



UNIVERSIDADE FEDERAL RURAL DE
PERNAMBUCO

*P*RÓ-*R*EITORIA DE *P*ESQUISA E *P*ÓS-*G*RADUAÇÃO



PROGRAMA DE PÓS-GRADUAÇÃO
EM FITOPATOLOGIA

Tese de Doutorado

**Montagem *de novo*, predição e anotação do genoma de
Colletotrichum musae e genômica comparativa do gênero**

Wilson José da Silva Junior

RECIFE – PE

2018

WILSON JOSÉ DA SILVA JUNIOR

Montagem *de novo*, predição e anotação e anodo genoma do *Colletotrichum musae* e genômica comparativa do gênero

Tese apresentada ao Programa de Pós-Graduação em Fitopatologia da Universidade Federal Rural de Pernambuco, como parte dos requisitos para obtenção do título de Doutor em Fitopatologia.

COMITÊ DE ORIENTAÇÃO:

Orientador: Professor Marcos Paz Saraiva Câmara, Ph.D.

Co-Orientador: Professor Doutor Valdir de Queiroz Balbino

Co-Orientadora: Dra. Waléria Guerreiro de Lima

RECIFE-PE
FEVEREIRO – 2018

**MONTAGEM *DE NOVO*, PREDIÇÃO E ANOTAÇÃO DO GENOMA DE
COLLETOTRICHUM MUSAE E GENÔMICA COMPARATIVA DO GÊNERO.**

WILSON JOSÉ DA SILVA JUNIOR

Tese defendida e aprovada pela Banca Examinadora em: 28/02/2018

ORIENTADOR(A):

Professor Marcos Paz Saraiva Câmara, Ph.D. (UFRPE)

EXAMINADORES:

Professor Doutor Marco Aurélio Siqueira da Gama (UFRPE)

Professora Doutora Elineide Barbosa de Souza (UFRPE)

Professora Doutora Ana Maria Benko-Iseppon (UFPE)

Professor Doutor Sérgio de Sá Leitão Paiva Júnior (UAST)

RECFE-PE

FEVEREIRO – 2018

"Faça o teu melhor, na condição que você tem, enquanto você não tem condições melhores, para fazer melhor ainda!"

Mário Sergio Cortella

AGRADECIMENTOS

À Universidade Federal Rural de Pernambuco e ao programa de Pós-graduação em Fitopatologia e aos professores: Gilvan Pio, Sônia Maria, Elineide Barbosa, Marco Gama, Rosa Mariano, Delson Laranjeira, Sami Michereff, Marcos Câmara, Elvira Pedrosa e Ailton Reis, pelos ensinamentos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos.

Aos meus orientadores Marcos Câmara e Valdir Balbino, pelos ensinamentos, orientação e pela amizade.

À Universidade Federal de Pernambuco e ao Programa de Pós-graduação em Genética e aos professores pelos conhecimentos transmitidos.

À rede INTERSYS e a Professora Ana Benko-Iseppon pela confiança e oportunidade.

À Universidade Federal do Rio Grande do Sul e ao programa de Pós-Graduação em Biologia Celular e Molecular e aos professores Fábio Carrer, Charley Staats, Augusto Schrank e Nicolau Sbaraini pelos ensinamentos.

Ao laboratório de Bioinformática e Biologia Evolutiva (LABBE) e a todos os colegas, pelo convívio e respeito.

Aos companheiros do grupo de Bioinformática: Lucas Christian, Raul Falcão, Dr. Sérgio Paiva e Roberto Farias pela amizade, respeito e companheirismo.

Aos amigos e colegas do programa de Pós-graduação em Fitopatologia pela amizade, respeito.

Aos companheiros do laboratório de Micologia: Waléria Guerreiro, Willie Anderson, Josiene Veloso e Geresa Rodrigues pela amizade e respeito.

À Ana Patrícia Gonçalves e aos Doutores Domingos Andrade e Cristina Assis que me iniciaram na fitopatologia

Aos meus pais Wilson José da Silva e Anadeje dos Santos Silva e ao meu irmão Williams Daniel dos Santos Silva, pelo carinho, amor, educação e paciência.

À minha esposa Rute Torres Portella e aos meus filhos Ian Francisco e Ivy Sofie, pelo carinho, amor, dedicação.

SUMÁRIO

AGRADECIMENTOS	v
RESUMO GERAL	vii
GENERAL ABSTRACT.....	viii
Montagem de novo, predição e anotação e anodo genoma do <i>Colletotrichum musae</i> e genômica comparativa do gênero.....	1
1. INTRODUÇÃO GERAL	1
1.1. A Cultura da banana	1
1.2. Importância econômica e distribuição geográfica	2
1.3. A antracnose da banana	4
1.4. Gênero <i>Colletotrichum</i>	5
1.5. Estilo de vida de <i>Colletotrichum</i>	7
1.6. Genômica de <i>Colletotrichum</i>	9
1.7. Genômica comparativa do gênero <i>Colletotrichum</i>	10
1.8. Famílias gênicas.....	11
1.9. Genes de metabolismo secundário.....	12
2. REFERÊNCIAS BIBLIOGRÁFICAS	15
CAPÍTULO II.....	21
Draft genome assembly of <i>Colletotrichum musae</i> , the pathogen of banana fruit	21
Abstract.....	22
Specifications Table	23
Values of the Data	23
1. Data	23
2. Experimental design, materials, and methods.....	24
2.1. DNA extraction and genome sequence	24
2.2. <i>De novo</i> assembly and genes annotation.....	24
Acknowledgments	25
References	25
CAPÍTULO III	27
Secondary metabolite biosynthetic gene clusters in the banana-specialist pathogen <i>Colletotrichum musae</i>	27
ABSTRACT	29
Prediction of secondary metabolite gene clusters.....	31
Phylogenetic analyses.....	32
RESULTS AND DISCUSSION.....	33
CONCLUSION	35
FIGURES, TABLES AND SUPPLEMENTARY DATA	38
CAPÍTULO IV	60
Conclusões Gerais	60
Conclusões gerais	61

RESUMO GERAL

Colletotrichum musae o agente etiológico da antracnose da banana é um importante patógeno por causar prejuízos no campo e principalmente na pós-colheita, afetando diretamente a comercialização do fruto. Ainda não é possível inferir sobre a interação planta-patógeno e quais fatores genéticos estariam ligados ao aparecimento de sintomas nos frutos a partir de um determinado sinalizador. O entendimento do ciclo da relação patógeno-hospedeiro, bem como dos fatores relevantes para o desenvolvimento da doença, são de fundamental importância para o manejo. Pouco se sabe sobre a genômica de *C. musae*, do patossistema ou do reflexo direto do arsenal gênico e quais aspectos moleculares podem ser responsáveis pelo surgimento da doença. O objetivo do presente trabalho foi sequenciar, montar, anotar o genoma, além de identificar famílias de genes, associadas aos fatores de patogenicidade de *C. musae* isolado CMM4420. A melhor montagem obtida pelo SPAdes, resultou em 50.635.197 pb com um conteúdo G + C de 53.74%, culminado em 2.763 *scaffolds*, abrigando 13.451 genes com um comprimento médio de 1.626 pb. A predição e a anotação de genes foram realizadas pelo *pipeline* do Funannotate, usando um padrão para identificação de genes com base no BUSCO. Notavelmente, 97% dos grupos de genes biossintéticos dos metabolitos secundários (GGBs) identificados de *C. musae* foram conservados em *C. gloeosporioides*, um patógeno generalista. Enquanto alguns *C. Gloeosporioides* GGBs estão ausentes no genoma de *C. Musae*. Estes resultados reforçam a importância dos metabolitos secundários na infecção por *C. musae*. Além disso, as análises detalhadas realizadas abrem o caminho para futuras pesquisas aprofundadas.

Palavras-chaves: Metabólitos secundários, estilo de vida, Banana e Antracnose.

GENERAL ABSTRACT

Colletotrichum musae, the etiological agent of banana anthracnose is an important pathogen due to its field and especially post-harvest damage, directly affecting the commercialization of the fruit. It is not yet possible to infer about a pathogenic plant interaction and what genetic factors would be related to the appearance of symptoms in the fruits from a particular molecular flag. Understanding the pathogen-host relationship cycle, as well as factors relevant to the development of the disease, are of fundamental importance for its management. Little is known about the genomics of *C. musae*, the pathosystem or the direct reflex of the gene pool and which can be used by the surgeon of the disease. The objective of this work was to sequence, assemble and annotate the genome, in addition to identify families which are associated to pathogenicity factors of *C. musae* genes isolated CMM4420. The best assembly obtained by SPAdes, resulted in 50,635,197 bp with a G + C content of 53.74%, culminating in 2,763 scaffolds, housing 13,451 genes with an average length of 1,626 bp. A prediction and annotation of genes were performed by the Funannotate pipeline, using a standard for identification of genes based on the BUSCO. Notably, 97% of the biosynthetic gene groups of the identified secondary metabolites (BGCs) of *C. musae* have been conserved in *C. gloeosporioides*, a generalist pathogen. While some *C. gloeosporioides* BGCs are absent in the genome of *C. musae*. These results reinforce an important of the secondary metabolites in *C. musae* infection. In addition, the detailed analyzes will abbreviate the way for further in-depth research.

Keywords: Secondary metabolites, lifestyle, banana and anthracnose.

CAPÍTULO I

Introdução Geral

1. INTRODUÇÃO GERAL

1.1. A Cultura da banana

Bananeira (*Musa* spp.) é a denominação genérica para diversas espécies do gênero *Musa*, dentro da família das Musáceas. As bananeiras produtoras de frutos comestíveis pertencem à classe das Monocotiledôneas, ordem Scitaminales, família Musacea, subfamília Musoidae e gênero *Musa*. No cenário evolutivo participaram as espécies diploides selvagens *M. acuminata* Colla e *M. balbisiana* Colla, ambas com 11 cromossomos ($n=11$). A espécie *M. acuminata* (Genoma A) possui ploidia AA, AAA e AAAA, já a *M. balbisiana* (Genoma B) pode ser, BB, BBB e BBBB, mas ambas as espécies podem cruzar entre si, seja naturalmente ou artificialmente, produzindo os híbridos: AB, AAB, ABBB, AABB e AAAB (DIAS, 2011).

As cultivares possuem três níveis cromossômicos distintos: diploide, triploide e tetraploide, respectivamente com dois, três e quatro múltiplos do número básico de cromossomos, 22, 33 e 44 respectivamente. Na Ásia onde predominam formas selvagens de *M. balbisiana*, houve um estágio evolucionário paralelo, onde surgiram cultivares diploides e triploides puramente *balbisiana*, os quais foram primeiramente reconhecidos nas Filipinas (ARVANITOYANNIS; MAVROMATIS, 2009). A natural distribuição de *M. acuminata* e *M. balbisiana* na região, associada a autocompatibilidade entre as duas espécies, levou ao surgimento das bananeiras AB, AAB e ABB. Os tetraploides dos grupos AAAA, AAAB, AABB e ABBB evoluíram de forma parecida (D'HONT et al., 2000)

A banana é um dos frutos mais consumidos no mundo, destacando-se como a principal fruteira tropical. É cultivada em mais de 120 países localizados nas regiões tropicais e subtropicais do mundo (SOUZA et al., 2000). Ao longo dos anos a produção de bananas teve uma base mais científica, a exploração botânica e comercial permitiu a distribuição de algumas variedades, a banana AAA Cavendish foi obtida primeiro na China em 1826, essas bananas foram originalmente multiplicadas em casas de vegetação na Europa e depois redistribuídas para jardins botânicos, estações de pesquisa e empreendimentos comerciais interessados. Esta variedade agora está disseminada em todo o mundo, não só como a principal banana do comércio de exportação, mas também rapidamente assumindo um papel importante na agricultura de algumas das regiões às quais foi introduzida (JEGER et al., 1995).

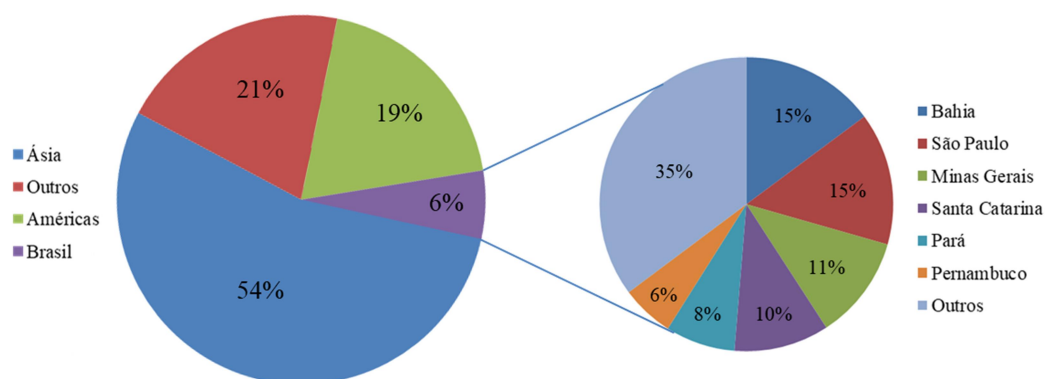
1.2. Importância econômica e distribuição geográfica

Embora existam muitas variedades de banana são poucas as cultivares com potencial agrônomo para serem usadas comercialmente levando em conta aspectos como preferência dos consumidores, produtividade, tolerância a pragas e doenças, resistência à seca, porte e resistência ao frio. As cultivares mais difundidas no Brasil são: Prata, Pacovan, Prata Anã, Maçã, Mysore, Terra e D'Angola, do grupo AAB, utilizadas unicamente para o mercado interno, e Nanica, Nanicão e Grande Naine, do grupo AAA, usadas principalmente no mercado para exportação. Em menor escala são plantadas 'Ouro' (AA), 'Figo Cinza' e 'Figo Vermelho' (ABB), 'Caru Verde' e 'Caru Roxa' (AAA). As cultivares Prata, Prata Anã e Pacovan são responsáveis por aproximadamente 60% da área cultivada com banana no Brasil (SILVA et al., 2001).

Em termos de produção é a quarta maior do mundo totalizando 5.494.005 hectare (ha) de área plantada e produção de 113.280.302 toneladas (t) equivalente a um rendimento de 20,62 t/ha. A Ásia (61.584.024 t) e a América (28.626.955 t) são os continentes que se destacam na produção desta cultura, sendo a Índia (29.124.000 t), China (13.066.778 t), Indonésia (7.007.125 t) e Brasil (6.764.324 t) os maiores produtores mundiais. No ano de 2016, o Brasil se estabeleceu como quarto maior produtor mundial com uma área colhida de 469.711 ha e uma produtividade de 14,4 t/ha, posicionando assim a bananeira como a segunda frutícola mais cultivada no país, atrás apenas da laranjeira (FAO, 2018).

Apesar de não ser autóctone a produção de banana ocorre em todos os estados da Brasil. O estado da Bahia (1.004.000 t) lidera o ranking de produção seguido por São Paulo (985.332 t), Minas Gerais (772.845 t), Santa Catarina (708.000 t) e Pará (519.348 t). Pernambuco ocupa a sexta posição com uma produção anual de 393.215 t e rendimento de 10,324 t/ha (IBGE, 2017). A região Nordeste do Brasil, detentora de polos de fruticultura irrigada, é a maior produtora de banana com 197.295 ha plantados e produção de 2.228.998 t, produzindo mais que as regiões norte, sul e centro-oeste somadas. Nesta região destacam-se os estados da Bahia (1.004.000 t), Pernambuco (393.215 t), Ceará (324.101 t), e Rio Grande do Norte (148.664 t) (IBGE, 2017) (Figura 1).

Figura 1. Produção Mundial e Nacional de banana.



Fonte: Elaborado a partir de dados disponibilizados pela FAO (2018) e IBGE (2018).

A região semi-árida do polo Petrolina-Juazeiro (Pernambuco/Bahia) apresenta excelentes condições para a produção de banana de alto padrão de qualidade, o bom desenvolvimento da cultura é favorecido pelas ótimas condições climáticas para a prática da agricultura irrigada e a alta luminosidade e temperatura. O polo conta com vários projetos de irrigação como Nilo Coelho, Mandacaru, Salitre, Tourão, Curaçá e Pontal que mantêm elevada a produção de frutas, destacando-se em nível nacional (LUCIANA; BORGES, 2009).

A banana consumida *in natura* é parte integrante da alimentação dado seu valor nutritivo e o baixo custo. Cerca de 90% produção mundial de banana é oriunda de pequenos agricultores, sendo este fator de grande importância socioeconômica pela geração de empregos e fixação do homem no campo (ALVES, 1999).

As bananeiras utilizadas para produção de frutos são de maneira geral partenocárpicas e estéreis, características que dificultam o melhoramento genético. As progênies obtidas a partir de cruzamentos entre genótipos apresentam tamanho reduzido e são compostas por um número variável de ploidias, inclusive aneuploides. Fato esse que dificulta o melhoramento genético impossibilitando o desenvolvimento de novas cultivares (SILVA et al., 2013).

A bananicultura no Brasil em geral apresenta um déficit de manejo propiciando o surgimento de problemas fitossanitários causados por fungos, bactérias, vírus, nematoides e insetos. Doenças como o moko, sigatoka-negra, sigatoka-amarela, fusariose e antracnose são fatores limitantes da cultura culminado na baixa produtividade e qualidade dos frutos (KIMATI, H.; AMORIM, BERGAMIN FILHO, A.; REZENDE, 1997). Dentre os organismos citados fungos do gênero *Colletotrichum*, agente etiológico da antracnose, destaca-se por causar perdas na pós-colheita, comprometendo cerca de 40% da produção (ROGÉRIO et al., 2007)

1.3. A antracnose da banana

A antracnose, causada por *Colletotrichum musae* (Berk & Curt.) Arx. ocorre em quase toda a banana cultivada no mundo. Considerada principalmente como uma doença de pós-colheita, as infecções dos frutos imaturos podem ocorrer ainda no campo, como já relatado na Austrália, Índia, Fiji e Filipinas (CORDEIRO, 2000).

Colletotrichum musae foi descrito por Berkeley (1874) inicialmente como *Myxosporium musae* Berk. & M.A. Curtis, logo depois foi nomeado *Colletotrichum* por von arx (1957). Sutton (1992) determinou ser uma espécie distinta e forneceu uma breve descrição morfológica. Mais tarde, Hyde et al. (2009) apontaram que a epidifização é necessária para esclarecer suas relações com outras espécies. Tradicionalmente, a identificação de *C. musae* baseava-se em caracteres morfológicos, como a esporulação abundante, conídios retos e cilíndricos e apressório de forma irregular (SUTTON, 1980). Esses caracteres morfológicos, no entanto, são muitas vezes sobrepostos e ambíguos entre o complexo de espécies *gloeosporioides*. Assim, análises moleculares são necessárias para um diagnóstico preciso (HYDE et al., 2009).

Colletotrichum musae parece ser hospedeiro específico para as espécies de *Musa*, no entanto, Mahadtanapuk et al. (2007) relataram este táxon como um agente de antracnose em flores de *Curcuma alismatifolia* Gagnep. na Tailândia. Porém a identificação baseou-se apenas nos caracteres morfológicos havendo assim a necessidade de uma identificação molecular (SU et al., 2011).

