

HILBERTY LUCAS NUNES CORREIA

**GENÔMICA COMPARATIVA POSSIBILITA A IDENTIFICAÇÃO DE FATORES DE
VIRULÊNCIA NO GENOMA DE *Colletotrichum lindemuthianum***

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*

HILBERTY LUCAS NUNES CORREIA

**GENÔMICA COMPARATIVA POSSIBILITA A IDENTIFICAÇÃO DE FATORES DE
VIRULÊNCIA NO GENOMA DE *Colletotrichum lindemuthianum***

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*

Orientadora: Marisa Vieira de Queiroz

Coorientadores:

Mateus Ferreira Santana

Pedro Marcos Pereira Vidigal

HILBERTY LUCAS NUNES CORREIA

**GENÔMICA COMPARATIVA POSSIBILITA A IDENTIFICAÇÃO DE FATORES DE
VIRULÊNCIA NO GENOMA DE *Colletotrichum lindemuthianum***

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:

Assentimento:

Hilberty Lucas Nunes Correia
Autor

Marisa Vieira de Querioz
Orientadora

AGRADECIMENTOS

Primeiramente a Deus pelo dom da vida e pelas inúmeras dádivas recebidas.

A minha família, amigos pessoais e a todos que sempre torceram, oraram por mim e me apoiaram.

Ao departamento de Microbiologia e ao programa de Pós-Graduação em Microbiologia Agrícola da Universidade Federal de Viçosa por todo o esforço empregado no intuito de entregar a melhor formação e condição de trabalho para seus alunos.

A Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro direto ou indireto.

A equipe do Laboratório de Genética Molecular de Microrganismos por todo o apoio e colaboração para com meu trabalho.

A minha comissão orientadora pelo valioso conhecimento compartilhado e por toda a colaboração com o trabalho.

A minha orientadora, professora Marisa Vieira de Queiroz por todos os ensinamentos compartilhados, muitos dos quais admito ter conseguido absorver somente após me retirar da Universidade e sem dúvida muitos dos quais irei ainda absorver muito ao longo dos próximos anos.

Aos doutores Pedro Marcos Pereira Vidigal e Jose Cleydson Ferreira da Silva, por compartilharem comigo seu tempo, conhecimento e paixão pela bioinformática.

A todos os membros que aceitaram participar da banca de defesa mesmo em meio ao momento delicado pelo qual a humanidade está passando.

A todos o meu Muito Obrigado!

SUMÁRIO

RESUMO.....	iv
ABSTRACT.....	vi
INTRODUÇÃO.....	1
Genomas de <i>C. lindemuthianum</i>	4
Anotação dos genomas.....	4
Anál. de seq. de DNA que cod. proteínas env. no ciclo sexual.....	5
Secretoma predito.....	6
Análise de grupos de ortólogos.....	6
Proteases e CAZymes preditas secretadas.....	7
Metabolismo secundário.....	8
RESULTADOS.....	9
Estrutura Geral dos Genomas de <i>C. lindemuthianum</i>	9
Repeat Induced Point (RIP).....	12
SNPs em Sequências Codificadoras de Efetores.....	13
Comparação Entre os Isolados de <i>C. lindemuthianum</i>	20
Análise de Grupos de Ortólogos.....	23
Transportadores.....	27
Enzimas Ativas em Carboidratos (CAZymes) Secretadas.....	47
Proteases secretadas.....	55
Metabolismo Secundário.....	56
DISCUSSÃO.....	69

RESUMO

CORREIA, Hilberty Lucas Nunes D.Sc., Universidade Federal de Viçosa, Dezembro de 2020. **Genômica comparativa possibilita a identificação de fatores de virulência no genoma de *Colletotrichum lindemuthianum***. Orientadora: Marisa Vieira de Queiroz. Coorientadores: Pedro Marcos Pereira Vidigal, Mateus Ferreira Santana.

O fungo *Colletotrichum lindemuthianum* é um fitopatógeno hemibiotrófico, responsável pela antracnose do feijoeiro-comum (*Phaseolus vulgaris* L.). Entre as características mais marcantes do *C. lindemuthianum* destacam-se o alto poder destrutivo e sua ampla variabilidade patogênica. O objetivo do presente trabalho foi realizar o estudo dos genomas pertencentes aos isolados 83.501 (raça 83) e A₂-2-3 (raça 89), de *C. lindemuthianum* por meio da análise comparativa, com o intuito de avaliar as suas características genômicas e identificar fatores de virulência envolvidos na infecção de plantas de feijoeiro. Considerando somente os dados de genomas obtidos com tecnologias de sequenciamento de primeira e segunda geração, *C. lindemuthianum* apresenta o maior percentual de elementos genéticos móveis já predito para os genomas de *Colletotrichum* spp. Esta espécie também apresentou os maiores índices de ocorrência de *repeat-induced point mutation* (RIP) dentre todas as espécies de *Colletotrichum* avaliadas, possuindo os genes que codificam as metiltransferases RID e DIM-2, que participam deste mecanismo silenciamento. *Colletotrichum lindemuthianum* possui toda a maquinaria necessária a ocorrência do ciclo sexual, no entanto, o locus *Mat1-1* não foi encontrado em seu genoma, bem como no genoma das demais espécies do gênero avaliadas. Foram identificados 55 genes no genoma do isolado 83.501 que estão ausentes no genoma do isolado-A₂-2-3 e 41 genes do isolado A₂-2-3 que estão ausentes no genoma do isolado 83.501. Os genomas de *C. lindemuthianum* apresentam alterações significativas no número de cópias de genes pertencentes a 20 famílias de P450 mono-oxigenases, 10 famílias de CAZymes envolvidas na degradação de material vegetal e em 99 tipos de transportadores quando comparados às demais espécies de *Colletotrichum*. Dentre as espécies do gênero *Colletotrichum* analisadas, o fungo *C.*

lindemuthianum é a espécie com as características genômicas mais distintas.

Palavras chave: patogenicidade, adaptação, secretoma.

ABSTRACT

CORREIA, Hilberty Lucas Nunes D.Sc., Universidade Federal de Viçosa, December, 2020. Comparative Genomics **Comparative genomics enables the identification of virulence factors in *Colletotrichum lindemuthianum* genome**. Advisor: Marisa Vieira de Queiroz. Co-advisors: Pedro Marcos Pereira Vidigal, Mateus Ferreira Santana.

Colletotrichum lindemuthianum is the etiological agent of anthracnose disease in common bean (*Phaseolus vulgaris* L.). Among the remarkable characteristics of this specie stands out its high damaging power and huge pathogenic variability. The goal of this work was to identify the striking genomic characteristics and virulence factors present in the genomes of two *C. lindemuthianum* isolates (83.501 and A2-2-3). To achieve this goal comparative genomic analysis was performed. Considering just second generation genomic data, *C. lindemuthianum* presents the highest mobile elements content of all *Colletotrichum* species. The highest repeat-induced point mutation (RIP) indexes and genes encoding RID and DIM-2, both important to the RIP silencing mechanism, were also found in *C. lindemuthianum* genome. Interestingly, despite the occurrence of the complete machinery required to the sexual cycle, the locus MAT-1-1 was not found, resembling other *Colletotrichum* species. In this work we identified 55 genes in the 83.501 isolate that are absent in the A2-2-3 isolate genome and 41 genes in the A2-2-3 isolate absent in the 83.501 isolate genome. The *C. lindemuthianum* genomes presents changes in the copy number of genes belonging to the 20 families of P450-monoxygenases, 10 families of Cazymes with action in plant material degrading as well as in 99 different type of transporters when compared to other *Colletotrichum* species. Together, these results show that *C. lindemuthianum* has the more distinctly genomic characteristics among the *Colletotrichum* genus species.

Keywords: Adaptation, Secretome, Patogenicity.

INTRODUÇÃO

O gênero *Colletotrichum* (*Sordariomycetes*, *Hypocreomycetidae*; *Glomerellaceae*; *Glomerellales*) apresenta espécies de fungos saprofitas, endofíticas e fitopatogênicas e possui 189 espécies catalogadas no banco de dados Q-Bank (<http://www.q-bank.eu/fungi/>, junho de 2018). Dentre as espécies do gênero *Colletotrichum*, destacam-se as fitopatogênicas. Espécies deste gênero foram classificadas em oitavo lugar em uma pesquisa realizada entre patologistas de fungos sobre os 10 fungos fitopatogênicos mais importante do mundo (Dean et al. 2012).

A taxonomia do gênero *Colletotrichum* ainda vem passando por reformulações e até o momento é bem estabelecido que as espécies são agrupadas em nove clados monofiléticos principais, referidos como agregados ou complexos de espécies (*acutatum*, *graminicola*, *spaethianum*, *destructivum*, *dematium*, *gloeosporioides*, *boninense*, *truncatum*, e *orbiculare*) (Dean et al., 2012).

O fungo *Colletotrichum lindemuthianum* (Sacc. & Magnus) Briosi & Cavara é um fitopatógeno hemibiotrófico, agente etiológico da antracnose do feijoeiro-comum (*Phaseolus vulgaris* L.). Esta espécie apresenta uma alta variabilidade genética, com pelo menos 182 raças fisiológicas identificadas em diversas regiões do mundo (Padder et al., 2017). Assim, *C. lindemuthianum* tem sido explorado ao longo dos anos em diversas pesquisas acerca da variabilidade genética e patogênica (Mahuku & Riascos, 2004; Sharma et al., 2007; Ishikawa et al., 2010; Pinto et al., 2012). Além disto, este fungo também tem sido explorado em outras áreas, como em estudos sobre a citologia e fisiologia da infecção do feijoeiro (O'Connell et al., 1985; O'Connell, 1987; O'Connell and Ride, 1990; O'Connell et al., 1996); anastomose de conídios e parassexualidade (Roca et al., 2003; Ishikawa et al., 2010; Rosada et al., 2010; Ishikawa et al., 2013); reprodução sexuada (Souza et al., 2010); expressão

diferencial de genes durante a patogênese (Fontenelle et al., 2017); identificação de genes requeridos para a patogenicidade (Parisot et al., 2002; Pellier et al., 2003; Hoi et al., 2007; Cnossen-Fassoni et al., 2013; Soares et al., 2014; Nogueira et al., 2019), análise de mtDNA e de genes que codificam proteínas efetoras (de Queiroz et al., 2018; de Queiroz et al., 2019), entre outros.

Estudos genômicos têm contribuído para o entendimento dos mecanismos de infecção e de adaptação no gênero *Colletotrichum*. Até o momento, os genomas de 38 espécies de *Colletotrichum* encontram-se disponíveis nos bancos de dados do NCBI (<https://www.ncbi.nlm.nih.gov/genome/?term=colletotrichum>), dos quais 13 foram anotados (O'Connel et al., 2012; Gan et al., 2013; Baroncelli et al., 2016; Gan et al., 2016; Hacquard et al., 2016; Dallery et al., 2017; Rao & Nandineni, 2017; Liang et al., 2018). Os genomas de *Colletotrichum* spp. variam de 46,75 (*Colletotrichum sublioneola*) a 109 Mb (*Colletotrichum trifolii*), com um conteúdo G+C variando de 37,5 (*Colletotrichum orbiculare*) a 55,1% (*Colletotrichum graminicola*) e de sequências repetitivas cobrindo de 0,75 (*Colletotrichum gloeosporioides*) a 17,7% (*Colletotrichum higginsianum*) do genoma. O número de proteínas codificadas nos genomas das espécies de *Colletotrichum* varia de 12.006 em *C. graminicola* a 17.827 em *Colletotrichum fruticola*, das quais 11,55 a 15,99% são secretadas.

O complexo orbiculare contém espécies de fungos hemibiotróficos, causadores de doenças em plantas herbáceas. Dentre estas, as espécies *C. orbiculare*, *C. lindemuthianum*, *C. trifolii*, *Colletotrichum spinosum* e *Colletotrichum sidae*, já tiveram seus genomas sequenciados (de Queiroz et al., 2017; Gan et al., 2019). Essas espécies apresentam características distintas das demais espécies de *Colletotrichum* spp., tais como uma expansão no tamanho de seus genomas, que não é acompanhado pelo aumento no número de genes preditos e uma redução no conteúdo G+C (%) (Gan et al., 2013; de Queiroz et al., 2017; Gan et al., 2019). Apesar disto, somente as sequências de *C. orbiculare* e de *C. lindemuthianum*

encontram-se como de domínio público para análises.

Várias famílias gênicas são mais abundantes em genomas de fungos do gênero *Colletotrichum* do que em outros fungos fitopatogênicos. Enzimas ativas em carboidratos (CAZymes) com atividade na degradação de celulose e pectina, e as famílias das carboidrato-esterases e glicosil-hidrolases estão expandidas nos genomas de *Colletotrichum* spp. que infectam plantas dicotiledôneas (O'Connell et al., 2012; Gan et al., 2013; Baroncelli et al., 2016; Liang et al., 2018). Algumas subfamílias de CAZymes conhecidas como fatores de virulência, como por exemplo as proteínas contendo do domínio CBM50 (LysM) também foram demonstradas como expandidas em *Colletotrichum* spp. (Liang et al., 2018). Outras proteínas expandidas nos genomas de *Colletotrichum* spp. são algumas classes de proteases envolvidas na virulência de fitopatógenos (O'Connell et al., 2012; Gan et al., 2013; Gan et al., 2016, Rao & Nandineni, 2017). Além disso, espécies de *Colletotrichum* apresentam um grande potencial para a produção de metabólitos secundários, ultrapassando vários outros gêneros de fungos fitopatógenos (O'Connell et al., 2012; Gan et al., 2013; Baroncelli et al., 2016; Liang et al., 2018).

Os genomas sequenciados de *C. lindemuthianum* estão entre os maiores já reportados para os fungos do gênero *Colletotrichum* (de Queiroz et al., 2017), sendo superados somente pelo genoma da espécie *C. trifolii* (Gan et al., 2019). Devido à disponibilidade desses genomas e a importância de *C. lindemuthianum* como fitopatógeno, o objetivo do presente trabalho foi desvendar as suas características genômicas marcantes e identificar fatores de virulência envolvidos na infecção de plantas de feijoeiro por *C. lindemuthianum* através da genômica comparativa dos genomas de dois isolados de *C. lindemuthianum* (83.501 e A₂-2-3) com fungos das demais espécies de *Colletotrichum*.

MATERIAL E MÉTODOS

Genomas de *C. lindemuthianum*

Os genomas dos isolados 83.501 (NCBI *Biosample*: SAMN05235603) e A₂-2-3 (NCBI *Biosample*: SAMN05235602) de *C. lindemuthianum*, montados pelo software CLC Genomics Workbench 6.5.1[®] (QIAGEN) foram disponibilizados por de Queiroz et al. (2017). Estes genomas foram sequenciados em plataforma Illumina Hi-seq 2500 com PE de 100 pb, a coberturas de 89 e 157x para os isolados 83.501 e A₂-2-3, respectivamente (Laboratório Central de Tecnologias de Alto Desempenho em Ciências de Vida [LaCTAD]). No presente trabalho foi realizada a predição de genes para os genomas dos isolados 83.501 e A₂-2-3, utilizando o software Augustus (Stanke et al., 2004) previamente calibrado com o genoma de *C. orbiculare*.

Anotação Dos Genomas

A anotação dos genes que compõem os proteomas preditos para cada genoma foi realizada utilizando-se o algoritmo BLASTp (E-value 10⁻¹⁰) para buscar sequências apresentando similaridade no banco de dados Uniprot/SwissProt e Uniprot/Trembl (<http://www.uniprot.org>). Para a identificação dos genes que codificam RNAs transportadores e RNAs ribossomais foram utilizados os softwares tRNAScan-SE e RNAmmer (Lowe & Eddy, 1997; Lagesen et al., 2007).

Os elementos transponíveis (ETs) foram identificados utilizando-se o RepeatMasker Open 4.0 (Tarailo & Chen, 2009) com o banco de dados RepBase (Bao; Kojima; Kohany, 2015). Adicionalmente, foi realizada uma busca pelos retroelementos *TCg1* (Braga et al., 2014), *Retrocl1* (dos Santos et al., 2011), *CgT1* (He et al., 1996) e *Cgret* (Zhu & Oudemans, 2000), previamente descritos em fungos do gênero *Colletotrichum*. Para tal, foi utilizada a ferramenta Blastn. Para a análise de *Repeat-induced point mutation* (RIP), um banco de dados *housemade* contendo sequências das proteínas RID (Freitag et al., 2002) e DIM2 (Kouzminova and Selker,

2001), que participam deste mecanismo de silenciamento, foi utilizado como consulta para comparação com sequências dos genomas de *Colletotrichum* spp. (Blastp E-vale 10^{-10}). Adicionalmente foi utilizado o software RIPCAL (Hane & Oliver, 2008) para o cálculo dos índices de RIP. Para a análise de correlação entre o tamanho dos genomas e número de genes preditos e tamanho dos genomas e conteúdo G+C (%), as informações dos genomas foram levantadas utilizando o software assembly-stats (disponível em: <https://github.com/sanger-pathogens/assembly-stats>).

Identificação de “Single Nucleotide Polimorfisms” (SNPS)

O genoma do isolado A₂-2-3 foi selecionado como referência para o mapeamento usando o algoritmo BWA-MEM do software BWA versão 0.7.17 (<http://bio-bwa.sourceforge.net/>) (Li & Durbin, 2009). Um marcador identificando cada isolado foi adicionado a cada arquivo do mapeamento. Em seguida, os arquivos do mapeamento foram processados utilizando as ferramentas SortSam, MarkDuplicates e BuildBamIndex do software Picard versão 2.18.27 (<https://github.com/broadinstitute/picard/>). Os polimorfismos foram preditos usando o software FreeBayes versão 1.2.0 (<https://github.com/ekg/freebayes>) (Garrison and Marth, 2012) com os seguintes parâmetros: minimum mapping quality: 20; minimum base quality: 20; minimum coverage: 20. Após a predição dos polimorfismos, os SNPs foram selecionados usando o software vcftools versão 0.16.15 (<https://vcftools.github.io/index.html>) e scripts shell.

Análise de sequências de DNA que codificam proteínas envolvidas no ciclo sexual

Foi realizada uma busca por sequências nucleotídeos que codificam proteínas envolvidas no ciclo sexual nos genomas de *C. lindemuthianum* que apresentam similaridade com as sequências disponibilizadas no banco de dados criado por Galagan et al.(2005), que classifica proteínas nas categorias: *mating process*, *mating signalling*, *fruiting body development*, *budding and fission yeast genes involved in*

mating, karyogamy and meiosis e core meiotic transcriptome conserved in Saccharomyces cerevisiae and Schizosaccharomyces pombe. Sequências duplicadas e que não foram encontradas nos bancos de dados quando foi realizada a triagem com os números de acesso disponibilizados pelos autores foram removidas da busca. Para esta busca foi utilizado o algoritmo Blastp (E-value 10^{-5}). Os resultados foram analisados manualmente, buscando-se diferenças no número de cópias e presença/ausência das sequências proteicas preditas analisadas (**Tabela Suplementar 1**).

Secretoma predito

Para a identificação do secretoma predito, os softwares Signal-P (Petersen et al., 2011) e Target-P (Emanuelsson et al., 2007) foram utilizados para a predição de peptídeos sinais nas sequências proteicas. As sequências de proteínas preditas contendo peptídeos sinais foram avaliadas quanto à sua localização subcelular utilizando o software WOLF-PSORT (Horton et al., 2007). As sequências que foram classificadas como extracelulares foram avaliadas quanto à presença de domínios transmembrana com os softwares TMHMM (Krogh et al., 2001) e Phobius (Käll et al., 2007), com o intuito de eliminar sequências codificadoras de proteínas com domínios transmembranares. As proteínas preditas como secretadas, extracelulares e que não apresentaram domínios transmembranares foram avaliadas quanto à presença de âncoras de glicosilfosfatidilinositol (GPI) utilizando o BIG-PI Fungi Predictor (Eisenhaber et al., 1998), sendo as positivas descartadas do conjunto de dados e as remanescentes classificadas como proteínas secretadas.

Análise de grupos de ortólogos

Para as análises de genômica comparativa foram utilizados os proteomas preditos das espécies *Arthrobotrys oligospora* (ADOT00000000.1), *Aspergillus fumigatus* (AAHF00000000.1), *Aspergillus nidulans* (AACD00000000.1), *Blumeria*

graminis (CAUH00000000.1), *Colletotrichum chlorophyti* (MPGH00000000.1), *Colletotrichum fiorinae* (JARH00000000.1), *Colletotrichum gloeosporioides* (GCA_000319635.1), *Colletotrichum graminicola* (ACOD00000000.1), *Colletotrichum higginsianum* (LTAN00000000.1), *Colletotrichum incanum* (JTLR00000000.1), *Colletotrichum nymphaeae* (JEMN00000000.1), *Colletotrichum orbiculare* (AMCV00000000.1), *Colletotrichum salicis* (JFFI00000000.1), *Colletotrichum simmondsii* (JFBX00000000.1), *Colletotrichum sublineola* (MQVQ00000000.1), *Fusarium graminearum* (AACM00000000.2), *Fusarium oxysporum* (AAXH00000000.1) e *Rhizoctonia solani* (JATN00000000.1). Para o agrupamento das sequências proteicas preditas em ortogrupos foi utilizado o software Orthofinder com o auxílio de *parsers* escritos em linguagem Perl. Adicionalmente, foi utilizado o software Geneious9[®] para análises filogenéticas utilizando o método de Neighbor-joining (Saitou & Nei, 1987), e a confiança dos ramos foi acessada pelo método de *bootstrap* com 4000 repetições.

Proteases e CAZymes preditas secretadas

Os bancos de dados MEROPS (Rawlings et al., 2017) e dbCAN (Yin et al., 2012) foram utilizados para a anotação de proteases e de CAZymes, respectivamente. Para tal, foi utilizada a ferramenta BLASTp (E-value 10^{-10}), empregando-se o secretoma predito de cada espécie (conforme previamente descrito) como consulta. A contagem das famílias de proteases e CAZymes presentes em cada espécie foi realizada utilizando-se *scripts housemade* escritos em linguagem Perl[®] e os gráficos gerados com o auxílio do Rstudio (RStudio Team, 2020) e Gnuplot (Janert, 2010). Foram anotadas como expansões e contrações somente aquelas alterações no número de cópias observadas em quatro ou menos espécies de *Colletotrichum* ou em casos onde observou-se incongruência entre os dois genomas de *C. lindemuthianum*.

Transportadores

Para a identificação dos transportadores foi utilizado o *Transporter Classification Database* (TCDB) (Saier et al., 2006). Para tal, com auxílio da ferramenta BLASTp (E-value 10^{-10}) os proteomas preditos de cada espécie analisada no trabalho foram comparados com sequências presentes no banco de dados completo do TCDB. A contagem de cada proteína do TCDB identificada nos proteomas das espécies foi realizada com o auxílio de um script em Perl® e os gráficos foram gerados com o auxílio do Rstudio (RStudio Team, 2020). Com o intuito de destacar as particularidades de *C. lindemuthianum*, foram anotadas as contrações e expansões de transportadores somente presente em no máximo quatro espécies de fungos dentre as analisadas ou em casos de incongruência entre os genomas de *C. lindemuthianum*.

Metabolismo secundário

As enzimas sintetases de peptídeos não ribossomais (NRPSs) envolvidas na biossíntese de metabólitos secundários foram identificadas realizando-se a comparação de sequências presentes nos proteomas preditos de cada fungo com um banco de dados contendo sequências de NRPSs conhecidas, por meio do algoritmo Blastp (E-value 10^{-10}). Posteriormente, utilizando o banco de dados *Conserved Domains Database* (CDD) (Marchler et al., 2016), as sequências protéicas positivadas pelo Blastp foram escaneadas para confirmar a presença de domínios típicos de NRPSs. Com o intuito de buscar ortólogos dos genes que codificam NRPSs em *C. lindemuthianum* foi realizada a análise filogenética das NRPSs de todas as espécies de *Colletotrichum* estudadas no presente trabalho, além de separadamente a análise filogenética das NRPSs dos isolados de *C. lindemuthianum* 83.501, A₂-2-3 e *C. orbiculare*. Foram adicionadas ao alinhamento as sequências apresentando alta similaridade (Blastp NCBI E-value 10^{-10}) a cada NRPS de *C. lindemuthianum*. Os genes que codificam P450 mono-oxigenases em *Colletotrichum*

spp. foram identificados submetendo os proteomas preditos a uma busca por sequências similares presentes no banco de dados Cytochrome P450 Homepage (Nelson, 2009), utilizando-se o algoritmo Blastp (E-value 10^{-5}). Para as proteínas do grupo das P450 mono-oxigenases foram anotadas somente expansões e contrações observadas em quatro ou menos espécies de *Colletotrichum* ou em casos onde observou-se incongruência entre os dois genomas de *C. lindemuthianum*. Para a identificação dos genes que codificam policetídeo sintases foi empregado o software antismash (Weber et al., 2015).

RESULTADOS

Estrutura Geral dos Genomas de *C. lindemuthianum*

Os genomas dos isolados 83.501 e A₂-2-3 de *C. lindemuthianum* (97,4 Mb e 99,16 Mb) estão entre os maiores reportados para o gênero, sendo superados somente pela espécie *C. trifolii*, que apresenta um genoma de 109,66 Mb (de Queiroz et al., 2017; Gan et al., 2019). No entanto, para os fungos do gênero *Colletotrichum* não foi observada uma correlação entre o tamanho de seus genomas e o número de genes preditos (**Figura 1**).

O genoma de *C. lindemuthianum* apresenta um baixo percentual de G+C (37,6 [83.501] e 37,3% [A₂-2-3]), quando comparado aos genomas das outras espécies do gênero *Colletotrichum* (**Figura 2**). Foi possível identificar que existe uma correlação entre a expansão no tamanho dos genomas de *Colletotrichum* spp. com a diminuição de seu conteúdo G+C (%) (**Figura 2**). Além disso, foi notada a separação das espécies de *Colletotrichum* em basicamente dois grupos: um contendo as espécies de fungos com genomas menores e mais compactados, com alto conteúdo G+C (%) e outro grupo contendo genomas maiores com um menor conteúdo G+C (%), sendo este último referente às espécies do clado orbiculare.

A partir das predições de ETs observou-se que os isolados de *C. lindemuthianum* 83.501 e A₂-2-3 apresentam 5,95 % e 6,18%, respectivamente, de seu genoma consistindo ETs (Tabelas Suplementares 2 e 3). Do total de sequências repetitivas identificadas nestes genomas, a maior porção corresponde a ETs (2,81% e 2,91%, do tamanho total dos genomas dos isolados 83.501 e A₂-2-3, respectivamente). Os elementos da Classe I são os mais abundantes nos genomas de ambos os isolados, cobrindo 2,35% do genoma do isolado 83.501 e 2,48% do genoma do isolado A₂-2-3.

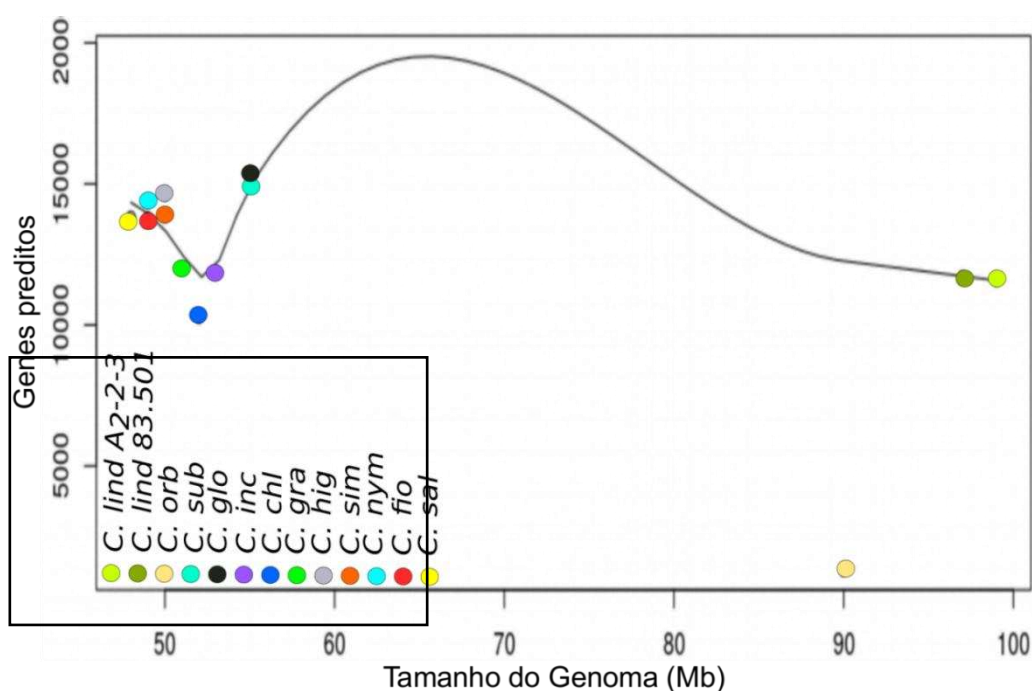


Figura 1. Relação entre o tamanho dos genomas de *Colletotrichum* spp. e o número de genes preditos em seus genomas. *C. lind A2-2-3* e *C. lind 83.501* – *C. lindemuthianum*; *C. orb* – *C. orbiculare*; *C. sub* – *C. sublineola*; *C. glo* – *C. gloeosporioides*; *C. inc* – *C. incanum*; *C. chl*: *C. chlorophyti*; *C. gra* – *C. graminicola*; *C. hig* – *C. higginsianum* *C. nym* – *C. nymphaeae*; *C. fio* – *C. fiorinae*; *C. sal* – *C. salicis*.

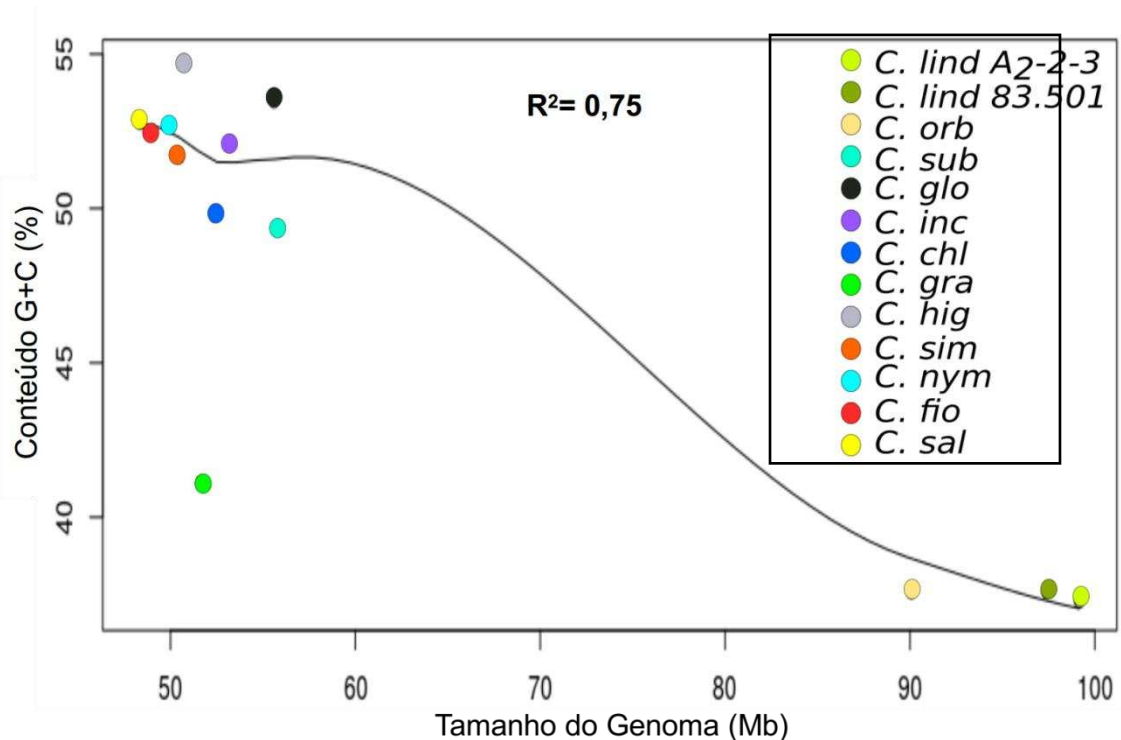


Figura 2. Correlação entre o tamanho dos genomas de *Colletotrichum* e aumento das regiões ricas em A+T. *C. lind A2-2-3* e *C. lind 83.501* – *C. lindemuthianum*; *C. orb* – *C. orbiculare*; *C. sub* – *C. sublineola*; *C. glo* – *C. gloeosporioides*; *C. inc* – *C. incanum*; *C. chl* - *C. chlorophyti*; *C. gra* – *C. graminicola*; *C. hig* – *C. higginsianum* *C. nym* – *C. nymphaeae*; *C. fio* – *C. fiorinae*; *C. sal* – *C. salicis*.

