

CASLEY BORGES DE QUEIROZ

**SEQUENCIAMENTO DO GENOMA DE *Colletotrichum lindemuthianum* E  
IDENTIFICAÇÃO DE GENES CANDIDATOS A EFETORES**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Microbiologia Agrícola, para obtenção do título de *Doctor Scientiae*.

VIÇOSA  
MINAS GERAIS – BRASIL  
2017

**Ficha catalográfica preparada pela Biblioteca Central da Universidade  
Federal de Viçosa - Câmpus Viçosa**

T

Queiroz, Casley Borges de, 1987-  
Q3s Sequenciamento do genoma de *Colletotrichum*  
2017 *lindemuthianum* e identificação de genes candidatos a efetores /  
Casley Borges de Queiroz. – Viçosa, MG, 2017.  
x, 137f. : il. (algumas color.) ; 29 cm.

Inclui anexos.

Orientador: Marisa Vieira de Queiroz.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. *Colletotrichum lindemuthianum*. 2. Fungos  
fitopatogênicos. 3. Genoma. 4. Antracnose. I. Universidade  
Federal de Viçosa. Departamento de Microbiologia. Programa de  
Pós-graduação em Meteorologia Agrícola. II. Título.

CDD 22 ed. 632.4

CASLEY BORGES DE QUEIROZ

**SEQUENCIAMENTO DO GENOMA DE *Colletotrichum lindemuthianum* E  
IDENTIFICAÇÃO DE GENES CANDIDATOS A EFETORES**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Microbiologia Agrícola, para obtenção do título de *Doctor Scientiae*.

APROVADA: 24 de maio de 2017.



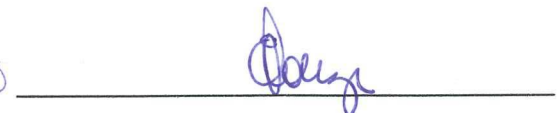
Mateus Ferreira Santana  
(Coorientador)



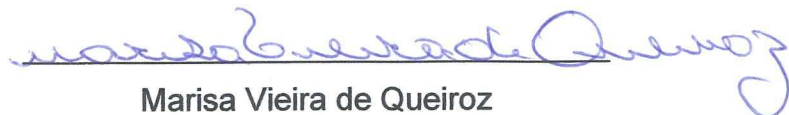
Sérgio Hermínio Bromonschenkel



Diego Silva Batista



Elaine Aparecida de Souza



Marisa Vieira de Queiroz  
(Orientadora)

Dedico

Aos meus pais Clementino e Eliane.

Às minhas irmãs Elane, Ellen e Eline

## **AGRADECIMENTOS**

Primeiramente agradeço a Deus pela saúde, força e determinação para superar os obstáculos e atingir meus objetivos.

À minha família pela compreensão, e todo apoio durante minha jornada.

À Universidade Federal de Viçosa e ao Programa de Microbiologia Agrícola, pela oportunidade de realizar este curso.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pelo apoio financeiro.

À professora Marisa Vieira de Queiroz pelos ensinamentos, pela orientação e principalmente pela confiança na realização deste trabalho.

Ao professor Mateus Ferreira Santana pela coorientação, amizade, ensinamentos, e por toda a contribuição na execução deste trabalho.

Ao Dr. Gilvan Ferreira da Silva, pelos ensinamentos, por todo apoio e incentivo.

Ao Técnico Pedro M. Pereira Vidigal pelos ensinamentos e auxílio na bioinformática. Sem ele, estes trabalhos seriam extremamente laboriosos.

Aos colegas de laboratório pela excelente convivência e amizade.

A todos aqueles que me apoiaram e torceram por mim.

Minha gratidão.

## SUMÁRIO

RESUMO .....	vii
ABSTRACT.....	ix
INTRODUÇÃO .....	1
REFERÊNCIAS .....	7
CAPITULO 1.....	14
DRAFT GENOME SEQUENCES OF TWO ISOLATES OF <i>Colletotrichum lindemuthianum</i> , THE CAUSAL AGENT OF ANTHRACNOSE IN COMMON BEANS.....	14
ABSTRACT.....	16
GENOME ANNOUNCEMENT .....	17
ACKNOWLEDGMENTS .....	18
FOOTNOTES .....	18
REFERENCES.....	19
CAPITULO 2.....	21
ANÁLISE COMPARATIVA DO GENOMA MITOCONDRIAL DO FUNGO <i>Colletotrichum lindemuthianum</i> , AGENTE CAUSAL DA ANTRACNOSE EM FEIJOEIRO.....	21
RESUMO .....	23
INTRODUÇÃO .....	24
MATERIAL E MÉTODOS .....	26
Sequenciamento de DNA e montagem.....	26
Anotação e análise do genoma mitocondrial de <i>C. lindemuthianum</i> .....	27
Análise filogenética.....	28
RESULTADOS .....	28
O genoma mitocondrial de <i>C. lindemuthianum</i> .....	28

RNAs transportadores de <i>C. lindemuthianum</i> .....	34
Open reading frames não conservadas (ncORFs), introns e ORFs intrônicas em <i>Colletotrichum</i> spp. ....	35
Sobreposições de genes e regiões intergênicas .....	40
Composição de nucleotídeos e uso de códons preferenciais .....	40
Análise filogenética das espécies de <i>Colletotrichum</i> .....	44
DISCUSSÃO .....	44
REFERÊNCIAS.....	51
ARQUIVOS ADICIONAIS.....	60
CAPITULO 3.....	69
O REPERTÓRIO DE CANDIDATOS A EFETORES EM <i>Colletotrichum</i> spp. REVELA INFORMAÇÕES IMPORTANTES PARA SEU O ESTILO DE VIDA.....	69
RESUMO .....	71
INTRODUÇÃO .....	72
MATERIAL E MÉTODOS.....	76
Predição de genes de proteínas candidatas à efetoras em <i>Colletotrichum</i> spp. ....	76
Análise da transcrição de candidatos a efetores de <i>Colletotrichum lindemuthianum</i> .....	78
RESULTADOS.....	79
Sequências de candidatas à efetoras em <i>Colletotrichum lindemuthianum</i> .....	79
Candidatas a proteínas efetoras de <i>C. lindemuthianum</i> apresentam domínios conservados e similaridades com efetoras conhecidas de fitopatógenos .....	82

Análise comparativa entre o repertório de proteínas candidatas à efetoras em <i>Colletotrichum lindemuthianum</i> e outras nove espécies de pertencetes ao gênero <i>Colletotrichum</i> .....	84
Domínios conservados encontrados em proteínas candidatas a efetoras em <i>Colletotrichum</i> spp. ....	86
Candidatos a efetores em <i>Colletotrichum</i> spp. apresentam termos funcionais do Gene Ontology.....	89
Genes candidatos a efetores de <i>C. lindemuthianum</i> apresentam expressão induzida in planta.....	92
DISCUSSÃO .....	94
CONSIDERAÇÕES FINAIS .....	104
REFERÊNCIAS.....	105
ARQUIVOS ADICIONAIS.....	120



## RESUMO

QUEIROZ, Casley Borges, D.Sc., Universidade Federal de Viçosa, maio de 2017. **Sequenciamento do genoma de *Colletotrichum lindemuthianum* e identificação de genes candidatos a efetores**. Orientadora: Marisa Vieira de Queiroz. Coorientadores: Mateus Ferreira Santana e Denise Mara Soares Bazzolli.

*Colletotrichum* é um gênero de fungo que compreende centenas de espécies amplamente distribuídas pelo mundo. Muitas dessas espécies além de serem fitopatógenos de grande importância econômica, representam modelos para a identificação de genes relacionados com a patogenicidade e virulência de patógenos hemibiotróficos. Apesar disso, até o momento, poucos genomas de espécies de *Colletotrichum* foram sequenciados e anotados. *Colletotrichum lindemuthianum*, agente causal da antracnose no feijoeiro comum (*Phaseolus vulgaris*), adota um estilo de parasitismo hemibiotrófico durante a infecção do feijoeiro e representa um dos grupos de patógenos mais devastadores para essa cultura. Neste trabalho é reportado o sequenciamento do genoma nuclear e a anotação e análise do genoma mitocondrial de dois isolados de *C. lindemuthianum* (83.501 e 89 A<sub>2</sub> 2-3). Além disso, realizou-se a predição *in silico* dos genes relacionados ao repertório de proteínas candidatas à efetoras no genoma nuclear de dez espécies de *Colletotrichum*, incluindo *C. lindemuthianum*. Os genomas nucleares dos isolados de *C. lindemuthianum* possuem um tamanho total de 97,4 Mb (83.501) e 99,16 Mb (89 A<sub>2</sub> 2-3), sendo até o momento os maiores genomas de espécies de *Colletotrichum* já sequenciados, e codificam um total de 11.627 (89 A<sub>2</sub> 2-3) e 11.673 (83.501) proteínas. Os tamanhos dos genomas mitocondriais foram 37.440 pb (83.501) e 37.446 pb (89 A<sub>2</sub> 2-3). A diferença de seis nucleotídeos entre os dois genomas é o resultado de uma deleção no gene *rps3* no isolado 83.501 e foi observada uma troca de adenina por guanina na posição 3.457 no genoma mitocondrial do isolado 83.501, dentro do gene *rps3* (proteína ribossômica S3). Ambos possuem um total de 53 genes, incluindo genes que codificam proteínas, RNAs transportadores, RNAs ribossômicos e *open reading frame* não conservadas (ncORF). O genoma mitocondrial de *C.*

*lindemuthianum* foi comparado com os genomas mitocondriais de sete espécies do gênero pertencentes aos clados acutatum, graminicola, orbiculare e gloeosporioides. Quatorze genes *core* mitocondriais em *Colletotrichum* spp. não apresentam sintonia. As ncORF variam em número e são codificadas por ambas as fitas. As regiões intergênicas possuem tamanho similares, exceto nas espécies do clado gloeosporioides, e o número de introns foi altamente variável e codificam um total de até nove endonucleases *homing*. A variação no tamanho dos genomas mitocondriais é devida principalmente as ncORF e introns do grupo I. O total de genes candidatos a efetores preditos a partir da montagem dos genomas nucleares em ambos isolados de *C. lindemuthianum* foi de 324. Entre as nove espécies analisadas, o total variou de 247 em *Colletotrichum graminicola* até 446 em *Colletotrichum orbiculare*. A maioria das sequências de aminoácidos codificados pelos genes candidatos à efetores (até 91,4% no caso de *Colletotrichum fioriniae*) apresentou homologia com proteínas não caracterizadas. Um total de 28 grupos de sequências de aminoácidos (total de 43 sequências de aminoácidos) codificadas por genes candidatos a efetores em *C. lindemuthianum* possuem sequências homólogas em todas as outras nove espécies de *Colletotrichum*. Foram encontrados 12 domínios conservados em todas as 10 espécies. Foram identificados também sequências de aminoácidos com presença de domínios conservados ainda não relatado para proteínas efetoras de *Colletotrichum* spp., o que permitiu sugerir papéis putativos para essas efetoras. A transcrição de seis genes candidatos a efetores de *C. lindemuthianum* foi avaliada por meio de Real-time RT-PCR. Todos os genes selecionados tiveram o mesmo padrão de expressão, sendo maior na fase biotrófica de crescimento do fungo no feijoeiro. Este trabalho disponibiliza informações importantes que podem auxiliar a compreensão da biologia e das estratégias de infecção utilizadas por espécies do gênero *Colletotrichum*.

## ABSTRACT

QUEIROZ, Casley Borges, D.Sc., Universidade Federal de Viçosa, May, 2017. **Sequencing of the genome of *Colletotrichum lindemuthianum* and identification of candidate genes for effectors**. Adviser: Marisa Vieira de Queiroz. Co-advisers: Denise Mara Soares Bazzolli and Mateus Ferreira Santana.

*Colletotrichum* is a genus of fungus that comprises hundreds of species widely distributed throughout the world. Many of these species, besides being phytopathogens of great economic importance, represent models for the identification of genes related to the pathogenicity and virulence of hemibiotrophic pathogens. Despite this, to date, few genomes of *Colletotrichum* species have been sequenced and annotated. *Colletotrichum lindemuthianum*, the causal agent of anthracnose in common bean (*Phaseolus vulgaris*), adopts a style of hemibiotrophic parasitism during bean infection and represents one of the most devastating groups of pathogens for this crop. In this work, we report the sequencing of the nuclear genome and the annotation and analysis of the mitochondrial genome of two isolates of *C. lindemuthianum* (83.501 and 89 A<sub>2</sub> 2-3). In addition, *in silico* prediction of genes related to the repertoire of effector candidate proteins in the nuclear genome of ten *Colletotrichum* species, including *C. lindemuthianum*, was performed. Nuclear genomes of *C. lindemuthianum* isolates have a total size of 97.4 Mb (83.501) and 99.16 Mb (89 A<sub>2</sub> 2-3), and the largest genomes of *Colletotrichum* species have been sequenced to date and encode one total of 11,627 (89 A<sub>2</sub> 2-3) and 11,673 (83,501) proteins. The sizes of the mitochondrial genomes were 37,440 bp (83,501) and 37,446 bp (89 A<sub>2</sub> 2-3). The difference of six nucleotides between the two genomes is the result of a deletion in the *rps3* gene in the 83501 isolate and an adenine exchange for guanine at position 3,457 was observed in the mitochondrial genome of the 83,501 isolate within the *rps3* gene (ribosomal protein S3). Both have a total of 53 genes, including genes encoding uncorrected proteins, transporter RNAs, ribosomal RNAs and open reading frame (ncORF). The mitochondrial

genome of *C. lindemuthianum* was compared to the mitochondrial genomes of seven species of the genus belonging to the clades acutatum, graminicola, orbiculare and gloeosporioides. Fourteen mitochondrial core genes in *Colletotrichum* spp. They do not have synteny. The ncORFs vary in number and are encoded by both strand. The intergenic regions have similar size, except in the gloeosporioid clade species, and the number of introns was highly variable and encode a total of up to nine homing endonucleases. The variation in the size of mitochondrial genomes is mainly due to ncORF and type I introns. The total of effector candidate genes predicted from the assembly of nuclear genomes in both *C. lindemuthianum* isolates was 324. Among the nine species analyzed, the total ranged from 247 in *Colletotrichum graminicola* to 446 in *Colletotrichum orbiculare*. Most of the amino acid sequences encoded by the effector candidate genes (up to 91.4% in the case of *Colletotrichum fioriniae*) showed homology with uncharacterized proteins. A total of 28 groups of amino acid sequences (total of 43 amino acid sequences) encoded by effector candidate genes in *C. lindemuthianum* have homologous sequences in all nine other *Colletotrichum* species. We found 12 conserved domains in all 10 species. It was also identified amino acid sequences with presence of conserved domains not yet reported for effectors of *Colletotrichum* spp., which allowed to suggest putative roles for these effectors. The expression of six candidate genes for effectors of *C. lindemuthianum* was evaluated by means of Real-time RT-PCR. All selected genes had the same expression pattern, being higher in the biotrophic phase of fungus growth in common bean. This work provides important information that can help to understand the biology and infection strategies used by species of the genus *Colletotrichum*.

## INTRODUÇÃO

As espécies do gênero *Colletotrichum* estão entre os patógenos de plantas mais disseminados e economicamente danosos, sendo aproximadamente 600 espécies que infectam mais de 3200 espécies de plantas economicamente importantes como abacate, manga, citros, milho, sorgo, grãos de legumes, mandioca, inhame e muitas outras culturas. (Cannon et al., 2012). O gênero é classificado como o oitavo mais importante grupo de fungos fitopatogênicos no mundo, baseado em suas características biológicas e importância econômica (Dean et al. 2012). A antracnose é a doença normalmente causada por *Colletotrichum* spp., porém outras doenças são também reportadas como a podridão vermelha da cana-de-açúcar, podridão da coroa do morango e banana e a mancha marrom de feijão-caupi (Cannon et al., 2012; Lenné, 2002). Os sintomas da antracnose incluem lesões necrotróficas nas folhas, caules, flores e frutos.

As espécies de *Colletotrichum* utilizam duas principais estratégias de parasitismo: colonização intracelular e a colonização intramural subcuticular, porém, a primeira é a mais comum entre as espécies deste gênero (Bailey et al., 1992; Perfect et al., 1999). O processo infeccioso de *Colletotrichum* spp. que possuem um estilo de parasitismo hemibiotrófico, envolve a formação sequencial de uma série de tipos de células especializadas (Perfect et al., 1999). Quando o conídio de espécies fitopatogênicas desse gênero encontra a superfície da planta, ele adere e germina, formando um tubo germinativo que subsequentemente forma o apressório escuro em formato de cúpula que medeia a penetração inicial no hospedeiro. A hifa de penetração emerge da base do apressório, que rompe a cutícula e a parede celular da célula da planta por meio da força mecânica gerada pela pressão de turgor e por enzimas líticas (Bechinger et al., 1999). Dentro das células epidérmicas da planta, a hifa de penetração se diferencia em hifas primárias bulbosas, que se expandem e são envelopadas pela membrana plasmática intacta do hospedeiro (O'Connell et al., 2012; Perfect et al., 1999). É mantida uma matriz interfacial separando a membrana plasmática da parede celular do fungo e células recém infectadas

permanecem viáveis por algum tempo (Perfect et al., 1999). Dessa maneira, esse estágio assintomático de infecção é nomeado de biotrófico em razão das células hospedeiras invadidas permanecerem vivas (O'Connell et al., 2012). Entre 48 a 72 horas após a penetração inicial, células infectadas contendo as hifas primárias e secundárias começam a se desintegrar. Esse estágio marca o início da fase necrotrófica, na qual hifas secundárias emergem a partir das hifas primárias e invadem células vizinhas, causando a degradação da parede e morte celular (Perfect et al., 1999).

A espécie *Colletotrichum lindemuthianum* (Sacc. & Magnus) Briosi & Cavara é o agente causal da antracnose no feijão comum (*Phaseolus vulgaris* L.), atacando folhas, caules, ramos, vagens e sementes (Mahuku & Riascos, 2004). Em condições favoráveis de umidade e temperatura, a antracnose resulta em desfolhamento prematuro, queda prematura de flores e vagens e em casos extremos pode levar a morte da planta, diminuindo a produtividade e causando depreciação dos grãos (Campa et al., 2014; Schwartz et al., 2005; Sharma et al., 2008). *Colletotrichum lindemuthianum* tem uma alta variabilidade genética (Dos Santos et al., 2012; Mota et al., 2016; Padder et al., 2007), caracterizada pela existência de pelo menos 100 variantes patogênicos ou raças fisiológicas reportadas entre isolados coletados em várias regiões do mundo (Alzate-Marin et al., 1999; Ferreira et al., 2013; González et al., 2015; Rodríguez-Guerra et al., 2006). Esta alta variabilidade genética é uma das principais barreiras para o controle deste patógeno, uma vez que dificulta a obtenção de cultivares com resistência durável (Ferreira et al., 2013).

Desde o sequenciamento do primeiro genoma de fungo, a levedura *Saccharomyces cerevisiae* (Goffeau et al., 1996), o número de sequências de genomas nucleares de fungos disponíveis tem aumentado consideravelmente. Entretanto, entre os membros de *Colletotrichum* spp., até o momento foram sequenciados os genomas de apenas 15 espécies, sendo elas: *Colletotrichum graminicola* com um genoma de 57,4 Mpb que é distribuído em 13 cromossomos, onde foram preditos 12.006 genes que codificam proteínas (GCPs) (O'Connell et al., 2012); *Colletotrichum higginsianum* com um genoma de 50,72 Mpb presente em 12 cromossomos com 14.651 GCPs (O'Connell et al., 2012; Zampounis et al., 2016);

*Colletotrichum orbiculare* com um genoma de 90,09 Mpb com 13.479 GCPs (Gan et al., 2013); *Colletotrichum gloeosporioides* Nara-gc5 que recentemente foi classificado como *Colletotrichum fructicola* (Crouch et al., 2014) com um genoma de 55,6 Mpb com 15.469 GCPs (Gan et al., 2013); *Colletotrichum sublineola* com um genoma de 46,75 Mpb com 12.699 GCPs (Baroncelli et al., 2014a); *Colletotrichum fioriniae* com um genoma de 49,01 Mpb com 13.759 GCPs (Baroncelli et al., 2014b); *Colletotrichum nymphaeae* com um genoma de 49,96 Mpb com 14.404 GCPs (Baroncelli et al., 2016); *Colletotrichum simmondsii* com um genoma de 50,47 Mpb com 13.884 GCPs (Baroncelli et al., 2016); *Colletotrichum salicis* com um genoma com 48,37 Mpb com 13.783 GCPs (Baroncelli et al., 2016); *Colletotrichum tofieldiae* com um genoma de 52,83 Mpb com 12.501 GCPs (Hacquard et al., 2016); *Colletotrichum incanum* com um genoma de 58,92 Mpb com 11.852 GCPs (Gan et al., 2016) e *Colletotrichum falcatum* com um genoma de 48,16 Mpb com 12.270 GCPs (Viswanathan et al., 2016). Além dessas espécies supracitadas, existem até o momento (abril de 2017) mais três genomas depositados no *GenBank*, mas estes ainda não tiveram trabalhos publicados: *Colletotrichum godetiae* com 35,03 Mpb (genes não preditos), *Colletotrichum chlorophyti* com 52,39 Mpb e 10.419 GCPs e *Colletotrichum orchidophilum* com 48,56 Mpb e 14.496 GCPs.

Quanto aos genomas mitocondriais de *Colletotrichum* spp., foram sequenciados apenas de nove espécies, incluindo *C. lindemuthianum* com tamanho de 36.957 pb (Gutiérrez et al., 2014); *C. acutatum* com tamanho de 30.892 pb (Kim et al., 2015); *C. graminicola* com tamanho de 39.649 pb (Liang et al., 2017); *C. lupini* com tamanho de 36.554 pb (Liang et al., 2017); *C. tamarilloi* com tamanho de 30.823 pb (Liang et al., 2017); *C. gloeosporioides* com tamanho de 55.169 pb (Liang et al., 2017); *C. fructicola* com tamanho de 56.051 pb (Liang et al., 2017); *C. siamense* com tamanho de 58.666 pb (Liang et al., 2017) e *C. fioriniae* com tamanho de 30.020 pb (não publicado). Cada genoma mitocondrial das oito espécies possui 14 genes conservados que codificam proteínas envolvidas na fosforilação oxidativa e síntese de ATP, dois genes da subunidade maior e menor ribossomal, um gene da proteína ribossomal S3 e um total variável de genes de RNAs transportadores (tRNA), de 25 a 29 genes. No genoma

mitocondrial de *C. lindemuthianum* foram preditos somente 13 genes conservados, além dos dois genes das subunidades ribossomais, um gene da proteína ribossomal S3 e 28 genes de RNAs transportadores (Gutiérrez et al., 2014).

Durante a tentativa de invasão do patógeno ao hospedeiro, ocorre uma "batalha" intensa entre os dois. O sistema de defesa da planta apresenta duas linhas de atuação (Dodds and Rathjen, 2010; Liu et al., 2013). A primeira linha envolve a resposta basal de defesa orquestrada por receptores transmembrana, extracelulares ou receptores de reconhecimento de padrões (PRRs). Os PRRs detectam a presença de moléculas amplamente conservadas, que são características de micro-organismos patogênicos, nomeadas como Padrão Molecular Associado ao Patógeno (PAMPs) e ativam uma resposta específica de imunidade ativada por PAMPs (PTI – imunidade ativada por PAMPs) (Selin et al., 2016; Wang et al., 2017). As PAMPs referem-se a qualquer tipo de molécula estrutural amplamente conservada entre os patógenos, por exemplo, componentes da parede celular do patógeno como quitina em fungos e peptidoglicano em bactérias (Giraldo & Valent, 2013; Motion et al., 2015).

Patógenos filamentosos de plantas, biotróficos ou hemibiotróficos, que obtêm nutrientes de células vivas da planta, precisam bloquear as defesas dos hospedeiros e manter os processos dos hospedeiros que necessitam para alimentação e crescimento (Giraldo & Valent, 2013; Rafiqi et al., 2012). Uma vez que o sistema de defesa da planta detecta a presença invasiva do patógeno, a planta pode induzir a produção de espécies reativas de oxigênio (ROS), bem como a secreção de compostos antimicrobianos, enzimas hidrolíticas que são danosas para o patógeno e inibidores de enzimas hidrolíticas do patógeno que são danosas para planta (Asai and Shirasu, 2015; Giraldo & Valent, 2013). Como contra-ataque, o patógeno secreta os efetores, que geralmente são pequenas proteínas solúveis que manipulam a estrutura e fisiologia da célula hospedeira para infectar o hospedeiro. Os efetores permitem que o patógeno escape ou desarme o mecanismo de defesa, por exemplo, inibindo a PTI, e suprime a morte programada das células, o que poderá restringir a colonização do patógeno (Göhre & Robatzek, 2008; Rafiqi et al., 2012). Por exemplo, diversos



patógenos secretam efetores que contem domínios LysM (módulos ligantes de carboidratos que geralmente liga a N-acetilglicosamina) que evitam a liberação de oligossacarídeos de quitina da parede celular do fungo ou sequestram estes oligossacarídeos para evitar o seu reconhecimento pelo sistema de defesa da planta (Akcapinar et al., 2015; Giraldo & Valent, 2013; Kombrink et al., 2016), sendo estes os casos dos efetores SLP1 de *Magnaporthe oryzae* (Mentlak et al., 2011) e ECP6 e AVR4 de *Cladosporium fulvum* (de Jonge et al., 2010). A planta por sua vez, desenvolveu uma segunda linha de defesa chamada de imunidade desencadeada por efector (ETI), que inclui proteínas de resistência (R), e este sistema é ativado pela detecção de proteínas efectoras altamente variáveis chamadas de proteínas de avirulência (efetores AVR). Este sistema de defesa está frequentemente associado ao desencadeamento da reação de hipersensibilidade (HR) (Bigeard et al., 2015; Liu et al., 2013;).

No que diz respeito a efetores em *Colletotrichum* spp., o estudo mais recente e completo tem sido feito com *C. higginsianum*. Neste fungo, os genes que codificam proteínas candidatas a efectoras são expressos em quatro ondas principais referentes a efetores expressos antes da penetração do apressório, logo após a penetração do apressório, durante a fase biotrófica e no momento da mudança para fase necrotrófica ou durante a fase necrotrófica, sendo assim mostrado que diferentes efetores são expressos nos diferentes estágios de patogênese (Kleemann et al., 2012). Em *C. graminicola*, o gene que codifica a proteína CgEP1 foi altamente expresso durante a fase biotrófica e CgEP1 é transportada para o núcleo de células do milho, sendo capaz de se ligar ao DNA, incluindo regiões *upstream* de genes que codificam fatores de transcrição, podendo, desta forma, está envolvida nos processos regulatórios de prevenção de resposta defensiva do hospedeiro (Vargas et al., 2015). Também em *C. graminicola* foi demonstrado que Cgfl é uma proteína com função de fungolisina que possui papel na virulência deste patógeno, podendo suprimir a defesa hospedeira por meio da degradação de quitinases (Sanz-Martín et al., 2016). A proteína CtNUDIX de *C. truncatum* induz uma morte celular severa no tabaco, similar à resposta de hipersensibilidade. O gene que codifica essa proteína é transcrito no final da biotrofia, podendo assim está envolvido na

mudança da fase biotrófica para fase necrotrófica (Bhadauria et al., 2013). É provável que a proteína efetora CltoxB de *C. lentis* também atue nesta etapa, amplificando o sinal de morte celular desencadeado pelo efector CINUDIX (Bhadauria et al., 2015). Em *C. higginsianum*, a proteína extracelular LysM 2 (ChELP2) se liga a quitina e oligômeros de quitina e assim suprime a ativação de defesa em *Arabidopsis*, sendo essencial para a patogenicidade (Takahara et al., 2016). Apesar dos relatos de estudos de genes relacionados à efetores em *Colletotrichum* spp. e outras espécies de fungos fitopatogênicos, uma abordagem mais descritiva e comparativa ainda se faz necessária, principalmente se levarmos em consideração a diversidade de espécies de *Colletotrichum* causadoras de doença em uma grande variedade de plantas hospedeiras, incluindo culturas de importância econômica.

Assim, a análise comparativa dos genomas de isolados de *C. lindemuthianum* com os genomas das espécies pertencentes a esse gênero pode ajudar no entendimento da biologia dessas espécies fitopatogênicas, bem como, fornecer alvos conservados que possam ser utilizados no controle da antracnose em diversas espécies de plantas.

## REFERÊNCIAS

- Akcapinar GB, Kappel L, Sezerman OU, Seidl-Seiboth V. (2015) Molecular diversity of LysM carbohydrate-binding motifs in fungi. *Curr Genet.* 61:103-113.
- Alzate-Marin AL, Menarim H, de Carvalho GA, de Paula J´unior TJ, de Barros EG, Moreira MA. (1999) Improved selection with newly identified RAPD markers linked to resistance gene to four pathotypes of *Colletotrichum lindemuthianum* in common bean. *Phytopathology* 89:281–285.
- Asai S, Shirasu, K. (2015) Plant cells under siege: Plant immune system versus pathogen effectors. *Curr Opin Plant Biol.* 28:1-8.
- Bailey JA, O’Connell RJ, Pring RJ, Nash C. (1992) Infection strategies of *Colletotrichum* species. In *Colletotrichum: Biology, Pathology and Control* (J. A. Bailey, and M. J. Jeger, Eds), pp. 88–120. CAB International, Wallingford.
- Baroncelli R, Amby DB, Zapparata A, Sarrocco S, Vannacci G, Le Floch G, Harrison RJ, Holub E, Sukno SA, Sreenivasaprasad S, Thon MR. (2016) Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC Genomics.* 17:555.
- Baroncelli R, Sanz-Martín JM, Rech GE, Sukno SA, Thon MR. (2014a) Draft genome sequence of *Colletotrichum sublineola*, a destructive pathogen of cultivated sorghum. *Genome Announc.* 2:e00540-14.
- Baroncelli R, Sanz-Martín JM1, Rech GE1, Sukno SA, Thon MR. (2014b) Draft Genome Sequence of *Colletotrichum sublineola*, a Destructive Pathogen of Cultivated Sorghum. *Genome Announc.* 2:e00540-14.

- Bechinger C1, Giebel KF, Schnell M, Leiderer P, Deising HB, Bastmeyer M. (1999) Optical measurements of invasive forces exerted by appressoria of a plant pathogenic fungus. *Science*. 285:1896-1899.
- Bhadauria V, Banniza S, Vandenberg A, Selvaraj G, Wei Y. (2013) Overexpression of a novel biotrophy-specific *Colletotrichum truncatum* effector, CtNUDIX, in hemibiotrophic fungal phytopathogens causes incompatibility with their host plants. *Eukaryot Cell*. 12:2-11.
- Bhadauria V, MacLachlan R, Pozniak C, Banniza S. (2015) Candidate effectors contribute to race differentiation and virulence of the lentil anthracnose pathogen *Colletotrichum lentis*. *BMC Genomics*. 16:628.
- Bigeard J, Colcombet J, Hirt H. (2015) Signaling mechanisms in pattern-triggered immunity (PTI). *Mol Plant*. 8:521-539.
- Campa A, Rodríguez-Suárez C, Giraldez R, Ferreira JJ. (2014) Genetic analysis of the response to eleven *Colletotrichum lindemuthianum* races in a RIL population of common bean (*Phaseolus vulgaris* L.). *BMC Plant Biol*. 14:115.
- Cannon PF, Damm U, Johnston PR, Weir BS. (2012). *Colletotrichum* – current status and future directions. *Stud Mycol*. 73:181-213.
- Crouch JA, O'Connell R, Gan P, Buiate E, Torres MF, Beirn L, Shirasu K, Vaillancourt L. (2014) The genomics of *Colletotrichum*. In: *Genomics of plant-associated fungi: monocot pathogens*. Berlin: Springer Verlag. 69-102.
- de Jonge R, van Esse HP, Kombrink A, Shinya T, Desaki Y, Bours R, van der Krol S, Shibuya N, Joosten MH, Thomma BP. (2010) Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *Science*. 329:953–955.
- Dean R, Van Kan JA, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J, Foster GD. (2012)

The Top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol.* 13:414-430.

Dos Santos LV, de Queiroz MV, Santana MF, Soares MA, de Barros EG, de Araújo EF, Langin T. 2012. Development of new molecular markers for the *Colletotrichum* genus using RetroCl1 sequences. *World J Microbiol Biotechnol.* 28:1087-1095.

Ferreira JJ, Campa A, Kelly JD. Organization of genes conferring resistance to anthracnose in common bean (2013). pp. 151–181. In: Varshney R. K. and Tuberosa R. (eds.) *Translational Genomics for Crop Breeding, Volume I: Biotic Stresses.* John Wiley & Sons, Inc.

Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell RJ, Narusaka Y, Takano Y, Kubo Y, Shirasu K. (2013) Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytol.* 197:1236-1249.

Gan P, Narusaka M, Kumakura, N, Tsushima A, Takano Y, Narusaka Y, Shirasu K. (2016) Genus-Wide Comparative genome analyses of *Colletotrichum* species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. *Genome Biol Evol.* 8:1467-1481.

Giraldo MC, Valent B. (2013) Filamentous plant pathogen effectors in action. *Nat Rev Microbiol.* 11:800-814.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG. (1996) Life with 6000 genes. *Science.* 274:563-567.

Göhre V, Robatzek S. (2008) Breaking the barriers: microbial effector molecules subvert plant immunity. *Annu Rev Phytopathol.* 46:189-215.

González AM, Yuste-Lisbona FJ, Rodiño AP, De Ron AM, Capel C, García-Alcázar M, Lozano R, Santalla M. (2015) Uncovering the genetic

architecture of *Colletotrichum lindemuthianum* resistance through QTL mapping and epistatic interaction analysis in common bean. *Front Plant Sci.* 6:141.

Gutiérrez P, Alzate J, Yepes MS, Marín M. (2014) Complete mitochondrial genome sequence of the common bean anthracnose pathogen *Colletotrichum lindemuthianum*. *Mitochondrial DNA*, Early Online: 1–2

Hacquard S, Kracher B, Hiruma K, Münch PC, Garrido-Oter R, Thon MR, Weimann A, Damm U, Dallery JF, Hainaut M, Henrissat B, Lespinet O, Sacristán S, Ver Loren van Themaat E, Kemen E, McHardy AC, Schulze-Lefert P, O'Connell RJ. (2016) Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nat Commun.* 7:11362.

Kim JO, Choi KY, Han JH, Choi IY, Lee YH, Kim KS. (2015) The complete mitochondrial genome sequence of the ascomycete plant pathogen *Colletotrichum acutatum*. *Mitochondrial DNA*, Early Online: 1–2.

Kleemann J, Rincon-Rivera LJ, Takahara H, Neumann U, Ver Loren van Themaat E, van der Does HC, Hacquard S, Stüber K, Will I, Schmalenbach W, Schmelzer E, O'Connell RJ. (2012) Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathog.* 8:e1002643.

Kombrink A, Rovenich H, Shi-Kunne X, Rojas-Padilla E, van den Berg GC, Domazakis E, de Jonge R, Valkenburg DJ, Sánchez-Vallet A, Seidl MF, Thomma BP. (2016) *Verticillium dahliae* LysM effectors differentially contribute to virulence on plant hosts. *Mol Plant Pathol.* 8:596–608.

Lenné JM. (2002) Some major plant diseases. In: *Plant Pathologist's Pocketbook* (Waller JM, Lenné JM, Waller SJ, eds). 3 rd edn. CABI, Wallingford, UK: 4–18.

Liang X, Tian X, Liu W, Wei T, Wang W, Dong Q, Wang B, Meng Y, Zhang R, Gleason ML, Sun G. (2017) Comparative analysis of the mitochondrial

- genomes of *Colletotrichum gloeosporioides* sensu lato: insights into the evolution of a fungal species complex interacting with diverse plants. *BMC Genomics*. 18:171.
- Liu W, Liu J, Ning Y, Ding B, Wang X, Wang Z, Wang GL. (2013) Recent progress in understanding PAMP- and effector-triggered immunity against the rice blast fungus *Magnaporthe oryzae*. *Mol Plant*. 6:605-620.
- Mahuku SG, Riascos JJ. (2004) Virulence and molecular diversity within *Colletotrichum lindemuthianum* isolates from Andean and Mesoamerican bean varieties and regions. *Eur J Plant Pathol*. 110:253–63.
- Mentlak TA, Kombrink A, Shinya T, Ryder LS, Otomo I, Saitoh H, Terauchi R, Nishizawa Y, Shibuya N, Thomma BP, Talbot NJ. (2011) Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. *Plant Cell*. 24:322–335.
- Mota SF, Barcelos QL, Dias MA, Souza EA. (2016) Variability of *Colletotrichum* spp. in common bean. *Genet Mol Res*. 15. doi: 10.4238/gmr.15027176.
- Motion GB, Amaro TM, Kulagina N, Huitema E. (2015) Nuclear processes associated with plant immunity and pathogen susceptibility. *Brief Funct Genomics*. 14:243-52.
- O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J et al. (2012). Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat Genet*. 44:1060-1065.
- Padder BA, Sharma PN, Sharma OP, Kapoor V. (2007) Genetic diversity and gene flow estimates among five populations of *Colletotrichum lindemuthianum* across Himachal Pradesh. *Physiol Mol Plant Pathol*. 70:8–12.
- Perfect SE, Hughes HB, O'Connell RJ, Green JR. (1999) *Colletotrichum*: a model genus for studies on pathology and fungal–plant interactions. *Fungal Genet Biol*. 27:186–198.

- Rafiqi M, Ellis JG, Ludowici VA, Hardham AR, Dodds PN. (2012) Challenges and progress towards understanding the role of effectors in plant-fungal interactions. *Curr Opin Plant Biol.* 15:477-482.
- Rodríguez-Guerra R, Acosta-Gallegos J, González-Chavira M, Simpson J. (2006) Patotipos de *Colletotrichum lindemuthianum* y su implicación en la generación de cultivares resistentes de frijol. *Agricultura técnica en México.* 32:101-114.
- Sanz-Martín J, Pacheco-Arjona J, Bello-Rico, V, Vargas W, Monod M, Díaz-Mínguez, J, Thon M, Sukno S. (2016) A highly conserved metalloprotease effector enhances virulence in the maize anthracnose fungus *Colletotrichum graminicola*. *Mol. Plant Pathol.* 36:1–15.
- Schwartz HF, Steadman JR, Hall R, Forster RL. (2005) *Compendium of bean diseases.* 2nd ed. St Paul, MN: APS Press.
- Selin C, Kievit TR de, Belmonte MF, Fernando WGD. (2016) Elucidating the role of effectors in plant-fungal interactions: Progress and Challenges *Front. Microbiol.* 7:1–21.
- Sharma PN, Sharma OP, Padder BA, Kapil R. (2008) Yield loss assessment in common bean due to anthracnose (*Colletotrichum lindemuthianum*) under sub temperate conditions of North-Western Himalayas. *Indian Phytopath.* 61:323-330.
- Takahara H, Kombrink A, Hughes HB, et al. (2016) *Colletotrichum higginsianum* extracellular LysM proteins play dual roles in appressorial function and suppression of chitin-triggered plant immunity. *New Phytol.* 11:1323-337.
- Vargas WA, Sanz-Martín JM, Rech GE, Armijos-Jaramillo VD, Rivera LP., Echeverria, MM, Díaz-Mínguez JM, Thon MR, Sukno SA. (2016) A fungal effector with host nuclear localization and DNA-binding properties is required for maize anthracnose development. *Mol. Plant-Microbe Interact.* 29:83–95.