O agente etiológico infecta os frutos ainda verdes e permanece em estado quiescente até a maturação, nesse estágio a infecção quiescente inicia-se concomitantemente ao período de transporte acarretando infecções secundárias. Sintomas característicos da doença como manchas na casca (cor marrom a preta) e lesões deprimidas, aparecem no período de amadurecimento dos frutos. Além disso, sob alta umidade pode ocorrer à formação de acérvulos (frutificações mucilaginosas) de cor salmão que podem ser observadas na casca do fruto (RANASINGHE; JAYAWARDENA; ABEYWICKRAMA, 2003). Essas manchas coalescem tornando-se maiores, nessas lesões também é possível observar conídios do fungo, que apresentam coloração laranja. A poupa do fruto somente é atingida já no estágio final da maturação (PLOETZ; THOMAS; SLABAUGH, 2003). Além dos frutos, há relatos do fungo causando podridão na flor, podridão da coroa e podridão da ponta da banana (INTAN SAKINAH; SUZIANI; LATIFFAH, 2014). A infecção, portanto, representa uma redução na viabilidade da fruta durante os períodos de maturação, transporte e armazenamento

(SLABAUGH; GROVE, 1982), o que leva a uma depreciação comercial e redução da vida útil da fruta (SILVA JUNIOR et al., 2018). Dentre os vários fatores responsáveis por perdas pós-colheita em banana, a antracnose, merece destaque, pois afeta tanto mercado local, como a exportação (JEGER M.J., EDEN-GREEN S., THRESH J.M., JOHANSON A., WALLER J.M., 1995).

O diagnóstico da doença e identificação do patógeno por métodos convencionais envolvem o isolamento e caracterização do agente patogênico por meio de testes de inoculação os quais são demorados e intensivos. Ao longo das últimas décadas, métodos imunológicos e moleculares têm recebido cada vez mais atenção como uma alternativa ou para complementar aos métodos convencionais (RIGOTTI et al., 2002). A detecção de patógenos fúngicos por vias moleculares tem sido utilizada com eficácia (SU et al., 2011). A identificação precisa e rápida de patógenos é necessária para o manejo adequado das doenças de plantas. Em particular, a caracterização genética e fenotípica dos fitopatógenos prevalentes em determinada área é requerida para um manejo eficiente (TSEDALEY, 2015).

As perdas em pós-colheita tem sido evitadas há tempos, porém, apesar de todos os esforços um método eficiente ainda não foi obtido. Embora o desenvolvimento de fungicidas modernos e a melhora em tecnologias de armazenamento tenham estendido amplamente a vida de prateleira de frutos, as perdas ainda são significativas.

1.4. Gênero *Colletotrichum*

Colletotrichum (Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Glomerellales; Glomerellaceae;) (RÉBLOVÁ; GAMS; SEIFERT, 2011; ZHANG et al., 2006) ocupa um lugar digno de nota na história da fitopatologia e micologia. Por sua ampla distribuição, capacidade substancial para destruição e importância científica como patossistemas modelo, espécies do gênero *Colletotrichum* estão ranqueadas entre os 10 principais fitopatógenos fúngicos mais importantes pela Comunidade Internacional de Fitopatologia (DEAN et al., 2012).

O gênero engloba mais de 100 espécies responsáveis por causar antracnoses e manchas foliares em quase todas as espécies cultivadas para alimentação, fibras e forragem em todo o mundo (CANNON et al., 2012; HYDE et al., 2009). Doenças economicamente importantes causadas por *Colletotrichum* são generalizadas, ocorrendo no campo em milho, feijão, morango, café, pimentas, pimentão, batata e inúmeras outras plantas cultivadas (CANNON et al., 2012) na pós-colheita produzem danos em culturas importantes de subsistência, incluindo, lentilha, feijão-caupi, inhame, banana, sorgo e mandioca

(ADEGBITE; AMUSA, 2008; CHONGO; GOSSEN; BERNIER, 2002; GREEN; SIMONS, 1994; NASH; STRANGE, 1996) resultando em grandes perdas econômicas, com graves infecções, sendo responsável por 100% das perdas durante o armazenamento (PRUSKY, 1996).

A adoção da nomenclatura de nome único para fungos pleomórficos estabelecida pelo Código 2013 do *the International Code of Nomenclature for algae, fungi, and plants* (<http://www.iapt-taxon.org/nomen/main.php>) sugere que o nome de *Glomerella* gradativamente entrará em desuso. Embora sejam conhecidas várias espécies no gênero que produzem o teleomorfo (*G. cingulata* e *G. acutata*), as espécies de *Colletotrichum* são predominantemente observadas no estado vegetativo ou assexuado, com a morfologia sexual raramente identificada para a maioria das espécies (THON; NUCKLES; VAILLANCOURT, 2000). Fitopatologistas e micologistas que trabalham com espécies do fungo normalmente encontram o anamorfo, sendo assim o nome *Colletotrichum* representa com precisão informações biológicas sobre o organismo. Além disso, *Colletotrichum* é o mais antigo dos dois gêneros e tem prioridade (1831 vs. 1903; www.mycobank.org). Desta forma deve-se referir a *Colletotrichum* tanto nas fases anamórficas quanto à teleomórfica.

Colletotrichum é o único membro da Glomerellaceae, uma das três famílias que compõem a ordem Glomerellales em Sordariomycete subclasse Hypocreomycetidae (RÉBLOVÁ; GAMS; SEIFERT, 2011). Estudos sugeriram *Colletotrichum* como um grupo irmão para *Verticillium* (ZHANG et al., 2006), porém as bases que sustentavam essa hipótese eram insuficientes e não representavam uma associação filogenética segura, pois *Verticillium* é membro da Plectosphaerellaceae (CANNON et al., 2012; RÉBLOVÁ; GAMS; SEIFERT, 2011).

A taxonomia do gênero tem sido alvo de várias revisões e o conceito de espécie, ainda que bem definido está em um estado transitório, porém foi estabelecido que o gênero é composto por vários grandes clados monofiléticos que são referidos como complexo de espécies, descritos pelo nome e os atributos das espécies representativas mais proeminentes no grupo (CANNON et al., 2012; O'CONNELL et al., 2012).

Os nove clados, nomeadamente *acutatum*, *graminicola*, *spaethianum*, *destructivum*, *dematium*, *gloeosporioides*, *boninense*, *truncatum* e *orbiculare* estão descritos com base molecular e filogenética (CANNON et al., 2012). O clado *gloeosporioides*, por exemplo, consiste em pelo menos 22 espécies tradicionalmente conhecidas como *C. gloeosporioides*, incluindo *C. gloeosporioides stricto sensu* (WEIR; JOHNSTON; DAMM, 2012). Pela perspectiva taxonômica de base molecular, *C. gloeosporioides stricto sensu* agora é conhecido

por ser menos comum no ambiente do que se pensava anteriormente (PHOULIVONG et al., 2010; WEIR; JOHNSTON; DAMM, 2012). A terminologia agregada é especialmente útil para diagnóstico de doenças que ainda dependem de fragmentos de DNA para identificar agentes causais. Vários dos grupos agregados foram amplamente caracterizados através de filogenias de várias regiões do genoma (HYDE et al., 2009; WEIR; JOHNSTON; DAMM, 2012), produzindo uma estrutura para a compreensão de relações evolutivas em todo o gênero (CANNON et al., 2012).

1.5. Estilo de vida de *Colletotrichum*

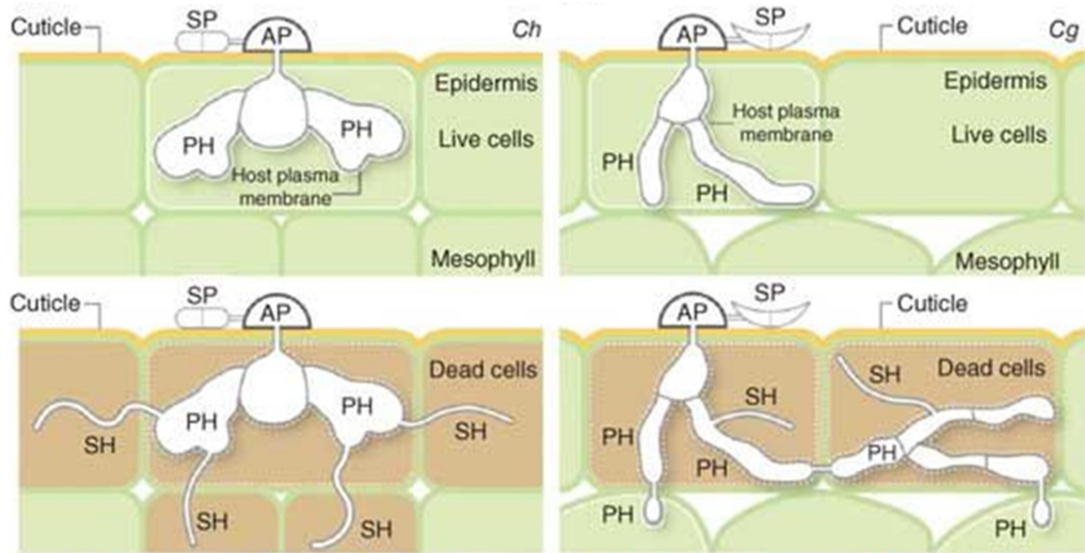
Grande parte dos fungos fitopatogênicos possuem uma variedade de interações com seus hospedeiros, alterando sua relação em diferentes estágios do seu ciclo de vida, dependendo da maturidade fisiológica, da resistência do hospedeiro, do meio ambiente e dos genes associados à virulência do patógeno (DELAYE; GARCÍA-GUZMÁN; HEIL, 2013; STERGIPOULOS; GORDON, 2014)

Espécies do gênero *Colletotrichum* dispõem de uma gama de estratégias nutricionais e estilos de vida, inclusive associações com plantas podendo ser necrotrófica, hemibiotrófica e endofítica. Porém algumas espécies empregam um estilo de vida saprofítico, retirando os nutrientes necessários do solo e da matéria orgânica (DEAN; LICHENS-PARK; KOLE, 2014). Espécies fitopatogênicas usam apressório melanizado para penetrar no tecido do hospedeiro (KUBO; TAKANO, 2013), sendo a melanina requerida para garantir o funcionamento do apressório, permitindo um acúmulo significativo de pressão de turgor na parede celular do hospedeiro (BECHINGER et al., 2016; KUBO, Y; FURUSAWA, 1991). O apressório também secreta pectinases e enzimas que degradam a parede celular do hospedeiro, que são passíveis de atuar em diversas etapas da infecção, adesão, sinalização e suavização da parede celular do hospedeiro (KLEEMANN et al., 2008; MENDGEN; HAHN; DEISING, 1996).

Fungos do gênero *Colletotrichum* possuem a habilidade de penetrar no tecido do hospedeiro de forma direta (pela utilização de mecanismo próprio) ou indiretamente pelas aberturas naturais e lesões decorrentes de terceiros. A relação patógeno-hospedeiro é caracterizada por uma fase biotrófica, a qual o patógeno sobrevive em contato com o hospedeiro, passando à fase necrotrófica, a qual obtém nutrientes. O período de uma fase para outra pode variar de espécie para espécie e em função do patossistema. Essa interação se dá pela germinação dos esporos, em seguida são formadas as estruturas diferenciadas, como o apressório, estrutura requerida para a penetração direta. Após a penetração, o patógeno pode

permanecer quiescente ou desenvolver-se intra e/ou intercelularmente, apresentando lesões com subsequente formação de hifas secundárias (AGRIOS, 2005; DEAN et al., 2012)

Figura 2. À esquerda: Processo de infecção de *C. higginsianum* (Ch), antracnose foliar em Brássicas e Arabidopsis e à direita: Processo de infecção de *C. graminicola* (Cg), e da queima das folhas, principais sintomas em milho. SP, esporo; AP, apressório; PH, hifas primárias biotróficas; SH, hifas secundárias necrotróficas.



Fonte: (O'CONNELL et al., 2012).

Na hemibiotrofia intracelular, o estágio inicial da infecção é a fase biotrófica assintomática na qual o patógeno invade as células epidérmicas e as hifas primárias produzem vesículas de infecção dentro de células epidérmicas e mesófilas. Nesta fase, as células hospedeiras permanecem vivas e as plantas não apresentam sintoma. Em seguida inicia-se a fase necrotrófica em que as hifas secundárias crescem intracelular e intercelularmente enquanto secretam enzimas degradantes da parede celular que matam as células hospedeiras (O'CONNELL et al., 2012) (Figura 2). Na infecção subcutânea, necrotrófica, o fungo cresce sob a cutícula dentro das paredes das células epidérmicas sem penetrar nos protoplastos. Mesmo que isso seja chamado de infecção subcutânea e intramural, pode haver também uma fase biotrófica muito curta antes das hifas intramurais iniciarem agressivamente a destruição dos tecidos hospedeiros colonizados (ARROYO et al., 2005; PERES et al., 2005; WHARTON, 2004).

O estágio necrotrófico também pode apresentar variações no tempo quanto ao surgimento de sintomas, por exemplo, não necessariamente as hifas secundárias estão diretamente relacionadas ao surgimento de sintomas sendo essa condição determinada pela espécie do patógeno e do hospedeiro (O'CONNELL et al., 2012).

Os padrões de estilo de vida das espécies de *Colletotrichum* são altamente regulados por famílias de genes específicas e interações bioquímicas que ocorrem através de enzimas específicas e metabólitos secundários produzidos na relação patógeno-hospedeiro. As análises

comparativas dos genomas em geral mostraram que as espécies de *Colletotrichum* possuem perfis adaptados de algumas enzimas de acordo com seu estilo de vida específico (GAN et al., 2016). As pesquisas de genômica e transcriptômica na última década forneceu informações valiosas sobre a base genética dos estilos de vida variáveis de fungos fitopatogênicos e a importância do arsenal gênico utilizados por esses agentes (CROUS et al., 2016; MCTAGGART et al., 2016; ZEILINGER et al., 2016).

1.6. Genômica de *Colletotrichum*

A análise em escala genômica de linhagens de *Colletotrichum* com estilos de vida contrastantes pode ajudar na identificação de pontos em comum e pode fornecer subsídios para o controle e manejo desses fungos. Na literatura especializada, há poucos relatos de espécies de *Colletotrichum* estudadas a nível genômico, e relativamente pouco se sabe sobre as espécies de maior importância econômica. Achados experimentais em um patossistema raramente são validados ou testados em outros sistemas e por isso não está claro se os mecanismos de patogenicidade são semelhantes em todas as linhagens (DEAN; LICHENS-PARK; KOLE, 2014). Genomas de outros fungos fitopatogênicos de importância econômica foram sequenciados, começando com *Magnaporthe oryzae* em 2005 (DEAN, 2005). Contudo, *Colletotrichum* permaneceu no fundo da revolução genômica, principalmente porque foi difícil escolher o sequenciamento de uma única espécie para representar todos os outros, o que torna o panorama para este gênero pouco elucidado em relação a estudos de patogenicidade. A taxonomia problemática, aliada a uma compreensão inadequada da relação evolutiva entre as espécies, também limitou inicialmente a capacidade de identificar os pontos mais adequados para a análise do genoma. O desenvolvimento de sequenciadores de nova geração (NGS – *Next-generation sequence*) possibilitou o sequenciamento de genomas completos a baixo custo e em pequena escala temporal, possibilitando a geração de dados e a publicação dos genomas de quatro espécies de *Colletotrichum*: *C. higginsianum*, *C. graminicola*, *C. orbiculare*, e *C. fructicola* (anteriormente *C. gloeosporioides*) (GAN et al., 2013; O'CONNELL et al., 2012).

De acordo com a compreensão atual da taxonomia de *Colletotrichum*, essas quatro espécies são posicionadas cada uma dentro de distintas linhagens monofiléticas deste gênero (CANNON et al., 2012). *C. graminicola* é membro do complexo graminicola. *C. higginsianum* faz parte do complexo destructivum, o clado irmão ao complexo graminicola. *C. orbiculare* é um parente próximo do agente causal da antracnose do feijão, *C. lindemuthianum*, os quais formam o complexo de espécies orbiculare. *C. fructicola* é um

membro do complexo *C. gloeosporioides*, um dos mais diversos grupos de *Colletotrichum*, abrangendo numerosas espécies associadas a um número enorme dos hospedeiros em todo o mundo (WEIR; JOHNSTON; DAMM, 2012).

A disponibilidade de sequências genômicas completas de fungos fitopatogênicos possibilitou uma revolução no campo da fitopatologia molecular. Nos últimos anos tem ocorrido um avanço significativo na aplicação da genômica comparativa para estudos da função estrutural e evolução de genomas de fungos. O valor da genômica comparativa de fungos fitopatogênicos já foi demonstrado pela identificação de genes relacionados à virulência e funções específicas de hospedeiros (O'CONNELL et al., 2012).

As comparações entre membros de diferentes espécies, e dentro de agregados, podem ser informativo na determinação de fatores que contribuem para adaptação a determinados hospedeiros e estilos de vida incluindo desenvolvimento do fungo.

1.7. Genômica comparativa do gênero *Colletotrichum*

As diferenças nos comprimentos de *reads* gerados pelos diferentes métodos de sequenciamento são refletidas na qualidade dos genomas resultantes. A utilização de tecnologias com *reads* curtos produzem montagens mais fragmentadas do que aquelas que incorporaram dados gerados por Sanger e Roche. Apesar das diferenças nas abordagens de sequenciamento, a cobertura geral de genes se mantém alta para indivíduos do gênero (GAN et al., 2013; O'CONNELL et al., 2012) quando avaliada usando o *pipeline* de CEGMA (*Core Eukaryotic Genes Mapping Approach*, Parra et al., 2007). No entanto, as comparações diretas entre os genomas do gênero *Colletotrichum* devem ser feitas com cautela, uma vez que são empregadas diferentes estratégias e metodologias de sequenciamento. Adicionalmente são utilizadas diferentes ferramentas computacionais para reunir e anotar os genomas, bem como analisar famílias gênicas e metabólitos secundários. (Tabela 1).

Tabela 1. Estatísticas de montagens de genomas de quatro espécies de *Colletotrichum*.

	<i>C. higginsianum</i>	<i>C. graminicola</i>	<i>C. fruticola</i>	<i>C. orbiculare</i>
Assembly size (Mb)	53.4	57.4	55.6	88.3
Coverage	101x	9.1x	37x	55x
Sequence technology	Roche 454 (25x) Illumina (76x) Sanger (0.2x)	Sanger (7.9x) Roche 454 (1.2x)	Illumina (37x)	Roche 454 (22x) Illumina (34x)
Number of scaffolds	653	367	1,241	525
N50 contig length	265.5	579.2	112.8	428.9
Number of contigs	10,269	1,151	53,335	10,545
Gene space coverage	95.1 %	99.2 %	96.4 %	98%
Number of predicted genes	16,172	12,006	15,469	13,479

Overall GC content	55.1 %	49.1 %	53.6 %	37.5 %
GC content of genes	58.4 %	59.3 %	56%	57.1 %
Number of chromossomes	10 major, 2 "B" chromossomes	10 major, 3 "B" chromossomes	Unknown	10 major, no "B" chromossomes
Multicopy genes	9,713	6,468	14,933	7,475
Conserved single copy genes	4,725	4,767	372	4,553
Repeat elements	1.2 %	12.2 %	0.75 %	8.3 %
Public access to genome	CACQ2000000	ACOD01000001	PRJNA171218	PRJNA171217

Fonte: (GAN et al., 2013; O'CONNELL et al., 2012)

1.8. Famílias gênicas

Família gênica é um grupo de genes que compartilha características importantes. Em muitos casos, os genes de uma família compartilham uma sequência semelhante de DNA. Esses genes fornecem instruções para produzir produtos (como proteínas) que têm uma estrutura ou função similar. Em outros casos, os genes dissimilares são agrupados em uma família porque as proteínas produzidas a partir desses genes trabalham juntas como uma unidade ou participam do mesmo processo. (OSTELL; MCENTYRE, 2007).