Dentre os ETs identificados nos genomas de *C. lindemuthianum*, os retroelementos da ordem LTR são os mais abundantes e entre estes, os elementos da superfamília *Gypsy* são mais representados nos genomas de ambos os isolados. Os retroelementos da superfamília *Copia* consistem no segundo grupo mais abundante nos genomas de ambos os isolados. Os elementos da Classe II mais abundantes nos dois genomas de *C. lindemuthianum* pertencem à superfamília *Tc1-Mariner*.

Por meio da busca utilizando a ferramenta tBlastn foram identificadas regiões nos genomas de *C. lindemuthianum* com similaridade às sequências das regiões *gag* e *pol*, contidas no retroelemento *Cgret* (Zhu & Oudemans, 2000). Enquanto os *hits* relacionados à região *gag* apresentam-se incompletos, cobrindo somente a porção desta região, os *hits* relacionados à região *pol* apresentaram cobertura de até 90%.

A busca pelos elementos da Classe II revelou que os genomas de *C. lindemuthianum* apresentam trechos curtos de similaridade com as porções terminais finais dos elementos *Tcg1* e *Cgt1* (Nourse et al., 1996; Braga et al., 2014). Observou-se que existem ao menos 160 sequências apresentando similaridade à sequência do elemento *RetroCl1* (Santos et al., 2012) no genoma do isolado A₂-2-3 e 151 sequências no genoma do isolado 83.501, a maioria das quais apresenta uma cobertura de 100% de sua sequência.

Os resultados da análise de distribuição do conteúdo G+C (%) nos genomas de *C. lindemuthianum* analisados demonstram que esta espécie possui um genoma bipartido, com regiões ricas em AT em mais da metade dos seus genomas (aproximadamente 55% em ambos os isolados) (**Figura 3**). O mesmo padrão foi observado para *C. orbiculare*, que por sua vez apresenta aproximadamente 50% de seu genoma formado por regiões ricas em AT. Para as demais espécies de *Colletotrichum* avaliadas no presente trabalho, não foi observada uma distribuição de conteúdo G+C(%) típica de genomas bipartidos (**Figura 3**).

Repeat Induced Point (RIP)

Dados os indícios de ocorrência de RIP nos genomas de *C. lindemuthianum*, foi realizada uma busca pelos genes que codificam as metiltransferases RID e DIM-2 (**Tabela 1**). O fungo *C. lindemuthianum*, assim como todas as demais espécies de *Colletotrichum* avaliadas, possui genes que codificam tanto RID quanto DIM-2. Posteriormente, foi calculado o índice *TpA/ApT*, um indicador da ocorrência de RIP. Os resultados (**Tabela 2**) demonstram que *C. lindemuthianum* apresenta fortes assinaturas de RIP em seu genoma, com valores *TpA/ApT* de aproximadamente 1,3 em ambos os isolados. Valores *TpA/ApT* acima de 0,89 também foram encontrados para as espécies *C. sublioneola* (1,21), *C. graminicola* (1,09) e *C. orbiculare* (1,20).

SNPs em Sequências Codificadoras de Efetores

Uma descrição detalhada acerca dos genes que codificam proteínas efetoras foi realizada anteriormente (de Queiroz et al., 2019). Foram identificados 356 e 353 genes que codificam proteínas candidatas a efetores nos isolados 83.501 e A₂-2-3, respectivamente. Visto que os dois genomas de *C. lindemuthianum* estudados no presente trabalho pertencem a isolados de duas raças patogênicas distintas, foi realizada uma análise de SNPs (*Single nucleotide polymorphisms*). Utilizando-se o genoma do isolado A₂-2-3 como referência foi possível identificar 156.758 *loci* apresentando polimorfismos, dos quais 18.842 encontram-se dentro de genes contidos nestes genomas. As análises indicam que os genes que codificam proteínas candidatas a efetores identificados por de Queiroz et al. (2019) apresentam uma frequência de ocorrência de SNPs (0,0021 e 0,0009) maior que a frequência encontrada para a totalidade de genes presentes nestes genomas (**Figura 4**). Quando os dois genomas de *C. lindemuthianum* foram comparados, foram encontradas 313.516 substituições do tipo C:T, que são substituições típicas do mecanismo de silenciamento RIP, das quais 37.704 encontram-se em regiões gênicas. Nos genes que codificam proteínas candidatas a efetores foram encontradas 930 mutações do tipo C:T, sendo a frequência destas quatro vezes maior do que o encontrado no conjunto total de genes destes genomas (0,004242952 em genes de efetores e 0,001948745 no conjunto total de genes). Além disto, observou-se que os genes que codificam proteínas candidatas a efetores em *C. lindemuthianum* estão mais próximos a regiões ricas em A+T do que os demais genes (**dados não apresentados**).

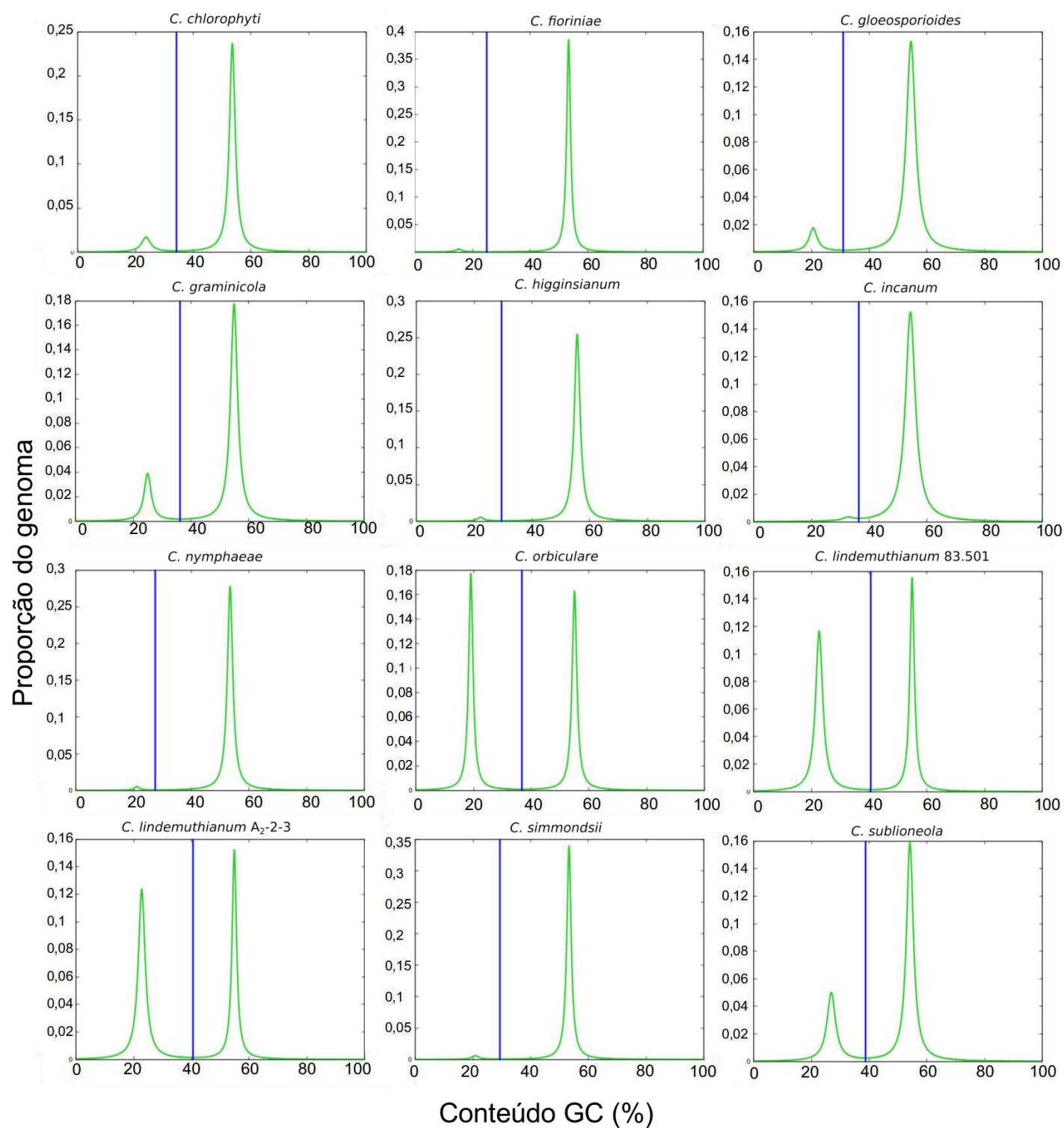


Figura 3. Distribuição do conteúdo GC dos genomas das espécies de *Colletotrichum* spp. Barras azuis indicam o limiar de divisão dos genomas em regiões ricas em A+T (à esquerda da barra) / regiões ricas em G+C (à direita da barra).

Tabela 1. Metiltransferases envolvidas em *repeat-induced point mutation* (RIP) nos genomas de diferentes fungos

Espécie	DIM2	RID	Acesso NCBI
<i>Aspergillus fumigatus</i>	-	+	ADOT00000000.1
<i>Arthrobotrys oligospora</i>	-	+	AAHF00000000.1
<i>Aspergillus nidulans</i>	-	+	AACD00000000.1
<i>Blumeria graminis</i>	-	-	CAUH00000000.1
<i>Colletotrichum chlorophyti</i>	+	+	MPGH00000000.1
<i>C. fioriniae</i>	+	+	JARH00000000.1
<i>C. gloeosporioides</i>	+	+	GCA_000319635.1
<i>C. graminicola</i>	+	+	ACOD00000000.1
<i>C. higginsianum</i>	+	+	LTAN00000000.1
<i>C. incanum</i>	+	+	JTLR00000000.1
<i>C. lindemuthianum 83.501</i>	+	+	JEMN00000000.1
<i>C. lindemuthianum A₂-2-3</i>	+	+	AMCV00000000.1
<i>C. nymphaeae</i>	+	+	JFFI00000000.1
<i>C. orbiculare</i>	+	+	JFBX00000000.1
<i>C. salicis</i>	+	+	MQVQ00000000.1
<i>C. simmondsii</i>	+	+	JFBX00000000.1
<i>C. sublioneola</i>	+	+	MQVQ00000000.1
<i>Fusarium graminearum</i>	+	+	AACM00000000.2
<i>Fusarium oxysporum</i>	+	+	AAXH00000000.1
<i>Rhizoctonia solani</i>	+	+	JATN00000000.1

(+) indica a presença do gene-e (-) indica a ausência do gene.

Tabela 2. Índices de *repeat-induced point mutation* (RIP) calculados para as espécies de *Colletotrichum*

Espécies	Índice RIP (TpA/ApT)	Acesso NCBI
<i>C. fioriniae</i>	0,713506903	JARH00000000.1
<i>C. gloesporioides</i>	0,851873112	GCA_000319635.1
<i>C. graminicola</i>	1,090737931	ACOD00000000.1
<i>C. higginsianum</i>	0,734752775	LTAN00000000.1
<i>C. incanum</i>	0,762483086	JTLR00000000.1
<i>C. lindemuthianum</i> 83.501	1,291407058	JEMN00000000.1
<i>C. lindemuthianum</i> A2-2-3	1,297013071	AMCV00000000.1
<i>C. nymphaeae</i>	0,675952016	JFFI00000000.1
<i>C. orbiculare</i>	1,209194398	JFBX00000000.1
<i>C. salicis</i>	0,682497348	MQVQ00000000.1
<i>C. simmondsii</i>	0,728485131	JFBX00000000.1
<i>C. sublineola</i>	1,211824617	MQVQ00000000.1

Análise de Proteínas Envolvidas no Ciclo Sexual

A descrição detalhada das proteínas envolvidas no ciclo sexual identificadas nas espécies avaliadas é apresentada na **Tabela Suplementar 1**. O número de genes que codificam proteínas envolvidas no ciclo sexual identificadas em *C. lindemuthianum* foi de 143 nos isolados 83.501 e A₂-2-3 (**Figura 5**). Em ambos genomas, não foi possível detectar o gene que codifica uma proteína homóloga a MAT1-1-1 (domínio α), estando presentes somente o gene que codifica uma proteína homóloga a MAT1-2-1 (HMG-box). Dentre as demais espécies de *Colletotrichum* foram encontradas de 142 (*C. fioriniae*) a 145 (*C. orbiculare* e *C. graminicola*) proteínas relacionadas como envolvidas na reprodução sexuada. Não foi possível

identificar ortólogos de MAT1-1-1 também nos genomas das outras espécies pertencentes a este gênero.

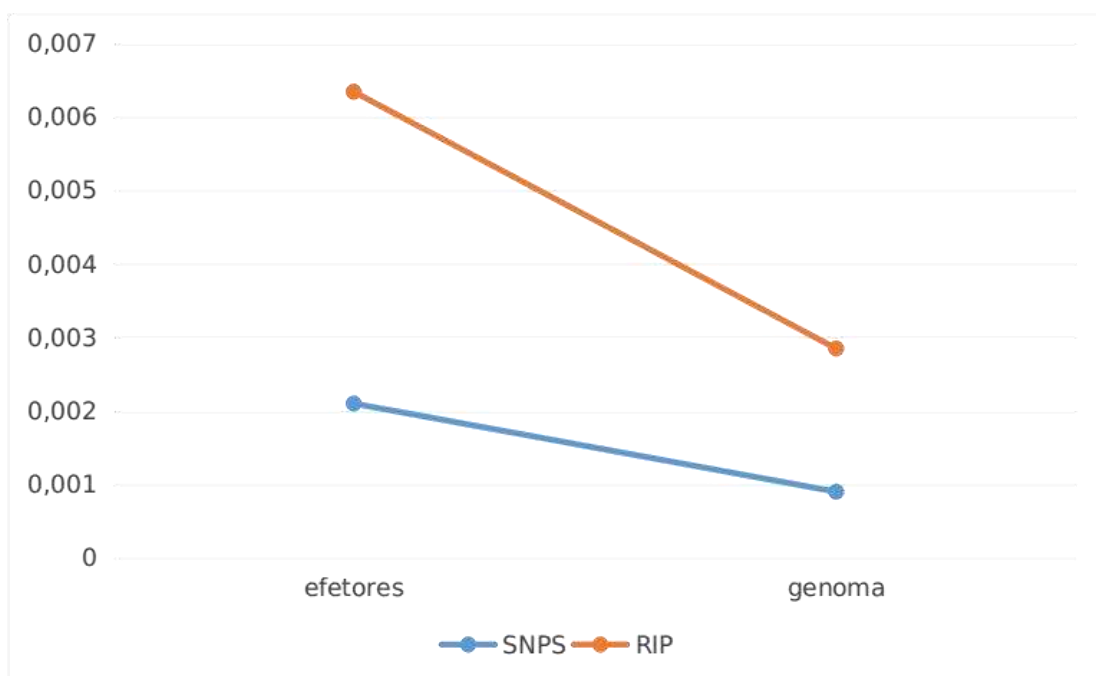


Figura 4. Frequência de ocorrência de SNPS (*Single nucleotide polymorphisms*) e substituições típicas de *repeat-induced point mutation* (RIP) em todos os genes e em genes que codificam efetores entre os isolados de *C. lindemuthianum* 83.501 e A₂-2-3.

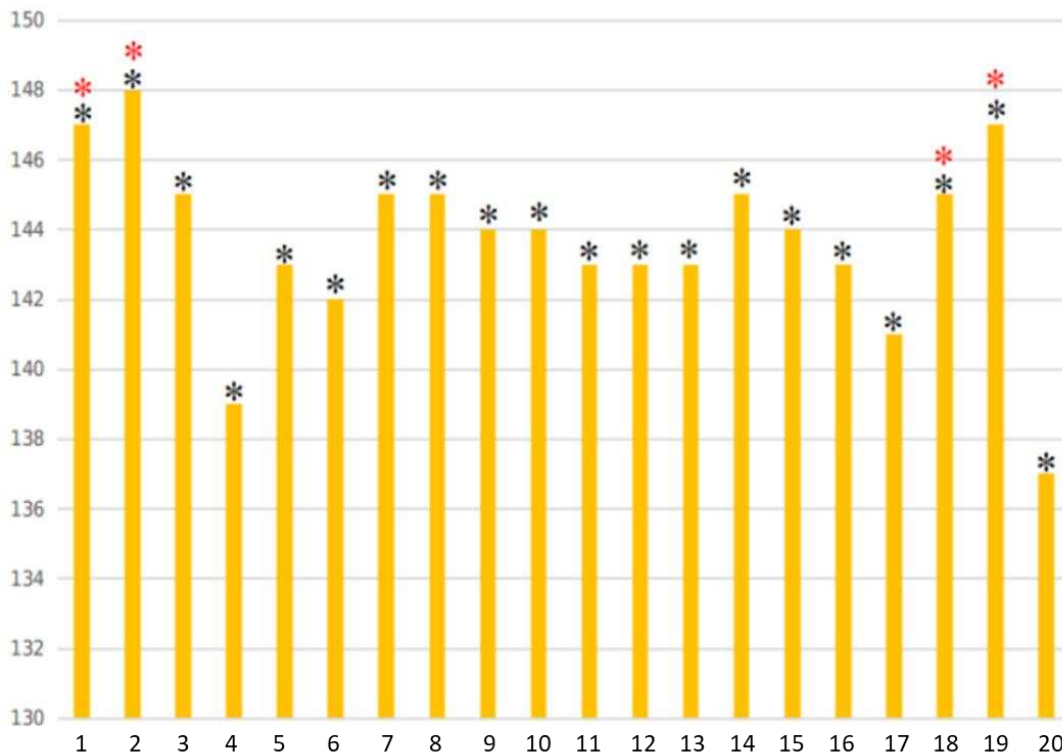


Figura 5. Número de genes que codificam proteínas envolvidas no ciclo sexual por genomas fúngicos. Asteriscos pretos indicam a presença de proteínas com homologia a MAT1-2-1 e asteriscos vermelhos indicam a presença de homólogos de MAT1-1-1. 1 – *Aspergillus fumigatos*; 2 – *Aspergillus nidulans*; 3 – *Arthrotritys oligospora*; 4 - *Blumeria graminis*; 5 – *Colletotrichum chlorophyti*; 6 – *C. fiorinae*; 7 – *C. gloeosporioides*; 8 – *C. graminicola*; 9 – *C. higginsianum*; 10 – *C. incanum*; 11 -*C. lindemuthianum* 83.501; 12 - *C. lindemuthianum* A₂-2-3; 13 – *C. nymphaeae*; 14 – *C. orbiculare*; 15 – *C. salicis*; 16 – *C. simmondsii*; 17 – *C. sublineola*; 18 - *Fusarium graminearum*; 19 - *F. oxysporum*; 20 - *Rhizoctonia solani*.

Já em relação a outras proteínas envolvidas no processo de *mating* que foram avaliadas, 9 das 10 que foram avaliadas encontram-se codificadas nos genomas dos dois isolados de *C. lindemuthianum*, estando ausente somente o gene que codifica a proteína PpgA (Dyer et al., 2003), um precursor de feromônio (*alpha-factor like*). Este resultado foi comum a todas as demais espécies do gênero *Colletotrichum* e de outros fungos avaliados neste trabalho, sendo a única exceção o fungo *A. nidulans* que possui um gene que codificando esta proteína em seu genoma.

Em tratando-se dos genes envolvidos na sinalização de *mating*, ambos os isolados de *C. lindemuthianum* não apresentam em seus genomas os dois genes que codificam as proteínas receptoras PreA e PreB (Pöggeler & Kück, 2001). A ausência de ambos os receptores foi também observada nos genomas de *C. chlorophyti*, *C. fiorinae*, *C. nymphaeae*, *C. simmondsi* e *C. sublioneola*. Nos genomas das espécies *C. graminicola*, *C. higginsianum*, *C. incanum* e *C. orbiculare* foi somente encontrada genes que codificam homólogos de *preA*, enquanto no genoma de *C. salicis* foi possível identificar somente genes codificando homólogos de PreB. No genoma do fungo *C. gloeosporioides* foram encontrados genes que codificam ambas as proteínas receptoras. Para as demais espécies avaliadas foram encontradas cópias de ambos os genes, sendo exceções os fungos *R. solani* e *B. graminis*, que possuem os genes *preA* e *preB* seus genomas, respectivamente.

Quanto aos demais genes envolvidos no processo de sinalização de *mating* avaliados, dentre eles *steA* e *steC*, cujos produtos são requeridos para a ocorrência do ciclo sexual em *A. nidulans* (Vallim et al., 2000; Wei et al., 2003), todos estão presentes nos genomas de *C. lindemuthianum* e nas demais espécies avaliadas neste trabalho, sendo a única exceção *R. solani* que não apresentou o gene que codifica a proteína SteA em seu genoma.

Os genomas de *C. lindemuthianum* apresentam todos os genes que codificam proteínas envolvidas no desenvolvimento dos corpos de frutificação que foram procurados. Nesta categoria, a variação de presença/ausência observada foi mínima, restringindo-se ao gene que codifica a proteína Prot1†† que se encontra presente em algumas espécies e ausente em outras (**Tabela Suplementar 1**).

Comparação entre os Isolados de *C. lindemuthianum*

Do total de 11.673 genes preditos para o genoma do isolado 83.501, 11.618 possuem homologia com genes presentes no genoma do isolado A₂-2-3. Em relação ao genoma do isolado A₂-2-3, de seus 11.627 genes preditos, 11.592 apresentam homologia com genes presentes no genoma do isolado 83.501. Assim, foram encontrados 55 e 45 genes exclusivos dos isolados 83.501 e A₂-2-3, respectivamente. Três dentre estes, que estão presentes no genoma do isolado 83.501 codificam proteínas candidatas a efetores (g2809, g2615 eg9014) Dentre os genes exclusivos do isolado A₂-2-3, dois codificam proteínas candidatas a efetores (g1217 e g11064) (de Queiroz et al., 2019).

Dentre os genes exclusivos do isolado 83.501, 38 apresentam *hits* no banco de dados SWISS-PROT, incluindo as três sequências de genes que codificam proteínas candidatas a efetores, vinte e um possuem similaridade com sequências de proteínas não caracterizadas preditas para genomas de outros ascomicetos (**Tabela 4**). As demais sequências apresentam similaridade com alfa-beta hidrolases, arilesterases, proteínas contendo o domínio dedo C6, hidrolases de éster carboxílico, cadeia pesada da clatrina, *decaprenylcistransferase*, proteínas contendo o domínio DUF1640, componente de exocisto, proteína Hex2, transportadores MFS, fosfotransferases, proteína contendo o domínio RNB, proteína de biossíntese da tiamina, polimerase envolvida de interação com *trf1* e tRNA-metiltransferase.

Já em relação aos 45 genes exclusivos do isolado A₂-2-3, 31 apresentam *hits* com sequências no banco de dados SWISS-PROT (**Tabela 5**), não tendo sido encontrada qualquer sequência que codifique um dos candidatos a efetores preditos anteriormente. Dentre as proteínas apresentando similaridade com sequências no banco de dados, 22 são similares a sequências de proteínas preditas classificadas como não caracterizadas. Destas, 21 consistem em sequências preditas em genomas de fungos de outras espécies do gênero *Colletotrichum* e uma foi

identificada na diatomácea *Thalassiosira oceanica*. As 23 sequências que foram classificadas pelos bancos de dados apresentam similaridade a adenilato cinases, proteínas contendo o domínio CFEM, proteínas da família isomerase hidratase enoyl, hidrolases de epóxidos, proteínas M, proteínas relacionadas à resistência multidrogas, fatores de transcrição SPT4, protease da família Up1 e proteína de parede celular gp1.

Dentre as 14 sequências remanescentes, 8 também não apresentam similaridade (Blastp E-value 10^{-5}) com o banco de dados *Non-redundant protein sequences* do NCBI nem tampouco domínios preditos pelo Pfam. As sequências proteicas preditas para os genes S89_g10663, S89_g11505, S89_g1776, S89_g4456, S89_g497 e S89_g7287 apresentam similaridade com sequências proteicas depositadas no banco "*Non-redundant protein sequences*". As descrições das sequências apresentando similaridade com estas consistem em proteínas histidina cinases, proteínas hipotéticas, RNA polimerase fator sigma-54, oxidase de aldeído 4 isoformas X1 e X2 e proteases da família Ulp1.

Análise de Grupos de Ortólogos

Os proteomas preditos de todas as espécies avaliadas neste trabalho foram agrupados em ortogrupos. Das 267.991 sequências codificadoras das proteínas analisadas, 241.378 foram agrupados em 15.530 ortogrupos. O restante (26.613) das proteínas não foi atribuído a qualquer ortogrupo.

Tabela 4. Genes que não foram atribuídos a ortogrupos no genoma do isolado 83.501

Gene ID	Subject Description	Identity(%)	E-value	
S83_g11174	Aconitate hydratase [Colletotrichum spinosum]	81	6,00E-58	
S83_g5852	hypothetical protein FOC1_g10004364 [Fusarium oxysporum f. sp. cubense race 1]	100	2,00E-86	
S83_g75	No significant similarity found	-	-	
S83_g561	No significant similarity found	-	-	
S83_g1515	No significant similarity found	-	-	
S83_g1873	No significant similarity found	-	-	
S83_g3463	No significant similarity found	-	-	
S83_g4195	No significant similarity found	-	-	
S83_g5061	No significant similarity found	-	-	
S83_g5064	No significant similarity found	-	-	
S83_g5129	No significant similarity found	-	-	
S83_g5956	No significant similarity found	-	-	
S83_g7031	No significant similarity found	-	-	
S83_g8724	No significant similarity found	-	-	
S83_g8877	No significant similarity found	-	-	(-)
S83_g9014	No significant similarity found	-	-	significa
S83_g9117	No significant similarity found	-	-	que os
S83_g10218	No significant similarity found	-	-	valores
S83_g10867	No significant similarity found	-	-	em
S83_g11500	No significant similarity found	-	-	questão
S83_g11486	Oleate hydroxylase FAH12 [Colletotrichum trifolii]	100	2,0E-111	não
S83_g8830	putative methyltransferase tdiE [Colletotrichum trifolii]	98	2,0E-23	22

foram obtidos para tal sequência.

Tabela 5. Genes (BLASTp) que não foram atribuídos a ortogrupos no genoma do isolado A₂-2-3

Gene ID	Subject Description	Identity(%)	E-value
S89_g497	aldehyde oxidase 4 isoform X2 [Nannospalax galili]	38	8.9
S89_g11557	Bifunctional epoxide hydrolase 2 [Colletotrichum orbiculare MA...]	83	9e-24
S89_g3294	Glyoxylate reductase [Colletotrichum spinosum]	82	2e-81
S89_g11505	hypothetical protein CDV36_010325 [Fusarium kuroshium]	37	0.039
S89_g1873	hypothetical protein THAOC_31069 [Thalassiosira oceanica]	58	5e-14
S89_g4260	Multidrug resistance protein [Colletotrichum higginsianum IMI...]	-	-
S89_g3746	No significant similarity found.	-	-
S89_g5769	No significant similarity found.	-	-
S89_g5879	No significant similarity found.	-	-
S89_g7288	No significant similarity found.	-	-
S89_g11095	No significant similarity found.	-	-
S89_g11500	No significant similarity found.	-	-
S89_g11502	No significant similarity found.	-	-
S89_g11506	No significant similarity found.	-	-
S89_g3242	putative amino-acid permease [Colletotrichum spinosum]	72	3e-28
S89_g1776	Redox sensor histidine kinase response regulator devS...	38	4.3
S89_g4456	RNA polymerase sigma-54 factor [Deltaproteobacteria bacterium]	33	1.9
S89_g7366	vegetative cell wall protein gp1 [Colletotrichum fructicola Na...]	73	8e-11
S89_g4477	Zinc-type alcohol dehydrogenase-like protein	89	1e-68

(-) significa que os valores em questão não foram obtidos para tal sequência.

Dois mil novecentos e dois ortogrupos são compartilhados entre todas as espécies de fungos analisadas no presente trabalho, enquanto 1.169 ortogrupos consistem em ortogrupos com cópias únicas. Em relação a *C. lindemuthianum*, para os isolados 83.501 e A₂-2-3, 11.651 e 11.608 proteínas foram agrupadas em ortogrupos, respectivamente, resultando em 99,8% das proteínas atribuídas a ortogrupos nos dois isolados. Nove mil novecentos e dezenove e 9.882 ortogrupos foram atribuídas as proteínas dos isolados 83.501 e A₂-2-3, respectivamente. O fungo *C. lindemuthianum* compartilha mais ortogrupos com *C. orbiculare* (9.371) e menos com *R. solani* (4.472) (**Tabela 3**), sendo que dentre as espécies de *Colletotrichum*, *C. lindemuthianum* compartilha menos ortogrupos com a espécie *C. chlorophyti* (7741).

Duzentos e oitenta e um ortogrupos exclusivos do clado orbiculare foram encontrados nos genomas de *C. lindemuthianum* (isolados 83.501 e A₂-2-3) e *C. orbiculare*. Adicionalmente, foram identificados 231 ortogrupos exclusivos de *C. lindemuthianum*, sendo estes compartilhados entre os dois isolados. A anotação funcional e atribuição de categorias “GO” destas proteínas são apresentadas nas **Figuras 6 e 7**.

As proteínas pertencentes a ortogrupos específicos de *C. lindemuthianum* na categoria “Processo Biológico”, apresentam uma expansão em termos relacionados ao transporte transmembranar. A atividade de transporte também teve uma maior representação na categoria “Função Molecular”.

Adicionalmente, utilizando o Interproscan foi possível atribuir termos do *gene ontology* a 55 genes nos dois isolados de *C. lindemuthianum*. Em ambos os genomas o termo mais frequente foi o de transportadores MFS, seguido de proteínas contendo domínios ligação ao NAD(P) e P450 mono-oxigenases. Cinco e seis genes consistem em candidatos a efetores, nos isolados 83.501 (S83_g3214, S83_g3641, S83_g5701, S83_g6565 e S83_g9172) e A₂-2-3 (S89_g4574, S89_g574, S89_g3178 S89_g7582, S89_g11321 e S89_g9457).

