UNIVERSIDADE ESTADUAL DE SANTA CRUZ

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

ANÁLISES COMPARATIVAS DOS EFETORES DE *Ceratocystis* sp COM ÊNFASE EM *C. cacaofunesta*

GABRIELA NICOLLE RAMOS LIZARDO

ILHÉUS – BAHIA – BRASIL

Dezembro de 2022

GABRIELA NICOLLE RAMOS LIZARDO

ANÁLISES COMPARATIVAS DOS EFETORES DE *Ceratocystis* sp COM ÊNFASE EM *C. cacaofunesta*

Dissertação apresentada à Universidade Estadual de Santa Cruz, como parte das exigências para obtenção do título de Mestre em Genética e Biologia Molecular.

Área de Concentração: Genética e Biologia Molecular

ILHÉUS – BAHIA – BRASIL

Dezembro de 2022

GABRIELA NICOLLE RAMOS LIZARDO

ANÁLISES COMPARATIVAS DOS EFETORES DE *Ceratocystis* sp COM ÊNFASE EM *C. cacaofunesta*

Dissertação apresentada à Universidade Estadual de Santa Cruz, como parte das exigências para obtenção do título de Mestre em Genética e Biologia Molecular.

Área de Concentração: Genética e Biologia Molecular

APROVADA: 20 de dezembro de 2022

Dra. Fátima Cerqueira Alvim

(UESC)

Dra. Francisca Feitosa Jucá Santos

(CEPLAC)

Dr. Raner José Santana Silva

(UESC)

Dr. Ronan Xavier Corrêa

(UESC/ Orientador)

A minha mãe, minha tia e avó por ser meu maior exemplo de força e superação!
Dedico.

AGRADECIMENTOS

A meu orientador, Dr. Ronan Xavier Corrêa, pela orientação, dedicação, suporte, ensinamento e confiança em mim;

A meus co-orientadores, Dr. Eric Roberto Guimarães Rocha Aguiar e o Dr. Carlos Priminho Pirovani pelos valiosos conselhos;

A colaboração de Jonathan Javier Mucherino Muñoz que foi um enorme suporte na pesquisa. Por toda paciência, dedicação e atenção;

À Universidade Estadual de Santa Cruz e ao programa de pós-graduação de Genética e Biologia Molecular pela oportunidade do mestrado;

Às secretárias do programa da pós-graduação, Fabrícia e Mara, por todo o empenho e dedicação com que exercem o trabalho;

À OEA e CAPES pelas concessões das bolsas de mestrado;

Ao NBCGIB, pelas instalações e suporte técnico para a pesquisa;

Ao meu amado padre celestial Deus, por cada uma de suas bençõe por nunca me deixar sozinha, pela inteligência e a sabedoria;

À minha mãe Sandra Lizardo, por todo o carinho, apoio, amor e conselhos. Por acreditar em mim.

À minha querida avó Blacina Lizardo e minha tia Lourdes Lizardo, por sempre estar presentes por cada uno dos conselhos e apoio, e por seu amor incondicional;

A meu pai Nicolas Ramos, minhas irmãs e minha prima Claudia Lizardo pelo apoio e carinho e a cada um dos meus familiares envolvidos amo vocês;

Aos meus amigos (Gabriela, Daniel, Eduardo, Ismael, Javier) por sua amizade em um pais totalmente novo para mim, pôr o apoio e cada sorriso

A todos que contribuíram, seja de forma direta ou indireta, muito obrigada!

ÍNDICE

Extrato

RAMOS-LIZARDO, Gabriela Nicolle, Universidade Estadual de Santa Cruz, Ilhéus, dezembro de 2022. **ANÁLISES COMPARATIVAS DOS EFETORES DE *Ceratocystis* sp COM ÊNFASE EM *C. cacaofunesta.*** Orientador: Ronan Xavier Corrêa. Co-orientadores: Eric Roberto Guimarães Rocha Aguiar, Carlos Priminho Pirovani. Colaborador: Jonathan Javier Mucherino Muñoz.