Algumas famílias gênicas são expandidas nos genomas do gênero *Colletotrichum*, em relação a outros ascomicetos. Essas expansões incluem genes que são preditos para codificar *carbohydrate-activeenzymes* (CAZymes), enzimas de metabólitos secundários (MS), proteases, e efetores putativos (GAN et al., 2013; O'CONNELL et al., 2012).

CAZymes são arsenais de enzimas expandidas capazes de degradar celulose e outros polissacarídeos que existem na parede celular das plantas, cenário comum para patógenos hemibiotróficos e necrotróficos, que incluem espécies de *Colletotrichum*, *Fusarium* e *M. oryzae* (CUOMO et al., 2007; DEAN, 2005; MA et al., 2010). A expansão dessas enzimas é uma característica observada nos genomas do gênero *Coletotrichum* sendo e a abundância geral destas são inigualável em qualquer outro ascomiceto sequenciado, até mesmo nos fungos necrotróficos como *Botrytis cinerea* e *Sclerotinia sclerotiorum*. (AMSELEM et al., 2011; GAN et al., 2013; O'CONNELL et al., 2012).

A pectina é um dos principais componentes da parede celular de plantas. Em dicotiledôneas, a pectina constitui 35% da parede celular, enquanto em monocotiledôneas essa proporção é de 10% (VOGEL, 2008). As anotações dos genomas de *C. higginsianum*, *C. fructicola* e *C. orbiculare* revelaram um número notável de pectinases, o que provavelmente indica um caráter adaptativo desses agentes patogênicos, destacando-se *C. orbiculare* e *C. fructicola* que codificaram mais de 100 pectinases diferente, muito superior a outros fungos com genoma sequenciado (DEAN; LICHENS-PARK; KOLE, 2014).

1.9. Genes de metabolismo secundário

Genes relacionados ao metabolismo secundário compreendem uma classe gênica bem representativa no genoma de *Colletotrichum* que codifica enzimas putativas de metabólitos secundários (MS). As enzimas MS são moléculas de baixo peso molecular que não são essenciais para o crescimento e a sobrevivência do organismo, mas podem se tornar importantes para a adaptação ao ambiente. A produção desses metabólitos está frequentemente associada à competição por recursos através de efeitos tóxicos e/ou inibitórios em outros organismos (BOLKER et al., 2009; SHWAB; KELLER, 2008). Existem inúmeros exemplos de genes MS que funcionam dessa maneira, incluindo alguns bem conhecidos como o grupo de genes de tricoteno de *Fusarium graminearum* e o grupo de toxina AAL de *Alternaria alternata*, patogênica em tomate (AKAGI et al., 2009). Um grande número de MS geralmente é encontrado nos genomas de patógenos necrotróficos (AMSELEM et al., 2011), e as enzimas MS são frequentemente identificadas como fitotoxinas com função direta na patogenicidade (MATTHEWS; GREGORY; GRACEN, 1979; SCOTT-CRAIG et al., 1992). Em contraste, esses mesmos metabólitos estão reduzidos em fungos biotróficos como *Blumeria graminis* (SPANU et al., 2010) e *Ustilago maydis* (BOLKER et al., 2009; KÄMPER et al., 2006).

O papel dos MS em patógenos hemibiotróficos de plantas ainda é desconhecido (BOHNERT, 2004; COLLEMARE et al., 2008). Espécies de *Colletotrichum* produzem uma variedade de genes MS, que incluem flavonas, peptídeos e terpenos, e também a melanina de DHN (1,8-dihidroxi-naftaleno), requisito essencial para utilização do apressório (KUBO, Y; FURUSAWA, 1991; SINGH et al., 2010). Em *C. gloeosporioides* pode-se destacar sideróforo ferricrocina, que tem atividade fitotóxica em cotilédones de gramíneas (OHRA et al., 1995), colletotrichinas A, B e C de *C. nicotianae*, produzem sintomas parecidos ao da antracnose de tabaco quando presentes em folhas de tabaco (GARCÍA-PAJÓN; COLLADO, 2003; GODDARD et al., 1976; KIMURA; GOHBARA; SUZUKI, 1977) e um composto tetra-hidroxiado com propriedades antioxidantes de *C. gloeosporioides*. Vários metabólitos secundários foram caracterizados a partir de *C. graminicola*, incluindo os compostos antifúngicos monorden e monicilinas I, II e III (WICKLOW; JORDAN; GLOER, 2009) e micosporina-alanina, um inibidor de germinação de esporos (LEITE; NICHOLSON, 1992). Estudos evidenciaram que a deleção de PPT1, um gene que codifica um cofator essencial para a função enzimática de todas as policétido sintases (PKS) e sintetases de péptidos não ribossômicas (NRPS), resultaram em uma diminuição da patogenicidade em *C. graminicola*, sugerindo que os genes MS desempenham um papel importante na regulação da

patogenicidade (HORBACH et al., 2009). Estudos também mostraram que algumas espécies de *Colletotrichum*, incluindo *C. gloeosporioides*, são capazes de sintetizar a auxina, hormônio de crescimento da planta, o que provavelmente é importante na manipulação do hospedeiro (CHUNG et al., 2003; ROBINSON; RIOV; SHARON, 1998). Todas as quatro espécies de *Colletotrichum* com genomas sequenciados contêm genes com potencial para codificar para produção e efluxo de auxina (GAN et al., 2016; O'CONNELL et al., 2012).

Os genes para síntese de auxina através de um intermediário indole-3-acetamida (IAM) estão presentes em *C. fructicola* e *C. graminicola*, já a síntese de auxina por *C. higginsianum* e *C. obiculare* pode ocorrer através de um intermediário diferente (DEAN; LICHENS-PARK; KOLE, 2014). Os genes que codificam os putativos PKS e os genes tipo PKS e dimetiltransferases (DMAT) são especialmente abundantes nos quatro genomas de *Colletotrichum* sequenciados (GAN et al., 2013; O'CONNELL et al., 2012). *C. higginsianum* e *C. fructicola* possuem muitos mais genes de PKS, do que qualquer outro fungo sequenciado, embora é importante ressaltar que essas duas montagens apresentam baixa qualidade, por isso pode interferir nos números total de PKS pela fragmentação dos genes (DEAN; LICHENS-PARK; KOLE, 2014). Além dos genes envolvidos na produção de melanina, alguns dos genes preditos de MS possuem ortólogos ausente do gênero, e muitos parecem ser específicos para cepas individuais. Em todas as espécies de fungos, os genes MS tendem a ser organizados em clusters; muitas vezes esses clusters incluem enzimas adicionais, genes do citocromo P450, genes do fator de transcrição e genes transportadores (SHWAB; KELLER, 2008). O número de clusters MS previsto nas quatro cepas de *Colletotrichum* sequenciadas é excepcionalmente alto, variando de 42 clusters em *C. orbiculare* a 56 grupos em *C. fructicola*. Esta abundância de agrupamentos MS é substancialmente maior do que outros fungos patogênicos (GAN et al., 2013; O'CONNELL et al., 2012).

Fungos do gênero *Colletotrichum* são importantes como modelos experimentais em estudos que abrangem diversos aspectos da fitopatologia, incluindo o desenvolvimento do fungo, enzimas que degradam carboidratos, processos infectivos, resistência do hospedeiro e biologia molecular das interações planta-patógeno (SUKNO et al., 2007, 2008; THON et al., 2002).

Os dados deste estudo fornecem uma compreensão básica da estrutura genômica de um isolado de *C. musae* associados à antracnose da banana; na perspectiva que esses dados ajudem no desenvolvimento de manejo efetivo da doença, pois o manejo da doença consiste atualmente na aplicação de fungicidas e uso de cultivares resistentes. A estrutura genômica de um agente patogênico e o conhecimento do seu conteúdo gênico poderão auxiliar no

entendimento evolutivo e adaptativo do patógeno auxiliando cada vez mais o manejo da doença. Portanto, o conhecimento da genômica de patógenos é um pré-requisito para programas efetivos de implementação para a seleção de cultivares resistentes à doença e para evitar a resistência a fungicidas.

2. REFERÊNCIAS BIBLIOGRÁFICAS

ADEGBITE, A. A.; AMUSA, N. A. The major economic field diseases of cowpea in the humid agro-ecologies of South-western Nigeria. v. 7, n. 25, p. 4706–4712, 2008.

AGRIOS, G. Plant Pathology 5th Edition. **San Diego: Academic Press**, p. 922, 2005.

AKAGI, Y. et al. Horizontal chromosome transfer, a mechanism for the evolution and differentiation of a plant-pathogenic fungus. **Eukaryotic Cell**, v. 8, n. 11, p. 1732–1738, 2009.

ALTEKAR, G. et al. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. **Bioinformatics**, v. 20, n. 3, p. 407–415, 2004.

ALVES, E. J. **A CULTURA DA BANANA**. 2. ed. Brasília: [s.n.].

AMSELEM, J. et al. Genomic analysis of the necrotrophic fungal pathogens sclerotinia sclerotiorum and botrytis cinerea. **PLoS Genetics**, v. 7, n. 8, 2011.

ARROYO, F. T. et al. Ultrastructure of the early stages of *Colletotrichum acutatum* infection of strawberry tissues. v. 500, p. 491–500, 2005.

ARVANITOYANNIS, I. S.; MAVROMATIS, A. Banana cultivars, cultivation practices, and physicochemical properties. **Critical Reviews in Food Science and Nutrition**, v. 49, n. 2, p. 113–135, 2009.

BECHINGER, C. et al. Optical Measurements of Invasive Forces Exerted by Appressoria of a Plant Pathogenic Fungus Published by : American Association for the Advancement of Science Stable URL : <http://www.jstor.org/stable/2897988> Optical Measurements of Invasive Forces Exerted. v. 285, n. 5435, p. 1896–1899, 2016.

BERKELEY, M. J. Notices of North American Fungi. **Grevillea**, v. 3, n. 25, p. 1–17, 1874.

BOHNERT, H. U. A Putative Polyketide Synthase/Peptide Synthetase from *Magnaporthe grisea* Signals Pathogen Attack to Resistant Rice. **the Plant Cell Online**, v. 16, n. 9, p. 2499–2513, 2004.

BOLKER, B. M. et al. Generalized linear mixed models: a practical guide for ecology and evolution. **Trends in ecology & evolution**, v. 24, n. 3, p. 127–35, 2009.

CANNON, P. F. et al. *Colletotrichum* - current status and future directions. **Studies in Mycology**, v. 73, p. 181–213, 2012.

CHONGO, G.; GOSSEN, B. D.; BERNIER, C. C. Note : Genetics and resistance Infection by *Colletotrichum truncatum* in resistant and susceptible lentil genotypes. v. 85, p. 81–85, 2002.

CHUNG, K. R. et al. Indole derivatives produced by the fungus *Colletotrichum acutatum* causing lime anthracnose and postbloom fruit drop of citrus. **FEMS Microbiology Letters**, v. 226, n. 1, p. 23–30, 2003.

COLLEMARE, J. et al. *Magnaporthe grisea* avirulence gene ACE1 belongs to an infection-

specific gene cluster involved in secondary metabolism. **New Phytologist**, v. 179, n. 1, p. 196–208, 2008.

CORDEIRO, Z. ORG. Banana: fitossanidade. **Produção: aspectos técnicos. Brasília: Embrapa Comunicação para Transferência de Tecnologia**, p. 121, 2000.

CROUCH, J. et al. The genomics of *colletotrichum*. In: **Genomics of Plant-Associated Fungi: Monocot Pathogens**. [s.l: s.n.]. p. 69–102.

CROUS, P. W. et al. Global food and fibre security threatened by current inefficiencies in fungal identification. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 371, n. 1709, p. 20160024, 2016.

DEAN, R. et al. The Top 10 fungal pathogens in molecular plant pathology. **Molecular Plant Pathology**, v. 13, n. 4, p. 414–430, 2012.

DEAN, R. A. The genome sequence of the rice blast fungus *Magnaporthe grisea*. **Nature**, v. 434, n. 7036, p. 980–986, 2005.

DEAN, R. A.; LICHENS-PARK, A.; KOLE, C. **Genomics of plant-associated fungi: Monocot pathogens**. [s.l: s.n.].

DELAYE, L.; GARCÍA-GUZMÁN, G.; HEIL, M. Endophytes versus biotrophic and necrotrophic pathogens-are fungal lifestyles evolutionarily stable traits? **Fungal Diversity**, v. 60, n. 1, p. 125–135, 2013.

DIAS, J. DO S. A. A cultura da bananeira. **Aspectos agronômicos, fitopatológicos e socioeconômicos da sigatoka-negra na cultura da bananeira no Estado do Amapá**, p. 95, 2011.

FAO Statistics. Disponível em:

<<http://www.fao.org/faostat/en/#datanload/Q/QC/E%5Cnhttp://faostat3.fao.org/>>.

GAN, P. et al. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. **New Phytologist**, v. 197, n. 4, p. 1236–1249, 2013.

GAN, P. et al. 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**, v. 8, n. 5, p. 1467–1481, 2016.

GARCÍA-PAJÓN, C. M.; COLLADO, I. G. Secondary metabolites isolated from *Colletotrichum* species. **Nat. Prod. Rep.**, v. 20, n. 4, p. 426–431, 2003.

GODDARD, R. et al. X-Ray crystal and molecular structure of acetylcolletotrichin (colletotrichin), a metabolite of *Colletotrichum capsici*. **Journal of the Chemical Society, Chemical Communications**, n. 11, p. 408, 1976.

GREEN, K. R.; SIMONS, S. A. "Dead skin" on yams (*Dioscorea alata*) caused by *Colletotrichum gloeosporioides*. p. 1062–1065, 1994.

HORBACH, R. et al. Sfp-Type 4'-Phosphopantetheinyl Transferase Is Indispensable for Fungal Pathogenicity. **The Plant Cell**, v. 21, n. 10, p. 3379–3396, 2009.

- HYDE, K. D. et al. *Colletotrichum* – names in current use. **Fungal Diversity - online advanced**, v. 39, n. November, p. 147–182, 2009.
- IBGE. IBGE. **Levantamento Sistemático Da Produção Agrícola**, v. 30, p. 115, 2017.
- INTAN SAKINAH, M. A.; SUZIANI, I. V.; LATIFFAH, Z. Phenotypic and molecular characterization of *Colletotrichum* species associated with anthracnose of banana (*Musa* spp) in Malaysia. **Genetics and Molecular Research**, v. 13, n. 2, p. 3627–3637, 2014.
- JEGER M.J., EDEN-GREEN S., THRESH J.M., JOHANSON A., WALLER J.M., B. A. E. Banana diseases. In: **Banana diseases**. World Crop ed. Dordrecht: Springer, 1995. p. 317–381.
- KÄMPER, J. et al. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. **Nature**, v. 444, n. 7115, p. 97–101, 2006.
- KIMATI, H.; AMORIM, BERGAMIN FILHO, A.; REZENDE, L. E. A. **Manual de Fitopatologia: Doenças das plantas cultivadas**. [s.l: s.n.]. v. 2
- KIMURA, Y.; GOHBARA, M.; SUZUKI, A. Assignment of ¹³C-nmr spectrum and biosynthesis of colletotrichin. **Tetrahedron Letters**, v. 18, n. 52, p. 4615–4618, 1977.
- KLEEMANN, J. et al. Identification of soluble secreted proteins from appressoria of *Colletotrichum higginsianum* by analysis of expressed sequence tags. **Microbiology**, v. 154, n. 4, p. 1204–1217, 2008.
- KUBO, Y; FURUSAWA, I. . **Melanin biosynthesis: prerequisite for successful invasion of the plant host by appressoria of *Colletotrichum* and *Pyricularia***. Plenum Press, , 1991.
- KUBO, Y.; TAKANO, Y. Dynamics of infection-related morphogenesis and pathogenesis in *Colletotrichum orbiculare*. **Journal of General Plant Pathology**, v. 79, 2013.
- LEITE, B.; NICHOLSON, R. L. Mycosporine-alanine: A self-inhibitor of germination from the conidial mucilage of *Colletotrichum graminicola*. **Experimental Mycology**, v. 16, n. 1, p. 76–86, 1992.
- LUCIANA, A.; BORGES, A. N. A. L. B. Sistema de produção da bananeira irrigada. **Embrapa Semiárido**, 2009.
- MA, L.-J. et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. **Nature**, v. 464, n. 7287, p. 367–373, 2010.
- MAHADTANAPUK, S. et al. Control of anthracnose caused by *Colletotrichum musae* on *Curcuma alismatifolia* Gagnep. using antagonistic *Bacillus* spp. **American Journal of Agricultural and Biological Science**, v. 2, n. 2, p. 54–61, 2007.
- MATTHEWS, D. E.; GREGORY, P.; GRACEN, V. E. *Helminthosporium maydis* Race T Toxin Induces Leakage of NAD from T Cytoplasm Corn Mitochondria. **Plant physiology**, v. 63, n. 6, p. 1149–1153, 1979.
- MCTAGGART, A. R. et al. Fungal Genomics Challenges the Dogma of Name-Based Biosecurity. **PLoS Pathogens**, v. 12, n. 5, p. 10–12, 2016.

- MEDEMA, M. H.; TAKANO, E.; BREITLING, R. Detecting sequence homology at the gene cluster level with multigeneblast. **Molecular Biology and Evolution**, v. 30, n. 5, p. 1218–1223, 2013.
- MENDGEN, K.; HAHN, M.; DEISING, H. Morphogenesis and Mechanisms of Penetration By Plant Pathogenic Fungi. **Annual Review of Phytopathology**, v. 34, n. 1, p. 367–386, 1996.
- NASH, E. M. C.; STRANGE, R. N. *Colletotrichum gloeosporioides* as the cause of stem tip dieback of cassava. 1996.
- O'CONNELL, R. J. et al. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. **Nature Genetics**, v. 44, n. 9, p. 1060–1065, 2012.
- OHRA, J. et al. Production of the phytotoxic metabolite, ferricrocin, by the fungus *Colletotrichum gloeosporioides*. **Bioscience, Biotechnology and Biochemistry**, v. 59, n. 1, p. 113–114, 1995.
- OSTELL, J.; MCENTYRE, J. The NCBI Handbook. **NCBI Bookshelf**, p. 1–8, 2007.
- PERES, N. A. et al. Lifestyles of *Colletotrichum acutatum*. **Plant Disease**, v. 89, n. 8, p. 784–796, 2005.
- PHOULIVONG, S. et al. *Colletotrichum gloeosporioides* is not a common pathogen on tropical fruits. **Fungal Diversity**, v. 44, p. 33–43, 2010.
- PLOETZ, R. C.; THOMAS, J. E.; SLABAUGH, W. R. Diseases of banana and plantain. **America**, p. 73–134, 2003.
- PRUSKY, D. PATHOGEN QUIESCENCE IN POSTHARVEST DISEASES. 1996.
- RANASINGHE, L. S.; JAYAWARDENA, B.; ABEYWICKRAMA, K. Use of waste generated from cinnamon bark oil (*Cinnamomum zeylanicum* Blume) extraction as a post harvest treatment for Embul banana. **Journal of Food, Agriculture & Environment**, v. 1, n. 2, p. 340–344, 2003.
- RÉBLOVÁ, M.; GAMS, W.; SEIFERT, K. A. Monilochaetes and allied genera of the Glomerellales, and a reconsideration of families in the Microascales. p. 163–191, 2011.
- RIGOTTI, S. et al. Characterization of molecular markers for specific and sensitive detection of *Botrytis cinerea* Pers. : Fr. in strawberry (*Fragaria x ananassa* Duch.) using PCR. **Microbiology Letters**, v. 209, p. 169–174, 2002.
- ROBINSON, M.; RIOV, J.; SHARON, A. Indole-3-Acetic Acid Biosynthesis in *Colletotrichum gloeosporioides* f. sp. *aeschynomene*. v. 64, n. 12, p. 5030–5032, 1998.
- ROGÉRIO, W. et al. Efeito da temperatura e período de molhamento sobre o desenvolvimento de lesões de *Colletotrichum musae* em banana. p. 147–151, 2007.
- SCOTT-CRAIG, J. S. et al. The cyclic peptide synthetase catalyzing HC-toxin production in the filamentous fungus *Cochliobolus carbonum* is encoded by a 15.7-kilobase open reading frame. **Journal of Biological Chemistry**, v. 267, n. 36, p. 26044–26049, 1992.

