https://doi.org/10.1098/rspb.2002.2218

Hermosa R, Cardoza RE, Rubio MB, et al (2014) Secondary Metabolism and Antimicrobial Metabolites of Trichoderma. In: Biotechnology and Biology of Trichoderma. Elsevier, pp 125–137

Holoch D, Moazed D (2015) RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet 16:71–84. https://doi.org/10.1038/nrg3863

Kanzi AM, Wingfield BD, Steenkamp ET, et al (2016) Intron Derived Size Polymorphism in the Mitochondrial Genomes of Closely Related Chrysoporthe Species. PLoS ONE 11:e0156104. https://doi.org/10.1371/journal.pone.0156104

Kolesnikova AI, Putintseva YA, Simonov EP, et al (2019) Mobile genetic elements explain size variation in the mitochondrial genomes of four closely-related Armillaria species. BMC Genomics 20:351. https://doi.org/10.1186/s12864-019-5732-z

Kolondra A, Labedzka-Dmoch K, Wenda JM, et al (2015) The transcriptome of Candida albicans mitochondria and the evolution of organellar transcription units in yeasts. BMC Genomics 16:827. https://doi.org/10.1186/s12864-015-2078-z

Kulik T, Bilska K, Żelechowski M (2020) Promising Perspectives for Detection, Identification, and Quantification of Plant Pathogenic Fungi and Oomycetes through Targeting Mitochondrial DNA. IJMS 21:2645. https://doi.org/10.3390/ijms21072645

Kulik T, Ostrowska A, Buśko M, et al (2015) Development of an FgMito assay: A highly sensitive mitochondrial based qPCR assay for quantification of Fusarium graminearum sensu stricto. International Journal of Food Microbiology 210:16–23. https://doi.org/10.1016/j.ijfoodmicro.2015.06.012

Kulik T, Van Diepeningen AD, Hausner G (2021) Editorial: The Significance of Mitogenomics in Mycology. Front Microbiol 11:628579. https://doi.org/10.3389/fmicb.2020.628579

Lee H-C, Li L, Gu W, et al (2010) Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. Mol Cell 38:803–814. https://doi.org/10.1016/j.molcel.2010.04.005

Martijn J, Vosseberg J, Guy L, et al (2018) Deep mitochondrial origin outside the sampled alphaproteobacteria. Nature 557:101–105. https://doi.org/10.1038/s41586-018-0059-5

Mosbach A, Edel D, Farmer AD, et al (2017) Anilinopyrimidine Resistance in Botrytis cinerea Is Linked to Mitochondrial Function. Front Microbiol 8:2361. https://doi.org/10.3389/fmicb.2017.02361

Mukherjee M, Mukherjee PK, Horwitz BA, et al (2012) Trichoderma-plant-pathogen interactions: advances in genetics of biological control. Indian J Microbiol 52:522–529. https://doi.org/10.1007/s12088-012-0308-5

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

Nakayashiki H, Kadotani N, Mayama S (2006) Evolution and diversification of RNA silencing proteins in fungi. J Mol Evol 63:127–135. https://doi.org/10.1007/s00239-005-0257-2

Nicolás FE, Ruiz-Vázquez RM (2013) Functional diversity of RNAi-associated sRNAs in fungi. Int J Mol Sci 14:15348–15360. https://doi.org/10.3390/ijms140815348

Nolan T, Cecere G, Mancone C, et al (2008) The RNA-dependent RNA polymerase essential for post-transcriptional gene silencing in Neurospora crassa interacts with replication protein A. Nucleic Acids Res 36:532–538. https://doi.org/10.1093/nar/gkm1071

Pogoda CS, Keepers KG, Nadiadi AY, et al (2019) Genome streamlining via complete loss of introns has occurred multiple times in lichenized fungal mitochondria. Ecol Evol 9:4245–4263. https://doi.org/10.1002/ece3.5056