O gênero *Ceratocystis* inclui um grande número de fungos fitopatogênicos, a maioria dos quais ataca diferentes espécies de plantas. A Murcha de *Ceratocystis* é uma das principais doenças que afetam o cacaueiro, causada pelo patógeno *Ceratocystis cacaofunesta*, que acarreta o entupimento do xilema, levando as plantas à morte. Os fungos fitopatogênicos interagem com seus hospedeiros principalmente através da secreção de efetores. Portanto, a identificação *in silico* e a caracterização funcional de candidatos a efetores é um passo importante na compreensão do modo de ação dos fitopatógenos. Este estudo teve como objetivo avaliar a interação planta-patógeno de *C. cacaofunesta* com *Theobroma cacao*, por meio da análise comparativa de genes de diferentes espécies de *Ceratocystis* envolvidas na patogênese e especificidade das plantas. Para isso foi realizada uma nova anotação do genoma *C. cacaofunesta*, permitindo prever o secretoma de proteínas candidatas a efetoras através de uma análise *in silico*, utilizando características dos peptídeos quanto a presença de peptídeo de sinal para secreção, ausência de domínio transmembrana, tamanho pequeno, sem retenção no reticulo endoplasmático e riqueza de resíduos de cisteína. As proteínas secretadas também foram preditas usando o software EffectorP, que é utilizado para distinguir proteínas efetoras secretadas em

fungos fitopatogênicos. Um total de 160 candidatos a efetores foram identificados no genoma de *C. cacaofunesta*, sendo 102 citoplasmáticos e 58 apoplásticos. A caracterização funcional das proteínas efetoras foi realizada através de InterProScan, das quais somente 45 delas mostrarem informação associada com características funcional envolvidas nas interações com as plantas incluindo CAZymes, hidrolases, linases e oxidoreductases. O repertório de candidatos a efetores das especies de *Ceratocystis* estudadas nos permitiu fazer uma comparação por ortologia entre eles, e foi realizada através de OrthoVem, apresentando 2 grupos exclusivos para *C. cacaofunesta*. Nossa base de dados foi gerada utilizando uma estratégia robusta de bioinformática, seguida de curadoria manual, gerando um recurso valioso para o estudo dos mecanismos de infecção destes patógenos em seus hospedeiros.

Palavras-chave: fator de patogenicidade, fitopatógeno, murcha vascular, Efetor, *Theobroma cacao*.

Abstract

RAMOS-LIZARDO, Gabriela Nicolle, Universidade Estadual de Santa Cruz, Ilhéus, dezembro de 2022. **COMPARATIVE ANALYSIS OF EFFECTS OF *Ceratocystis* sp WITH EMPHASIS ON *C. cacaofunesta.*** Orientador: Ronan Xavier Corrêa. Co-orientadores: Eric Roberto Guimarães Rocha Aguiar, Carlos Priminho Pirovani. Colaborador: Jonathan Javier Mucherino Muñoz.

The genus *Ceratocytis* includes a large number of phytopathogenic fungi, most of which attack different plant species. *Ceratocystis* wilt is one of the main diseases affecting the cacao tree, caused by the pathogen *Ceratocystis cacaofunesta*, which causes clogging of the xylem, leading the plants to death. Phytopathogenic fungi interact with their hosts mainly through the secretion of effectors. Therefore, *in silico* identification and functional characterization of candidate effectors is an important step in understanding the mode of action of phytopathogens. This study aimed to evaluate the plant-pathogen interaction of *C. cacaofunesta* with *Theobroma cacao* by comparative analysis of genes from different *Ceratocystis* species involved in plant pathogenesis and specificity. To this end, a new annotation of the *C. cacaofunesta* genome was performed, allowing the secretome of candidate effector proteins to be predicted by an *in silico* analysis using peptide features such as the presence of a signal peptide for secretion, absence of a transmembrane domain, small size, no retention in the endoplasmic reticulum, and richness of cysteine residues. Secreted proteins were also predicted using EffectorP software, which is used to distinguish secreted effector proteins in phytopathogenic fungi. A total of 160 effector candidates were identified in the *C. cacaofunesta* genome, of which 102 were cytoplasmic and 58

apoplastic. Functional characterization of the effector proteins was performed using InterProScan, of which only 45 of them showed information associated with functional features involved in interactions with plants including CAZymes, hydrolases, lyases and oxidoreductases. The repertoire of candidate effectors of the *Ceratocystis* species studied allowed us to make a comparison by orthology between them, and was performed using OrthoVem, presenting 2 groups unique to *C. cacaofunesta*. Our database was generated using a robust bioinformatics strategy followed by manual curation, generating a valuable resource for studying the infection mechanisms of these pathogens in their hosts.