- SHWAB, E. K.; KELLER, N. P. Regulation of secondary metabolite production in filamentous ascomycetes. **Mycological Research**, v. 112, n. 2, p. 225–230, 2008.
- SILVA, S. DE O. E et al. Melhoramento genético da bananeira: estratégias e tecnologias disponíveis. **Revista Brasileira de Fruticultura**, v. 35, n. 3, p. 919–931, 2013.
- SILVA, S. O. et al. Banana Breeding Program at Embrapa. **Cropp Breeding and Applied Biotechnology**, v. 1, n. 4, p. 399–436, 2001.
- SILVA JUNIOR, W. J. DA et al. Draft genome assembly of *Colletotrichum musae*, the pathogen of banana fruit. **Data in Brief**, v. 17, p. 256–260, 2018.
- SINGH, J. et al. Production and Extraction of Phytotoxins from *Colletotrichum dematium* FGCC# 20 effective against *Parthenium hysterophorus* L. **Brazilian Archives of Biology and Technology**, v. 53, n. 3, p. 669–678, 2010.
- SLABAUGH, W. R.; GROVE, M. D. Postharvest diseases of bananas and their control. **Plant Disease**, v. 66, n. 8, p. 746–750, 1982.
- SOUZA, A. DA S. et al. **Banana. Produção: aspectos técnicos**. Brasília: Embrapa, 2000.
- SPANU, P. D. et al. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. **Science**, v. 330, n. 6010, p. 1543–1546, 2010.
- STERGIOPOULOS, I.; GORDON, T. R. Cryptic fungal infections: the hidden agenda of plant pathogens. **Frontiers in Plant Science**, v. 5, n. September, p. 10–13, 2014.
- SU, Y. Y. et al. Epitypification of *Colletotrichum musae*, the causative agent of banana anthracnose. **Mycoscience**, v. 52, n. 6, p. 376–382, 2011.
- SUTTON, B. C. The Coelomycetes. Fungi imperfecti with pycnidia, acervuli and stromata. **Commonwealth Mycol. Inst., Surrey, England**, p. 696, 1980.
- SUTTON, B. C. **The genus Glomerella and its anamorph Colletotrichum. Colletotrichum: biology, pathology and control**. Wallingford CAB International, , 1992.
- THON, M. R.; NUCKLES, E. M.; VAILLANCOURT, L. J. Restriction Enzyme-Mediated Integration Used to Produce Pathogenicity Mutants of *Colletotrichum graminicola*. **Molecular Plant-Microbe Interactions**, v. 13, n. 12, p. 1356–1365, 2000.
- TSEDALEY, B. A Review on Disease Detection, Pathogen Identification and Population Genetics in Fungi. **Journal of Biology, Agriculture and Healthcare**, v. 5, n. 1, p. 6–20, 2015.
- VOGEL, J. Unique aspects of the grass cell wall. **Current Opinion in Plant Biology**, v. 11, n. 3, p. 301–307, 2008.
- VON ARX, J. A. **Die Arten der Gattung Colletotrichum Cda: [s.l: s.n.]**.
- WEIR, B. S.; JOHNSTON, P. R.; DAMM, U. The *Colletotrichum gloeosporioides* species complex. **Studies in Mycology**, v. 73, p. 115–180, 2012.

WHARTON, P. S. The biology of *Colletotrichum acutatum* by. v. 61, n. 1, p. 3–22, 2004.

WICKLOW, D. T.; JORDAN, A. M.; GLOER, J. B. Antifungal metabolites (monorden, monocillins I, II, III) from *Colletotrichum graminicola*, a systemic vascular pathogen of maize. **Mycological Research**, v. 113, n. 12, p. 1433–1442, 2009.

ZEILINGER, S. et al. Friends or foes? Emerging insights from fungal interactions with plants. **FEMS Microbiology Reviews**, v. 40, n. 2, p. 182–207, 2016.

ZHANG, N. et al. An overview of the systematics of the Sordariomycetes based on a four-gene phylogeny. v. 98, n. 6, p. 1076–1087, 2006.

CAPÍTULO II

Draft genome assembly of *Colletotrichum musae*, the pathogen of banana fruit



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Draft genome assembly of *Colletotrichum musae*, the pathogen of banana fruit

Wilson José da Silva Junior^{a,*}, Raul Maia Falcão^b,
 Lucas Christian de Sousa-Paula^b, Nicolau Sbaraini^c,
 Willie Anderson dos Santos Vieira^a, Waléria Guerreiro Lima^d,
 Sérgio de Sá Leitão Paiva Junior^b, Charley Christian Staats^c,
 Augusto Schrank^c, Ana Maria Benko-Iseppon^b,
 Valdir de Queiroz Balbino^{b,1}, Marcos Paz Saraiva Câmara^{a,1}

^a Departamento de Agronomia, Universidade Federal Rural de Pernambuco, Recife, PE, Brazil

^b Departamento de Genética, Universidade Federal de Pernambuco, Recife, PE, Brazil

^c Programa de Pós-Graduação em Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul, RS, Brazil

^d Faculdade Guararapes, Pernambuco, PE, Brazil

ARTICLE INFO

Article history:

Received 4 December 2017

Received in revised form

19 December 2017

Accepted 2 January 2018

Available online 8 January 2018

ABSTRACT

Colletotrichum musae is an important cosmopolitan pathogenic fungus that causes anthracnose in banana fruit. The entire genome of *C. musae* isolate GM20 (CMM 4420), originally isolated from infected banana fruit from Alagoas State, Brazil, was sequenced and annotated. The pathogen genomic DNA was sequenced on HiSeq Illumina platform. The *C. musae* GM20 genome has 50,635,197 bp with G + C content of 53.74% and in its present assembly has 2763 scaffolds, harboring 13,451 putative genes with an average length of 1626 bp. Gene prediction and annotation was performed by Funannotate pipeline, using a pattern for gene identification based on BUSCO.

© 2018 Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author.

E-mail addresses: wilson_jsjunior@hotmail.com (W.J.d. Silva Junior), rmf4@cin.ufpe.br (R.M. Falcão), lcousapaula@gmail.com (L.C.d. Sousa-Paula), nicolausbaraini@icloud.com (N. Sbaraini), andersonvieira12@gmail.com (W.A.d.S. Vieira), wagueli@hotmail.com (W.G. Lima), sslpaiva@gmail.com (S.d.S.L. Paiva Junior), staats@ufrgs.br (C.C. Staats), aschrank@cbiot.ufrgs.br (A. Schrank), ana.iseppon@gmail.com (A.M. Benko-Iseppon), valdir@ufpe.br (V.d.Q. Balbino), marcos.camara@ufrpe.br (M.P.S. Câmara).

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.dib.2018.01.002>

2352-3409/© 2018 Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area	Biology
More specific subject area	Microbiology, Agricultural, Genomics.
Type of data	Genome sequence data
How data was acquired	Illumina HiSeq. 2500 Next Generation Platforms
Data format	Assembled genome sequence.
Experimental factors	Genomic DNA was extract from mycelial growth in culture medium.
Experimental features	Genome of <i>Colletotrichum musae</i> strain GM20 was sequenced and assembled.
Data source location	<i>Colletotrichum musae</i> strain GM20 was isolated from banana lesions, in Maceio, Pernambuco Brazil.
Data accessibility	The <i>Colletotrichum musae</i> GM20 genome is available in DDBJ/ENA/GenBank under the accession number NWMS01000000.
Related research article	
Data accessibility	https://www.ncbi.nlm.nih.gov/nucore/NWMS00000000

Value of the Data

- *Colletotrichum musae* is the causal agent of anthracnose in banana fruits, the main disease post-harvest worldwide.
- This is the first genome sequence of *Colletotrichum musae* using next-generation sequencing available in public database.
- The published genome data herein will facilitate biology, pathogenicity, evolution and interaction pathogen-host studies of *Colletotrichum musae*, through comparative genomes studies of *Colletotrichum* spp. and related species.

1. Data

Fungi infection in plants is the most frequent cause of extensive loses in Agriculture. The fact that many endophytic fungi can case infection adds further complexity to fungal plant pathogens. Banana (*Musa* sp.) is one of the world's important food crops and a staple food for more than 400 million people [1]. Over 100 million tons are produced worldwide at some 5 million hectares and the cultivated area is expected to increase in the future [2]. However, banana fruits are highly susceptible to pathogens, and anthracnose disease caused by fungi from *Colletotrichum* genus is amongst the most frequents. *Colletotrichum* comprises over 100 species that are able to infect and damage diverse crops around the world [3].

Due to its ubiquity, substantial destruction capacity and scientific importance as a model of pathosystems, *Colletotrichum* spp. are among the top 10 of most important plant pathogens according to the international community of plant pathology researchers [4]. *Colletotrichum musae* (Berk. and M. A. Curtis), the causative agent of anthracnose, is a major post-harvest pathogen of banana fruits and causes severe global crop losses [5]. The disease develops from a latent fungal infection during pre-harvest, originated from spores that are present in immature fruits in the field. Symptoms, such as patches on the bark (brown to black color) and depressed lesions, appear in the ripening of the fruits. Furthermore, under high humidity, the formation of salmon-colored acervuli can be observed [6]. The infection thus accounts for a reduction in fruit viability during maturation, transport and storage periods [7], leading to a commercial depreciation and shortening fruit's shelf life.

To circumvent post-harvest losses, chemical fungicides are usually adopted, but other side-methods (e.g., radiation treatment, hot water removal, refrigeration, induced resistance and biological control agents) have also been applied [8]. However, chemical fungicide usage has been limited by potential harmful effects to human health and environment. Besides, fungal pathogens are known to quickly develop resistance to chemical defensives [9].

Furthermore, the absence of available genomic sequences from *C. musae* is one of the main limitations for best characterization of fungal virulence determinants and development of improved management strategies. Here we report, for the first time, the whole genome sequence of the *C. musae* strain GM20 (CMM 4420) isolated from infected banana fruit from Alagoas, Brazilian Northeast State.

In recent years, several phytopathogenic fungal genomes have been published boosting the discovery of virulence determinants in these species. Expectedly, our analysis will encourage further studies of *C. musae* biology, which should provide better details about host-pathogen interaction, leading to new management measures.

2. Experimental design, materials, and methods

2.1. DNA extraction and genome sequence

The GM20 isolate of *C. musae* was cultured, and DNA was extracted as previously described [10]. Whole shotgun genome sequence of *C. musae* GM20 was generated using the Illumina HiSeq. 2500 platform (Illumina, San Diego, CA) at the Center for Functional Genomics - Universidade de São Paulo (Piracicaba, Brazil). The libraries were prepared with the Illumina Nextera XT DNA Library Prep Kit (Illumina, San Diego, CA) and the sequencing was performed on a HiSeq Flow Cell v4 with HiSeq SBS Kit v4 (Illumina, San Diego, CA), leading to 100 bp paired-reads (2×).

2.2. De novo assembly and genes annotation

The shotgun sequencing produced 13,273,851 paired reads. Initially, FastQC [11] was applied to analyse reads quality, and adapters were trimmed using FASTX-Toolkit 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit). Originally, three assemblers were tested: ABySS 2.0.2 [12]; SPAdes 1.10 [13]; Velvet 1.1 [14], with SPAdes showing the best results (12,435 contigs > 500 bp). Additionally, Redundans [15] posteriorly ran for scaffolds assembly.

Assembly statistics were generated by QUAST 3.9 (Table 1) [16]. Gene prediction and annotation was carried out with Funannotate pipeline [17] BUSCO 2.0 [18] [parameters: Sordariomycetes database (*Verticillium longporum* selected as closely-related species)] to generate the training files for two genome predictors: GeneMark-ES [19] and AUGUSTUS [20]. Moreover, BUSCO 2.0 was employed to evaluate genome completeness, based on conservation of single-copy benchmarking universal single-copy orthologs (BUSCOs).

Table 1
Genome assembly statics for *Colletotrichum musae* GM20.

	<i>C. musae</i> GM20
Assembly size	50.7 Mb
Coverage sequencing	100×
Sequencing technology	Illumina HiSeq. 2500
Number of scaffolds	2763
N50 scaffolds length	32,818
Number of contigs	10,618
Number of predicts genes	13,451
Overall GC content	53.74
Public access to genome	NWMS01000000

The final assembly of the *C. musae* GM20 genome was determined to be 50,635,197 bp with a G+C content of 53.74% in 2763 scaffolds (maximum 208,119 bp; N50 32,818 bp), and 13,451 genes were predicted. This whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number NWMS00000000. The version described in this paper is version NWMS01000000.

BUSCO analysis showed a high degree of completeness with a BUSCO score of 96.3%, of which 1263 genes were complete BUSCOs, four were complete duplicated BUSCOs, 23 were fragmented BUSCOs, and 25 were missing BUSCO orthologs out of the 1315 BUSCO groups searched.

Acknowledgments

The authors thank to CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil) (23038.010050/2013-04) Bio-Computational Program and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) (310871/2014-0) for financial support and fellowships.

Transparently document. Supporting information

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2018.01.002>.

References

- [1] D. Holscher, S. Dhakshinamoorthy, T. Alexandrov, M. Becker, T. Bretschneider, A. Buerkert, A.C. Crecelius, D. De Waele, A. Elsen, D.G. Heckel, H. Heklau, C. Hertweck, M. Kai, K. Knop, C. Krafft, R.K. Maddula, C. Matthaus, J. Popp, B. Schneider, U. S. Schubert, R. a Sikora, A. Svato, R.L. Swennen, Phenalenone-type phytoalexins mediate resistance of banana plants (*Musa* spp.) to the burrowing nematode *Radopholus similis*, *Proc. Natl. Acad. Sci.* 111 (2014) 105–110. <http://dx.doi.org/10.1073/pnas.1314168110>.
- [2] FAO, Food and agricultural organization. (<http://www.fao.org/home/en/>), 2017 (Accessed 01 Jan 2017).
- [3] P.F. Cannon, U. Damm, P.R. Johnston, B.S. Weir, Colletotrichum - current status and future directions, *Stud. Mycol.* 73 (2012) 181–213. <http://dx.doi.org/10.3114/sim0014>.
- [4] R. Dean, J.A.L. Van Kan, Z.A. Pretorius, K.E. Hammond-Kosack, A. Di Pietro, P.D. Spanu, J.J. Rudd, M. Dickman, R. Kahmann, J. Ellis, G.D. Foster, The Top 10 fungal pathogens in molecular plant pathology, *Mol. Plant Pathol.* (2012), <http://dx.doi.org/10.1111/j.1364-3703.2012.00822.x> (804–804).
- [5] M. Maqbool, A. Ali, S. Ramachandran, D.R. Smith, P.G. Alderson, Control of postharvest anthracnose of banana using a new edible composite coating, *Crop Prot.* 29 (2010) 1136–1141. <http://dx.doi.org/10.1016/j.cropro.2010.06.005>.
- [6] L.S. Ranasinghe, B. Jayawardena, K. Abeywickrama, Use of waste generated from cinnamon bark oil (*Cinnamomum zeylanicum* Blume) extraction as a post harvest treatment for Embul banana, *J. Food Agric. Environ.* 1 (2003) 340–344 (<http://www.world-food.net>).
- [7] W.R. Slabaugh, M.D. Grove, Postharvest diseases of bananas and their control, *Plant Dis.* 66 (1982) 746–750. <http://dx.doi.org/10.1094/PD-66-746>.
- [8] V.Y. Zhimo, D. Dilip, J. Sten, V.K. Ravat, D.D. Bhutia, B. Panja, J. Saha, Antagonistic Yeasts for Biocontrol of the banana postharvest anthracnose pathogen *Colletotrichum musae*, *J. Phytopathol.* 165 (2017) 35–43. <http://dx.doi.org/10.1111/jph.12533>.
- [9] H. Sonah, R.K. Deshmukh, R.R. Bélanger, Computational prediction of effector proteins in fungi: opportunities and challenges, *Front. Plant Sci.* 7 (2016) 1–14. <http://dx.doi.org/10.3389/fpls.2016.00126>.
- [10] J.J.D.J.L.J.J. Doyle, J.J.D.J.L.J.J. Doyle, Isolation of plant DNA from fresh tissue, *Focus (Madison)* 12 (1990) 13–15. <http://dx.doi.org/10.3923/rjmp.2012.65.73>.
- [11] S. Andrews, FastQC: a quality control tool for high throughput sequence data, Babraham Bioinform. (2010) (citeulike-article-id:11583827) <http://www.bioinformatics.babraham.ac.uk/projects/>.
- [12] J.T. Simpson, K. Wong, S.D. Jackman, J.E. Schein, S.J.M. Jones, I. Birol, ABySS: a parallel assembler for short read sequence data, *Genome Res.* 19 (2009) 1117–1123. <http://dx.doi.org/10.1101/gr.089532.108>.
- [13] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A. D. Pribludski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.* 19 (2012) 455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.
- [14] D.R. Zerbino, Using the Velvet de novo assembler for short-read sequencing technologies, *Curr. Protoc. Bioinform.* (2010), <http://dx.doi.org/10.1002/0471250953.bi1105s31>.
- [15] L.P. Prysacz, T. Gabaldón, Redundans: an assembly pipeline for highly heterozygous genomes, *Nucleic Acids Res.* 44 (2016) e113. <http://dx.doi.org/10.1093/nar/gkw294>.

- [16] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QAST: quality assessment tool for genome assemblies, *Bioinformatics*. 29 (2013) 1072–1075. <http://dx.doi.org/10.1093/bioinformatics/btt086>.
- [17] J.M. Palmer, Funannotate: a Fungal Genome Annotation and Comparative Genomics Pipeline(<https://github.com/nextgenusfs/funannotate>), 2016.
- [18] F.A. Simão, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics* 31 (2015) 3210–3212. <http://dx.doi.org/10.1093/bioinformatics/btv351>.
- [19] J. Besemer, GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions, *Nucleic Acids Res.* 29 (2001) 2607–2618. <http://dx.doi.org/10.1093/nar/29.12.2607>.
- [20] M. Stanke, R. Steinkamp, S. Waack, B. Morgenstern, AUGUSTUS: a web server for gene finding in eukaryotes, *Nucleic Acids Res.* (2004), <http://dx.doi.org/10.1093/nar/gkh379>.