Prasad RD, Rangeshwaran R, Hegde SV, Anuroop CP (2002) Effect of soil and seed application of Trichoderma harzianum on pigeonpea wilt caused by Fusarium udum under field conditions. Crop Protection 21:293–297. https://doi.org/10.1016/S0261-2194(01)00100-4

Razew M, Warkocki Z, Taube M, et al (2018) Structural analysis of mtEXO mitochondrial RNA degradosome reveals tight coupling of nuclease and helicase components. Nat Commun 9:97. https://doi.org/10.1038/s41467-017-02570-5

Rehner SA, Samuels GJ (1995) Molecular systematics of the Hypocreales: a teleomorph gene phylogeny and the status of their anamorphs. Can J Bot 73:816–823. https://doi.org/10.1139/b95-327

Roger AJ, Muñoz-Gómez SA, Kamikawa R (2017) The Origin and Diversification of Mitochondria. Curr Biol 27:R1177–R1192. https://doi.org/10.1016/j.cub.2017.09.015

Rongvaux A, Jackson R, Harman CCD, et al (2014) Apoptotic caspases prevent the induction of type I interferons by mitochondrial DNA. Cell 159:1563–1577. https://doi.org/10.1016/j.cell.2014.11.037

Rubinoff D, Holland BS (2005) Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. Systematic Biology 54:952–961. https://doi.org/10.1080/10635150500234674

Sagan L (1967) On the origin of mitosing cells. J Theor Biol 14:255–274. https://doi.org/10.1016/0022-5193(67)90079-3

Samuels GJ (2006) Trichoderma: systematics, the sexual state, and ecology. Phytopathology 96:195–206. https://doi.org/10.1094/PHYTO-96-0195

Schmoll M, Dattenböck C, Carreras-Villaseñor N, et al (2016) The Genomes of Three Uneven Siblings: Footprints of the Lifestyles of Three Trichoderma Species. Microbiol Mol Biol Rev 80:205–327. https://doi.org/10.1128/MMBR.00040-15

Sellem CH, d'Aubenton-Carafa Y, Rossignol M, Belcour L (1996) Mitochondrial intronic open reading frames in Podospora: mobility and consecutive exonic sequence variations. Genetics 143:777–788

Smith DR (2016) The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? Brief Funct Genomics 15:47–54. https://doi.org/10.1093/bfgp/elv027

Spinelli JB, Haigis MC (2018) The multifaceted contributions of mitochondria to cellular metabolism. Nat Cell Biol 20:745–754. https://doi.org/10.1038/s41556-018-0124-1

Srivastava M, Shahid M (2014) Trichoderma Genome to Genomics: A Review. J Data Mining Genomics Proteomics 05: https://doi.org/10.4172/2153-0602.1000162

Stepien PP, Margossian SP, Landsman D, Butow RA (1992) The yeast nuclear gene suv3 affecting mitochondrial post-transcriptional processes encodes a putative ATP-dependent RNA helicase. Proc Natl Acad Sci U S A 89:6813–6817. https://doi.org/10.1073/pnas.89.15.6813

Stoddard BL (2014) Homing endonucleases from mobile group I introns: discovery to genome engineering. Mob DNA 5:7. https://doi.org/10.1186/1759-8753-5-7

Szczesny RJ, Borowski LS, Brzezniak LK, et al (2010) Human mitochondrial RNA turnover caught in flagranti: involvement of hSuv3p helicase in RNA surveillance. Nucleic Acids Res 38:279–298. https://doi.org/10.1093/nar/gkp903

Tang G, Zhang C, Ju Z, et al (2018) The mitochondrial membrane protein FgLetm1 regulates mitochondrial integrity, production of endogenous reactive oxygen species and mycotoxin biosynthesis in Fusarium graminearum. Mol Plant Pathol 19:1595–1611. https://doi.org/10.1111/mpp.12633

Varshney D, Jaiswar A, Adholeya A, Prasad P (2016) Phylogenetic analyses reveal molecular signatures associated with functional divergence among Subtilisin like Serine Proteases are linked to lifestyle transitions in Hypocreales. BMC Evol Biol 16:220. https://doi.org/10.1186/s12862-016-0793-y