Tabela 3. Número médio de ortogrupos de diferentes fungos compartilhados com *C. lindemuthianum*.

Espécies	Número de ortogrupos compartilhados	Acesso NCBI
<i>Arthrotrrys oligospora</i>	5720	JARH00000000.1
<i>Aspergillus fumigatus</i>	6224	GCA_000319635.1
<i>A. nidulans</i>	6423	ACOD00000000.1
<i>Blumeria graminis</i>	4694	LTAN00000000.1
<i>Colletotrichum chlorophyti</i>	7741	JTLR00000000.1
<i>C. fioriniae</i>	8643	JEMN00000000.1
<i>C. gloeosporioides</i>	8629	AMCV00000000.1
<i>C. graminicola</i>	8313	JFFI00000000.1
<i>C. higginsianum</i>	8295	JFBX00000000.1
<i>C. incanum</i>	8086	MQVQ00000000.1
<i>C. nymphaeae</i>	8668	JFBX00000000.1
<i>C. orbiculare</i>	9371	MQVQ00000000.1
<i>C. salicis</i>	8603	JARH00000000.1
<i>C. simmondsi</i>	8666	GCA_000319635.1
<i>C. sublineola</i>	8412	ACOD00000000.1
<i>Fusarium graminearum</i>	7380	LTAN00000000.1
<i>F. oxysporum</i>	7654	JTLR00000000.1
<i>Rhizoctonia solani</i>	4472	JEMN00000000.1

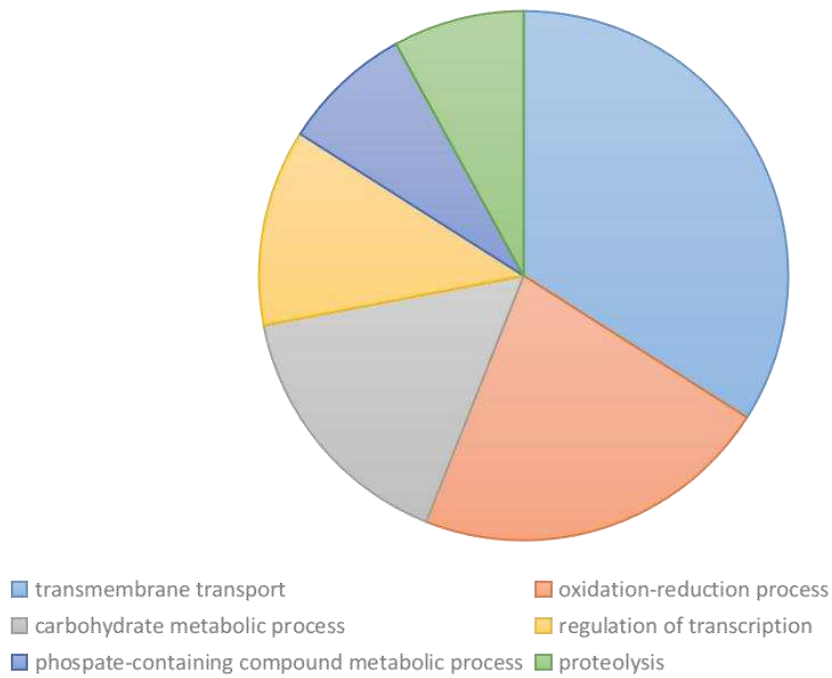


Figura 6. Anotação funcional dos ortogrupos exclusivamente encontrados nos genomas de *Colletotrichum lindemuthianum*. Genes do isolado A₂-2-3.

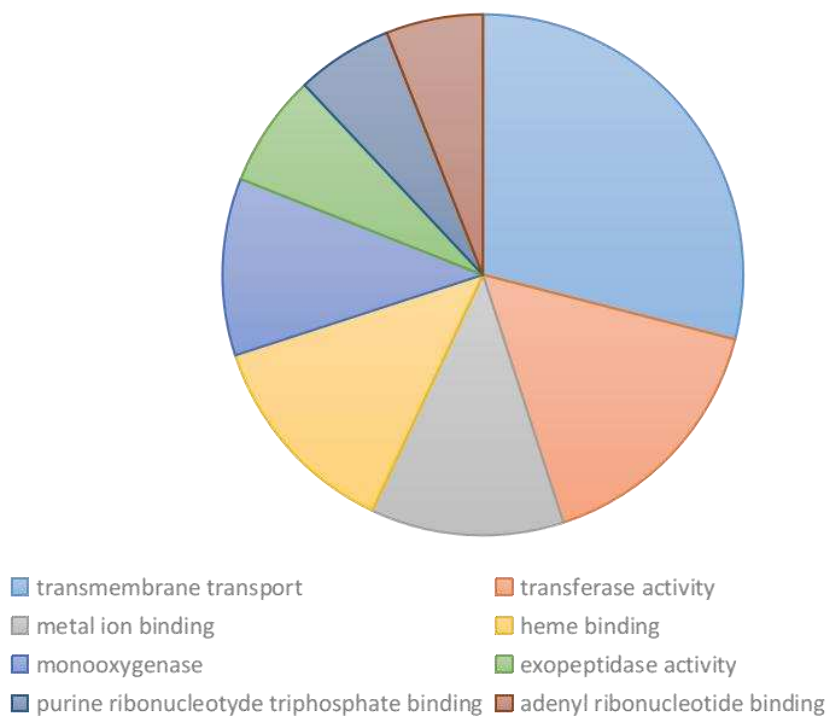


Figura 7. Anotação funcional dos ortogrupos exclusivos de *Colletotrichum lindemuthianum*. Genes do isolado 83.501.

Transportadores

Os genomas de *C. lindemuthianum* apresentam um dos menores números de genes que codificam proteínas preditas como transportadoras dentre os genomas dos fungos do gênero *Colletotrichum* analisados, superando somente a espécie *C. graminicola* neste termo (**Figura 8**). hospedeiros também apresentaram um maior número de genes que codificam proteínas transportadoras que as espécies com gamas mais restritas como é o caso de *C. lindemuthianum* e *C. graminicola*.

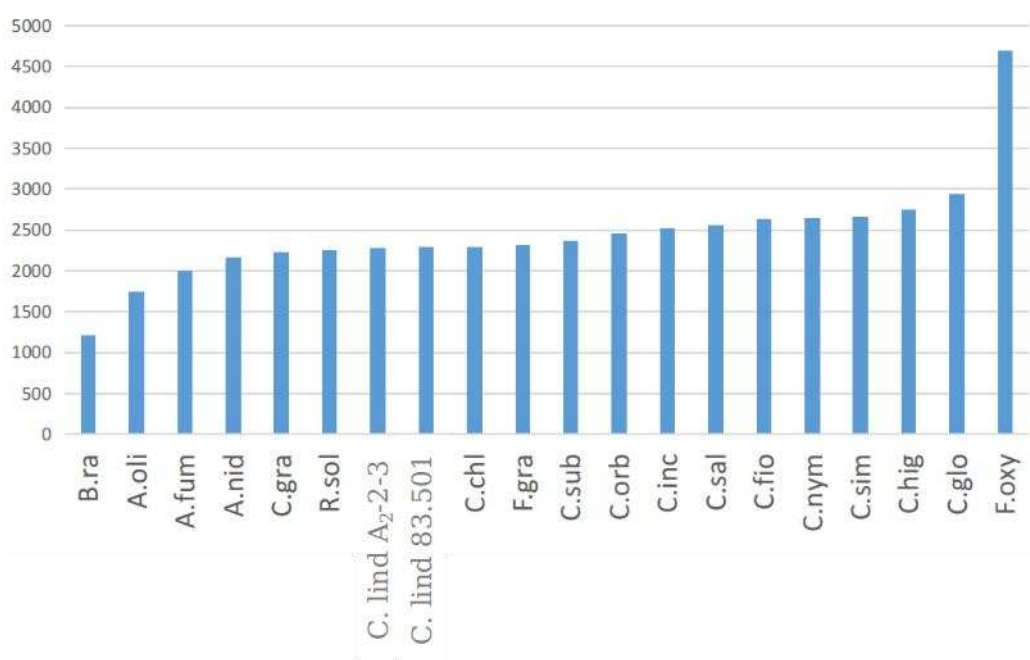


Figura 8. Número de transportadores MFS para cada genoma de fungo. Os transportadores MFS foram classificados utilizando-se o banco de dados TCDB com o auxílio da ferramenta Blastp. B.gra - *Blumeria graminis*; A.oli - *Arthrotrypis oligospora*; A.fum - *Aspergillus fumigatus*; A.nid - *Aspergillus nidulans*; C.gra - *C. graminicola*; R.sol - *Rhizoctonia solani*; C. lind 83.501 e C. lind A₂-2-3 - *C. lindemuthianum*; Cchl - *C. chlorophyti*; C.fio - *C. fiorinae*; C.glo - *C. gloeosporioides*; C.hig - *C. higginsianum*; C.inc - *C. incanum*; C.nym - *C. nymphaeae*; C.orb - *C. orbiculare*; C.sal - *C. salicis*; C.sim - *C. simmondsii*; C.sub - *C. sublineola*; F.gram - *Fusarium graminearum*; F.oxy - *F. oxysporum*.

Em relação às demais espécies de *Colletotrichum*, o número destes genes variou de 2.230 em *C. graminicola* a 2.942 no genoma de *C. gloeosporioides*, e foi possível observar que espécies de fungos como *C. gloeosporioides* e *F. oxysporum*, que possuem maiores gamas de

No presente trabalho foram identificadas 1.433 transportadores diferentes do TCDB, os quais encontram-se distribuídos entre os genomas analisados (**Figuras 9-22**). Em relação ao seu número de cópias, 99 destes apresentam-se contraídos ou expandidos nos genomas de *C. lindemuthianum*, dos quais 47 encontram-se contraídos e 52 expandidos (**Figuras 9-22, Tabela Suplementar 4**). Destas alterações no número de cópias, 80 são compartilhadas com mais espécies analisadas e 19 consistem em alterações exclusivas de *C. lindemuthianum* (presentes no genoma de um ou dos dois isolados).

Foi possível observar que os 99 transportadores expandidos ou contraídos nos genomas de *C. lindemuthianum* pertencem à 59 famílias distintas de transportadores, dentre as quais se destacaram as famílias de transportadores *Major Facilitator Superfamily* (MFS), *Mitochondrial Carrier Family* (MCM), *Drug/Metabolite Transporter Family* (DMT), *The Putative, Vectorial Glycosyl Polymerization Family* (VGP), *ATP-binding Cassete Superfamily* (ABC) e *Signaling Adaptor Protein KARP/DAP12/TYROBP Family* (SAP), sendo a superfamília MFS a mais representada dentre estas (12% do total de transportadores contraídos/expandidos) (**Figura 23**).

Dezenove dentre estes apresentam a contração ou expansão exclusivamente nos genomas de *C. lindemuthianum* (**Tabela Suplementar 4**), sendo que 11 estão contraídos e 8 expandidos em relação a seu número de cópias. A classificação dos mesmos revelou que estes pertencem a 13 famílias de transportadores, sendo mais frequentemente associados às famílias de transportadores *The Drug/Metabolite Transporter (DMT) Superfamily*, *The Major Facilitator Superfamily* (MFS),

Mitochondrial Carrier Family (MC), *The Huntington-interacting protein 14 (HIP14) Family*, *The Chloride Carrier/Channel (CIC) Family* e *The Transient Receptor Potential Ca²⁺ (TRP-CC) Family* (**Figura 24**). Dentre os substratos preditos como sendo carregados por estas proteínas estão: Cl⁻, H⁺, K⁺, Ca²⁺, radical superóxido, íons, açúcares ligados a nucleotídeos, nitrato, proteína PaaT, CMP-ácido siálico, Quinidina, Barban, Cisplatina, Bleomicina, AMP, ADP, ATP e fosfato.

Foram encontradas cinco alterações no número de cópias que divergem entre os genomas dos dois isolados de *C. lindemuthianum* (**Tabela Suplementar 4**). Dentre estas, duas consistem em contrações (transportadores 8.A.51.1.1 e 1.A.1.20.8), e três em expansões (3.A.1.121.19, 3.A.1.17.3 e 2.A.100.1.3). Enquanto os dois contraídos pertencem às famílias *The Dipeptidyl-aminopeptidase-like Protein 6 beta subunit of Kv4 channels (DPP6) Family* e *The Voltage-gated Ion Channel (VIC) Superfamily*, os transportadores expandidos que divergem entre os dois isolados de *C. lindemuthianum* pertencem às famílias ABC e *The ferroportin (Fpn) Family*.

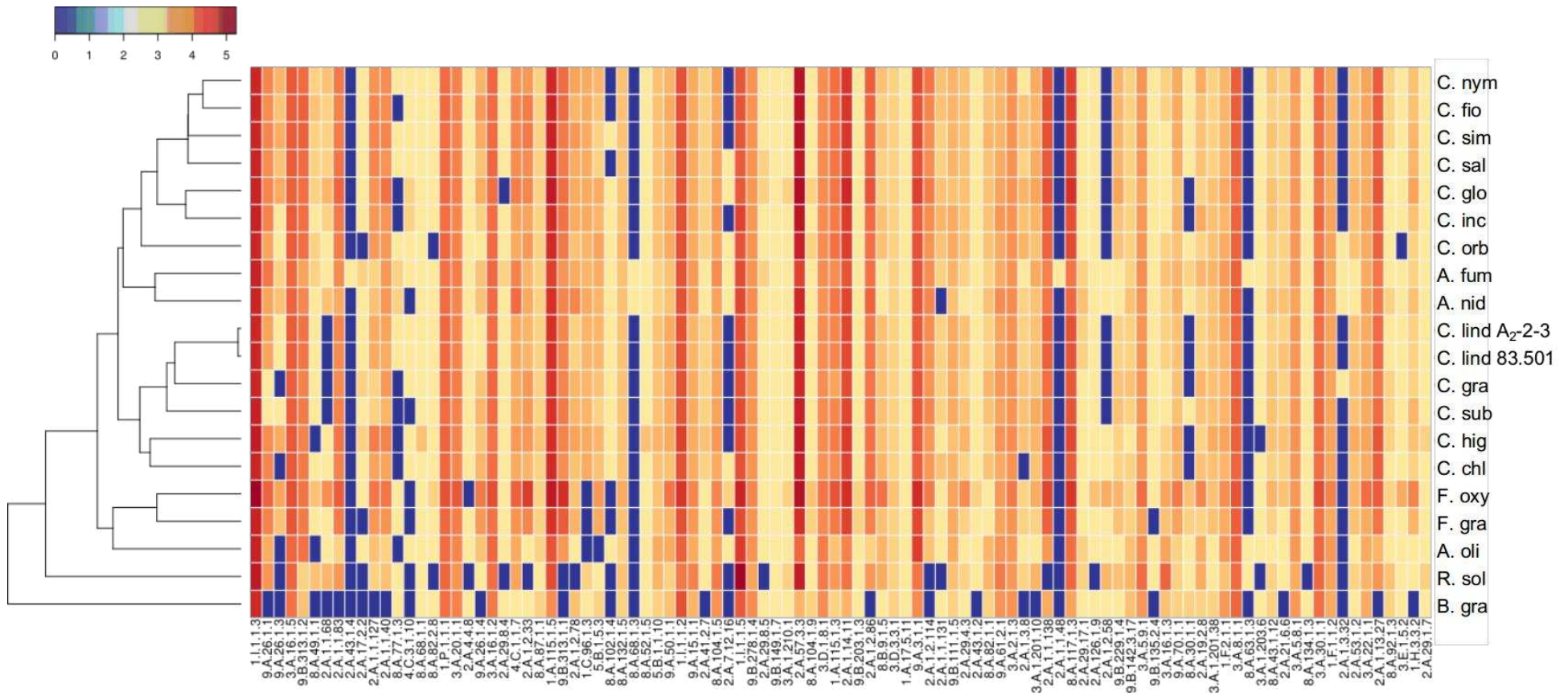


Figura 9. Contrações e expansões de famílias de transportadores em *Colletotrichum lindemuthianum*.

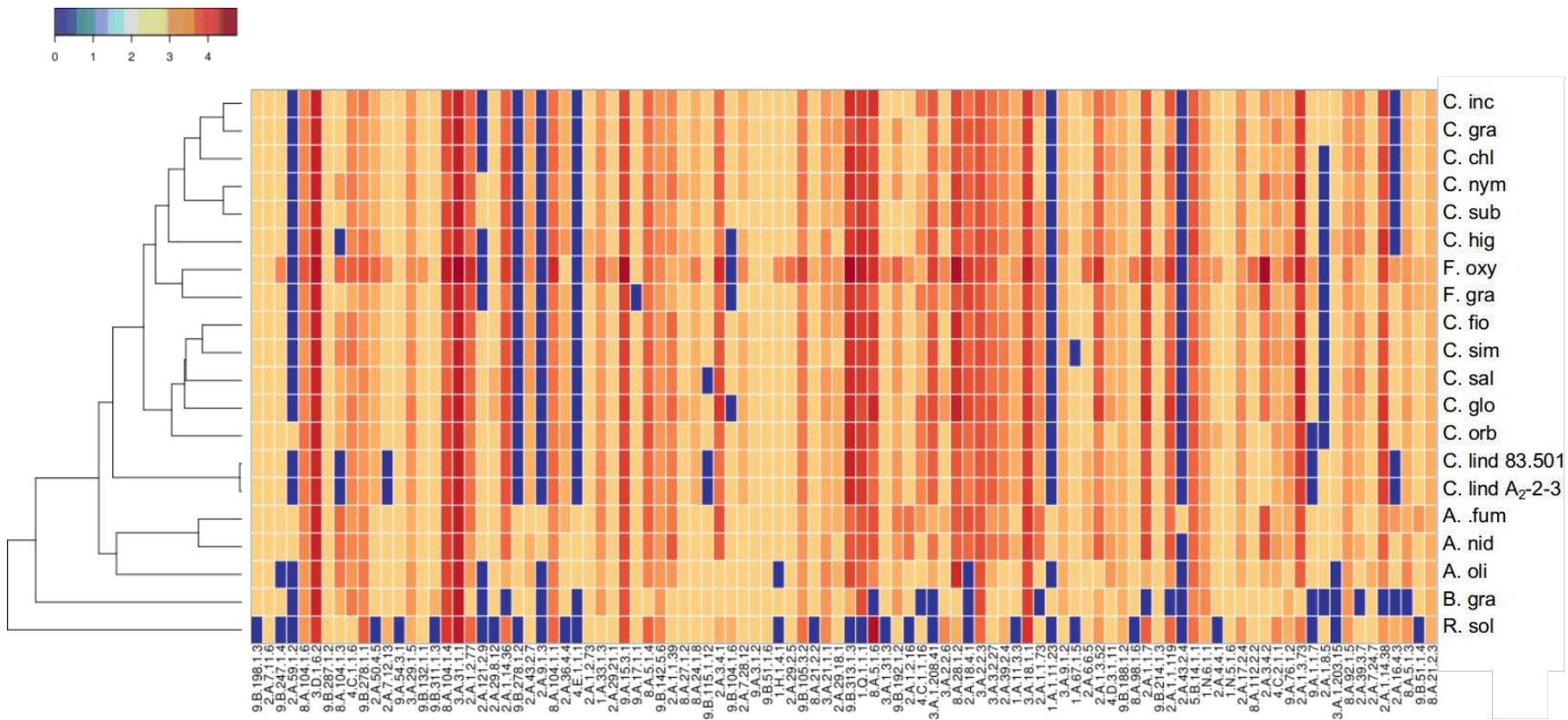


Figura 10. Contrações e expansões de famílias de transportadores em *C. lindemuthianum*.

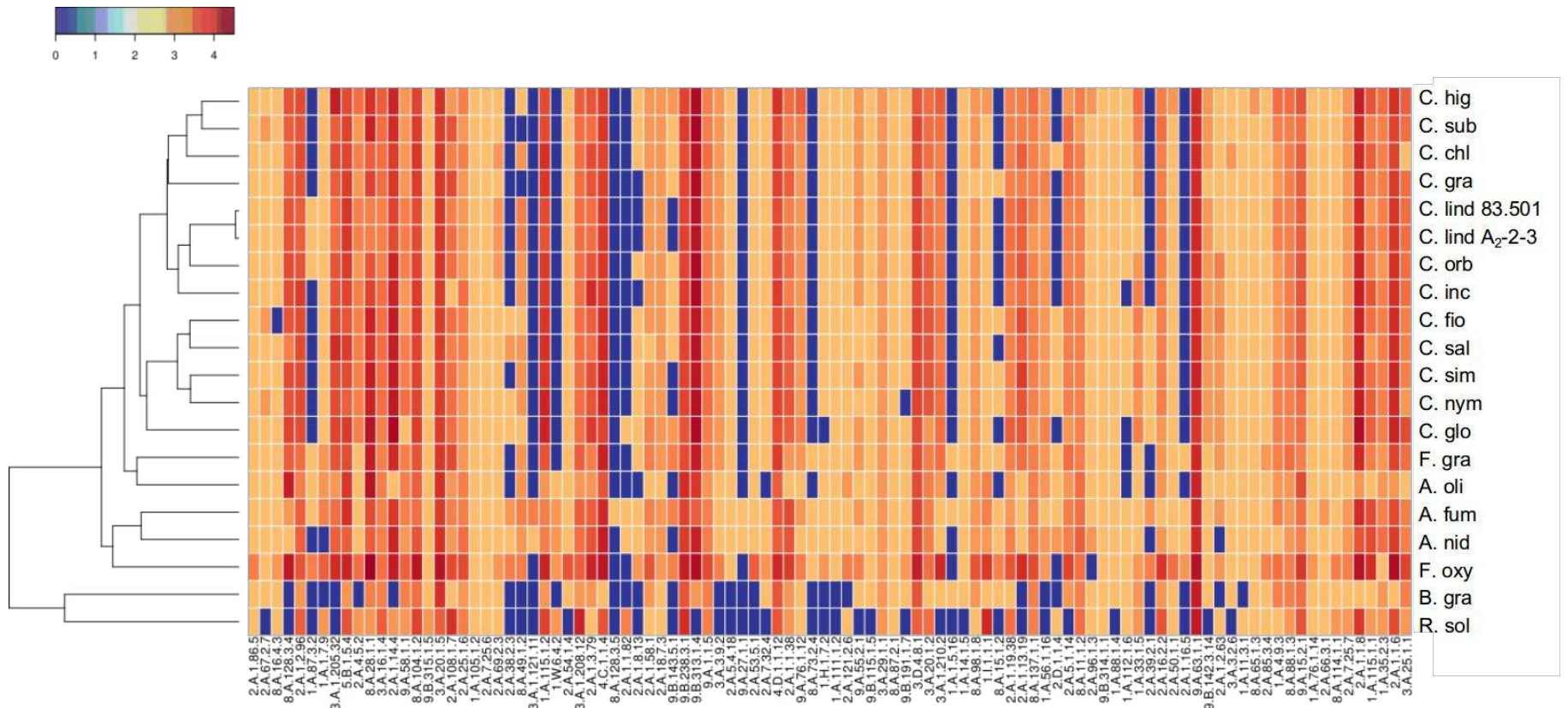


Figura 11. Contrações e expansões de famílias de transportadores em *C. lindemuthianum*.

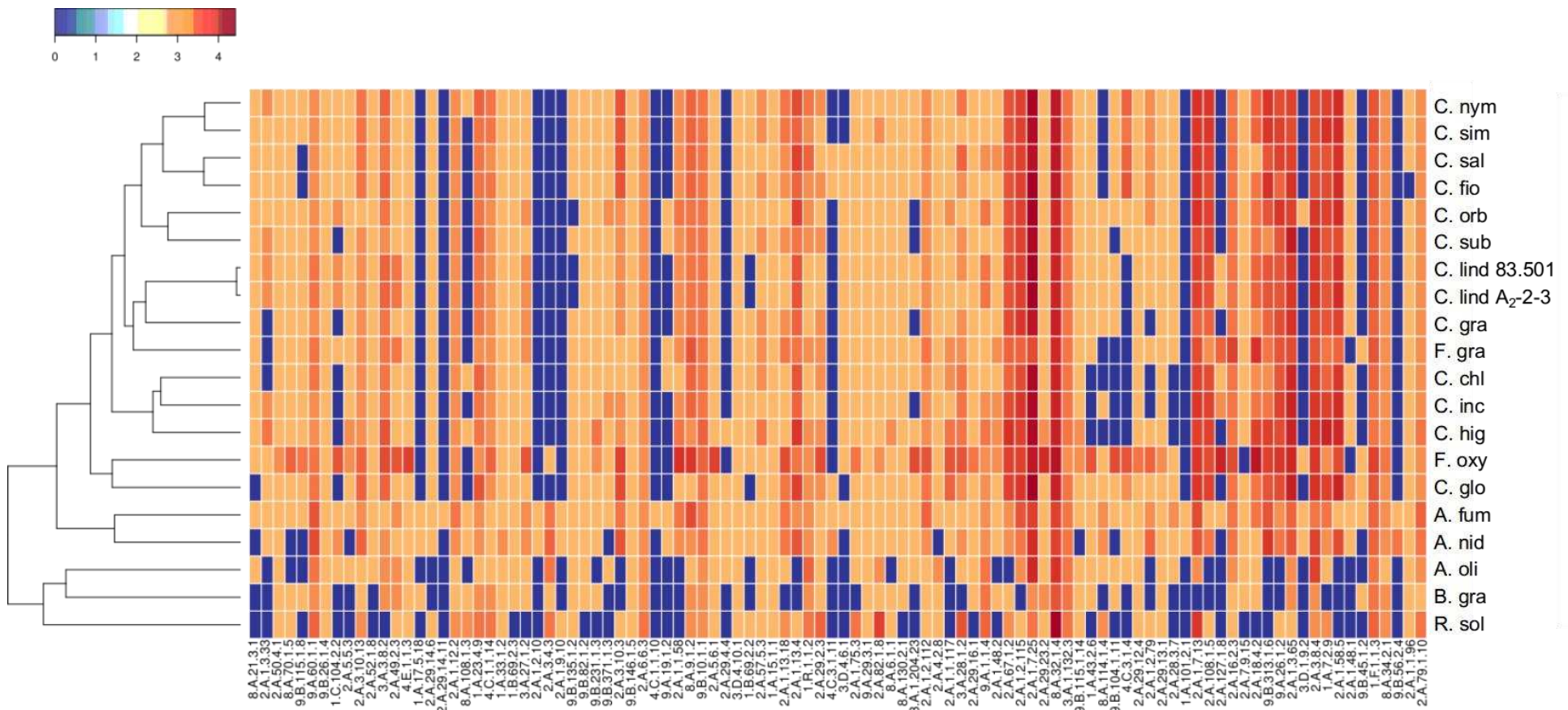


Figura 12. Contrações e expansões de famílias de transportadores em *C. lindemuthianum*.

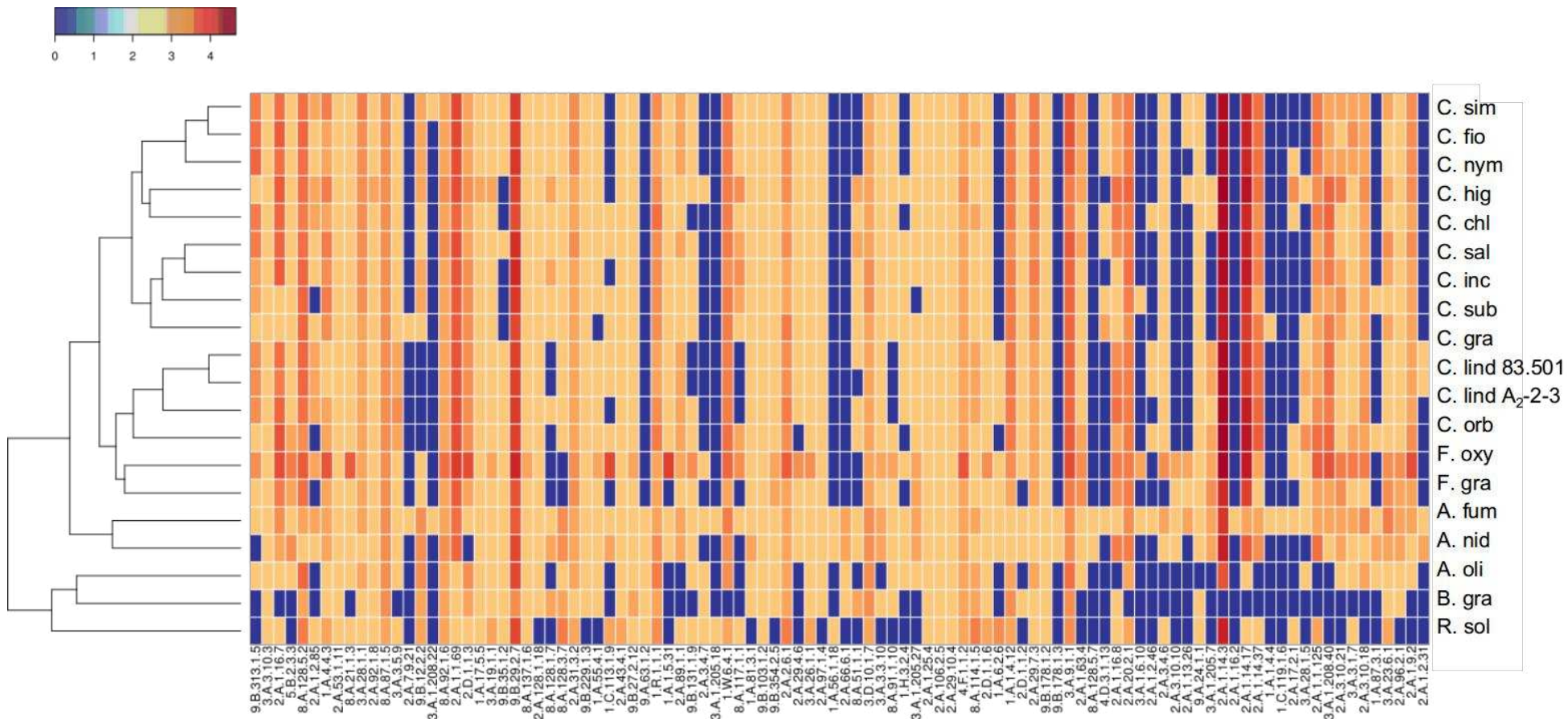


Figura 13. Contrações e expansões de famílias de transportadores em *C. lindemuthianum*.