- Viswanathan R, Prasanth CN, Malathi P, Sundar AR. (2016) Draft Genome Sequence of *Colletotrichum falcatum* - A Prelude on Screening of Red Rot Pathogen in Sugarcane. J Genomics. 4:1-3.
- Wang C, Wang G, Zhang C, Zhu P, Dai H, Yu N, He Z, Xu L, Wang E. (2017) OsCERK1-mediated chitin perception and immune signaling requires Receptor-like Cytoplasmic Kinase 185 to activate a MAPK cascade in rice. Mol Plant. pii: S1674-2052(17)30007-2.
- Zampounis A, Pigné S, Dallery JF, Wittenberg AH, Zhou S, Schwartz DC, Thon MR, O'Connell RJ. (2016) Genome sequence and annotation of *Colletotrichum higginsianum*, a causal agent of crucifer anthracnose disease . Genome Announc. 4. pii: e00821-16.

## CAPITULO 1

### **DRAFT GENOME SEQUENCES OF TWO ISOLATES OF *Colleotrichum lindemuthianum*, THE CAUSAL AGENT OF ANTHRACNOSE IN COMMON BEANS**

Queiroz CB, Correia HLN, Menicucci RP, Vidigal PMP, Queiroz MV (2017). Draft genome sequences of two isolates of *Colleotrichum lindemuthianum*, the causal agent of anthracnose in common beans. Publicado em Genome Announcements. 5: pii: e00214-17. doi: 10.1128/genomeA.00214-17.

**Draft genome sequences of two isolates of *Colleotrichum lindemuthianum*, the causal agent of anthracnose in common beans**

Casley Borges de Queiroz<sup>a</sup>, Hilbert L. Nunes Correia<sup>a</sup>, Renato Pedrozo Menicucci<sup>a</sup>, Pedro M. Pereira Vidigal<sup>b</sup> and Marisa Vieira de Queiroz <sup>#a</sup>.

<sup>a</sup>Departamento de Microbiologia, Instituto de Biotecnologia Aplicada a Agropecuária (BIOAGRO), CEP: 36570-900, Universidade Federal de Viçosa, Minas Gerais, Brazil

<sup>b</sup>Núcleo de Análise de Biomoléculas (NuBioMol), Centro de Ciências Biológicas, Universidade Federal de Viçosa, Minas Gerais, Brazil

<sup>#</sup>Corresponding author.

## **ABSTRACT**

*Colletotrichum lindemuthianum* is the causal agent of anthracnose in common beans, one of the main limiting factors of their culture. Here, we report for the first time, to our knowledge, a draft of the complete genome sequences of two isolates belonging to 83.501 and 89 A2 2-3 of *C. lindemuthianum*.

## GENOME ANNOUNCEMENT

*Colletotrichum* is one of the most widespread and economically important genera among the various damaging plant pathogens (1). This is especially true in tropical and subtropical regions, where *Colletotrichum* spp. limit the production of important crops, causing severe social and economic impacts (2). *Colletotrichum lindemuthianum* is the causal agent of anthracnose in the common bean (*Phaseolus vulgaris* L.) and attacks its leaves, stems, branches, string beans, and seeds (3). Under favorable conditions of humidity and temperature, in extreme cases, anthracnose can cause death of the plant (4). *C. lindemuthianum* has a high genetic variability, manifested by the presence of many physiologically diverse races (5). This high genetic variability is one of the main limiting factors to combating this pathogen, because it prevents the long-term use of resistant cultivars (6). The genomes of two isolates, 83.501 and 89 A<sub>2</sub> 2-3, of *C. lindemuthianum* were sequenced using the HiSeq 2500 Illumina platform with paired-end reads of 100 bp and an average coverage of 89.3 for isolate 83.501 and 157.7 for isolate 89 A<sub>2</sub> 2-3 (Laboratório Central de Tecnologias de Alto Desempenho em Ciências de Vida [LaCTAD]). The genes were predicted using Augustus version 3.2.2 (7). For de novo assembly, we used CLC Genomics Workbench 6.5.1. The completeness of the genome was estimated using Benchmarking Universal Single-Copy Orthologs (BUSCO) (8). The contig assembly of isolate 83.501 using the CLC Genomics Workbench software resulted in 1,857 contigs (N<sub>50</sub>, 111.275 kbp), the smallest containing a 1.02-kbp and the largest containing a 623.12-kbp, for a total of 97.4 Mbp and a GC content of 37.6%. The contig assembly of isolate 89 A<sub>2</sub> 2-3 resulted in 1,276 contigs (N<sub>50</sub>, 158.217 kb), the smallest containing 1.03 kbp and the largest containing 1.11 Mbp, for a total of 99.16 Mbp, with a GC content of 37.3%. Assessment of the completeness of the genome using

1,438 BUSCO groups for fungi resulted in 98% complete (C) (6.1% duplicated [D]), 1.0% fragmented (F), 0.2% missed (M), and 1,438 genes (n) for isolate 83.501 and 98% C (6.0% D), 1.1% F, 0.4% M, and 1,438 n for isolate 89 A<sub>2</sub> 2-3, indicating that the assembled genome covered most of the coding regions. There were 11,673 (isolate 83.501) and 11,627 (isolate 89 A<sub>2</sub> 2-3) predicted genes.

In this study, we present a draft of the genome sequences belonging to two isolates from different races of *C. lindemuthianum*. This information will provide a valuable resource for identifying the mechanisms of pathogenesis and the plant-pathogen interactions, as well as new pathogenicity factors. Our data will also allow for comparative genomic studies with other fungal plant pathogens.

**Accession number(s).** The sequences from this whole-genome shotgun project for the two strains have been deposited in the DDBJ/ENA/GenBank database under accession numbers MASO00000000 and MASP00000000. The versions described in this paper are the first versions.

## **ACKNOWLEDGMENTS**

This work was supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG, and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

Research conducted with the help of the Cluster UFV (Universidade Federal de Viçosa).

## **FOOTNOTES**

Address correspondence to Marisa Vieira de Queiroz, mvqueiro@hotmail.com.

## REFERENCES

1. O'Connell RJ<sup>1</sup>, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, Damm U, Buiate EA, Epstein L, Alkan N, Altmüller J, Alvarado-Balderrama L, Bauser CA, Becker C, Birren BW, Chen Z, Choi J, Crouch JA, Duvick JP, Farman MA, Gan P, Heiman D, Henrissat B, Howard RJ, Kabbage M, Koch C, Kracher B, Kubo Y, Law AD, Lebrun MH, Lee YH, Miyara I, Moore N, Neumann U, Nordström K, Panaccione DG, Panstruga R, Place M, Proctor RH, Prusky D, Rech G, Reinhardt R, Rollins JA, Rounsley S, Schardl CL, Schwartz DC, Shenoy N, Shirasu K, Sikhakolli UR, Stüber K, et al. 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nat Genet. 44:1060-1065. doi:10.1038/ng.2372.
2. Cannon PF, Damm U, Johnston PR, Weir BS. 2012. *Colletotrichum* - current status and future directions. Stud Mycol. 73:181-213. doi:10.3114/sim0014.
3. Mahuku SG, Riascos JJ. 2004. Virulence and molecular diversity within *Colletotrichum lindemuthianum* isolates from Andean and Mesoamerican bean varieties and regions. Eur J Plant Pathol 110: 253–63. doi:10.1023/B:EJPP.0000019795.18984.74.
4. Campa A, Rodríguez-Suárez C, Giraldez R, Ferreira JJ. 2014. Genetic analysis of the response to eleven *Colletotrichum lindemuthianum* races in a RIL population of common bean (*Phaseolus vulgaris* L.). BMC Plant Biol. 14:115. doi:10.1186/1471-2229-14-115.

5. Santos LV, de Queiroz MV, Santana MF, Soares MA, de Barros EG, de Araújo EF, Langin T. 2012. Development of new molecular markers for the *Colletotrichum* genus using RetroCl1 sequences. *World J Microbiol Biotechnol.* 28:1087-1095. doi:10.1007/s11274-011-0909-x.
6. Ansari KI, Palacios N, Araya C, Langin T, Egan D, Doohan FM. 2004. Pathogenic and genetic variability among *Colletotrichum lindemuthianum* isolates of different geographic origins. *Plant Pathol.* 53:635-642. doi:10.1111/j.0032-0862.2004.01057.x.
7. Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* 32:W309–W312. doi:10.1093/nar/gkh379.
8. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 31::3210-3212. doi: 10.1093/bioinformatics/btv351.



## CAPITULO 2

### **ANÁLISE COMPARATIVA DO GENOMA MITOCONDRIAL DO FUNGO *Colletotrichum lindemuthianum*, AGENTE CAUSAL DA ANTRACNOSE EM FEIJOEIRO**

Queiroz CB, Santana MF, Vidigal PMP, Queiroz MV. Análise comparativa do genoma mitocondrial do fungo *Colletotrichum lindemuthianum*, agente causal da antracnose em feijoeiro. Applied Microbiology and Biotechnology. Em preparação.

**Análise comparativa do genoma mitocondrial do fungo *Colletotrichum lindemuthianum*, agente causal da antracnose em feijoeiro**

Casley Borges de Queiroz<sup>1</sup>, Mateus Ferreira Santana<sup>1</sup>, Pedro M. Pereira Vidigal<sup>2</sup>, Marisa Vieira de Queiroz<sup>1\*</sup>

<sup>1</sup>Laboratório de Genética Molecular de Fungos (LGMF)/BIOAGRO, CEP: 36570-900, Universidade Federal de Viçosa, Minas Gerais, Brazil

<sup>2</sup>Núcleo de Análise de Biomoléculas (NuBioMol), Centro de Ciências Biológicas, Universidade Federal de Viçosa, Minas Gerais, Brazil

\*Corresponding author

Laboratório de Genética Molecular de Fungos (LGMF)/BIOAGRO, CEP: 36570-900, Universidade Federal de Viçosa, Minas Gerais, Brazil. +55 (31) 3899-2975. E-mail address: [mvqueiro@ufv.br](mailto:mvqueiro@ufv.br).

## RESUMO

Fungos do gênero *Colletotrichum* são importantes economicamente e são usados como modelos em estudos da interação planta-patógeno. Neste trabalho foram anotados e analisados os genomas mitocondriais completos de dois isolados de *Colletotrichum lindemuthianum* e foram comparados com os genomas mitocondriais de sete espécies do gênero pertencentes aos clados acutatum, graminicola, orbiculare e gloeosporioides. O genoma mitocondrial de *C. lindemuthianum* é uma típica molécula circular de 3.7446 pb (isolado 89 A<sub>2</sub> 2-3) e 3.7440 pb (isolado 83.501). A diferença de seis nucleotídeos entre os dois genomas é o resultado de uma deleção no gene *rps3* no isolado 83.501 e foi observada a troca de adenina por guanina na posição 3.457 no genoma mitocondrial do isolado 83.501 dentro do gene *rps3* (proteína ribossomal S3). Foram encontrados 14 genes *core* que codificam proteínas relacionados a atividade de fosforilação oxidativa e síntese de ATP, dois genes das subunidades ribossomais maior e menor, um gene da proteína ribossomal S3 codificada por um íntron, 29 genes de RNA transportadores (tRNAs), cinco *open read frame* não conservadas (ncORF) e um total de cinco íntrons que codificam duas endonucleases *homing*. A maioria dos genes *core* apresentam sintênia. As ncORFs variam em número e são codificadas por ambas as fitas. O número de íntrons é altamente variável e codificam um total de até nove endonucleases *homing*. A variação no tamanho dos genomas pode ser atribuída principalmente as ncORF e íntrons do tipo I. Os dados de filogenia revelaram que os genes *core* mitocondriais podem ser utilizados como alternativa em estudos de relação filogenética em *Colletotrichum* spp. Este trabalho contribue para os estudos de genética e biologia de *Colletotrichum* spp. que são de grande importância em âmbito econômico e científico.

**Palavras chave:** Genoma mitocondrial, antracnose, introns que codificam proteínas, ORFs não conservadas (ncORF)

## INTRODUÇÃO

A principal função da mitocôndria está relacionada com a produção de energia por meio da fosforilação oxidativa (Saraste 1999). O conteúdo do DNA mitocondrial é bastante conservado entre os organismos eucariotos e tipicamente codifica 14 proteínas relacionadas ao complexo de cadeia respiratória, duas subunidades ribossomais, um grupo variável de RNA transportadores e de *open read frame* de função desconhecida (uORF) (Aguileta et al. 2014; Torriani et al. 2014) também chamadas de ORFs não conservadas (ncORF) por serem pobremente conservadas, mesmo entre espécies próximas e seus papéis ainda são pouco conhecidos. Genomas mitocondriais normalmente são moléculas de DNA circulares caracterizadas por um alto conteúdo de A e T (Burger et al. 2003; Torriani et al. 2014). O tamanho dos genomas mitocondriais de fungos varia de 12 kb no micoparasita *Rozella allomyces* (James et al. 2013) até 235,8 kb no fungo *Rhizoctonia solani* (Losada et al. 2014). Vários fatores contribuem para variação no tamanho dos genomas mitocondriais, incluindo a proliferação de sequências não-codificantes como sequências repetitivas curtas, duplicação gênica seguida por inativação, expansão de introns ou outras sequências de diferentes fontes (Burger et al. 2003; Losada et al. 2014).

Íntrons mitocondriais são classificados em dois grupos, I e II, de acordo com suas estruturas secundárias conservadas de RNA, sendo predominante para genomas mitocôndrias de fungos os íntrons do grupo I, enquanto que o grupo II é dominante em genomas mitocondriais de plantas (Lang et al., 2007). Íntrons do grupo I são divididos em seis principais subgrupos (IA, IA3, IB, IC1, IC2 e ID) baseado em análises filogenéticas. Íntrons do grupo I e II são capazes de se mover e proliferar nos genomas mitocondriais por meio de um mecanismo denominado de íntron *homing*, no qual o movimento do íntron para um gene sem íntron é mediado por

endonucleases codificadas por uma ORF geralmente presente no próprio intron, denominadas de endonucleases *homing* (LAGLIDADG e GIY-YIG, nos introns do grupo I) (Losada et al., 2014; Lang et al., 2007). O número de introns é altamente variável entre os genomas mitocondriais, por exemplo, *Fusarium graminearum* possui 34 introns do grupo I (Al-Reedy et al. 2012), enquanto que *Mycosphaerella graminicola* não apresenta íntrons (Torriani et al. 2008).

Genomas mitocondriais possuem um maior número de cópias e evolução mais rápida do que o genoma nuclear (Ballard e Whitlock 2004) devido a uma grande exposição a espécies reativas de oxigênio e um sistema de reparo de DNA menos eficiente (Kang e Hamasaki 2002). Isso tem permitido que os genes mitocondriais sejam utilizados para estudos evolutivos em diversas espécies de fungos, como no fungo endofítico *Phialocephala scopiformis* (Robicheau et al. 2016), no patógeno do bambu *Shiraia bambusicola* (Shen et al. 2015), no entomopatógeno *Metarhizium anisopliae* (Ghikas et al. 2006), além de outras 38 espécies de fungos (Aguileta et al. 2014) e em outros organismos eucariotos, por exemplo, em *Somniosus microcephalus* conhecido como tubarão-da-groenlândia (Santaquiteria et al. 2017), em insetos (Yang et al. 2015) e em pássaros (Slack et al. 2007).

O gênero *Colletotrichum* consiste de vários clados monofiléticos que são referidos como um agregado de espécies nomeados de acordo com a espécie mais proeminente do grupo, sendo que atualmente existem nove clados nomeados de acutatum, graminicola, spaethianum, destructivum, dematium, gloeosporioides, boninense, truncatum e orbiculare (Cannon et al. 2012; Crouch et al. 2014). Embora existam aproximadamente 600 espécies de *Colletotrichum* que atacam mais de 3.200 espécies de plantas (O'Connell et al. 2012), ainda existe pouco estudo sobre genomas mitocondriais de *Colletotrichum* spp. Até o momento, existem 442 genomas mitocôndrias de fungos sequenciados e depositado no *NCBI Organelle Genome database*, porém existem somente nove espécies de *Colletotrichum* com genomas mitocôndrias sequenciados, incluindo espécies do clado acutatum, graminicola, orbiculare e gloeosporioides (Gutiérrez et al. 2014; Kim et al. 2015; Liang et al. 2017).

Recentemente o genoma mitocondrial de *Colletotrichum lindemuthianum* (Sacc. & Magnus) Briosi & Cavara foi sequenciado (Gutiérrez et al. 2014). Este fungo é o causador da antracnose no feijão comum (*Phaseolus vulgaris*), sendo um dos patógenos mais disseminado e destrutivo para essa cultura, principalmente em regiões tropicais da América do Sul e África Oriental, onde as condições climáticas favorecem o desenvolvimento da doença, que causa grandes perdas na produção (Fernandez et al. 2000; Pastor-Corrales e Tu 1989; Sharma et al. 2008). *Colletotrichum lindemuthianum* possui uma alta variabilidade genética (dos Santos et al. 2012; Mota et al. 2016; Padder et al. 2007) possuindo mais de 100 raças patogênicas (Alzate-Marin et al. 1999; Ferreira et al. 2013; González et al. 2015; Rodríguez-Guerra et al. 2006;). Em consequência dessa alta variabilidade do patógeno, a resistência de cultivares é frequentemente suplantada (Ferreira et al. 2013).

Neste trabalho, foram anotados e analisados os genomas mitocondriais completos de mais dois isolados de *C. lindemuthianum* pertencentes às raças fisiológicas 83 e 89, e foram comparados com mais oito genomas mitocondriais de sete espécies de *Colletotrichum* pertencentes aos clados acutatum, graminicola, orbiculare e gloeosporioides. Os dados deste trabalho completam os resultados obtidos anteriormente e demonstram que existe uma grande variedade em números de ncORFs e introns entre os genomas mitocôndriais de *Colletotrichum* spp. Entretanto, com exceção das espécies do clado gloeosporioides, as regiões intergênicas em termos de tamanho, são similares.

## **MATERIAL E MÉTODOS**

### **Sequenciamento de DNA e montagem**

Foram usados os genomas mitocondriais dos isolados 83.501 e 89 A<sub>2</sub> 2-3 de *C. lindemuthianum* pertencentes à coleção do Laboratório de Genética Molecular de Fungos da Universidade Federal de Viçosa. As sequências dos DNAs mitocôndriais foram obtidas e montadas juntamente

com as sequências dos genomas nucleares realizadas por Queiroz et al. (2017) usando a plataforma Illumina HiSeq2500 com *paired-end reads* de 100 bp com média de cobertura para os genomas mitocondriais de 6020,98 e 12376,93X para os isolados 83.501 e 89 A<sub>2</sub> 2-3, respectivamente. A montagem *de novo* foi realizada por meio do programa CLC Genomics Workbench 6.5.1.

### **Anotação e análise do genoma mitocondrial de *C. lindemuthianum***

Os genomas mitocondriais dos isolados 89 A<sub>2</sub> 2-3 e 83.501 de *C. lindemuthianum* foram anotados automaticamente usando o MITOS web server (Bernt et al., 2013) e a ferramenta MFannot (<http://megasun.bch.umontreal.ca/cgi-in/mfannot/mfannotInterface.pl>) usando código genético 4, que inclui código genético de fungos filamentosos. Cada gene anotado foi manualmente revisado utilizando alinhamento com sequências de genes mitocondriais de espécies próximas e por meio de BLASTp (*Basic Local Alignment Search Tool*) usando sequências traduzidas pela ferramenta ExPASy (<http://web.expasy.org/translate/>) usando o código 4 para cada gene que codifica proteína. MITOS web server também foi utilizado para re-anotação de genes de RNA transportadores para os genomas mitocondriais de outras espécies de *Colletotrichum* analisadas e a ferramenta MFannot foi utilizada para identificação dos introns e das ncORFs também nas outras espécies de *Colletotrichum*. As estruturas secundárias dos RNAs transportadores foram preditas usando o programa ARWEN (Laslett e Canbäck, 2008) com parâmetros default. O uso de codons e os aminoácidos mais frequentes foram preditos por meio da ferramenta Codon Usage ([http://www.bioinformatics.org/sms2/codon\\_usage.html](http://www.bioinformatics.org/sms2/codon_usage.html)) com o código genético 4. A ferramenta GenomeVx (<http://wolfe.ucd.ie/GenomeVx/>) foi utilizada para plotar a posição e orientação de cada gene dos genomas mitocondriais de *C. lindemuthianum*. As regiões conservadas dos genomas mitocondriais de *C. lindemuthianum* em relação aos genomas mitocondriais das sete espécies de *Colletotrichum* aqui analisadas foram visualizadas por meio da ferramenta BRIG (Alikhan et al. 2011). O mapeamento das regiões

com alterações nas sequências de bases dos genomas mitocondriais de *C. lindemuthianum* foi feita por meio do programa CLC Genomics Workbench 6.5.1 usando as *reads* do sequenciamento de cada isolado.

### **Análise filogenética**

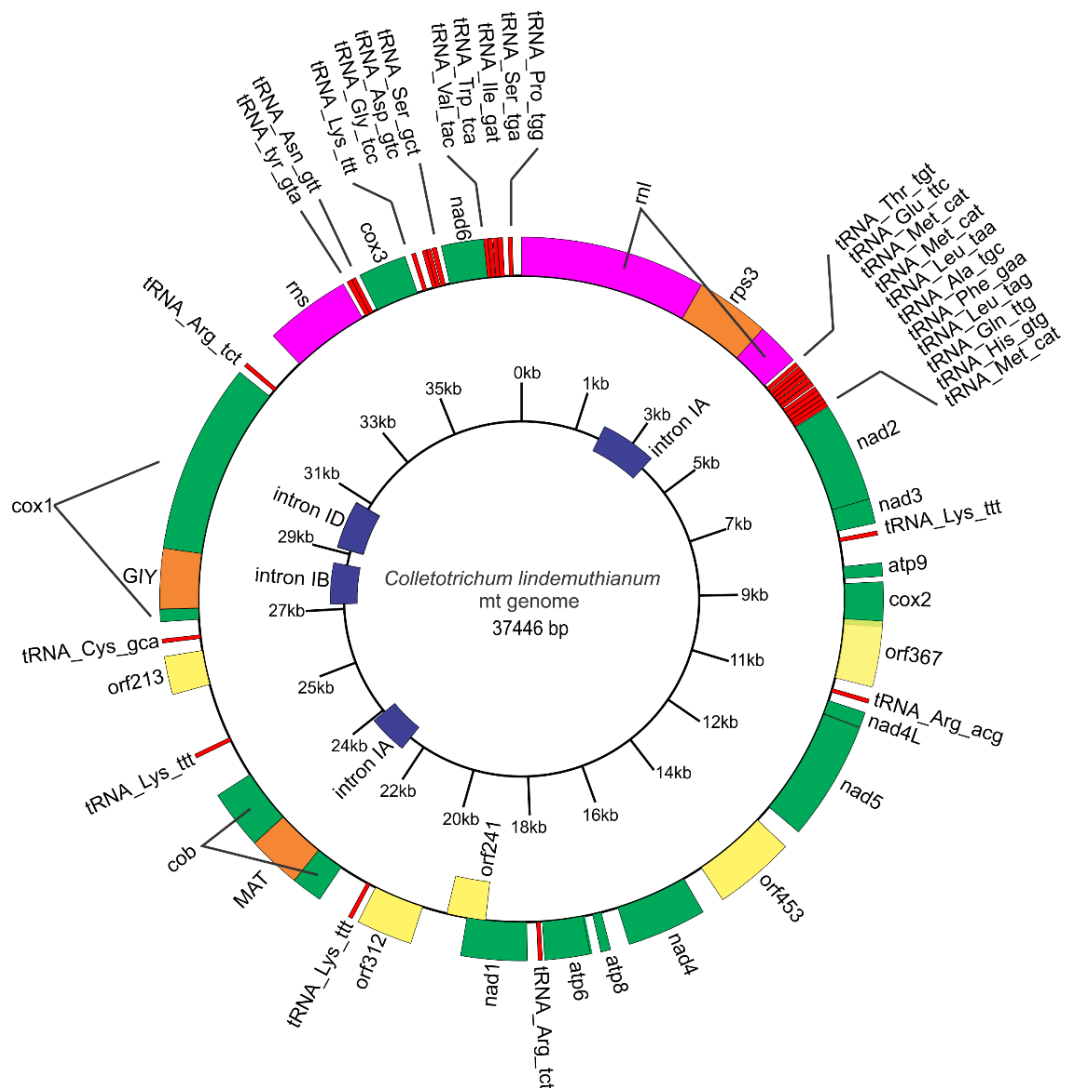
Para analisar se os genes mitocondriais de *Colletotrichum* spp. seriam úteis para determinar as relações filogenéticas entre essas espécies, foram utilizadas as sequências de nucleotídeos de 14 genes conservados que codificam proteínas, também chamados de genes *core*, incluindo *cox1*, *cox2*, *cox3*, *cob*, *atp6*, *atp8*, *atp9*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* e *nad6* e cada gene foi alinhado pelo ClustalW implementado no programa MEGA versão 7 (Kumar et al. 2016). Posteriormente, os alinhamentos foram tratados com o trimAl (Capella-Gutiérrez et al. 2009) para remover as regiões das sequências de nucleotídeos mal alinhadas. De acordo com o critério de Informação de Akaike (AIC) o modelo de melhor ajuste para o alinhamento das sequências de nucleotídeos foi determinado pelo programa MrModeltest versão 2.3 (Nylander, 2004), que foram GTR+G para *nad2*, GTR para *ad3*, GTR+I para *atp9*, GTR+I para *cox2*, GTR+I para *nad4L*, GTR+G para *nad5*, GTR+I para *nad4*, F81+G para *atp8*, GTR+G para *atp6*, GTR+G para *nad1*, HKY+G para *cob*, GTR para *cox1*, GTR+G para *cox3* e GTR+G para *nad6*. Os alinhamentos foram concatenados por meio do programa Mesquite versão 3.2 (Maddison e Maddison, 2017) e a árvore filogenética foi construída baseada na inferência Bayesiana com o programa MrBayes versão 3.2 (Ronquist et al. 2012). Análise Bayesiana incluiu duas corridas separadas de  $1 \times 10^7$  gerações, e amostradas a cada 1.000 gerações e 25% das gerações iniciais foram descartadas como *burn-in*.

## **RESULTADOS**

### **O genoma mitocondrial de *C. lindemuthianum***



Os genomas mitocondriais de *C. lindemuthianum* (GenBank accession – em preparação) são uma típica molécula de DNA dupla fita circular. O tamanho é de 37.446 pb para o isolado 89 A<sub>2</sub> 2-3 e 37.440 pb para o isolado 83.501, sendo semelhante ao tamanho dos outros sete genomas mitocondriais das outras espécies de *Colletotrichum* analisadas neste trabalho que variam de 30.020 pb em *C. fiorinae* até 58.666 pb em *C. siamense* (Arquivo adicional: Tabela S1). O genoma mitocondrial de *C. lindemuthianum* contém um total de 53 genes, incluindo um grupo de 14 genes que codificam proteínas conservadas em genomas mitocondriais de fungos. Foram também encontrados dois RNAs ribossomais, 29 RNAs transportadores, três ORFs intrônicas e cinco ncORFs (Figura 1 e Tabela 1). Os 14 genes que codificam proteínas mitocôndriais conservadas, codificam sete subunidades do complexo I da cadeia de transporte de elétrons (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* e *nad6*), três citocromo oxidases (*cox1-3*), uma subunidade do complexo I (*cob*) e três da subunidade F<sub>0</sub> do complexo ATP sintase (*atp6*, *8* e *9*). Em *C. lindemuthianum*, exceto uma ncORF (*orf241*), os genes são codificados pela fita positiva orientados no sentido horário.



**Figura 1.** O genoma mitocondrial de *Colletotrichum lindemuthianum* 89 A<sub>2</sub> 2-3. Os 14 genes *core* são apresentados em verde; genes de RNA transportadores são apresentados em vermelho; genes de RNA ribossomais são apresentados em rosa; ORFs intrônicas são apresentadas em laranja; Introns são apresentados em azul no anel interno; ncORFs são apresentadas em amarelo. ORF no lado interno do anel maior representa gene localizado na fita negativa do genoma mitocondrial.

**Tabela 1.** Características e posição dos genes no genoma mitocondrial de *Colletotrichum lindemuthianum*.

Gene	Região	Tamanho	GC%	Codon inicial	Codon de parada	Anticodon
<i>rnl</i>	1-5064	5064	32,66	-	-	-

<i>rps3</i>	3101-4345	1245	23,45/23,16*	ATG	TAA	-
<i>trnT</i>	5119-5189	71	42,25	-	-	UGU
<i>trnE</i>	5206-5278	73	43,83	-	-	UUC
<i>trnM</i>	5280-5348	69	31,88	-	-	CAU
<i>trnM</i>	5354-5426	73	34,24	-	-	CAU
<i>trnL2</i>	5438-5520	83	39,76	-	-	UAA
<i>trnA</i>	5530-5601	72	37,50	-	-	UGC
<i>trnF</i>	5628-5700	73	39,72	-	-	GAA
<i>trnL1</i>	5701-5784	84	41,66	-	-	UAG
<i>trnQ</i>	5791-5863	73	31,50	-	-	UUG
<i>trnH</i>	5865-5937	73	45,20	-	-	GUG
<i>trnM</i>	5939-6011	73	32,87	-	-	CAU
<i>nad2</i>	6026-7693	1668	22,84	ATG	TAA	-
<i>nad3</i>	7693-8103	411	23,11	ATG	TAA	-
<i>trnK</i>	8233-8305	73	31,50	-	-	UUU
<i>atp9</i>	8764-8988	225	36,00	ATG	TAA	-
<i>cox2</i>	9078-9824	747	29,85	ATG	TAA	-
<i>orf367</i>	9728-10831	1104	40,58	ATG	TAA	
<i>trnR</i>	11035-11105	71	38,02	-	-	ACG
<i>nad4L</i>	11265-11534	270	24,07	ATG	TAA	-
<i>nad5</i>	11534-13531	1998	28,07	ATG	TAA	-
<i>orf453</i>	13850-15211	1362	32,96	ATG	TAA	
<i>nad4</i>	15590-16921	1332	25,52	ATG	TAA	-
<i>atp8</i>	17240-17404	165	23,63	ATG	TAA	-
<i>atp6</i>	17548-18339	792	26,64	ATG	TAA	-

<i>trnR</i>	18379-18449	71	35,21	-	-	UCU
<i>nad1</i>	18628-19734	1107	29,81	ATG	TAA	-
<i>orf241</i>	19389-20114	726	25,48	ATG	TAA	
<i>orf312</i>	20578-21516	939	32,27	ATG	TAA	
<i>trnK</i>	21628-21699	72	27,77	-	-	UUU
<i>cob</i>	22247-24661	2415	27,12	ATG	TAA	
<i>Mat</i>	22769-23665	897	22,52	ATG	TAA	
<i>trnK</i>	25371-25441	71	36,62	-	-	UUU
<i>orf213</i>	26495-27136	642	41,74	ATG	TAA	
<i>trnC</i>	27354-27422	69	36,23	-	-	GCA
<i>cox1</i>	27706-32121	4416	30,73	ATG	TAG	-
<i>Giy</i>	27919-28923	1005	28,45	GAC	TAA	-
<i>trnR</i>	32246-32316	71	35,21	-	-	UCU
<i>rns</i>	32935-34376	1442	29,88	-	-	-
<i>trnY</i>	34443-34529	87	37,93	-	-	GUA
<i>trnN</i>	34532-34602	71	36,62	-	-	GUU
<i>cox3</i>	34682-35491	810	31,85	ATG	TAA	-
<i>trnK</i>	35602-35674	73	30,13	-	-	UUU
<i>trnG</i>	35785-35855	71	42,25	-	-	UCC
<i>trnD</i>	35858-35930	73	42,46	-	-	GUC
<i>trnS1</i>	35949-36029	81	41,97	-	-	GCU
<i>nad6</i>	36118-36825	708	27,02	ATG	TAG	-
<i>trnV</i>	36816-36887	72	41,66	-	-	UAC
<i>trnW</i>	36890-36960	71	39,43	-	-	UCA
<i>trnI</i>	36972-37042	71	28,16	-	-	GAU

<i>trnS2</i>	37048-37133	86	48,83	-	-	UGA
<i>trnP</i>	37221-37293	73	50,68	-	-	UGG

---

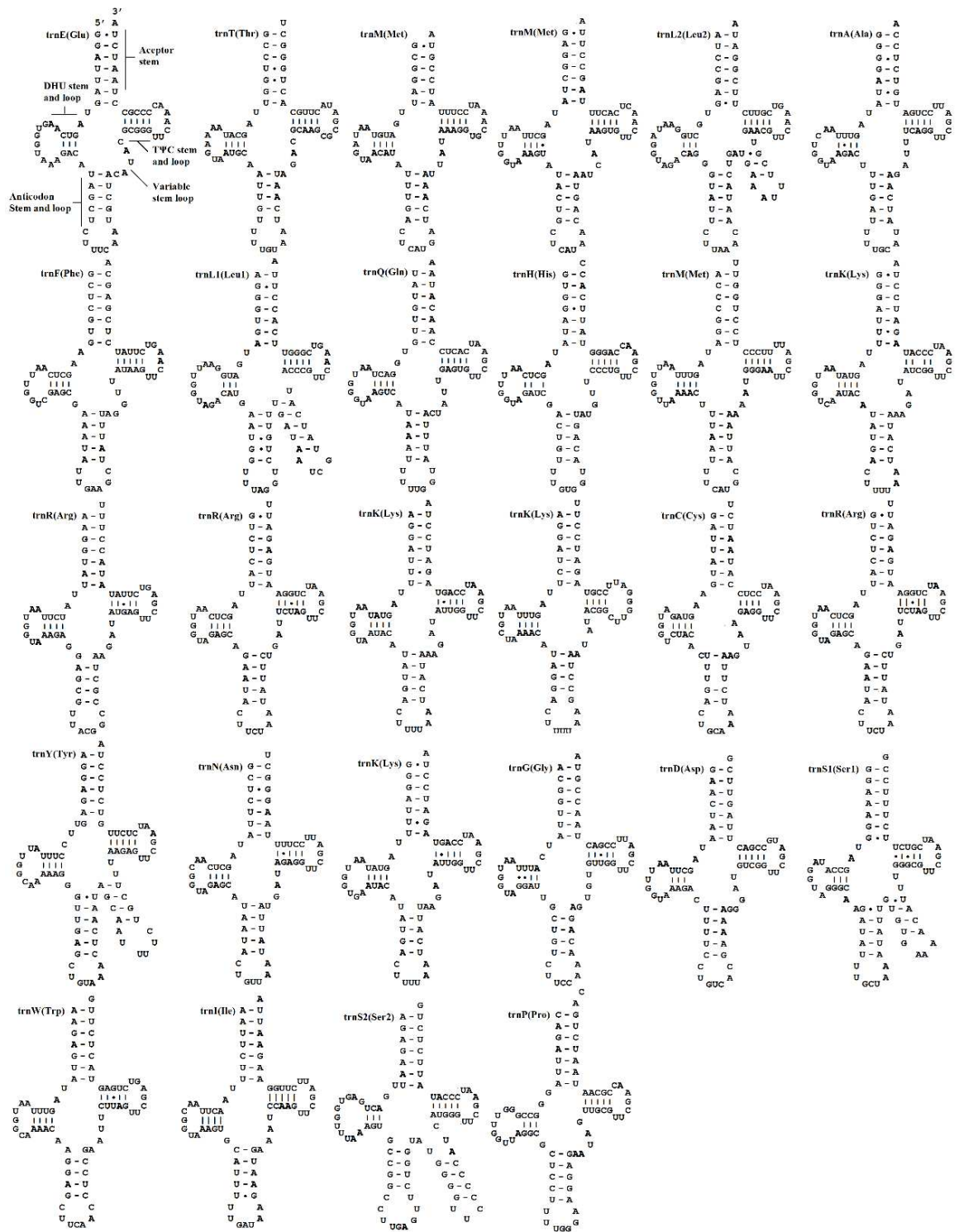
\**rps3* de i89 e i83 respectivamente

Foi observado no genoma mitocondrial do isolado de *C. lindemuthianum* sequenciado e anotado por Gutiérrez et al. (2014), a ausência do gene *nad3*. Entretanto, ao compararmos a sequência de *nad3* dos isolados 89 A<sub>2</sub> 2-3 e 83.501 com a sequência do genoma mitocondrial do isolado de *C. lindemuthianum* previamente publicado por Gutiérrez et al. (2014), foi verificado que as sequências nestes isolados estão conservadas, exceto pela ausência de uma timina que levou a perda do quadro de leitura no isolado sequenciado por Gutiérrez et al. (2014). Foi realizado um mapeamento da sequência do *nad3* no isolado 89 A<sub>2</sub> 2-3 com as *reads* de nosso sequenciamento e foi observado que em nossos isolados essa região foi montada corretamente com a presença de timina na maioria das *reads* sobrepostas na posição 7.702 pb (Arquivo adicional: Figura S1). Quando foi realizada uma comparação entre os genomas mitocondriais dos isolados 89 A<sub>2</sub> 2-3 e 83.501, foi observado uma deleção de seis bases (GCGGCG) entre as posições 3.457 e 3.458 pb e uma troca de adenina por guanina na posição 3.457 no genoma mitocondrial do isolado 83.501 dentro do gene *rps3* (proteína ribossomal S3). Estas foram as únicas diferenças entre os dois genomas. Para analisar se isso era consequência de erros de montagem, foi realizado um mapeamento das sequências dos genomas dos dois isolados com suas respectivas *reads* do sequenciamento e foi observada a presença das seis bases na maioria das *reads* sobrepostas do isolado 89 A<sub>2</sub> 2-3 e ausência nas *reads* do isolado 83.501 com média de cobertura de 14 mil e 6 mil, respectivamente, o mesmo foi observado na troca de A por G e com a mesmas coberturas da deleção (Arquivo adicional: Figura S1). Como a deleção ocorreu em duas trincas de nucleotídeos, não ocorreu perda do quadro de leitura. Além disso, a troca de A por G no isolado 83.501 não levou a troca de aminoácidos, pois ocorreu na terceira base do códon CCA levando a substituição pela códon CCG, onde ambos codificam o aminoácido prolina. No genoma mitocondrial de *C. lindemuthianum*

previamente publicado por Gutiérrez et al. (2014), essa região é idêntica a região correspondente no genoma mitocondrial do isolado 89 A<sub>2</sub> 2-3.