Vialle A, Feau N, Allaire M, et al (2009) Evaluation of mitochondrial genes as DNA barcode for Basidiomycota. Molecular Ecology Resources 9:99–113. https://doi.org/10.1111/j.1755-0998.2009.02637.x

Wang G, Chen H-W, Oktay Y, et al (2010) PNPASE regulates RNA import into mitochondria. Cell 142:456–467. https://doi.org/10.1016/j.cell.2010.06.035

Wijayawardene N (2020) Outline of Fungi and fungus-like taxa. Mycosphere 11:1060–1456. https://doi.org/10.5943/mycosphere/11/1/8

Wu B, Hao W (2014) Horizontal transfer and gene conversion as an important driving force in shaping the landscape of mitochondrial introns. G3 (Bethesda) 4:605–612. https://doi.org/10.1534/g3.113.009910

Zardoya R (2020) Recent advances in understanding mitochondrial genome diversity. F1000Res 9:270. https://doi.org/10.12688/f1000research.21490.1

ANEXOS

Anexo A - Artigos publicados em revistas internacionais (n=12)

- **Paula Fonseca,** Flavia Ferreira, Felipe da Silva, Liliane Santana Oliveira, João Trindade Marques, Aristóteles Goes-Neto, Eric Aguiar & Arthur Gruber. (2021) A novel mitovirus from the sand fly *Lutzomyia longipalpis* shows sRNA profiles consistent with siRNA pathway activation. *Viruses*, 13:9. https://doi.org/10.3390/v13010009.
- Vaz, A. B., **Fonseca, P. L.C.**, Silva, F. F., Quintanilha-Peixoto G., Sampedro, I., Siles, J. A., Carmo, A., Kato, R. B., Azevedo, V., Badotti, F., Ocampo, J. A., Rosa, C. A., Góes-Neto A. (2020) Foliar mycoendophytome of an endemic plant of the Mediterranean biome (*Myrtus communis*) reveals the dominance of basidiomycete woody saprotrophs. *PeerJ*, 8:e10487. https://doi.org/10.7717/peerj.10487.
- Dos Santos, A. B., Ribeiro Tomé, L. M., De Carvalho, G. C., **Camargos Fonseca, P. L.**, Pereira Figuiredo, H. C., Pereira, F., ... & Góes-Neto, A. (2020). Effect of biogas production on characteristics of municipal solid waste landfill sites. In *Proceedings of the Institution of Civil Engineers-Waste and Resource Management* 173:3. https://doi.org/10.1680/jwarm.19.00006.
- **Fonseca, P. L.C.**, Badotti, F., De-Paula, R. B., Araújo, D. S., Bortolini, D. E., Del-Bem, L. E., ... Aguiar. E., R., G., R. & Góes-Neto, A. (2020). Exploring the Relationship Among Divergence Time and Coding and Non-coding Elements in the Shaping of Fungal Mitochondrial Genomes. *Frontiers in microbiology*, 11:765. https://doi.org/10.3389/fmicb.2020.00765.
- Tomé, L. M. R., Badotti, F., Assis, G. B. N., **Fonseca, P. L. C.**, da Silva, G. A., da Silveira, R. M. B., ... & Góes-Neto, A. (2019). Proteomic fingerprinting for the fast and accurate identification of species in the Polyporoid and Hymenochaetoid fungi clades. *Journal of proteomics*, 203:103390. https://doi.org/10.1016/j.jprot.2019.103390.
- **Fonseca, P. L.**, Badotti, F., De Oliveira, T. F., Fonseca, A., Vaz, A. B., Tomé, L. M., ... & Aguiar, E. R. (2018). Virome analyses of Hevea brasiliensis using small RNA deep sequencing and PCR techniques reveal the presence of a potential new virus. *Virology journal*, 15:184. https://doi.org/10.1186/s12985-018-1095-3.
- Vaz, A. B., **Fonseca, P. L.**, Badotti, F., Skaltsas, D., Tomé, L. M., Silva, A. C., ... & Chaverri, P. (2018). A multiscale study of fungal endophyte communities of the foliar endosphere of native rubber trees in Eastern Amazon. *Scientific reports*,

8:16151. https://doi.org/10.1038/s41598-018-34619-w (divido a primeira autoria neste trabalho).