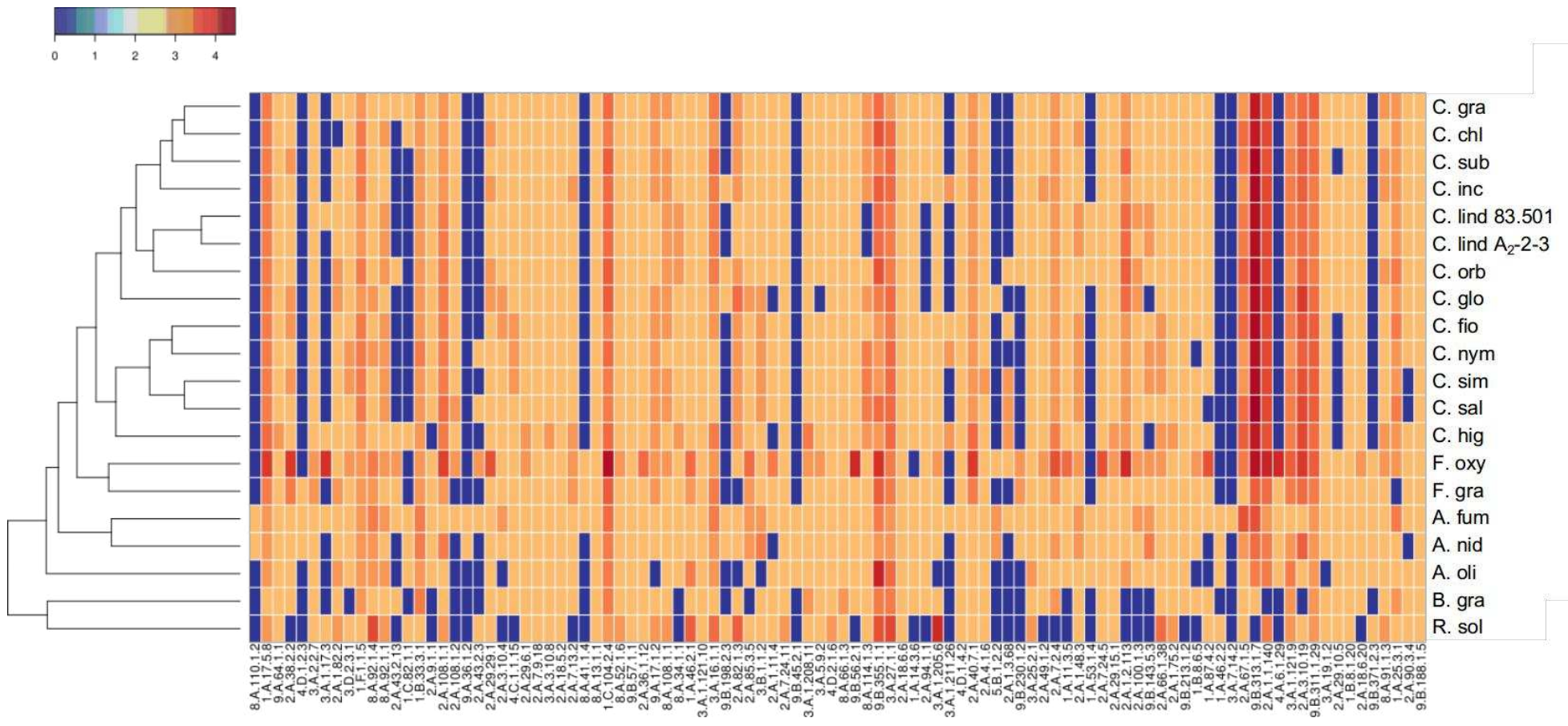


Figura 14. Contrações e expansões de famílias de transportadores em *C. lindemuthianum*.

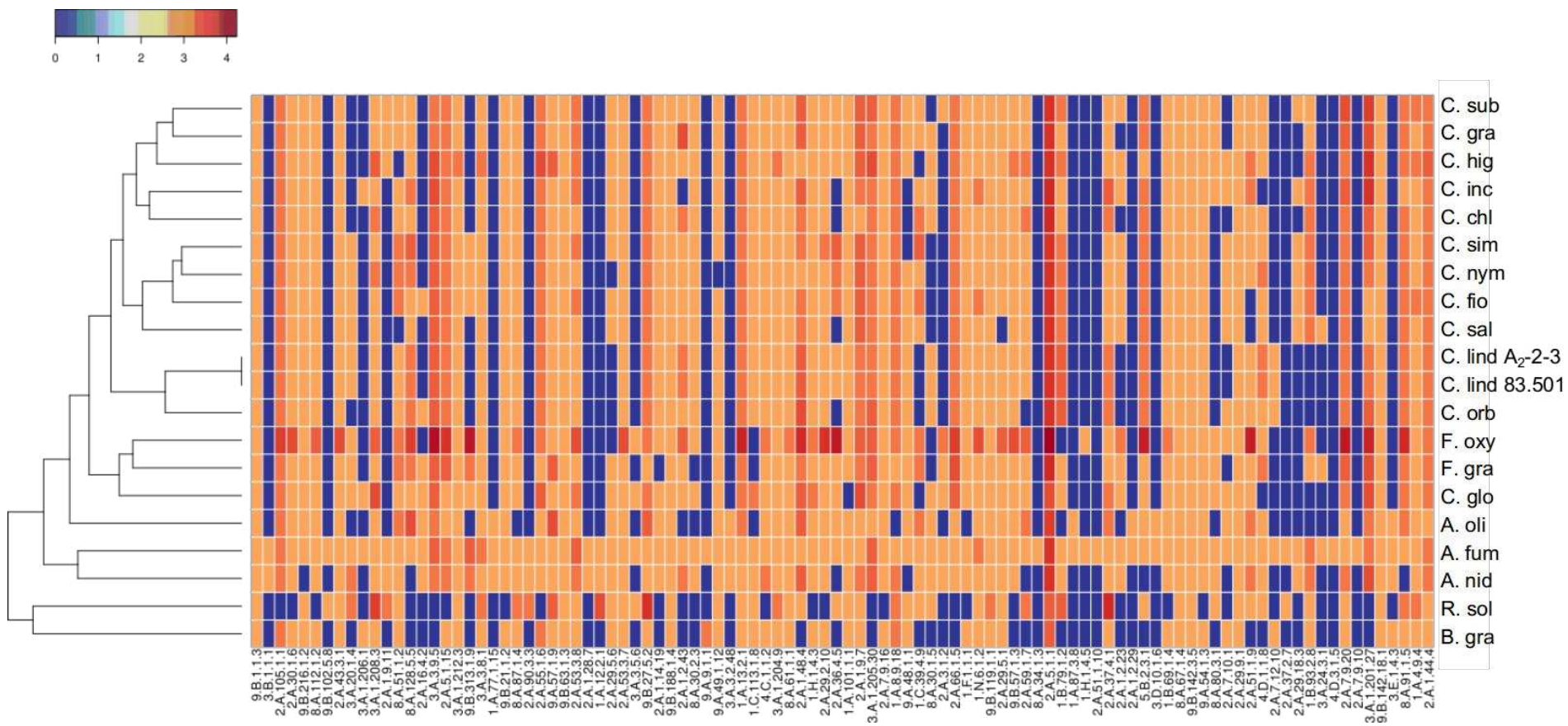


Figura 15. Contrações e expansões de famílias de transportadores em *C. lindemuthianum*.

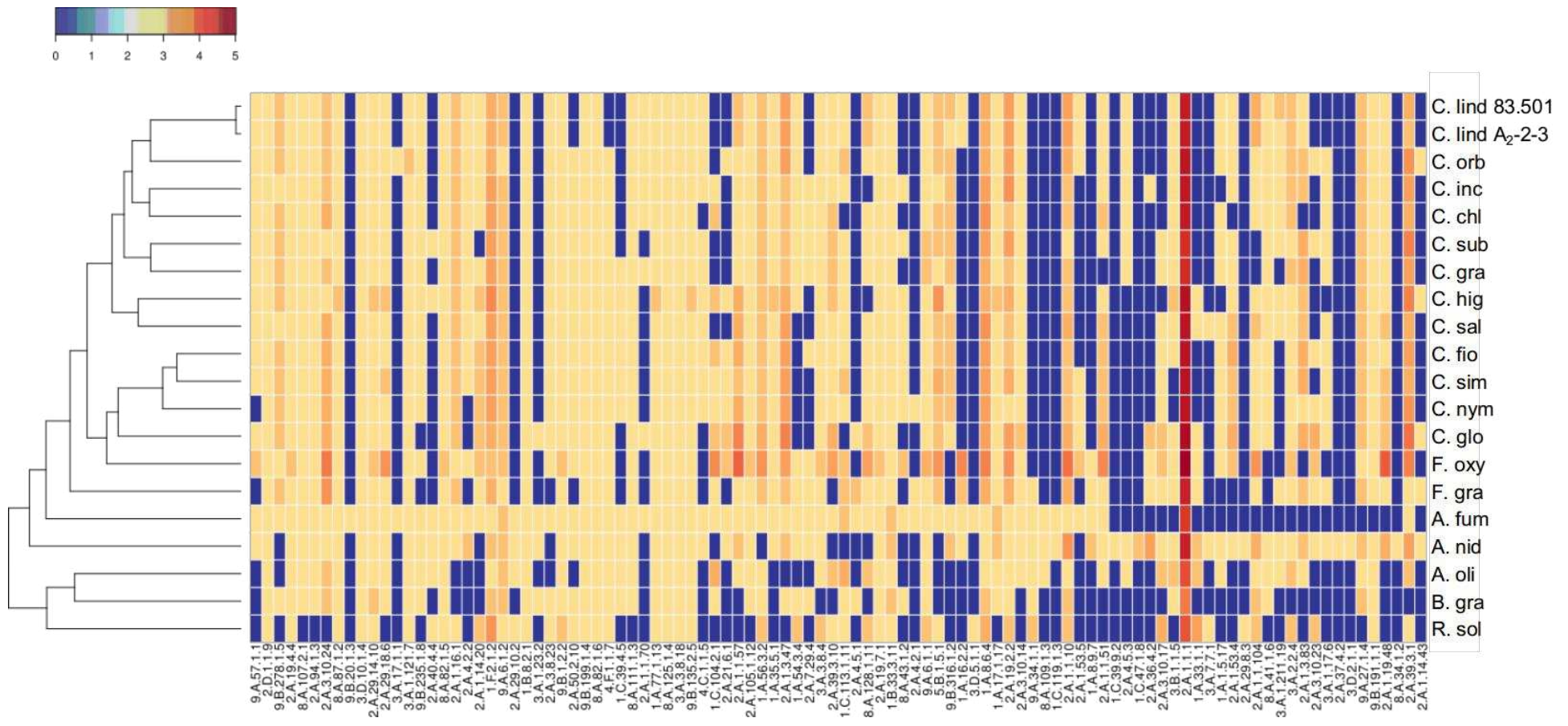


Figura 16. Contrações e expansões de famílias de transportadores em *C. lindemuthianum*.

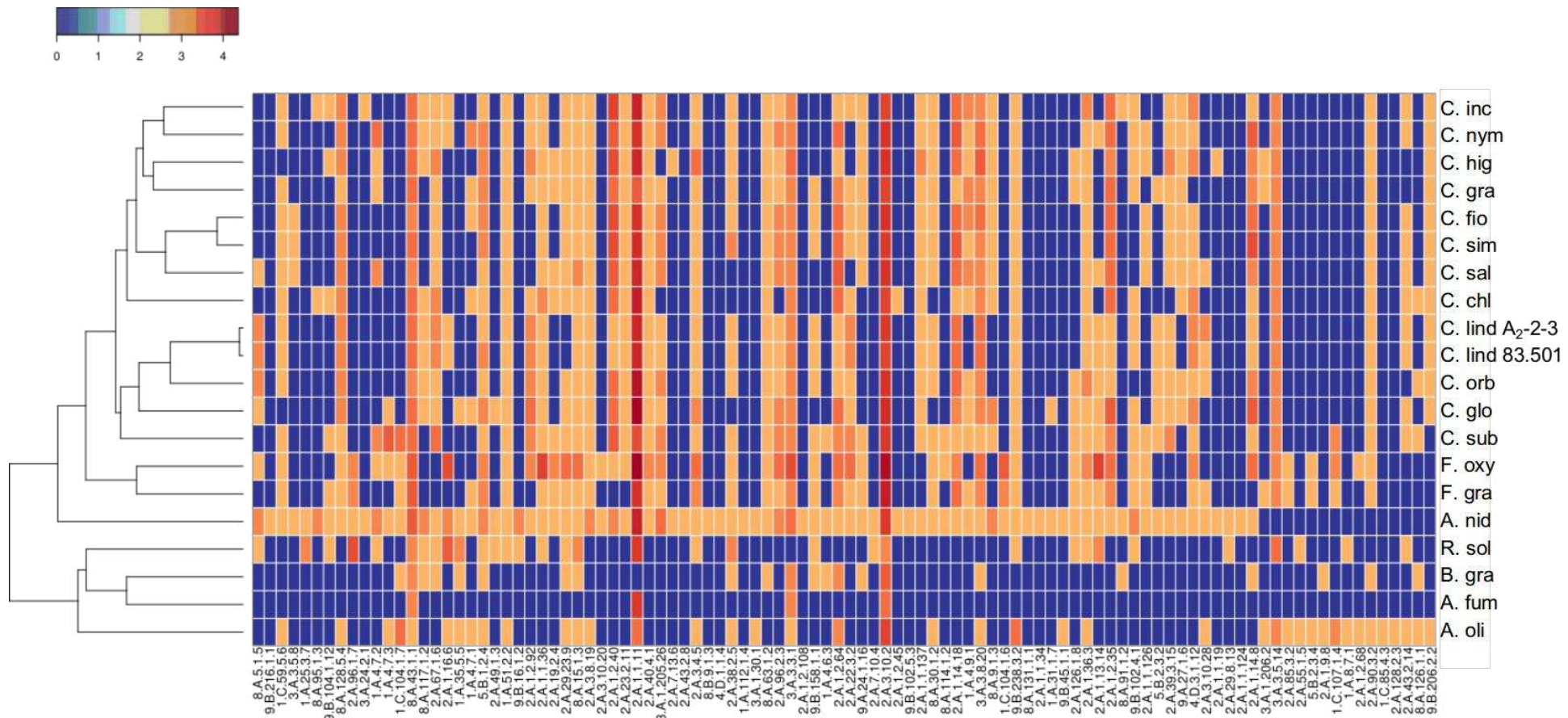


Figura 17. Contrações e expansões de famílias de transportadores em *C. lindemuthianum*.

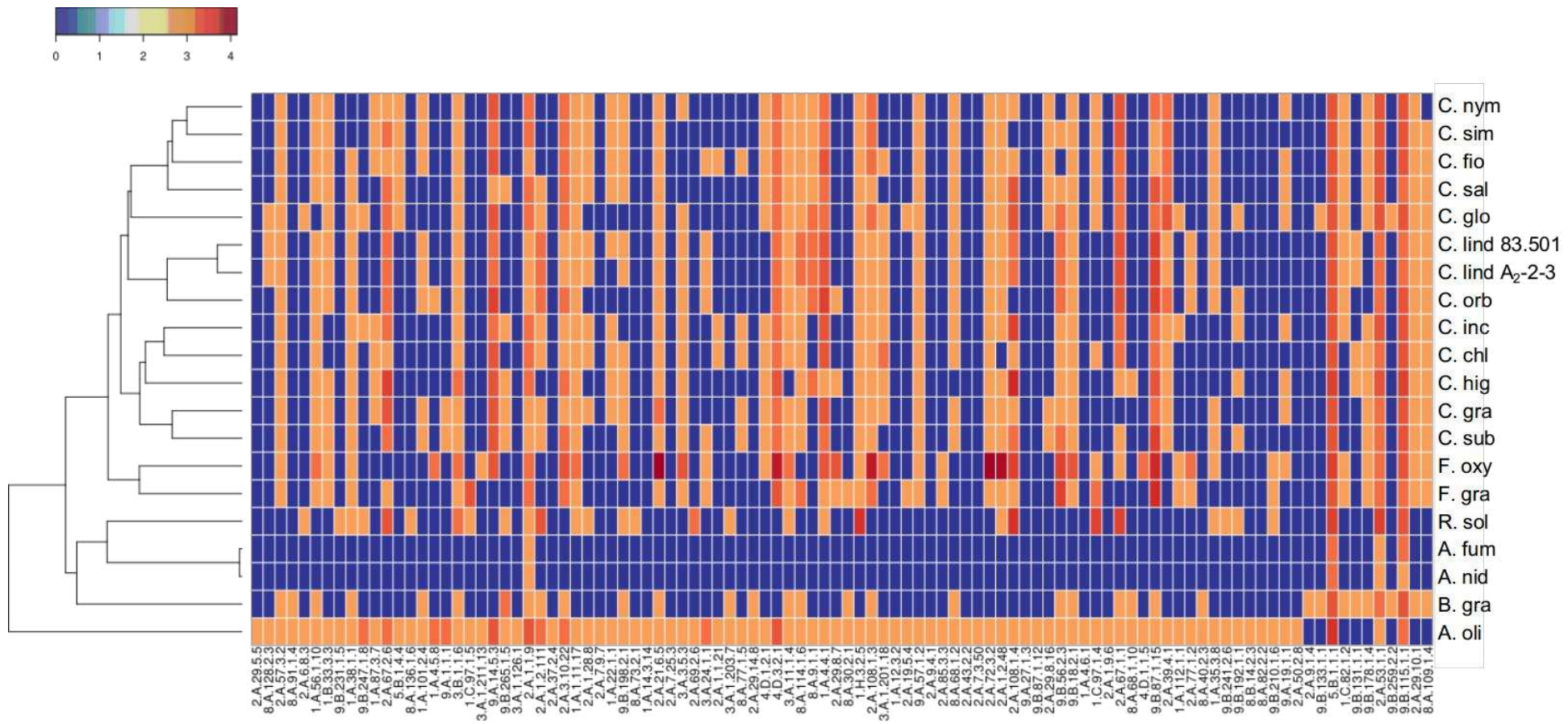


Figura 18. Contrações e expansões de famílias de transportadores em *C. lindemuthianum*.

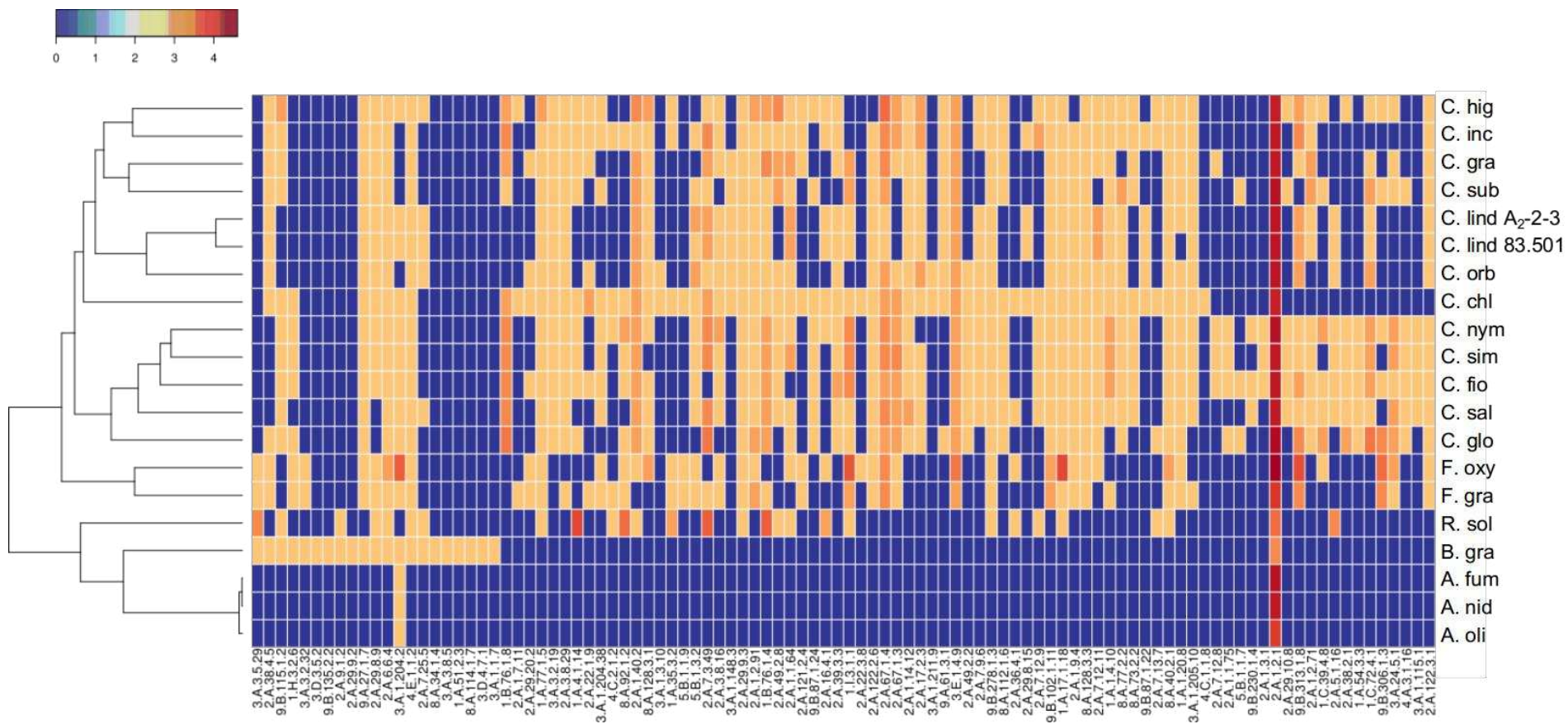


Figura 19. Contrações e expansões de famílias de transportadores em *C. lindemuthianum*.

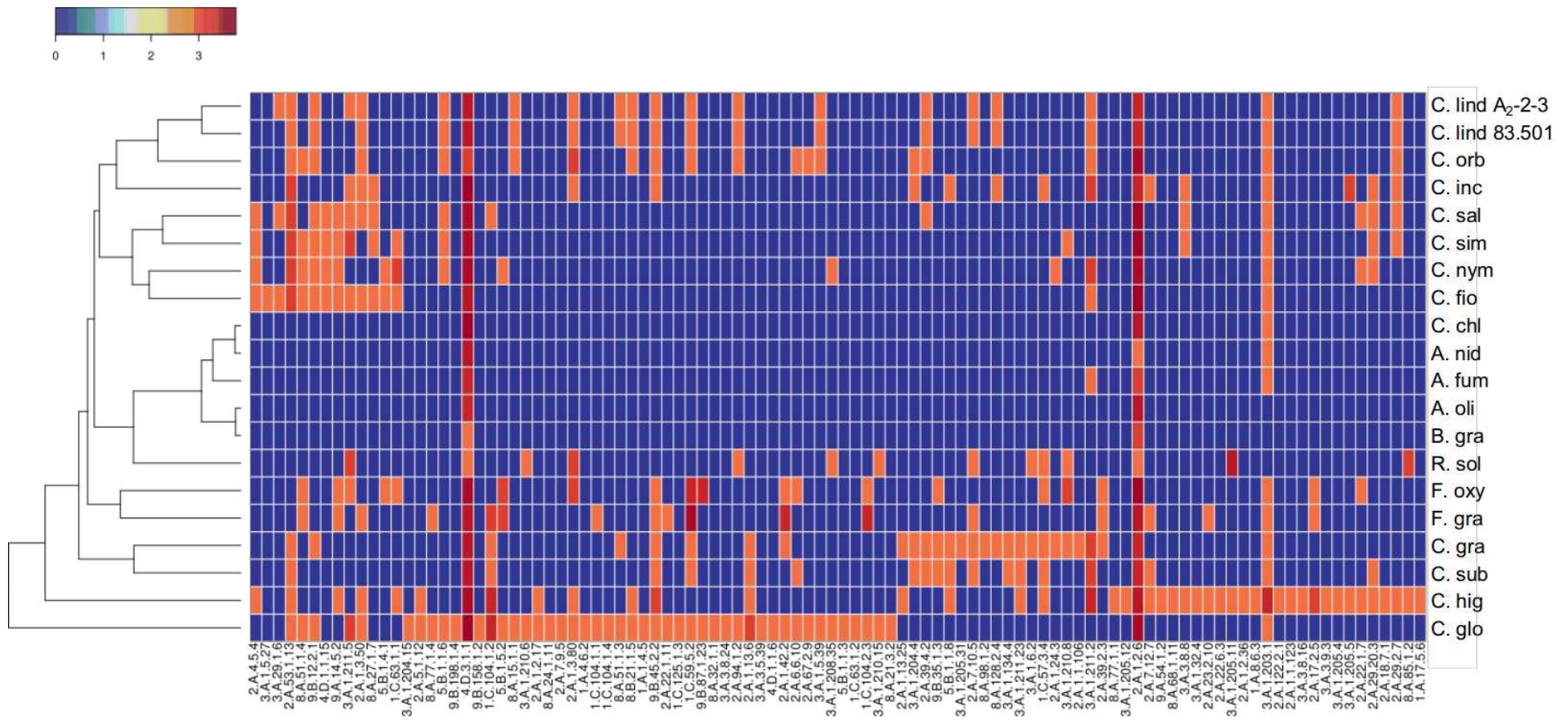


Figura 20. Contrações e expansões de famílias de transportadores em *C. lindemuthianum*.

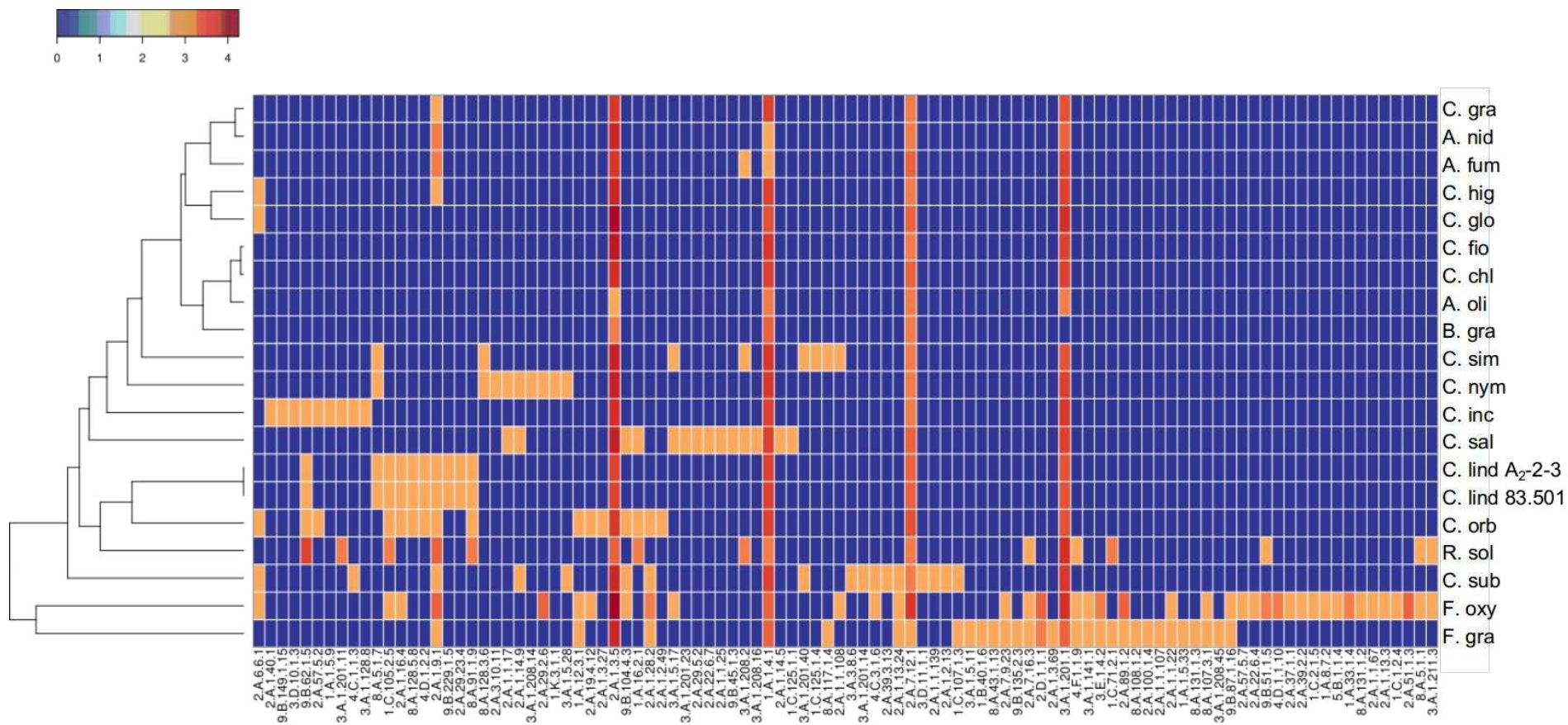


Figura 21. Contrações e expansões de famílias de transportadores em *C. lindemuthianum*.

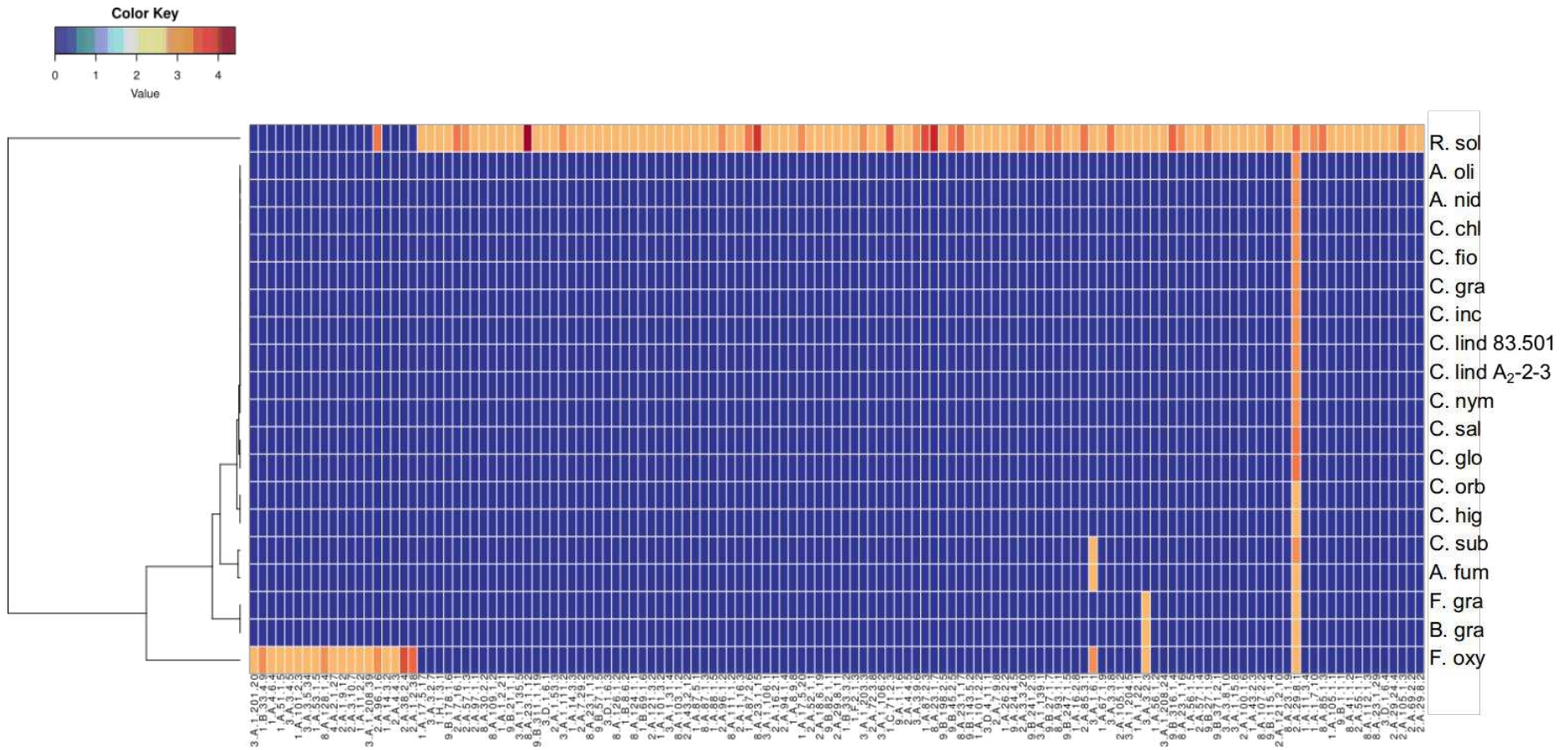


Figura 22. Contrações e expansões de famílias de transportadores em *C. lindemuthianum*.