Keywords: pathogenicity factor, phytopathogen, vascular wilt, effector, *Theobroma cacao*

1. Introdução

O fitopatógeno *Ceratocystis cacaofunesta* é um ascomiceto, necrotrófico específico do xilema, de difícil controle, geralmente entra nos tecidos da planta através de aberturas provocadas no caule por ferramentas ou insetos. No cacaueiro (*Theobroma cacao*), a doença recebe o nome de Mal do facão, em referência a esta ferramenta utilizada na poda. Esse fungo penetra também através de aberturas feitas por insetos, que atuam como vetores de disseminação da doença. Essa doença causa danos irreversíveis ao sistema vascular da planta, após a instalação do fungo nos vasos condutores, surgem estrias na medula que se espalham pelo parênquima radial, causando a morte de parte do câmbio vascular e impedindo o movimento da seiva bruta para as partes aéreas da planta, causando murchamento e, em pouco tempo, a morte da planta. As folhas mortas permanecem presas por muito tempo à planta, e o fungo cresce no tecido em decomposição onde se formam os esporos sexuados e assexuais (MALHEIRO, 2015, FERREIRA *et al*., 2020).

No desenvolvimento de doença na planta, o patógeno expressa genes de patogenicidade, especificidade e virulência contra um hospedeiro em particular (MORA-OCAMPO, 2016). De acordo com Barbosa *et al*. (2018), durante o processo de infecção, a interação patógeno-hospedeiro é modulada por moléculas da planta e do microrganismo que interagem entre si. Para promover a patogênese, os patógenos secretam efetores, como fatores de virulência, que suprimem as defesas do hospedeiro e modulam a resposta celular (PRADHAN, *et al.*, 2021, BORGES, 2018). Os efetores têm sido considerados como “moléculas liberadas por um organismo ou ligadas a ele, que alteram a fisiologia de outro organismo" usados para obter nutrientes, suprimir as defesas da planta, facilitando assim infecções (fatores de virulência ou

toxinas) e/ou que provocam as defesas da planta (fatores de avirulência: Avr) (MALHEIRO, 2015).

As plantas suscetíveis ao patógeno não reconhecem as moléculas do invasor e, como resultado, o sistema imunológico da planta é derrotado e a doença se desenvolve. Em situação de resistência, a planta reconhece as moléculas do patógeno e ativa suas defesas, mantendo-se saudável (BARBOSA *et al*., 2018). Por causa da complexidade da interação planta-patógeno, diferentes estudos têm sido realizados, utilizando-se diversas abordagens tais como genética clássica, biologia celular e bioquímica, auxiliadas com ferramentas de bioinformática. A análise do genoma de fungos tem sido imprescindível para ter um melhor entendimento desta dinâmica de interação, assim como estudos subsequentes ao do genoma que pode ser o do proteoma combinado com ferramentas de bioinformática têm sido métodos poderosos para prever os possíveis candidatos a efetores em um organismo (MALHEIRO, 2015).

Nesta perspectiva, o presente trabalho teve como objetivo fazer a previsão das proteínas efetoras de cinco espécies de *Ceratocystis; C. albifundus*, *C. cacaofunesta*, *C. fimbriata*, *C. manginecans e C. platani*, a partir da análise comparativa dos genomas disponível na literatura. Os objetivos específicos foram: (I). Anotação estrutural dos genomas das espécies *C. albifundus, C. cacaofunesta e C. manguinecans*; (II). Identificar *in-silico* o secretomas e efetoroma proteico de *C. cacaofunesta;* (III). Apresentar uma lista do repertório de candidatos a proteínas efetoras secretadas das cinco espécies de *Ceratocystis;* (IV). Identificar proteínas efetoras secretadas comuns entre as cinco espécies de *Ceratocystis.*