### **RNAs transportadores de *C. lindemuthianum***

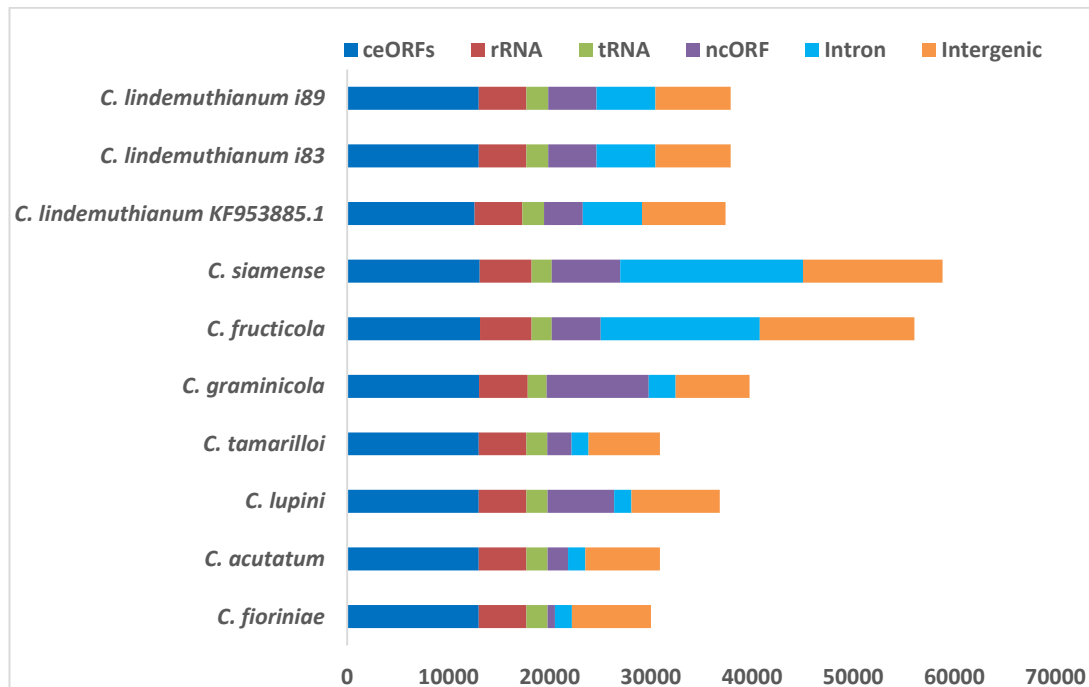
Um total de 29 tRNAs foram encontrados nos genomas mitocondriais de *C. lindemuthianum* correspondentes aos 20 aminoácidos (Tabela 1 e Figura 2). Em todas as espécies de *Colletotrichum* analisadas, o total de tRNA variou de 25 em *C. graminicola* até 29 em *C. lindemuthianum* (Arquivo adicional: Tabela S1), correspondendo aos 20 aminoácidos. O tamanho dos genes de tRNA de *C. lindemuthianum* variam de 69 pb (*trnM*) a 87 pb (*trnY*). Foram encontrados três e quatro genes de tRNA para metionina e lisina, respectivamente, onde os tRNAs para o mesmo aminoácido apresentam o mesmo anticódon. Para arginina foram encontrados três genes de tRNAs, mas somente dois com o mesmo anticódon. Dois genes de tRNA foram encontrados para leucina e serina, cada um com dois tipos de anticódons. Cinco dos 29 tRNAs de *C. lindemuthianum* apresentam um longo braço variável (*trnL*, *trnL2*, *trnS1*, *trnS2* e *trnY*). Nas estruturas secundárias de 29 tRNAs de *C. lindemuthianum*, um total de 57 pb preditas não pareiam corretamente. O total de 22 genes de tRNA foram encontrados em quatro *clusters*. O primeiro e maior *cluster* consiste de 11 genes de tRNA confinados em uma área de 893 pb entre os genes *rnl* e *nad2* (Figura 1). O segundo grupo consiste de dois genes, o terceiro inclui quatro genes e o quarto possui cinco genes de tRNA. Três genes de tRNAs estão dispersos no genoma.



**Figura 2.** Estrutura secundária putativa dos 29 tRNAs mitocondriais de *C. lindemutianum*.

**Open reading frames não conservadas (ncORFs), introns e ORFs intrônicas em *Colletotrichum* spp.**

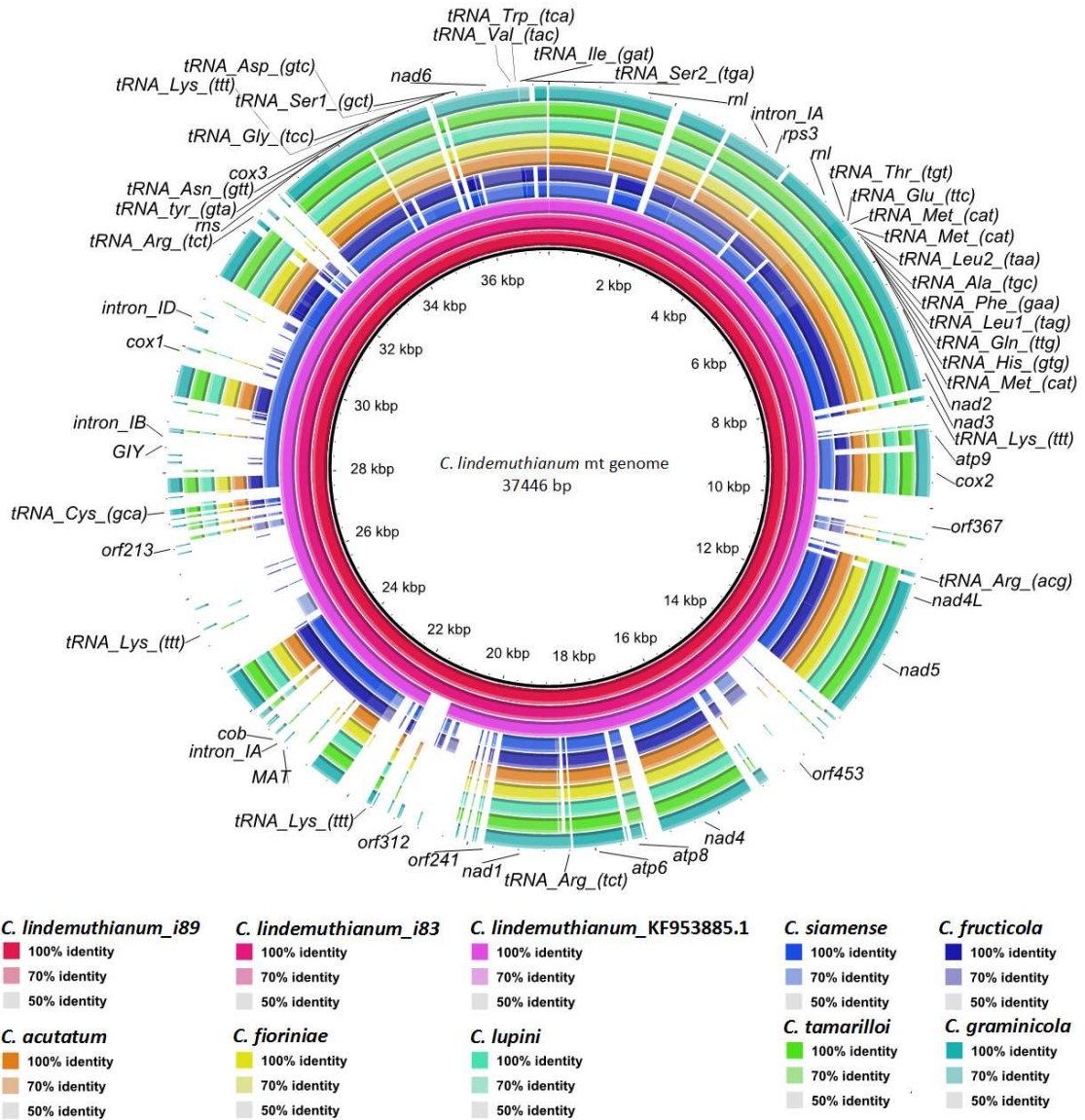
O tamanho total das regiões intrônicas no genoma mitocondrial de cada espécie de *Colletotrichum* foi altamente variável, com 1.694 pb em *C. tamarilloi* até 18.084 pb em *C. siamense*, representando 5,5% e 30,8% do tamanho total do genoma mitocondrial, respectivamente (Figura 3 e Arquivo adicional: Tabela S2). Em *C. lindemuthianum*, o total de quatro íntrons foram encontrados nos genes que codificam um RNA ribossomal e dois genes que codificam proteínas, incluindo um íntron do grupo IA no gene *cob* (*cytochrome b*) e *rnl* (*large subunit ribosomal RNA*). Na predição gênica realizada neste trabalho, foi observado no gene *cox1* um exon a mais de 212 pb *upstream* comparado a sequência do gene *cox1* predito no isolado de *C. lindemuthianum* de Gutiérrez et al. (2014). Foi realizado um alinhamento da sequência de *cox1* dos isolados 89 A<sub>2</sub> 2-3 e 83.501 com o genoma mitocondrial do isolado de Gutiérrez et al. (2014), e foi encontrada a conservação desse exon extra. Portanto, em *C. lindemuthianum* o gene *cox1* (*cytochrome c oxidase I*) possui um tamanho total de 4.416 pb incluindo dois íntrons, um do grupo IB e um do grupo ID.



**Figura 3.** Variação dos tamanhos das regiões gênicas, intergênicas e intrônicas nos genomas mitocondriais de *Colletotrichum* spp. As ceORFs são os exons dos genes *core*; as ncORFs são os genes não conservados. i89: isolado 89 A<sub>2</sub> 2-3; i83: isolado 83.501.



Sequências codificadoras intrônicas foram identificadas em três íntrons e elas codificam a proteína ribossomal S3 (IA no gene *rnl*), uma endonuclease LAGLIDADG (IA no gene *cob*) e uma endonuclease GIY (IB no gene *cox1*). Em *C. graminicola* foram encontrados dois íntrons, um do grupo IA dentro do gene *rnl* com uma ORF para a proteína ribossomal S3 e um íntron do grupo IB no gene *nad5* com uma ORF para uma endonuclease LAGLIDADG. Em *C. siamense* foram encontrados 11 íntrons, sendo dois do grupo IA em *cob* e *rnl*, cinco do grupo IB (quatro em *cox1* e um em *cox3*), um do grupo IC2 em *nad2*, dois do tipo ID em *cob* e *cox1* cada, e um íntron de grupo desconhecido em *cox1*. Em *C. fructicola* foram encontrados 10 íntrons no total, sendo dois do grupo IA em *cob* e *rnl*, seis do grupo IB (cinco em *cox1* e um em *nad5*), um do grupo ID em *cob* e um de grupo desconhecido em *cox1*. Para as demais espécies de *Colletotrichum* foi encontrado somente o íntron do grupo IA no gene *rnl*, contendo uma ORF para proteína ribossomal S3. As regiões de localização dos três íntrons presentes em *C. lindemuthianum*, mas ausentes nos genomas mitocondriais de *Colletotrichum* spp. são altamente divergentes em relação a conservação das bases, em contraste as regiões gênicas que são altamente conservadas, exceto nas regiões das ncORFs (Figura 4).



**Figura 4.** Comparação do genoma mitocondrial de *Colletotrichum lindemuthianum* 89 A<sub>2</sub> 2-3 e sete espécies de *Colletotrichum* por meio de BLASTn. Cada anel representa um genoma mitocondrial. A variação de cores de cada anel representa identidade de 70% a 100% em relação ao genoma mitocondrial de *C. lindemuthianum*. Os genes pertencem ao genoma mitocondrial de *C. lindemuthianum*. i89: isolado 89 A<sub>2</sub> 2-3; i83: isolado 83.501.

Foram também encontradas cinco ncORFs nos genomas mitocondriais nos isolados 89 A<sub>2</sub> 2-3 e 83.501 de *C. lindemuthianum*. Para a validação dessas cinco ncORFs, foi utilizada a ferramenta BLASTn contra um transcriptoma montado obtido de micélio de *C. lindemuthianum* crescido

em meio de cultura (dados não publicados) e todas as ncORFs foram encontradas, mostrando que todos os cinco genes são transcritos. Foi também utilizada a ferramenta BLASTn contra um banco de dados local construído com *reads* dos transcriptoma de *C. graminicola* (GEO Acession: GSM852201; GSM852203; GSM852204), sendo também observado que todas as ncORFs encontradas no genoma mitocondrial de *C. graminicola* são transcritas (Arquivo adicional: Tabela S4). Por isso, foi presumido que as ncORFs preditas nas outras espécies de *Colletotrichum* também podem ser transcritas e por esta razão, estas foram consideradas em nossas análises. O total de ncORFs foi altamente variável em cada espécie, sendo que *C. fioriniae* possui apenas uma ncORF com um tamanho de 723 pb que corresponde a apenas 2,41% de seu genoma mitocondrial, enquanto que o tamanho total das sequências de ncORFs em *C. graminicola* é de 10.092 pb que equivale a 25,4% de seu genoma mitocondrial (Arquivo adicional: Tabela S2). Em *C. lupini* e *C. siamense* foram encontradas até 10 ncORFs cada (Arquivo adicional: Tabelas S1 e S3). A *orf312* dos genomas mitocondriais dos isolados 89 A<sub>2</sub> 2-3 e 83.501 não foi encontrada no genoma mitocondrial do isolado de *C. lindemuthianum* de Gutiérrez et al. (2014). Foi realizado um alinhamento da sequência da *orf312* com a sequência do genoma mitocondrial de *C. lindemuthianum* de Gutiérrez et al. (2014) e foi observada ausência de 517 pb entre as posições 3.6956 e 3.6957, destas, 445 pb pertenciam a sequência do *orf312* que possui 939 pb (Figura 4).

A maioria das sequências de aminoácidos codificadas em ncORFs não possuem similaridade significativa com qualquer proteína, porém *orf747* de *C. graminicola* codifica uma proteína com domínio de DNA polimerase tipo B (pfam03175/*e-value* 2,3E-89) e já anotada como DNA polimerase no genoma mitocondrial de *C. graminicola* depositado no *GenBank*. No genoma mitocondrial de *C. lupini*, os genes *orf204* e *orf300* codificam proteínas putativas contendo o domínio endonuclease GIY-YIG (cl27895/*e-value* 1.02E-23 e 8.92E-16, respectivamente), porém nenhum íntron adjacente às ORFs foi encontrado.

## **Sobreposições de genes e regiões intergênicas**

Ao todo foram encontrados 455 pb como sobreposições em 10 genes do genoma mitocondrial dos isolados de *C. lindemuthianum*: duas sobreposições de uma base entre os genes *nad2* e *nad3* e entre *nad4l* e *nad5*; 10 pb entre os genes *nad6* e *trnV* e 97 pb entre os genes *orf367* e *cox2*. As outras 346 pb são referentes a sobreposição do gene *orf241* com *nad1*, embora, *orf241* seja transcrito no sentido contrário ao de *nad1*. Igual ao observado em *C. lindemuthianum*, exceto em *C. graminicola*, as duas sobreposições de 1 pb que ocorrem entre *nad2* e *nad3* e entre *nad4l* e *nad5* e de 10 pb entre *nad6* e *trnV* são encontradas nos genomas mitocondriais das outras seis espécies. Em *C. graminicola*, a sobreposição entre *nad6* e *trnV* foi de 44 pb e entre *nad4l* e *nad5* foi de 1 pb, semelhante as demais espécies. Além disso, ocorre uma sobreposição em 69 pb entre o gene *orf309* e *cox3*, onde *orf309* tem o sentido da transcrição inverso a *cox3*. Em *C. siamense*, além das sobreposições que é comum para todos *Colletotrichum* spp. ocorre uma de 14 pb entre *orf313* e *orf106*.

Os genomas mitocondriais de *C. lindemuthianum* dos isolados 89 A<sub>2</sub> 2-3 e 83.501 tem um total de 7.414 pb de sequências intergênicas em 52 localizações, variando de 1 a 1.053 pb, sendo a maior localizada entre os genes *orf213* e *trnK*. Os tamanhos das sequências intergênicas foram similares entre os genomas mitocondrias de seis espécies de *Colletotrichum*, com 7.035 pb em *C. tamarilloi* até 8.749 pb em *C. lupini* (Figura 3 e Arquivo adicional: Tabela S2). *Colletotrichum siamense* e *C. fructicola*, além de possuírem os maiores genomas mitocondrias entre as espécies de *Colletotrichum*, possuem regiões intergênicas maiores que as demais espécies de *Colletotrichum*, com 13.792 pb e 15.302 pb, respectivamente.

## **Composição de nucleotídeos e uso de códons preferenciais**

A composição de nucleotídeos de genomas mitocondriais de fungos é geralmente determinada pelos nucleotídeos G e C. Nos genomas mitocondriais dos isolados 89 A<sub>2</sub> 2-3 e 83.501 a composição de G+C foi de

30,8% similar às outras espécies de *Colletotrichum* que variou de 29,9% em *C. graminicola* e *C. lupini* até 34,04% em *C. fructicola*. No genoma mitocondrial de *C. lindemuthianum*, o conteúdo G+C é de 29,1% em todos os genes que codificam proteínas, 33,4% nos genes de RNAs ribossomais e 38,1% nos genes de tRNAs (Arquivo adicional: Tabela S1). Um total de 6.953 e 6.951 de aminoácidos são codificados pelos genes nos genomas mitocôndriais dos isolados 89 A<sub>2</sub> 2-3 e 83.501, respectivamente. O baixo conteúdo de G+C é reflexo do uso preferencial de codons, principalmente na terceira posição: códons contendo A ou U (A=41,27%, U=43,07%) são fortemente preferidos a C ou G (C=6,53%, G=9,13 %) na terceira posição. Por exemplo, o uso do codon AAA (87,6%) para lisina é fortemente preferido em relação a AAG (12,4%). Os códons mais utilizados consistem somente de A e U, e foram UUA (9,46%), AUA (5,50%), AAU (5,07%), UUU (4,80%), AAA (4,14%), UAU (3,67%) e AUU (3,47%) (Tabela 2). Os aminoácidos mais frequentes nas proteínas mitocondriais de *C. lindemuthianum* são leucina (12,36%), isoleucina (9,81%), serina (9,42%) e fenilalanina (7,10%) (Figura 5).

**Tabela 2.** Uso de códons para os genes que codificam proteínas do genoma mitocondrial de *C. lindemuthianum*.

Códon	N*	%	Fração	Códon	N	%	Fração
GCG(Ala)	24 <sup>a</sup>	0,34 <sup>a</sup>	0.06	CCG(Pro)	11 <sup>b</sup>	0,16 <sup>b</sup>	0.04
GCA(Ala)	110	1,58	0.29	CCA(Pro)	64 <sup>c</sup>	0,92 <sup>c</sup>	0.24 <sup>c</sup>
GCT(Ala)	204	2,92	0.54	CCT(Pro)	169	2,42	0.63
GCC(Ala)	43	0,62	0.11	CCC(Pro)	25	0,36	0.09
TGT(Cys)	45	0,65	0.94	CAG(Gln)	16	0,23	0.09
TGC(Cys)	3	0,04	0.06	CAA(Gln)	155	2,22	0.91
GAT(Asp)	195	2,80	0.84	AGG(Arg)	5	0,07	0.03
GAC(Asp)	38	0,54	0.16	AGA(Arg)	131	1,88	0.72
GAG(Glu)	46	0,66	0.18	CGG(Arg)	4	0,06	0.02

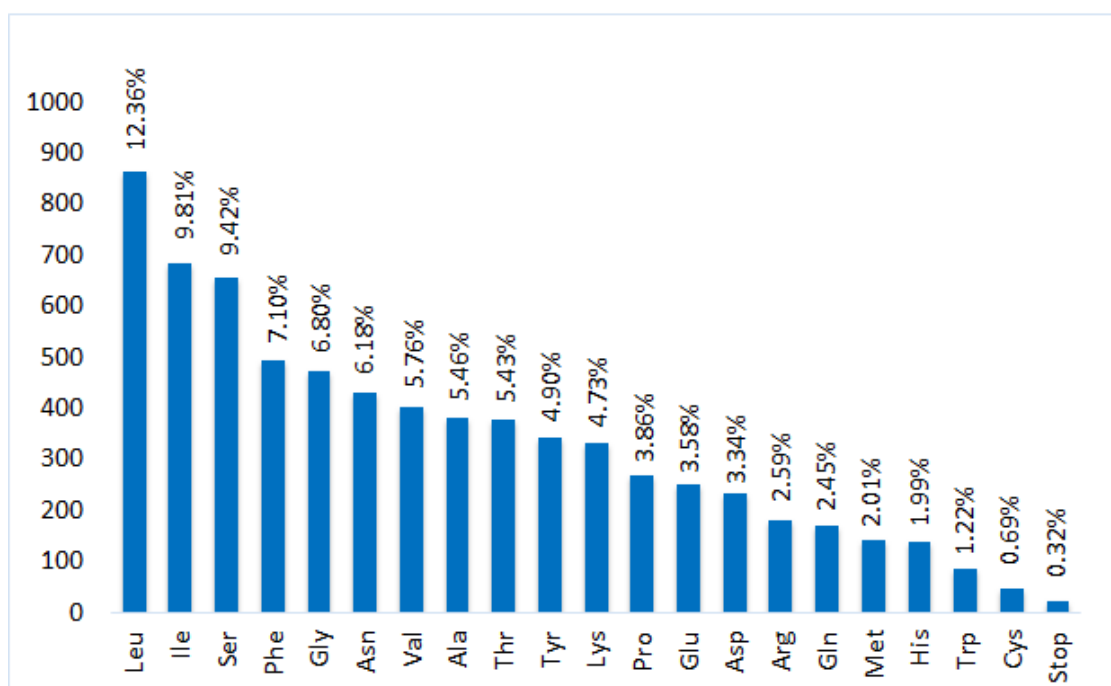
GAA(Glu)	204	2,92	0.82	CGA(Arg)	7	0,10	0.04
TTT(Phe)	335	4,80	0.68	CGT(Arg)	32	0,46	0.18
TTC(Phe)	160	2,29	0.32	CGC(Arg)	2	0,03	0.01
GGG(Gly)	36	0,52	0.08	AGT(Ser)	215	3,08	0.33
GGA(Gly)	199	2,85	0.42	AGC(Ser)	55	0,79	0.08
GGT(Gly)	227	3,25	0.48	TCG(Ser)	29	0,42	0.04
GGC(Gly)	12	0,17	0.03	TCA(Ser)	153	2,19	0.23
CAT(His)	109	1,56	0.78	TCT(Ser)	186	2,67	0.28
CAC(His)	30	0,43	0.22	TCC(Ser)	19	0,27	0.03
ATA(Ile)	383	5,49	0.56	ACG(Thr)	7	0,10	0.02
ATT(Ile)	242	3,47	0.35	ACA(Thr)	166	2,38	0.44
ATC(Ile)	59	0,85	0.09	ACT(Thr)	193	2,77	0.51
AAG(Lys)	41	0,59	0.12	ACC(Thr)	13	0,19	0.03
AAA(Lys)	289	4,14	0.88	GTG(Val)	34	0,49	0.08
TTG(Leu)	45	0,65	0.05	GTA(Val)	188	2,70	0.47
TTA(Leu)	660	9,46	0.77	GTT(Val)	168	2,41	0.42
CTG(Leu)	11	0,16	0.01	GTC(Val)	12	0,17	0.03
CTA(Leu)	69	0,99	0.08	TGG(Trp)	4	0,06	0.05
CTT(Leu)	74	1,06	0.09	TGA(Trp)	81	1,16	0.95
CTC(Leu)	3	0,04	0.00	TAT(Tyr)	256	3,67	0.75
ATG(Met)	140	2,01	1.00	TAC(Tyr)	86	1,23	0.25
AAT(Asn)	354	5,08	0.82	TAG(End)	2	0,03	0.09
AAC(Asn)	77	1,10	0.18	TAA(End)	20	0,29	0.91

<sup>a</sup> Isolado 83.501: N = 24; % = 0,32.

<sup>b</sup> Isolado 83.501: N = 12; % = 0,17.

<sup>c</sup> Isolado 83.501: N = 63; % = 0,90, Fraction = 0,23.

\* Total de cada códon.



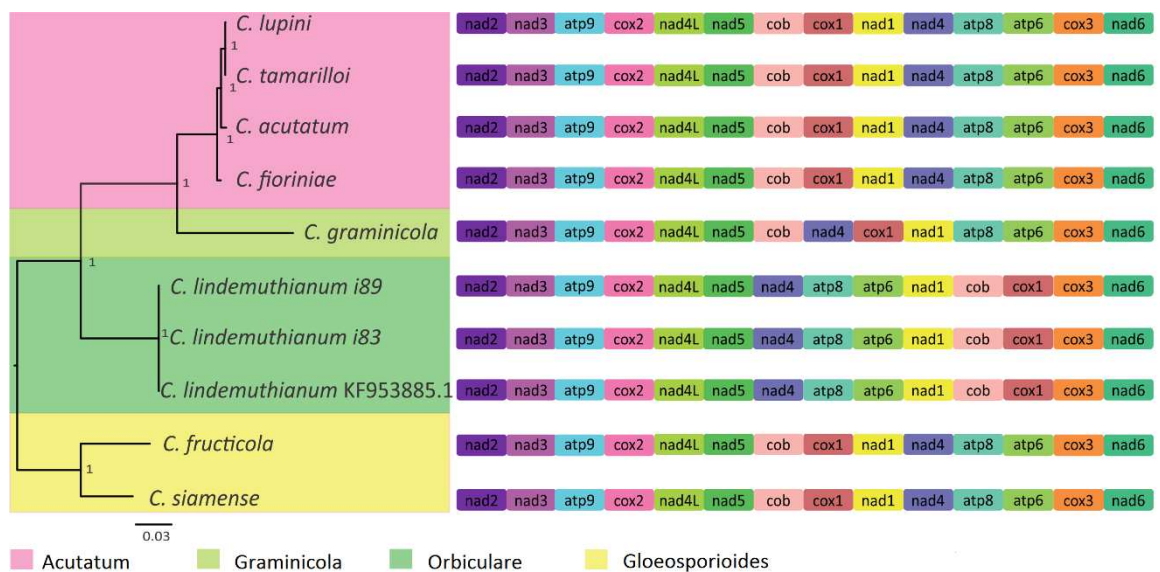
**Figura 5.** Aminoácidos mais frequentes codificados pelos genes mitocondriais de *C. lindemuthianum*.

O tamanho total das regiões codificadoras é de 27.339 pb (89 A<sub>2</sub> 2-3) e 27.333 pb (83.501), que correspondem a 73,0% de todo o genoma mitocondrial de *C. lindemuthianum*. Com exceção ao gene da endonuclease GIY-YIG, todos os genes que codificam proteínas iniciam a tradução usando o códon ATG (Tabela 1). O gene da endonuclease GIY possivelmente utiliza um codon de iniciação não usual (GAC), porém, existe um codon alternativo ATG à 6 pb no final do primeiro exon do gene *cox1*, *upstream* ao gene da endonuclease GIY. Para as seis espécies de *Colletotrichum*, além dos isolados de *C. lindemuthianum* aqui analisados, na maioria dos genes que codificam proteínas, o ATG é o códon de iniciação da tradução. A exceção foi em *C. siamense* e *C. fructicola*, onde o códon de iniciação ATG não foi encontrado em *cox3* (TTG) e nos genes de todas as endonucleases LAGLIDADG e GIY-YIG. No genoma mitocondrial dos isolados de *C. lindemuthianum*, 20 dos 22 genes que codificam proteínas, possuem códon de parada TAA, enquanto *cox1* e *nad 6* usam o TAG. Nas outras espécies de *Colletotrichum*, o códon de parada TAG foi também encontrado em *C. graminicola* (*nad6*), *C. lupini* (*orf367* e *orf270*), *C. siamense* (*nad6*, *cox3*,

*orf106*, *orf116*, *orf137* e *orf341*) e *C. fructicola* (*nad6*, *cox3*, *orf336*, *orf101* e *orf99*). Em *C. fiorinae*, *C. acutatum* e *C. tamarilloi* o códon de parada TAG foi encontrado somente em *nad6*.

### Análise filogenética das espécies de *Colletotrichum*

Para determinar as relações filogenéticas entre as oito espécies de *Colletotrichum*, foi utilizado o método de inferência Bayesiana com um total de 13.826 nucleotídeos pertencentes a 14 genes mitocondriais conservados (Figura 6). O resultado revelou quatro grupos que correspondem aos clados gloeosporioides, orbiculare, graminicola e acutatum, todos com fortes suportes de probabilidade posterior. As espécies do clado gloeosporioides foram mais relacionadas ao clado orbiculare que inclui *C. lindemuthianum*.



**Figura 6.** Árvore filogenética Bayesiana baseada em 14 genes *core* concatenados de quatro clados de espécies de *Colletotrichum*. A ordem dos 14 genes *core* foram listadas pela ordem sintênica começando pelo gene *nad2*, pois é o primeiro gene que codifica proteína nas sequências dos genomas mitocondriais dos isolados de *C. lindemuthianum*. i89: isolado 89 A<sub>2</sub> 2-3; i83: isolado 83.501.

## DISCUSSÃO



O genoma mitocondrial de *C. lindemuthianum* tem um tamanho similar a maioria das espécies de *Colletotrichum* que possuem genomas mitocôndriais disponibilizados até o momento, sendo o maior o de *C. siamense* com 58.666 pb e o menor de *C. fiorinae* com 30.020 pb. O tamanho do genoma mitocondrial de *C. lindemuthianum* também é equivalente a maioria dos outros fungos filamentosos que varia de 20.000 pb a 65.000 pb (Aguileta et al. 2014). Porém, também existe uma grande variação no genoma mitocondrial de fungos, onde o menor genoma possui 12.000 pb, do micoparásita *R. allomyces* (James et al. 2013) e o maior é de 235.000 pb, do basidiomicota *R. solani*, patógeno da batata (Losada et al. 2014). Tem sido reportado que a variação do tamanho de genomas mitocôndriais pode ser causada pelo tamanho de regiões intergênicas ou presença de íntrons de vários tamanhos (Aguileta et al. 2014; Burger et al. 2003). Neste trabalho, foi observado que a variação nos genomas mitocôndrias das espécies de *Colletotrichum*, exceto em *C. fructicola* e *C. siamense*, não está relacionada com as sequências intergênicas, e que para *C. lindemuthianum*, *C. fructicola* e *C. siamense*, as sequências intrônicas contribuem em parte para a expansão em comparação às outras espécies de *Colletotrichum* de genomas mitocôndrias menores. Além disso, foi observado uma grande variação na quantidade de ncORFs em cada espécie, o que também contribui para as diferenças nos tamanhos dos genomas mitocôndriais de *Colletotrichum*. Assim, a variação existente nos tamanhos dos genomas mitocôndrias de *Colletotrichum* é devido principalmente, mas não exclusivamente, à presença ou ausência de ncORFs e presença de sequências intrônicas, e não devido às sequências intergênicas como tem sido demonstrado para outros fungos (Aguileta et al. 2014; Burger et al. 2003). Entretanto, sequências intergênicas parecem ser importantes na expansão genômica de *C. fructicola* e *C. siamense* que possuem essas regiões quase duas vezes maiores que a maioria dos genomas mitocôndrias das outras espécies de *Colletotrichum*. Em um trabalho prévio com essas duas espécies realizado por Liang et al. (2017), as regiões intergênicas foram de 8,8% (*C. fructicola*) e 11,4% (*C. siamense*) maiores que o encontrado em nosso trabalho. Contudo, essas diferenças são em razão de Liang et al. (2017) não terem considerado as ncORFs

como parte da região codificante. Transcritos de uma ncORF nomeada de *orf414* (função desconhecida) foram detectados por meio de real-time RT-PCR no fungo *Verticillium nonalfalfae* (Jelen et al., 2016). Além disso, foram comparadas as sequências das cinco ncORFs encontradas em *C. lindemuthianum* de nossos isolados com um transcriptoma de micélios de *C. lindemuthianum* crescido em meio de cultura (dados não publicados). Também foi comparada as ncORFs de *C. graminicola* com as *reads* de seu transcriptoma. Foi observado transcritos de todas as cinco ncORFs de *C. lindemuthianum* e de todas as ncORFs de *C. graminicola*, indicando que estas também podem ser transcritas nas demais espécies de *Colletotrichum*.

As ncORFs têm sido frequentemente reportadas em genomas mitocondriais de fungos (Jelen et al. 2015; Li et al. 2015; Salavirta et al. 2014; Torriani et al. 2014; Zhang et al. 2017), porém sua origem e função ainda não foram elucidadas. Além disso, não é comum encontrar ncORFs compartilhadas entre espécies distantes filogeneticamente. Neste trabalho, foi encontrada uma ncORF em *C. fioriniae* e 10 em *C. lupini*, mas, somente a ncORFs *orf240* foi compartilhada entre *C. fioriniae*, *C. acutatum*, *C. lupini* e *C. tamarilloi*, que são espécies pertencentes ao clado *acutatum*. A *orf347* somente foi encontrada em *C. lupini* e *C. tamarilloi*. Nas duas espécies do clado *gloeosporiodes*, somente *orf313* é compartilhada entre elas. Torriani et al. (2014) propõe que a origem de ncORFs em genomas mitocondriais poderia ser por mutação de ponto e por pequenos *indels* em um gene mitocondrial pré-existente que levaria à criação de uma nova ORF. Também tem sido discutido a hipótese que ncORFs tenham sido transferidas do genoma nuclear para o genoma mitocondrial. Quanto a função das ncORFs, seria auxiliar as funções do *core* mitocondrial ou codificar novas funções que não estão diretamente conectadas com a fosforilação oxidativa. No fungo endofítico *Pestalotiopsis fici* foram identificadas nove ncORFs e três delas tiveram homologia com transcriptase reversa, aminotransferase e N-acetiltransferase (Zhang et al. 2017). Além disso, os autores relataram que as sequências das ncORFs com homologia a aminotransferase e N-acetiltransferase foram encontradas em diversos genomas mitocondriais de outros fungos, e destacam que esses dois genes são comumente encontrados em genomas nucleares de fungos. Além disso, essas duas

ncORFs estão ausentes no genoma de uma linhagem *P. fici* e foram identificadas no genoma mitocondrial de outra linhagem (Zhang et al. 2017). Isto corrobora com uma das hipóteses da origem das ncORFs, que neste caso teriam sido transferidas do genoma nuclear para o genoma mitocondrial. Em *C. graminicola*, *orf747* codifica uma sequência de aminoácidos com domínio de DNA polimerase tipo B. Genes de DNA polimerase tipo B tem sido reportado em plasmídeos de fungos, genomas mitocondriais, DNA de cloroplasto e em genomas virais (Ferandon et al. 2008; Pasharawipasa et al. 2009; Salavirta et al. 2014). Isso leva a indícios de que este gene seria resultado de uma integração de um plasmídeo no genoma mitocondrial durante a evolução, como sugerido por Ferandon et al. (2008) e que isso seria uma possível origem das ncORF.

A maioria dos genes que codificam proteínas mitocondriais envolvidos na fosforilação oxidativa e síntese de ATP é altamente conservado em genomas mitocôndrias de fungos (Aguileta et al. 2014; Jelen et al. 2015; Robicheau et al. 2016; Zhang et al. 2017). Este também foi o caso dos genomas mitocôndriais de *Colletotrichum* spp. analisados neste trabalho, que codificam 14 genes conservados envolvidos em fosforilação oxidativa e síntese de ATP, além dos genes das subunidades ribossomais maior e menor, *rnl* e *rns*, respectivamente, e o gene da proteína ribossomal S3. Todos esses genes conservados são codificados pela mesma fita de DNA, enquanto que alguns das ncORFs foram encontradas na fita negativa. Genes encontrados na fita negativa têm sido relatados em genomas mitocondriais de fungos, incluindo os genes conservados (Nadimi et al. 2016; Salavirta et al. 2014; Stone et al. 2014).

Em *C. lindemuthianum* foi encontrado um total de quatro íntrons, todos do grupo I, e três deles presentes em genes que codificam proteínas, entre esses, um intron codifica uma sequência de aminoácidos com domínio LAGLIDADG e outro codifica uma sequência de aminoácidos contendo o domínio GIY-YIG. Neste trabalho foi observado que íntrons que codificam ORFs para endonucleases LAGLIDADG e GIY-YIG são ausentes nas espécies do clado acutatum, porém em *C. lupini*, duas ncORFs codificam sequências de aminoácidos contendo o domínio GIY-YIG. Endonucleases *homing* codificadas fora de íntrons foram encontradas também no fungo *R.*

*solani*, onde 11 de 33 endonucleases *homing* foram encontradas em regiões intergênicas (Losada et al. 2014). Um íntron do grupo IA foi encontrado no gene *rnl* nos genomas mitocondriais de todas as espécies de *Colletotrichum* aqui estudadas. Esse resultado difere do relatado para os genomas mitocondriais das espécies do clado gloeosporioides (Liang et al. 2017), onde não foi descrito qualquer intron presente em *rnl*. O gene *rps3* é encontrado dentro deste intron, que por sua vez é encontrado dentro do gene *rnl* e isto é reportado como uma característica comum em genomas mitocondriais da classe sordariomycetes, que inclui o gênero *Colletotrichum* (Jelen et al. 2016; Sethuraman et al. 2009). O produto do gene *rps3* em humanos e em *Drosophila melanogaster* tem sido relatado com múltiplas atividades além da síntese protéica, como sendo envolvido no reparo de DNA e com atividade potencial de endonuclease (Yadavillia et al. 2009; Jang et al. 2004). Com base nesses relatos, Sethuraman et al. (2009) especularam que é possível que o ancestral do gene *rps3* codificado por um íntron tenha existido como *free-ORF*, mas seu produto tinha atividade de endonuclease e teria atuado semelhante a uma endonuclease *homing* e que por meio de uma recombinação inseriu uma cópia de si mesmo dentro de um íntron do grupo I. Além disso, o fato do íntron que codifica *rps3* estar inserido dentro do gene *rnl* em genomas mitocondriais de muitos fungos, pode ser uma estratégia para assegurar a transcrição do gene *rps3* juntamente com a do gene *rnl* e isso levaria a manutenção deste íntron (Sethuraman et al. 2009).