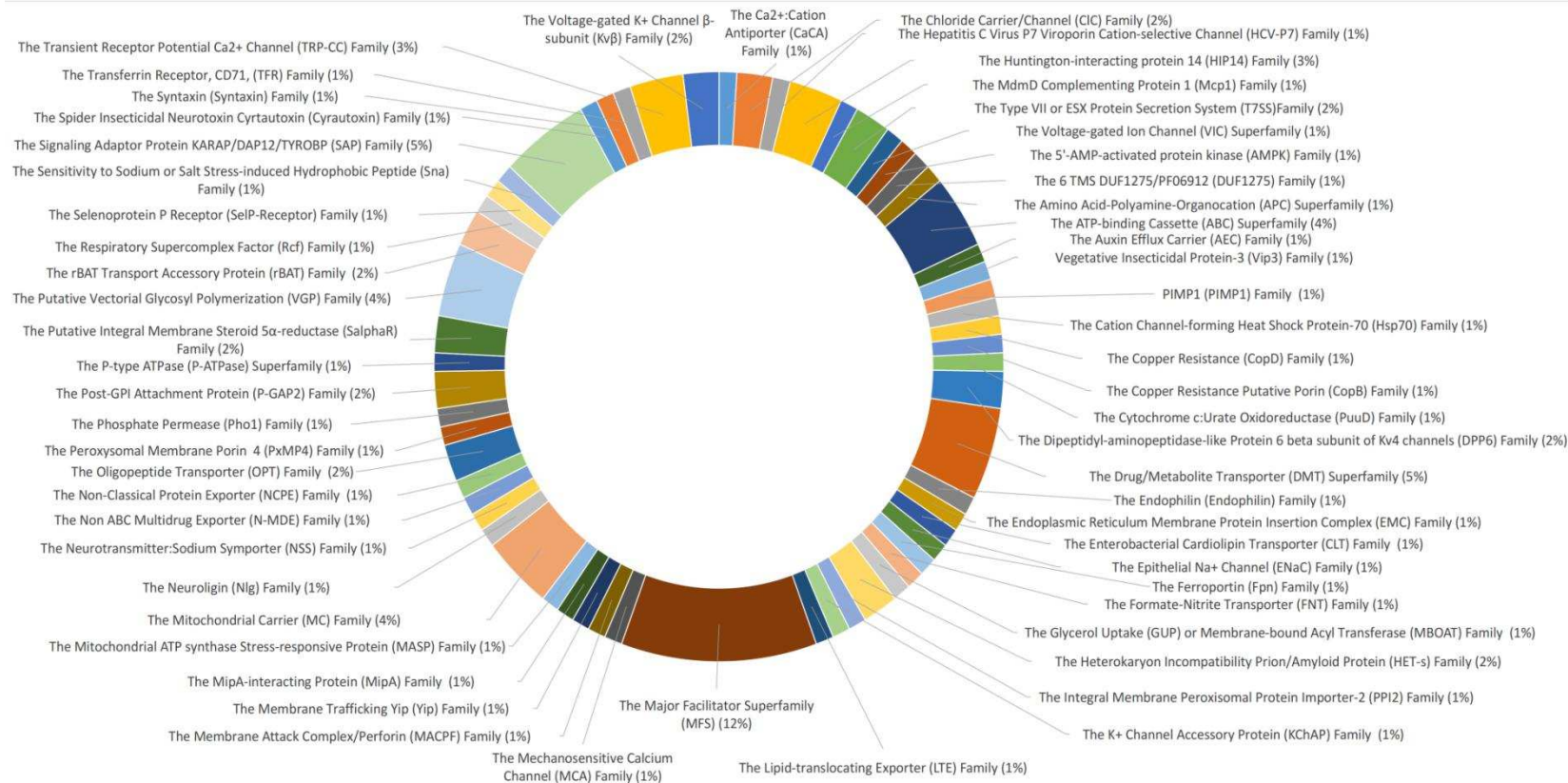


Figura 23. Classificação dos transportadores contraídos/expandidos nos genomas de *Colletotrichum lindemuthianum* em famílias do TCDB. A classificação foi obtida por meio dos códigos do TCDB.

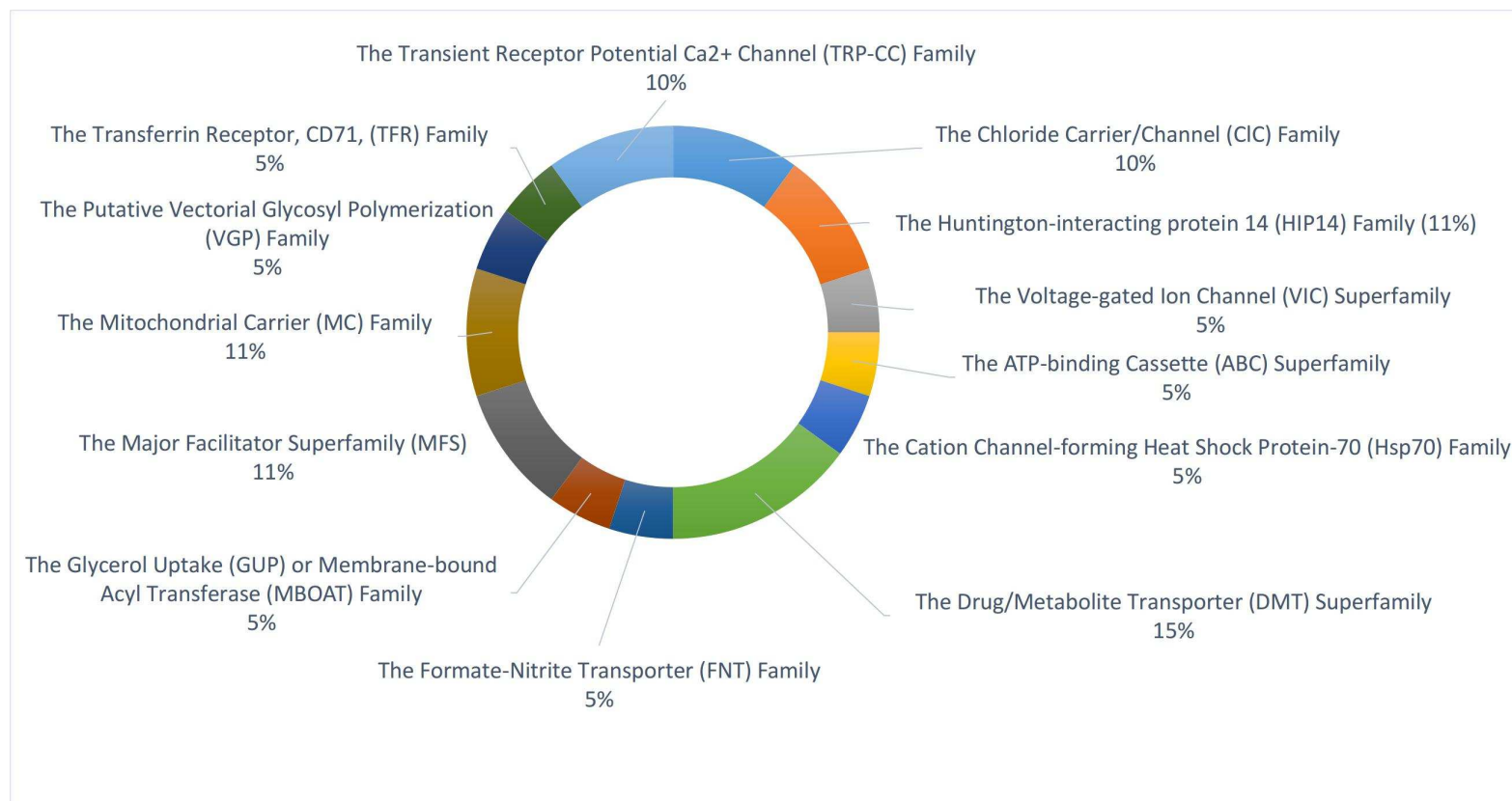


Figura 24. Classificação dos transportadores contraídos/expandidos exclusivamente nos genomas de *Colletotrichum lindemuthianum* em famílias do TCDB. A classificação foi obtida por meio dos códigos do TCDB.

Enzimas Ativas em Carboidratos (CAZymes) Secretadas

Foram identificados 415 genes codificando CAZymes secretadas nos genomas de *C. lindemuthianum*. Este número é superior ao que foi encontrado nos genomas de algumas espécies de *Colletotrichum* que infectam um número maior de espécies vegetais, tais como *C. higginsianum* (317), *C. graminicola* (342), *C. sublioneola* (392) e *C. chlorophyti* (413). Entre os genomas analisados, o número de CAZymes secretadas variou de 43 em *B. graminis* até 539 em *C. gloeosporioides*. A maioria das espécies de *Colletotrichum* analisadas apresenta um maior número de CAZymes em seus secretomas preditos do que os fungos saprofiticos e biotróficos avaliados, sendo uma exceção a espécie *C. incanum* (214) (**Figura 25**).

A média do número de CAZymes secretadas em *Colletotrichum* (409) é superior à média das CAZymes encontradas em *F. graminicola* e *F. oxysporum* e próxima a observada no genoma de *R. solani*. As glicosídeo hidrolases (GH) foram as CAZymes mais abundantes nas espécies analisadas neste trabalho, seguidas de enzimas com atividades auxiliares (AA), módulos de ligação a carboidratos (CBM), carboidrato esterases (CE), polissacarídeo liases (PL) e glicosil-transferases (GT).

A análise de agrupamentos baseada no perfil de distribuição de famílias de CAZymes secretadas demonstrou a formação de três principais grupos de espécies. As espécies de *Colletotrichum* não agruparam de acordo com o esperado pela proximidade filogenética, como no caso de *C. lindemuthianum* (agregado orbiculare) que agrupou com *C. chlorophyti* (agregado dematium) (**Figura 25**). Ainda neste grupo, estão as espécies *R. solani* e *F. oxysporum* não pertencentes ao gênero *Colletotrichum*.

Dentre as 20 famílias de CAZymes mais abundantes nos genomas de *C. lindemuthianum*, ao menos 12 são conhecidas como sendo envolvidas na degradação da biomassa vegetal (**Tabela Suplementar 5**). A maior parte das

espécies de *Colletotrichum*, assim como *F. oxysporum* apresenta uma expansão no repertório de genes que codificam xilanases quando comparado às demais espécies avaliadas, sendo a exceção a espécie *C. incanum* (**Figura 26**). Os dados demonstram também que as espécies de *Colletotrichum* patógenas de plantas não gramíneas, com exceção de *C. incanum* apresentam uma expansão no repertório de genes que codificam pectinases quando comparado as espécies que causam doenças em gramíneas.

Adicionalmente foram analisadas as famílias de CAZymes envolvidas na degradação de biomassa vegetal nos secretomas de *Colletotrichum* spp. (**Figura 27**). O dendrograma de agrupamento com base no padrão da distribuição e abundância destas famílias demonstra que as espécies *C. lindemuthianum* e *C. orbiculare* não apresentam uma distribuição similar das famílias de CAZymes avaliadas, não tendo, portanto, estas agrupado conjuntamente. A espécie *C. lindemuthianum* de acordo com este agrupamento é mais similar à *C. incanum*, enquanto *C. orbiculare* apresenta-se mais proximamente relacionado à *C. salicis*.

Foram encontradas alterações no número de cópias de 10 famílias de CAZymes secretadas nos genomas de *C. lindemuthianum* (**Figura 27**). Destas, quatro consistem em contrações acordantes entre os genomas dos isolados 83.501 e A₂-2-3 (famílias GH29, GH51, GH62 e GH115) (**Tabela 6**). Seis destas alterações são divergentes entre os isolados-83.501 e A₂-2-3, estando todas estas contraídas no genoma do isolado 83.501 e expandidas no genoma do isolado A₂-2-3 (GH36, GH53, GH67, GH74, GH93 e PL4). A única alteração no número de cópias exclusivamente encontrada na espécie *C. lindemuthianum* foi a contração na família GH62, e é compartilhada nos genomas dos dois isolados analisados.

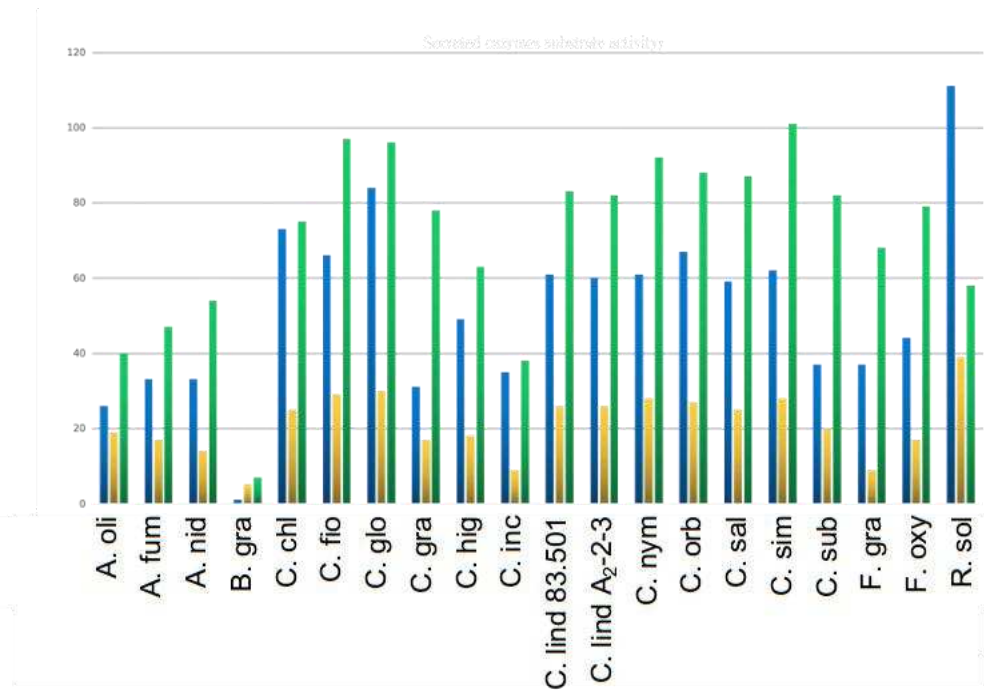


Figura 26. Classificação das cazymes secretadas quanto ao substrato alvo. As enzimas foram classificadas em: Pectinases, celulasas e xilanases. B. gra - *Blumeria graminis*; A. oli – *Arthrobotrys oligospora*; A. fum – *Aspergillus fumigatus*; A. nid – *Aspergillus nidulans*; C. gra – *C. graminicola*; R. sol - *Rhizoctonia solani*; C. lind 83.501 e C. lind A₂-2-3– *C. lindemuthianum*; C. chl – *C. chlorophyti*; C. fio – *C. fiorinae*; C. glo – *C. gloeosporioides*; C. hig – *C. higginsianum*; C. inc – *C. incanum*; C. nym – *C. nymphaeae*; C. orb – *C. orbiculare*; C. sal – *C. salicis*; C. sim – *C. simmondsii*; C. sub – *C. sublineola*; F. gram - *Fusarium graminearum*; F. oxy - *F. oxysporum*.

Tabela 6. Famílias de cazymes que sofreram contrações e expansões nos genomas de *C. lindemuthianum*.

Família	Substrato alvo	Situação
GH29	hemicelulose	Contraída em <i>C. lind</i> 83.501 e A ₂ -2-3 e em <i>C. hig</i>
GH51	hemicelulose	Contraída em <i>C. lind</i> 83.501e A ₂ -2-3, <i>C. inc</i> e em <i>C. orb</i>
GH62	hemicelulose	Contraída exclusivamente em <i>C. lind</i> 83.501 e A ₂ -2-3
GH115	hemicelulose	Contraída em <i>C. lind</i> 83.501e A ₂ -2-3 e em <i>C. inc</i>
GH36	hemicelulose	*Contraída em <i>C. lind</i> 83.501e outras e expandida em <i>C. lind</i> A ₂ -2-3 e outras
GH67	hemicelulose	*Contraída em <i>C. lind</i> 83.501e outras e expandida em <i>C. lind</i> A ₂ -2-3 e outras
GH74	celulose	*Contraída em <i>C. lind</i> 83.501e outras espécies e expandida em <i>C. lind</i> A ₂ -2-3 e outras
GH93	hemicelulose	*Contraída em <i>C. lind</i> 83.501e outras, expandida em <i>C. lind</i> A ₂ -2-3 e outras
GH53	pectina	*Contraída em <i>C. lind</i> 83.501e outras, expandida em <i>C. lind</i> A ₂ -2-3 e outras
PL4	pectina	*Contraída em <i>C. lind</i> 83.501e outras espécies e expandida em <i>C. lind</i> A ₂ -2-3 e outras

“ * ” significam que a alteração em questão é divergente entre os genomas dos dois isolados de *C. lindemuthianum*. “outras” indicam que a alteração é compartilhada com outros fungos pertencentes ao gênero *Colletotrichum*. *C. lind* 83.501 e *C. lind* A₂-2-3 – *C. lindemuthianum*; *C. hig* – *C. higginsianum*; *C. inc* – *C. incanum*; *C. orb* – *C. orbiculare*.

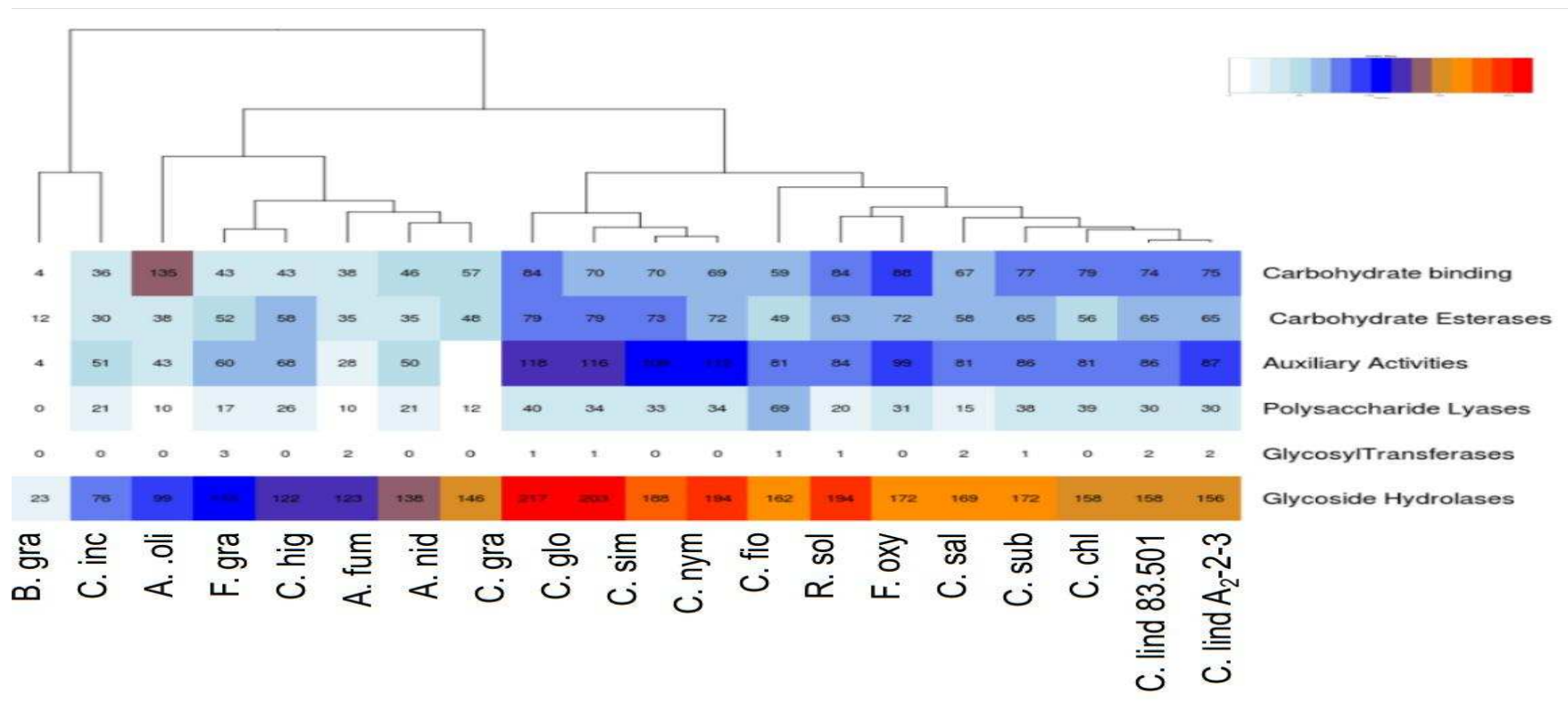


Figura 25. Número de genes que codificam **CAZymes** secretadas totais, codificadas nos genomas de *Colletotrichum* spp. e demais espécies.. B. gra - *Blumeria graminis*; A. oli – *Arthrotritys oligospora*; A. fum – *Aspergillus fumigatos*; A. nid – *Aspergillus nidulans*; C. gram – *C. graminicola*; R. sol - *Rhizoctonia solani*; C. lind 83.501 e C. lind A₂-2-3 – *C. lindemuthianum*; C. chl – *C.chlorophyti*; C. fio – *C. fiorinae*;

C. glo – *C. gloeosporioides*; C. hig – *C. higginsianum*; C. inc – *C. incanum*; C. nym – *C. nymphaeae*; C. orb – *C. orbiculare*; C. sal – *C. salicis*; C. sim – *C. simmondsii*; C. sub – *C. sublineola*; F. gram - *Fusarium graminearum*; F. oxy - *F. oxysporum*.

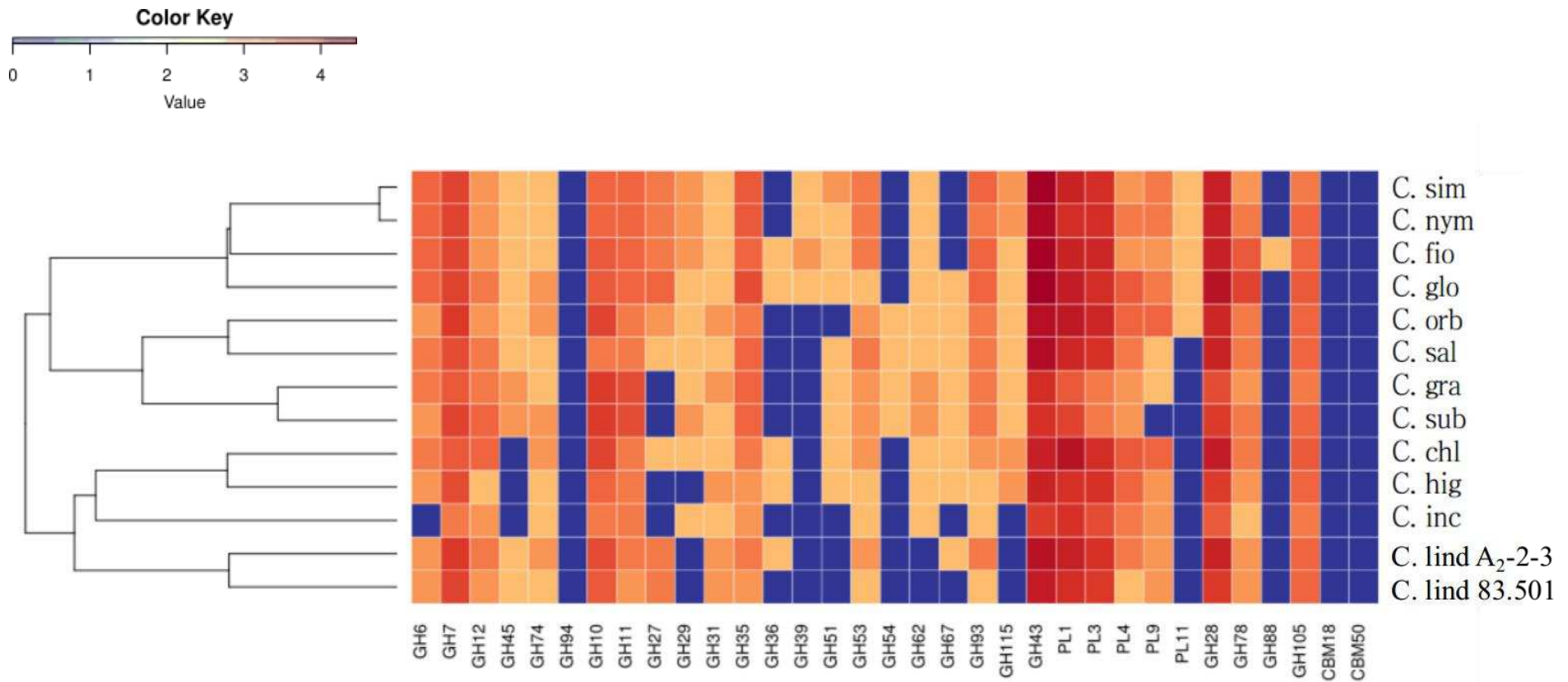


Figura 27. Expansões e contrações de famílias de genes que codificam **CAZymes** secretadas envolvidas na degradação de biomassa vegetal, codificadas nos genomas de *Colletotrichum*. Os valores são expostos em escala logarítmica crescente do azul para o vermelho. Valores azuis

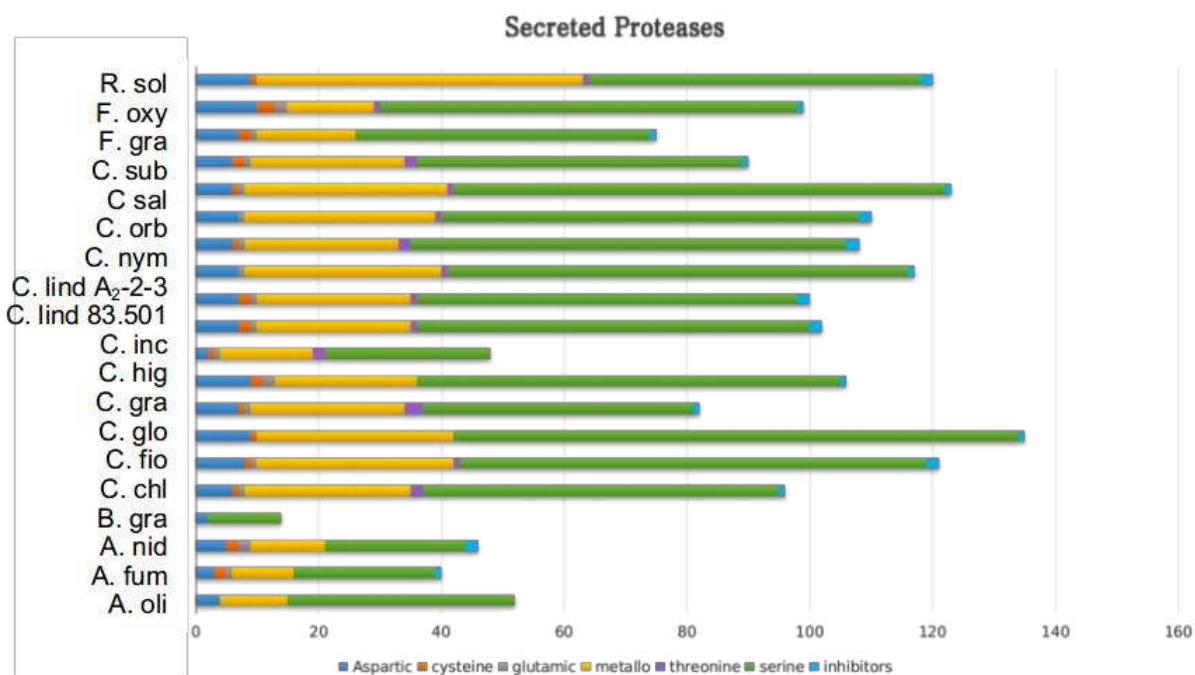
indicam contrações e vermelhos expansões. B. gra - *Blumeria graminis*; A. oli – *Arthrobotrys oligospora*; A. fum – *Aspergillus fumigatus*; A. nid – *Aspergillus nidulans*; C. gram – *C. graminicola*; R. sol - *Rhizoctonia solani*; C. lind 83.501 e C. lind A₂-2-3 – *C. lindemuthianum*; C. chl – *C. chlorophyti*; C. fio – *C. fiorinae*; C. glo – *C. gloeosporioides*; C. hig – *C. higginsianum*; C. inc – *C. incanum*; C. nym – *C. nymphaeae*; C. orb – *C. orbiculare*; C. sal – *C. salicis*; C. sim – *C. simmondsii*; C. sub – *C. sublineola*; F. gram - *Fusarium graminearum*; F. oxy - *F. oxysporum*.

Proteases secretadas

O total de proteases secretadas encontrado em *C. lindemuthianum* foi de 100 e 102, nos isolados 83.501 e A₂ 2-3, respectivamente. Entre as demais espécies de *Colletotrichum*, o total variou de 48 em *C. incanum* até 135 em *C. gloeosporioides* (**Figura 28**). Nas espécies não pertencentes ao gênero *Colletotrichum*, o total de proteases secretadas variou de apenas 14 em *B. graminis* até 120 em *R. solani*. Foi observada uma expansão no repertório de serina proteases em todas as espécies de fungos analisadas neste trabalho, o que é mais evidente nas espécies pertencentes ao gênero *Colletotrichum*.

A família de metalloproteases foi a segunda família de proteases mais abundante, variando de 15 genes no genoma de *C. incanum* até 33 em *C. simmondsii*. Baseando-se no perfil de proteases secretadas, *C. lindemuthianum*, *C. salicis* e *C. higginsianum*, que pertencem aos clados orbiculare, acutatum e destructivum, respectivamente, agruparam juntos em vez de agruparem com outras espécies de *Colletotrichum* filogeneticamente mais próximas. Este perfil também foi observado para as espécies não pertencentes ao gênero *Colletotrichum*, como *F. graminearum* que agrupou com *C. graminicola*.

Foram encontradas alterações no número de cópias em quatro famílias de proteases secretadas em *C. lindemuthianum*, sendo todas estas compartilhadas com uma ou mais espécies do gênero *Colletotrichum* (**Figura 29**). A família S12 está contraída nos genomas de *C. lindemuthianum*, *C. graminicola* e em *C. higginsianum*. A família M12 está contraída nos genomas de *C. lindemuthianum* e no genoma de *C. graminicola* analisado. A família C69 encontra-se expandida nos genomas de *C. lindemuthianum* e no genoma de *C. fioriniae*. E por último, a família S01 encontra-se expandida em *C. lindemuthianum* e no genoma de *C. fioriniae*.



Figura

28. Número de genes codificando proteases secretadas, colocar as referências codificadas nos genomas de *Colletotrichum* spp. e outras espécies de fungos. B. gra - *Blumeria graminis*; A. oli - *Arthrobotrys oligospora*; A. fum - *Aspergillus fumigatus*; A. nid - *Aspergillus nidulans*; C. gra - *C. graminicola*; R. sol - *Rhizoctoniasolani*; C. lind 83.501 e C. lind A₂-2-3 - *C. lindemuthianum*; C. chl - *C. chlorophyti*; C. fio - *C. fiorinae*; C. glo - *C. gloeosporioides*; C. hig - *C. higginsianum*; C. inc - *C. incanum*; C. nym - *C. nymphaeae*; C. orb - *C. orbiculare*; C. sal - *C. salicis*; C. sim - *C. simmondsii*; C. sub - *C. sublineola*; F. gra - *Fusarium graminearum*; F. oxy - *F. oxysporum*.