2. Revisão de Literatura

O cacaueiro (*Theobroma cacao* L., Malvaceae) é uma planta comum nas bacias dos rios Amazonas e Orinoco, conhecido a mais de mil anos pelos Maias e Astecas. Trata-se de uma árvore frutífera, perene, típica de clima tropical, encontrada em seu habitat tanto nas terras baixas sob a proteção de grandes árvores, dentro de bosques úmidos como em altitudes mais elevadas (até 1.000 metros), em florestas relativamente menos úmidas e exuberante (MODA *et al.* 2019). Suas sementes são matéria prima para as indústrias alimentícia, cosmética e farmacêutica, portanto representa grande importância econômica em vários países do mundo (MORA-OCAMPO, 2016).

Atualmente, o Brasil ocupa o sétimo lugar na produção de cacau no mundo, e o estado do Pará é o maior produtor do Brasil com uma produção nacional de 51% sendo aproximadamente 26 mil produtores envolvidos no cultivo do cacau, gerando 64 mil empregos diretos e 225 mil indiretos o que permite movimentar cerca de R$ 1,1 bilhão (HOLANDA *et al*., 2021). Segundo ICCO (2017), a produção mundial de cacau atingiu 3,9 milhões de toneladas (t) de amêndoas secas no período 2015/2016. Além disso, a produção brasileira de cacau mudou ligeiramente nos últimos 15 anos, não só porque a área colhida aumentou marginalmente, mas também por causa de tensões abióticas e bióticas. Com respeito a essas tensões bióticas, a incidência de várias doenças tem contribuído para diminuir e tem levado a um desequilíbrio ainda maior na oferta e demanda de amêndoas de cacau (FERNANDES *et al*., 2018).

Pragas ou doenças são um grande problema na produção de cacau porque afetam diretamente a qualidade e a produtividade e suas perdas dependem da cultura e da gravidade do

ataque. Estima-se que o mundo perde entre um 20% e 40% dos rendimentos das suas lavouras ano após ano devido aos danos causados por pragas e doenças, enquanto no Brasil as perdas anuais podem chegar a os R$ 55 bilhões (HOLANDA *et al*., 2021). Os cacaueiros são constantemente confrontados com patógenos importantes que atacam ramos vegetativos, frutos, almofadas de flores e o sistema vascular (MORA-OCAMPO, 2016). De acordo com Holanda *et al* (2021), as doenças fúngicas que mais atingem o cacaueiro no Brasil são a Vassoura de bruxa (*Moniliophthora perniciosa*), a Podridão parda (*Phytophthora spp*) e o Mal do facão (*C. cacaofunesta*). Em 2022, a doença Monilíase foi constatada em plantações de cacau no Acre (SIVIERO *et al.,* 2022).

2.1 O gênero *Ceratocystis*

Ceratocystis é um dos numerosos gêneros pertencentes à família *Ceratocystidaceae (Microascales, Sordariomycetes, Ascomycota)* (KANZI *et al*., 2020). Essa espécie fúngica foi originalmente descrita em 1890 por Halsted como o agente causal de podridão da raiz negra da batata doce (*Ipomoea batatas Lam*.) recebendo o nome de *Ceratocystis fimbriata* (CABRERA *et al*., 2016). Os primeiros relatos da Murcha de *Ceratocystis* no Brasil foram em São Paulo na cultura de *Crotalaria juncea,* que é uma leguminosa usada para proteção do solo contra erosões, posteriormente também foi encontrada em manga, figo e inhame. Na Bahia e Minas Gerais foi encontrado afetando plantações de Eucalipto e a cultura do cacau no Alto Amazonas e na Bahia (MALHEIRO, 2015).

O gênero contém vários patógenos vegetais significativos, que causam murcha e cancro através da degradação do parênquima, xilema e destruição vascular em uma ampla gama de espécies vegetais economicamente importantes (DHARMARAJ *et al*., 2022, MOLANO *et al*., 2018). Por exemplo, o *C. platani* causa a murcha severa de Platanus na Europa, *Ceratocystis manginecans* produz murcha da mangueira (WILKEN *et al*., 2013). No Brasil, foram registradas três espécies: *C. paradoxa*, que ataca principalmente as monocotiledôneas; *C. cacaofunesta*, que causa grandes problemas nas plantações de cacau (*T. cacao*); e *C. fimbriata*, que afeta a batata-doce (*Ipomoea batatas)* e o *eucalipto* (*Eucalyptus* spp.) (FERREIRA *et al*., 2020).