CAPÍTULO III

Secondary metabolite biosynthetic gene clusters in the banana-specialist pathogen *Colletotrichum musae*

Title:

Secondary metabolite biosynthetic gene clusters in the banana-specialist pathogen *Colletotrichum musae*

Authors:

Wilson José da Silva Junior¹ - wilson_jsjunior@hotmail.com

Lucas Christian de Sousa-Paula¹ - lcsousapaula@gmail.com

Sérgio de Sá Leitão Paiva Junior² - sslpaiva@gmail.com

Antonio Roberto Gomes de Farias¹ - rfariasagro@gmail.com

Raul Maia Falcão² - rmf4@cin.ufpe.br

Ana Maria Benko-Isenpon² - ana.iseppon@gmail.com

Nicolau Sbaraini³ - nicolausbaraini@me.com

Fábio Carrer Andreis³ - fabio.andreis@gmail.com

Augusto Schrank³ - aschrank@cbiot.ufrgs.br

Charley Christian Staats³ - staats@cbiot.ufrgs.br

Valdir de Queiroz Balbino^{#,*,2} - valdir@ufpe.br

Marcos Paz Saraiva Câmara^{#,*,1} - marcos.camara@ufrpe.br

Affiliations:

¹Departamento de Agronomia, Universidade Federal Rural de Pernambuco, Recife, PE, Brazil

²Departamento de Genética, Universidade Federal de Pernambuco, Recife, PE, Brazil

³Programa de Pós-Graduação em Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul, RS, Brazil

Correspondence author contact:

valdir@ufpe.br and marcos.camara@ufrpe.br

#These authors contributed equally to this work

ABSTRACT

Bananas (*Musa* spp.) are an important food crop and a staple food for more than 400 million people. Bananas are highly susceptible to infections, and anthracnose disease caused by fungus from *Colletotrichum* genus (especially *Colletotrichum musae* described as banana-specialist pathogen) is a major post-harvest problem. The unavailability of genomic sequences from *C. musae* is a hindrance for characterization of fungal virulence determinants. Thus, the genome of *C. musae* GM20 (GM 4420) was sequenced and annotated. Among the several virulence determinants described we focused in secondary metabolites biosynthetic gene clusters (BGCs) which are enrolled in production of several compounds. The genes involved in secondary metabolite synthesis are almost underexploited in *Colletotrichum* spp., despite the progress in the study of *Colletotrichum*-host relationships. A deep survey and description of BGCs was conducted. Among 72 BGCs identified in *C. musae*, a total of 15 % have already characterized orthologs in other fungal species, allowing insights on their final products. From this work, apicidin/HC-toxin- and alternapyrone-like compounds gene clusters were identified in *C. musae* with presumably virulence-related roles. Additionally, the conservation of *C. musae*' BGCs was evaluated in 19 *Colletotrichum* spp. and *Verticillium dahliae*. Notably, 97 % of the identified BGCs of *C. musae* were conserved in *C. gloeosporioides*, a generalist pathogen. While some *C. gloeosporioides* BGCs are absent in *C. musae* genome. These results reinforce the putative importance of secondary metabolites in *C. musae* infection. Moreover, the detailed analyses conducted paves the way for future in-depth research.

INTRODUCTION

Plant pathogenic species are responsible for major losses to economically important crops . The impact of these destructive plant pathogens is exacerbated in developing countries, especially in tropic and subtropic regions. Species of *Colletotrichum* genus are important pathogenic fungi, infecting virtually every cultivated crop (CANNON et al., 2012; HYDE et al., 2009). Most *Colletotrichum* species, initially, establish a latent infection, switching for a destructive and necrotrophic phase secondarily. Notably, similar to other pathogens, while some *Colletotrichum* species display a broad host range, other species exhibit signs of host specialization. Understand the genetic differences between generalist and specialist pathogens is a major goal in the study of *Colletotrichum* genus as well as other pathogenic fungi. *Colletotrichum musae* is a specialist pathogens of bananas (*Musa* sp.), an essential food in Africa and Americas (DE SILVA et al., 2017). Additionally, *C. musae* is phylogenetically related to *C. gloeosporioides*, a generalist pathogen, which infection more than six host (DEAN; LICHENS-PARK; KOLE, 2014). In recent years, the genome of *C. gloeosporioides* and several other *Colletotrichum* species was sequenced, allowing interesting sources for comparative genomic analysis. Here, we analyzed the genomic sequences the genome of *C. musae* strain GM20 and seek secondary metabolite biosynthetic gene clusters (BGCs). The conservation of the described BGCs was assessed within the nineteen available *Colletotrichum* genomes. A special focus was kept in relation to *C.musae-C. gloeosporioides* BGCs presence/absence. Some interesting BGCs (i.e. linked with putative apicidin/HC-toxin-like and alternapyrone-like compounds) were explored through phylogenetic inference and comparative genomic analyses.

MATERIAL AND METHODS

Prediction of secondary metabolite gene clusters

All fungal genomes from *Colletotrichum* spp. and *Verticillium dahliae* strain VdLs.17 were downloaded from the NCBI Genome Database, with the exception of the *C. musae* GM20 genome isolated and sequenced by our group (SILVA JUNIOR et al., 2018). The descriptions and accession numbers are displayed in Table 1. To assess the completeness and quality of genome sequence assembly/predicted proteins, BUSCO version 2.0 (SIMÃO et al., 2015) was used with the Sordariomyceta dataset and with species parameters set to the nearest represented species (*V. dahliae*). The quality assessment are displayed in Supplementary Table 1.

Putative BGCs in the *C. musae* GM20 genome were identified with antiSMASH 4.0 (using the genome sequence assembly as input) (WEBER et al., 2015) and SMIPS 2016-07-26 (using the predicted proteins as input) (WOLF et al., 2016) algorithms. Each BGC was manually analyzed and sorted according to BGC and backbone gene features. The BGCs were sorted in five categories: PKS (type 1 and 3), NRPS, terpenes (TERP), indoles (IND) and OTHER (a generic class of gene clusters grouping BGCs with unusual backbone genes [e.g., NRPS-like and PKS-like]). Afterward, BLASTP (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>; non-redundant protein sequence database) (ALTSCHUL et al., 1990) was used to search for and curate orthologous BGCs among other genomes of filamentous fungi and to select putative orthologous for phylogenetic analysis (amino acid sequences were used and three cutoffs were defined: e -value $< 1 \times 10^{-5}$, query coverage ≥ 50 %, and identity ≥ 45 %). Additionally, several fungal genome sequences have been deposited at NCBI as raw or incomplete assemblies. These projects do not have predicted proteins/genes/mRNAs deposited at NCBI, and are therefore inaccessible through routine BLASTp against the non-redundant protein sequences (nr) database and BLASTn against the nucleotide collection (nr/nt) database. Thus, a screening using BLASTn against the WGS database was also performed in order to select putative orthologous for phylogenetic analysis. When a putative positive match was found, the genome sequences were downloaded and putative genes were predicted with FGENESH (gene-finding parameters for the closest species presented in the algorithm). Genes that satisfied the previously fixed cutoffs were included in further analyses. Furthermore, in order to compare, putative BGCs in the *Colletotrichum gloeosporioides* strain Nara gc5 genome were also predicted.

The conservation of predicted BGCs among *Colletotrichum* spp. (*Colletotrichum gloeosporioides* strain Nara gc5, *Colletotrichum fiorinae* strain PJ7, *Colletotrichum*

nymphaeae strain SA-01, *Colletotrichum chlorophyti* strain NTL11, *Colletotrichum orbiculare* strain MAFF 240422, *Colletotrichum incanum* strain MAFF 238704, *Colletotrichum salicis* strain CBS 607.94, *Colletotrichum simmondsii* strain CBS122122, *Colletotrichum sublineola* strain TX430BB, *Colletotrichum acutatum* strain 1, *Colletotrichum acutatum* strain C71, *Colletotrichum higginsianum* strain IMI 349063, *Colletotrichum godetiae* strain C184, *Colletotrichum falcatum* strain Cf671, *Colletotrichum tofieldiae* strain CBS127615, *Colletotrichum orchidophilum* strain IMI 309357, *Colletotrichum graminicola* strain M1.001, *Colletotrichum lindemuthianum* strain 83.501, *Colletotrichum lindemuthianum* strain 89) and *Verticillium dahliae* strain VdLs.17 was assessed using MultiGeneBlast 1.1.14 (analyses of genomic loci with predicted proteins) (MEDEMA; TAKANO; BREITLING, 2013), based primarily on backbone gene conservation (amino acid sequences; e -value $< 1 \times 10^{-5}$, query coverage ≥ 50 %, and identity ≥ 50 %). Additionally, MultiGeneBLAST 1.1.14 was also employed to compare putative orthologous biosynthetic gene clusters conserved in other fungal species.

Phylogenetic analyses

For phylogenetic analyses, the backbone genes of specific BGCs were used (i. e. backbone gene for apicidin/HC-toxin-like biosynthetic gene cluster [CMUNRPS4] and backbone gene for alternapyrone-like biosynthetic gene cluster [CMUPKS10]). BLAST-collected amino acid sequences for CMUNRPS4 and CMUPKS10 backbone genes, in addition to FGENESH-annotated entries, were subjected to alignment reliability analyses using GUIDANCE 2.0 (alignments are shown in Supplementary Data Sheet 1 and Supplementary Data Sheet 2), with PRANK (LÖYTYNOJA, 2014) for sequence alignment with 100 bootstrap replicates and variable gap penalties and a GUIDANCE2 score cutoff of 0.80 for site removal. The best-fit evolutionary model was estimated using Prottest 3.4 (DARRIBA et al., 2011) (Supplementary Table 2), and phylogenetic reconstruction was conducted using two different methods: Maximum Likelihood using PhyML 3.1 (GUINDON et al., 2010) with aLRT SH-like (approximate likelihood ratio test Shimodaira–Hasegawa) branch support estimation (ANISIMOVA et al., 2011; GASCUE, 2006) and Bayesian Inference using MrBayes (ALTEKAR et al., 2004; AYRES et al., 2012; RONQUIST et al., 2012). The latter method was executed for up to 10^7 Markov Chain Monte Carlo (MCMC) generations (sampled every 100 steps), applying an average standard deviation of split frequencies < 0.01 , calculated every thousandth generation, as the main convergence criterion. The first 25% of samples,

corresponding to the initial stages in MCMC sampling where likelihood values increase rapidly, were discarded as burn-in, summarizing parameters and trees afterwards.

Moreover, a species tree based on concatenated sequences of nuclear ribosomal internal transcribed spacer (*ITS*), partial glyceraldehyde-3-phosphate dehydrogenase gene (*GAPDH*), partial calmodulin gene (*CAL*), partial actin gene (*ACT*), partial beta-tubulin gene (*TUB*), partial chitin synthase 1 gene (*CHS-1*), partial manganese superoxide dismutase gene (*SOD*) and partial glutamine synthetase gene (*GS*) sequences was constructed to clarify the relationships between the *Colletotrichum* species used in comparative genomic analysis. Initially, the nucleotide sequences for each DNA barcode were subjected to alignment reliability analyses using GUIDANCE 2.0, with MAFFT (KATO; ROZEWICKI; YAMADA, 2017) for sequence alignment with 100 bootstrap replicates and variable gap penalties and a GUIDANCE2 score cutoff of 0.93 for site removal, and GBLOCKS (CASTRESANA, 2000) were evaluated. Alignments were concatenated with Sequencematrix (SCHEFFER; LEWIS; JOSHI, 2006). The best-fit evolutionary model was assessed using MrModeltest 2.3 (NYLANDER, 2004) (Supplementary Table 2), and a phylogenetic tree was constructed inferred with MrBayes, as described above. Additionally, information about host range accordingly to Crouch and coworkers (2014) was amended to the phylogenetic tree (Figure 1). (Alignment is shown in Supplementary Data Sheet 3).

In a addition we performed the average nucleotide identity using both best hits (one-way ANI) and reciprocal best hits (two-way ANI) between groups genomic datasets of *Colletotrichum* species, as calculated by Goris et al (2007). Typically, the ANI values between genomes of the same species are above 95%. Values below 75% are not to be trusted, and AAI should be used instead. This tool supports both complete and draft genomes (multi-fasta). (Figure 2).

RESULTS AND DISCUSSION

The survey in *C. musae* GM20 genome identified 72 putative backbone genes/BGCs (Table 2), of which nine are NRPS, twenty-six are PKS (Type 1), one PKS (Type 1 and 3), one PKS (Type 3), ten TERPs, fifteen INDs and nine OTHERs. Noteworthy, several BGCs that had been found in *C. musae* GM20 are presented in *C. gloeosporioides* strain Nara gc5 genome (70 conserved gene cluster [cgc]) (Table 2). Accordingly, *C. gloeosporioides* is the *Colletotrichum* species is the closest related species with available genome analyzed (Figure 1). Additionally, usually more than half of these BGCs are conserved in other *Colletotrichum* species: 48 cgc in *C. fioriniae* strain PJ7 genome, 48 cgc in *C. nymphaeae* strain SA-01

genome, 42 cgc in *C. chlorophyti* strain NTL11 genome, 49 cgc in *C. orbiculare* strain MAFF 240422 genome, 47 cgc in *C. incanum* strain MAFF 238704 genome, 50 cgc in *C. salicis* strain CBS 607.94 genome, 50 cgc in *C. simmondsii* strain CBS122122 genome, 43 cgc in *C. sublineola* strain TX430BB genome, 48 cgc in *C. acutatum* strain 1 genome, 44 cgc in *C. acutatum* strain C71 genome, 48 cgc in *C. higginsianum* strain IMI 349063 genome, 33 cgc in *C. godetiae* strain C184 genome, 37 cgc in *C. falcatum* strain Cf671 genome, 47 cgc in *C. tofieldiae* strain CBS127615 genome, 42 cgc in *C. orchidophilum* strain IMI 309357 genome, 46 cgc in *C. graminicola* strain M1.001 genome, 44 cgc in *C. lindemuthianum* strain 83.501 genome, 44 cgc in *C. lindemuthianum* strain 89 genome. Moreover, only 19 BGCs are presented in *V. dahliae* strain VdLs.17 genome (Table 2). In general the results shown that *C. musae* GM20 harbor a great potential for SM production. This SM potential is partially shared with *Colletotrichum* species with diverse host range.

Comparative genomic analysis was employed to clarify the predictable final products of the assigned BGCs. These comparisons revealed certain interesting gene clusters, which are listed in Table 2. CMUNRPS1 and CMUNRPS2 are putatively enrolled in the biosynthesis of the siderophores metachelin and ferricrocin, respectively. Siderophores are vital metabolites for iron uptake and are described as indirect virulence factors. CMUNRPS4 was assigned as apicidin-like/HC-toxin-like gene cluster. CMUNRPS4 is related to the apicidin biosynthetic gene cluster from *Fusarium incarnatum* (45–58 % identity in protein-by-protein comparisons of constituent genes of these BGCs) and HC-toxin biosynthetic gene cluster from *Bipolaris zeicola* (52–54 % identity in protein-by-protein comparisons of constituent genes of these BGCs) (Figure 2). Both compounds are described as histone deacetylase inhibitors. Furthermore, CMUNRPS4 backbone gene orthologous are usually presented in phytopathogenic fungi (Figure 2a). CMUPKS1 was assigned for production of melanin. Melanin is an important metabolite for virulence and environmental maintenance in several fungal species. Another interesting BGCs is CMUPKS10 linked with production of a putative alternapyrone-like compound. Some genes are conserved between CMUPKS10 gene cluster and the alternapyrone biosynthetic gene cluster from *Alternaria solani* (49-53 % identity in protein-by-protein comparisons of constituent genes of these BGCs) and the backbone gene are phylogenetically related (Figure 3). *A. solani* is a destructive pathogen and alternapyrone/alternapyrone-like gene clusters are only conserved in plant-related fungi (Figure 3a). This result suggest a putative importance as virulence factor for alternapyrone/alternapyrone-like molecules. Other BGCs from *C. musae* GM20 with orthologous characterized in other fungal species were: CMUPKS4 linked with production of

brefeldin/fumagillin-like compound; CMUPKS9 linked with production of monacolin J-like compound; CMUPKS11, which backbone gene is similar to *azaB*; CMUTERP5 linked with production of botrydial-like compound and CMUTERP9 linked with production of PR-toxin-like compound (Table 2). In short, almost 15 % of the identified BGCs from *C. musae* GM20 shown characterized orthologous in other fungal species (Table 2).

Furthermore, once 70 (97.22 %) of the BGCs that had been found in *C. musae* GM20 are presented in *C. gloeosporioides* strain Nara gc5 genome, the conservation of *C. gloeosporioides* BGCs was also evaluated in *C. musae* GM20 and the already cited *Colletotrichum* and *Verticillium* species. Notably, 25 BGCs identified in *C. gloeosporioides* strain Nara gc5 are absent in *C. musae* GM20 (Table 3). Additionally, the conservation of these 25 BGCs among the *Colletotrichum* species is sparse (Table 3). These almost exclusive BGCs from *C. gloeosporioides* may this fungus to infect a broad host range.

CONCLUSION

In recent years, SM have been explored in several fungal species. The SM potential of pathogenic species have been a main focus, once SM can act as virulence determinants. Notably, conservation and diversity of BGCs have been suggested as a factor for broad host range. The genome of specialist pathogen *C. musae* GM20 shown several BGCs and some putative gene clusters can be suggested as virulence determinants, as the apicidin-like/HC-toxin-like gene cluster and alternapyrone-like gene cluster. Notably, the SM potential of *C. musae* GM20 seems strictly similar to *C. gloeosporioides* strain Nara gc5, although several BGCs presented in *C. gloeosporioides* strain Nara gc5 genome are not conserved in *C. musae* GM20 genome. Thus, those BGCs not conserved in *C. musae* GM20, but presented in *C. gloeosporioides* strain Nara gc5, can be essential to attain a successful infection in non-banana hosts.

REFERENCES

- Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics*. 20:407–415. doi: 10.1093/bioinformatics/btg427.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–10. doi: 10.1016/S0022-2836(05)80360-2.
- Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst. Biol.* 60:685–699. doi: 10.1093/sysbio/syr041.
- Ayres DL et al. 2012. BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics. *Syst. Biol.* 61:170–173. doi: 10.1093/sysbio/syr100.
- Cannon PF, Damm U, Johnston PR, Weir BS. 2012. *Colletotrichum* - current status and future directions. *Stud. Mycol.* 73:181–213. doi: 10.3114/sim0014.
- Castresana J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. 540–552.
- Crouch J et al. 2014. The genomics of *Colletotrichum*. In: *Genomics of Plant-Associated Fungi: Monocot Pathogens*. pp. 69–102. doi: 10.1007/978-3-662-44053-7_3.
- Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest-HPC: Fast selection of best-fit models of protein evolution. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*. 6586 LNCS:177–184. doi: 10.1007/978-3-642-21878-1_22.
- Dean RA, Lichens-Park A, Kole C. 2014. Genomics of plant-associated fungi: Monocot pathogens. doi: 10.1007/978-3-662-44053-7.
- GASCUE MAAO. 2006. Approximate Likelihood-Ratio Test for Branches : A Fast , Accurate ., 55:539–552. doi: 10.1080/10635150600755453.
- Goris J et al. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57:81–91. doi: 10.1099/ijms.0.64483-0.
- Guindon S et al. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 2.0. *Syst. Biol.* 59:307–321. doi: 10.1093/sysbio/syq010.
- Hyde KD et al. 2009. *Colletotrichum* – names in current use. *Fungal Divers.* - online Adv. 39:147–182. <http://www.scopus.com/inward/record.url?eid=2-s2.0-84920491130&partnerID=tZOtx3y1>.
- Katoh K, Rozewicki J, Yamada KD. 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 1–7. doi: 10.1093/bib/bbx108.

- Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* 1079:155–170. doi: 10.1007/978-1-62703-646-7_10.
- Medema MH, Takano E, Breitling R. 2013. Detecting sequence homology at the gene cluster level with multigeneblast. *Mol. Biol. Evol.* 30:1218–1223. doi: 10.1093/molbev/mst025.
- Nylander JAA. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University. citeulike-article-id:7578084.
- Ronquist F et al. 2012. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61:539–542. doi: 10.1093/sysbio/sys029.
- Scheffer SJ, Lewis ML, Joshi RC. 2006. DNA Barcoding Applied to Invasive Leafminers (Diptera: Agromyzidae) in the Philippines. *Ann. Entomol. Soc. Am.* 99:204–210. doi: 10.1603/0013-8746(2006)099[0204:DBATIL]2.0.CO;2.
- De Silva DD, Crous PW, Ades PK, Hyde KD, Taylor PWJ. 2017. Life styles of *Colletotrichum* species and implications for plant biosecurity. *Fungal Biol. Rev.* 31:155–168. doi: 10.1016/j.fbr.2017.05.001.
- Silva Junior WJ da et al. 2018. Draft genome assembly of *Colletotrichum musae*, the pathogen of banana fruit. *Data Br.* 17:256–260. doi: 10.1016/j.dib.2018.01.002.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 31:3210–3212. doi: 10.1093/bioinformatics/btv351.
- Weber T et al. 2015. AntiSMASH 3.0-A comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 43:W237–W243. doi: 10.1093/nar/gkv437.
- Wolf T, Shelest V, Nath N, Shelest E. 2016. CASSIS and SMIPS: Promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes. *Bioinformatics.* 32:1138–1143. doi: 10.1093/bioinformatics/btv713.

FIGURES, TABLES AND SUPPLEMENTARY DATA

Table 1: Comparison of the primary genome features between *Colletotrichum* species

Table 2: *Colletotrichum musae*' BGCs conservation in the *Colletotrichum* genus.

Table 3: *Colletotrichum gloeosporioides*' BGCs conservation in the *Colletotrichum* genus.

Figure 1: Bayesian tree based on concatenated nucleotide sequences of nuclear ribosomal internal transcribed spacer (*ITS*), partial glyceraldehyde-3-phosphate dehydrogenase gene (*GAPDH*), partial calmodulin gene (*CAL*), partial actin gene (*ACT*), partial beta-tubulin gene (*TUB*), partial chitin synthase 1 gene (*CHS-1*), partial manganese superoxide dismutase gene (*SOD*) and partial glutamine synthetase gene (*GS*). This tree was constructed to clarify the relationships between the *Colletotrichum* species used in a comparative genomic analysis. *Colletotrichum boninense* strain CBS 128547 was used as outgroup and the resulting tree was rooted at this species (outgroup). The species were classified according number of susceptible hosts. The branch support values (Bayesian posterior probability) are associated with nodes. The Bayesian inference ran for 10^7 generations.

Figure 2: ANY analysis between all *Colletotrichum* species founded at GBANK until 02/19/2018. The ANI calculator estimates the average nucleotide identity using both best hits (one-way ANI) Values below 75% are not to be trusted, and AAI should be used instead. This tool supports both complete and draft genomes (multi-fasta).

Figure 3: Apicidin/HC-toxin-like biosynthetic BGC (CMUNRPS4) phylogeny, conservation and synteny. (A) Phylogenetic analyses were performed using Maximum-likelihood and Bayesian Inference, based on apicidin/HC-toxin-like backbone gene and orthologous sequences exhibited by several fungi. The orthologous sequences were classified according to fungal lifestyle trait, represented by different colors. The Maximum-likelihood is displayed and branch support values (aLRT SH-like supports and Bayesian posterior probability) are associated with nodes. The Bayesian inference ran for 100,000 generations. *C. musae* strain GM20, *Fusarium fujikuroi* strain IMI 58289 (apicidin producer strain), *F. incarnatum* (apicidin producer strain), *B. zeicola* (HC-toxin producer strain) and *Cochliobolus carbonum* (HC-toxin producer strain) are highlighted in bold. (B) Some genes from *C. musae*

CMUNURPS4 resemble the characterized apicidin BGC from *F. incarnatum* (45–58 % identity) and HC-toxin BGC from *B. zeicola* (52–54 % identity). Orthologous genes were assigned with the same color; white boxes represent genes that were not predicted to be part of *C. musae* GM20 gene cluster.

Figure 4: Alternapyrone-like biosynthetic BGC (CMUPKS10) phylogeny, conservation and synteny. (A) Phylogenetic analyses were performed using Maximum-likelihood and Bayesian Inference, based on alternapyrone-like backbone gene and orthologous sequences exhibited by several fungi. The orthologous sequences were classified according to fungal lifestyle trait, represented by different colors. The Maximum-likelihood is displayed and branch support values (aLRT SH-like supports and Bayesian posterior probability) are associated with nodes. The Bayesian inference ran for 100,000 generations. *C. musae* strain GM20 and *Alternaria solani* (alternapyrone producer strain) are highlighted in bold. (B) Some genes from *C. musae* CMUPKS10 resemble the characterized alternapyrone BGC from *A. solani* (49-53 % identity) as well as putative BGCs from other fungal species. Orthologous genes were assigned with the same color; white boxes represent genes that were not predicted to be part of *C. musae* GM20 gene cluster.

Supplementary Table 1: Fungal genomes used in this work.

Supplementary Data Sheet 1: Amino acid alignment for the apicidin/HC-toxin-like backbone gene and related orthologs.

Supplementary Data Sheet 2: Amino acid alignment for the alternapyrone-like backbone gene and related orthologs.

Supplementary Data Sheet 3: Concatenated nucleotide sequences of nuclear ribosomal internal transcribed spacer (ITS), partial glyceraldehyde-3-phosphate dehydrogenase gene (GAPDH), partial calmodulin gene (CAL), partial actin gene (ACT), partial beta-tubulin gene (TUB), partial chitin synthase 1 gene (CHS-1), partial manganese superoxide dismutase gene (SOD) and partial glutamine synthetase gene (GS).

Supplementary Table 2: Best-fit evolutionary models predicted with Prottest 3.4 or jModeltest-2.1.9 for each alignment.

Supplementary Table 3: Best-fit evolutionary models predicted with MrModeltest-2.3 for each alignment

Supplementary Table 4 - Sequences of Colletotrichum isolates used in this article

Table 1

Species/Strain	Assembly size	Coverage	Sequencing technology	scaffolds	N50 scaffolds length	Contigs	Genes	Annotation strategy	GC content	Reference	Public access to genome
<i>Colletotrichum musae</i> GM20	50.7 Mb	100x	Illumina	2.763	32.818	10.618	13.451	Funannotate	53.74	Present study	NWMS01000000.1
<i>Colletotrichum gloeosporioides</i> Nara ge5	55.6 Mb	37x	Illumina	1.241	112.809	5.335	15.381	-	53.52	-	ANPB00000000.1
<i>Colletotrichum fioriniae</i> PJ7	49.01 Mb	48.72x	Illumina	1.096	137.254	1.096	13.759	CEGMA	52.50	[1]	JARH00000000.1
<i>Colletotrichum nymphaeae</i> SA-01	49.96 Mb	46.94x	Illumina	1.884	91.051	2.247	14.404	-	52.77	-	JEMN00000000.1
<i>Colletotrichum chlorophyti</i> NTL11	52.38 Mb	77x	Illumina	558	644.295	3.613	10.419	-	50.18	[2]	MPGH00000000.1
<i>Colletotrichum orbiculare</i> MAFF 240422	90.08 Mb	55x	Illumina/454	3.37	423.81	8.092	13.347	-	37.67	-	AMCV00000000.1
<i>Colletotrichum incanum</i> MAFF 238704	53.15 Mb	153x	Illumina	1.036	292.512	6.302	11.852	MAKER	52.15	[3]	JTLR00000000.1
<i>Colletotrichum salicis</i> CBS 607.94	48.37 Mb	40.4x	Illumina	2.776	46.166	5.121	13.783	-	52.91	-	JFFI00000000.1
<i>Colletotrichum simmondsii</i> CBS122122	50.47 Mb	39.38x	Illumina	929	292.136	3.860	13.884	-	51.99	-	JFBX00000000.1
<i>Colletotrichum sublineola</i> TX430BB	46.75 Mb	92x	Illumina	1.625	70.717	6.383	12.699	CEGMA	52.70	[4]	JMSE00000000.1
<i>Colletotrichum acutatum</i> 1	52.12 Mb	20x	PacBio/Illumina	27	4,416.304	34	13.559	GeneMark-ES/EvidenceModeler	51.73	[5]	LUXP00000000.1
<i>Colletotrichum acutatum</i> C71	44.92 Mb	47x	Illumina	2.614	37.362	9.732	-	-	52.12	-	LVCK00000000.1
<i>Colletotrichum higginsianum</i> IMI 349063	57.44 Mb	76x	454/Sanger/Illumina	10.235	6.147	10.235	14.650	Broad pipeline/EvidenceModeler	54.41		CACQ00000000.2
<i>Colletotrichum godetiae</i> C184	35.03 Mb	83x	Illumina	1.819	48.837	7.086	-	-	52.60	-	LZRM00000000.1
<i>Colletotrichum falcatum</i> Cf671	48.18 Mb	150x	Illumina	4.358	29.329	6.400	12.270	AUGUSTUS/FGENESH/GeneMarkS	51.65	[6]	LPVI00000000.1
<i>Colletotrichum tofieldiae</i> CBS127615	52.72 Mb	273x	Illumina	791	377.287	794	13.134	CEGMA	52.93	[7]	LFHR00000000.1
<i>Colletotrichum orchidophilum</i> IMI 309357	48.55 Mb	200x	Illumina	321	284.473	328	14.496	-	51.07	-	MJBS00000000.1
<i>Colletotrichum graminicola</i> M1.001	50.9 Mb	9.09x	454/Sanger	654	579.194	1.152	12.020	Broad pipeline	49.11	-	ACOD00000000.1

<i>Colletotrichum lindemuthianum</i> 83.501	97.41 Mb	89.3x	Illumina	1.857	111.275	2.282	11.67 3	-	37.56	[8]	MASO0000000.2
<i>Colletotrichum lindemuthianum</i> 89	99.16 Mb	157.7x	Illumina	1.276	158.217	1.767	11.62 7	-	37.32	[8]	MASP0000000.2

1. Baroncelli R, Sreenivasaprasad S, Sukno SA, Thon MR, Holub E: **Draft Genome Sequence of *Colletotrichum acutatum* Sensu Lato (*Colletotrichum fioriniae*)**. *Genome Announcements* 2014, **2**(2):e00112-00114.
2. Gan P, Narusaka M, Tsushima A, Narusaka Y, Takano Y, Shirasu K: **Draft Genome Assembly of *Colletotrichum chlorophyti*, a Pathogen of Herbaceous Plants**. *Genome Announcements* 2017, **5**(10):e01733-01716.
3. Gan P, Narusaka M, Kumakura N, Tsushima A, Takano Y, Narusaka Y, Shirasu K: **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* 2016, **8**(5):1467-1481.
4. Baroncelli R, Sanz-Martín JM, Rech GE, Sukno SA, Thon MR: **Draft Genome Sequence of *Colletotrichum sublineola*, a Destructive Pathogen of Cultivated Sorghum**. *Genome Announcements* 2014, **2**(3):e00540-00514.
5. Han J-H, Chon J-K, Ahn J-H, Choi I-Y, Lee Y-H, Kim KS: **Whole genome sequence and genome annotation of *Colletotrichum acutatum*, causal agent of anthracnose in pepper plants in South Korea**. *Genomics Data* 2016, **8**:45-46.
6. Viswanathan R, Prasanth CN, Malathi P, Sundar AR: **Draft Genome Sequence of *Colletotrichum falcatum* - A Prelude on Screening of Red Rot Pathogen in Sugarcane**. *Journal of Genomics* 2016, **4**:1-3.
7. Hacquard S, Kracher B, Hiruma K, Munch PC, Garrido-Oter R, Thon MR, Weimann A, Damm U, Dallery JF, Hainaut M *et al*: **Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi**. *Nature Communications* 2016, **7**:12.
8. de Queiroz CB, Correia HLN, Menicucci RP, Vidigal PMP, de Queiroz MV: **Draft Genome Sequences of Two Isolates of *Colletotrichum lindemuthianum*, the Causal Agent of Anthracnose in Common Beans**. *Genome Announcements* 2017, **5**(18):e00214-00217.

Cluster/Backbone	CGL	CFI	CNY	CCH	COR	CIN	CSA	CSI	CSU	CACI	CAC71	CHI	CGO	CFA	CTO	CRC	CGR	CL183	CL189	VDA	Locus Tag	Putative product	Backbone gene		
CMUNRPS1																						>FUN_006814	Metachelin	A-PCP-C-PCP-C	
CMUNRPS2																							>FUN_003228	Ferrirococin	A-PCP-C-A-ACP-C-PCP-C-A-PCP-C
CMUNRPS3																							>FUN_012931		PCP-E-C-A-PCP-C
CMUNRPS4																							>FUN_010952	Apicidin-like	A-PCP-E-C-A-PCP-C-A-PCP-C
CMUNRPS5																							>FUN_002739		A-PCP-C-A-TD
CMUNRPS6																							>FUN_001917		KR-C-A-PCP-TD
CMUNRPS7																							>FUN_012932		A-C-A-PCP-C
CMUNRPS8																							>FUN_012933		A-PCP-C-A-TD
CMUNRPS9																							>FUN_012934		C-A-PP-E-C-A-PCP-C
CMUPKS1 (Type 1)																							>FUN_005730	Melanin	KS-AT-DH-ACP-ACP-TE
CMUPKS2 (Type 1)																							>FUN_011329		AT-KS-ACP
CMUPKS3 (Type 1)																							>FUN_002787		KS-AT-DH-ER-KR-ACP
CMUPKS4 (Type 1)																							>FUN_003642	Brefeldin/Fumagillin-like	KS-AT-DH
CMUPKS5 (Type 1)																							>FUN_000896		KS-AT-DH-MT-ER-KR-ACP
CMUPKS6 (Type 1)																							>FUN_000875		KS-AT-ACP-TE
CMUPKS7 (Type 1)																							>FUN_012935/ >FUN_008048		KS-AT-DH-MT-ER-KR/
CMUPKS8 (Type 1)																							>FUN_006744		KS-AT-DH-KR-ACP
CMUPKS9 (Type 1)																							>FUN_012936	Monacolin J-like	KS-AT-DH-MT-ER-KR
CMUPKS10 (Type 1)																							>FUN_012937	Alternapyrone-like	KS-AT-DH-MT-ER-KR-ACP
CMUPKS11 (Type 1)																							>FUN_012481	azaB-like	KS-AT-DH-MT-ER-KR-ACP
CMUPKS12 (Type 1)																							>FUN_004772		KS-AT-DH-ER-KR
CMUPKS13 (Type 1)																							>FUN_012938		AT-ER-KR
CMUPKS14 (Type 1)																							>FUN_012939		KS-AT-DH-ACP-MT
CMUPKS15 (Type 1)																							>FUN_009066		KS-AT-DH-MT-ER-KR-ACP
CMUPKS16 (Type 1)																							>FUN_012597		KS-AT-DH-MT-ER-KR-ACP
CMUPKS17 (Type 1)																							>FUN_012940		KS-AT-DH-MT-ER-KR
CMUPKS18 (Type 1)																							>FUN_004704		KS-AT-DH-KR-ACP
CMUPKS19 (Type 1)																							>FUN_011413		KS-AT-DH-MT

CMUPKS20 (Type 1)																				>FUN_012941		KS-AT-DH-MT-EP-KR
CMUPKS21 (Type 1)																				>FUN_012942		KS-AT-DH-MT-EP-KR-ACP
CMUPKS22 (Type 1)																				>FUN_012943		KS-AT-DH-MT-EP-KR-ACP
CMUPKS23 (Type 1)																				>FUN_012944		KS-AT-DH-KR
CMUPKS24 (Type 1)																				>FUN_012945		KS-AT-DH-MT-EP-ACP
CMUPKS25 (Type 1)																				>FUN_012946		KS-AT-DH
CMUPKS26 (Type 1)																				>FUN_012947		KS-ACP-MT
CMUPKS27 (Type 1 and 3)																				>FUN_000009/ >FUN_000010		KS-DH-MT-ER-KR-ACP
CMUPKS28 (Type 2)																				>FUN_007035		
CMUTERP1																				>FUN_011316		
CMUTERP2																				>FUN_009477		
CMUTERP3																				>FUN_004923		
CMUTERP4																				>FUN_013402		
CMUTERP5																				>FUN_009953	Botrydial-like	
CMUTERP6																				>FUN_002452		
CMUTERP7																				>FUN_009343		
CMUTERP8																				>FUN_002317		
CMUTERP9																				>FUN_012474	PR toxin-like	
CMUTERP10																				>FUN_012948		
CMUIND1																				>FUN_009575		
CMUIND2																				>FUN_012468		
CMUIND3																				>FUN_001353		
CMUIND4																				>FUN_001353		
CMUIND5																				>FUN_006232		
CMUIND6																				>FUN_003092		
CMUIND7																				>FUN_002984		
CMUIND8																				>FUN_008196		
CMUIND9																				>FUN_012949		
CMUIND10																				>FUN_012324		
CMUIND11																				>FUN_003683/ >FUN_003685		
CMUIND12																				>FUN_008004		
CMUIND13																				>FUN_004567		

CMUIND14																					>FUN_012593			
CMUIND15																						>FUN_001056		
CMUOTHER1																						>FUN_004970		
CMUOTHER2																						>FUN_010419		
CMUOTHER3																						>FUN_005711		
CMUOTHER4																						>FUN_004993		
CMUOTHER5																						>FUN_002393		
CMUOTHER6																						>FUN_008059		
CMUOTHER7																						>FUN_002558		
CMUOTHER8																						>FUN_002579		
CMUOTHER9																						>FUN_009466		
CMUOTHER10																						>FUN_012564		
TOTAL=72	70	48	48	42	49	47	50	50	43	48	44	48	33	37	47	42	46	44	44	19				

Table 2: *Colletotrichum musae*' BGCs conservation in the *Colletotrichum* genus. *Colletotrichum musae* strain GM20 (CMU), *Colletotrichum gloeosporioides* strain Nara gc5 (CGL), *Colletotrichum fioriniae* strain PJ7 (CFI), *Colletotrichum nymphaeae* strain SA-01 (CNY), *Colletotrichum chlorophyti* strain NTL11 (CCH), *Colletotrichum orbiculare* strain MAFF 240422 (COR), *Colletotrichum incanum* strain MAFF 238704 (CIN), *Colletotrichum salicis* strain CBS 607.94 (CSA), *Colletotrichum simmondsii* strain CBS122122 (CSI), *Colletotrichum sublineola* strain TX430BB (CSU), *Colletotrichum acutatum* strain 1 (CAC1), *Colletotrichum acutatum* strain C71 (CAC71), *Colletotrichum higginsianum* strain IMI 349063 (CHI), *Colletotrichum godetiae* strain C184 (CGO), *Colletotrichum falcatum* strain Cf671 (CFA), *Colletotrichum tofieldiae* strain CBS127615 (CTO), *Colletotrichum orchidophilum* strain IMI 309357 (CRC), *Colletotrichum graminicola* strain M1.001 (CGR), *Colletotrichum lindemuthianum* strain 83.501 (CLI83), *Colletotrichum lindemuthianum* strain 89 (CLI89), *Verticillium dahliae* strain VdLs.17 (VDA). Domains for PKS and NRPS backbone gene are listed. Adenylation (A), Condensation (C), Peptidyl Carrier Protein (PCP), Keto-synthase (KS), Acyltransferase (AT), Acyl Carrier Protein (ACP), Dehydrogenase (DH), Enoylreductase (ER), Ketoreductase (KR), Methyltransferase O- or C- (MT), Thioesterase (TE), Thioester reductase (TD). The conservation of predicted BGCs among *Colletotrichum*

species was assessed using MultiGeneBlast 1.1.14 (analyses of genomic loci with predicted proteins) [1], based primarily on backbone gene conservation (e-value $< 1 \times 10^{-5}$, query coverage ≥ 50 %, and identity ≥ 50 %).