Metabolismo Secundário

O fungo *C. lindemuthianum* apresenta um número intermediário de genes que codificam NRPSs em seus genomas (12) quando comparado as demais espécies de *Colletotrichum*, e as espécies deste gênero coletivamente apresentam um número superior de genes de NRPSs e PKSs (policetídeo sintases) ao encontrado nas demais espécies de fungos avaliados (**Figura 30**). O número de genes que codificam NRPSs identificadas entre as espécies de fungos avaliadas variou de um em *B. graminis* a 22 em *C. higginsianum*. O número de genes que codificam sintetases de peptídeo não ribossomais não é expandido nos genomas de fitopatógenos generalistas, quando comparado a espécies mais específicas.

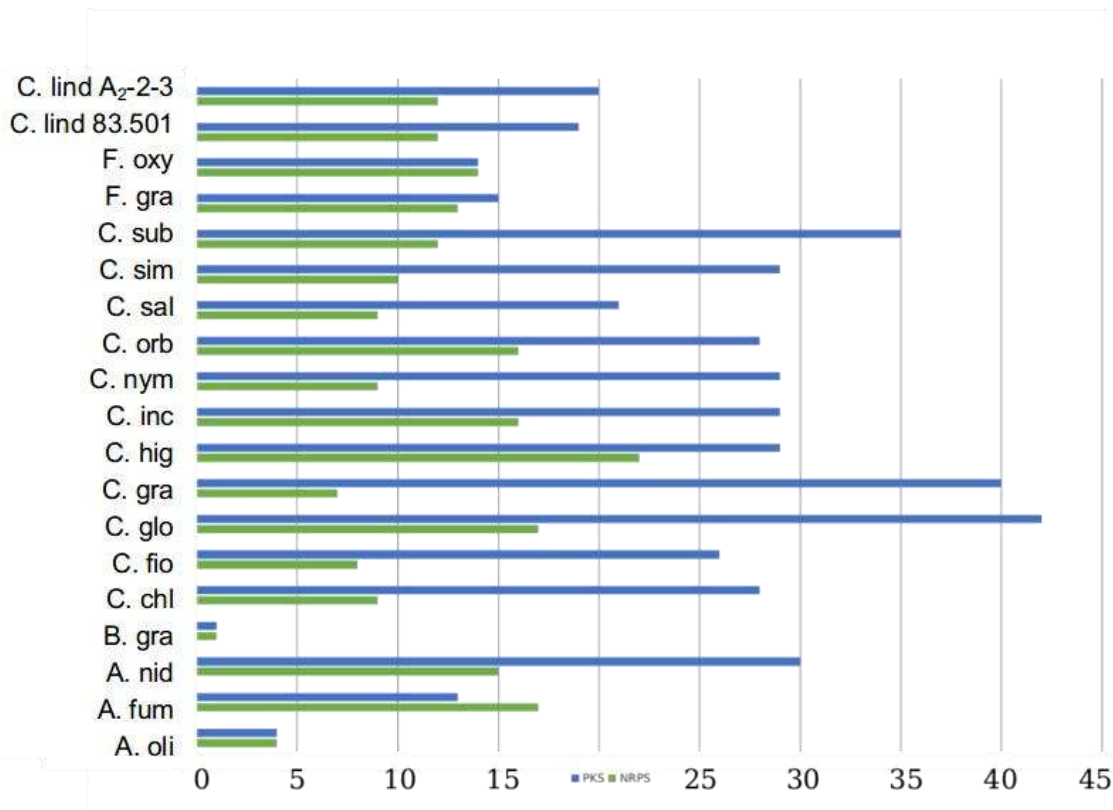


Figura 30. Número de genes que codificam sintetases de peptídeo não ribossomais (NRPS) e policetídeo sintases (PKS) nos genomas de *Colletotrichum lindemuthianum* e de outras espécies de fungos. Barras azuis correspondem ao número de PKS e barras verdes correspondem ao número de NRPS identificadas nos genomas em questão. B. gra - *Blumeria graminis*; A. oli - *Arthrobotrys oligospora*; A. fum - *Aspergillus fumigatus*; A. nid - *Aspergillus nidulans*; C. gram - *C. graminicola*; R. sol - *Rhizoctonia solani*; C. lind 83.501 e C. lind A₂-2-3 - *C. lindemuthianum*; C. chl - *C. chlorophyti*; C. fio - *C. forinae*; C. glo - *C. gloeosporioides*; C. hig - *C. higginsianum*; C. inc - *C. incanum*; C. nym - *C. nymphaeae*; C. orb - *C. orbiculare*; C. sal - *C. salicis*; C. sim - *C. simmondsii*; C. sub - *C. sublineola*; F. gram - *Fusarium graminearum*; F. oxy - *F. oxysporum*.

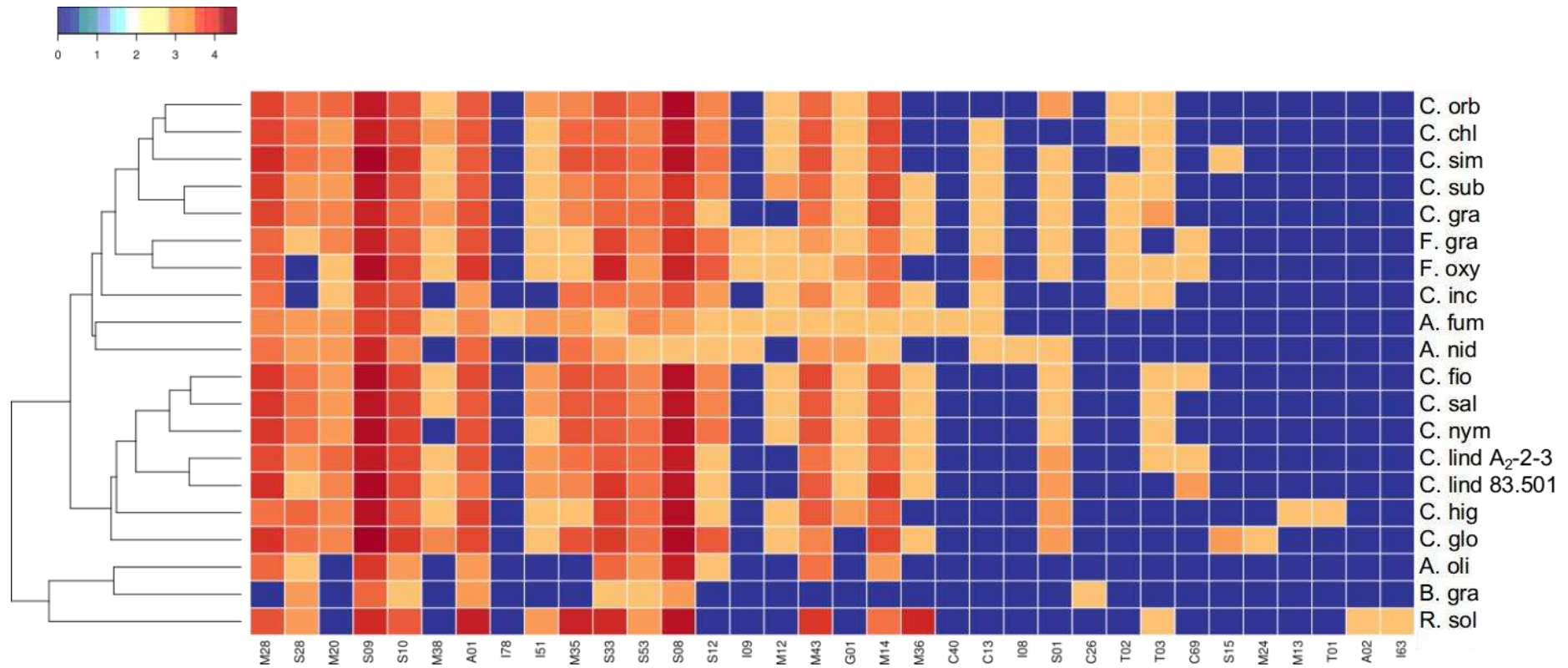


Figura 29. Contagem das famílias de proteases secretadas encontradas nos genomas de fungos analisados. Valores em log₂ do número

de genes pertencentes a cada família. B. gra - *Blumeria graminis*; A. oli – *Arthrobotrys oligospora*; A. fum – *Aspergillus fumigatus*; nid – *Aspergillus nidulans*; C. gram – *C. graminicola*; R. sol - *Rhizoctonia solani*; C. lind 83.501 e C. lind A₂-2-3 – *C. lindemuthianum*; C. chl

C. chlorophyti; *C. fio* – *C. fiorinae*; *C. glo* – *C. gloeosporioides*; *C. hig* – *C. higginsianum*; *C. inc* – *C. incanum*; *C. nym* – *C. nymphaeae*; *C. orb* – *C. orbiculare*; *C. sal* – *C. salicis*; *C. sim* – *C. simmondsii*; *C. sub* – *C. sublineola*; *F. gram* - *Fusarium graminearum*; *F. oxy* - *F. oxysporum*.

Foram identificados 19 (isolado 83.501) e 20 (isolado A₂-2-3) genes que codificam PKSs nos genomas de *C. lindemuthianum*, sendo estes os menores repertórios de PKSs encontrado dentre as espécies de *Colletotrichum* avaliadas. O número de genes que codificam PKSs encontrados nas demais espécies variou de 1 em *B. graminis* a 42 em *C. gloeosporioides* (**Figura 30**). As PKSs também são mais abundantes nas espécies de *Colletotrichum* do que nas demais espécies de fungos avaliadas, e tratando-se das espécies do gênero *Colletotrichum*, diferentemente do que foi encontrado para as sintetases de peptídeo não ribossomais o número de genes que codificam PKSs parece ser expandido nas espécies mais generalistas quando comparado às espécies com uma gama de hospedeiros mais restritas, como é o exemplo das espécies *C. gloeosporioides* e *C. graminicola* (maiores gamas de hospedeiros e maiores números de genes que codificam PKSs) e *C. lindemuthianum* e *C. salicis* (menores gamas de hospedeiros e menores números de genes que codificam PKSs).

Com o intuito de buscar ortólogos dos genes que codificam NRPSs em *C. lindemuthianum* foi realizada a análise filogenética das NRPSs dos dois isolados, além de *C. orbiculare*. Os resultados da reconstrução filogenética apontam para uma capacidade de produção de peptídeos não ribossomais altamente diversos por parte das espécies *C. lindemuthianum* e *C. orbiculare* (**Figura 31**). Sendo a única exceção a proteína correspondente ao gene *S89_g2060*, que não apresentou ortólogos no isolado 83.501 nem tampouco no genoma de *C. orbiculare*, todas as demais sequências proteicas preditas como sendo NRPSs em *C. lindemuthianum* apresentaram correspondência entre os dois isolados. Foram encontrados ortólogos deste gene nos genomas de *C. higginsianum* e de *C. nymphaeae* (**Figura Suplementar**

Apesar de sempre estarem agrupadas em clados contendo uma ou mais sequências protéicas de *C. orbiculare*, as sequências proteicas de NRPSs de *C. lindemuthianum* (de ambos isolados) se distanciam das demais, sempre convergindo para um ancestral comum entre si mais recentemente do que com ancestrais em comum com sequências de outras espécies, como é o caso de *C. orbiculare*.

As sequências proteicas preditas para os genes *S89_g8157* e *S83_3024* agruparam em um clado contendo *Cob11288* (*C. orbiculare*) e a sequência da *atnA* sintase, responsável pela biossíntese da aspercriptina, uma fitotoxina produzida pelo fungo *A. nidulans* (Chiang et al., 2016). Em um clado irmão altamente relacionado com uma alta proximidade filogenética (100% de confiança pelo método de *bootstrap*) foram agrupadas as sequências de *S83_5702*, *S89_3179* e *Cob_06950*.

As sequências das proteínas *S83_9364* e *S89_9844* agruparam em um clado irmão do clado contendo as sequências proteicas codificadas pelos genes *Cob_01351* e *NRPS8*, de *C. orbiculare* e *C. spinosum*, respectivamente. As sequências contidas nestes clados não agruparam com qualquer sequência proteica de função conhecida.

As sequências de *S89_1561* e *S83_180* agruparam no mesmo clado de *SidD*, requerida para a biossíntese de um sideróforo em *Aspergillus fumigatus* (Reiber et al., 2005). As sequências proteicas preditas para os genes *S89_6490* e *S83_6717* agruparam no mesmo clado do gene que codifica a *apicidin F sintetase*, responsável pela produção do tetrapeptídeo com atividade de histona deacetilase, produzido por *Fusarium fujikuroi* (Von Bargen et al., 2013).

As sequências *S89_10665* e *S83_4688* apresentaram-se agrupadas num clado contendo *Cob_01942* e *ChyA*, uma NRPS envolvida biossíntese do pigmento *Chrysogine* e outros 13 compostos em *Penicillium chrysogenum* (Viggiano et al., 2018). As sequências *S83_5858* e *S89_9373* agruparam no mesmo clado da sequência de

uma sintase de sideróforo putativa de *C. gloeosporioides*.

As sequências S89_4811 e S83_2984 agruparam num clado contendo uma *gramicidin synthase* putativa de *C. orbiculare*, possivelmente envolvida na biossíntese do antibiótico gramidicina, assim como na *Bacteria Bacillus brevis*. Em um clado proximamente relacionado a este foram encontradas as sequências da S89_g2060 e da NRPS14 de *Tolypocladium ophiglossoides*.

As sequências proteicas preditas para os genes S89_8176 e S83_3043 agruparam no mesmo clado da acIP sintase putativa de *C. trifolii* (Chankhamjon et al., 2014), responsável pela biossíntese de uma toxina em *A. oryzae*. As sequências de S89_8993 e S83_1408 agruparam no clado contendo a sequência putativa da FUB8 sintase de *C. orbiculare*. Em um clado próximo foram agrupadas as sequências de S89_8419 e S83_6123.

As sequências das proteínas preditas para os genes S83_1413 e S89_8998 agruparam separadamente como *outgroup*, não tendo, portanto, mostrado uma estreita relação filogenética com nenhuma das demais sequências contempladas na análise.

Os genomas de *C. lindemuthianum* apresentam um número de genes que codificam P450 mono-oxigenases inferior ao que foi identificado na maior parte das espécies do gênero, superando somente a espécie *C. graminicola* (**Figura 32**). As **Figuras 33 e 34** apresentam as 208 famílias de P450 mono-oxigenases identificadas nas espécies avaliadas neste trabalho. Foram encontradas alterações no número de cópias de 21 destas (**Tabela 7**), das quais três consistem em contrações e as demais em expansões. Somente cinco das alterações no número de cópias encontradas convergem entre os genomas dos dois isolados de *C. lindemuthianum*, e uma das alterações no número de cópias de famílias de P450 mono-oxigenases é exclusiva desta espécie, neste caso sendo compartilhada entre os genomas dos dois isolados.

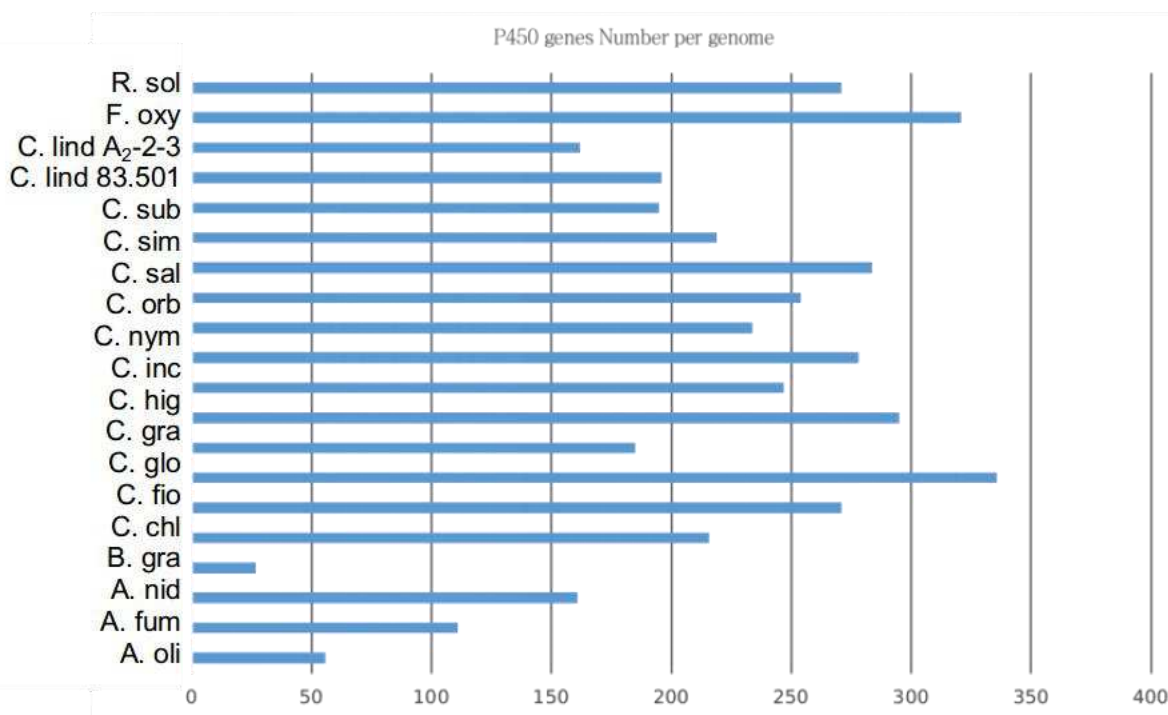


Figura 32. Número de P450 mono-oxigenases por genoma. B. gra - *Blumeria graminis*; A. oli – *Arthrotrrys oligospora*; A. fum – *Aspergillus fumigatos*; A. nid – *Aspergillus nidulans*; C. gram – *C. graminicola*; R. sol - *Rhizoctonia solani*; C. lind 83.501 e C. lind A₂-2-3 – *C. lindemuthianum*; C. chl – *C.chlorophyti*; C. fio – *C. fioriniae*; C. glo – *C. gloeosporioides*; C. hig – *C. higginsianum*; C. inc – *C. incanum*; C. nym – *C. nymphaeae*; C. orb – *C. orbiculare*; C. sal – *C. salicis*; C. sim – *C. simmondsii*; C. sub – *C. sublineola*; F. gram - *Fusarium graminearum*; F. oxy - *F. oxysporum*.

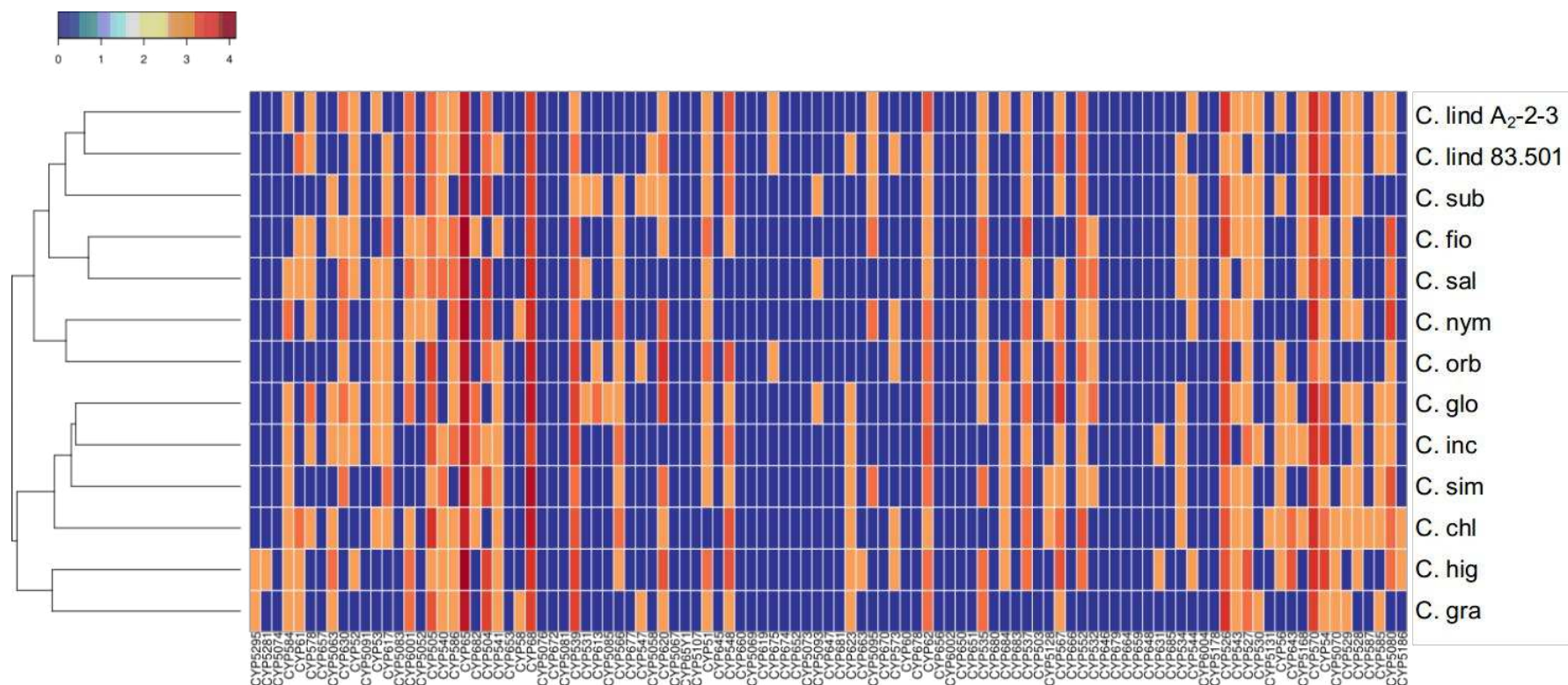


Figura 33. Contrações e expansões de famílias de citocromo P450 mono-oxigenases em *C. lindemuthianum*. Os valores consistem no Log₂ do número de genes. B. gra - *Blumeria graminis*; A. oli – *Arthrotrrys oligospora*; A. fum – *Aspergillus fumigatos*; A. nid – *Aspergillus nidulans*; C. gram – *C. graminicola*; R. sol - *Rhizoctonia solani*; C. lind 83.501 e C. lind A₂-2-3 – *C. lindemuthianum*; C. chl – *C. chlorophyti*;

C. fio – *C. fioriniae*; *C. glo* – *C. gloeosporioides*; *C. hig* – *C. higginsianum*; *C. inc* – *C. incanum*; *C. nym* – *C. nymphaeae*; *C. orb* – *C. orbiculare*; *C. sal* – *C. salicis*; *C. sim* – *C. simmondsii*; *C. sub* – *C. sublineola*; *F. gram* - *Fusarium graminearum*; *F. oxy* - *F. oxysporum*.

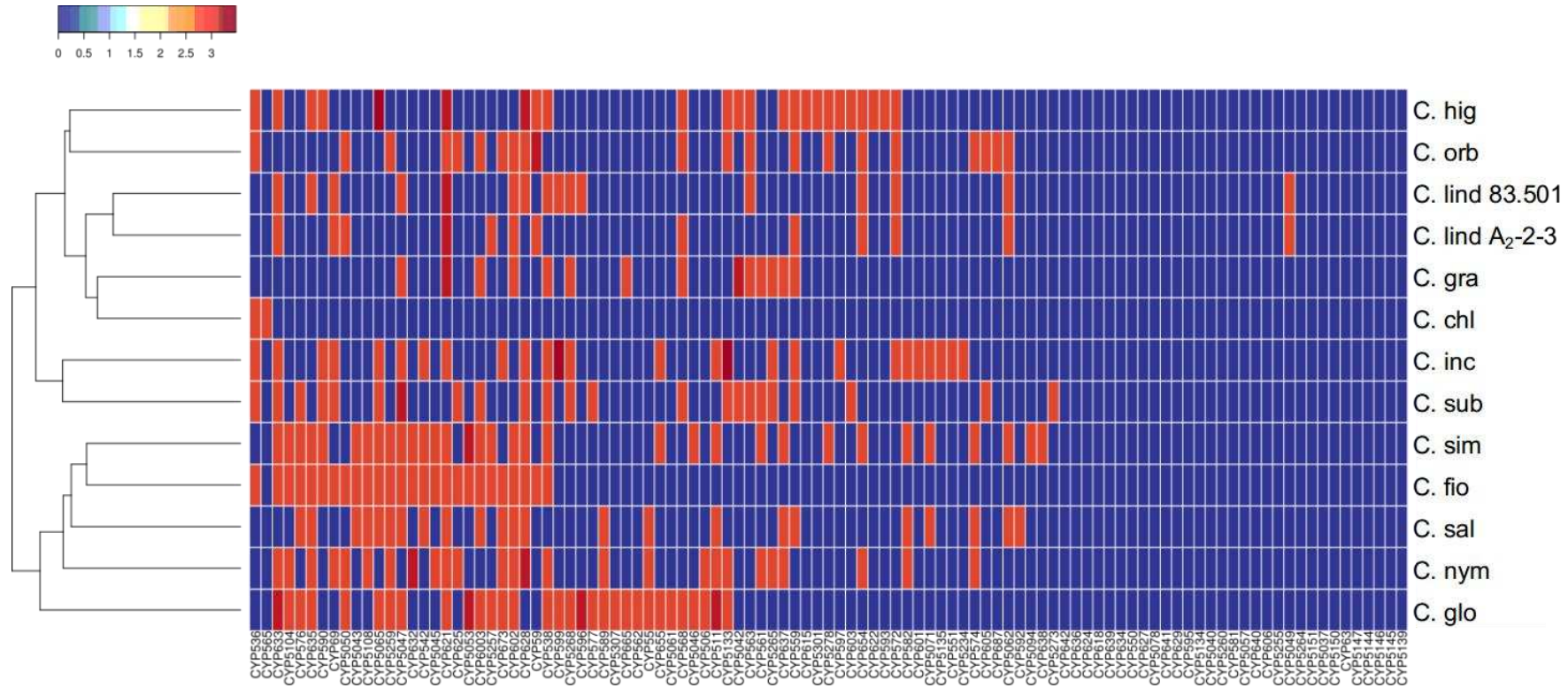


Figura 34. Contrações e expansões de famílias de citocromo P450 mono-oxigenases em *C. lindemuthianum*..

Tabela 7. Contrações e expansões observadas em famílias de P450 mono-oxigenases nos genomas de *C. lindemuthianum*.

Família	Situação
CYP584	*Contraída em C. lind 83, C. sub, C. fio e em C. orb, expandido em C. lind 89 e demais espécies
CYP61	*Expandida em C. lind 83 e outras espécies e contraída em C. lind 89 e outras espécies
CYP630	*Expandida em C. lind 89 e outras espécies e contraída em C. lind 89 e outras espécies
CYP53	*Expandida em C. lind 89 e outras e contraída em C. lind 83 e outras
CYP617	*Contraída em C. lind 89, C. hig e em C. gra, expandida em C. lind 83 e demais espécies
CYP541	*Expandida em C. lind 83 e outras e contraída em C. lind 89 e outras
CYP566	Contraída em C. lind 83 e 89 e em C. gra
CYP5058	*Expandida em C.lind 83 e em C. sub, contraída em C. lind 89 e demais espécies
CYP675	Expandida em C. lind 83 e 89 e em C. orb
CYP623	*Expandida em C. lind 83 e outras e contraída em C. lind 89 e outras
CYP573	*Expandida em C. lind 83 e outras e contraída em C. lin 89 e outras
CYP62	*Expandida em C. lind 89 e outras e contraída em C. lind 83 e outras

CYP684	*Expandida em C. lind 89 e outras e contraída em C. lind 83 e outras
CYP567	*Expandida em C. lind 83 e outras e contraída em C. lind 89 e outras
CYP544	*Expandida em C. lind 89 e outras e contraída em C. lind 83 e outras
CYP526	*Expandida em C. lind 83 e outras e contraída em C. lind 89 e outras
CYP527	*Expandida em C. lind 89 e outras e contraída em C. lind 83 e outras
CYP56	*Expandida em C. lind 89 e outras e contraída em C. lind 83 e outras
CYP5049	Expandida exclusivamente em C. lind 83 e 89
CYP5062	Expandida em C. lind 83 e 89, C. orb, C. sim e C. sal
CYP572	Expandida em C. lind 83 e 89, C. orb, C. hig e C. inc

“ * ” significam que a alteração em questão é divergente entre os genomas dos dois isolados de *C. lindemuthianum*. “outras” indicam que a alteração é compartilhada com outros fungos pertencentes ao gênero *Colletotrichum*. C. gram – *C. graminicola*; C. lind 83.501 e C. lind A₂-2-3 – *C. lindemuthianum*; C. fio – *C. fioriniae*; hig – *C. higginsianum*; C. inc – *C. incanum*; C. orb – *C. orbiculare*; C. sal – *C. salicis*; C. sim – *C. simmondsii*; C. sub – *C. sublineolax*

DISCUSSÃO

No presente trabalho foi realizada a anotação dos genomas de *C. lindemuthianum* (isolados 83.501 e A₂-2-3), que apresenta até o momento o segundo maior genoma reportado para fungos do gênero *Colletotrichum*. Em organismos eucariotos que possuem genomas pequenos, geralmente o tamanho dos genomas é relacionado ao número de genes (Rafaelle & Kamoun, 2012). No entanto, não foi encontrada uma correlação linear entre o tamanho dos genomas e o número de genes preditos para os genomas de fungos do gênero *Colletotrichum*, apesar de existir uma correlação entre o aumento do tamanho dos genomas e a diminuição no conteúdo G+C (%) (**Figura 2**).

Conforme esperado, também observou-se que os fungos do clado orbiculare apresentam uma estrutura de genomas distante daquela apresentada pelas demais espécies de *Colletotrichum*, que de modo geral apresentam genomas compactos com conteúdo G+C (%) superior às espécies do clado orbiculare. Além disto, *C. lindemuthianum* e *C. orbiculare* apresentam genomas com um padrão bimodal, com até 55,5% de sua totalidade sendo referentes a regiões ricas em A+T (**Figura 3**). Esta diminuição no conteúdo G+C (%) e a redução da densidade gênica são indicadores de que a expansão no tamanho dos genomas de *C. lindemuthianum* e *C. orbiculare* pode ter seguido o modelo “mutation-hazard” (Kelkar & Ochman, 2012), onde a expansão do genoma seria dirigida pela deriva genética, possibilitando o acúmulo de elementos móveis, que poderiam contribuir para a existência de regiões genômicas com baixo conteúdo de G+C (%).

Os genomas de *C. lindemuthianum* apresentam um alto percentual de sequências repetitivas considerando-se que os dados utilizados para suas montagens consistem em dados de sequenciamento de segunda geração (Rao et al., 2018). Estes, consistem nos maiores percentuais de sequências repetitivas encontradas para fungos do gênero *Colletotrichum* em dados de genomas montados exclusivamente com dados de tecnologias de sequenciamento de primeira e segunda geração (**dados não apresentados**), e são ainda assim, comparáveis aos conteúdos de montagens de genomas quase completas de fungos do gênero em questão, como por exemplo o das espécies *C. lentis* (4,56%), *C. truncatum* (6,8%) e *C. orbiculare* (8.3%) (Rao et al., 2018; Bhaudaria et al., 2019).