Neste trabalho foram encontrados 25 a 29 genes de tRNA em *Colletotrichum* spp., que é similar ao relatados em outros fungos (Aguileta et al. 2014; Torriani et al. 2014), e estes correspondem aos 20 aminoácidos. Em *C. lindemuthianum*, as ORFs mitocondriais apresentam forte preferência para códons contendo U ou A, principalmente na terceira posição. A tendência do uso de U ou A tem sido atribuída em outras espécies ao pareamento *wobble* (Deng et al. 2016; Stone et al. 2010; Wendy e Sande 2012). Embora haja 61 códons de aminoácidos, a maioria (81,6%) das espécies de bactérias e organismos eucariotos possuem menos de 45 tipos de tRNA (Chan e Lowe 2009), assim um tipo de tRNA precisa parear com mais de um tipo de códon (Novoa e Pouplana, 2012). Isto acontece devido a

presença de bases *wobble* na terceira posição do códon ou na primeira posição do anticódon do tRNA (Novoa e Pouplana 2012). Normalmente o local da base *wobble* do anticódon deve ter um G para um códon NNY (N = qualquer base; Y = C ou U), porque G pode parear com C ou U, e deve ter um U para parear com NNR (R = A ou G) para parear com A ou G (Carullo e Xia 2008; Xia 2005). Outra hipótese para o alto conteúdo de A e T é devida ao fato das células terem uma alta disponibilidade de ATP e relativamente baixa disponibilidade dos outros três tipos de NTPs. Assim, a eficiência da transcrição seria aumentada por maximizar o uso de A na terceira posição do códon (Sun et al. 2009; Xia 2005). Normalmente existe uma relação entre o uso de códon e a velocidade com que os ribossomos decodificam esses códons com a abundância de tRNA no núcleo da célula (Novoa e Pouplana 2012; Sun et al. 2009; Wei et al. 2014). Embora os códons mais abundantes dos genes que codificam proteínas mitocondriais de *C. lindemuthianum* tenham sido para os aminoácidos leucina, isoleucina e serina, os genes de tRNA mais encontrados foram para lisina e metionina, com quatro e três genes, respectivamente. Em alguns organismos o genoma mitocondrial não tem genes para todos os aminoácidos, e para compensar isso, a mitocôndria importa diversos tRNAs codificados pelo núcleo (Schneider 2011). Se em *C. lindemuthianum* também existir uma relação entre o uso de códons e a abundância de tRNA, para traduzir os códons mais frequentes nos genes mitocondriais, a mitocôndria dessa espécie deve importar os tRNA codificados pelo núcleo, correspondentes a esses aminoácidos.

A análise filogenética revelou quatro grupos que correspondem aos quatros clados de espécies de *Colletotrichum* utilizadas neste trabalho, estando de acordo com resultados filogenéticos prévios realizados com genes nucleares (Cannon et al. 2012; Gan et al. 2016). Nesses trabalhos, as análises filogenéticas revelaram que o clado gloeosporioides é mais relacionado ao clado acutatum. No entanto, os dados deste trabalho demonstram que o clado gloeosporioides está mais relacionado ao clado orbiculare, isso possivelmente se deve ao fato dessas espécies compartilharem íntrons exclusivos a estes clados. Como discutido anteriormente, o único intron encontrado nas espécies do clado acutatum foi no gene *rnl*, e no clado gramínicola existe apenas um intron a mais presente

no gene *nad5*, além daquele presente no gene *rnl*. Porém o gene *rnl* não foi utilizado em nossas análises de filogenia. *Colletotrichum lindemuthianum* que pertence ao clado orbiculare, possui íntrons nos genes *rnl*, *cob* e *cox1*. *C. fructicola* e *C. siamense*, pertencentes ao clado gloeosporioides, possuem íntrons em 4 e 5 genes, respectivamente. Nossos dados de filogenia demonstram que os genes mitocondriais conservados são mais uma alternativa para determinação de relações filogenéticas entre espécies de *Colletotrichum*, assim como têm sido utilizados em outras espécies de fungos e outros organismos (Aguileta et al. 2014; Robicheau et al. 2016; Shen et al. 2015; Yang et al. 2015).

Os genomas mitocondriais de *Colletotrichum* spp. analisados neste trabalho possuem sequências de DNA altamente conservadas principalmente nas regiões que codificam os genes considerados padrões de genomas mitocondriais de fungos. Por outro lado, existe uma grande variação no número de ncORFs e íntrons, que juntos contribuem para as diferenças no tamanho de cada genoma mitocondrial. Além disso, a presença de transcritos de ncORF em *C. lindemuthianum* (dados não mostrados) e de *C. graminicola* foi suportada por análise de transcriptoma. Provavelmente essas ncORFs possuem papéis importantes nessas espécies. Futuros trabalhos podem demonstrar os papéis desses genes. Além disso, os resultados da análise filogenética demonstram que os genes mitocondriais podem ser utilizados como mais uma alternativa para inferências filogenéticas em espécies de *Colletotrichum*. Os resultados deste trabalho contribuem para os estudos de genética e biologia de *Colletotrichum* spp. que são de grande importância em âmbito econômico e científico.

## REFERÊNCIAS

- Aguileta G, De Vienne DM, Ross ON, Hood ME, Giraud T, Petit E, Gabaldón T (2014) High variability of mitochondrial gene order among fungi. *Genome Biol Evol* 6:451–465. doi: 10.1093/gbe/evu028
- Al-Reedy RM, Malireddy R, Dillman CB, Kennell JC (2012) Comparative analysis of *Fusarium mitochondrial* genomes reveals a highly variable region that encodes an exceptionally large open reading frame. *Fungal Genet Biol* 49:2–14. doi: 10.1016/j.fgb.2011.11.008
- Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402
- Alzate-Marin AL, Gonçalves De Barros E, Moreira MA (1999) Co-evolution model of *Colletotrichum lindemuthianum* (Melanconiaceae, Melanconiales) races that occur in some Brazilian regions. *Genet Mol Biol* 22:115–118. <http://dx.doi.org/10.1590/S1415-47571999000100022>
- Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Mol Ecol* 13:729–744. doi: 10.1046/j.1365-294X.2003.02063.x
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritsch G, Pütz J, Middendorf M, Stadler PF (2013) MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol* 69:313–319. doi: 10.1016/j.ympev.2012.08.023

- Burger G, Gray MW, Lang BF (2003) Mitochondrial genomes: Anything goes. *Trends Genet* 19:709–716. doi: 10.1016/j.tig.2003.10.012
- Cannon PF, Damm U, Johnston PR, Weir BS (2012) *Colletotrichum* - current status and future directions. *Stud Mycol* 73:181–213. doi: 10.3114/sim0014
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. doi: 10.1093/bioinformatics/btp348
- Carullo M, Xia X (2008) An extensive study of mutation and selection on the wobble nucleotide in tRNA anticodons in fungal mitochondrial genomes. *J Mol Evol* 66:484–493. doi: 10.1007/s00239-008-9102-8
- Chan PP, Lowe TM (2009) GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* 37:93–97. doi: 10.1093/nar/gkn787
- Crouch JA., O’Connell R, Gan P, Buiate E, Torres MF, Beirn L, Shirasu K, Vaillancourt L. (2014) The genomics of *Colletotrichum*. In: *Genomics of plant-associated fungi: monocot pathogens*. Berlin: Springer Verlag. 69-102.
- Deng Y, Zhang Q, Ming R, Lin L, Lin X, Lin Y, Li X, Xie B, Wen Z (2016) Analysis of the mitochondrial genome in *Hypomyces aurantius* reveals a novel twintron complex in fungi. *Int J Mol Sci*. doi: 10.3390/ijms17071049
- dos Santos LV, de Queiroz MV, Santana MF, Soares MA, de Barros EG, de Araújo EF, Langin T (2012) Development of new molecular markers for the *Colletotrichum* genus using RetroCl1 sequences. *World J Microbiol Biotechnol* 28:1087–1095. doi: 10.1007/s11274-011-0909-x
- Ferandon C, Chatel SEK, Castandet B, Castroviejo M, Barroso G (2008) The *Agrocybe aegerita* mitochondrial genome contains two inverted repeats of the nad4 gene arisen by duplication on both sides of a linear plasmid



integration site. *Fungal Genet Biol* 45:292–301. doi: 10.1016/j.fgb.2007.10.012

Fernández MT, Fernández M, Casares A, Rodríguez R, Fueyo M (2000) Bean germplasm evaluation for anthracnose resistance and characterization of agronomic traits: A new physiological strain of *Colletotrichum lindemuthianum* infecting *Phaseolus vulgaris* L. in Spain. *Euphytica* 114:143–149. doi: 10.1023/A:1003937812700

Ferreira JJ, Campa A, Kelly JD (2013) Organization of Genes Conferring Resistance to Anthracnose in Common Bean. *Transl Genomics Crop Breed I*:151–176. doi: 10.1002/9781118728475

Gan P, Narusaka M, Kumakura N, Tsushima A, Takano Y, Narusaka Y, Shirasu K (2016) Genus-wide comparative genome analyses of *Colletotrichum* species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. *Genome Biol Evol* 8(5):1467–1481. doi: 10.1093/gbe/eww089

Ghikas D V., Kouvelis VN, Typas MA (2006) The complete mitochondrial genome of the entomopathogenic fungus *Metarhizium anisopliae* var. *anisopliae*: Gene order and trn gene clusters reveal a common evolutionary course for all Sordariomycetes, while intergenic regions show variation. *Arch Microbiol* 185:393–401. doi: 10.1007/s00203-006-0104-x

González AM, Yuste-Lisbona FJ, Rodiño AP, De Ron AM, Capel C, García-Alcázar M, Lozano R, Santalla M (2015) Uncovering the genetic architecture of *Colletotrichum lindemuthianum* resistance through QTL mapping and epistatic interaction analysis in common bean. *Front Plant Sci* 6:13. doi: 10.3389/fpls.2015.00141

Gutiérrez P, Alzate J, Yepes MS, Marín M (2014) Complete mitochondrial genome sequence of the common bean anthracnose pathogen *Colletotrichum lindemuthianum*. *Mitochondrial DNA* 1736:1–2. doi: 10.3109/19401736.2013.878912

- Hu X, Zhang YJ, Xiao GH, Zheng P, Xia YL, Zhang XY, St Leger RJ, Liu XZ, Wang CS (2013) Genome survey uncovers the secrets of sex and lifestyle in caterpillar fungus. *Chinese Sci Bull* 58:2846–2854. doi: 10.1007/s11434-013-5929-5
- Pastor-Corrales MA, Tu JC (1989) Anthracnose. In: Schwartz HF, Pastor-Corrales MA (eds) *Bean production problems in the tropics*, 2nd edn. Centro Internacional de Agricultura Tropical (CIAT), Colombia, pp 77–104.
- James TY, Pelin A, Bonen L, Ahrendt S, Sain D, Corradi N, Stajich JE (2013) Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia. *Curr Biol* 23:1548–1553. doi: 10.1016/j.cub.2013.06.057
- Jang CY, Lee JY, Kim J (2004) RpS3, a DNA repair endonuclease and ribosomal protein, is involved in apoptosis. *FEBS Lett* 560:81–85. doi: 10.1016/S0014-5793(04)00074-2
- Jelen V, De Jonge R, Van De Peer Y, Javornik B, Jakše J (2016) Complete mitochondrial genome of the Verticillium-wilt causing plant pathogen *Verticillium nonalfalfae*. *PLoS One* 11:1–18. doi: 10.1371/journal.pone.0148525
- Kang D, Hamasaki N (2002) Maintenance of mitochondrial DNA integrity: Repair and degradation. *Curr Genet* 41:311–322. doi: 10.1007/s00294-002-0312-0
- Kim J-O, Choi K-Y, Han J-H, Choi I-Y, Lee Y-H, Kim KS (2015) The complete mitochondrial genome sequence of the ascomycete plant pathogen *Colletotrichum acutatum*. *Mitochondrial DNA* 1736:1–2. doi: 10.3109/19401736.2015.1101556
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:msw054. doi: 10.1093/molbev/msw054

- Lang BF, Laforest MJ, Burger G (2007) Mitochondrial introns: a critical view. *Trends Genet* 23:119–125. doi: 10.1016/j.tig.2007.01.006
- Laslett D, Canbäck B (2008) ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* 24:172–175. doi: 10.1093/bioinformatics/btm573
- Li Y, Hu X-D, Yang R-H, Hsiang T, Wang K, Liang D-Q, Liang F, Cao D-M, Zhou F, Wen G, Yao Y-J (2015) Complete mitochondrial genome of the medicinal fungus *Ophiocordyceps sinensis*. *Sci Rep* 5:13892. doi: 10.1038/srep13892
- Liang X, Tian X, Liu W, Wei T, Wang W, Dong Q, Wang B, Meng Y, Zhang R, Gleason ML, Sun G (2017) Comparative analysis of the mitochondrial genomes of *Colletotrichum gloeosporioides* sensu lato: insights into the evolution of a fungal species complex interacting with diverse plants. *BMC Genomics* 18:171. doi: 10.1186/s12864-016-3480-x
- Losada L, Pakala SB, Fedorova ND, Joardar V, Shabalina SA, Hostetler J, Pakala SM, Zafar N, Thomas E, Rodriguez-Carres M, Dean R, Vilgalys R, Nierman WC, Cubeta MA (2014) Mobile elements and mitochondrial genome expansion in the soil fungus and potato pathogen *Rhizoctonia solani* AG-3. *FEMS Microbiol Lett* 352:165–173. doi: 10.1111/1574-6968.12387
- Mota SF, Barcelos QL, Dias MA, Souza EA (2016) Variability of *Colletotrichum* spp in common bean. *Genet Mol Res*. doi: 10.4238/gmr.15027176
- Nadimi M, Stefani FOP, Hijri M (2016) The large (134.9 kb) mitochondrial genome of the glomeromycete *Funneliformis mosseae*. *Mycorrhiza* 26:747–755. doi: 10.1007/s00572-016-0710-7
- Novoa EM, Ribas de Pouplana L (2012) Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends Genet* 28:574–581. doi: 10.1016/j.tig.2012.07.006

O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, Damm U, Buiate E a, Epstein L, Alkan N, Altmüller J, Alvarado-Balderrama L, Bauser C a, Becker C, Birren BW, Chen Z, Choi J, Crouch JA, Duvick JP, Farman M a, Gan P, Heiman D, Henrissat B, Howard RJ, Kabbage M, Koch C, Kracher B, Kubo Y, Law AD, Lebrun M-H, Lee Y-H, Miyara I, Moore N, Neumann U, Nordström K, Panaccione DG, Panstruga R, Place M, Proctor RH, Prusky D, Rech G, Reinhardt R, Rollins J a, Rounsley S, Schardl CL, Schwartz DC, Shenoy N, Shirasu K, Sikhakolli UR, Stüber K, Sukno S a, Sweigard J a, Takano Y, Takahara H, Trail F, van der Does HC, Voll LM, Will I, Young S, Zeng Q, Zhang J, Zhou S, Dickman MB, Schulze-Lefert P, Ver Loren van Themaat E, Ma L-J, Vaillancourt LJ (2012) Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nat Genet 44:1060–5. doi: 10.1038/ng.2372

Pasharawipas T, Thaikua S, W. Flegel T (2009) Intriguing phylogenetic arrangement of tailed bacteriophages based on putative DNA polymerase sequences. ScienceAsia 35:125. doi: 10.2306/scienceasia1513-1874.2009.35.125

Queiroz CB, Correia HLN, Menicucci RP, Vidigal PMP, Queiroz MV (2017). Draft genome sequences of two isolates of *Colleotrichum lindemuthianum*, the causal agent of anthracnose in common beans. Genome Announc 5: pii: e00214-17. doi: 10.1128/genomeA.00214-17

Robicheau BM, Young AP, Labutti K, Grigoriev I V, Walker AK (2016) The complete mitochondrial genome of the conifer needle endophyte, *Phialocephala scopiformis* DAOMC 229536 confirms evolutionary division within the fungal *Phialocephala fortinii* s.l. – *Acephala appalanata* species complex. Fungal Biol species complex Fungal Biol 121:212–221. doi: 10.1016/j.funbio.2016.11.007

Rodríguez-Guerra R, Acosta-Gallegos JA, González-Chavira MM, Simpson J (2006) Patotipos de *Colletotrichum lindemuthianum* y su implicación en

la generación de cultivares resistentes de frijol. *Agric. técnica en México* 32:101–114.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542. doi: 10.1093/sysbio/sys029

Salavirta H, Oksanen I, Kuuskeri J, Mäkelä M, Laine P, Paulin L, Lundell T (2014) Mitochondrial genome of *Phlebia radiata* is the second largest (156 kbp) among fungi and features signs of genome flexibility and recent recombination events. *PLoS One*. doi: 10.1371/journal.pone.0097141

Santaquiteria A, Nielsen J, Klemetsen T, Willassen NP, Præbel K (2017) The complete mitochondrial genome of the long-lived Greenland shark (*Somniosus microcephalus*): characterization and phylogenetic position. *Conserv Genet Resour* 0:0. doi: 10.1007/s12686-016-0676-y

Saraste M (1999) Oxidative phosphorylation at the fin de siècle. *Science* 283:1488–1493. doi: 10.1126/science.283.5407.1488

Schneider A (2011) Mitochondrial tRNA import and its consequences for mitochondrial translation. *Annu Rev Biochem* 80:1033–53. doi: 10.1146/annurev-biochem-060109-092838

Sethuraman J, Majer A, Iranpour M, Hausner G (2009) Molecular evolution of the mtDNA encoded rps3 gene among filamentous ascomycetes fungi with an emphasis on the ophiostomatoid fungi. *J Mol Evol* 69:372–385. doi: 10.1007/s00239-009-9291-9

Sharma P, Sharma O (2008) Yield loss assessment in common bean due to anthracnose (*Colletotrichum lindemuthianum*) under sub temperate conditions of North-Western Himalayas. *Indian Phytopathol* 61:323–330.

Shen XY, Li T, Chen S, Fan L, Gao J, Hou CL (2015) Characterization and phylogenetic analysis of the mitochondrial genome of *Shiraia*

- bambusicola* reveals special features in the order of Pleosporales. PLoS One 10:1–20. doi: 10.1371/journal.pone.0116466
- Slack KE, Delsuc F, Mclenachan PA, Arnason U, Penny D (2007) Resolving the root of the avian mitogenomic tree by breaking up long branches. Mol Phylogenet Evol 42:1–13. doi: 10.1016/j.ympev.2006.06.002
- Specht CA, DiRusso CC, Novotny CP, Ullrich RC (1982) A method for extracting high-molecular-weight deoxyribonucleic acid from fungi. Anal Biochem 119:158–163. doi: 10.1016/0003-2697(82)90680-7
- Stone CL, Buitrago MLP, Boore JL, Frederick RD (2010) Analysis of the complete mitochondrial genome sequences of the soybean rust pathogens *Phakopsora pachyrhizi* and *P. meibomia*. Mycologia 102:887–897. doi: 10.3852/09-198
- Sun Z, Wang D, Murphy R, Ma L, Zhang X, Huang D (2009) Comparison of base composition and codon usage in insect mitochondrial genomes. Genes Genomics 31:67–67. doi: 10.1007/BF03191139
- Torriani SFF, Goodwin SB, Kema GHJ, Pangilinan JL, McDonald BA (2008) Intraspecific comparison and annotation of two complete mitochondrial genome sequences from the plant pathogenic fungus *Mycosphaerella graminicola*. Fungal Genet Biol 45:628–637. doi: 10.1016/j.fgb.2007.12.005
- Torriani SFF, Penselin D, Knogge W, Felder M, Taudien S, Platzer M, McDonald BA, Brunner PC (2014) Comparative analysis of mitochondrial genomes from closely related *Rhynchosporium* species reveals extensive intron invasion. Fungal Genet Biol 62:34–42. doi: 10.1016/j.fgb.2013.11.001
- van de Sande WWJ (2012) Phylogenetic analysis of the complete mitochondrial genome of *Madurella mycetomatis* confirms its taxonomic position within the order sordariales. PLoS One. doi: 10.1371/journal.pone.0038654

- Wei L, He J, Jia X, Qi Q, Liang Z, Zheng H, Ping Y, Liu S, Sun J (2014) Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. BMC Evol Biol 14:262. doi: 10.1186/s12862-014-0262-4
- Xia X (2005) Mutation and selection on the anticodon of tRNA genes in vertebrate mitochondrial genomes. Gene 345:13–20. doi: 10.1016/j.gene.2004.11.019
- Yang X, Cameron SL, Lees DC, Xue D, Han H (2015) A mitochondrial genome phylogeny of owlet moths (Lepidoptera: Noctuoidea), and examination of the utility of mitochondrial genomes for lepidopteran phylogenetics. Mol Phylogenet Evol 85:230–237. doi: 10.1016/j.ympev.2015.02.005
- Zhang S, Wang X-N, Zhang X-L, Liu X-Z, Zhang Y-J (2017) Complete mitochondrial genome of the endophytic fungus *Pestalotiopsis fici*: features and evolution. Appl Microbiol Biotechnol 101:1593–1604. doi: 10.1007/s00253-017-8112-0

## ARQUIVOS ADICIONAIS

**Tabela S1.** Características dos genomas mitocondriais de *Colletotrichum* spp.

Espécies	Acesso	Tamanho (bp)	GC%	PCGC	tRNA	rRNA	ncORF	IO	Intron	N.gene	PCG (%)	Região codificante (%)	GC%- PCG	GC% rRNA	GC% tRNA
<i>C. lindemuthianum</i>	Xxxx	37446	30,86	14	29	2	5	3	4	53	60,43	73,01	29,13	33,40	38,10
89 A <sub>2</sub> 2-3															
<i>C. lindemuthianum</i>	Xxxx	37440	30,86	14	29	2	5	3	4	53	60,42	73,00	29,11	33,40	38,10
83.501															
<i>C. lindemuthianum</i>	KF953885.1	36957	30,88	13	29	2	4	3	4	51	57,61	70,33	29,17	33,40	38,10
<i>C. fioriniae</i>	KU375885.1	30020	30,04	14	29	2	1	1	1	47	50,46	73,06	27,51	32,23	38,26
<i>C. acutatum</i>	KR349346.1	30892	30,51	14	28	2	2	1	1	47	53,23	75,26	28,57	32,28	38,41
<i>C. lupini</i>	KT918406.1	36554	29,90	14	28	2	10	1	1	55	57,44	76,07	28,32	32,30	38,55
<i>C. tamarilloi</i>	KU196965.1	30823	30,50	14	28	2	3	1	1	48	54,49	76,56	28,59	32,25	38,36
<i>C. graminicola</i>	CM001021.1	39649	29,90	14	25	2	9	2	2	52	64,11	80,63	28,23	32,57	38,20
<i>C. fructicola</i>	KX034082.1	56051	34,04	14	27	2	7	9	10	59	59,16	62,00	31,03	33,75	39,30
<i>C. siamense</i>	KX885101.1	58666	33,83	14	27	2	10	10	11	63	61,57	64,17	31,80	33,68	39,30



**Tabela S2.** Características gerais dos genomas mitocondriais de *Colletotrichum* spp.

<b>Espécies</b>	<b>Sobreposição (N)</b>	<b>Tamanho de sobreposição</b>	<b>Tamanho de região não codificante (%)</b>	<b>Tamanho das regiões intrônicas (%)</b>	<b>Tamanho das regiões intergênicas (%)</b>	<b>ncORF (%)</b>
<i>C. lindemuthianum</i> 89 A <sub>2</sub> 2-3	5	455	10107 (26,99)	5840 (15,60)	7414 (19,80)	4773 (12,75)
<i>C. lindemuthianum</i> 83.501	5	455	10107 (27,00)	5834 (15,58)	7414 (19,80)	4773 (12,75)
<i>C. lindemuthianum</i> KF953885.1	4	454	10966 (29,67)	5851 (15,83)	8262 (22,36)	3834 (10,37)
<i>C. fiorinae</i>	3	12	8087 (26,94)	1700 (5,66)	7809 (26,01)	723 (2,41)
<i>C. acutatum</i>	3	12	7644 (24,74)	1701 (5,51)	7365 (23,83)	2031 (6,57)
<i>C. lupini</i>	3	12	8749 (23,93)	1694 (4,63)	8749 (23,17)	6588 (18,02)
<i>C. tamarilloi</i>	3	12	7226 (23,44)	1694 (5,50)	7035 (22,82)	2385 (7,74)
<i>C. graminicola</i>	3	114	7679 (19,37)	2677 (6,75)	7291 (18,39)	10092 (25,45)
<i>C. fructicola</i>	3	12	21316 (38,03)	15779 (28,04)	15302 (27,30)	4827 (8,61)
<i>C. siamense</i>	4	26	21020 (35,83)	18084 (30,82)	13792 (23,51)	6786 (11,57)

**Tabela S3.** Anotação das ORFs não conservadas nos genomas mitocondriais de *Colletotrichum* spp.

<b>Espécies</b>	<b>ncORF</b>	<b>Posição</b>	<b>Fita</b>	<b>Função putativa</b>
<i>C. lindemuthianum_KF953885.1</i>	orf213	4912-5553	+	Unknown
<i>C. lindemuthianum_KF953885.1</i>	orf367	25613-26716	+	Unknown
<i>C. lindemuthianum_KF953885.1</i>	orf453	29736-31097	+	Unknown
<i>C. lindemuthianum_KF953885.1</i>	orf241	35274-35999	-	Unknown
<i>C. fioriniae</i>	orf240	1464-2186	+	Unknown
<i>C. acutatum</i>	orf240	1358-2080	+	Unknown
<i>C. acutatum</i>	orf435	26135-27442	+	Unknown
<i>C. lupini</i>	orf240	7263-7985	+	Unknown
<i>C. lupini</i>	orf437	12788-14101	+	Unknown
<i>C. lupini</i>	orf193	27038-27619	+	Unknown
<i>C. lupini</i>	orf270	28443-29255	+	Unknown
<i>C. lupini</i>	orf170	29308-29820	+	Unknown
<i>C. lupini</i>	orf107	29914-30237	+	Unknown
<i>C. lupini</i>	orf204	30350-30964	+	GIY-YIG endonuclease
<i>C. lupini</i>	orf151	31005-31460	+	Unknown
<i>C. lupini</i>	orf300	31635-32537	+	GIY-YIG endonuclease
<i>C. lupini</i>	orf114	35882-36226	-	Unknown
<i>C. tamarilloi</i>	orf115	1646-1993	-	Unknown
<i>C. tamarilloi</i>	orf240	9589-10311	+	Unknown
<i>C. tamarilloi</i>	orf437	15114-16427	+	Unknown
<i>C. graminicola</i>	orf264	3871-4665	+	Unknown
<i>C. graminicola</i>	orf295	6383-7270	+	Unknown
<i>C. graminicola</i>	orf376	8301-9431	+	Unknown
<i>C. graminicola</i>	orf747	13546-11468	-	DNA polymerase type B
<i>C. graminicola</i>	orf203	15818-16429	+	Unknown
<i>C. graminicola</i>	orf321	17905-18870	+	Unknown
<i>C. graminicola</i>	orf336	19102-20112	+	Unknown
<i>C. graminicola</i>	orf309	24863-25792	-	Unknown
<i>C. graminicola</i>	orf559	36393-38072	+	Unknown
<i>C. fructicola</i>	orf336	54515-55525	+	Unknown
<i>C. fructicola</i>	orf217	4145-4798	+	Unknown
<i>C. fructicola</i>	orf101	5130-5435	-	Unknown
<i>C. fructicola</i>	orf313	20421-21362	+	Unknown

<i>C. fructicola</i>	orf357	25140-26213	+	Unknown
<i>C. fructicola</i>	orf99	26681-26980	+	Unknown
<i>C. fructicola</i>	orf179	47519-48058	-	Unknown
<i>C. siamense</i>	orf268	4608-5414	+	Unknown
<i>C. siamense</i>	orf121_1	5891-6256	-	Unknown
<i>C. siamense</i>	orf313	21403-22344	+	Unknown
<i>C. siamense</i>	orf106	22331-22651	+	Unknown
<i>C. siamense</i>	orf116	24315-24665	+	Unknown
<i>C. siamense</i>	orf121_2	25003-25368	+	Unknown
<i>C. siamense</i>	orf402	26220-27428	+	Unknown
<i>C. siamense</i>	orf137	34305-34718	+	Unknown
<i>C. siamense</i>	orf327	50018-51001	-	Unknown
<i>C. siamense</i>	orf341	57115-58140	+	Unknown

---

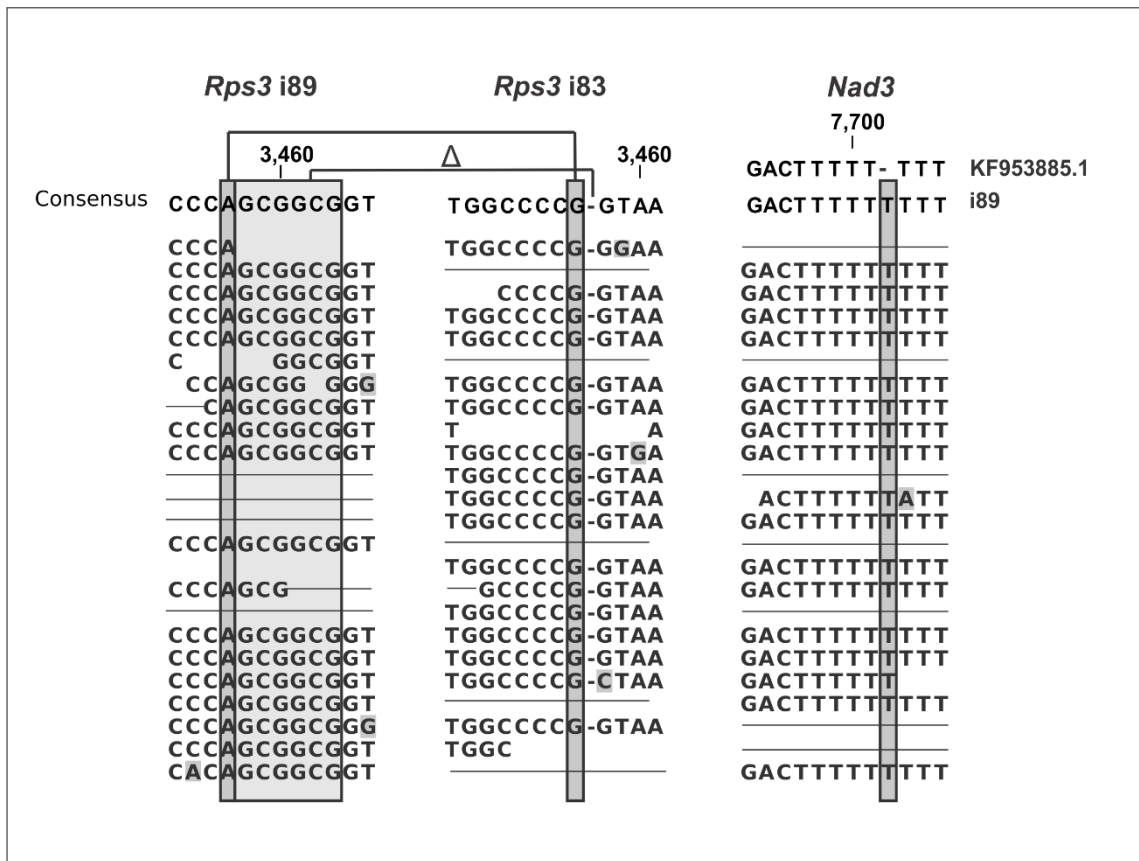
**Tabela S4.** Reads do transcriptoma de *Colletotrichum graminicola* que correspondem as ncORFs mitocondriais desta espécie.

Query id	Acesso GEO	Read id	Identidade	
			(%)	E-value
<i>orf747</i>	GSM852204	SRR391826.23467912	100	2.35908E-30
<i>orf747</i>	GSM852203	SRR391825.5766604	100	2.35908E-30
<i>orf747</i>	GSM852203	SRR391825.5635263	100	2.35908E-30
<i>orf747</i>	GSM852203	SRR391825.402952	100	2.35908E-30
<i>orf747</i>	GSM852203	SRR391825.26826827	100	8.48475E-30
<i>orf747</i>	GSM852203	SRR391825.26826481	100	8.48475E-30
<i>orf747</i>	GSM852203	SRR391825.9141724	97.37	5.10651E-27
<i>orf747</i>	GSM852203	SRR391825.3934602	97.37	5.10651E-27
<i>orf747</i>	GSM852203	SRR391825.26404987	98.48	3.9756E-23
<i>orf747</i>	GSM852203	SRR391825.22870921	100	1.8628E-16
<i>orf264</i>	GSM852204	SRR391826.10785059	100	9.21703E-31
<i>orf295</i>	GSM852204	SRR391826.32842087	100	1.0101E-30
<i>orf295</i>	GSM852204	SRR391826.29613761	100	1.0101E-30
<i>orf295</i>	GSM852204	SRR391826.25754766	100	1.0101E-30
<i>orf295</i>	GSM852204	SRR391826.19927166	100	1.0101E-30
<i>orf295</i>	GSM852204	SRR391826.19611415	100	1.0101E-30
<i>orf295</i>	GSM852204	SRR391826.17554436	100	1.0101E-30
<i>orf295</i>	GSM852204	SRR391826.14057987	100	1.0101E-30
<i>orf295</i>	GSM852204	SRR391826.13162368	100	1.0101E-30
<i>orf295</i>	GSM852204	SRR391826.7554324	100	1.0101E-30
<i>orf295</i>	GSM852204	SRR391826.5056561	100	1.0101E-30
<i>orf295</i>	GSM852204	SRR391826.3769004	100	1.0101E-30
<i>orf295</i>	GSM852204	SRR391826.1826307	100	1.0101E-30
<i>orf295</i>	GSM852204	SRR391826.1244210	100	1.0101E-30
<i>orf295</i>	GSM852203	SRR391825.29068378	100	1.0101E-30
<i>orf295</i>	GSM852203	SRR391825.4542511	100	1.0101E-30
<i>orf295</i>	GSM852201	SRR391823.19156743	100	1.0101E-30
<i>orf295</i>	GSM852203	SRR391825.15861840	100	3.63297E-30
<i>orf295</i>	GSM852204	SRR391826.19376444	100	1.30665E-29
<i>orf295</i>	GSM852204	SRR391826.22916300	98.68	4.69955E-29
<i>orf295</i>	GSM852204	SRR391826.21257183	100	4.69955E-29
<i>orf295</i>	GSM852201	SRR391823.1409121	97.37	2.18649E-27

<i>orf295</i>	GSM852204	SRR391826.8725915	94.74	4.73291E-24
<i>orf376</i>	GSM852204	SRR391826.33953141	100	1.29651E-30
<i>orf376</i>	GSM852204	SRR391826.33716030	100	1.29651E-30
<i>orf376</i>	GSM852204	SRR391826.31587407	100	1.29651E-30
<i>orf376</i>	GSM852204	SRR391826.31587241	100	1.29651E-30
<i>orf376</i>	GSM852204	SRR391826.15328823	100	1.29651E-30
<i>orf376</i>	GSM852204	SRR391826.3920876	100	1.29651E-30
<i>orf376</i>	GSM852203	SRR391825.31190256	100	1.29651E-30
<i>orf376</i>	GSM852203	SRR391825.22347081	100	1.29651E-30
<i>orf376</i>	GSM852203	SRR391825.18322061	100	1.29651E-30
<i>orf376</i>	GSM852203	SRR391825.15548032	100	1.29651E-30
<i>orf376</i>	GSM852203	SRR391825.2865682	100	1.29651E-30
<i>orf376</i>	GSM852204	SRR391826.32037917	100	4.66309E-30
<i>orf376</i>	GSM852204	SRR391826.12745515	98.68	6.03209E-29
<i>orf376</i>	GSM852201	SRR391823.24974341	98.68	6.03209E-29
<i>orf376</i>	GSM852204	SRR391826.2179668	100	7.803E-28
<i>orf203</i>	GSM852204	SRR391826.32551667	100	7.01217E-31
<i>orf203</i>	GSM852204	SRR391826.30275736	100	7.01217E-31
<i>orf203</i>	GSM852204	SRR391826.28314807	100	7.01217E-31
<i>orf203</i>	GSM852204	SRR391826.27530829	100	7.01217E-31
<i>orf203</i>	GSM852204	SRR391826.27235675	100	7.01217E-31
<i>orf203</i>	GSM852204	SRR391826.24873603	100	7.01217E-31
<i>orf203</i>	GSM852204	SRR391826.21298494	100	7.01217E-31
<i>orf203</i>	GSM852204	SRR391826.8914520	100	7.01217E-31
<i>orf203</i>	GSM852204	SRR391826.3480332	100	7.01217E-31
<i>orf203</i>	GSM852204	SRR391826.2959956	100	7.01217E-31
<i>orf203</i>	GSM852204	SRR391826.2410209	100	7.01217E-31
<i>orf203</i>	GSM852203	SRR391825.16900995	100	7.01217E-31
<i>orf203</i>	GSM852203	SRR391825.16292216	100	7.01217E-31
<i>orf203</i>	GSM852203	SRR391825.12695324	100	7.01217E-31
<i>orf203</i>	GSM852203	SRR391825.11259027	100	7.01217E-31
<i>orf203</i>	GSM852203	SRR391825.7858969	100	7.01217E-31
<i>orf203</i>	GSM852203	SRR391825.7594988	100	7.01217E-31
<i>orf203</i>	GSM852203	SRR391825.7518779	100	7.01217E-31
<i>orf203</i>	GSM852203	SRR391825.6377258	100	7.01217E-31
<i>orf203</i>	GSM852203	SRR391825.3508323	100	7.01217E-31
<i>orf203</i>	GSM852203	SRR391825.1785664	100	7.01217E-31

<i>orf203</i>	GSM852204	SRR391826.29882780	98.68	3.26245E-29
<i>orf203</i>	GSM852204	SRR391826.26986038	98.68	3.26245E-29
<i>orf203</i>	GSM852203	SRR391825.5237152	98.63	1.51787E-27
<i>orf203</i>	GSM852204	SRR391826.3681684	100	1.96349E-26
<i>orf203</i>	GSM852203	SRR391825.11073128	100	1.19855E-13
<i>orf203</i>	GSM852203	SRR391825.11073128	97.37	4.34137E-08
<i>orf203</i>	GSM852204	SRR391826.219416	100	3.35609E-09
<i>orf203</i>	GSM852204	SRR391826.219416	97.3	1.56143E-07
<i>orf321</i>	GSM852204	SRR391826.27689916	100	1.10204E-30
<i>orf321</i>	GSM852204	SRR391826.24664999	100	1.10204E-30
<i>orf321</i>	GSM852204	SRR391826.18161262	100	1.10204E-30
<i>orf321</i>	GSM852204	SRR391826.16582146	100	1.10204E-30
<i>orf321</i>	GSM852204	SRR391826.14840163	100	1.10204E-30
<i>orf321</i>	GSM852204	SRR391826.3915672	100	1.10204E-30
<i>orf321</i>	GSM852204	SRR391826.2203649	100	1.10204E-30
<i>orf321</i>	GSM852204	SRR391826.297985	100	1.10204E-30
<i>orf321</i>	GSM852203	SRR391825.27980497	100	1.10204E-30
<i>orf321</i>	GSM852203	SRR391825.24224013	100	1.10204E-30
<i>orf321</i>	GSM852203	SRR391825.18350760	100	1.10204E-30
<i>orf321</i>	GSM852203	SRR391825.15517022	100	1.10204E-30
<i>orf321</i>	GSM852203	SRR391825.6119124	100	1.10204E-30
<i>orf321</i>	GSM852201	SRR391823.13504077	100	1.10204E-30
<i>orf321</i>	GSM852204	SRR391826.16120035	100	8.57975E-27
<i>orf336</i>	GSM852204	SRR391826.32972265	100	1.15508E-30
<i>orf336</i>	GSM852204	SRR391826.31086318	100	1.15508E-30
<i>orf336</i>	GSM852204	SRR391826.12498466	100	1.15508E-30
<i>orf336</i>	GSM852204	SRR391826.10457639	100	1.15508E-30
<i>orf336</i>	GSM852204	SRR391826.6819416	100	1.15508E-30
<i>orf336</i>	GSM852203	SRR391825.29445948	100	1.15508E-30
<i>orf336</i>	GSM852203	SRR391825.25760618	100	1.15508E-30
<i>orf336</i>	GSM852203	SRR391825.6032413	100	1.15508E-30
<i>orf336</i>	GSM852201	SRR391823.2398414	98.68	5.37404E-29
<i>orf336</i>	GSM852201	SRR391823.1390196	98.68	5.37404E-29
<i>orf336</i>	GSM852201	SRR391823.7962852	97.37	2.5003E-27
<i>orf336</i>	GSM852203	SRR391825.1282940	93.22	5.48932E-14
<i>orf309</i>	GSM852204	SRR391826.33773115	100	1.00067E-30
<i>orf309</i>	GSM852204	SRR391826.28390545	100	1.00067E-30

<i>orf309</i>	GSM852204	SRR391826.20665421	100	1.00067E-30
<i>orf309</i>	GSM852204	SRR391826.17390890	100	1.00067E-30
<i>orf309</i>	GSM852204	SRR391826.15044630	100	1.00067E-30
<i>orf309</i>	GSM852204	SRR391826.10374727	100	1.00067E-30
<i>orf309</i>	GSM852204	SRR391826.4774452	100	1.00067E-30
<i>orf309</i>	GSM852204	SRR391826.479277	100	1.00067E-30
<i>orf309</i>	GSM852204	SRR391826.92052	100	1.00067E-30
<i>orf309</i>	GSM852203	SRR391825.24307641	100	1.00067E-30
<i>orf309</i>	GSM852203	SRR391825.23742523	100	1.00067E-30
<i>orf309</i>	GSM852203	SRR391825.16874544	100	1.00067E-30
<i>orf309</i>	GSM852203	SRR391825.13112683	100	1.00067E-30
<i>orf309</i>	GSM852203	SRR391825.3134464	100	1.00067E-30
<i>orf309</i>	GSM852203	SRR391825.899023	100	1.00067E-30
<i>orf309</i>	GSM852204	SRR391826.19303143	98.68	4.65568E-29
<i>orf309</i>	GSM852204	SRR391826.16824074	100	4.65568E-29
<i>orf309</i>	GSM852204	SRR391826.16823937	100	4.65568E-29
<i>orf309</i>	GSM852201	SRR391823.16826105	98.68	1.67448E-28
<i>orf559</i>	GSM852204	SRR391826.34050164	100	1.94359E-30
<i>orf559</i>	GSM852204	SRR391826.34049800	100	1.94359E-30
<i>orf559</i>	GSM852204	SRR391826.33175923	100	1.94359E-30
<i>orf559</i>	GSM852204	SRR391826.30565942	100	1.94359E-30
<i>orf559</i>	GSM852204	SRR391826.27749884	100	1.94359E-30
<i>orf559</i>	GSM852204	SRR391826.24709301	100	1.94359E-30
<i>orf559</i>	GSM852204	SRR391826.18569445	100	1.94359E-30
<i>orf559</i>	GSM852204	SRR391826.14492573	100	1.94359E-30
<i>orf559</i>	GSM852204	SRR391826.14421600	100	1.94359E-30
<i>orf559</i>	GSM852204	SRR391826.9757513	100	1.94359E-30
<i>orf559</i>	GSM852204	SRR391826.4271302	100	1.94359E-30
<i>orf559</i>	GSM852204	SRR391826.2957747	100	1.94359E-30
<i>orf559</i>	GSM852204	SRR391826.268654	100	1.94359E-30
<i>orf559</i>	GSM852203	SRR391825.15951781	100	1.94359E-30
<i>orf559</i>	GSM852203	SRR391825.13366656	100	1.94359E-30
<i>orf559</i>	GSM852203	SRR391825.2311731	100	1.94359E-30
<i>orf559</i>	GSM852204	SRR391826.26475135	100	6.9904E-30
<i>orf559</i>	GSM852201	SRR391823.7139368	96.05	1.95739E-25
<i>orf559</i>	GSM852203	SRR391825.4359644	100	1.54561E-11



**Figura S1.** Parte dos mapeamentos dos genes *rps3* e *nad3* com as *reads* do sequenciamento. As caixas cinzas no gene *rps3* destacam a deleção de seis bases e uma troca de A para G no isolado 83.501. No gene *nad3*, a caixa cinza destaca a ausência de um T no isolado de *C. lindemuthianum* de Gutiérrez et al. (2014) que desfaz o quadro de letra aberta. i89: isolado 89 A<sub>2</sub> 2-3; i83: isolado 83.501.