1. Medema MH, Takano E, Breitling R: **Detecting Sequence Homology at the Gene Cluster Level with MultiGeneBlast**. *Molecular Biology and Evolution* 2013, **30**(5):1218-1223.

Absent	CFI	CNY	CCH	COR	CIN	CSA	CSI	CSU	CAC1	CAC71	CHI	CGO	CFA	CTO	CRC	CGR	CLI83	CLI89	VDA	Locus Tag	Backbone gene
<i>CGLNRPS1</i>																				XP_007274757	A-PCP-C
<i>CGLNRPS2</i>																				XP_007280798	A-A-PCP-C
<i>CGLNRPS3</i>																				XP_007280822	A-PCP-C-A-
<i>CGLNRPS4</i>																				XP_007286740	A-PCP-C
<i>CGLPKS1</i> (Type 1)																				XP_007274878/ XP_007274880	KS-AT-DH- MT-KR-ACP/
<i>CGLPKS2</i> (Type 1)																				XP_007277375/ XP_007277378	KS-AT-DH-ER- KR-ACP/
<i>CGLPKS3</i>																				XP_007280354	KS-AT-DH-ER-
<i>CGLPKS4</i>																				XP_007287036	KS-AT-DH-
<i>CGLPKS5</i>																				XP_007283082	
<i>CGLPKS-NRPS1</i>																				XP_007275317/ XP_007275318	C-A-PP-C-A/ KS-AT-DH-
<i>CGLPKS-</i>																				XP_007275649	KS-AT-DH-
<i>CGLPKS-NRPS3</i>																				XP_007284065	KS-AT-DH- MT-KR-ACP-C-
<i>CGLPKS-NRPS4</i>																				XP_007285713/ XP_007285714	C-A-PCP-C/ KS-AT-DH-
<i>CGLPKS-</i>																				XP_007279389	KS-AT-DH-
<i>CGLIND1</i>																				XP_007278693	
<i>CGLOTHER1</i>																				XP_007281503	
TOTAL=16	2	2	2	4	1	1	2	2	1	1	6	2	3	2	0	3	1	1	0		
Possible	CFI	CNY	CCH	COR	CIN	CSA	CSI	CSU	CAC1	CAC71	CHI	CGO	CFA	CTO	CRC	CGR	CLI83	CLI89	VDA	Locus Tag	
<i>CGLNRPS5</i>																				XP_007274673	C-A-PCP-C
<i>CGLPKS5</i> (Type 1)																				XP_007280259/ XP_007280260	KS-AT-DH- ACP-TE/
<i>CGLPKS6</i>																				XP_007287364	KS-AT-DH-
<i>CGLPKS-</i>																				XP_007278547	KS-AT-DH-
<i>CGLPKS-</i>																				XP_007287453	KS-AT-DH-
<i>CGLTERP1</i>																				XP_007277705	
<i>CGLTERP2</i>																				XP_007277903	
<i>CGLTERP3</i>																				XP_007283729	
TOTAL=8	0	0	3	4	1	0	0	3	0	0	4	0	1	1	0	1	2	2	0		

Table 3: *Colletotrichum gloeosporioides*' BGCs conservation in the *Colletotrichum* genus. *Colletotrichum gloeosporioides* strain Nara gc5 (CGL), *Colletotrichum fioriniae* strain PJ7 (CFI), *Colletotrichum nymphaeae* strain SA-01 (CNY), *Colletotrichum chlorophyti* strain NTL11 (CCH), *Colletotrichum orbiculare* strain MAFF 240422 (COR), *Colletotrichum incanum* strain MAFF 238704 (CIN), *Colletotrichum salicis* strain CBS 607.94 (CSA), *Colletotrichum simmondsii* strain CBS122122 (CSI), *Colletotrichum sublineola* strain TX430BB (CSU), *Colletotrichum acutatum* strain 1 (CAC1), *Colletotrichum acutatum* strain C71 (CAC71), *Colletotrichum higginsianum* strain IMI 349063 (CHI), *Colletotrichum godetiae* strain C184 (CGO), *Colletotrichum falcatum* strain Cf671 (CFA), *Colletotrichum tofieldiae* strain CBS127615 (CTO), *Colletotrichum orchidophilum* strain IMI 309357 (CRC), *Colletotrichum graminicola* strain M1.001 (CGR), *Colletotrichum lindemuthianum* strain 83.501 (CLI83), *Colletotrichum lindemuthianum* strain 89 (CLI89), *Verticillium dahliae* strain VdLs.17 (VDA). Domains for PKS and NRPS backbone gene are listed. Adenylation (A), Condensation (C), Peptidyl Carrier Protein (PCP), Keto-synthase (KS), Acyltransferase (AT), Acyl Carrier Protein (ACP), Dehydrogenase (DH), Enoylreductase (ER), Ketoreductase (KR), Methyltransferase O- or C- (MT), Thioesterase (TE), Thioester reductase (TD). The conservation of predicted BGCs among *Colletotrichum* species was assessed using MultiGeneBlast 1.1.14 (analyses of genomic loci with predicted proteins) [1], based primarily on backbone gene conservation (e-value $< 1 \times 10^{-5}$, query coverage ≥ 50 %, and identity ≥ 50 %). *Possible truncated backbone gene.

1. Medema MH, Takano E, Breitling R: **Detecting Sequence Homology at the Gene Cluster Level with MultiGeneBlast.** *Molecular Biology and Evolution* 2013, **30**(5):1218-1223.

Figure 1.

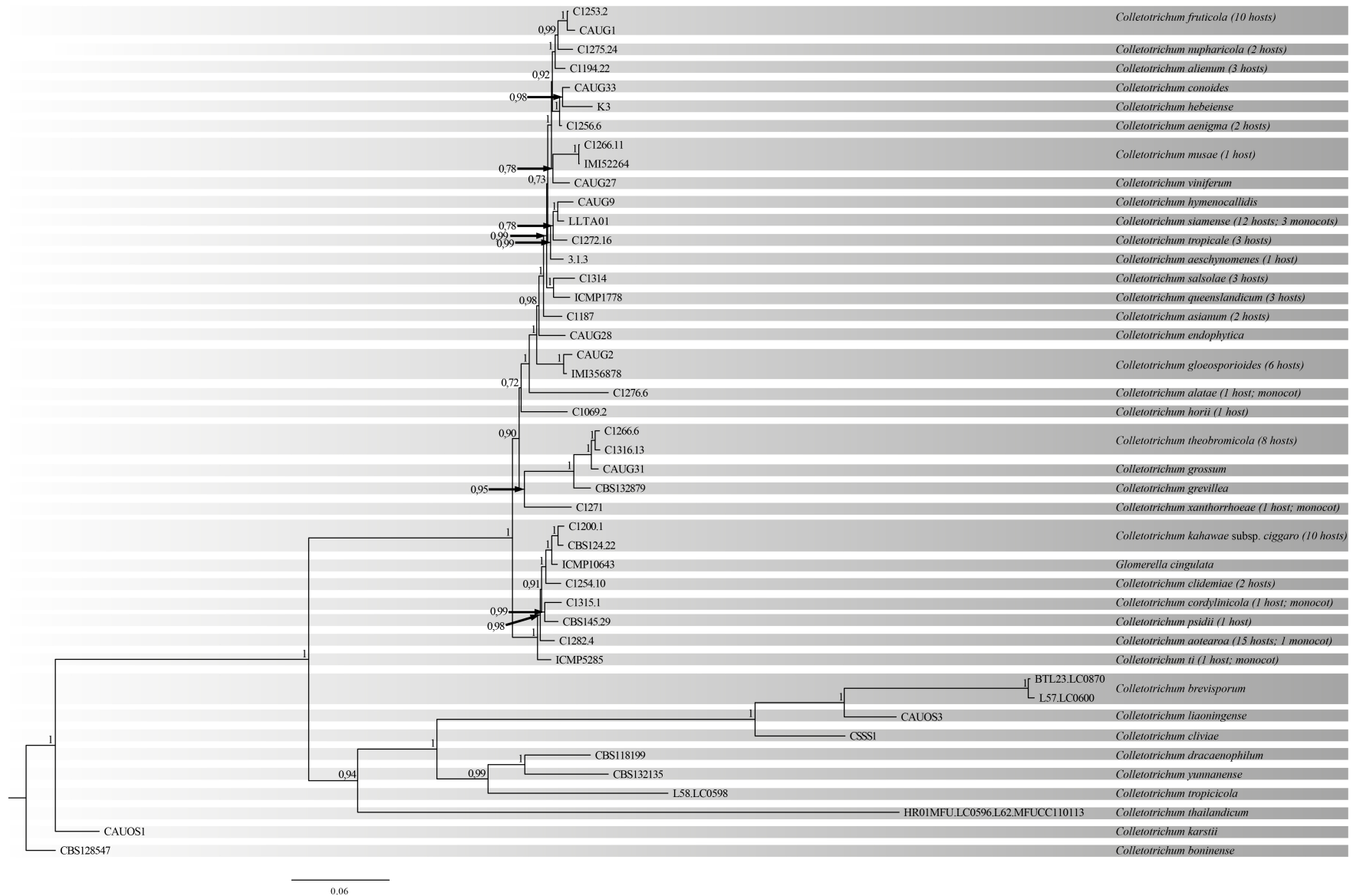


Figure 2.

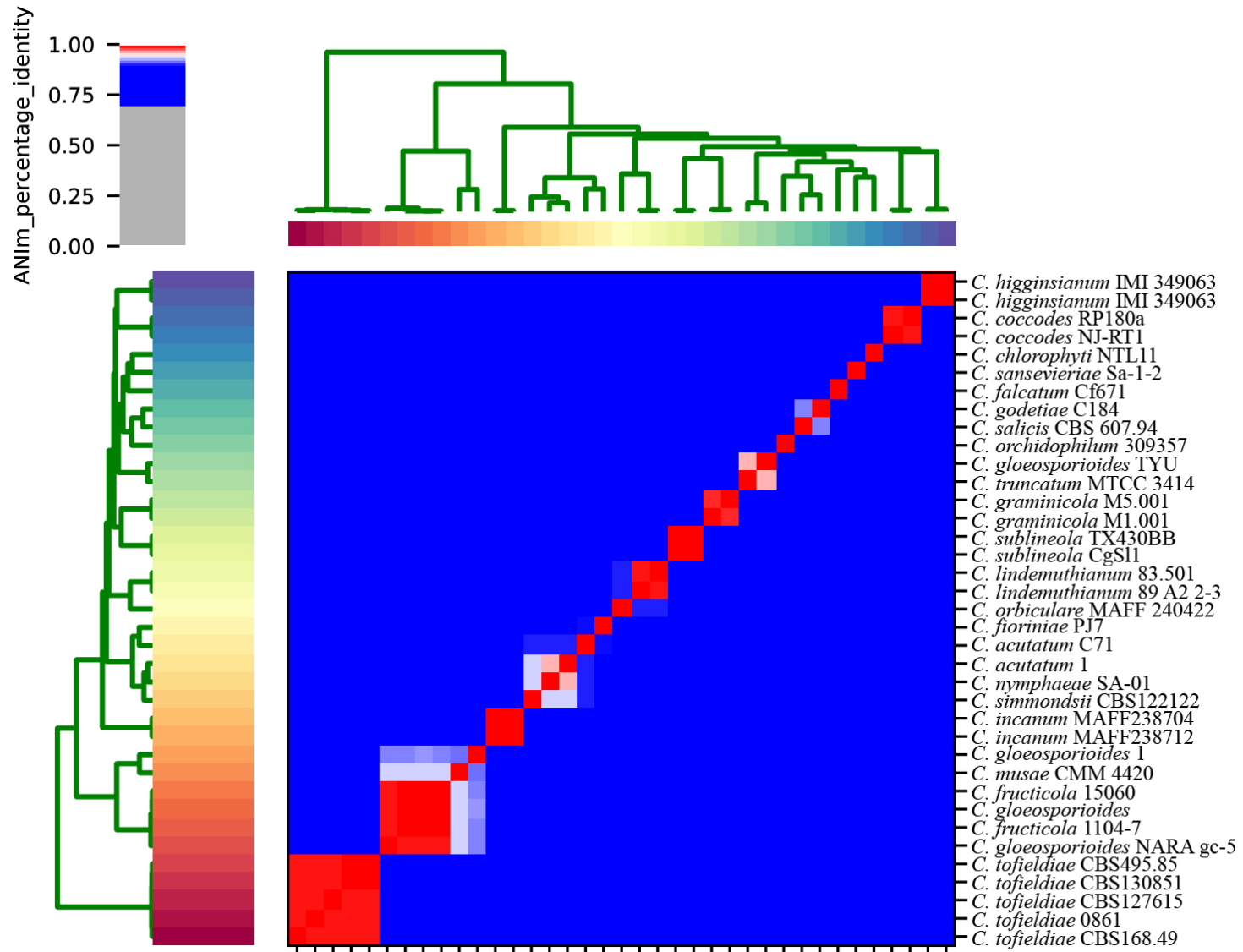
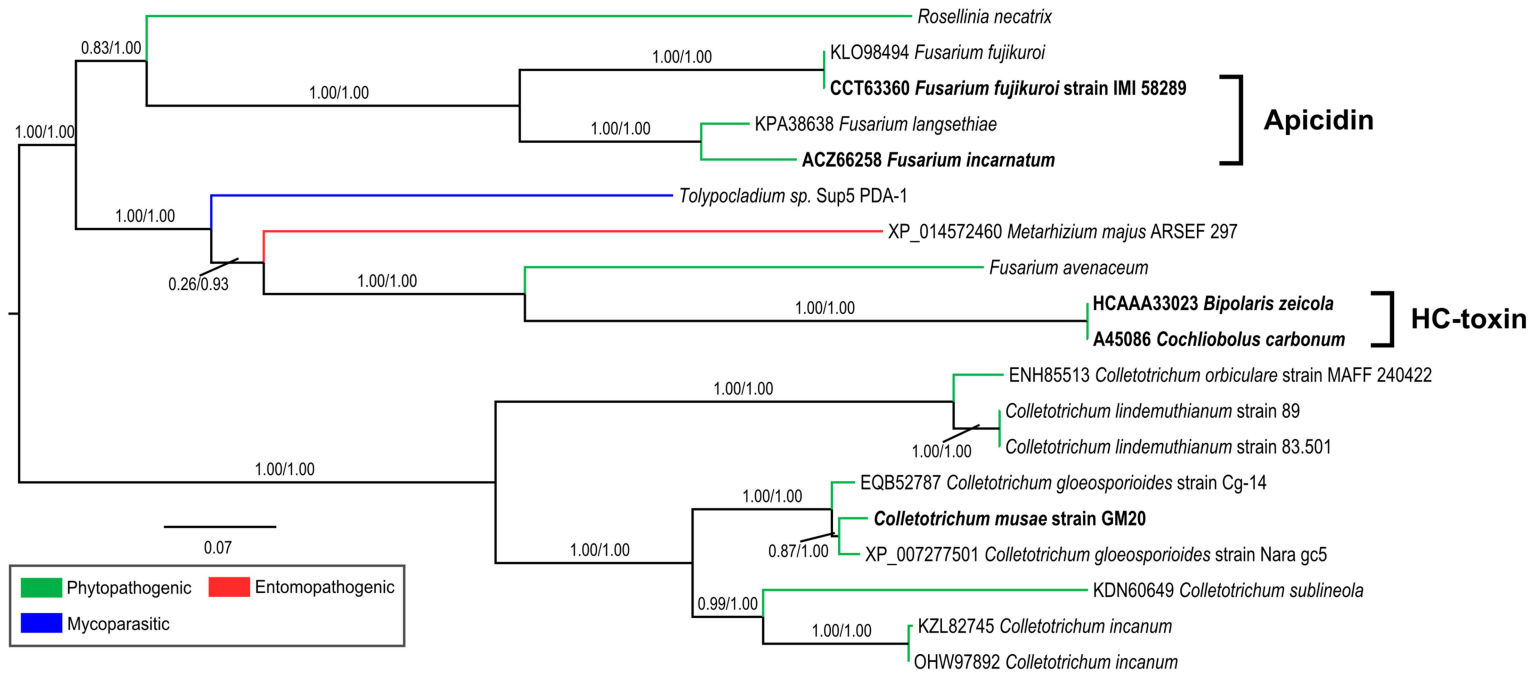


Figure 3.