Dallery et al (2017) em trabalho que empregou as tecnologias PacBio de sequenciamento de molécula única em tempo real e mapeamento óptico combinadas para o sequenciamento do genoma de *C. higginsianum* obtiveram uma montagem praticamente livre de *gaps*. Os autores identificaram que os ETs, que na montagem anterior, obtida por meio de dados de *reads* pequenas, correspondiam a somente 1,2 % passaram a corresponder a 7,0 % no genoma completo de *C. higginsianum*. Além disto, as repetições simples cobrem 12,7% do genoma desse. Estas informações indicam que provavelmente o conteúdo real de sequências repetitivas nos genomas de *C. lindemuthianum* é superior ao observado no presente trabalho, e pode estar entre os maiores do gênero.

Dentre as sequências repetitivas identificadas nos genomas de *C. lindemuthianum*, os TEs da Classe I são os mais abundantes, sendo os elementos da ordem LTR pertencentes às superfamílias *Gypsy* e *Copia* os mais representados dentre estes. Estes resultados corroboram com o que foi encontrado para as espécies *C. truncatum*, *C. higginsianum*, *C. graminicola*, *C. scovillei*, *C. chlorophyti* e *C. orchidophilum* (Rao et al., 2018). Apesar da proximidade filogenética e de apresentar diversas características genômicas similares entre si, as espécies *C. lindemuthianum* e *C. orbiculare* parecem apresentar diferenças na distribuição de

seus ETs. Enquanto elementos da superfamília *Gypsy* são os mais abundantes nos genomas de *C. lindemuthianum*, no caso de *C. orbiculare*, Rao et al. (2018) relataram que os elementos da superfamília *Copia* são os mais abundantes, sendo seguidos por elementos da superfamília *Gypsy*. Além disto, observou-se que a maior fração dos TEs no genoma de *C. orbiculare* consiste em elementos desconhecidos, o que contrasta com os dados observados no presente trabalho para a espécie *C. lindemuthianum*, onde foi possível classificar a maior parte dos elementos identificados. No entanto, a abordagem de análise aplicada por Rao et al. (2018) é distinta da abordagem utilizada para *C. lindemuthianum*. Enquanto no presente trabalho foi aplicada a ferramenta *RepeatMasker*, Rao et al. (2018) utilizaram o *RepeatModeler* que realiza a predição “de novo”. Dada a fragmentação das montagens dos genomas de *C. lindemuthianum* não foi possível aplicar tal ferramenta no presente trabalho. Além do mais, a montagem do genoma de *C. orbiculare* está mais próxima da completude do que a de *C. lindemuthianum*, que possui genoma maior do que o de *C. orbiculare*. No futuro, uma nova montagem dos genomas de *C. lindemuthianum* serão realizadas aplicando dados de tecnologias de sequenciamento mais recentes, o que vai possibilitar um melhor acesso ao conteúdo de DNA repetitivo nos genomas desta espécie.

Os genomas de todas as espécies de *Colletotrichum* avaliadas no presente trabalho codificam as metiltransferases RID e DIM-2 (**Tabela 1**). A análise de assinaturas de RIP nos genomas revelou que *C. lindemuthianum* possui os maiores índices de RIP dentre todas as espécies avaliadas, sendo seguido pelas espécies *C. orbiculare*, *C. sublioneola* e *C. graminicola*. Previamente, foi realizada a identificação, anotação e análise comparativa do repertório de genes que codificam proteínas candidatas a efetores em *C. lindemuthianum* e outras espécies do gênero *Colletotrichum* (Queiroz et al., 2019). Assim, foi realizada a identificação de possíveis SNPS entre os genomas dos dois isolados. Os resultados demonstram que a frequência de SNPS em genes que codificam efetores é 2,3 vezes maior do que em

outros genes. Além disto, a frequência de mutações típicas de RIP em genes que codificam candidatos a efetores de *C. lindemuthianum* é quatro vezes maior do que em outras classes de genes (**Figura 4**). Quando expostos em conjunto, o baixo conteúdo G+C (%) dos genomas de *C. lindemuthianum*, associado ao elevado teor de elementos repetitivos, assim como à presença dos genes que codificam as metiltransferases envolvidas em RIP em seus genomas indicam que o mecanismo RIP ocorre ou já ocorreu nos genomas de *C. lindemuthianum* com uma alta frequência e pode ter sido importante-para a modelagem do genoma desta espécie.

Em *C. lindemuthianum* a ocorrência do ciclo sexual foi primeiramente descrita no ano de 1913 (Shear & Wood 1913), sendo este fungo primeiramente descrito como homotálico (Shear and Wood, 1913) e posteriormente como heterotálico (Batista & Chaves, 1982). No entanto, um estudo mais recente demonstrou que o reconhecimento entre os parceiros em *C. lindemuthianum* difere do que foi descrito para outros fungos, pois foi possível a realização de cruzamentos entre isolados que apresentam apenas o idiomorfo *Mat1-2* (Rodriguez-Guerra et al., 2005). No presente trabalho, também não foi encontrado o gene *Mat1-1-1* nos genomas de *C. lindemuthianum* avaliados, nem tampouco nos genomas das demais espécies de *Colletotrichum*. Estes resultados foram comuns a todas as demais espécies do gênero *Colletotrichum*, e quando somados aos vários relatos de ocorrência do ciclo sexual nestes fungos (Shear & Wood 1913; Cisar & TeBeest, 1999; Guerber & Correll, 2001; Guerra et al., 2005; Armstrong-Cho and Banniza, 2006) corroboram com o que foi previamente descrito em trabalhos anteriores (Wheeler, 1954; Cisar & TeBeest 1999; Vaillancourt et al., 2000), que postularam que espécies de *Colletotrichum* não apresentam um sistema de compatibilidade vegetativa bipolar, sendo estes por sua vez detentores de apenas um *locus* de *mating type* com múltiplos alelos alternados. Assim, os fungos pertencentes a este gênero não se encaixam no modelo de base genética para a compatibilidade sexual descrito como 'single locus/two idiomorphs'

descrito como sendo o mais frequente para fungos pertencentes ao filo *Ascomycota* (Glass & Kulda, 1992; Menat et al., 2016).

Os genes que codificam as proteínas PpgA, PreA e PreB não foram encontrados nos genomas de *C. lindemuthianum*. O gene *ppgA* codifica um precursor de feromônio *alpha-factor like*, enquanto, as proteínas PreA e PreB consistem em receptores dos feromônios envolvidos no ciclo sexual. Seo et al. (2004) avaliaram o efeito da mutação dos genes de receptores *preA* e *preB* na espécie *A. nidulans* e observaram que ambos os mutantes produziram um número reduzido de cleistotécios em relação à linhagem selvagem (aproximadamente 5% da produção da linhagem selvagem). Além disto, os autores relataram que o tamanho dos cleistotécios também diminuiu. Guerra et al (2005) observaram que mesmo ao parear isolados de *C. lindemuthianum* sexualmente compatíveis ocorre uma baixa produção de cleistotécios. Esta baixa taxa de produção de estruturas sexuais mesmo quando isolados compatíveis são cruzados poderia ser resultante a ausência dos genes *preA* e *preB* em seu genoma, mas também pode ocorrer devido à ausência de condições ótimas para a ocorrência do cruzamento sexual. O fungo *A. nidulans* por exemplo, reprime o desenvolvimento sexual quando sob incidência de luz no comprimento de onda do vermelho (Blumenstein et al., 2005). Similarmente, Zentmyer et al (2018) observaram diferenças significativas no número e tamanho das estruturas sexuais do Oomiceto *Phytophthora cinanamomi* utilizando o mesmo cruzamento de isolados em diferentes meios de cultivo. Para provar esta hipótese seria requerida a análise funcional destes genes e de seu impacto em *C. lindemuthianum*. Alternativamente, é possível que *C. lindemuthianum* possua outro mecanismo de reconhecimento dos parceiros no cruzamento, ou ainda que este tipo de reconhecimento não seja necessário.

Genes que codificam proteínas homólogas a Sc_TOS7 não foram encontrados nos genomas de *C. lindemuthianum*, resultado este comum à espécie *F. graminearum*, que apresenta um ciclo sexual bem estudado. O papel desta proteína

no processo de meiose ainda não é bem elucidado e apesar de ser listada no banco de dados utilizado como sendo importante para a ocorrência da meiose, o fato de não ser encontrada em *F. graminearum*, demonstra que ela não é essencial para a ocorrência do ciclo sexual, pelo menos neste fungo. Dado que *C. lindemuthianum* consiste na única espécie pertencente ao gênero *Colletotrichum* onde homólogos desta proteína não foram encontrados, são necessários mais estudos acerca do impacto da ausência de *Sc_TOS7* nestes genomas.

Apesar de todos os indícios de ocorrência de RIP e da maquinaria necessária para a ocorrência do ciclo sexual estar conservada em *C. lindemuthianum*, seu ciclo sexual é descrito como raro, o que limitaria o impacto deste mecanismo de silenciamento em seu genoma. Este impacto, que seria em tese fortemente limitado contrasta com as fortes assinaturas de RIP apresentadas pelo genoma de *C. lindemuthianum* (as mais altas para os fungos do gênero avaliados). Gladyshev & Kleckner (2017) descreveram um possível mecanismo de geração de mutações de RIP o qual seria independente da atividade de RID, e por consequente independente da ocorrência do ciclo sexual em fungos. Este mecanismo, dirigido pela proteína DIM-2, em interação com as proteínas DIM-5 (histona metiltransferase) e HP1 (proteína da heterocromatina 1) poderia então ser o principal responsável pela geração de mutações RIP no genoma de *C. lindemuthianum*, explicando assim as fortes assinaturas de RIP em seu genoma mesmo na ausência de um ciclo sexual frequente. Visto que *C. lindemuthianum* é o fungo do gênero *Colletotrichum* com as mais fortes assinaturas de RIP e apresenta um ciclo sexual de rara ocorrência, esta espécie apresenta-se como um modelo interessante para estudos acerca do mecanismo de geração de RIP descrita por Gladyshev & Kleckner (2017).

A comparação dos proteomas dos isolados 83.501 e A₂-2-3 revelaram 52 e 45 genes exclusivos de cada raça, respectivamente. Dentre os genes exclusivos do isolado 83.501 (raça 83) estão genes que codificam proteínas das classes das arielsterases, hidrolases de éster carboxílico, proteínas contendo o domínio

DUF1640, componentes de exocisto, transportadores MFS e proteína de biossíntese de tiamina. Parte destas classes proteicas já foram descritas anteriormente como sendo envolvidas na virulência de fungos fitopatogênicos.

Em estudo de proteômica, Wickramasekara et al. (2011) identificaram uma arilesterase que degrada bisfenolA, um composto poluente, demonstrando a provável capacidade de *C. lindemuthianum* de lidar com compostos poluentes. Pedras et al. (2007) realizaram a identificação do gene que codifica a proteína que é responsável por destoxificar a fitoalexina brassisina em isolados de *Leptosphaeria maculans*, o que resultou na identificação de uma hidrolase de éster carboxílico, induzível responsável por este processo.

Similarmente, proteínas componentes de exocisto foram encontradas dentre os genes exclusivos do isolado 83.501. Sabe-se que os fungos fitopatogênicos manipulam os componentes do exocisto, de maneira a alterar o tráfego vesicular, altamente importante para o funcionamento do sistema de defesa basal das plantas. Transportadores MFS também têm sido relatados como fatores de virulência em fungos, e em *Cercospora nicotianae* um transportador é responsável pelo transporte de cercosporina, uma toxina importante no processo infeccioso (Choquer et al., 2007).

Outro gene exclusivo do isolado 83.501 consiste num gene que codifica uma proteína de biossíntese de tiamina, previamente caracterizada por Roldán et al. (2008) como sendo requerido para a produção de tiamina, e conseqüente maior tolerância ao estresse oxidativo na espécie *F. oxysporum*. A deleção deste gene levou ao fenótipo de auxotrofia para tiamina e de maior sensibilidade ao estresse oxidativo.

Entre os genes exclusivos do genoma do isolado A₂-2-3 estão genes que codificam proteínas adenilato cinases, que são conhecidas como envolvidas no processo de resposta em redes de sinalização (Dzeja & Terzic, 2009). Proteínas

contendo o domínio CFEM também foram previamente descritas como fatores de virulência. A proteína Pth11p de *Manaportha grisea*, que contém o domínio CFEM tem ação no desenvolvimento do apressório, uma fase essencial para a infecção de plantas de arroz (Dezwaan et al., 1999).

Em seu trabalho Gomi, Hiroyuky & Kazuya (2003) realizaram a identificação de uma hidrolase de epóxidos no genoma de *Alternaria alternata*. Acredita-se que esta proteína seja responsável por interferir na biossíntese de cutina e na produção de ácido jasmônico, envolvidos no processo de defesa vegetal contra a invasão de fitopatógenos. Um gene codificando uma proteína da classe MDR (do inglês “multidrug resistance”) também foi classificado como sendo exclusivo do isolado A₂-2-3. Esta classe de proteínas é conhecidamente envolvida com a virulência de fungos fitopatogênicos, tendo ação no fenômeno de resistência por efluxo, onde a resistência contra compostos de defesa vegetais se dá pelo efluxo destas substâncias, processo este mediado por transportadores ABC (De Waard et al., 2006).

Uma das características bastante intrigantes na espécie *C. lindemuthianum* é sua alta variabilidade patogênica (Padder et al., 2017). Os mecanismos genéticos responsáveis por este fenômeno ainda não foram elucidados e podem ser responsáveis pela perda ou aquisição de proteínas efetoras, pelas mutações em genes que codificam proteínas efetoras e fatores de virulência em geral. Assim, possivelmente algumas das proteínas exclusivas dos isolados 83.501 e A₂-2-3 podem ser responsáveis pela capacidade ou incapacidade destes fungos de infectar variedades de feijoeiro.

Conforme o esperado, *C. lindemuthianum* compartilha mais ortogrupos com *C. orbiculare*, a espécie mais proximamente relacionada em termos filogenéticos dentre as analisadas. Foram encontrados 231 ortogrupos exclusivos de *C. lindemuthianum*, dentre os quais através da anotação funcional foi possível notar que existe uma

expansão na classe de proteínas envolvidas no transporte transmembranar, sendo que estas proteínas foram atribuídas as classes de transportadores, proteínas contendo domínios de ligação do NAD(P) e P450 mono-oxigenases, além das atividades de oxidação e redução e proteínas efetoras.

Visto que todas estas classes já são conhecidas como envolvidas na patogenicidade em um grande número de fitopatógenos (Siewers et al., 2005; Choquer et al., 2007; Wilson et al., 2010; Urquiza & Keyhani, 2015), é possível que grande parte das proteínas exclusivas de *C. lindemuthianum* sejam fatores requeridos para o sucesso da infecção de plantas de feijoeiro.

Transportadores consistem em proteínas essenciais a todos os organismos celulares, sendo responsáveis pelo estabelecimento e manutenção de gradientes químicos e elétricos ao longo das membranas celulares (Burgstaller, 1997). Além disto, estas proteínas também são envolvidas na captação de nutrientes e sinalização celular, atuando em vias de transdução de sinal. Alguns transportadores contribuem para a virulência de fitopatógenos, participando da secreção de toxinas e efetores, e outros participam do processo de resistência de fungos a compostos tóxicos presentes no ambiente, sendo este processo denominado de resistência mediada por efluxo (de Waard et al., 2006; Coleman & Mylonakis, 2009).

No presente trabalho foram analisados todos os transportadores descritos no TCDB em 20 espécies de fungos, sendo 12 do gênero *Colletotrichum*. Foi possível observar que *C. lindemuthianum* apresenta um menor repertório de transportadores quando comparado com espécies de fungos mais generalistas como *C. gloeosporioides*. Esta redução pode ser atribuída ao fato desse fungo ser um patógeno especializado na cultura do feijão, o que em tese reduz a quantidade de compostos tóxicos com os quais o mesmo teria de lidar durante o processo de infecção.

Foram identificados 99 transportadores com alterações no número de cópias nos genomas de *C. lindemuthianum*, dentre os quais 19 possuem um número de cópias alterados exclusivamente nesta espécie. Estes transportadores pertencem às famílias *The Drug/Metabolite Transporter (DMT) Superfamily*, *The Major Facilitator Superfamily (MFS)*, *Mitochondrial Carrier Family (MC)*, *The Huntington-interacting protein 14 (HIP14) Family*, *The Chloride Carrier/Channel (CIC) Family* e *The Transient Receptor Potential Ca²⁺ (TRP-CC) Family*.

Transportadores ABC consistem em uma das maiores famílias de proteínas conhecidas na natureza e diversos trabalhos demonstraram que atuam na resistência a drogas e a compostos de defesa vegetal em fungos fitopatogênicos (Kovalchuk & Driessen, 2010; Nakaune et al., 1998; Nakaune et al., 2002; Roohparvar et al., 2007). Similarmente, a superfamília de transportadores MFS também é constituída de proteínas importantes, envolvidas na virulência de fungos fitopatogênicos, tendo sido associadas à secreção de diversas substâncias, tais como toxinas, metabólitos oriundos do ciclo de Krebs, permeases bacterianas, etc (Perlin & San Toh, 2014). Recentemente alguns trabalhos têm demonstrado a contribuição de transportadores MFS na virulência de espécies do gênero *Colletotrichum* e em outros importantes gêneros de fitopatógenos (Liu et al., 2017; Wang et al., 2018; Chen, Zhou, Liu, 2019; Vela-Corcía et al., 2019;).

A família de transportadores mitocondriais (MC) é responsável pelo transporte de cetoácidos, aminoácidos, nucleotídeos, íons inorgânicos e cofatores ao longo da membrana interna (Kunji & Robinson, 2010). Além disto, em fungos filamentosos esses carreadores são também requeridos para fenômenos como a tolerância a stress oxidativo, aquisição de ferro e virulência (Inoue, Namiki, Tsuge, 2002; Yang et al., 2012; Huang et al., 2019).

Adicionalmente, as famílias de transportadores DMT, HIP14, CIC e TRP-CC estão envolvidas nos processos de resistência a drogas, (Martin & Kirk, 2004).

Transportadores da família HIP14 estão relacionados ao processo de *Palmitoylation*, que é um processo de modificação reversível de lipídeos, que desempenha um papel chave em patógenos de todos os tipos (Blanc, Blaskovic, der Goot, 2013). Em trabalho realizado por Sweigard et al (1998), a mutação de uma palmitoil transferase levou à redução dos sintomas da infecção de *Magnaporthe grisea* em folhas destacadas de arroz (*Oryza sativa*).

Transportadores de *Chloride* estão envolvidos com o processo de formação e maturação das estruturas sexuais de *F. graminearum* (Hallen et al., 2007). Adicionalmente, linhagens mutantes do fungo *M. oryzae* defectivas em canais permeáveis de cálcio apresentaram deficiências nos processos de crescimento micelial, esporulação e formação de apressórios, além de apresentarem uma redução na agressividade a diferentes cultivares de arroz.

Quando expostos em conjunto, os dados reforçam a já demonstrada importância do repertório de transportadores para a fisiologia dos fungos e em especial chamam a atenção para sua importância como possíveis fatores de virulência. De fato, para uma conclusão com melhor embasamento, a análise funcional dos transportadores contraídos ou expandidos nos genomas de *C. lindemuthianum* se faz necessária para a comprovação de sua contribuição no processo infeccioso em plantas de feijoeiro. No entanto, foi demonstrado previamente com análises funcionais de genes de transportadores, que estes em geral podem contribuir para virulência em *C. lindemuthianum*, seja em casos onde a inativação destes genes diminuem a agressividade deste fungo ou em casos onde a inativação de um transportador propicia o aumento da virulência neste fitopatógeno (Oliveira, 2009; Pereira, 2011). Além disto, diversos outros trabalhos também demonstram sua participação no processo infeccioso de diferentes fungos fitopatogênicos (de Waard et al., 2006; Coleman & Mylonakis, 2009). Como as alterações se tratam de contrações e expansões e não de presença e ausência de determinados transportadores nos genomas de *C. lindemuthianum*, o estudo da contribuição dos produtos destes genes

requer sistemas de marcadores recicláveis para a transformação sequencial destes fungos, como por exemplo o sistema proposto por Kumakura, Ueno, Shirasu (2018) para o fungo *C. orbiculare*.

O repertório de CAZymes secretadas das espécies de *Colletotrichum* é de modo geral maior do que as demais espécies de fungos avaliadas. Este grande potencial de produção das espécies de *Colletotrichum* já havia sido demonstrado anteriormente (O' Connel et al., 2012; Gan et al., 2013) e pode ser interessante do ponto de vista de aplicações biotecnológicas para estes fungos. O dendrograma baseado no padrão de distribuição de todas as famílias de CAZymes secretadas identificadas nas espécies de *Colletotrichum* demonstra que o conteúdo dos genes que codificam tais enzimas não é moldado pelas relações filogenéticas nem tampouco pela gama de hospedeiros (**Figura 25**).

No entanto, ao restringir a análise às famílias de CAZymes secretadas que são conhecidas pela atividade na degradação de material vegetal, *C. lindemuthianum* agrupou com *C. incanum*, uma espécie que causa doença nas plantas de soja (*Glycine max*) (**Figura 27**). A soja é uma planta pertencente a mesma família do feijoeiro (*Fabaceae*) e compartilha diversas características fisiológicas e estruturais com este, sendo portanto esperado que o repertório de enzimas degradadoras de carboidratos vegetais empregados por estas duas espécies de fitopatógenos seja bastante semelhante.

Foram encontradas 10 famílias de CAZymes secretadas envolvidas na degradação de biomassa vegetal apresentando alterações no número de cópias nos genomas de *C. lindemuthianum*, das quais quatro consistem em contrações e seis em expansões (**Tabela 6**). A maioria destas famílias participa da degradação de hemicelulose (7), enquanto duas estão envolvidas na degradação de pectina, e uma está envolvida na degradação de celulose (**Tabela 6**). Diferentemente do que foi encontrado para transportadores, proteases e NRPSs, onde a maioria das alterações

no número de cópias é convergente nos genomas dos dois isolados de *C. lindemuthianum*, no caso das CAZymes secretadas envolvidas na degradação de biomassa vegetal a maior parte destas foi divergente entre os isolados, sendo contraída no genoma de um e expandida no genoma do outro.

Apesar destes fungos pertencerem à mesma espécie e infectarem o mesmo hospedeiro, as raças de *C. lindemuthianum* são altamente especializadas por cultivares de feijoeiro específicos. Sendo assim, não seria surpreendente que a alteração no repertório de CAZymes nestes fungos dirigisse sua especialização por cultivares de feijoeiro.

Foram encontradas alterações no número de cópias em quatro famílias de proteases secretadas em *C. lindemuthianum*, sendo todas estas compartilhadas com uma ou mais espécies do gênero *Colletotrichum* (**Figura 29**). A família S12 está contraída nos genomas dos dois isolados de *C. lindemuthianum*, em *C. graminicola* e em *C. higginsianum*. A família M12 está contraída nos genomas de *C. lindemuthianum* e em *C. graminicola*. A família C69 encontra-se expandida nos genomas de *C. lindemuthianum* e em *C. fioriniae*. E por último, a família S01 encontra-se expandida em *C. lindemuthianum* e em *C. fioriniae*.

A família de serina peptidase S12, contraída nos genomas de *C. lindemuthianum* está envolvida na degradação de quitinases, e de acordo com os resultados obtidos por Muszewska et al (2017) em trabalho que analisou os proteomas preditos de mais de 600 fungos, não faz parte do *core* de serina peptidases fúngicas. Esta família encontra-se expandida no genoma de *C. truncatum* (Rao & Nandineni, 2017), sendo possivelmente atuante na defesa contra quitinases vegetais.

Metaloproteases consistem em enzimas hidrolíticas que atuam na interação de fitopatógenos com o hospedeiro, podendo consistir em fatores de toxicidade (Zhou et al., 2018). De acordo com Lévesque et al (2010), a família de metaloproteases M12,

contraída nos genomas de *C. lindemuthianum* é notavelmente expandida no genoma do fitopatógeno necrotrófico *Pythium ultimum* em relação a espécies de oomicetos hemibiotróficos. Assim, é previsível que esta família esteja contraída nas espécies de fungos hemibiotróficos analisados no presente trabalho. No entanto, a contração foi encontrada também na espécie de fungo necrotrófico *R. solani*, levantando a hipótese da atuação ou não das proteínas desta família no processo infeccioso.

De acordo com Dubovenko et al (2010) *trypsins* peptidases como as que pertencem à família S01 podem ser utilizadas como marcadores da fitopatogenicidade em fungos. Estas proteínas podem possuir as funções de chaperonas e atuar na degradação extracelular (Monod et al., 2002; Padmanabhan et al., 2008). É provável que a expansão da família S01 esteja relacionada à virulência de *C. lindemuthianum*. Assim, todos os indícios apontam para uma contribuição importante da classe das proteases secretadas na especialização e virulência de *C. lindemuthianum*.

Metabólitos secundários (MS) consistem em produtos não requeridos para o crescimento, os quais são produzidos com o intuito de obter alguma vantagem adaptativa em relação ao ambiente ou em relação a outras formas de vida, tais como sideróforos, micotoxinas, antibióticos e pigmentos (Keller & Bennet, 2005). No presente trabalho foram analisados os genes de NRPSs, PKSs e de P450 mono-oxigenases, classes importantes de genes que atuam na síntese metabólitos secundários 'MS'. As espécies do gênero *Colletotrichum* em geral demonstraram um poder de produção de peptídeos não ribossomais superior ao apresentado pelas outras espécies de fungos analisados neste trabalho. No entanto, *C. lindemuthianum* apresenta um número intermediário de NRPSs em relação aos fungos analisados. A busca por ortólogos de genes que codificam NRPSs revelou uma provável capacidade de produção de uma gama diversa de peptídeos por este fungo, sendo a maioria destes conservado entre os genomas dos dois isolados, com exceção do gene *S89_g2060* que encontra-se ausente no isolado 83.501 e em *C. orbiculare*.

Dentre as funções dos produtos das proteínas NRPSs codificadas pelos ortólogos de *C. lindemuthianum* destacam-se toxinas como a AcIP, aspercriptina e sideróforos como SidD.

Apesar do repertório de P450 mono-oxigenases em *C. lindemuthianum* ser um dos menores dentro do gênero *Colletotrichum* (**Figura 32**), foram encontradas alterações no número de cópias em 21 famílias pertencentes a esta classe (**Figuras 33 e 34, Tabela 7**). A maior parte destas alterações consiste em expansões e intrigantemente não são convergentes nos genomas dos dois isolados, apontando para uma possível importância na interação com os genótipos de feijoeiro, dado que a única diferença em relação ao ciclo de vida destes fitopatógenos são os genótipos os quais os mesmos podem infectar.

CONCLUSÃO

O genoma de *C. lindemuthianum* apresenta uma arquitetura bipartida típica de genomas de duas velocidades, com uma maior frequência de ocorrência de SNPS e mutações típicas de RIP em genes que codificam proteínas candidatas a efetores, bem como com uma maior proximidade destes à regiões ricas em repetições. A expansão no tamanho de seus genomas não se deu em termos de aumento de número de genes, nem tampouco de famílias de genes que codificam fatores de virulência conhecidos, sendo muito provavelmente associada ao aumento no teor de elementos repetitivos. *Colletotrichum lindemuthianum* apresenta em seu genoma alterações no números de cópias de importantes famílias de genes, quando comparado às demais espécies de *Colletotrichum*. Estas alterações foram observadas em famílias gênicas previamente relatadas como sendo envolvidas na virulência de fungos fitopatogênicos. Alterações nos números de cópias destas classes já haviam sido relatadas para fungos do gênero *Colletotrichum* e outros gêneros. No entanto, no presente foi possível identificar contrações e expansões em diversas famílias gênicas exclusivas de *C. lindemuthianum*, e além disto foram

apontadas divergências entre os genomas dos isolados que pertencem a duas raças patogênicas distintas, as quais poderiam explicar a variação patogênica desta espécie. Estas informações podem ser exploradas futuramente em trabalhos de análise funcional visando elucidar quais são os mecanismos responsáveis pela especificidade de *C. lindemuthianum* por plantas de feijoeiro e também acerca dos mecanismos geradores da variabilidade patogênica desta espécie.

Em relação ao conteúdo de sequências repetitivas no genoma de *C. lindemuthianum*, trabalhos, empregando tecnologias de sequenciamento que forneçam uma montagem de seu genoma com uma maior resolução, poderiam possibilitar a melhor identificação dos elementos e sua localização nos cromossomos.

REFERÊNCIAS

- Armstrong-Cho, C. L., & Banniza, S. (2006). *Glomerella truncata* sp. nov., the teleomorph of *Colletotrichum truncatum*. *Mycological research*, 110(8), 951-956.
- Balardin, R. S., Jarosz, A. M., & Kelly, J. D. (1997). Virulence and molecular diversity in *Colletotrichum lindemuthianum* from South, Central, and North America. *Phytopathology*, 87(12), 1184-1191.
- Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile Dna*, 6(1), 11.
- Baroncelli, R., Amby, D. B., Zapparata, A., Sarrocco, S., Vannacci, G., Le Floch, G., ... & Thon, M. R. (2016). Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC genomics*, 17(1), 555.
- Barrus, M. F. (1918). Varietal Susceptibility of Beans to Strains of *Colletotrichum lindemuthianum* (Sacc. and Magn.) B. Et C.
- Bell, J. N., Dixon, R. A., Bailey, J. A., Rowell, P. M., & Lamb, C. J. (1984). Differential induction of chalcone synthase mRNA activity at the onset of phytoalexin accumulation in compatible and incompatible plant-pathogen interactions. *Proceedings of the National Academy of Sciences*, 81(11), 3384-3388.
- Blanc, M., Blaskovic, S., & van der Goot, F. G. (2013). Palmitoylation, pathogens and their host.
- Blumenstein, A., Vienken, K., Tasler, R., Purschwitz, J., Veith, D., Frankenberg-Dinkel, N., & Fischer, R. (2005). The *Aspergillus nidulans* phytochrome FphA represses sexual development in red light. *Current Biology*, 15(20), 1833-1838.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.
- Braga, RM; Santana, MF; Costa, RV; Brommonschenkel, SH; Araújo, EF; Queiroz, MV. (2014). Transposable elements belonging to the Tc1-Mariner superfamily are heavily mutated in *Colletotrichum graminicola*. *Mycologia*, 106(4), 2014, pp. 000–000. DOI: 10.3852/13–262
- Burgstaller, W. (1997). Transport of small ions and molecules through the plasma membrane of filamentous fungi. *Critical reviews in microbiology*, 23(1), 1-46.
- Cannon, P. F., Damm, U., Johnston, P. R., & Weir, B. S. (2012). *Colletotrichum*—current status and future directions. *Studies in mycology*, 73, 181-213.
- Chankhamjon, P., Boettger - Schmidt, D., Scherlach, K., Urbansky, B., Lackner, G., Kalb, D., ... & Hertweck, C. (2014). Biosynthesis of the halogenated mycotoxin aspirochlorine in koji mold involves a cryptic amino acid conversion. *Angewandte Chemie*, 126(49), 13627-13631.

Chen, N. (2004). Using Repeat Masker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics*, 5(1), 4-10.

Chen, N. (2004). Using Repeat Masker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics*, 5(1), 4-10.