## CAPITULO 3

### **O REPERTÓRIO DE CANDIDATOS A EFETORES EM *Colletotrichum* spp. REVELA INFORMAÇÕES IMPORTANTES PARA SEU O ESTILO DE VIDA**

Queiroz CB, Santana MF, Batista DS, Correia HLN, Vidigal PMP, Queiroz MV. O repertório de candidatos a efetores em *Colletotrichum* revela informações importantes para o seu estilo de vida. Molecular Plant Pathology. Em preparação.

**O repertório de candidatas a efetoras em *Colletotrichum lindemuthianum* revela informações importantes para o seu estilo de vida**

Casley Borges de Queiroz<sup>1</sup>, Mateus Ferreira Santana<sup>1</sup>, Diego Silva Batista<sup>2</sup>, Hilbert L. Nunes Correia<sup>1</sup>, Pedro M. Pereira Vidigal<sup>3</sup>, Marisa Vieira de Queiroz<sup>1\*</sup>

<sup>1</sup>Laboratório de Genética Molecular de Fungos (LGMF)/BIOAGRO, CEP: 36570-900, Universidade Federal de Viçosa, Minas Gerais, Brazil

<sup>2</sup>Laboratório de Cultura de Tecidos (LCTII)/BIOAGRO, CEP: 36570-900, Universidade Federal de Viçosa, Minas Gerais, Brazil

<sup>3</sup>Núcleo de Análise de Biomoléculas (NuBioMol), Centro de Ciências Biológicas, Universidade Federal de Viçosa, Minas Gerais, Brazil

\*Corresponding author

Laboratório de Genética Molecular de Fungos (LGMF)/BIOAGRO, CEP: 36570-900, Universidade Federal de Viçosa, Minas Gerais, Brazil. +55 (31) 3899-2975. E-mail address: [mvqueiro@ufv.br](mailto:mvqueiro@ufv.br).

## RESUMO

O fungo *Colletotrichum lindemuthianum* é o agente causal da antracnose no feijoeiro (*Phaseolus vulgaris*), uma das doenças mais devastadoras nessa cultura. Entretanto, pouco se sabe sobre as proteínas essenciais para a interação fungo-planta. Por causa disso, o conhecimento a cerca do arsenal de proteínas efetoras do fungo se torna de grande importância para o melhor conhecimento desse patossistema. Foi realizada neste trabalho a predição de genes de proteínas candidatas à efetoras de isolados de duas raças fisiológicas de *C. lindemuthianum*. Foram preditos ao total 324 genes que codificam proteínas candidatas a efetoras, sendo que 77,1% apresentam homologia às sequências não caracterizadas e não apresentam qualquer domínio conservado. Entretanto, 67% das candidatas a efetoras são ricas em cisteína, possuem sequências repetitivas de aminoácidos e/ou possuem sequências de localização nuclear. Diversos domínios conservados foram encontrados nas candidatas a efetoras de *C. lindemuthianum*. Com o intuito de comparar o arsenal de proteínas candidatas a efetoras de *C. lindemuthianum* com os de outras espécies de *Colletotrichum*, foi realizada também a predição de efetores de nove espécies pertencentes a este gênero. Nas outras espécies o número de genes de proteínas candidatas à efetoras variou de 247 em *Colletotrichum graminicola* até 446 em *Colletotrichum orbiculare*. Doze domínios conservados foram preditos nas proteínas em todas as espécies de *Colletotrichum* analisadas. Foram identificados também proteínas candidatas a efetoras com presença de domínios conservados ainda não relatados para efetoras de *Colletotrichum* spp. A análise da expressão de seis genes de proteínas candidatas a efetoras de *C. lindemuthianum* revelou indução durante a fase biotrófica do fungo no feijoeiro. Este trabalho apresenta diversas genes codificadores de proteínas candidatas à efetoras e fornece uma valiosa fonte de estudo para elucidar os mecanismos de infecção desses patógenos em seus hospedeiros.

## INTRODUÇÃO

Recentemente o gênero *Colletotrichum* foi classificado como o oitavo gênero de fungo fitopatogênico mais importante do mundo, devido ao impacto econômico das doenças causadas pelas diferentes espécies e por ser um gênero modelo para estudos da interação planta-patógeno (Crouch *et al.*, 2014; Dean *et al.*, 2012). Este é o caso da espécie *Colletotrichum lindemuthianum* (Sacc. & Magnus) Briosi & Cavara, agente causal da antracnose no feijoeiro comum (*Phaseolus vulgaris* L) (Mahuku and Riascos, 2004). Em condições favoráveis de umidade e temperatura, a antracnose pode levar a uma redução drástica na produção do feijão (Richard *et al.*, 2014; Schwartz *et al.*, 2005). Os estudos de interação entre *C. lindemuthianum* e *P. vulgaris* iniciaram com o trabalho de Barrus, (1918), sendo este o primeiro estudo da interação patógeno-hospedeiro para espécies fitopatogênicas pertencentes ao gênero *Colletotrichum*. *Colletotrichum lindemuthianum* tem uma alta variabilidade genética (Mota *et al.*, 2016; Padder *et al.*, 2007), caracterizada pela existência de pelo menos 100 variantes patogênicas ou raças fisiológicas reportadas entre isolados coletados em várias regiões do mundo (Alzate-Marin *et al.*, 1999; Ferreira *et al.*, 2013; González *et al.*, 2015; Rodríguez-Guerra *et al.*, 2006). Esta alta variabilidade genética é uma das principais barreiras para o controle deste patógeno, uma vez que dificulta a obtenção de cultivares com resistência durável (Ferreira *et al.*, 2013).

*Colletotrichum lindemuthianum*, assim como a maioria das espécies fitopatogênicas pertencentes ao gênero *Colletotrichum*, é um fitopatógeno com estilo de parasitismo hemibiotrófico (Gonzaga *et al.*, 2015; Hiruma *et al.*, 2016; Mejía *et al.*, 2014). No caso dos hemibiotróficos, o processo de infecção, envolve a formação sequencial de uma série de tipos celulares especializados (Perfect *et al.*, 1999). Após a aderência e germinação do conídio, ocorre a formação de um tubo germinativo que subsequentemente forma o apressório. A hifa de penetração emerge da base do apressório, que rompe a cutícula e a parede celular da célula da planta por meio da força mecânica gerada pela pressão de turgor associada à ação de enzimas

líticas secretadas (Bechinger *et al.*, 1999). Dentro das células epidérmicas da planta, a hifa de penetração se diferencia em hifas primárias bulbosas, que se expandem e são envelopadas pela membrana plasmática do hospedeiro (O'Connell *et al.*, 2012; Perfect *et al.*, 1999). É mantida uma matriz interfacial separando a membrana plasmática hospedeira da parede celular do fungo e células recém-infectadas permanecem viáveis por algum tempo (Perfect *et al.*, 1999). Esse estágio assintomático de infecção é denominado de biotrófico, pelo fato das células hospedeiras invadidas permanecerem vivas (O'Connell *et al.*, 2012). Após a fase biotrófica, o fungo muda para a fase necrotrófica. Nesta fase, hifas secundárias emergem das hifas primárias e invadem células vizinhas à célula inicialmente infectada, matando-as e degradando a parede celular para prosseguir a infecção (O'Connell *et al.*, 2012).

Durante o processo de interação planta-patógeno, ocorre uma disputa intensa entre os dois organismos pela sobrevivência (Giraldo and Valent, 2013). A planta apresenta duas linhas de defesa (Dodds and Rathjen, 2010). A primeira envolve a resposta basal de defesa orquestrada por receptores de reconhecimento de padrões (PRRs) localizados na membrana, que detectam a presença de moléculas amplamente conservadas, que são características de micro-organismos patogênicos, nomeadas de padrão molecular associado ao patógeno (PAMPs) e ativam uma resposta específica de imunidade desencadeada por PAMPs (PTI) (Oliveira-Garcia and Valent, 2015; Selin *et al.*, 2016). As PAMPs referem-se a qualquer tipo de molécula estrutural amplamente conservada entre os patógenos, por exemplo, componentes da parede celular, como a quitina (Giraldo and Valent, 2013; Motion *et al.*, 2015). Uma vez que o sistema de defesa da planta detecta a presença invasiva do patógeno, a planta pode induzir a produção de espécies reativas de oxigênio (ROS), bem como a secreção de compostos antimicrobianos, enzimas hidrolíticas, que são danosas para o patógeno, e inibidores de enzimas hidrolíticas do patógeno, que são danosas para a planta (Asai and Shirasu, 2015; Giraldo and Valent, 2013). Em resposta a essa defesa da planta, o patógeno secreta os efetores, que geralmente são pequenas moléculas, proteínas solúveis ou pequenos RNAs, que manipulam a estrutura e fisiologia da célula

hospedeira para conseguir infectá-la (Asai and Shirasu, 2015; Giraldo and Valent, 2013; Selin et al., 2016). Os efetores permitem que o patógeno escape ou desarme os mecanismos de defesa, por exemplo, inibindo a PTI, e suprimindo a morte programada das células que podem restringir a colonização do patógeno (Göhre and Robatzek, 2008; Rafiqi et al., 2012). A planta por sua vez, desenvolve uma segunda linha de defesa chamada de imunidade desencadeada por efector (ETI), regida por proteínas receptoras intracelulares NB-LRR (*nucleotide-binding site leucine-rich repeat*) codificadas por genes de resistências, capazes de reconhecer proteínas efectoras chamadas de proteínas de avirulência (Jones and Dangl, 2006). O sistema ETI frequentemente desencadeia uma resposta de morte celular localizada, chamada de resposta de hipersensibilidade (Dodds and Rathjen, 2010).

A presença dessas proteínas efectoras fúngicas dentro das células infectadas implica em um importante papel na virulência, mas ainda não se sabe com clareza a função da maioria delas. Geralmente, os efetores fúngicos são evolutivamente diversos e muitos deles não apresentam qualquer sequência com homologia à proteínas depositadas em banco de dados (Rafiqi et al., 2012). Além disso, a função biológica dos efetores é difícil para elucidar, provavelmente em razão da redundância funcional, ou seja, diversos efetores podendo exercer funções iguais (Giraldo and Valent, 2013; Rafiqi et al., 2012). Contudo, para alguns efetores, o estudo de suas funções tem sido realizado com sucesso. O efector CMU1 de *Ustilago maydis* foi o primeiro efector fúngico a ter uma clara descrição de sua função, que é de corismato mutase (Djamei et al., 2011). Diversos patógenos secretam efetores com domínio LysM (*lysine motif*) que protegem a parede celular do fungo da ação de quitinases da planta, ou sequestram os quitooligosacarídeos que foram liberados, para evitar o reconhecimento pelo sistema de defesa da planta (Kombrink et al., 2016), como no caso de SLP1 de *Magnaporthe oryzae* (Mentlak et al., 2012), Mg1LysM e Mg3LysM de *Mycosphaerella graminicola* (Marshall et al., 2011), ECP6 e AVR4 de *Cladosporium fulvum* (Jonge et al., 2010).

Em *C. higginsianum* foi mostrado que os genes relacionados a proteínas candidatas a efectoras foram expressos em quatro ondas

principais, referentes a efetores expressos antes da penetração na planta, logo após a penetração, durante a fase biotrófica e transição da fase biotrófica para fase necrotrófica ou durante a fase necrotrófica, sendo assim mostrado que os efetores foram necessários em diferentes estágios de invasão (Kleemann *et al.*, 2012). Em *C. graminicola*, foi relatado que o gene que codifica a proteína efetora CgEP1 foi altamente expresso na hifa primária durante a fase biotrófica. Foi demonstrado que CgEP1 é endereçada para o núcleo do milho, e também que é capaz de se ligar ao DNA de forma não específica, podendo está envolvida nos processos regulatórios de prevenção de resposta defensiva do hospedeiro. Entretanto, essa sequência codificadora parece ser exclusiva do clado graminicola (Vargas *et al.*, 2016). Foi demonstrado também em *C. graminicola* que Cgfl é uma proteína efetora com função de fungolisina que possui papel na virulência deste patógeno, podendo suprimir a defesa hospedeira por meio da degradação de quitinases (Sanz-Martín *et al.*, 2016). Tem sido demonstrado em *Colletotrichum truncatum* que a proteína CtNUDIX induz severa morte celular no tabaco, similar a resposta de hipersensibilidade, e que sua expressão ocorre na biotrofia tardia podendo assim possuir um mecanismo sofisticado para troca da biotrofia para necrotrofia (Bhadauria *et al.*, 2013). Os autores especulam que CtNUDIX perturba a função da membrana plasmática hospedeira podendo causar extravazamentos, e isto permitiria a entrada de proteínas não-nativas que seriam reconhecidas pelo sistema de defesa da planta, que induziriam a morte celular (Bhadauria *et al.*, 2013). É provável que a proteína efetora CltoxB de *Colletotrichum lentis* atue durante a troca da biotrofia para necrotrofia para amplificar o sinal de morte celular causada pelo efector CINUDIX (Bhadauria *et al.*, 2015).

Apesar dos relatos supracitados de estudos de genes relacionados à efetores em *Colletotrichum* spp., uma abordagem mais descritiva e comparativa ainda se faz necessária, principalmente se levarmos em consideração a diversidade de espécies deste gênero causadoras de doenças. Neste sentido, espécies economicamente importantes pertencentes a este gênero foram ainda pouco estudadas quanto ao seu repertório de possíveis efetores. Neste trabalho, foi realizada a predição do repertório de genes de proteínas candidatas a efetoras no genoma de dois

isolados pertencentes a duas raças fisiológicas de *C. lindemuthianum* e a comparação com nove espécies de *Colletotrichum*. Diversos genes de proteínas candidatas a efetoras com funções putativas foram identificados, que podem auxiliar o fungo durante o processo de infecção em seus hospedeiros. Além disso, a expressão diferencial *in planta* de seis genes de proteínas candidatas a efetoras de *C. lindemuthianum* foi determinada.

## MATERIAL E MÉTODOS

### **Predição de genes de proteínas candidatas à efetoras em *Colletotrichum* spp.**

A predição de genes de proteínas candidatas a efetoras foi feita por meio de sequências de aminoácidos de proteínas preditas de onze genomas de *Colletotrichum*, contabilizando dez espécies: *C. lindemuthianum* (MASO00000000.2 e MASP00000000.2), *Colletotrichum fructicola* Nara gc5 (ANPB00000000.1), *Colletotrichum fioriniae* PJ7 (JARH00000000.1), *Colletotrichum nymphaea* SA-01 (JEMN00000000.1), *Colletotrichum orbiculare* MAFF 240422 (AMCV00000000.1), *Colletotrichum graminicola* M1.001 (ACOD00000000.1), *Colletotrichum sublineola* TX430BB (JMSE00000000.1), *Colletotrichum salicis* CBS 607.94 (JFFI00000000.1), *Colletotrichum simmondsii* CBS122122 (JFBX00000000.1) e *Colletotrichum higginsianum* IMI 349063 (GCA\_000313795.2).

As sequências de aminoácidos preditas foram submetidas a um fluxo de análise constituído de uma série de filtragens, iniciando com SignalP 4.1 (Petersen *et al.*, 2011) para detectar sequências de aminoácidos que possuam peptídeos sinais com sítio de clivagem na região N-terminal e TargetP 1.1 (Emanuelsson *et al.*, 2000) e WoLF-PSORT (Horton *et al.*, 2007) para identificar e excluir proteínas de localização sub-celular. Foram descartadas sequências de aminoácidos com domínios transmembrana usando os programas TMHMM 2.0 (Emanuelsson *et al.*, 2007) e Phobius (Käll *et al.*, 2007), pois tais proteínas são endereçadas para a membrana da célula do fungo. Também foram excluídas por meio do programa BIG-PI



Fungal Predictor (Eisenhaber *et al.*, 2004) as sequências de aminoácidos que possuem âncoras GPI (glicofosfatidilinositol) que representam proteínas de superfície celular do fungo em vez de proteínas efetoras. Foram excluídas todas as sequências de proteínas com mais de 300 aminoácidos (Hacquard *et al.*, 2012) com exceção daquelas que apresentaram homologia com efetores conhecidos de patógenos de plantas. As sequências de aminoácidos de proteínas solúveis pequenas secretadas que foram preditas e que não apresentaram qualquer similaridade com sequências presentes no banco de dados UniProt com o limite (*threshold*) de  $1e^{-3}$ , ou que apresentaram *hits* com outras sequências de proteínas efetoras descritas em outros fungos foram definidas como candidatas à efetoras (O'Connell *et al.*, 2012).

Para predição de sequência sinal de localização nuclear, foram usados os programas NLStradamus (Nguyen Ba *et al.*, 2009), Myhits (Pagni *et al.*, 2004) e NLS Mapper (Kosugi *et al.*, 2009) com parâmetros *default*. Uma limitação na identificação de efetores *in silico* é a ausência de domínios conservados que possam ser preditos em bancos de dados de domínios como o Pfam, com exceção daqueles envolvidos na patogenicidade. Contudo, diversos efetores já comprovados experimentalmente possuem domínios conservados. Para demonstrar isto, nós fizemos a busca por domínios no banco de dados Pfam usando 142 sequências de efetores comprovados experimentalmente, cujas sequências foram adquiridas com buscas na literatura e no PHI-base (*Pathogen Host Interactions*) que é um banco de dados que possui sequências de aminoácidos no qual o envolvimento com a patogenicidade foi validado experimentalmente (Urban *et al.*, 2015). Da mesma forma, a presença de domínios conservados para candidatas a efetoras em *Colletotrichum* spp. foi determinada usando os bancos de dados Pfam e NCBI CDD (*Conserved Domain Database*) com limite de corte para *e-value* de  $1e^{-5}$ . Como proteínas efetoras são conhecidas também por serem ricas em cisteína, foi utilizado um *script* em *awk* para identificar sequências de aminoácidos candidatas à efetoras com conteúdo de cisteína maior que 3% em *C. lindemuthianum*. Proteínas contendo repetições (RCP) de resíduos de aminoácidos são frequentemente encontradas em efetoras de fungos e oomicetos (Saunders *et al.*, 2012).

Foram utilizados os programas T-REKS (Jorda and Kajava, 2009) e XSTREAM (Newman and Cooper, 2007) com parâmetros *default*, para identificar sequências contendo repetições em tandem. Os termos do *Gene ontology* (GO) foram atribuídos às sequências usando o programa Blast2go (Conesa *et al.*, 2005).

A análise da presença ou ausência de grupos de candidatos a efetores homólogos entre as espécies de *Colletotrichum* foi obtida por meio do programa CD-HIT (Huang *et al.*, 2010) com limite de identidade de 30%. Para identificar possíveis similaridade entre candidatos a efetores identificados em *C. lindemuthianum* com outras sequências de efetores já descritos em fungos e oomicetos fitopatógenos, foi utilizada a ferramenta BLASTp com valor de corte de *e-value* de 1e-5 contra um banco de dados local construído com 142 sequências de proteínas efetoras.

### **Análise da transcrição de candidatos a efetores de *Colletotrichum lindemuthianum***

Foram coletadas folhas do feijoeiro às 48 e 94 h após a inoculação, que correspondem às fases biotrófica e necrotrófica, respectivamente. Também foram coletados micélios vegetativos crescidos durante sete dias em meio de cultura GPYECH (20g de glicose, 5g de peptona, 1g de extrato de levedura e 1 g de caseína L<sup>-1</sup>). O RNA total foi extraído com o kit NucleoSpin<sup>®</sup> RNA Plant (Macherey-Nagel) seguindo os procedimentos recomendados pelo fabricante. A qualidade e quantidade do RNA foi determinada por espectrofotômetro NanoDrop<sup>®</sup> 2000 da Thermo Scientific. Os cDNAs foram sintetizados utilizando o kit ImProm-II Reverse Transcription System (Promega Corporation, Madison, WI, USA). Os genes de proteínas candidatas a efetoras analisados foram *clcc10*, *clcc53*, *clcc97*, *clcc147*, *clcc178* e *clcc229*, além do gene da actina como normalizador (Arquivo adicional: Tabela S4). As reações de RT-qRT-PCR (*Real-Time Quantitative Reverse Transcription PCR*) foram realizadas utilizando o kit SYBR<sup>®</sup> Green PCR-Master Mix (Applied Biosystems, FosterCity,CA,USA) e as leituras foram realizadas por meio do equipamento CFX96Touch<sup>™</sup> Real-Time PCR Detection System(BIO-RAD<sup>®</sup>) sob as seguintes condições: 50°C

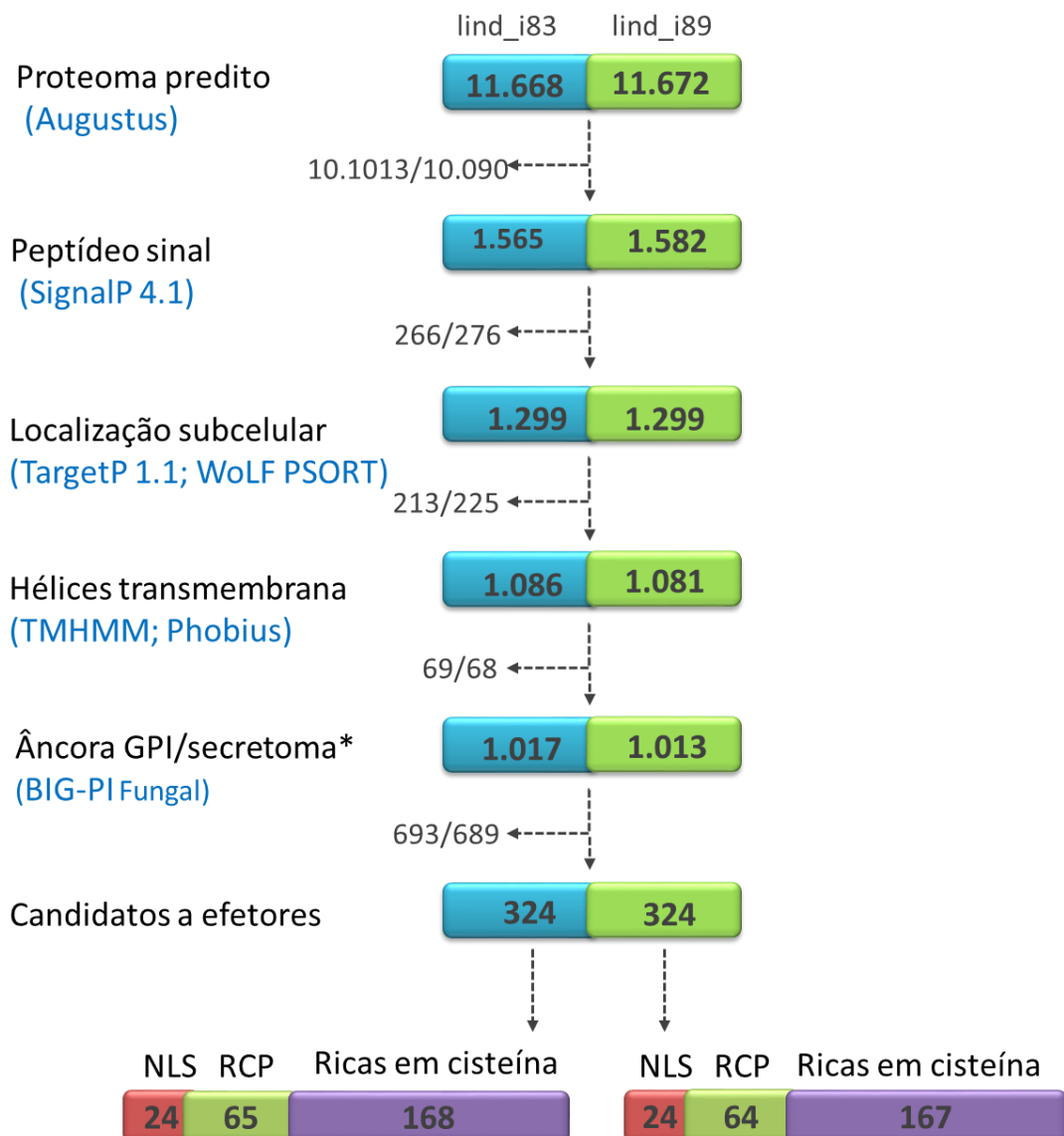
por 2 minutos seguido por 95°C por 10 minutos; 40 ciclos de 95°C por 15 segundos, 60°C por 1 minuto e 95 °C por 0,01 segundo; a curva de *melting* foi obtida por um aumento progressivo em temperatura de 0,5°C/0,05 segundos a partir de 60°C até 95°C. O perfil de expressão gênica foi normalizada para o nível de expressão do gene que codifica actina (em *C. lindemuthianum*) e os cálculos dos valores dos níveis relativos de expressão foram feitos usando o método comparativo ( $2^{-\Delta\Delta Ct}$ ) (Livak & Schmittgen 2001), com três replicatas biológicas para cada tratamento e as médias foram comparadas usando o intervalo de confiança por meio do teste de Dunnet ( $P \leq 0.05$ ).

## RESULTADOS

### **Sequências de candidatas à efetoras em *Colletotrichum lindemuthianum***

Para identificar as sequências aminoácidos de proteínas candidatas à efetoras nos genomas de *C. lindemuthianum*, foi construído um fluxo de análise com base nas características típicas de efetores conhecidos de patógenos de plantas. Inicialmente, foi efetuada a predição para dois isolados de *C. lindemuthianum* (Figura 1) pertencentes às raças fisiológicas 83 e 89. Nesta espécie foram preditas 1.299 sequências de proteínas com sinal de secreção e que não possuem como destino final as mitocôndrias e outras regiões sub-celulares em ambos os isolados. Destas, 1.086 e 1.081 proteínas nos isolados 83.501 e 89 A<sub>2</sub> 2-3, respectivamente, não apresentam hélices transmembranas. Foram excluídas 69 e 68 proteínas dos isolados 83.501 e 89 A<sub>2</sub> 2-3, respectivamente, que possuem motivos de ancoragem de glicofosfatidilinositol (âncora-GPI). Assim, um total de 1.017 (8,7%) e 1.013 (8,6%) sequências de aminoácidos de proteínas dos isolados 83.501 e 89 A<sub>2</sub> 2-3, respectivamente, foram preditas como sendo secretadas. Como a maioria das proteínas efetoras possuem menos que 300 aminoácidos (Guyon et al., 2014; Pattemore et al., 2014), todas as sequências com mais de 300 aminoácidos foram excluídas das análises,

com exceção de 18 sequências que apresentaram similaridades com efetores conhecidos de patógenos de plantas. Dessa forma, um total de 324 proteínas foram preditas como candidatas a efetoras de *C. lindemuthianum* (CIEC) em ambos os isolados 83.501 e 89 A<sub>2</sub> 2-3 (Tabela S1 e S2), representando 31,8% de todo o secretoma predito e com um conteúdo médio de cisteína de 3,1%. Sequências codificadoras de candidatos a efetores exclusivas de cada isolado foram referidas como CIEC\_i83 para o isolado 83.501 e CIEC\_i89 para o isolado 89 A<sub>2</sub> 2-3 (Tabela S1 e S2).



**Figura 1.** Fluxo de trabalho das análises *in silico* para predição de candidatos a efetores em *Colletotrichum lindemuthianum*. lind\_i83 e lind\_i89, referem-se aos dois isolados de *C. lindemuthianum*. Os números dentro das caixas representam a

quantidade de sequências de aminoácidos selecionadas em cada etapa. Os números ao lado da seta tracejada indicam o número de sequências de aminoácidos retiradas em cada etapa. NLS (sinal de localização nuclear), RCP (proteínas contendo repetições), cysteine rich (proteínas ricas em resíduos de cisteína).

\*Sequências de proteínas maiores que 300 aminoácidos que apresentaram similaridade com efetores foram adicionadas de volta, e aquelas com *cut-off* menor que  $1e-3$  para proteínas não efetoras foram excluídas.

Quatro CIECs foram encontradas somente em *C. lindemuthianum*, em ambos os isolados, podendo ser efetores espécie-específicos (Tabela S1 e S2). A comparação de todos os genes CIECs entre os dois isolados por meio da ferramenta BLASTn revelou que 64 CIECs possuem identidade menor ou igual a 99,96% e a mínima com 75,09% (Tabela S3). Além desses, dois CIECs foram preditos como exclusivos em cada isolado: CIEC80 e CIEC119 no isolado 83.501 e CIEC118b e CIEC281 no isolado 89 A<sub>2</sub> 2-3. Foram realizados mapeamentos com as *reads* de cada isolado na seguinte forma: *reads* do sequenciamento do isolado 89 A<sub>2</sub> 2-3 contra os genes CIECs preditos somente no isolado 83.501 e *reads* do sequenciamento do isolado 83.501 contra os genes CIECs preditos somente no isolado 89 A<sub>2</sub> 2-3. Nenhuma *read* do 89 A<sub>2</sub> 2-3 foi mapeada para o gene CIEC80, enquanto que para o gene CIEC119 as *reads* foram mapeadas somente em 62 bases a partir da posição 709 na região 3', sendo o tamanho total de CIEC119 de 768 pb. Para o gene CIEC118b, na posição 25 as *reads* do isolado 83.501 apresentam dois nucleotídeos a mais (CT), os quais levaram a perda do quadro de leitura. Já o mapeamento no gene CIEC281 revelou que o gene está completo também no isolado 83.501.

Neste trabalho também foram realizadas buscas por três características marcantes encontradas em alguns tipos de proteínas efetoras já descritas na literatura: proteínas contendo repetições (RCPs) de resíduos de aminoácidos em tandem, proteínas com conteúdo de cisteína maior que 3% (proteínas ricas em cisteína) e sinal de localização celular (NLS). Foram encontradas 65 e 64 CIECs que são RCPs, nos isolados 83.501 e 89 A<sub>2</sub> 2-3, respectivamente. Foram identificadas 168 e 167 CIECs ricas em cisteínas

nos isolados 83.501 e 89 A<sub>2</sub> 2-3, respectivamente. (Arquivo adicional: Tabela S1). Em ambos os isolados foram identificados 24 CIECs contendo NLS que podem atuar no núcleo da planta. No total foram identificadas 257 (83.501) e 255 (89 A<sub>2</sub> 2-3) proteínas contendo pelo menos uma dessas características (Figura 1, Tabela S1 e S2 e Arquivo adicional: Tabela S1).

### **Candidatas a proteínas efetoras de *C. lindemuthianum* apresentam domínios conservados e similaridades com efetoras conhecidas de fitopatógenos**

Diversas proteínas efetoras já comprovadas experimentalmente possuem domínios conservados (Ma *et al.*, 2015; Pareja-Jaime *et al.*, 2008; Fellbrich *et al.*, 2002; Rose *et al.*, 2002). Para demonstrar isto, foi realizada uma busca por domínios no banco de dados Pfam usando 142 sequências de aminoácidos de proteínas efetoras comprovadas experimentalmente, adquiridas por meio de busca na literatura e no PHI-base. Entre as proteínas efetoras comprovadas, 59 apresentaram domínios conservados como, por exemplo, o domínio PAN\_1 (PF00024) presente no efector CBEL de *Phytophthora parasítica* var *nicotianae* (Mateos *et al.*, 1997); o domínio glycosyl hydrolase 10 (PF00331) presente no efector AvrLm4 de *Leptosphaeria maculans* (Huang *et al.*, 2006); o domínio dedo de zinco FYVE (PF01363) presente no efector Avr1b-1 *Phytophthora sojae* (Shan *et al.*, 2004); o domínio Lipase\_3 (PF01764) presente no efector FGL1 de *Fusarium graminearum* (Blumke *et al.*, 2014); além de outros efetores contendo outros domínios (Arquivo adicional: Tabela S3).

Como demonstrado nas análises supracitadas, alguns fungos e oomicetos possuem proteínas efetoras que possuem domínios conservados. Para identificar a presença de domínios conservados em CIECs, foram realizadas buscas nos bancos de dados Pfam e CDD com limite de corte de *e-value* 1e-5. Foram identificadas 62 (83.501) e 61 (89 A<sub>2</sub> 2-3) CIECs contendo domínios conservados, sendo os mais abundantes: o domínio de homologia a Cianovirina-N (PF08881.7), o domínio LysM (PF01476.17), o

domínio metaloprotease Peptidase\_M35 (PF02102.12) e o domínio de NPP1 (PF05630.8) (Tabela S1 e S2 e arquivo adicional: Tabela S2). Também foi construído um banco de dados local de proteínas efetoras conhecidas por meio de busca na literatura e no PHI-base (Urban *et al.*, 2015), pelo qual foi possível identificar 23 CIECs em ambos os isolados que apresentam similaridades à 12 proteínas efetoras conhecidas de outros fitopatógenos (Tabela 1).

**Tabela 1.** Candidatas a efetoras de *Colletotrichum lindemuthianum* homólogos à efetoras conhecidas de patógenos de planta

CIECs	e-value	Efetores	Domínios (Pfam)	Tamanho	Referências
CIEC139	6,38e-12	AvrLm6	-	154	Fudal et al., 2007
CIEC264	3,5e-15	AvrLm6	-	159	Fudal et al., 2007
CIEC82	1,05e-10	AVR-Pita	Peptidase_M35 (PF02102.12)	355	Kang et al., 2001
CIEC19	3,76e-11	AVR-Pita	Aspzincin_M35 (PF14521.3)	355	Kang et al., 2001
CIEC143	1,14e-11	AVR-Pita	Peptidase_M35 (PF02102.12)	359	Kang et al., 2001
CIEC147	1,13e-17	AVR-Pita	Peptidase_M35 (PF02102.12)	359	Kang et al., 2001
CIEC118b_i89	5,74e-37	BAS2	-	100	Mosquera et al., 2009
CIEC118	1,52e-18	BAS2	-	177	Mosquera et al., 2009
CIEC166	5,00e-07	BAS3	-	120	Mosquera et al., 2009
CIEC12	1,60e-08	BAS4	-	94	Mosquera et al., 2009
CIEC211	1,77e-08	BAS4	-	117	Mosquera et al., 2009
CIEC93	3,27e-06	ECP6	LysM (PF01476.17)	91	Bolton et al., 2008
CIEC70	3,05e-76	MgSM1	Cerato-platanin (PF07249.9)	138	Yang et al., 2009
CIEC266	1,22e-27	NPP1	NPP1 (PF05630.8)	379	Fellbrich et al., 2002
CIEC1	5,0e-71	PaNie	NPP1 (PF05630.8)	239	Veit et al., 2001
CIEC180	3,18e-26	PaNie	NPP1 (PF05630.8)	302	Veit et al., 2001
CIEC144	2,15e-51	PaNie	NPP1 (PF05630.8)	244	Veit et al., 2001
CIEC234	8,71e-28	PaNie	NPP1 (PF05630.8)	279	Veit et al., 2001
CIEC280	1,13e-28	PaNie	NPP1 (PF05630.8)	578	Veit et al., 2001
CIEC84	1,74e-34	PsojNIP	NPP1 (PF05630.8)	337	Qutob et al., 2002
CIEC165	7,70e-36	SIX1	-	261	Rep et al., 2004
CIEC205	3,58e-20	SIX1	-	287	Rep et al., 2004
CIEC188	9,97e-29	SLP1	LysM (PF01476.17)	230	Mentlak et al., 2012

## Análise comparativa entre o repertório de proteínas candidatas à efetoras em *Colletotrichum lindemuthianum* e outras nove espécies de pertencetes ao gênero *Colletotrichum*

Além da predição supracitada de efetores para dois isolados de *C. lindemuthianum*, realizou-se a predição de candidatos a efetores para nove espécies de *Colletotrichum* que possuem genomas sequenciados disponíveis. O total de candidatos a efetores encontrados foi: de 247 para *C. graminicola*; 288 para *C. sublineola*; 348 para *C. higginsianum*; 365 para *C. salicis*; 372 para *C. simmondsii*; 383 para *C. fioriniae* e *C. nymphaea*; 404 para *C. fructicola* e 446 para *C. orbiculare*. O número de proteínas candidatas a efetoras contendo NLS preditas variaram de 16 em *C. nymphaea* até 32 em *C. orbiculare* (Tabelas S4 a S12). Para as sequências contendo RCPs preditas, o número total variou de 53 em *C. graminicola* até 70 para *C. orbiculare* e *C. sublineola*. Para as sequências ricas em cisteína, foram encontradas 105 em *C. graminicola* até 262 em *C. orbiculare*. O número total de proteínas contendo as NLS, RCPs e ricas em cisteína, considerando as redundâncias (uma proteína pode conter mais de uma característica), variou de 171 em *C. graminicola* até 364 em *C. orbiculare*, e sem as redundâncias o número total variou de 151 para *C. graminicola* até 311 para *C. orbiculare* (Tabelas 2, Tabela S4 a Tabela S12).