- Badotti, F., **Fonseca, P. L. C.**, Tomé, L. M. R., Nunes, D. T., & Góes-Neto, A. (2018). ITS and secondary biomarkers in fungi: review on the evolution of their use based on scientific publications. *Brazilian Journal of Botany*, *41*(2), 471-479.
- Ferreira, D. S. S., Kato, R. B., Miranda, F. M., da Costa Pinheiro, K., **Fonseca, P. L. C.**, Tomé, L. M. R., ... & de Carvalho Azevedo, V. A. (2018). Draft genome sequence of *Trametes villosa* (Sw.) Kreisel CCMB561, a tropical white-rot Basidiomycota from the semiarid region of Brazil. *Data in brief*, 18. https://doi.org/10.1016/j.dib.2018.04.074.
- Vaz, A.B.M., dos Santos, D.S., Cardoso, D., Van den Berg, C., de Queiroz, L.P., Badotti, F., **Fonseca, P.L.C.**, Cavalcanti, L.H., Góes-Neto, A. (2017). Corticolous myxomycetes assemblages in a seasonally dry tropical forest in Brazil. Mycoscience, 58:4. https://doi.org/10.1016/j.myc.2017.04.004.
- Badotti, F., de Oliveira, F. S., Garcia, C. F., Vaz, A. B. M., **Fonseca, P. L. C.**, Nahum, L. A., ... & Góes-Neto, A. (2017). Effectiveness of ITS and sub-regions as DNA barcode markers for the identification of Basidiomycota (Fungi). *BMC microbiology*, 17:41. https://doi.org/10.1007/s40415-018-0471-y.
- Vaz, A. B., **Fonseca, P. L.**, Leite, L. R., Badotti, F., Salim, A. C., Araujo, F. M., ... & Goes-Neto, A. (2017). Using Next-Generation Sequencing (NGS) to uncover diversity of wood-decaying fungi in neotropical Atlantic forests. *Phytotaxa*, 295:1. https://doi.org/10.11646/phytotaxa.295.1.1.

Anexo B - Artigo e Capítulo de livro aceitos para publicação (n=2)

- Daniel S. Araújo, Ruth B. De-Paula, Luiz M. R. Tomé, Gabriel Quintanilha-Peixoto, Carlos A. Salvador-Montoya, Luiz-Eduardo Del-Bem, Fernanda Badotti, Vasco A. C. Azevedo, Bertram Brenig, Eric R. G. R. Aguiar, Elisandro R. Drechsler-Santos, **Paula L. C. Fonseca** & Aristóteles Góes-Neto. (2021) Comparative mitogenomics of Basidiomycota: diversity, abundance, impact and coding potential of putative open-reading frames. Mitochondrion. (Divido a correspondência neste artigo).
- **Paula Luize Camargos Fonseca**, Thairine Mendes-Pereira, Luiz Marcelo Ribeiro Tomé, Felipe Ferreira da Silva, Gabriel Quintanilha-Peixoto, Daniel Silva Araújo, … Eric Roberto Guimarães Rocha Aguiar & Aristóteles Góes-Neto.

(2021) Microbial Consortia: Using *Omics* sciences for understanding the role of microorganisms and their potentialities in agriculture and environmental sustainability. In: Environmental and Agricultural Microbiology: Advances and Applications. 1ed. Wiley-Scrivener.