Ili, R., Amby, D. B., Zapparata, A., Sarrocco, S., Vannacci, G., Le Floch, G., ... & Thon, M. R. (2016). Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC genomics*, 17(1), 555.

Chen, Y. Y., Conner, R. L., Gillard, C. L., McLaren, D. L., Boland, G. J., Balasubramanian, P. M., ... & Babcock, C. (2013). A quantitative real - time PCR assay for detection of *Colletotrichum lindemuthianum* in navy bean seeds. *Plant pathology*, 62(4), 900-907.

Chen, Y., Zhou, G., & Liu, J. (2019). A major facilitator superfamily transporter in *Colletotrichum fructicola* (CfMfs1) is required for sugar transport, appressorial turgor pressure, conidiation and pathogenicity. *Forest Pathology*, 49(6), e12558.

Cisar, C. R., & TeBeest, D. O. (1999). Mating system of the filamentous ascomycete, *Glomerella cingulata*. *Current genetics*, 35(2), 127-133.

Crossen-Fassoni, A., Bazzolli, D. M. S., Brommonschenkel, S. H., de Araújo, E. F., & de Queiroz, M. V. (2013). The pectate lyase encoded by the *pecCI1* gene is an important determinant for the aggressiveness of *Colletotrichum lindemuthianum*. *Journal of Microbiology*, 51(4), 461-470.

Coleman, J. J., & Mylonakis, E. (2009). Efflux in fungi: la piece de resistance. *PLoS Pathog*, 5(6), e1000486.

Dallery, J. F., Lapalu, N., Zampounis, A., Pigné, S., Luyten, I., Amselem, J., ... & Auger, A. (2017). Gapless genome assembly of *Colletotrichum higginsianum* reveals chromosome structure and association of transposable elements with secondary metabolite gene clusters. *BMC genomics*, 18(1), 667.

de Queiroz, C. B., Correia, H. L. N., Menicucci, R. P., Vidigal, P. M. P., & de Queiroz, M. V. (2017). Draft genome sequences of two isolates of *Colletotrichum lindemuthianum*, the causal agent of anthracnose in common beans. *Genome Announc.*, 5(18), e00214-17.

de Queiroz, C. B., Correia, H. L. N., Santana, M. F., Batista, D. S., Vidigal, P. M. P., Brommonschenkel, S. H., & de Queiroz, M. V. (2019). The repertoire of effector candidates in *Colletotrichum lindemuthianum* reveals important information about *Colletotrichum* genus lifestyle. *Applied microbiology and biotechnology*, 103(5), 2295-2309.

de Queiroz, C. B., Santana, M. F., Vidigal, P. M. P., & de Queiroz, M. V. (2018). Comparative analysis of the mitochondrial genome of the fungus *Colletotrichum lindemuthianum*, the causal agent of anthracnose in common beans. *Applied*

microbiology and biotechnology, 102(6), 2763-2778.

De Waard, M. A., Andrade, A. C., Hayashi, K., Schoonbeek, H. J., Stergiopoulos, I., & Zwiars, L. H. (2006). Impact of fungal drug transporters on fungicide sensitivity, multidrug resistance and virulence. *Pest Management Science: formerly Pesticide Science*, 62(3), 195-207.

Dean, R., Van Kan, J. A., Pretorius, Z. A., Hammond - Kosack, K. E., Di Pietro, A., Spanu, P. D., ... & Foster, G. D. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Molecular plant pathology*, 13(4), 414-430.

DeZwaan, T. M., Carroll, A. M., Valent, B., & Sweigard, J. A. (1999). Magnaporthe grisea pth11p is a novel plasma membrane protein that mediates appressorium differentiation in response to inductive substrate cues. *The Plant Cell*, 11(10), 2013-2030.

Dos Santos, L. V., de Queiroz, M. V., Santana, M. F., Soares, M. A., de Barros, E. G., de Araújo, E. F., & Langin, T. (2012). Development of new molecular markers for the Colletotrichum genus using RetroCl1 sequences. *World Journal of Microbiology and Biotechnology*, 28(3), 1087-1095.

Dow, J. M., Davies, H. A., & Daniels, M. J. (1998). A metalloprotease from Xanthomonas campestris that specifically degrades proline/hydroxyproline-rich glycoproteins of the plant extracellular matrix. *Molecular plant-microbe interactions*, 11(11), 1085-1093.

Dubovenko, A. G., Dunaevsky, Y. E., Belozersky, M. A., Oppert, B., Lord, J. C., & Elpidina, E. N. (2010). Trypsin-like proteins of the fungi as possible markers of pathogenicity. *Fungal biology*, 114(2-3), 151-159.

Dyer, P. S., Paoletti, M., & Archer, D. B. (2003). Genomics reveals sexual secrets of Aspergillus. *Microbiology*, 149(9), 2301-2303.

Dzeja, P., & Terzic, A. (2009). Adenylate kinase and AMP signaling networks: metabolic monitoring, signal communication and body energy sensing. *International journal of molecular sciences*, 10(4), 1729-1772.

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

Eisenhaber, B., Bork, P., & Eisenhaber, F. (1998). Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. *Protein engineering*, 11(12), 1155-1161.

Emanuelsson, O., Brunak, S., Von Heijne, G., & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols*, 2(4), 953.

Fontenelle, M. R., Santana, M. F., Cnossen, A., Bazzolli, D. M., Bromonschenkel, S. H., de Araújo, E. F., & de Queiroz, M. V. (2017). Differential expression of genes

during the interaction between *Colletotrichum lindemuthianum* and *Phaseolus vulgaris*. *European journal of plant pathology*, 147(3), 653-670.

Freitag, M., Williams, R. L., Kothe, G. O., Selker, E. U. (2002). A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. *Proceedings of the National Academy of Sciences USA* 99(13), 8802–8807.

Gacura, M. D., Sprockett, D. D., Heidenreich, B., & Blackwood, C. B. (2016). Comparison of pectin-degrading fungal communities in temperate forests using glycosyl hydrolase family 28 pectinase primers targeting Ascomycete fungi. *Journal of microbiological methods*, 123, 108-113.

Galagan, J. E., Calvo, S. E., Cuomo, C., Ma, L. J., Wortman, J. R., Batzoglou, S., ... & Kapitonov, V. (2005). Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*, 438(7071), 1105.

Gan, P., Ikeda, K., Irieda, H., Narusaka, M., O'Connell, R. J., Narusaka, Y., ... & Shirasu, K. (2013). Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytologist*, 197(4), 1236-12

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 biology and evolution*, 8(5), 1467-1481.

Gan, P., Tsushima, A., Narusaka, M., Narusaka, Y., Takano, Y., Kubo, Y., & Shirasu, K. (2019). Genome sequence resources for four phytopathogenic fungi from the *Colletotrichum orbiculare* species complex. *Molecular Plant-Microbe Interactions*, 32(9), 1088-1090.

Garrison, E., & Marth, G. (2012). FreeBayes. arXiv preprint1207. 3907 [q-bio. GN][Internet].

Glass, N. L., & Kuldau, G. A. (1992). Mating type and vegetative incompatibility in filamentous ascomycetes. *Annual review of phytopathology*, 30(1), 201-224.

Gomi, K., Yamamoto, H., & Akimitsu, K. (2003). Epoxide hydrolase: a mRNA induced by the fungal pathogen *Alternaria alternata* on rough lemon (*Citrus jambhiri* Lush). *Plant molecular biology*, 53(1-2), 189-199.

Guerber, J. C., & Correll, J. C. (2001). Characterization of *Glomerella acutata*, the teleomorph of *Colletotrichum acutatum*. *Mycologia*, 93(1), 216-229.

Hacquard, S., Kracher, B., Hiruma, K., Münch, P. C., Garrido-Oter, R., Thon, M. R., ... & Henrissat, B. (2016). Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nature communications*, 7, 11362.

Hallen, H. E., Huebner, M., Shiu, S. H., Güldener, U., & Trail, F. (2007). Gene

expression shifts during perithecium development in *Gibberella zeae* (anamorph *Fusarium graminearum*), with particular emphasis on ion transport proteins. *Fungal Genetics and Biology*, 44(11), 1146-1156.

Hane, J. K., & Oliver, R. P. (2008). RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC bioinformatics*, 9(1), 478.

Hane, J. K., Anderson, J. P., Williams, A. H., Sperschneider, J., & Singh, K. B. (2014). Genome sequencing and comparative genomics of the broad host-range pathogen *Rhizoctonia solani* AG8. *PLoS genetics*, 10(5), e1004281.

He C, Nourse JP, Kelemu S, Irwin JA, Manners JM (1996) Cgt1: a nonLTR retrotransposon with restricted distribution in the fungal phytopathogen *Colletotrichum gloeosporioides*. *Mol Gen Genet* 252(3): 320–331. <https://doi.org/10.1007/BF02173778>

Hoi, J. W. S., Herbert, C., Bacha, N., O'Connell, R., Lafitte, C., Borderies, G., ... & Dumas, B. (2007). Regulation and role of a STE12 - like transcription factor from the plant pathogen *Colletotrichum lindemuthianum*. *Molecular microbiology*, 64(1), 68-82.

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic acids research*, 35(suppl_2), W585-W587.

Huang, J., Ma, Z., Zhong, G., Sheppard, D. C., Lu, L., & Zhang, S. (2019). The mitochondrial thiamine pyrophosphate transporter TptA promotes adaptation to low iron conditions and virulence in fungal pathogen *Aspergillus fumigatus*. *Virulence*, 10(1), 234-247.

Hyde, K. D., Cai, L., Cannon, P. F., Crouch, J. A., Crous, P. W., Damm, U., ... & Liu, Z. Y. (2009). *Colletotrichum*—names in current use. *Fungal Diversity*, 39(1), 147-182.

Inoue, I., Namiki, F., & Tsuge, T. (2002). Plant colonization by the vascular wilt fungus *Fusarium oxysporum* requires FOW1, a gene encoding a mitochondrial protein. *The Plant Cell*, 14(8), 1869-1883.

Irelan, J. T., & Selker, E. U. (1996). Gene silencing in filamentous fungi: RIP, MIP and quelling. *Journal of Genetics*, 75(3), 313-324.

Ishikawa, F. H., Souza, E. A., Read, N. D., & Roca, M. G. (2010). Live-cell imaging of conidial fusion in the bean pathogen, *Colletotrichum lindemuthianum*. *Fungal biology*, 114(1), 2-9.

Ishikawa, F. H., Souza, E. A., Read, N. D., & Roca, M. G. (2013). *Colletotrichum lindemuthianum* exhibits different patterns of nuclear division at different stages in its vegetative life cycle. *Mycologia*, 105(4), 795-801.

Jashni, M. K., Dols, I. H., Iida, Y., Boeren, S., Beenen, H. G., Mehrabi, R., ... & de Wit,

P. J. (2015). Synergistic action of a metalloprotease and a serine protease from *Fusarium oxysporum* f. sp. *lycopersici* cleaves chitin-binding tomato chitinases, reduces their antifungal activity, and enhances fungal virulence. *Molecular Plant-Microbe Interactions*, 28(9), 996-1008.

Janert, P. K. (2010). *Gnuplot in action: understanding data with graphs*. Manning.

Jashni, M. K., Mehrabi, R., Collemare, J., Mesarich, C. H., & De Wit, P. J. (2015). The battle in the apoplast: further insights into the roles of proteases and their inhibitors in plant-pathogen interactions. *Frontiers in plant science*, 6, 584.

Jia, Y., McAdams, S. A., Bryan, G. T., Hershey, H. P., & Valent, B. (2000). Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *The EMBO journal*, 19(15), 4004-4014.

Käll, L., Krogh, A., & Sonnhammer, E. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic acids research*, 35(suppl_2), W429-W432.

Kashyap, D. R., Vohra, P. K., Chopra, S., & Tewari, R. (2001). Applications of pectinases in the commercial sector: a review. *Bioresource technology*, 77(3), 215-227.

Kelkar, Y. D., & Ochman, H. (2012). Causes and consequences of genome expansion in fungi. *Genome Biology and Evolution*, 4(1), 13-23.

Keller, N. P., Turner, G., & Bennett, J. W. (2005). Fungal secondary metabolism—from biochemistry to genomics. *Nature Reviews Microbiology*, 3(12), 937-947.

keto acids, amino acids, nucleotides, inorganic ions and co-factors across the mitochondrial inner membrane

Kovalchuk, A., & Driessen, A. J. (2010). Phylogenetic analysis of fungal ABC transporters. *BMC genomics*, 11(1), 1-21.

Kovalchuk, A., & Driessen, A. J. (2010). Phylogenetic analysis of fungal ABC transporters. *BMC genomics*, 11(1), 1-21.

Kouzminova, E., Selker, E. U. (2001). *dim-2* encodes a DNA methyltransferase responsible for all known cytosine methylation in *Neurospora*. *EMBO J*, 20, 4309–4323.

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, 305(3), 567-580.

Kumakura, N., Ueno, A., & Shirasu, K. (2019). Establishment of a selection marker recycling system for sequential transformation of the plant - pathogenic fungus *Colletotrichum orbiculare*. *Molecular plant pathology*, 20(3), 447-459.

- Kunji, E. R., & Robinson, A. J. (2010). Coupling of proton and substrate translocation in the transport cycle of mitochondrial carriers. *Current opinion in structural biology*, 20(4), 440-447.
- Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research*, 35(9), 3100-3108.
- Lengeler, K. B., Davidson, R. C., D'souza, C., Harashima, T., Shen, W. C., Wang, P., ... & Heitman, J. (2000). Signal transduction cascades regulating fungal development and virulence. *Microbiol. Mol. Biol. Rev.*, 64(4), 746-785.
- Lévesque, C. A., Brouwer, H., Cano, L., Hamilton, J. P., Holt, C., Huitema, E., ... & Zerillo, M. M. (2010). Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome biology*, 11(7), R73.
- Liang, X., Wang, B., Dong, Q., Li, L., Rollins, J. A., Zhang, R., & Sun, G. (2018). Pathogenic adaptations of *Colletotrichum* fungi revealed by genome wide gene family evolutionary analyses. *PloS one*, 13(4), e0196303.
- Li, H., & Durbin, R. (2009). Making the Leap: Maq to BWA. *Mass Genomics*, 25, 1754-1760.
- Liu, G., Kennedy, R., Greenshields, D. L., Peng, G., Forseille, L., Selvaraj, G., & Wei, Y. (2007). Detached and attached *Arabidopsis* leaf assays reveal distinctive defense responses against hemibiotrophic *Colletotrichum* spp. *Molecular Plant-Microbe Interactions*, 20(10), 1308-1319.
- Liu, L., Yan, Y., Huang, J., Hsiang, T., Wei, Y., Li, Y., ... & Zheng, L. (2017). A novel MFS transporter gene ChMfs1 is important for hyphal morphology, conidiation, and pathogenicity in *Colletotrichum higginsianum*. *Frontiers in microbiology*, 8, 1953.
- Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research*, 25(5), 955-964.
- Mahuku, G. S., & Riascos, J. J. (2004). Virulence and molecular diversity within *Colletotrichum lindemuthianum* isolates from Andean and Mesoamerican bean varieties and regions. *European Journal of Plant Pathology*, 110(3), 253-263.
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., ... & Gwadz, M. (2016). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic acids research*, 45(D1), D200-D203.
- Martin, R. E., & Kirk, K. (2004). The malaria parasite's chloroquine resistance transporter is a member of the drug/metabolite transporter superfamily. *Molecular biology and evolution*, 21(10), 1938-1949.
- Menat, J., Armstrong-Cho, C., & Banniza, S. (2016). Lack of evidence for sexual

- reproduction in field populations of *Colletotrichum lentis*. *Fungal Ecology*, 20, 66-74.
- Monod, M., Capoccia, S., Léchenne, B., Zaugg, C., Holdom, M., & Jousson, O. (2002). Secreted proteases from pathogenic fungi. *International Journal of Medical Microbiology*, 292(5-6), 405-419.
- Moran, G. P., Sanglard, D., Donnelly, S. M., Shanley, D. B., Sullivan, D. J., & Coleman, D. C. (1998). Identification and expression of multidrug transporters responsible for fluconazole resistance in *Candida dubliniensis*. *Antimicrobial Agents and Chemotherapy*, 42(7), 1819-1830.
- Muszevska, A., Stepniewska-Dziubinska, M. M., Steczkiewicz, K., Pawlowska, J., Dziedzic, A., & Ginalski, K. (2017). Fungal lifestyle reflected in serine protease repertoire. *Scientific reports*, 7(1), 1-12.
- Nakaune, R., Adachi, K., Nawata, O., Tomiyama, M., Akutsu, K., & Hibi, T. (1998). A novel ATP-binding cassette transporter involved in multidrug resistance in the phytopathogenic fungus *Penicillium digitatum*. *Applied and Environmental Microbiology*, 64(10), 3983-3988.
- Nakaune, R., Hamamoto, H., Imada, J., Akutsu, K., & Hibi, T. (2002). A novel ABC transporter gene, PMR5, is involved in multidrug resistance in the phytopathogenic fungus *Penicillium digitatum*. *Molecular Genetics and Genomics*, 267(2), 179-185.
- Nelson, D. R. (2009). The cytochrome p450 homepage. *Human genomics*, 4(1), 59.
- Nogueira, G. B., dos Santos, L. V., de Queiroz, C. B., Corrêa, T. L. R., Menicucci, R. P., Bazzolli, D. M. S., ... & de Queiroz, M. V. (2019). The histidine kinase *slnCl1* of *Colletotrichum lindemuthianum* as a pathogenicity factor against *Phaseolus vulgaris* L. *Microbiological research*, 219, 110-122.
- O'Connell, R. J., Bailey, J. A., & Richmond, D. V. (1985). Cytology and physiology of infection of *Phaseolus vulgaris* by *Colletotrichum lindemuthianum*. *Physiological plant pathology*, 27(1), 75-98.
- O'CONNELL, R. J. (1987). ABSENCE OF A SPECIALIZED INTERFACE BETWEEN INTRACELLULAR HYPHAE OF COLLETOTRICHUM LINDEMUTHIANUM AND CELLS OF PHASEOLUS VULGARIS. *New Phytologist*, 107(4), 725-734.
- O'Connell, R. J., & Ride, J. P. (1990). Chemical detection and ultrastructural localization of chitin in cell walls of *Colletotrichum lindemuthianum*. *Physiological and Molecular Plant Pathology*, 37(1), 39-53.
- O'CONNELL, R. J., Pain, N. A., Hutchison, K. A., Jones, G. L., & Green, J. R. (1996). Ultrastructure and composition of the cell surfaces of infection structures formed by the fungal plant pathogen *Colletotrichum lindemuthianum*. *Journal of Microscopy*, 181(2), 204-212.
- Ortiz-Urquiza, A., & Keyhani, N. O. (2015). Stress response signaling and virulence: insights from entomopathogenic fungi. *Current Genetics*, 61(3), 239-249.

- Padder, B. A., Sharma, P. N., Awale, H. E., & Kelly, J. D. (2017). *Colletotrichum lindemuthianum*, the causal agent of bean anthracnose. *Journal of Plant Pathology*, 99(2), 317-330.
- Padmanabhan, N., Fichtner, L., Dickmanns, A., Ficner, R., Schulz, J. B., & Braus, G. H. (2009). The yeast HtrA orthologue Ynm3 is a protease with chaperone activity that aids survival under heat stress. *Molecular biology of the cell*, 20(1), 68-77.
- Parisot, D., Dufresne, M., Veneault, C., Lauge, R., & Langin, T. (2002). *clap1*, a gene encoding a copper-transporting ATPase involved in the process of infection by the phytopathogenic fungus *Colletotrichum lindemuthianum*. *Molecular genetics and genomics*, 268(2), 139-151.
- Pellier, A. L., Laugé, R., Veneault - Fourrey, C., & Langin, T. (2003). CLNR1, the AREA/NIT2 - like global nitrogen regulator of the plant fungal pathogen *Colletotrichum lindemuthianum* is required for the infection cycle. *Molecular microbiology*, 48(3), 639-655.
- Perlin, M. H., Andrews, J., & San Toh, S. (2014). Essential letters in the fungal alphabet: ABC and MFS transporters and their roles in survival and pathogenicity. In *Advances in Genetics* (Vol. 85, pp. 201-253). Academic Press.
- Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*, 8(10), 785.
- Pinto, J. M. A., Pereira, R., Mota, S. F., Ishikawa, F. H., & Souza, E. A. (2012). Investigating phenotypic variability in *Colletotrichum lindemuthianum* populations. *Phytopathology*, 102(5), 490-497.
- Pitkin, J. W., Panaccione, D. G., & Walton, J. D. (1996). A putative cyclic peptide efflux pump encoded by the TOXA gene of the plant-pathogenic fungus *Cochliobolus carbonum*. *Microbiology*, 142(6), 1557-1565.
- Pöggeler, S., & Kück, U. (2001). Identification of transcriptionally expressed pheromone receptor genes in filamentous ascomycetes. *Gene*, 280(1-2), 9-17.
- Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology*, 10(6), 417-430.
- Rao, S., & Nandineni, M. R. (2017). Genome sequencing and comparative genomics reveal a repertoire of putative pathogenicity genes in chilli anthracnose fungus *Colletotrichum truncatum*. *PloS one*, 12(8), e0183567.
- Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., & Finn, R. D. (2017). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic acids research*, 46(D1), D624-D632.

RJ O'CONNELL absence of aspecialized interface between intracellular hyphae of *Colletotrichum lindemuthianum* and cells of *Phaseolus vulgaris*. *New Phytologist*, 107: 725-734. 1987

RJ O'Connell, JA Bailey, DV Richmond Cytology and physiology of infection of *Phaseolus vulgaris* by *Colletotrichum lindemuthianum*. *Physiological Plant Pathology*, 27, Issue 1, July 1985, Pages 75-98.

RJ O'Connell, JP Ride. xChemical detection and ultrastructural localization of chitin in cell walls of *Colletotrichum lindemuthianum*. *Physiological and Molecular Plant Pathology*, 37, Issue 1, July 1990, Pages 39-53.

RJ O'CONNELL, NA Pain, KA Hutchison, GL Jones, JR Green. Ultrastructure and composition of the cell surfaces of infection structures formed by the fungal plant pathogen *Colletotrichum lindemuthianum*. *Journal of Microscopy* 181 (2): 204-212.

Roca, M. G., Davide, L. C., Mendes-Costa, M. C., & Wheals, A. (2003). Conidial anastomosis tubes in *Colletotrichum*. *Fungal Genetics and Biology*, 40(2), 138-145.

Rodriguez-Guerra, R., Ramírez-Rueda, M. T., Cabral-Enciso, M., García-Serrano, M., Lira-Maldonado, Z., Guevara-González, R. G., ... & Simpson, J. (2005). Heterothallic mating observed between Mexican isolates of *Glomerella lindemuthiana*. *Mycologia*, 97(4), 793-803.

Roohparvar, R., De Waard, M. A., Kema, G. H., & Zwiers, L. H. (2007). MgMfs1, a major facilitator superfamily transporter from the fungal wheat pathogen *Mycosphaerella graminicola*, is a strong protectant against natural toxic compounds and fungicides. *Fungal Genetics and Biology*, 44(5), 378-388.

ROSADA, L. J.; FRANCO, C. C. S.; SANT'ANNA, J. R.; KANESHIMA, E. N.; GONÇALVES - VIDIGAL, M. C.; CASTRO - PRADO, M. A. A. Parasexuality in Race 65 *Colletotrichum lindemuthianum* Isolates. *J. Eukaryot. Microbiol.*, 57(4), 2010 pp. 383-384.

Rosada, L. J., FRANCO, C. C., Sant'Anna, J. R., Kaneshima, E. N., GONÇALVES - VIDIGAL, M. C., & CASTRO - PRADO, M. A. (2010). Parasexuality in Race 65 *Colletotrichum lindemuthianum* isolates. *Journal of Eukaryotic Microbiology*, 57(4), 383-384.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.

Saier Jr, M. H., Tran, C. V., & Barabote, R. D. (2006). TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic acids research*, 34(suppl_1), D181-D186.

Sanz - Martín, J. M., Pacheco - Arjona, J. R., Bello - Rico, V., Vargas, W. A., Monod,

- M., Díaz - Mínguez, J. M., ... & Sukno, S. A. (2016). A highly conserved metalloprotease effector enhances virulence in the maize anthracnose fungus *Colletotrichum graminicola*. *Molecular plant pathology*, 17(7), 1048-1062.
- Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6), 863-864.
- Seo, J. A., Han, K. H., & Yu, J. H. (2004). The *gprA* and *gprB* genes encode putative G protein - coupled receptors required for self - fertilization in *Aspergillus nidulans*. *Molecular microbiology*, 53(6), 1611-1623.
- Sharma, P. N., Padder, B. A., Sharma, O. P., Pathania, A., & Sharma, P. (2007). Pathological and molecular diversity in *Colletotrichum lindemuthianum* (bean anthracnose) across Himachal Pradesh, a north-western Himalayan state of India. *Australasian Plant Pathology*, 36(2), 191-197.
- Shear, C. L., & Wood, A. K. (1913). Studies of fungus parasites belonging to the genus *Glomerella*. *Studies of fungus parasites belonging to the genus Glomerella.*, (252).
- Siewers, V., Viaud, M., Jimenez-Teja, D., Collado, I. G., Gronover, C. S., Pradier, J. M., ... & Tudzynski, P. (2005). Functional analysis of the cytochrome P450 monooxygenase gene *bcbot1* of *Botrytis cinerea* indicates that botrydial is a strain-specific virulence factor. *Molecular plant-microbe interactions*, 18(6), 602-612.
- Smit, A. F., & Hubley, R. (2008). RepeatModeler Open-1.0. Available from <http://www.repeatmasker.org>.
- Soares, M. A., Nogueira, G. B., Bazzolli, D. M. S., de Araújo, E. F., Langin, T., & de Queiroz, M. V. (2014). PacCl, a pH-responsive transcriptional regulator, is essential in the pathogenicity of *Colletotrichum lindemuthianum*, a causal agent of anthracnose in bean plants. *European journal of plant pathology*, 140(4), 769-785.
- Souza, E. A., Camargo Jr, O. A., & Pinto, J. M. (2010). Sexual recombination in *Colletotrichum lindemuthianum* occurs on a fine scale. *Genetics and Molecular Research*, 9(3), 1759-1769.
- Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic acids research*, 32(suppl_2), W309-W312.
- Sweigard, J. A., Carroll, A. M., Farrall, L., Chumley, F. G., & Valent, B. (1998). Magnaporthe grisea pathogenicity genes obtained through insertional mutagenesis. *Molecular plant-microbe interactions*, 11(5), 404-412.
- Tapre, A. R., & Jain, R. K. (2014). Pectinases: Enzymes for fruit processing industry. *International Food Research Journal*, 21(2).

- Tarailo - Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics*, 25(1), 4-10.
- Vaillancourt, L., Du, M., Wang, J., Rollins, J., & Hanau, R. (2000). Genetic analysis of cross fertility between two self-sterile strains of *Glomerella graminicola*. *Mycologia*, 92(3), 430-435.
- Vela-Corcía, D., Srivastava, D. A., Dafa-Berger, A., Rotem, N., Barda, O., & Levy, M. (2019). MFS transporter from *Botrytis cinerea* provides tolerance to glucosinolate-breakdown products and is required for pathogenicity. *Nature communications*, 10(1), 1-11.
- Wang, Q., Chen, D., Wu, M., Zhu, J., Jiang, C., Xu, J. R., & Liu, H. (2018). MFS transporters and GABA metabolism are involved in the self-defense against DON in *Fusarium graminearum*. *Frontiers in plant science*, 9, 438.
- Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., ... & Breitling, R. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids research*, 43(W1), W237-W243.
- Wheeler, H. E. (1954). Genetics and evolution of heterotallism in *Glomerella*. *Am J Bot*, 44, 342-345.
- Yang, J., Kong, L., Chen, X., Wang, D., Qi, L., Zhao, W., ... & Peng, Y. L. (2012). A carnitine–acylcarnitine carrier protein, MoCrc1, is essential for pathogenicity in *Magnaporthe oryzae*. *Current genetics*, 58(3), 139-148.
- Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., & Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic acids research*, 40(W1), W445-W451.
- Zhou, R., Zhou, X., Fan, A., Wang, Z., & Huang, B. (2018). Differential functions of two metalloproteases, Mrmep1 and Mrmep2, in growth, sporulation, cell wall integrity, and virulence in the filamentous fungus *Metarhizium robertsii*. *Frontiers in Microbiology*, 9, 1528.
- Zentmyer, G. A., Klure, L. J., & Pond, E. C. (1979). The influence of temperature and nutrition on formation of sexual structures by *Phytophthora cinnamomi*. *Mycologia*, 71(1), 55-67.
- Zhu P, Oudemans PV (2000) A long terminal repeat retrotransposon Cgret from the phytopathogenic fungus *Colletotrichum gloeosporioides* on cranberry. *Curr Genet* 38(5):241–247.

Tabela Suplementar 2 - Predição do conteúdo de DNA repetitivo no genoma de *C. lindemuthianum* 83.501.

* Elementos representando menos de 0,01% do genoma foram omitidos da tabela.

Classe	Número de cópias	Tamanho do elemento no genoma	Proporção do Genoma (%)
Total de Repetições no genoma	-	-	5,95
Retroelementos:	6851	2284585 pb	2.35
SINEs:	15	846 pb	0.00
LINEs:	933	67542 pb	0.07
L2/CR1/Rex	280	18131 pb	0.02
R1/LOA/Jockey	135	11807 pb	0.01
L1/CIN4	146	9696 pb	0.01
Elementos LTR:	5903	2216197 pb	2.28
BEL/Pao	76	5255 pb	0.01
Ty1/Copia	2103	685978 pb	0.70
Gypsy/DIRS1	3521	1512164 pb	1.55
Transposons de DNA:	1871	449708 pb	0.46
hobo-Activator	322	22048 pb	0.02
Tc1-IS630-Pogo	631	358589 pb	0.37
Tourist/Harbinger	102	6209 pb	0.01
Não Classificadas	245	25830 bp	0.03
Repetições Simples	53489	2654877 bp	2.73

* Elementos representando menos de 0,01% do genoma foram omitidos da tabela.

Tabela Suplementar 3 - Predição do conteúdo de DNA repetitivo no genoma de *C. lindemuthianum* A₂-2-3.

Classe	Número de cópias	Tamanho do elemento no genoma	Proporção do Genoma (%)
Total de Repetições no genoma	-	-	6,18
Retroelementos:	7188	2455674bp	2.48
LINEs:	956	68923 bp	0.07
L2/CR1/Rex	287	19351 bp	0.02
R1/LOA/Jockey	135	11609 bp	0.01
L1/CIN4	143	9671 bp	0.01
Elementos LTR:	6213	2385767 bp	2.41
BEL/Pao	72	5273 bp	0.01
Ty1/Copia	2222	738639 bp	0.74
Gypsy/DIRS1	3709	1628804 bp	1.64
Retroviral	134	8010 bp	0.01
Transposons de DNA:	1998	494793 bp	0.50
hobo-Activator	315	21666 bp	0.02
Tc1-IS630-Pogo	662	395979 bp	0.40
PiggyBac	15	876 bp	0.00
Tourist/Harbinger	109	6499 bp	0.01
Não Classificadas	233	25376 bp	0.03
Repetições Simples	54987	2729125 bp	2.75

* Elementos representando menos de 0,01% do genoma foram omitidos da tabela.