**Tabela 2.** Candidatas a efetoras preditas nas espécies de *Colletotrichum* contendo as três características comuns à proteínas efetoras já estudadas em outros fungos e oomicetos.

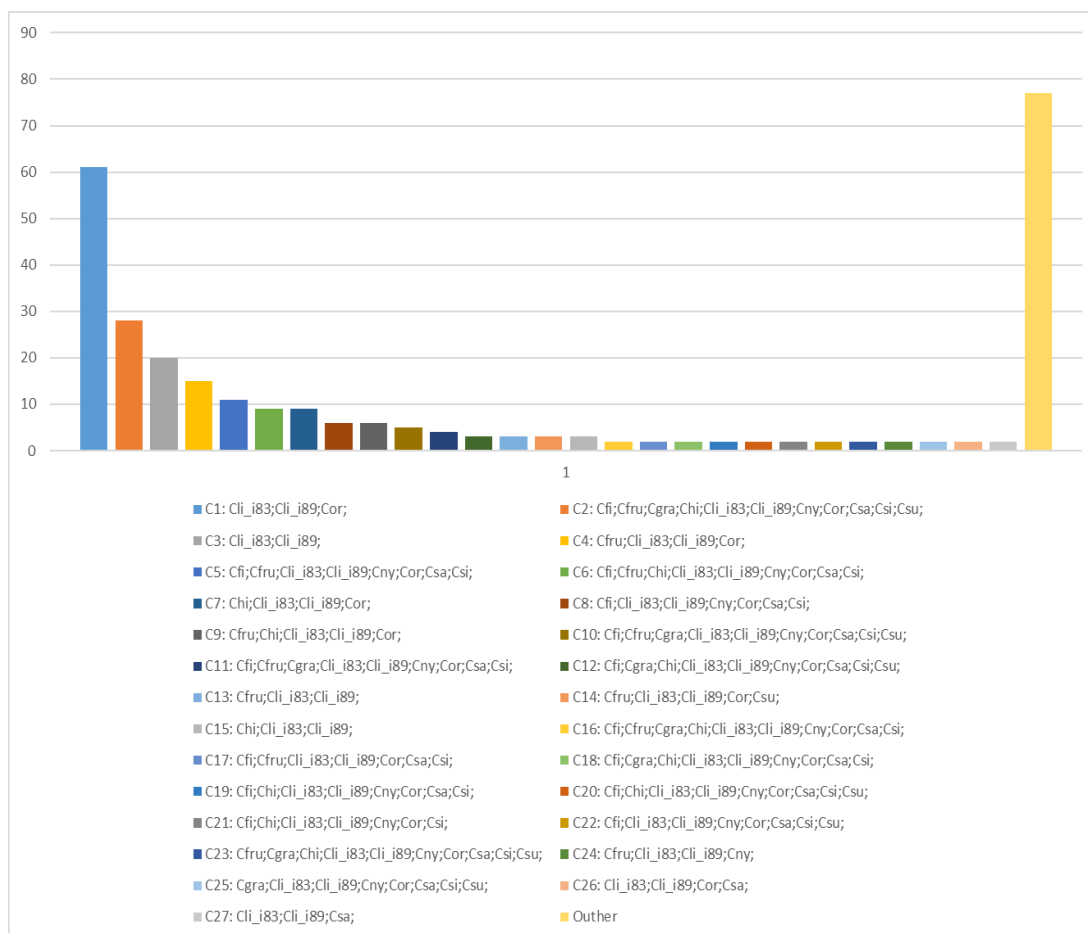
Espécie	N.CE	%/Gen*	NLS	RCP	Cys >3%	Redundante	Não-redundante (%)
<i>C. graminicola</i>	247	2	13	53	105	171	151(61,1)
<i>C. sublineola</i>	288	2,2	19	70	125	214	175(60,8)
<i>C. lindemuthianum</i> i83	324	2,7	24	65	168	257	219(67,6)
<i>C. lindemuthianum</i> i89	324	2,7	24	64	167	255	217(67,0)
<i>C. higginsianum</i>	348	2,1	27	62	178	267	227(65,2)
<i>C. salicis</i>	365	2,6	19	66	160	245	216(59,2)
<i>C. simmondsii</i>	372	2,6	21	58	185	264	239(64,2)
<i>C. fioriniae</i>	383	2,8	20	64	201	285	255(66,6)



<i>C. nymphaeae</i>	383	2,6	16	66	200	282	254(66,3)
<i>C. fructicola</i>	404	2,6	22	65	237	324	282(69,8)
<i>C. orbiculare</i>	446	3,3	32	70	262	364	311(69,7)

\* Porcentagem de candidatas a efetoras em relação ao genoma. **N.CE**: número de candidatas a efetoras; **NLS**: sinal de localização nuclear; **RCP**: proteína contendo repetições.

Foi analisada a presença de homólogos a candidatos a efetores entre as dez espécies de *Colletotrichum* por meio do CD-HIT com limite de corte com identidade de 30%. Foi identificado um total de 1.116 clusters (Tabela S13). Em razão do grande número de *clusters* e pelo fato de que 44% deles serem representados por uma única sequência, foi realizada uma comparação somente com as sequências de candidatos a efetores de *C. lindemuthianum* que possuem sequências homólogas nas outras nove espécies de *Colletotrichum*. Sessenta e um *clusters* de CIECs (Candidatos a efetores em *Colletotrichum lindemuthianum*) possuem homólogos em *C. orbiculare*; 28 *clusters* CIECs com mais de 30% de identidade foram encontrados em todas as 10 espécies; três *cluster* estão presente em *C. lindemuthianum* e *C. higginsianum*; três *cluster* são compartilhados com *C. fructicola* e dois *cluster* com *C. salicis* (Figura 2).

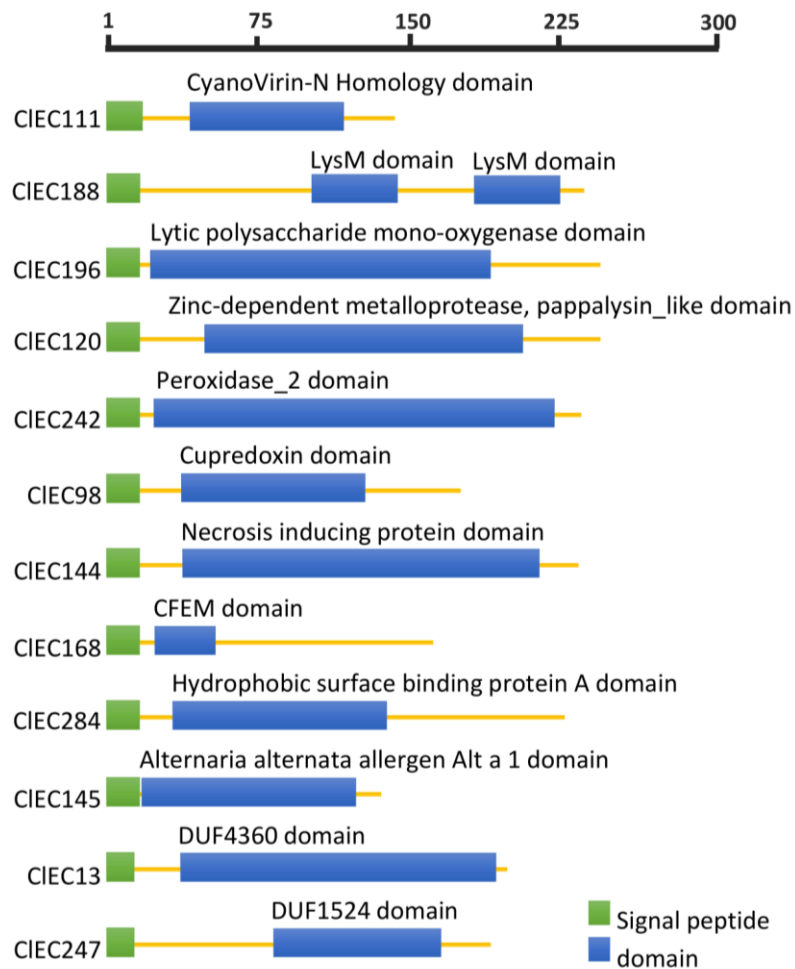


**Figura 2.** Efetores de *Colletotrichum lindemuthianum* presentes nas outras espécies de *Colletotrichum*. Outros: Agrupamentos de espécies de *Colletotrichum* que possuem um único efector homólogo compartilhado, ou seja, foram plotados no gráfico como *cluster* somente grupos que possuem mais que duas sequências por grupo. *Clusters* que possuem sequências presentes em apenas uma espécie foram considerados como outros. Cli\_i83: *Colletotrichum lindemuthianum* isolado 83.501; Cli\_i89: *C. lindemuthianum* isolado 89 A<sub>2</sub> 2-3; Cfi: *C. fioriniae*; Cfru: *C. fructicola*; Cgra: *C. graminicola*; Chi: *C. higginsianum*; Cny: *C. nymphaea*; Cor: *C. orbiculare*; Csa: *C. salicis*; Csi: *simmondsii*; Csu: *C. sublineola*.

### **Domínios conservados encontrados em proteínas candidatas a efectoras em *Colletotrichum* spp.**

Foi realizada a predição da presença de domínios conservados em sequências de aminoácidos de proteínas candidatas a efectoras para todas as espécies estudadas neste trabalho. O total de candidatas a efectoras

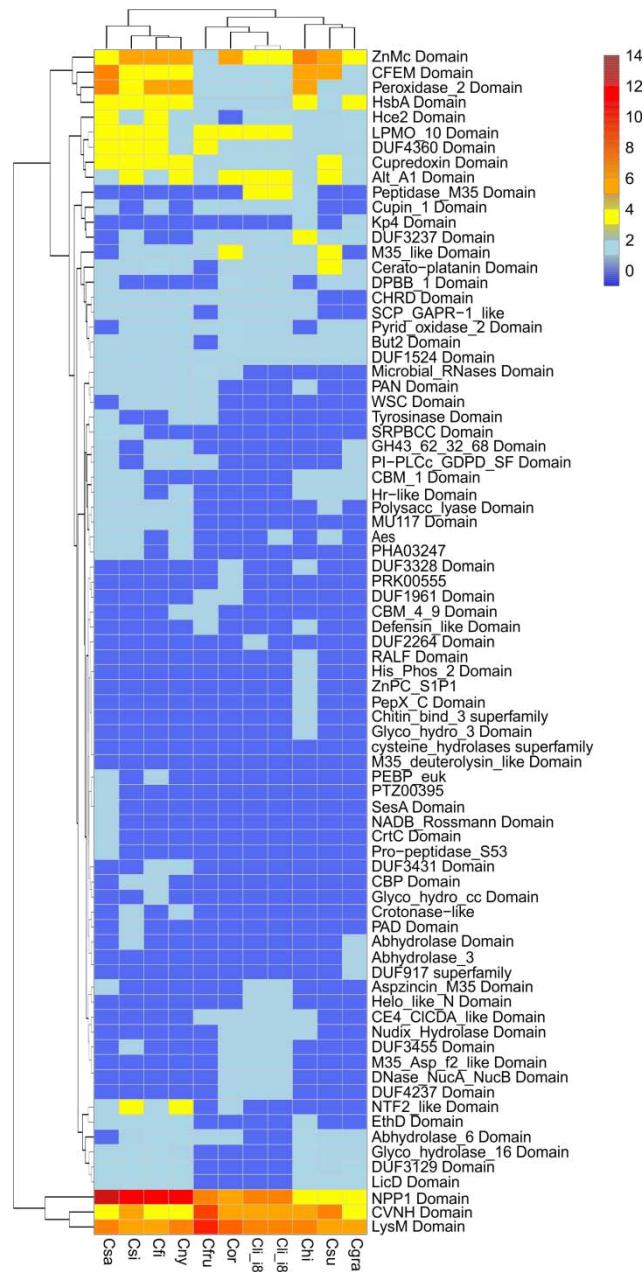
contendo domínios conservados preditos foi de 53 para *C. graminicola* até 95 para *C. salicis* (Tabelas S4 a Tabela S12; Arquivo adicional: Tabela S2). Doze domínios conservados foram preditos nas proteínas candidatas a efetoras de todas as espécies de *Colletotrichum* aqui analisadas (Figura 3). Por outro lado, diversos domínios foram encontrados somente em duas espécies e também como único dentre as espécies (Tabela S2).



**Figura 3.** Domínios conservados (12) que estão presentes em proteínas candidatas à efectoras em todas dez espécies de *Colletotrichum*. Sequências de proteínas candidatas à efectoras de *C. lindemuthianum* foram selecionadas para representar os 12 domínios.

Foi realizado um agrupamento hierárquico com base na presença de todos os domínios conservados das sequências candidatas a efectoras das dez espécies de *Colletotrichum* para determinar se havia alguma relação entre presença destes domínios e clados filogenéticos de cada espécie ou

com a capacidade da espécie infectar diversos hospedeiros. No dendrograma formado houve um agrupamento das espécies de acordo com as suas relações filogenéticas (Figura 4). Assim, foi obtido o grupo com as espécies pertencentes ao clado acutatum (*C. salicis*; *C. simmondsii*; *C. nymphaea* e *C. fiorinae*); o grupo com as espécies do clado orbiculare (*C. lindemuthinaum* e *C. orbiculare*) e o grupo do clado graminícola e destructivum (*C. gramicola* e *C. sublineola* e *C. higginsianum*) que é filogeneticamente próximo ao clado graminícola. A exceção foi para *C. fructicola* que é filogeneticamente mais próximo do clado acutatum e ficou posicionado no grupo do clado orbiculare.



**Figura 4.** Agrupamento hierárquico de proteínas candidatas a efetoras em *Colletotrichum* spp. que apresentam domínios conservados. O agrupamento hierárquico foi realizado usando o programa R com o pacote pheatmap 1.0.8. Cor azul escuro representa ausência de proteínas com domínios conservados, azul claro até amarelo representam de 1 a 3, laranja até ao marrom de 4 a 13 domínios.

**Candidatos a efetores em *Colletotrichum* spp. apresentam termos funcionais do Gene Ontology**

As sequências de candidatas à efetoras foram também funcionalmente anotadas por atribuição de termos GO usando o Blast2GO (Figura 5, Tabela S14). O total de proteínas com termos GO atribuído foi de 90 em *C. graminicola*; 93 em *C. sublineola*; 111 em *C. fructicola*; 112 em *C. orbiculare*; 113 em *C. fiorinae*; 115 em *C. simmondsii*; 116 em *C. nymphaea*; 120 em *C. higginsianum*; 122 em *C. salicis* e 98 e 100 em *C. lindemuthianum* nos isolados 83.501 e 89 A<sub>2</sub> 2-3, respectivamente. Para a categoria de processos biológicos foram atribuídos 21 termos GO, já o total de sequências atribuídas a esses termos variou de 46 em *C. sublineola* até 111 em *C. higginsianum* (Tabela S14M). Entre esses termos foram comuns para todos: processo metabólico primário, processos biológicos de substâncias orgânicas e resposta ao estresse. O termo patogênese foi atribuído somente para *C. graminicola*. Os termos de resposta celular ao estímulo, que abrange proteínas transportadoras e proteínas envolvidas com a sinalização celular e de processos catabólicos foram encontrados somente em *C. higginsianum* com quatro sequências de aminoácidos.

Na categoria de função molecular foram atribuídos 13 termos GO e o total de sequências em cada espécie com estes termos variou de 31 em *C. sublineola* até 57 em *C. higginsianum* (Tabela S14N). Os termos GO que foram comuns a todas as espécies foram de atividade de hidrolases, ligação a íons, atividade de oxidoreductase, ligações a compostos heterocíclicos e ligação de compostos orgânicos cíclicos. Os termos GO de ligação a cofator (proteínas de ligação a coenzimas como dinucleotídeo de nicotinamida e adenina [NAD]) e atividade de liase foram encontrados somente em *C. fructicola* e *C. salicis*, respectivamente. O termo GO de atividade de transportador transmembrana, tais como proteínas que atuam formando canais de íons e o termo de atividade de transportador de substrato específico, foram atribuídos somente em *C. graminicola*, ambos os termos com duas sequências de proteínas.

Na categoria de componente celular, foram atribuídos dez termos GO e o total de sequências atribuídas a esses termos variou de 13 em *C. sublineola* até 39 em *C. nymphaea* (Tabela S14O). O termo de componente celular que foi comum a todos foi somente o de componente intrínseco de membrana, que englobam, por exemplo, proteínas de membrana plasmática

e também de membrana de organelas, a distribuição dessas sequências de aminoácidos nas espécies, variou de seis em *C. sublineola* até 15 em *C. nymphaea*.

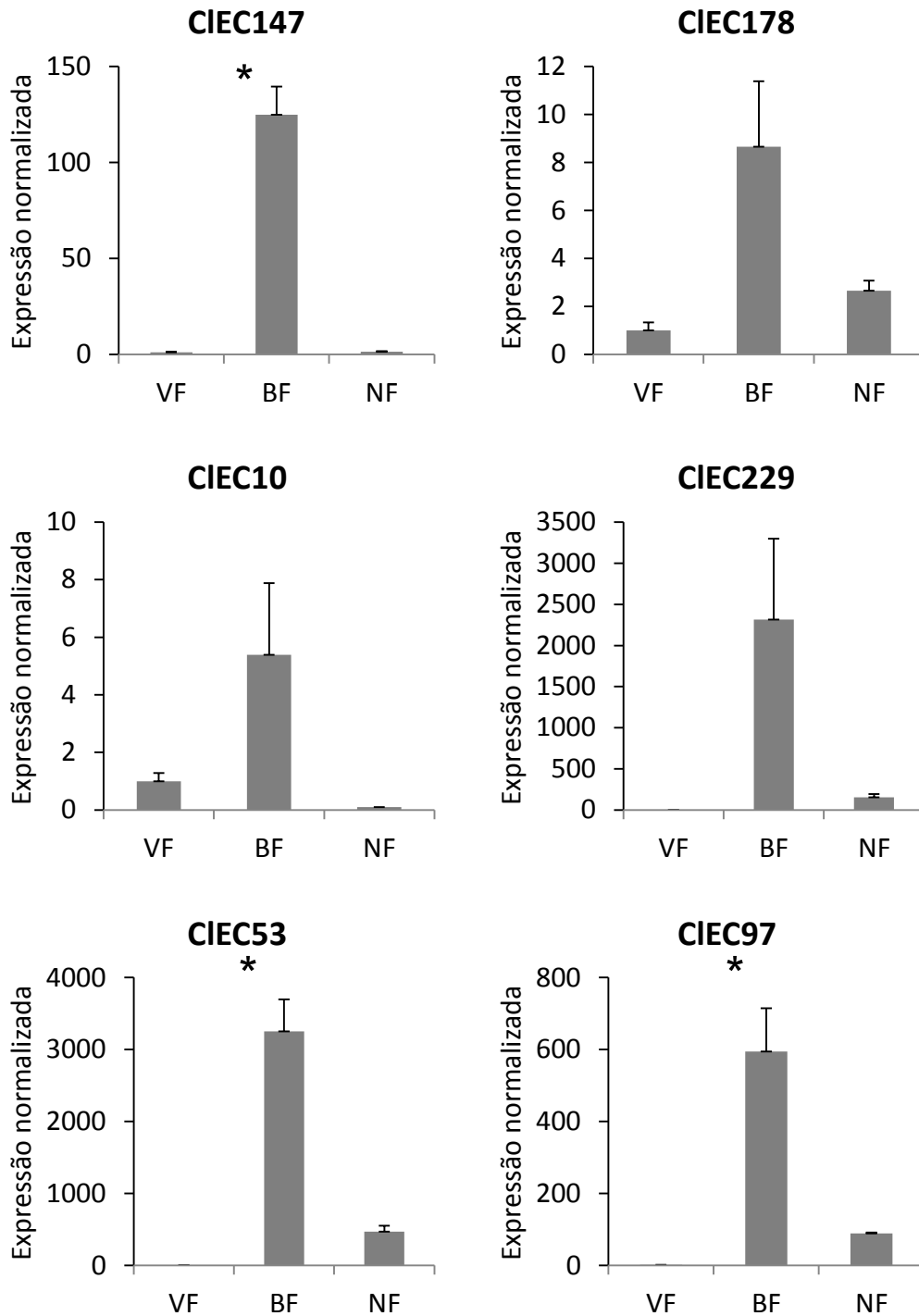


**Figura 5.** Anotação funcional (Gene Ontology) para as proteínas candidatas à efetores nas espécies de *Colletotrichum*. O eixo y representa o número de proteínas para cada termo GO funcional atribuído. O eixo x representa todos os termos GO em cada espécie de *Colletotrichum*.

## **Genes candidatos a efetores de *C. lindemuthianum* apresentam expressão induzida *in planta***

Para testar se as proteínas candidatas a efetoras aqui preditas para *C. lindemuthianum* apresentariam papel durante a interação com o hospedeiro, foram selecionadas seis sequências usando como suporte o transcriptoma de *C. lindemuthianum* durante a interação com o feijoeiro (dados não mostrados) nos quais os genes candidatos a efetores foram mapeados. Foram analisados seis candidatos a efetores expressos diferencialmente durante a fase biotrófica, três sem domínio conservado (CIEC10, CIEC97 e CIEC178) e três apresentando domínios conservados, CIEC53 (DUF4360), CIEC147 (Peptidase\_M35) e CIEC229 (DPBB\_1) em suas sequências de aminoácidos. Com base em nossos critérios de predição de candidatos a efetores, CIEC97 e CIEC178 possuem homólogos somente entre as espécies de *Colletotrichum*, enquanto que CIEC10, CIEC53, CIEC147 e CIEC229 possuem homólogos em diversas espécies de fungos patógenos de plantas. Os seis candidatos a efetores tiveram um padrão de expressão semelhante, sendo que três deles (CIEC53, CIEC97 e CIEC147) estatisticamente foram diferencialmente expressos *in planta* durante a fase biotrófica (Figura 6). CIEC53 apresentou expressão maior que 3 mil vezes, CIEC97 foi expresso cerca de 600 vezes e CIEC147 cerca de 120 vezes, todos durante a fase biotrófica e em comparação às amostras obtidas do crescimento em meio de cultura. Quatro CIECs (CIEC178, CIEC 229, CIEC53 e CIEC97) apresentaram indução durante a fase necrotífica, porém, essa expressão foi menor em comparação a fase biotrófica.





**Figura 6.** Expressão relativa de genes CIECs (CIEC10, CIEC97, CIEC178, CIEC53 e CIEC229) em *Phaseolus vulgaris*. VF (fase vegetativa), BF (fase biotrófica) e NF (fase necrotrófica). Expressão gênica relativa ao gene que codifica actina. Dados apresentados com média de desvio padrão de (n = 3). O asterisco representa amostras que foram significativamente diferentes da fase vegetativa ( $p < 0.05$  pelo teste de Dunnett).

## DISCUSSÃO

Neste trabalho foi realizada a predição de candidatos a efetores em dois isolados de *C. linduthianum* pertencentes as raças 83 e 89, que resultou em 324 CIECs em ambos os isolados. As análises revelaram que os isolados possuem um repertório de candidatos a efetores muito similar. Foi encontrado que 64 CIECs possuem divergências no conteúdo de nucleotídeos, com identidade igual ou menor que 99,96%, os restantes possuem 100% de identidade. Apesar disso, os domínios presentes em algumas das sequências de aminoácidos entre os 64 CIECs estão conservados nos dois isolados. Assim, é possível que estes domínios sejam funcionais nos CIECs de ambas as raças, contudo, também é possível que as divergências fora da região dos domínios conservados afetem a interação com outras moléculas, como por exemplo a interação com proteínas de resistência da planta. Assim, é possível que as diferenças entre os dois isolados sejam mais em razão da variabilidade das sequências dos genes CIECs e não pela presença de genes específicos em cada isolado.

A análise de transcrição de seis genes que codificam proteínas candidatas a efetoras demonstrou que os seis genes apresentaram um mesmo padrão de transcrição, durante a fase biotrófica indicando que a metodologia usada neste trabalho de predição de candidatos a efetores em *C. lindemuthianum* foi eficaz. O gene candidato a efetor apresentou expressão maior que 3 mil vezes durante a fase biotrófica em comparação com o crescimento *in vitro*. Sua sequência de aminoácido possui um domínio conservado ainda de função desconhecida, DUF4360 (*domain of unknown function*). No transcriptoma do fungo ectomicorrizico *Hebeloma cylindrosporum* durante a interação com *Pinus pinaster*, uma pequena proteína secretada contendo este domínio também foi relatada, e seu gene teve a expressão induzida, porém os autores não discutem qualquer função putativa para esta proteína (Doré et al., 2017). Foi também observado que as sequências de aminoácidos de proteínas candidatas a efetoras contendo DUF4360 são conservadas em todas as espécies de *Colletotrichum* aqui analisadas, indicando que esta proteína é potencialmente requerida para biotrofia em *Colletotrichum* spp. Neste trabalho, é demonstrado pela primeira

vez em fungos fitopatogênicos que proteína contendo o domínio DUF4360 tem a expressão induzida *in planta*.

O candidato a efetor CIEC147 também foi altamente induzido *in planta*. A sua sequência de aminoácidos possui homologia a Avr-Pita de *M. oryzae* (Jia et al., 2000). Avr-Pita ativa a resposta de defesa contra o fungo mediada pela proteína de resistência Pi-ta LRD do arroz. Para nossas análises de expressão, nós utilizamos uma linhagem suscetível de feijoeiro (variedade Rosinha). Futuras análises como inoculação de linhagens de *C. lindemuthianum* com deleção do homólogo de Avr-Pita (CIEC147) em linhagens resistentes de feijoeiro poderão ser realizadas para verificar se haverá ganho de virulência e em linhagens suscetíveis de feijoeiro para observar se haverá perda de virulência. Isto poderá demonstrar se CIEC147 também atua como proteína de avirulência em plantas resistentes e como fator de virulência em plantas suscetíveis. Outro gene CIEC que teve a expressão induzida na planta durante a fase biotrófica foi CIEC97, contudo nenhum domínio conservado foi predito para a sequência de aminoácidos e nenhuma similaridade foi observada com algum fator de virulência em outros patógenos. O mesmo foi observado para os genes CIEC10 e CIEC178. Os três genes possuem sequências homólogas em diversas espécies de *Colletotrichum*, mas em todas as espécies esses genes foram anotados como codificando proteínas não caracterizadas. Os resultados indicam que em *C. lindemuthianum* CIEC97, CIEC10 e CIEC178 podem estar associados a biotrofia. Os resultados de análise de expressão demonstram que apesar de alguns autores considerarem como característica de efetor a ausência de domínios conservados, essa não é uma estratégia confiável visto que três genes codificadores de CIECs possuem domínios conservados e seus genes foram altamente expressos *in planta*.

Análises funcionais por meio do *Gene Ontology* (GO) revelaram que para vários dos candidatos a efetores de *Colletotrichum* spp. foram atribuídas às três categorias: para processos biológicos, onde os termos mais abundantes em todas as espécies foram para processo metabólico primário, processos biológicos de substâncias orgânicas; para função molecular, os termos mais abundantes foram os de atividade de hidrolases e ligação à íons; para componente celular, o termo de componente intrínseco

de membrana foi o mais abundante. Isso indica que os candidatos a efetores de *Colletotrichum* spp. estão envolvidos principalmente na interação com estruturas (ex. membranas) e metabolismo da célula hospedeira. Isso está de acordo com o conceito de efetor que o define como moléculas que alteram a estrutura e função da célula hospedeira, facilitando a infecção, e/ou desencadeando a resposta de defesa (Selin *et al.*, 2016), bem como, demonstra que nossa estratégia de predição de efetores foi eficiente.

Muitas proteínas efetoras de fungos e oomicetos fitopatogênicos possuem várias características comuns: sinal de secreção localizada na região N-terminal; sinal de localização nuclear (NLS), a maioria é rica em resíduos de cisteína e também são proteínas contendo repetições (RCPs) de resíduos de aminoácidos. Porém, não necessariamente uma proteína efetora possui as três características. As repetições encontradas em proteínas possuem diversas funções biológicas como direcionar a localização do efetor e intermediar a interação com outras moléculas como RNA, DNA, proteínas ou carboidratos (Mesarich *et al.*, 2015). Neste trabalho foram relatadas pela primeira vez várias RCPs em todas as espécies de *Colletotrichum* aqui estudadas. Além disso, por exemplo, 16 (66%) das NLS preditas em *C. lindemuthianum* também possuem repetições características de RCPs, consistente com o papel destas regiões na localização do efetor no seu alvo na célula hospedeira. Os domínios de repetições podem também contribuir para a evolução adaptativa do efetor, por meio de mudanças na ordem e número de repetições por meio de deleções e inserções (Mesarich *et al.*, 2015). Isto pode fornecer um recurso de diversidade funcional e flexibilidade para evitar o seu reconhecimento pelo sistema imune dos hospedeiros.

Diversas candidatas a efetoras contendo possíveis NLSs foram identificadas em todas as espécies de *Colletotrichum* aqui analisadas, indicando que estes tipos de proteínas efetoras estão amplamente difundidas neste gênero e que podem interferir no mecanismo de transcrição ou em componentes regulatórios nucleares da planta para beneficiar estes patógenos durante a infecção. Considerando que os processos que ocorrem no núcleo da célula exercem funções centrais na sua fisiologia, não é surpresa que o fungo tente atingi-lo usando seus efetores durante o

processo de infecção. Várias proteínas efetoras de bactérias e oomicetos são importadas para o núcleo usando a maquinaria da célula hospedeira, aonde elas podem interagir com componentes e/ou funções essenciais para o sistema de defesa da planta (Hicks and Galán, 2013; Schornack et al., 2010). Para fungos fitopatogênicos, as proteínas contendo NLS, têm sido identificadas e estudadas em *Uromyces fabae* (Kemen, 2005), *Fusarium oxysporum* (Ma et al., 2013), *M. larici-populina* (Petre et al., 2015), *U. maydis* (Redkar et al., 2015) e *C. graminicola* (Vargas et al., 2016).

Durante as análises *in silico*, a exclusão de proteínas contendo domínios conservados cujo seu envolvimento com a patogenicidade ainda não foi relatado, pode acarretar a exclusão de um novo efector. De fato, neste trabalho é demonstrado que das 142 sequências de efetores comprovadas experimentalmente, 59 possuem domínios conservados. Em relação à *Colletotrichum* spp., foram identificados diversos domínios conservados nas candidatas a efetoras de todas as espécies aqui analisadas. Além disso, alguns destes domínios foram encontrados em efetoras confirmadas experimentalmente e em candidatas de outros micro-organismos fitopatogênicos.

Alguns estudos têm mostrado um número maior de enzimas secretadas em patógenos quando comparado com não patógenos e em fungos necrotróficos e hemibiotróficos quando comparado com fungos biotróficos, o que sugere que patógenos podem ser associados mais fortemente de acordo com o seu estilo de vida do que com suas relações filogenéticas (Gan et al., 2016; Zhao et al., 2013). O agrupamento hierárquico com o número de proteínas candidatas a efetoras contendo domínios conservados revelou que diferente dos estudos prévios supracitados, no caso de candidatas a efetoras houve agrupamento de acordo com suas relações filogenéticas. Isto indica que a expansão ou perda dos genes dessas proteínas está mais relacionada com sua relação filogenética do que a especificidade de hospedeiro.

Entre as proteínas candidatas a efetoras de *C. lindemuthianum* foram encontrados doze domínios que são conservados em candidatas a efetoras de todas as espécies de *Colletotrichum* analisadas. Entre eles, o domínio NPP1 que é encontrado em proteínas de bactérias, oomicetos e fungos,

incluindo os patógenos de planta, como a proteína efetora NPP1 de *Phytophthora parasitica* (Fellbrich *et al.*, 2002), PsojNIP de *Phytophthora sojae* (Qutob *et al.*, 2002) e MbNPP1 do fungo *Marssonina brunnea* (Cheng *et al.*, 2014). Tem sido proposto que este domínio está envolvido na indução da morte celular da planta hospedeira. Os mecanismos pelos quais a morte celular é manipulada por estas proteínas ainda são pouco conhecidos. Ainda assim, tem sido proposto que as proteínas efetoras com domínio NPP1 (Santhanam *et al.*, 2013; Schouten *et al.*, 2008) agem como toxinas formadoras de poros na membrana celular ou nuclear, que causaria a despolarização da membrana levando a morte celular (Santhanam *et al.*, 2013; Schouten *et al.*, 2008). Contudo estas proteínas não possuem região de ligação à membrana, assim, estes tipos de efetoras poderiam se ligar a uma proteína da planta como uma lectina para interagir com a membrana. Ottmann *et al.* (2009) também observaram no oomiceto *Pythium aphanidermatum* que efetora contendo o domínio NPP1 causa danos a membrana celular, e isto poderia liberar padrões moleculares associados ao dano, capazes de desencadear a ativação do sistema de defesa da planta induzindo a morte celular. Contudo, Oome *et al.* (2014), demonstraram que o dano celular causado por NPP1 do oomiceto *Hyaloperonospora arabidopsidis* não é requerido para a ativação da defesa da planta induzida por NPP1 em *Arabidopsis*, e que a ativação da resposta de defesa pode ser por um mecanismo diferente. Considerando que o domínio NPP1 é conservado em diversas candidatas a efetoras em *Colletotrichum* spp., estas proteínas poderiam atuar de forma semelhante, desestabilizando a polaridade da membrana que seria suficiente para levar a morte celular da célula hospedeira, durante a fase necrotrófica.

Proteases secretadas são essenciais para degradação da parede celular da planta ou outros componentes orgânicos por fungos fitopatogênicos e saprofíticos (Xu *et al.*, 2016), adicionalmente para os fitopatógenos, estas também podem suprimir ou ativar a defesa do hospedeiro (Jia *et al.*, 2000; Takahashi *et al.*, 2010). Futuros estudos podem determinar se proteínas efetoras contendo domínios de proteases da família M35 em *Colletotrichum* spp. atuam como proteínas de avirulência ou como

manipuladoras da defesa hospedeira, assim como tem sido demonstrado em *M. oryzae* (Jia *et al.*, 2000) e *C. graminicola* (Sanz-Martín *et al.*, 2016).

O LPMO\_10 da família glicosil hidrolase 61 está envolvido na ligação e degradação de celulose. Na proteína efetora MoCDIP4 de *M. oryzae* (Chen *et al.*, 2013), qual é uma glicosil hidrolase da família 61, induz a morte celular no arroz, porém, como MoCDIP4 atua, ainda não foi esclarecido. Como MoCDIP4 é conservado em algumas candidatas a efetoras de *Colletotrichum* spp., há a possibilidade dessas proteínas atuarem como enzimas líticas para auxiliar na penetração do fungo na parede celular da planta, ou induzir a morte celular da planta para que o fungo entre na fase necrotrófica.

O domínio CFEM tem sido encontrado em pequenas proteínas de diversos secretomas de patógenos de plantas (Guyon *et al.*, 2014; Lu and Edwards, 2016; Saunders *et al.*, 2012). Este domínio foi identificado primeiramente em *M. oryzae* (Guyon *et al.*, 2014) na proteína ACI1, que também possui uma âncora GPI. Para esta proteína foi proposto apresentar um papel importante na patogenicidade de *M. oryzae*, no qual interage com uma adenilato ciclase, que produz cAMP, que tem um importante papel na regulação do desenvolvimento do apressório deste fungo. Em plantas, cAMP apresenta diversos papéis em vias de transdução de sinais, incluindo sinalização ao ataque de patógenos (Gehring, 2010). De fato, em *Arabidopsis*, adenilato ciclase (At3g14460) é uma proteína de resistência da família NBS-LRR que detecta a invasão do patógeno levando a ativação da defesa da planta e morte celular (Deyoung and Innes, 2006). Uma vez que as sequências de proteínas candidatas à efetoras em *Colletotrichum* spp. contendo o domínio CFEM não tem âncoras de GPI, que as ligariam a parede celular do fungo, estas proteínas poderiam atuar de forma diferente de ACI1, contudo, ainda podendo interagir com a adenilato ciclase da planta, interferindo na sinalização celular, para que de alguma forma o fungo desarme a defesa para desenvolver a biotrofia ou que propositalmente ative a defesa do hospedeiro para entrar na fase necrotrófica.

O domínio CVNH (Cyanovirin-N Homology) é encontrado em diversos fungos (Percudani *et al.*, 2005), incluindo em candidatas a efetoras do fitopatógeno *Sclerotinia sclerotiorum* e em efetores de *M. oryzae* (Guyon *et*

*al.*, 2014; Koharudin *et al.*, 2015). Este domínio é homólogo ao domínio CVN (cyanovirin-N), que é um membro do grupo de proteínas lectina, as quais são encontradas em muitos organismos eucariotos, e são importantes componentes do complexo de interação proteínas-carboídratos (Percudani *et al.*, 2005). A caracterização desse domínio em proteínas efetoras tem sido realizada em MoCVNH3 de *M. oryzae* (Koharudin *et al.*, 2015), para qual foi proposta a ligação a oligômeros de quitina para evitar que esses atuem como PAMP. De fato, o domínio CVNH desta proteína é do tipo III, que possuem um domínio LysM adicional. Nas proteínas candidatas a efetoras de *Colletotrichum* spp. contendo CVNH foi identificado somente um domínio em cada proteína, ao que parece estar mais relacionado aos CVNH tipo I, mas estes, porém, apresentam domínios CVNH sequenciais. Considerando que essas proteínas em *Colletotrichum* spp. são provavelmente secretadas, mesmo não possuindo o domínio adicional LysM que se liga a quitina, e que é uma característica de CVNH interagir com carboidratos, é possível que estas proteínas contendo o domínio CVNH em *Colletotrichum* spp. se liguem à oligossacarídeos da planta para impedi-los de atuar como PAMPs para desencadear a resposta de defesa.