A



B

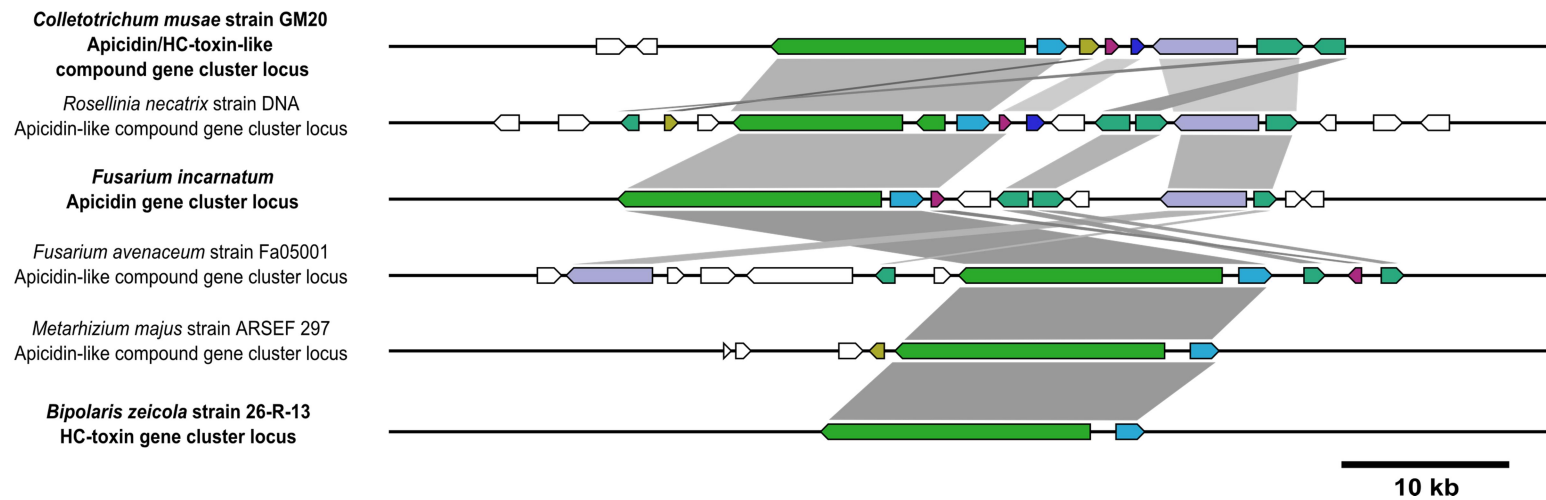
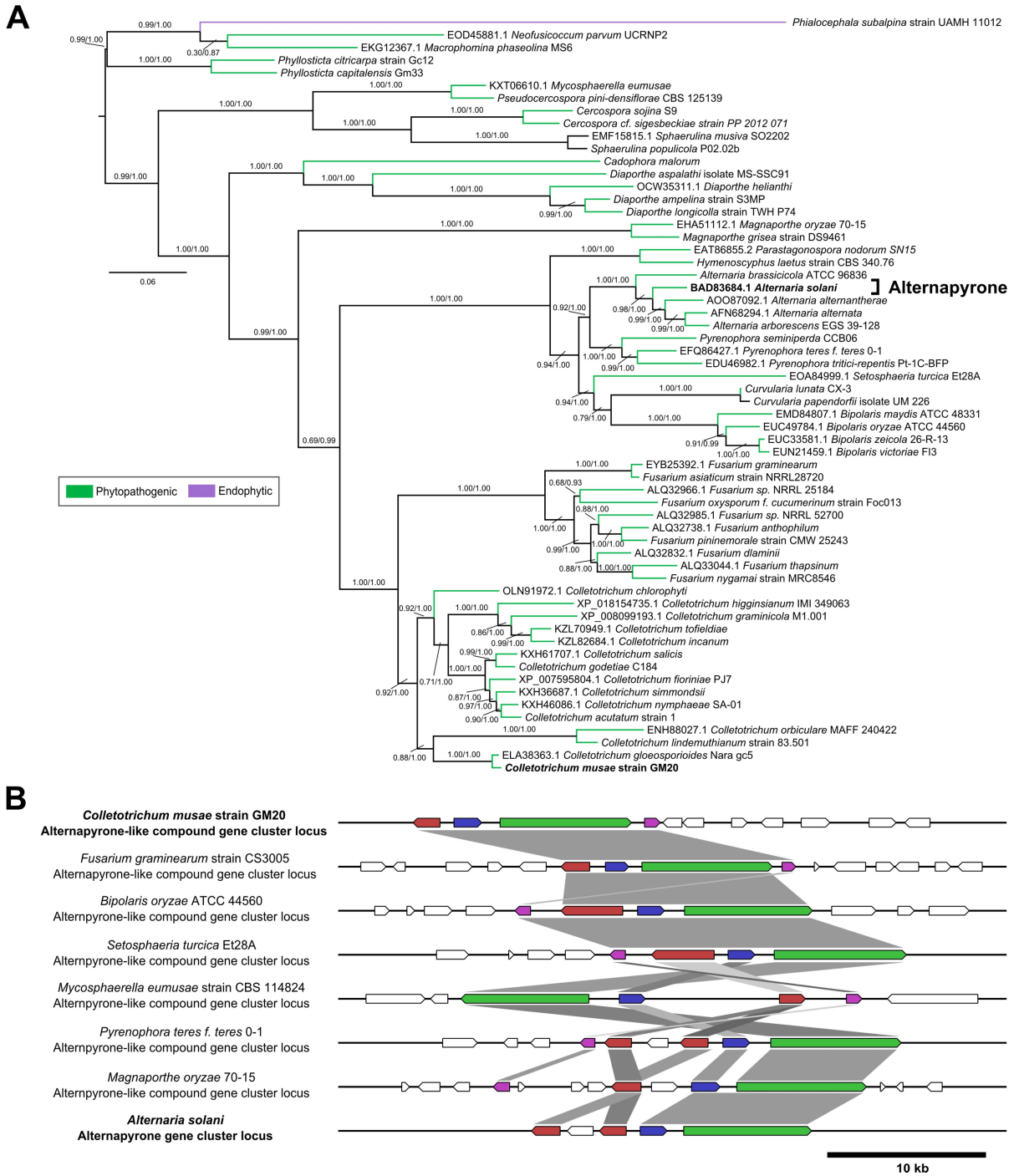


Figure 4.



Supplementary Table 1 - Fungal genomes used in this work.

Genome	NCBI's Bioproject Number	Genbank Assembly Accession Number	Nucleotide Sequence Accession Number
<i>Colletotrichum musae</i> GM20	PRJNA407219	GCA_002814275.1	NWMS00000000.1
<i>Colletotrichum gloeosporioides</i> Nara gc5	PRJNA171218	GCA_000319635.1	ANPB00000000.1
<i>Colletotrichum fioriniae</i> PJ7	PRJNA233987	GCA_000582985.1	JARH00000000.1
<i>Colletotrichum nymphaeae</i> SA-01	PRJNA237763	GCA_001563115.1	JEMN00000000.1
<i>Colletotrichum chlorophyti</i> NTL11	PRJNA350752	GCA_001937105.1	MPGH00000000.1
<i>Colletotrichum orbiculare</i> MAFF 240422	PRJNA171217	GCA_000350065.1	AMCV00000000.1
<i>Colletotrichum incanum</i> MAFF 238704	PRJNA267650	GCA_001855235.1	JTLR00000000.1
<i>Colletotrichum salicis</i> CBS 607.94	PRJNA238477	GCA_001563125.1	JFFI00000000.1
<i>Colletotrichum simmondsii</i> CBS122122	PRJNA239224	GCA_001563135.1	JFBX00000000.1
<i>Colletotrichum sublineola</i> TX430BB	PRJNA246670	GCA_000696135.1	JMSE00000000.1
<i>Colletotrichum acutatum</i> 1	PRJNA314171	GCA_001593745.1	LUXP00000000.1
<i>Colletotrichum acutatum</i> C71	PRJNA314187	GCA_001662755.1	LVCK00000000.1
<i>Colletotrichum higginsianum</i> IMI 349063	PRJNA47061	GCA_000313795.2	CACQ00000000.2

<i>Colletotrichum godetiae</i> C184	PRJNA315612	GCA_001663355.1	LZRM00000000.1
<i>Colletotrichum falcatum</i> Cf671	PRJNA272959	GCA_001484525.1	LPVI00000000.1
<i>Colletotrichum tofieldiae</i> CBS127615	PRJNA286721	GCA_001618715.1	LFHR00000000.1
<i>Colletotrichum orchidophilum</i> IMI 309357	PRJNA342923	GCA_001831195.1	MJBS00000000.1
<i>Colletotrichum graminicola</i> M1.001	PRJNA37879	GCA_000149035.1	ACOD00000000.1
<i>Colletotrichum lindemuthianum</i> 83.501	PRJNA325493	GCA_001693015.2	MASO00000000.2
<i>Colletotrichum lindemuthianum</i> 89	PRJNA325493	GCA_001693025.2	MASP00000000.2

Supplementary Table 2 - Best-fit evolutionary models predicted with Prottest 3.4 or jModeltest-2.1.9 for each alignment.

Gene/Tree	Model
CMUNRPS4	LG+I+G+F
CMUPKS10	JTT+I+G

Supplementary Table 3 - Statistical summary of the alignments followed by the substitution model used in phylogenetic analysis.

Gene	Treatment	Length	C	V	PI	PI(%)	BI
ACT	Masked	328	176	124	93	28,4	
	Trimmed	265	155	82	55	20,8	
	Gblocks	279	140	139	94	33,7	HKY+G
CAL	Masked	976	420	406	293	30,0	
	Trimmed	702	371	181	121	17,2	
	Gblocks	730	343	387	311	42,6	GTR+I+G
CHS	Guidance	301	227	72	52	17,3	
	Gblocks	299	224	75	52	17,4	K80+I+G
GAPDH	Masked	431	132	216	155	36,0	
	Trimmed	241	104	54	28	11,6	
	Gblocks	275	73	202	161	58,5	HKY+I+G
GS	Masked	1539	443	517	219	14,2	
	Trimmed	967	209	189	66	6,8	
	Gblocks	852	320	532	223	26,2	GTR+G
ITS	Masked	679	538	105	77	11,3	
	Trimmed	600	498	67	41	6,8	
	Gblocks	593	468	125	97	16,4	GTR+G
SOD2	Guidance	374	275	99	57	15,2	
	Gblocks	374	275	99	57	15,2	HKY+I+G

	Masked	918	439	384	265	28,9	
TUB2	Trimmed	730	385	253	151	20,7	
	Gblocks	689	288	401	277	40,2	HKY+I+G

Length of the sequence (Length), conserved sites (C), variable sites (V), parsimoniously informative sites (PI), percentage of parsimoniously informative sites (%PI), and substitution model used in Bayesian inferences (BI).

Supplementary Table 4 - Sequences of *Colletotrichum* isolates used in this article

Colletotrichum species	Strain	Host	Genes: GenBank accessions							
			ITS	GAPDH	CAL	ACT	TUB2	CHS	SOD2	GS
<i>Colletotrichum aenigma</i>	C1256.6	2	JX010243.1	JX009913.1	JX009684.1	JX009519.1	JX010390.1	JX009789.1	JX010312.1	JX010079.1
<i>Colletotrichum aeshynomenes</i>	3-1-3	1	JX010176.1	JX009930.1	JX009721.1	JX009483.1	JX010392.1	JX009799.1	JX010314.1	JX010081.1
<i>Colletotrichum alatae</i>	C1276.6	1	JX010191.1	JX010011.1	JX009739.1	JX009470.1	JX010449.1	JX009846.1	JX010371.1	JX010136.1
<i>Colletotrichum alienum</i>	C1194.22	3	JX010246.1	JX009959.1	JX009657.1	JX009552.1	JX010386.1	JX009755.1	JX010308.1	JX010075.1
<i>Colletotrichum aotearoa</i>	C1282.4	15	JX010205.1	JX010005.1	JX009611.1	JX009564.1	JX010420.1	JX009853.1	JX010345.1	JX010113.1
<i>Colletotrichum asianum</i>	C1187	2	JX010192.1	JX009915.1	JX009723.1	JX009576.1	JX010384.1	JX009753.1	JX010306.1	JX010073.1
<i>Colletotrichum boninense</i>*	CBS_128547		JQ005159.1	JQ005246.1	JQ005680.1	JQ005507.1	JQ005593.1	JQ005333.1	-	-
<i>Colletotrichum brevisporum</i>	BTL23/LC0870		JN050239.1	JN050228.1	-	JN050217.1	JN050245.1	-	-	-
	L57/LC0600		JN050238.1	JN050227.1	JN050222.1	JN050216.1	JN050244.1	KF687760.1	-	JN050233.1
<i>Colletotrichum clidemiae</i>	C1254.10	2	JX010274.1	JX009909.1	JX009639.1	JX009476.1	JX010439.1	JX009777.1	JX010353.1	JX010128.1
<i>Colletotrichum cliviae</i>	CSSS1		GU109479.1	GU085867.1	GU085863.1	GU085861.1	GU085869.1	GU085865.1	-	-
<i>Colletotrichum conoides</i>	CAUG33		KP890169.1	KP890163.1	KP890151.1	KP890145.1	KP890175.1	KP890157.1	-	-
<i>Colletotrichum cordylinicola</i>	C1315.1	1	JX010226.1	JX009975.1	JX009651.1	JX009586.1	JX010440.1	JX009864.1	JX010361.1	JX010122.1
<i>Colletotrichum dracaenophilum</i>	CBS_118199		JX519222.1	JX546707.1	-	JX519238.1	JX519247.1	JX519230.1	-	-
<i>Colletotrichum endophytica</i>	CAUG28		KP145441.1	KP145413.1	KP145357.1	KP145329.1	KP145469.1	KP145385.1	-	-
	CAUG1		KP145416.1	KP145388.1	KP145332.1	KP145304.1	KP145444.1	KP145360.1	-	-
<i>Colletotrichum fruticola</i>	C1253.2		JX010167.1	JX009998.1	JX009675.1	JX009491.1	JX010388.1	JX009772.1	JX010310.1	JX010077.1
	CAUG2	6	KP145417.1	KP145389.1	KP145333.1	KP145305.1	KP145445.1	KP145361.1	-	-
<i>Colletotrichum gloeosporioides</i>	IMI 356878	6	JX010152.1	JX010056.1	JX009731.1	JX009531.1	JX010445.1	JX009818.1	JX010365.1	JX010085.1
<i>Colletotrichum grevillea</i>	CBS 132879		KC297078.1	KC297010.1	KC296963.1	KC296941.1	KC297102.1	KC296987.1	-	KC297033.1
<i>Colletotrichum grossum</i>	CAUG31		KP890166.1	KP890160.1	KP890148.1	KP890142.1	KP890172.1	KP890154.1	-	-
<i>Colletotrichum hebeiense</i>	K3		-	KF377495.1	-	KF377532.1	KF288975.1	KF289008.1	KF678464.1	-
<i>Colletotrichum horii</i>	C1069.2		GQ329687.1	GQ329685.1	-	JX009533.1	JX010375.1	JX009748.1	JX010296.1	JX010072.1
<i>Colletotrichum hymenocallidis</i>	CAUG9		KP145423.1	KP145395.1	KP145339.1	KP145311.1	KP145451.1	KP145367.1	-	-
<i>Colletotrichum kahawae</i> subsp. <i>ciggaro</i>	C1200.1	10	JX010214.1	JX009971.1	-	JX009431.1	JX010426.1	JX009757.1	JX010348.1	-

	CBS 124.22		JX010228.1	JX009950.1	JX009744.1	JX009536.1	JX010433.1	JX009902.1	JX010367.1	JX010134.1
<i>Colletotrichum karstii</i>	CAUOS1		KP890103.1	KP890134.1	KP890118.1	KP890096.1	KP890110.1	KP890126.1	-	-
<i>Colletotrichum liaoningense</i>	CAUOS3		KP890105.1	KP890136.1	KP890120.1	-	KP890112.1	KP890128.1	-	-
	IMI 52264		KC566797.1	KC566651.1	KC566506.1	KC566943.1	KC566219.1	KC566364.1	-	-
<i>Colletotrichum musae</i>	C1266.11	1	JX010142.1	JX010015.1	JX009689.1	JX009432.1	JX010395.1	JX009815.1	JX010317.1	JX010084.1
<i>Colletotrichum nupharicola</i>	C1275.24	2	JX010189.1	JX009936.1	JX009661.1	JX009486.1	JX010397.1	JX009834.1	JX010319.1	JX010087.1
<i>Colletotrichum psidii</i>	CBS145.29	1	JX010219.1	JX009967.1	JX009743.1	JX009515.1	JX010443.1	JX009901.1	JX010366.1	JX010133.1
<i>Colletotrichum queenslandicum</i>	ICMP 1778	3	JX010276.1	JX009934.1	JX009691.1	JX009447.1	JX010414.1	JX009899.1	JX010336.1	JX010104.1
<i>Colletotrichum salsolae</i>	C1314	3	JX010242.1	JX009916.1	JX009696.1	JX009562.1	JX010403.1	JX009863.1	JX010325.1	JX010093.1
<i>Colletotrichum siamense</i>	LLTA01	12	JX010259.1	JX009974.1	JX009713.1	JX009549.1	JX010415.1	JX009895.1	JX010337.1	JX010105.1
<i>Colletotrichum thailandicum</i>	HR01MFU/LC0596/L62/MFUCC110113		JN050242.1	JN050231.1	JN050225.1	JN050220.1	JN050248.1	-	-	JN050236.1
	C1316.13		JX010294.1	JX010006.1	JX009591.1	JX009444.1	JX010447.1	JX009869.1	JX010372.1	JX010139.1
<i>Colletotrichum theobromicola</i>	C1266.6	8	JX010288.1	JX010003.1	JX009589.1	JX009448.1	JX010379.1	JX009819.1	JX010301.1	JX010062.1
<i>Colletotrichum ti</i>	ICMP5285	1	JX010267.1	JX009910.1	JX009650.1	JX009553.1	JX010441.1	JX009897.1	JX010363.1	JX010124.1
<i>Colletotrichum tropicale</i>	C1272.16	3	JX010275.1	JX010020.1	JX009722.1	JX009480.1	JX010396.1	JX009826.1	JX010318.1	JX010086.1
<i>Colletotrichum tropicicola</i>	L58/LC0598		JN050240.1	JN050229.1	JN050223.1	JN050218.1	JN050246.1	-	-	JN050234.1
<i>Colletotrichum viniferum</i>	CAUG27		KP145440.1	KP145412.1	KP145356.1	KP145328.1	KP145468.1	KP145384.1	-	-
<i>Colletotrichum xanthorrhoeae</i>	C1271	1	JX010261.1	JX009927.1	JX009653.1	JX009478.1	JX010448.1	JX009823.1	JX010369.1	JX010138.1
<i>Colletotrichum yunnanense</i>	CBS 132135		JX546804.1	JX546706.1	-	JX519239.1	JX519248.1	JX519231.1	-	-
<i>Glomerella cingulata</i>	ICMP10643		JX010224.1	JX009908.1	JX009630.1	JX009540.1	JX010436.1	JX009891.1	JX010358.1	JX010119.1

CAPÍTULO IV

Conclusões Gerais

Conclusões gerais

- O genoma de *C. musae* tem tamanho, conteúdo GC e o número de genes únicos preditos semelhante a outros representantes do gênero.
- A qualidade da montagem do genoma depende diretamente da cobertura do sequenciamento, no grau de pureza das sequências e no número total de estratégias de sequenciamento, esses fatores corroboram para uma boa predição consequentemente anotação gênica e de metabólitos secundários.
- *C. musae* GM20 mostrou vários grupos gênicos e alguns agrupamentos de genes putativos que podem ser sugeridos como determinantes para a virulência, como por exemplo o agrupamento de genes de tipo apicidina ou de toxina HC e o agrupamento de genes tipo alternapirona.
- Notavelmente, o potencial de metabólitos secundários de *C. musae* GM20 parece semelhante ao Nara gc5 de *C. gloeosporioides*, embora vários grupos gênicos apresentados no genoma Nara gc5 de *C. gloeosporioides* não estejam conservados no genoma GM20 de *C. Musae*. Assim, os grupos não conservados em *C. musae* GM20, mas apresentados na cepa Nara gc5 de *C. gloeosporioides*, podem ser essenciais para atingir uma infecção bem sucedida em outros hospedeiros.