Anexo C - Artigos submetidos para publicação (n=2)

- **Paula L. C. Fonseca**, Daniel S. Araújo, Ruth B. de-Paula, Luiz M. R. Tomé, Leonardo de Oliveira Barbosa, Thairine Mendes-Pereira, Fernanda Badotti, Ana C. F. Soares, Eric R. G. R. Aguiar and Aristóteles Góes-Neto. The potential of the mitochondrial gene *rns* as a molecular marker for Hypocreales (Fungi, Ascomycota). Submetido para publicação na revista *FEMS Microbiology Letters* em janeiro de 2021.
- **Paula L. C. Fonseca**, Joel A. M. Porto, Felipe da Silva, João Paulo P. de Almeida, Juliana N. Armache, Jonathan Muscherino, … & Eric R. G. R. Aguiar. Characterization of the miRNA pathway in the invasive ectoparasite mite *Varroa destructor*. Submetido para publicação na revista *Genomics* em dezembro de 2020.

Anexo D - Orientação de alunos de Iniciação Científica (n=4)

- **Estudante** Daniel Silva Araujo. Mitogenômica comparativa do Filo Basidiomycota. **Período:** 06/2019 a 06/2020.
- **Estudante** Ruth Barros de Paula. Detecção e identificação de RNAs longos não codificantes no genoma do fungo *Trichoderma harzianum*. **Período:** 06/2018 a 06/2019.
- **Estudante** Mayara Cabral Cunha. Análise *in silico* da proteína SUV3 na ordem Hypocreales (Fungi). **Período:** 01/2019 a 12/2019.
- **Estudante** Allefi Castro Silva. Caracterização das proteínas envolvidas na via de RNA de interferência no fungo *Trichoderma harzianum*. **Período:** 06/2017 a 06/2018.

Anexo E - Participação em projetos de pesquisa submetidos a agências de fomento (n=3)

- **Projeto**: Investigação do viroma global de abelhas Apis melífera utilizando uma abordagem baseada em pequenos RNAs. **Agência financiadora**: PROPP UESC. **Situação:** Em andamento.
- **Projeto**: Rede Cooperativa de Pesquisa Internacional em Biodiversidade e Potencial Biotecnológico de vírus de fungos endofíticos de seringueira. **Agência financiadora:** Fundação de Amparo à Pesquisa do Estado da Bahia. **Situação:** Concluído.
- **Projeto**: Mycota associated to native *Hevea* spp. in the Brazilian Amazon Region. **Agência financiadora:** National Science Foundation. **Situação:** Concluído.

Anexo F - Prêmios e títulos (n=1)

- Best Poster Award, AB3C, *X-meeting experience* (2020).

Anexo G - Cursos nacionais (n=1)

- Análise de Transcritomas e microRNAs. Laboratório Nacional de Computação Científica (LNCC), Petrópolis, 2019 (Carga horária: 36h).

Anexo H - Cursos internacionais (n=5)

- NGS analysis applied to virome sequencing in agricultural systems. Bioinformatic pratices. Universidad de Costa Rica (UCR), San Pedro, 2020 (Carga horária: 21h);
- NGS analysis applied to virome sequencing in agricultural systems. Wet lab pratices. Universidad de Costa Rica (UCR), San Pedro, 2020 (Carga horária: 36h);
- Electron Microscopy Cryo-techniques. Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, 2018 (Carga horária: 30h);
- Introdução às técnicas RNAi, CRISPR e microRNAs. Universidade de São Paulo (USP) – CBAB, Ribeirão Preto, 2017 (Carga horária 80 horas).
- Theoretical and Practical Approaches to Metagenomics and Viral Discovery. Universidade de São Paulo (USP), São Paulo, 2017 (Carga horaria 30 horas).

Anexo I - Simpósios e congressos

- X meeting Experience. 2020. Evento *online*;
- UFMG Jovem. 2019. Simpósio;

- VI Simpósio de Microbiologia da UFMG: Microbiologia Interligada. 2019. Simpósio;
- X-meeting. 2019. Congresso;
- V Simpósio de Microbiologia da UFMG. 2018;
- IV Simpósio de Microbiologia da UFMG. 2017.

Anexo J - Apoio

1. Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq;
2. National Science Foundation - NSF