Foi identificado o domínio Peroxidase\_2 também conhecido como cloroperoxidase ou heme haloperoxidase, em proteínas candidatas a efetoras de todas as dez espécies de *Colletotrichum* aqui estudadas. Cloroperoxidasas funcionam como catalases, facilitando a decomposição de H<sub>2</sub>O<sub>2</sub> em oxigênio e água (Hofrichter and Ullrich, 2006). Especula-se que cloroperoxidasas permitem ao fungo sobrepor as defesas iniciais da planta durante a infecção e/ou permitem a hifa tolerar à toxicidade do H<sub>2</sub>O<sub>2</sub> e o estresse oxidativo que ocorre nas folhas durante a atividade fotossintética (do Amaral *et al.*, 2012).

Em *Aspergillus niger*, a expressão de proteínas contendo o domínio HsbA foi fortemente induzida quando a fonte de carbono (glicose) foi trocada por palha de trigo, sugerindo que esta proteína apresenta papel importante no recrutamento de enzimas que degradam palha do trigo (Delmas *et al.*, 2012). Proteínas candidatas a efetoras com HsbA em *Colletotrichum* spp., podem estar envolvidos no recrutamento de enzimas que degradariam ou modificariam moléculas hidrofóbicas da planta interferindo no seu



metabolismo, ou poderiam de alguma forma auxiliar as enzimas líticas como cutinases durante o processo de penetração do fungo na planta.

Para proteínas secretadas de outros fungos, o domínio cupredoxina até o momento, foi relatado somente em *P. striiformis* f.sp. *tritici* (Garnica *et al.*, 2013), porém nenhuma função putativa foi atribuída para esta proteína neste fungo. Foi realizada uma análise usando a ferramenta BLASTp no banco de dados UniProt, usando como sequência *query* proteínas candidatas a efetoras de *Colletotrichum* spp. contendo o domínio Cupredoxina, e neste trabalho foram identificadas em diversos fungos fitopatogênicos e em patógenos de insetos e de animais, como por exemplo, *F. oxysporum*, *Gibberella fujikuroi*, *Metarhizium guizhouense* e *Stachybotrys chartarum* (dados não mostrados). Tem sido reportado que a proteína azurina, que é uma cupredoxina na bactéria *Pseudomonas aeruginosa*, é secretada e tem função de transportador de elétrons, ao qual forma um complexo com a proteína supressora de tumor p53 (um regulador transcricional), que a transporta para o núcleo de macrófagos de ratos. Esta interação gerou altos níveis de ROS, sendo importante para citotoxicidade de auzerina, que leva a morte celular dos macrófagos (Chakrabarty, 2016; Punj *et al.*, 2003; Punj *et al.*, 2004). Em *Colletotrichum* spp. estas proteínas poderiam atuar de forma semelhante induzindo a morte celular em suas plantas hospedeiras, por meio da interação com alguma molécula que também as conduziriam para o núcleo para manipular a regulação transcricional de genes envolvidos na morte celular ou mesmo exerceriam sua atividade citotóxica no citoplasma ou em outras localizações sub-celulares.

O domínio LysM já foi relatado, por exemplo, nas proteínas efetoras SLP1 de *M. oryzae* (Mentlak *et al.*, 2012), Ecp6 de *C. fulvum* (Bolton *et al.*, 2008), Mg1LysM e Mg3LysM de *M. graminicola* (Marshall *et al.*, 2011) e o recentemente estudado ChELP2 de *C. higginsianum* (Takahara *et al.*, 2016). O interessante é que domínios LysM são encontrados em todas as PRRs de plantas que estão envolvidas no reconhecimento de quitina (Cao *et al.*, 2014; Takahara *et al.*, 2016). Com base nesta evidência e nos experimentos realizados por estes autores, é proposto que estes tipos de proteínas efetoras podem atuar de duas formas, sequestrando os oligômeros de

quitina ou protegendo a parede celular contra ação de quitinases. Uma vez evitando o reconhecimento da quitina, o fungo impede a ativação da resposta de defesa da planta, podendo prosseguir na infecção.

Ainda não se conhece a função dos demais domínios Alt\_A1, DUF4360 e DUF1524, que estão entre os 12 presentes em todas as espécies de *Colletotrichum* aqui analisadas, porém, dada a conservação destes, sugere-se que estas candidatas a efetoras fazem parte de um core de virulência desse gênero.

Além dos 12 domínios conservados, o domínio Hce2 somente não foi identificado nas proteínas candidatas a efetoras de *C. orbiculare*. Foi verificado se alguma sequência de seu genoma seria homóloga a um candidato a efetor de *C. lindemuthianum*, uma vez que pertencem ao mesmo complexo. Foi identificada uma sequência que possui o domínio Hce2, mas não possui o peptídeo sinal predito. Por esta razão não foi incluído como candidata a proteína efetora. O domínio Hce2 tem sido relatado na proteína efetora ECP2 do *C. fulvum*, *Mycosphaerella fijiensis* e *F. graminearum* (Laugé, 1997; Lu and Edwards, 2016; Stergiopoulos *et al.*, 2010) e em diversas outras espécies de fungos (Stergiopoulos *et al.*, 2012). A função intrínseca deste tipo de proteína efetora ainda não é desconhecida. Porém, tem sido proposto que a proteína efetora ECP2 se liga a algum componente de defesa da planta e induz a necrose (Stergiopoulos *et al.*, 2010).

Por outro lado, diversos domínios conservados foram encontrados somente em algumas espécies. Foi identificado uma candidata a efetora somente em *C. lindemuthianum* e *C. orbiculare* contendo o domínio DNase\_NucA\_NucB. Sabe-se que DNA extracelular (exDNA) é o principal componente estrutural de matrizes celulares viscosas localizadas em regiões de infecções em humanos, denominada de complexo de armadilhas extracelulares dos neutrófilos (NET) (Izano *et al.*, 2008; Wartha *et al.*, 2007; Yan *et al.*, 2016). NET inibe a invasão de patógenos, incluindo fungos (Bruns *et al.*, 2010; Sánchez and Colom, 2010). É interessante o fato que exDNA, histonas e outras proteínas foram relatadas como componentes da mucilagem da coifa de ervilha (*Pisum sativum*), e a digestão do exDNA por exDNAse aboliu a resistência da raiz contra o fungo *Fusarium solani* (Wen *et*

*al.*, 2009). Neste trabalho foi realizada uma análise usando a ferramenta BLASTp contra o banco de dados UniProt usando CIEC90 como sequência *query*, e foram encontrados homólogos em diversas espécies de fungos, incluindo patógenos de plantas, entomopatógenos e endofíticos (dados não mostrados) fora do gênero *Colletotrichum*. Isto indica que exDNAses são importantes para estes fungos, pois mantêm-se conservadas. Estudos devem ser feito para analisar se exDNAs estão presentes em regiões extracelulares de células de outras regiões da planta, além da coifa, pois candidatas a efetoras como CIEC90 e sua homóloga em *C. orbiculare* (ENH76220), assim como homólogas em outros fitopatógenos, podem ser importantes para degradar NETs que possam existir em seus hospedeiros para que possam ter sucesso na infecção.

Candidatas a efetoras contendo o domínio *defensin\_like* foram identificadas nas espécies *C. fruticola* e *C. higginsianum*, que possuem uma ampla gama de plantas hospedeiras. *Defensin\_like* é um domínio encontrado amplamente em plantas contendo atividade antifúngica e antibacteriana, sendo importante para a defesa da planta contra esses patógenos (Terras *et al.*, 1995; De Coninck *et al.*, 2013). Também algumas dessas proteínas são inibidoras de enzimas digestivas como  $\alpha$ -amilases e proteases. Tem sido reportado que o fungo *Aspergillus giganteus* secreta um tipo de defensina que inibe as atividades das enzimas quitinase sintases III e V (Hagen *et al.*, 2007).

A proteína contendo o domínio RALF (*rapid alkalinisation fator*) apresenta papéis na regulação e desenvolvimento de raízes de plantas, pelos radiculares, nódulos de leguminosas e tubos de polens (Murphy and De Smet, 2014). A proteína efetora RALF-B de *F. oxysporum* f. sp. *lycopersici* causou inibição do crescimento de plântulas de tomate, rápida explosão oxidativa, alcalinização e ativação de MAPK (Mitogen Activated Protein Kinases) (Thynne *et al.*, 2016). Todos esses mecanismos são características da ativação da defesa de plantas. Contudo, a função intrínseca das proteínas efetoras com domínio RALF em fungos ainda permanece desconhecida. Neste trabalho, candidatas a efetoras com domínio RALF foram encontradas somente em *C. higginsianum*, e futuros trabalhos podem demonstrar se nesta espécie tal efetora também induz a

ativação da resposta de defesa de seus hospedeiros, que pode ser uma estratégia adicional para que esse entre na sua fase necrotrófica.

Os domínios Helo\_like\_N e SesA, foram presentes em candidatas a efetoras de *C. lindemuthianum* e *C. salicis*, respectivamente. Esses domínios pertencem à família de STAND NTPases que estão presentes em proteínas receptoras intracelulares, requeridas para reconhecimento e sinalização na resposta imune de plantas e animais (Leipe *et al.*, 2004; Sarris *et al.*, 2015). Em fungos, proteínas dessa família estão envolvidas no processo de incompatibilidade do heterocário (Daskalov *et al.*, 2012; Dyrka *et al.*, 2014; Thynne *et al.*, 2016). As candidatas a efetoras em que os domínios Helo\_like\_N e SesA foram identificados, apresentam duas características importantes que os diferem daquelas proteínas envolvidas na incompatibilidade do heterocário: são secretadas e não possuem o domínio PFD localizado na região C-terminal, qual é capaz de levar a incompatibilidade.

## **CONSIDERAÇÕES FINAIS**

Neste trabalho, foi realizada uma série de filtragens por meio de bioinformática para identificar genes que codificam proteínas candidatas a efetoras em dez espécies de *Colletotrichum* e analisar estas proteínas. Os dados revelaram diversas sequências codificadoras de proteínas candidatas a efetoras e forneceu uma valiosa fonte de estudo para elucidar os mecanismos de infecção desses patógenos em seus hospedeiros. A maioria das sequências candidatas a efetoras apresentou homologia à sequências não caracterizadas depositadas em bancos de dados de sequências, sugerindo que existem várias proteínas efetoras importantes na infecção dos fungos ainda pouco estudadas. Por outro lado, foram identificadas sequências candidatas a efetoras com presença de domínios conservados. A maioria desses domínios conservados ainda não tinha sido relatado para efetoras de *Colletotrichum* spp. o que permite sugerir papéis putativos para as efetoras aqui analisadas. Neste sentido, este trabalho se torna relevante como auxílio na identificação e estudos de novas proteínas efetoras em

*Colletotrichum* spp., disponibilizando informações importantes que permitam compreender a estratégia de estilo de vida de espécies de *Colletotrichum* que são patógenos de plantas.

## REFERÊNCIAS

- Alzate-Marin, A.L., Gonçalves De Barros, E. and Moreira, M.A.** (1999) Co-evolution model of *Colletotrichum lindemuthianum* (Melanconiaceae, Melanconiales) races that occur in some Brazilian regions. *Genet. Mol. Biol.* **22**, 115–118.
- Asai, S. and Shirasu, K.** (2015) Plant cells under siege: Plant immune system versus pathogen effectors. *Curr. Opin. Plant Biol.* **28**, 1–8.
- Barrus, M. F.** (1918). Varietal susceptibility of beans to strains of *Colletotrichum lindemuthianum* (Sacc. & Magn.) B & C. *Phytopathology*. **8**, 589–605.
- Bechinger, C., Giebel, K., Schnell, M., Leiderer, P., Deising, H. and Bastmeyer, M.** (1999) Optical measurements of invasive forces exerted by appressoria of a plant pathogenic fungus. *Science*, **285**, 1896–1899.
- Bhadauria, V., MacLachlan, R., Pozniak, C. and Banniza S.** (2015) Candidate effectors contribute to race differentiation and virulence of the lentil anthracnose pathogen *Colletotrichum lentis*. *BMC Genomics*, **16**:628.
- Blumke, A., Falter, C., Herrfurth, C., Sode, B., Bode, R., Schafer, W., Feussner, I. and Voigt, C.A.** (2014) Secreted Fungal Effector Lipase Releases Free Fatty Acids to Inhibit Innate Immunity-Related Callose Formation during Wheat Head Infection. *Plant Physiol.* **165**, 346–358.

- Bolton, M.D., Esse, H.P. Van, Vossen, J.H., et al.** (2008) The novel *Cladosporium fulvum* lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species. *Mol. Microbiol.* **69**, 119–136.
- Bruns, S., Kniemeyer, O., Hasenberg, M., et al.** (2010) Production of extracellular traps against *Aspergillus fumigatus* in vitro and in infected lung tissue is dependent on invading neutrophils and influenced by hydrophobin RodA. *PLoS Pathog.* **6**, e1000873.
- Cao, Y.1., Liang Y., Tanaka, K., Nguyen, C.T., Jedrzejczak, R.P., Joachimiak, A., Stacey G.** (2014). The kinase LYK5 is a major chitin receptor in *Arabidopsis* and forms a chitin-induced complex with related kinase CERK1. *Elife*, **3**, doi: 10.7554/eLife.03766.
- Chakrabarty, A.M.** (2016) Bacterial azurin in potential cancer therapy. *Cell Cycle*, **4101**, 12–14.
- Chen, S., Songkumarn, P., Venu, R.C., et al.** (2013) Identification and characterization of in planta-expressed secreted effector proteins from *Magnaporthe oryzae* that induce cell death in rice. *Mol. Plant. Microbe. Interact.* **26**, 191–202.
- Cheng, Q., Wang, H., Xu, B., Zhu, S., Hu, L. and Huang, M.** (2014) Discovery of a novel small secreted protein family with conserved N-terminal IGY motif in Dikarya fungi. *BMC Genomics* **15**, 1151.
- Conesa, A., Götz, S., García-Gómez, JM., Terol, J., Talón, M. and Robles, M.** (2005) Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.
- Coninck, B. De, Cammue, B.P.A. and Thevissen, K.** (2013) Modes of antifungal action and in planta functions of plant defensins and defensin-like peptides. *Fungal Biol. Rev.* **26**, 109–120.
- Crouch, J.A., O'Connell, R., Gan, P., Buiate, E., Torres, M.F., Beirn, L., Shirasu, K., Vaillancourt L.** (2014) The genomics of *Colletotrichum*. In: Genomics of plant-associated fungi: monocot pathogens. Berlin: Springer Verlag. 69-102.

- Daskalov, A., Paoletti, M., Ness, F. and Saupe, S.J.** (2012) Genomic clustering and homology between HET-S and the NWD2 stand protein in various fungal genomes. *PLoS One*, **7**, e34854.
- Dean, R., Kan, J.A.L. Van, Pretorius, Z.A., et al.** (2012) The Top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* **13**, 414–430.
- Delmas, S., Pullan, S.T., Gaddipati, S., et al.** (2012) Uncovering the Genome-Wide Transcriptional Responses of the Filamentous Fungus *Aspergillus niger* to Lignocellulose Using RNA Sequencing. *PLoS Genet.* **8**, e1002875.
- Deyoung, B.J. and Innes, R.W.** (2006) Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat Immunol.* **7**, 1243–1249.
- Dodds, P.N. and Rathjen, J.P.** (2010) Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.* **11**, 539–548.
- Doré, J., Kohler, A., Dubost, A., Hundley, H., Singan, V., Peng, Y., Kuo, A., Grigoriev, IV., Martin, F., Marmeisse, R., Gay, G.** (2017) The ectomycorrhizal basidiomycete *Hebeloma cylindrosporum* undergoes early waves of transcriptional reprogramming prior to symbiotic structures differentiation. *Environ Microbiol.* **19**, 1338-1354.
- Dyrka, W., Lamacchia, M., Durrens, P., Kobe, B., Daskalov, A., Paoletti, M., Sherman, D.J. and Saupe, S.J.** (2014) Diversity and variability of NOD-like receptors in fungi. *Genome Biol. Evol.* **6**, 3137–3158.
- Eisenhaber, B., Schneider, G., Wildpaner, M. and Eisenhaber, F.** (2004) A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for *Aspergillus nidulans*, *Candida albicans*, *Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *J. Mol. Biol.* **337**, 243–253.
- Emanuelsson, Henrik Nielsen, S.B. and G. von H.** (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Bio.* **300**, 1005–1016.

- Emanuelsson, O., Brunak, S., Heijne, G. von and Nielsen, H.** (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**, 953–971.
- Fellbrich, G., Romanski, A., Varet, A., et al.** (2002) NPP1, a *Phytophthora*-associated trigger of plant defense in parsley and *Arabidopsis*. *Plant J.* **32**, 375–390.
- Ferreira, J.J., Campa, a. and Kelly, J.D.** (2013) Organization of genes conferring resistance to anthracnose in common bean. *Transl. Genomics Crop Breed. I*, 151–176.
- Fudal, I., Ross, S., Gout, L., et al.** (2007) Heterochromatin-like regions as ecological niches for avirulence genes in the *Leptosphaeria maculans* genome: map-based cloning of AvrLm6. *Mol. Plant. Microbe. Interact.* **20**, 459–470.
- Gan, P., Ikeda, K., Irieda, H., Narusaka, M., O’Connell, R.J., Narusaka, Y., Takano, Y., Kubo, Y. and Shirasu, K.** (2013) Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytol.* **197**, 1236–1249.
- Gan, P., Narusaka, M., Kumakura, N., Tsushima, A., Takano, Y., Narusaka, Y., Shirasu, K.** (2016) Genus-wide comparative genome analyses of *Colletotrichum* species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. *Genome Biol Evol.* **8**, 1467-1481.
- Garnica, D.P., Upadhyaya, N.M., Dodds, P.N. and Rathjen, J.P.** (2013) Strategies for wheat stripe rust pathogenicity identified by transcriptome sequencing. *PLoS One*, **8**, e67150
- Gehring, C.** (2010) Adenyl cyclases and cAMP in plant signaling - past and present. *Cell Commun. Signal.* **8**, 15.
- Giraldo, M.C. and Valent, B.** (2013) Filamentous plant pathogen effectors in action. *Nat. Rev. Microbiol.* **11**, 800–14.



- Göhre, V. and Robatzek, S.** (2008) Breaking the barriers: microbial effector molecules subvert plant immunity. *Annu. Rev. Phytopathol.* **46**, 189–215.
- Gonzaga, L.L., Costa, L.E.O., Santos, T.T., Araújo, E.F. and Queiroz, M. V.** (2015) Endophytic fungi from the genus *Colletotrichum* are abundant in the *Phaseolus vulgaris* and have high genetic diversity. *J. Appl. Microbiol.* **118**, 485–496.
- González, A.M., Yuste-Lisbona, F.J., Rodiño, a P., Ron, A.M. De, Capel, C., García-Alcázar, M., Lozano, R. and Santalla, M.** (2015) Uncovering the genetic architecture of *Colletotrichum lindemuthianum* resistance through QTL mapping and epistatic interaction analysis in common bean. *Front. Plant Sci.* **6**, 13.
- Guyon, K., Balagué, C., Roby, D. and Raffaele, S.** (2014) Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen *Sclerotinia sclerotiorum*. *BMC Genomics* **15**, 336.
- Hacquard, S., Joly, D.L., Lin, Y.-C., et al.** (2012) A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *Mol. Plant. Microbe. Interact.* **25**, 279–293.
- Hagen, S., Marx, F., Ram, A.F. and Meyer, V.** (2007) The antifungal protein AFP from *Aspergillus giganteus* inhibits chitin synthesis in sensitive fungi. *Appl. Environ. Microbiol.* **73**, 2128–2134.
- Hicks, S.W. and Galán, J.E.** (2013) Exploitation of eukaryotic subcellular targeting mechanisms by bacterial effectors. *Nat. Rev. Microbiol.* **11**, 316–326.
- Hiruma, K., Gerlach, N., Sacristán, S., et al.** (2016) Root Endophyte *Colletotrichum tofieldiae* Confers Plant Fitness Benefits that Are Phosphate Status Dependent. *Cell*, **165**, 464–474.

- Hofrichter, M. and Ullrich, R.** (2006) Heme-thiolate haloperoxidases: versatile biocatalysts with biotechnological and environmental significance. *Appl. Microbiol. Biotechnol.* **71**, 276–288.
- Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. and Nakai, K.** (2007) WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* **35**, 585–587.
- Huang, Y., Niu, B., Gao, Y., Fu, L. and Li, W.** (2010) CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics*, **26**, 680–682.
- Huang, Y.J., Li, Z.Q., Evans, N., Rouxel, T., Fitt, B.D.L. and Balesdent, M.H.** (2006) Fitness cost associated with loss of the AvrLm4 avirulence function in *Leptosphaeria maculans* (phoma stem canker of oilseed rape). *Eur. J. Plant Pathol.* **114**, 77–89.
- Izano, E.A., Amarante, M.A., Kher, W.B. and Kaplan, J.B.** (2008) Differential roles of poly-N-acetylglucosamine surface polysaccharide and extracellular DNA in *Staphylococcus aureus* and *Staphylococcus epidermidis* biofilms. *Appl. Environ. Microbiol.* **74**, 470–476.
- Jia, Y., McAdams, S.A., Bryan, G.T., Hershey, H.P. and Valent, B.** (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO J.* **19**, 4004–4014.
- Jones, J.D.G. and Dangl, J.L.** (2006) The plant immune system. *Nature*, **444**, 323–329.
- Jonge, R., van Esse, H.P., Kombrink., et al.** (2010) Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *Science.* **329**, 953–955.
- Jorda, J. and Kajava, A. V.** (2009) T-REKS: Identification of Tandem REpeats in sequences with a K-meanS based algorithm. *Bioinformatics*, **25**, 2632–2638.
- Käll, L., Krogh, A. and Sonnhammer, E.L.L.** (2007) Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Res.* **35**, 429–432.

- Kan, J.A. van, Ackerveken, G.F.J.M. van den and Wit, P.J.G.M. de** (1991) Cloning and characterization of cDNA of avirulence gene *avr9* of the fungal pathogen *Cladosporium fulvum*, causal agent of tomato leaf mold. *Mol. Plant. Microbe. Interact.* **4**, 52–59.
- Kang, S., Lebrun, M.H., Farrall, L. and Valent, B.** (2001) Gain of virulence caused by insertion of a *Pot3* transposon in a *Magnaporthe grisea* avirulence gene. *Mol. Plant. Microbe. Interact.* **14**, 671–674.
- Kemen, E.** (2005) Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Mol. Plant Microbe Interact.* **18**, 1130–1139.
- Kleemann, J. Rincon-Rivera, L.J., Takahara, H., Neumann, U., Ver Loren van Themaat, E., van der Does, H.C., Hacquard, S., Stüber, K., Will, I., Schmalenbach, W., Schmelzer, E., O'Connell, R.J.** (2012) Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathog.* **8**:e1002643.
- Kombrink, A., Rovenich, H., Shi-Kunne, X., Rojas-Padilla E., van den Berg, G.C., Domazakis, E., de Jonge, R., Valkenburg, D.J., Sánchez-Vallet, A., Seidl MF., Thomma, B.P.** (2016) *Verticillium dahliae* LysM effectors differentially contribute to virulence on plant hosts. *Mol Plant Pathol.* **8**, 596–608.
- Kosugi, S., Hasebe, M., Tomita, M. and Yanagawa, H.** (2009) Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 10171–10176.
- Laugé, R.** (1997) The in planta-produced extracellular proteins ECP1 and ECP2 of *Cladosporium fulvum* are virulence factors. *Mol. Plant. Microbe. Interact.* **10**, 725–734.
- Leipe, D.D., Koonin, E. V. and Aravind, L.** (2004) STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: Multiple, complex domain architectures, unusual phyletic

patterns, and evolution by horizontal gene transfer. *J. Mol. Biol.* **343**, 1–28.

**Lu, S. and Edwards, M.C.** (2016) Genome-Wide Analysis of Small Secreted Cysteine-Rich Proteins Identifies Candidate Effector Proteins Potentially Involved in *Fusarium graminearum*-Wheat Interactions. *Phytopathology*, **106**, 166–176.

**Ma, Z., Song T., Zhu, L., Ye, W., Wang, Y., Shao, Y., Dong, S., Zhang, Z., Dou, D., Zheng, X., Tyler, B.M., Wang, Y.** (2015) A *Phytophthora sojae* Glycoside Hydrolase 12 Protein Is a Major Virulence Factor during Soybean Infection and Is Recognized as a PAMP. *Plant Cell*, **27**, 2057–2072.

**Mahuku, G.S. and Riascos, J.J.** (2004) Virulence and molecular diversity within *Colletotrichum lindemuthianum* isolates from Andean and Mesoamerican bean varieties and regions. *Eur. J. Plant Pathol.* **110**, 253–263.

**Marshall, R., Kombrink, A., Motteram, J., Loza-Reyes, E., Lucas, J., Hammond-Kosack, K.E., Thomma, B.P.H.J. and Rudd, J.J.** (2011) Analysis of two in planta expressed LysM effector homologs from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. *Plant Physiol.* **156**, 756–69.

**Mateos, F. V, Rickauer, M. and Esquerré-Tugayé, M.T.** (1997) Cloning and characterization of a cDNA encoding an elicitor of *Phytophthora parasitica* var. *nicotianae* that shows cellulose-binding and lectin-like activities. *Mol. Plant. Microbe. Interact.* **10**, 1045–1053.

**Mejía, L.C., Herre, E.A., Sparks, J.P., et al.** (2014) Pervasive effects of a dominant foliar endophytic fungus on host genetic and phenotypic expression in a tropical tree. *Front. Microbiol.* **5**, 1–17.

**Mentlak, T.A., Kombrink, A., Shinya, T., et al.** (2012) Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. *Plant Cell*, **24**, 322–335.

- Mesarich, C.H., Bowen, J.K., Hamiaux, C. and Templeton, M.D.** (2015) Repeat-containing protein effectors of plant-associated organisms. *Front. Plant Sci.* **6**, 872.
- do Amaral, A.M., Antoniw, J., Rudd, J.J. and Hammond-Kosack, K.E.** (2012) Defining the Predicted Protein Secretome of the Fungal Wheat Leaf Pathogen *Mycosphaerella graminicola*. *PLoS One*, **7**, 1–19.
- Mosquera, G., Giraldo, M.C., Khang, C.H., Coughlan, S. and Valent, B.** (2009) Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as Biotrophy-associated secreted proteins in rice blast disease. *Plant Cell*, **21**, 1273–1290.
- Mota, S.F., Barcelos, Q.L., Dias, M.A. and Souza, E.A.** (2016) Variability of *Colletotrichum* spp. in common bean. *Genet. Mol. Res.* **15**.
- Motion, G.B., Amaro, T.M.M.M., Kulagina, N. and Huitema, E.** (2015) Nuclear processes associated with plant immunity and pathogen susceptibility. *Brief. Funct. Genomics*, **14**, 243–52.
- Murphy, E. and Smet, I. De** (2014) Understanding the RALF family: A tale of many species. *Trends Plant Sci.* **19**, 664–671.
- Newman, A.M. and Cooper, J.B.** (2007) XSTREAM: a practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. *BMC Bioinformatics*, **8**, 382.
- Nguyen Ba, A.N., Pogoutse, A., Provard, N. and Moses, A.M.** (2009) NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics*, **10**, 202.
- O’Connell, R.J., Thon, M.R., Hacquard, S., et al.** (2012) Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat. Genet.* **44**, 1060–1065.
- Oliveira-Garcia, E. and Valent, B.** (2015) How eukaryotic filamentous pathogens evade plant recognition. *Curr. Opin. Microbiol.* **26**, 92–101.
- Oome, S., Raaymakers, T.M., Cabral, A., Samwel, S., Böhm, H., Albert, I., Nürnbergger, T. and Van den Ackerveken G.** (2014) Nep1-like proteins

from three kingdoms of life act as a microbe-associated molecular pattern in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 16955–60.

**Ottmann, C., Luberacki, B., Kűfner, I., et al.** (2009) A common toxin fold mediates microbial attack and plant defense. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 10359–64.

**Padder, B.A., Sharma, P.N., Sharma, O.P. and Kapoor, V.** (2007) Genetic diversity and gene flow estimates among five populations of *Colletotrichum lindemuthianum* across Himachal Pradesh. *Physiol. Mol. Plant Pathol.* **70**, 8–12.

**Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C.V. and Falquet, L.** (2004) MyHits: A new interactive resource for protein annotation and domain identification. *Nucleic Acids Res.* **32**, 332–335.

**Pareja-Jaime Y., Roncero MI., Ruiz-Roldán MC.** (2008) Tomatinase from *Fusarium oxysporum* f. sp. *lycopersici* is required for full virulence on tomato plants. *Mol. Plant. Microbe. Interact.* **21**, 728–36.

**Percudani, R., Montanini, B. and Ottonello, S.** (2005) The anti-HIV cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes. *Proteins Struct. Funct. Genet.* **60**, 670–678.

**Perfect, S.E., Hughes, H.B., O'Connell, R.J. and Green, J.R.** (1999) *Colletotrichum*: A model genus for studies on pathology and fungal-plant interactions. *Fungal Genet. Biol.* **27**, 186–198.

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

**Petre, B., Saunders, D.G., Sklenar, J., Lorrain, C., Win, J., Duplessis, S. and Kamoun, S.** (2015) Candidate effector proteins of the rust pathogen *Melampsora Larici-Populina* target diverse plant cell compartments. *Mol. Plant. Microbe. Interact.* **28**, 689–700.

- Punj, V., Bhattacharyya, S., Saint-Dic, D., et al.** (2004) Bacterial cupredoxin azurin as an inducer of apoptosis and regression in human breast cancer. *Oncogene*, **23**, 2367–78.
- Punj, V., Gupta, T.K. Das and Chakrabarty, A.M.** (2003) Bacterial cupredoxin azurin and its interactions with the tumor suppressor protein p53. *Biochem. Biophys. Res. Commun.* **312**, 109–114.
- Qutob, D., Kamoun, S. and Gijzen, M.** (2002) Expression of a *Phytophthora sojae* necrosis-inducing protein occurs during transition from biotrophy to necrotrophy. *Plant J.* **32**, 361–373.
- Rafiqi, M., Ellis, J.G., Ludowici, V.A., Hardham, A.R. and Dodds, P.N.** (2012) Challenges and progress towards understanding the role of effectors in plant-fungal interactions. *Curr. Opin. Plant Biol.* **15**, 477–482.
- Redkar, A., Villajuana- Bonequi, M. and Doehlemann, G.** (2015) Conservation of the *Ustilago maydis* effector See1 in related smuts. *Plant Signal. Behav.* **10**, e1086855.
- Rep, M., Does, H.C. Van Der, Meijer, M., Wijk, R. Van, Houterman, P.M., Dekker, H.L., Koster, C.G. De and Cornelissen, B.J.C.** (2004) A small, cysteine-rich protein secreted by *Fusarium oxysporum* during colonization of xylem vessels is required for I-3-mediated resistance in tomato. *Mol. Microbiol.* **53**, 1373–1383.
- Richard, M.M.S., Pflieger, S., Sévignac, M., Thareau, V., Blanchet, S., Li, Y., Jackson, S.A. and Geffroy, V.** (2014) Fine mapping of Co-x, an anthracnose resistance gene to a highly virulent strain of *Colletotrichum lindemuthianum* in common bean. *Theor. Appl. Genet.* **127**, 1653–1666.
- Rodríguez-Guerra, R., Acosta-Gallegos, J.A., González-Chavira, M.M. and Simpson, J.** (2006) Patotipos de *Colletotrichum lindemuthianum* y su implicación en la generación de cultivares resistentes de frijol. *Agric. técnica en México* **32**, 101–114.
- Rose, J.K., Ham, K.S., Darvill, A.G., Albersheim, P.** (2002) Molecular cloning and characterization of glucanase inhibitor proteins: coevolution

of a counterdefense mechanism by plant pathogens. *Plant Cell*, **14**, 1329-45.

**Sánchez, M. and Colom, F.** (2010) Extracellular DNase activity of *Cryptococcus neoformans* and *Cryptococcus gattii*. *Rev. Iberoam. Micol. organo la Asoc. Esp. Espec. en Micol.* **27**, 10–13.

**Santhanam, P., Esse, P. Van, Albert, I., Faino, L., Nürnberger, T. and Thomma, B.P.H.J.** (2013) Evidence for Functional Diversification within a Fungal NEP1-Like Protein Family. *Mol. Plant. Microbe. Interact.* **26**, 278-86.

**Sanz-Martín, J., Pacheco-Arjona, J., Bello-Rico, V., Vargas, W., Monod, M., Díaz-Mínguez, J., Thon, M. and Sukno, S.** (2016) A highly conserved metalloprotease effector enhances virulence in the maize anthracnose fungus *Colletotrichum graminicola*. *Mol. Plant Pathol.* **36**, 1–15.

**Saunders, D.G.O., Win, J., Cano, L.M., Szabo, L.J., Kamoun, S. and Raffaele, S.** (2012) Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS One*, **7**, e29847

**Schorneck, S., Damme, M. van, Bozkurt, T.O., Cano, L.M., Smoker, M., Thines, M., Gaulin, E., Kamoun, S. and Huitema, E.** (2010) Ancient class of translocated oomycete effectors targets the host nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17421–17426.

**Schouten, A., Baarlen, P. Van and Kan, J.A.L. Van.** (2008) Phytotoxic Nep1-like proteins from the necrotrophic fungus *Botrytis cinerea* associate with membranes and the nucleus of plant cells. *New Phytol.* **177**, 493–505.

**Shan, W., Cao, M., Leung, D., Tyler, B.M.** (2004) The Avr1b locus of *Phytophthora sojae* encodes an elicitor and a regulator required for avirulence on soybean plants carrying resistance gene Rps1b. *Mol. Plant. Microbe. Interact.* **17**, 394-403.



- Selin, C., Kievit, T.R. de, Belmonte, M.F. and Fernando, W.G.D.** (2016) Elucidating the Role of Effectors in Plant-Fungal Interactions: Progress and Challenges. *Front. Microbiol.* **7**, 1–21.
- Stergiopoulos, I., Burg, H.A. van den, Okmen, B., Beenen, H.G., Liere, S. van, Kema, G.H.J. and Wit, P.J.G.M. de** (2010) Tomato Cf resistance proteins mediate recognition of cognate homologous effectors from fungi pathogenic on dicots and monocots. *Proc. Natl. Acad. Sci.* **107**, 7610–7615.
- Stergiopoulos, I., Kourmpetis, Y.A.I., Slot, J.C., Bakker, F.T., Wit, P.J.G.M. De and Rokas, A.** (2012) In silico characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. *Mol. Biol. Evol.* **29**, 3371–3384.
- Schwartz, H.F., Steadman, J.R., Hall, R., Forster, R.L.** (2005) Compendium of bean diseases. **2nd ed.** St Paul, MN: APS Press.
- Takahara, H., Kombrink, A., Hughes, H.B., et al.** (2016) *Colletotrichum higginsianum* extracellular LysM proteins play dual roles in appressorial function and suppression of chitin-triggered plant immunity. *New Phytol.* **11**,1323-1337.
- Takahashi, M., Ashizawa, T., Hirayae, K., et al.** (2010) One of two major paralogs of AVR-Pita1 is functional in Japanese rice blast isolates. *Phytopathology*, **100**, 612–618.
- Terras, F.R., Eggermont, K., Kovaleva, V., et al.** (1995) Small cysteine-rich antifungal proteins from radish: their role in host defense. *Plant Cell*, **7**, 573–588.
- Thynne, E., Saur, I.M.L., Simbaqueba, J., et al.** (2016) Fungal phytopathogens encode functional homologues of plant rapid alkalisation factor (RALF) peptides. *Mol. Plant Pathol.* doi: 10.1111/mpp.12444.
- Urban, M., Pant, R., Raghunath, A., Irvine, A.G., Pedro, H., Hammond-Kosack K.E.** (2015) The Pathogen-Host Interactions database (PHI-

base): additions and future developments. *Nucleic Acids Res.* **43**, D645-55.

**Vargas, W.A., Sanz-Martín, J.M., Rech, G.E., Armijos-Jaramillo, V.D., Rivera, L.P., Echeverria, M.M., Díaz-Mínguez, J.M., Thon, M.R. and Sukno, S.A.** (2016) A fungal effector with host nuclear localization and DNA-binding properties is required for maize anthracnose development. *Mol. Plant-Microbe Interact.* **29**, 83–95.

**Veit, S., Nu, T., Koch, W. and Seitz, H.U.** (2001) A novel protein elicitor (PaNie) from. *Society*, **127**, 832–841.

**Wartha, F., Beiter, K., Normark, S. and Henriques-Normark, B.** (2007) Neutrophil extracellular traps: casting the NET over pathogenesis. *Curr. Opin. Microbiol.* **10**, 52–56.

**Wen, F., White, G.J., VanEtten, H.D., Xiong, Z. and Hawes, M.C.** (2009) Extracellular DNA is required for root tip resistance to fungal infection. *Plant Physiol.* **151**, 820–829.

**Xu, C., Chen, H., Gleason, M.L., Xu, J.-R., Liu, H., Zhang, R. and Sun, G.** (2016) *Peltaster fructicola* genome reveals evolution from an invasive phytopathogen to an ectophytic parasite. *Sci. Rep.* **6**, 22926.

**Yan, H., Zhou, H.-F., Akk, A., Hu, Y., Springer, L.E., Ennis, T.L. and Pham, C.T.N.** (2016) Neutrophil proteases promote experimental abdominal aortic aneurysm via extracellular trap release and plasmacytoid dendritic cell activation. *Arterioscler. Thromb. Vasc. Biol.* **36**, 1660-1669.

**Yang, Y., Zhang, H., Li, G., Li, W., Wang, X. and Song, F.** (2009) Ectopic expression of MgSM1, a Cerato-platanin family protein from *Magnaporthe grisea*, confers broad-spectrum disease resistance in *Arabidopsis*. *Plant Biotechnol. J.* **7**, 763–777.

**Zhao, Z., Liu, H., Wang, C., Xu, J.-R.** (2013) Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*, **14**, 274.



## ARQUIVOS ADICIONAIS

**Tabela S1.** Candidatos efetores de *Colletotrichum lindemuthianum* contendo as três características marcantes encontradas em efetores já comprovados: RCP, NLS, e ricas em cisteína.

NLS						
CIECs	Identidade i83/i89	e-value i83/i89	Homologia a (Uniprot acesso)	Domínio e-value i83/i89	Acesso do domínio	Domínio conserva
CIEC21	79.59	9.57E-116	Uncharacterized protein (N4VCQ2)			
CIEC43	73.08	7.68E-40	Uncharacterized protein (N4W5Q9)			
CIEC74	88.68	4.30E-36	Uncharacterized protein (H1VS09)			
CIEC76	79.81	8.34E-44	Uncharacterized protein (T0KRU7)			
CIEC96	78.49	2.65E-43	Uncharacterized protein (N4UN51)			
CIEC130	70.03	4.34E-65	Uncharacterized protein (N4VR23)			
CIEC158	63.09	3.19E-116	Uncharacterized protein (N4V7Z2)			
CIEC189	88.51/89.53	5.78239E- 44/1.1E-43	Uncharacterized protein (N4VE68)			
CIEC218	87.50	1.36E-22	Uncharacterized protein (N4V5T4)			
CIEC248	60.55	1.56E-22	Uncharacterized protein (N4VAL9)			
CIEC265	71.61	2.18E-53	Uncharacterized protein (N4UZS5)			
CIEC282	90.65	1.53E-35	Uncharacterized protein (N4VGB6)			
CIEC283	96.53	1.42E-94	Small secreted protein (N4VKF4)			
CIEC289	84.43	1.67E-109	Small secreted protein (N4VJY4)			
CIEC291	94.53/93.75	1.03235E- 50/4.62E-50	Uncharacterized protein (N4V4L5)			

CIEC293	76.19	1.49E-30	Uncharacterized protein (A0A010QMR1)			
CIEC305	63.87	1.97534E-007	Uncharacterized protein (N4VXL5)			
CIEC306	93.28	8.11E-66	Uncharacterized protein (N4UWN4)			
CIEC307	84.04	1.63E-17	Uncharacterized protein (N4VP33)			
CIEC314	97.54	2.39E-67	Uncharacterized protein (N4V1A2)			
CIEC316	73.20	7.90E-40	Uncharacterized protein (N4W1Q3)			
CIEC318	78.18	7.97E-54	Uncharacterized protein (N4W6U9)			
CIEC320	84.36	5.39E-66	Uncharacterized protein (N4V706)			
CIEC323	91.89	1.00E-33	Uncharacterized protein (N4VAT0)			
<b>RCP</b>						
CIEC3	87.58	0	Uncharacterized protein (N4UW64)			
CIEC10	92.77	6.02E-96	Uncharacterized protein (N4V6I4)			
CIEC14	73.56	1.34E-98	Uncharacterized protein (N4VC65)			
CIEC17	88.51	7.89E-89	Uncharacterized protein (N4VKR2)			
CIEC35	78.85	1.84E-107	Uncharacterized protein (N4V219)			
CIEC38	77.90	1.40E-119	LysM domain-containing protein (N4W364)	2.86E-08	cd00118	LysM
CIEC44	92.31	0	Chitin binding protein (N4VW34)	3.25E-85	cd10951	CE4_CICDA_like
CIEC46	75.59	3.67E-56	Uncharacterized protein (N4V4M5)			
CIEC51	72.12	3.71E-163	LysM domain-containing protein (A0A010Q1X8)			
CIEC57_i83 <sup>a</sup>	70.51	1.71E-59	Nudix domain-containing protein (N4VA50)	7.3e-09	PF00293.25	NUDIX
CIEC58	95.28	2.80E-63	Uncharacterized protein (N4VAE9)	1.82E-22	cl17157	Alt_A1
CIEC66	96.75	0	Metalloprotease 1 (N4VHH7)	1.02E-104	cd04275	ZnMc_pappalysin_like

CIEC69	78.18/70.91	8.02E-13/1.16E-08	Uncharacterized protein (N4UX54)			
CIEC71	93.02	2.36E-92	Uncharacterized protein (N4UUK0)			
CIEC74	88.68/89.52	4.30E-41/2.2E-41	Uncharacterized protein (N4V7Z2)			
CIEC95	95.50/96.00	3.92E-122/6.6E-124	Secreted protein (N4V6Y8)			
CIEC106	83.75	1.52E-30	Uncharacterized protein (N4V3Q9)			
CIEC109	94.07	6.21E-140	Uncharacterized protein (N4VDZ3)			
CIEC110	84.17	1.93E-119	Uncharacterized protein (N4V330)			
CIEC121	56.01/56.47	2.81E-118/5E-118	LysM domain-containing protein (L2FDS5)	5.80E-08/1.56E-08	cd00118	LysM
CIEC127	96.05/81.88	0	LysM domain-containing protein (N4VCF3)	3.36E-08/5.36E-09	cd00118	LysM
CIEC128	82.52/82.86	1.28E-141/7.7E-141	Uncharacterized protein (N4V6F3)	2.31E-38/1.7e-30	PF14021.3	DUF4237
CIEC130	70.03	4.34E-70	Uncharacterized protein (N4VE68)			
CIEC132	93.38	0	Uncharacterized protein (N4VX50)	1.1e-40	PF13883.3	Pyrid_oxidase_2
CIEC138	68.70	6.55E-52	Uncharacterized protein (N4V876)			
CIEC143	92.76	0	Neutral protease 2 (N4V0V2)	1.2e-75	PF02102.12	Peptidase_M35
CIEC151	78.17	2.79E-106	Uncharacterized protein (N4UXZ9)			
CIEC153	76.92	2.16372E-007	Uncharacterized protein (H1VMV2)			
CIEC157	90.38	1.53E-106	Uncharacterized protein (N4V5T8)			
CIEC160	93.81	2.88E-150	Uncharacterized protein (N4VY54)			

CIEC168	74.44	6.75E-31	CFEM domain-containing protein (A0A010RBE7)	3.3e-09	PF05730.8	CFEM
CIEC176	73.10	9.24E-94	Putative extracellular serine-rich protein (A0A066XIK4)	1.75E-09	cd00920	Cupredoxin
CIEC184	81.33	8.89E-66	Uncharacterized protein (N4VAX1)			
CIEC188	99.13	6.89E-158	Intracellular hyphae protein 1 (CIH1)	2.79E-12	cd00118	LysM
CIEC189	88.51	5.78E-44	Uncharacterized protein (N4VAL9)			
CIEC196	94.34	6.37E-150	Chitin binding protein (N4VXL7)	3.8e-14	PF03067.12	LPMO_10
CIEC212	79.31	2.47E-41	Uncharacterized protein (N4W5J8)			
CIEC218	87.50	1.36E-27	Uncharacterized protein (N4UZS5)			
CIEC230	90.64	1.47E-89	Uncharacterized protein (N4UUD4)			
CIEC235	53.40	1.05E-25	Uncharacterized protein (N4VNF3)			
CIEC248	60.55	1.56E-27	Uncharacterized protein (T0KRU7)			
CIEC266	73.08	0	NPP1 domain-containing protein (N4VPV3)	4.2e-43	PF05630.8	NPP1
CIEC282	90.65	1.53E-40	Uncharacterized protein (N4VGB6)			
CIEC283	96.53	1.42E-99	Small secreted protein (N4VKF4)			
CIEC284	81.54	5.68E-119	Uncharacterized protein (N4VAE5)	2.1e-16	PF12296	HsbA
CIEC288	76.00	8.89E-53	Uncharacterized protein (L2FYL5)			
CIEC289	84.43	1.67E-114	Small secreted protein (N4VJY4)			
CIEC290	64.65/91.67	4.79E-62/5.38E-81	Uncharacterized protein (A0A010RTD1)			
CIEC291	94.53/93.75	1.03E-50/4.62E-50	Uncharacterized protein (N4V4L5)			

CIEC293	76.19	1.49E-35	Uncharacterized protein (A0A010QMR1)			
CIEC295	83.54/81.74	3.79E-95/3.02E-94	Pathogenesis-related protein (N4VPG7)	1.43E-51/1.60E-51	cd05382	SCP_GAPR-1_like
CIEC297	87.79	6.08E-50	Uncharacterized protein (N4V296)			
CIEC298	97.37	1.77E-67	Uncharacterized protein (N4ULR6)			
CIEC305	63.87	1.97534E-007	Uncharacterized protein (N4VXL5)			
CIEC307	84.04	1.63E-22	Uncharacterized protein (N4VP33)			
CIEC308	63.03	7.83E-41	Uncharacterized protein (N4V348)			
CIEC310	58.93	4.12E-18	Uncharacterized protein (N4VQ76)			
CIEC314	97.54	2.39E-72	Uncharacterized protein (N4V1A2)			
CIEC315	92.24	1.85E-71	Uncharacterized protein (N4V188)			
CIEC316	73.20	7.90E-45	Uncharacterized protein (N4W1Q3)			
CIEC317	84.91	9.67E-81	Uncharacterized protein (N4V771)			
CIEC318	78.18	7.97E-59	Uncharacterized protein (N4W6U9)			
CIEC321	91.63	6.04E-125	Uncharacterized protein (N4V3Q6)			
CIEC323	91.89	1E-33	Uncharacterized protein (N4VAT0)			
CIEC325	89.44	4.88E-125	Uncharacterized protein (N4V8Z0)			

---

**Cysteine-rich**

---

CIEC2	95.58/96.46	2.88E-73/5.88E-75	Uncharacterized protein (N4UUM3)			
CIEC3	87.58	0	Uncharacterized protein (N4UW64)			
CIEC4	90.62	1.08E-19	Uncharacterized protein (N4UW50)			
CIEC6	90.55	1.23E-62	Uncharacterized protein (N4W4U0)			
CIEC8	76.83	9.77E-24	Uncharacterized protein (N4VAD4)			



CIEC11	93.36	4.13E-156	Uncharacterized protein (N4UXW2)			
CIEC12	95.74	1.31E-55	Uncharacterized protein (N4V1C2)			
CIEC15	87.78	3.62E-102	Uncharacterized protein (N4VCA4)			
CIEC16	87.85	2.45E-155	Uncharacterized protein (N4V4L2)	6.8e-10	PF14856.3	Hce2
CIEC17	88.51	7.89E-89	Uncharacterized protein (N4VKR2)			
CIEC20	97.70	8.78E-52	Uncharacterized protein (N4V0M1)			
CIEC22	92.22	2.84E-52	Uncharacterized protein (N4VNU2)			
CIEC23	83.75	1.48E-36	Uncharacterized protein (N4VUT2)			
CIEC24	50.00	3.36E-18	Uncharacterized protein (N4VDP9)			
CIEC25	85.56/84.44	2.20E-51/1.57E-45	Cvnh domain-containing protein (N4VPS9)	2.87E-05/2.95E-05	cl07470	CVNH
CIEC28	97.31	2.44E-128	Chitin binding protein (N4VQK0)	2.4e-07	PF03067.12	LPMO_10
CIEC29	52.48/53.92	3.97E-20/9.69E-20	Uncharacterized protein (L2G0M0)			
CIEC30	94.65	1.42E-114	Uncharacterized protein (N4V9S2)			
CIEC31	82.35	2.44E-146	Cvnh domain-containing protein (N4VUN1)	1.2e-08	PF08881.7	CVNH
CIEC32	64.00	5.77E-24	Uncharacterized protein (N4VQQ8)			
CIEC33	98.23	2.24E-163	Secreted protein (N4UR87)			
CIEC34	97.89	7.04E-95	Uncharacterized protein (N4UR14)			
CIEC35	78.85	1.84E-107	Uncharacterized protein (N4V219)			
CIEC36	95.95	1.2E-99	Uncharacterized protein (N4W3D0)			
CIEC38	77.90	1.4E-119	LysM domain-containing protein (N4W364)	2.86E-08	cd00118	LysM

CIEC39	95.61	3.5E-65	Uncharacterized protein (N4VNG0)			
CIEC40	72.94	3.79E-22	Uncharacterized protein (N4W5I6)			
CIEC41	92.39	1.12E-109	Chitin binding protein (N4VNB2)	4.2e-07	PF03067.12	LPMO_10
CIEC42	81.95	6.8E-66	Uncharacterized protein (N4VKQ7)			
CIEC43	73.08	7.68E-45	Uncharacterized protein (N4VR23)			
CIEC44	92.31	0	Chitin binding protein (N4VW34)	8.64E-84	cd10951	CE4_CICDA_like
CIEC46	75.59	3.67E-56	Uncharacterized protein (N4V4M5)			
CIEC47	61.74	4.36E-44	Uncharacterized protein (N4V2R0)			
CIEC48	76.61	4.04E-79	Uncharacterized protein (N4V3P4)			
CIEC51	72.12	3.70E-163	LysM domain-containing protein (A0A010Q1X8)			
CIEC52	92.21	4.41E-42	Uncharacterized protein (N4VM39)			
CIEC54	93.20	5.24E-125	Uncharacterized protein (N4VP59)			
CIEC55	58.65	1.79E-27	Uncharacterized protein (N4VHN3)			
CIEC56	82.93/83.74	3.02E-62/1.13E-62	Uncharacterized protein (N4UQ10)			
CIEC60			not available			
CIEC61	92.47	2.55E-82	Small secreted protein (N4VSF1)			
CIEC62	84.00	2.18E-26	Uncharacterized protein (N4VMN4)			
CIEC67	66.07	2.84E-76	Uncharacterized protein (N4VD62)			
CIEC68	69.47	3.31E-37	Uncharacterized protein (N4VY65)			
CIEC71	93.02	2.36E-92	Uncharacterized protein (N4UUUK0)			
CIEC72	90.86	2.93E-97	Uncharacterized protein (N4V345)			
CIEC73	80.46	7.09E-36	Uncharacterized protein (N4V368)			
CIEC75	68.93	3.13E-38	Uncharacterized protein (N4VXQ8)			

CIEC76	79.81	8.34E-49	Uncharacterized protein (N4W5Q9)			
CIEC79	97.06	1.20E-37	Uncharacterized protein (N4VPA6)			
CIEC80	93.42	1.19E-95	Uncharacterized protein (N4V857)			
CIEC81	99.27	1.33E-89	Uncharacterized protein (N4V6C3)			
CIEC83	91.30/89.72	7.19E-118/1.84E-58	Uncharacterized protein (N4UMN6)			
CIEC85	78.26	3.08E-37	Uncharacterized protein (N4VJH4)			
CIEC87	90.98	1.83E-81	Cvnh domain-containing protein (N4VQD4)	3.8e-10	PF08881.7	CVNH
CIEC89	92.41	2.04E-96	Uncharacterized protein (N4V9A8)			
CIEC90	95.86	0	Uncharacterized protein (N4V1K4)	1.1e-17	PF14040.3	DNase_NucA_NucB
CIEC92	65.33	3.88E-24	Uncharacterized protein (N4VGC7)			
CIEC94	95.70	1.80E-47	Uncharacterized protein (N4VM86)			
CIEC95	95.50	3.91E-122	Secreted protein (N4V6Y8)			
CIEC96	78.49	2.65E-47	Uncharacterized protein (H1VS09)			
CIEC97	79.17	1.27E-18	Uncharacterized protein (N4V772)			
CIEC99	97.06/98.53	2.56E-37/7.58E-38	Uncharacterized protein (N4VB69)			
CIEC101	94.63	1.32E-83	Uncharacterized protein (N4V7J8)			
CIEC102	87.60	1.51E-66	Uncharacterized protein (N4VD42)			
CIEC103	93.65	1.51E-66	Uncharacterized protein (N4VYL6)			
CIEC105	99.46	1.35E-127	Uncharacterized protein (N4VNW4)			
CIEC106	83.75	1.51E-30	Uncharacterized protein (N4V3Q9)			
CIEC111	78.85	7.25E-72	Uncharacterized protein (N4V4W2)	8.2e-13	PF08881.7	CVNH
CIEC112	76.19	1.29E-46	Uncharacterized protein (N4VU13)			
CIEC115	93.65	3.75E-75	Uncharacterized protein (N4VQY3)			

CIEC116	94.71	5.48E-101	Uncharacterized protein (N4VV08)			
CIEC117	82.71	1.66E-61	Uncharacterized protein (N4VR04)			
CIEC118b <sup>b</sup>	97.00	7.77E-56	Biotrophy-associated secreted protein 2 (N4V676)			
CIEC118	65.71	2.62E-64	Putative biotrophy-associated secreted protein 2 (A0A066XFM9)			
CIEC121	56.01/56.47	2.80E-118/4.97E-118	LysM domain-containing protein (L2FDS5)	5.80E-08/1.56E-08	cd00118	LysM
CIEC122	91.80	8.30E-18	Uncharacterized protein (N4VCQ9)			
CIEC123	93.60	6.17E-67	Uncharacterized protein (N4VPQ8)			
CIEC125	96.45	6.01E-94	Uncharacterized protein (N4VCD8)			
CIEC126	94.07	0	LysM domain-containing protein (N4V1T7)	5.3e-05	PF01476.17	LysM
CIEC127	96.05/81.88	0	LysM domain-containing protein (N4VCF3)	3.36E-08/5.36E-09	cd00118	LysM
CIEC129	46.15	8.59E-13	Uncharacterized protein (N4VQT6)			
CIEC134	65.14	1.63E-40	Uncharacterized protein (L2FF21)			
CIEC136	84.09	1.33E-42	Uncharacterized protein (N4VAZ6)			
CIEC138	68.70	6.54E-52	Uncharacterized protein (N4V876)			
CIEC139	76.47	1.55E-78	M6 protein (N4W1J2)			
CIEC140	62.86	3.34E-31	Uncharacterized protein (N4VLV5)			
CIEC141	60.64	2.89E-33	Uncharacterized protein (H1UZ14)			
CIEC142	85.71	8.66E-89	Uncharacterized protein (N4UUZ9)			
CIEC149	95.33	2.74E-067	Uncharacterized protein (N4UK60)			

CIEC150	95.15/94.55	5.56E-105/1.19E-104	Uncharacterized protein (N4ULU1)			
CIEC151	78.17	2.79E-106	Uncharacterized protein (N4UXZ9)			
CIEC152	98.23	3.01E-69	Uncharacterized protein (N4UK84)			
CIEC156	80.00	2.40E-15	Uncharacterized protein (N4V7R0)			
CIEC162	95.11	5.76E-123	Uncharacterized protein (N4V2W5)			
CIEC164	95.04	2.07E-75	Uncharacterized protein (N4UY44)			
CIEC166	57.14	7.72E-31	CAD1 (Q9HFC2)			
CIEC167	97.46	1.27E-78	Uncharacterized protein (N4VMN2)			
CIEC168	74.44	6.75E-31	CFEM domain-containing protein (A0A010RBE7)	3.3e-09	PF05730.8	CFEM
CIEC169	73.53	3.82E-24	Uncharacterized protein (N4UYY2)			
CIEC172	60.87	4.78E-11	Uncharacterized protein (N4UQT5)			
CIEC174	90.55	6.47E-61	Uncharacterized protein (N4VEC4)			
CIEC175	80.68	1.01E-39	Uncharacterized protein (N4VUG8)			
CIEC177	86.67	3.89E-39	Uncharacterized protein (N4UKH8)			
CIEC178	82.19	1.39E-33	Uncharacterized protein (N4UKI6)			
CIEC179_i83 <sup>e</sup>	96.65	5.51E-122	Uncharacterized protein (N4UQX8)			
CIEC182	82.69	1.46E-127	Uncharacterized protein (N4W156)			
CIEC184	81.33	8.89E-66	Uncharacterized protein (N4VAX1)			
CIEC185	89.15	5.2E-69	Uncharacterized protein (N4VW72)			
CIEC186	89.95	1.34E-140	Uncharacterized protein (N4VSA4)			
CIEC187	98.48	2.36E-38	Uncharacterized protein (N4VN27)			
CIEC190	89.05	5.31E-83	Uncharacterized protein (N4V6S9)			
CIEC191	86.18	1.31E-74	Uncharacterized protein (N4UUH0)			
CIEC192	89.73	2.35E-119	Uncharacterized protein (N4UKJ4)	1.2e-25	PF16541.2	AltA1

CIEC193	69.49	2.53E-49	Uncharacterized protein (N4V5M2)			
CIEC194	47.12	7.35E-19	Uncharacterized protein (N4UVD0)			
CIEC195	84.03	1.58E-58	Uncharacterized protein (N4W5V5)			
CIEC197	70.33	1.30E-34	Uncharacterized protein (N4VVK2)			
CIEC199	86.96	4.94E-48	Uncharacterized protein (N4VUU3)			
CIEC200	96.92	5.13E-158	Uncharacterized protein (N4VH24)			
CIEC201	57.97	2.14E-18	Uncharacterized protein (T0LP40)			
CIEC202	89.55	8.53E-32	Uncharacterized protein (N4W760)			
CIEC203			Not available			
CIEC204	80.83/81.67	1.89E-59/5.98E-60	Uncharacterized protein (N4VL16)			
CIEC205	82.86	1.05E-121	Secreted in xylem 1 protein (N4UXB6)			
CIEC206	98.85	5.34E-51	Uncharacterized protein (N4W7A7)	2.2e-06	PF05730.8	CFEM
CIEC207	83.72	2.38E-42	Uncharacterized protein (N4V0X4)			
CIEC209	94.23	4.23E-58	Uncharacterized protein (N4UTP9)			
CIEC211	88.03	3.25E-69	Uncharacterized protein (N4VXG7)			
CIEC212	79.31	2.46E-41	Uncharacterized protein (N4W5J8)			
CIEC213	46.38	5.24E-9	Putative uncharacterized protein (E3QEY2)			
CIEC214	88.10	2.34E-103	Uncharacterized protein (N4VYW5)			
CIEC215	88.98	2.14E-69	Uncharacterized protein (N4UTS7)			
CIEC220	87.07	2.14E-64	Uncharacterized protein (N4VVG3)			
CIEC222	89.43	1.10E-154	Uncharacterized protein (N4VIZ5)			
CIEC223	75.90	1.37E-26	Uncharacterized protein (N4V4R1)			
CIEC224	91.28	8.04E-92	Proteinase inhibitor kazal (N4VF32)			

CIEC225	97.96	3.66E-98	Uncharacterized protein (N4W5T1)			
CIEC226	83.86/83.41	9.63E-128/5.4E-127	Secreted in xylem 6-like protein (N4VP26)			
CIEC227	67.57	1.10E-21	Uncharacterized protein (N4VEI4)			
CIEC229	47.54/52.73	8.33E-29/1.7E-34	Avirulent on Ve1 (H9CDN2)	4.61E-11/2.08E-11	cl04011	DPBB_1
CIEC230	90.64	1.47E-89	Uncharacterized protein (N4UUD4)			
CIEC233	93.33	1.14E-53	Uncharacterized protein (N4V989)			
CIEC235	53.40	1.04E-25	Uncharacterized protein (N4VNF3)			
CIEC237	92.38	1.52E-52	Uncharacterized protein (N4VI97)			
CIEC239	88.61	1.E-36	Uncharacterized protein (N4VNH5)			
CIEC241	92.59	1.72E-32	Uncharacterized protein (N4UQQ0)			
CIEC243	98.13	2.3E-65	Uncharacterized protein (N4USI6)			
CIEC249	45.59	0.006	Uncharacterized protein (N4VPA6)			
CIEC250	61.60	2.2E-37	Uncharacterized protein (N4VUX0)			
CIEC251	74.16	4.72E-36	Uncharacterized protein (N4VBG4)			
CIEC252	86.59	1.42E-100	Uncharacterized protein (N4VUL0)			
CIEC254	62.96	9.01E-27	Uncharacterized protein (A0A010QDW0)			
CIEC256	78.95	1.33E-63	Secreted protein (N4V6T4)			
CIEC257	92.05	9.72E-115	Uncharacterized protein (N4V8Y2)			
CIEC259	96.74	1.50E-127	Uncharacterized protein (N4VKF1)			
CIEC260	79.82	1.3E-58	Uncharacterized protein (T0KBD3)			
CIEC261	65.43	1.92E-11	Uncharacterized protein (N4V726)			
CIEC263	55.17	1.74E-21	Uncharacterized protein (N4VL67)			

CIEC264	66.67	3.0E-46	M6 protein (N4VN83)			
CIEC267	88.39	7.7E-63	Uncharacterized protein (N4VPC8)			
CIEC269	71.88	4.59E-57	Uncharacterized protein (T0JIR1)			
CIEC271	85.29	1.73E-22	Uncharacterized protein (N4UY53)			
CIEC272	60.54	5.54E-46	Uncharacterized protein (N4V6J2)	2.5e-10	PF08881.7	CVNH
CIEC273	76.71	1.49E-26	Uncharacterized protein (N4V767)			
CIEC274	94.02	0	LysM domain-containing protein (N4UNG2)	7.51E-06	cd00118	LysM
CIEC275	67.26	2.5E-47	Uncharacterized protein (N4V6G9)			
CIEC276	91.82	8.20E-66	Uncharacterized protein (T0KQU3)			
CIEC277	77.78	1.26E-60	Uncharacterized protein (N4VHB4)			

---

<sup>a</sup>CIEC57\_i89 does not have predicted RCP

<sup>b</sup>CIEC118b - 49 % of identity to Clec118\_i89 and Clec118\_i83

<sup>c</sup>CIEC179\_i89 has Cysteine content less 3%



**Tabela S2.** Domínios conservados encontrados em proteínas candidatas a efetores em *Colleotrichum* spp. O número de proteínas contendo o domínio por cada espécie é dado nas colunas a baixo dos códigos das espécies.

Domínios conservados	lind_83	lind_89	sub	Fio	nym	salic	sim	frut	gram	orb	higg	Função putativa
Aes	1	1	1		1	1	1					Acetyl esterase/lipase (Lipid transport and metabolism)
CrtC Domain						1						Lipid metabolism
crotonase-like					1		1					Crotonase/Enoyl-Coenzyme A (CoA) hydratase
CBP Domain				1			1					Calcium binding protein
CHRD Domain	1	1		1	1	1	1	1		1	1	Inhibitor of bone morphogenetic proteins
But2 Domain	1	1	1	1	1	1	1		1	2	1	Ubiquitin 3 binding protein
PEBP_euk				1		1						PhosphatidylEthanolamine-Binding Protein
PAN Domain				1	1	1	1	1			1	Mediating protein-protein or protein-carbohydrate interactions
WSC Domain				1	1		1	1				Carbohydrate binding
CBM_4_9 Domain					1			1				Carbohydrate binding domain
CVNH Domain	5	5	7	4	3	3	5	9	3	5	5	carbohydrate-binding
LysM Domain	7	7	5	6	7	7	6	10	6	8	7	Chitin binding
CE4_CICDA_like Domain	1	1						1		1	1	Chitin binding
Chitin_bind_3 superfamily											1	Chitin binding
GH43_62_32_68 Domain				1	1	1			1			Glycosyl hydrolase
Glyco_hydro_3 Domain											1	Glycosyl hydrolase

Glyco_hydro_cc Domain				1								Glycosyl hydrolase	
Glyco_hydrolase_16 Domain			1	2	1	2	2		1		2	Glycosyl hydrolase	
CBM_1 Domain			1			1	1		1		1	Fungal cellulose binding domain	
Polysacc_lyase Domain			1	1	1	1	1					Polysaccharide lyase	
LPMO_10 Domain	3	3	1	3	2	3	3	3	2	3	1	Lytic polysaccharide mono-oxygenase hydrolase activity	
Abhydrolase_3											1	hydrolase activity	
Abhydrolase_6 Domain			1	2	2			1	1	1	1	2	hydrolase activity
cysteine_hydrolases superfamily												1	
Nudix_Hydrolase Domain	1	1									1	1	Phosphohydrolases
PI-PLCc_GDPD_SF Domain				1	1	1			1	1			phosphoric diester hydrolase activity
Pro-peptidase_S53							1						Activation domain of S53 peptidases
Aspzincin_M35 Domain	1	1					1						Metalloendopeptidase activity
M35_deuterolysin_like Domain										2	1		
M35_like Domain	1	1	3	1	1			1	1		4	1	Metallopeptidase
Peptidase_M35 Domain	3	3										2	Metalloprotease
ZnMc Domain	3	3	5	5	5	3	6	2	3	5	7		Metalloprotease
PepX_C Domain												1	Dipeptidyl-peptidase
Cupin_1 Domain	1	1		1					2		1	2	Oxalate oxidases
Tyrosinase Domain						1	2		1				Oxidoreductase activity
Peroxidase_2 Domain	2	2	1	5	5	7	4	2	1	2	5		Peroxidase activity
Hr-like Domain			1		1	1	1			1		1	Oxygen (O2) transport
Cupredoxin Domain	2	2	3	3	3	4	3	1	2	2	1		Inter-molecular electron transfer

microbial_RNases Domain			1	1	1	1	2		1			Cleave phosphodiester bonds in RNA	
DNase_NucA_NucB Domain	1	1								1		Deoxyribonuclease	
ZnPC_S1P1											1	Zinc dependent phospholipase C/S1-P1 nuclease	
NTF2_like Domain			2	3	1	3				1		Nuclear import of RanGDP	
Helo_like_N Domain	1	1										STAND NTPases	
NADB_Rossmann Domain						1						NAD binding	
SesA Domain						1						N-terminal domain on NACHT_NTPase and P-loop NTPases	
His_Phos_2 Domain											1	Phosphatase	
Pyrid_oxidase_2 Domain	1	1	1	1	1		1	1	1	2		Synthesis of pyridoxine (vitamin B6) and pyridoxal phosphate	
NPP1 Domain	7	7	3	11	11	13	11	7	4	6	4	Necrosis inducing	
Hce2 Domain	2	2	1	3	2	3	2	1	2		1	Necrosis-inducing factor	
CFEM Domain	2	2	6	3	3	7	3	2	2	2	5	Fungal pathogenesis	
Cerato-platanin Domain	1	1	3	2	1	1	1			1	1	Phytotoxin	
Kp4 Domain										1	2	Killer toxin	
SCP_GAPR-1_like	1	1		1	1	1	1				1	1	Golgi-associated plant pathogenesis related protein (GAPR)-like
Defensin_like Domain									1			1	Antifungal protein
HsbA Domain	1	1	2	4	4	4	4	1	3	1	4	4	Stimulates PBSA degradation

PAD Domain							1					Protein-arginine deiminase
RALF Domain											1	Rapid alkalization factor
DPBB_1 Domain	1	1	1			1			1	2		Rare lipoprotein A
PTZ00395						1						Sec24-related protein
SRPBCC Domain						2	1					Bind diverse ligands
PRK00555										1		Galactokinase
Alt_A1 Domain	3	3	3	2	3	2	3	2	2	3	1	Unknown
DUF4360 Domain	2	2	1	3	2	3	3	3	1	2	1	Unknown
DUF917 superfamily									1			Unknown
DUF1524 Domain	1	1	1	1	1	1	1	1	1	1	1	Unknown
DUF1961 Domain								1		1		Unknown
DUF2264 Domain	1											Unknown
DUF3129 Domain			1	2	2	2	2		2		2	Unknown
DUF3237 Domain	1	1	1				1	1	2	1	3	Unknown
DUF3328 Domain										1	1	Unknown
DUF3431 Domain				1	1							Unknown
DUF3455 Domain	1	1					1			1		Unknown
DUF4237 Domain	1	1								1		Unknown
Abhydrolase Domain							1		1			Unknown
EthD Domain				2	2	1	2				1	Unknown
LicD Domain			2	2	2	2	2		2		2	Unknown
M35_Asp_f2_like Domain	1	1								1		Unknown
MU117 Domain				1	1	1	1					Unknown
PHA03247					1	1	1					Unknown
<b>Total</b>	<b>62</b>	<b>61</b>	<b>58</b>	<b>84</b>	<b>83</b>	<b>95</b>	<b>88</b>	<b>63</b>	<b>53</b>	<b>67</b>	<b>80</b>	

**Tabela S3.** Efeitores comprovados experimentalmente de fungos e oomicetos que possuem domínios funcionais presentes no banco de dados Pfam.

<b>Efedor</b>	<b>Acesso Uniprot</b>	<b>Pfam</b>	<b>Posição (E-value)</b>	<b>Referência</b>
NP1	Q02039	NIP_1 (PF08995)	1-82(3.8e-57)	Rohe et al., 1995
PcF	Q94FS7	PcF (PF09461)	24-68(3.7e-22)	Orsomando et al., 2001
CBEL	O42830	PAN_1 (PF00024)	61-165(1.6e-19); 197-257(1.6e-09)	Mateos et al., 1997
NPP1	Q9AT28	NPP1 (PF05630)	41-234(2.1e-72)	Fellbrich et al., 2002
AvrL567-A	Q6R661	AvrL567-A (PF11529)	24-150(1.5e-81)	Dodds et al., 2004
AVRL567-b	Q6R659	AvrL567-A (PF11529)	24-150(1.9e-87)	Dodds et al., 2004
AVR4	Q00363	CBM_14 (PF01607)	50-109(4.8e-07)	Joosten et al., 1994
ECP2	Q00365	Hce2 (PF14856)	34-139(1.9e-15)	Laugé et al., 1997
INF1	Q01905	Elicitin (PF00964)	23-115(1.6e-22)	Kamoun et al., 1998
AVR3a	Q572D3	RXLR (PF16810)	1-143(1.9e-30)	Armstrong et al., 2005
INF2A	O42718	Elicitin (PF00964)	23-115(2.8e-20)	Vleeshouwers et al., 2006
PsojNIP	Q9AXL8	NPP1 (PF05630)	41-234(3.2e-69)	Qutob et al., 2002
AvrLm4	O93976	Glyco_hydro_10 (PF00331)	28-325(1.8e-74)	Pareja-Jaime et al., 2008
Avr1b-1	Q6TA23	FYVE (PF01363)	26-103(3.2e-12)	Shan et al., 2004
EPI1	Q6PQH2	Kazal_2 (PF07648)	98-128(3.6e-07)	Tian et al., 2004
EPI10	Q6PQH1	Kazal_2 (PF07648)	101-133(2e-08)	Tian et al., 2005
GIP2	Q945T9	Trypsin (PF00089)	29-252(1.5e-39)	Rose et al., 2002
Avh5	E0W547	RXLR (PF16810)	1-127(5.5e-26)	Sun et al., 2013
Avh241	E0W4Y2	RXLR (PF16810)	1-130(2.1e-06)	Yu et al., 2012
Avr1b-1	G4ZRQ8	RXLR (PF16810)	1-135(6.6e-26)	Song et al., 2013
GIP1	Q945U0	Trypsin (PF00089)	31-251(1.2e-36)	Rose et al., 2002
SFI4	D0NCC1	TPR_12 (PF13424)	106-173(1.1e-11); 276-315(5.9e-06)	Zheng et al., 2014

SFI6	D0NN72	RXLR (PF16810)	2-121(4e-23)	Zheng et al., 2014
Avr1d	L7WWG5	RXLR (PF16810)	1-123(1.2e-20)	Yin et al., 2013
XEG1	G5A0G9	Glyco_hydro_12 (PF01670)	33-238(6.8e-40)	Ma et al., 2015
Avr1c	A4LAA7	RXLR (PF16810)	1-124(6.5e-11)	Na et al., 2014
Avr1a	A4L9T6	RXLR (PF16810)	1-107(3.3e-11)	Na et al., 2014
ATR13	Q5G7K8	ATR13 (PF16829)	54-154(1.3e-62)	Caillaud et al., 2012
AVR-Pita	Q9C478	Aspzincin_M35 (PF14521)	77-216(3.3e-20)	Kang et al., 2001
SLP1	G4N906	LysM (PF01476)	43-86(4.3e-06); 115-158(1.7e-07)	Mentlak et al., 2012
AVR-Pita1JA	C1KJH7	Aspzincin_M35 (PF14521)	78-217(1.1e-20)	Takahashi et al., 2010
AVR-Pita1JB	C1KJG7	Aspzincin_M35 (PF14521)	78-217(2.2e-19)	Takahashi et al., 2010
AVR-Pita1	C1KJH9	Aspzincin_M35 (PF14521)	78-217(1.1e-20)	Dai et al., 2010
AVR-Pita1	C1KJI7	Aspzincin_M35 (PF14521)	78-217(3.3e-19)	Khang et al., 2010
PemG1	A0N0D1	Stm1_N (PF09598)	4-59(2.7e-10)	Peng et al., 2011
MoCDIP4	G4MVX4	Glyco_hydro_61 (PF03443)	19-220(1.5e-50)	Chen et al., 2013
MgSM1	G4MKI0	Cerato-platanin (PF07249)	19-137(9e-57)	Yang et al., 2009
Iug18	A0A0A7DLV6	RelA_SpoT (PF04607)	96-231(6.2e-08)	Dong et al., 2015
FGL1	Q6WER3	Lipase_3 (PF01764)	106-236(3.9e-31)	Blümke et al., 2014
Ecp6	B3VBK9	LysM (PF01476)	44-87(1.6e-10); 117-161(3e-11); 174-217(3.7e-07)	Bolton et al., 2008
AvrL567-C	Q6R656	AvrL567-A (PF11529)	24-150(9.1e-83)	Dodds et al., 2004
VPS9	G3KLH0	VPS9 (PF02204)	372-471(2.7e-31)	Nirmala et al., 2011
Avr2	D0N6C5	RXLR (PF16810)	1-82(6.2e-08)	Haas et al., 2009; Lokossou et al., 2009
CRN8	Q2M405	Pkinase (PF00069)	447-535(1.3e-07)	Win et al., 2006; Schornack et al., 2010
Avr1b	Q8LSI1	RXLR (PF16810)	1-136(3.1e-21)	Shan et al., 2004

Avr3b	E0W4V8	NUDIX (PF00293)	174-229(7.1e-08)	Dong et al., 2011
Avr3c	C0LF25	RXLR (PF16810)	1-130(1.7e-09)	Dong et al., 2009
Avr1d	G1FQR7	RXLR (PF16810)	1-119(1.4e-18)	Na et al., 2013
PaNie	Q9SPD4	NPP1 (PF05630)	43-231(1.1e-68)	Veit et al., 2001

---

**Tabela S4.** Lista de primers utilizados nas análises de expressão por meio de *real-time* qRT-PCR.

<b>Exclusivo de</b>	<b>Sequências 5' – 3'</b>
<b><i>Colletotrichum</i> spp.</b>	
CIEC97F	TGCGTTTCATTCTGACACTCG
CIEC97R	GATCACCACGGTTGTTGCAG
CIEC178F	CGACCTTCCCAGATCGTTGT
CIEC178R	GAGCAAGTGTAACCGCCAAG
<b>Presente em outras espécies de patógenos</b>	
CIEC53F	CCCCGACTACTCCTTTGACG
CIEC53R	TACTGGTTCTGGTTCTTGGCC
CIEC229F	AACAACCCGGCCTATTACCC
CIEC229R	CGACCACCTTGACCTGGAAC
CIEC10F	CTTACTGTTGCTGGCGGTG
CIEC10R	TCTGGTCCTGGCTTTGGTTC
CIEC147F	TGCACCGACGTATACAGCTC
CIEC147R	CATGGCAGGTTTTGGTCAGG
<b>Controle endógeno</b>	
actF	TGTCTTCGAGACCTTCAACG
actR	AAGAGCGAAACCCTCGTAAA



CASLEY BORGES DE QUEIROZ

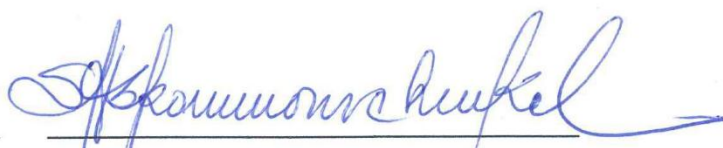
**SEQUENCIAMENTO DO GENOMA DE *Colletotrichum lindemuthianum* E  
IDENTIFICAÇÃO DE GENES CANDIDATOS A EFETORES**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Microbiologia Agrícola, para obtenção do título de *Doctor Scientiae*.

APROVADA: 24 de maio de 2017.



Mateus Ferreira Santana  
(Coorientador)



Sérgio Hermínio Bromonschenkel



Diego Silva Batista



Elaine Aparecida de Souza



Marisa Vieira de Queiroz  
(Orientadora)