Em 2005, Engelbrecht e Harrington descreveram o fungo *C. cacaofunesta* como uma nova espécie no complexo de espécies *C. fimbriata* que pode ser distinguida de outras linhagens do complexo *C. fimbriata* por ser específico para infectar *T. cacao* (CABRERA *et al*., 2016). Os primeiros relatos da Murcha de *Ceratocystis* do cacaueiro ocorreram no Equador (1918), Colômbia (1940s), Trinidad (1950s), Venezuela (1958) e Costa Rica (1958). Nos anos 70, começou a se espalhar da América Central para o Brasil (sudoeste da Amazônia (Rondônia) (1978) e Bahia (1990) (DELGADO-OSPINA *et al*., 2021). Segundo Ferreira e*t al* (2020), a murcha de *Ceratocystis* e responsável pela morte de aproximadamente um 85% de plantas em fazendas de cacau em na região sul da Bahia, ocupando o terceiro lugar de perdas econômicas no Brasil.

C. cacaofunesta é um fungo com estilo de vida necrotrófico de difícil controle, causa danos irreversíveis ao sistema vascular da planta e em pouco tempo a planta morre, mas as folhas mortas ficam presas por muito tempo permitindo ao fungo crescer no tecido em decomposição onde se formam os esporos sexuados e assexuais (FERREIRA *et al*., 2020). O fungo cresce otimamente entre 18 e 28 °C e pode produzir ascósporos e micélio dentro do hospedeiro, ou aleurioconídios presentes no solo ou detritos vegetais quando se encontra em uma variedade de ambientes desfavoráveis (PALADINES-REZABALA, *et al*., 2022).

O fungo produz uma camada superficial parcial ou totalmente embutida no substrato, esférica, com longas fimbriadas, de coloração marrom a preta. Os ascósporos são elipsoidais, hialinos e possuem uma casca gelatinosa que lhes confere uma aparência de chapéu. Eles são produzidos em ascos voláteis, cujas paredes se dissolvem facilmente e são então liberadas através do ostíolo em uma substância gelatinosa (**Figure 1**) (SANTOS, 2012; CABRERA *et al*., 2016).

Figure 1 Sintomatologia de murcha de *Ceratocystis* no cacau e características morfológicas do *C. cacaofunesta*. (a) cacaueiro mostrando os sintomas típicos: murcha e necroses, folhas murchas presas aos galhos por várias semanas após a morte da árvore. (b) Corte transversal vertical de cacaueiro infectado com *C. cacaofunesta* com regiões do parênquima apresentando sombras escuras. (c) Ascomata de *C. cacaofunesta* no Malt-Agar. Peritécio marrom escuro com uma base globosa imersa no substrato e massas pegajosas de ascósporos alaranjados que emergem de seus pescoços. (d) Peritécio com massas de ascósporos emergentes das hifas ostiolares. (e) Ascósporos hialinos em forma de chapéu. (f) Aleurioconidia marrom e globosa. (g) Micélio e endoconia cilíndrica de *C. cacaofunesta* (CABRERA *et al.*, 2016).

A principal fonte de inóculo do fungo *C. cacaofunesta* são plantas doentes e mortas, incluindo raízes, troncos e galhos, além disso o solo também e considerada como outra fonte de inóculo em espécies do gênero, como *C. fimbriata* (PALADINES-REZABALA, *et al*., 2022) . O mecanismo de infecção de *Ceratocystis* é diferente dos chamados "verdadeiros patógenos vasculares" que inclui fungos dos gêneros *Verticilium e Fusarium*. O fungo *C. cacaofunesta* penetra o cacaueiro através de orifícios e ferimentos causados por insetos vetores do gênero *Xyleborus* sp. (*Coleoptera-Scolytidae*) ou ferramentas de corte contaminadas, dando origem ao nome popular da doença Mal do facão (**Figure 2**) (MAGALHÃES *et al*., 2019, MOLANO, 2019).

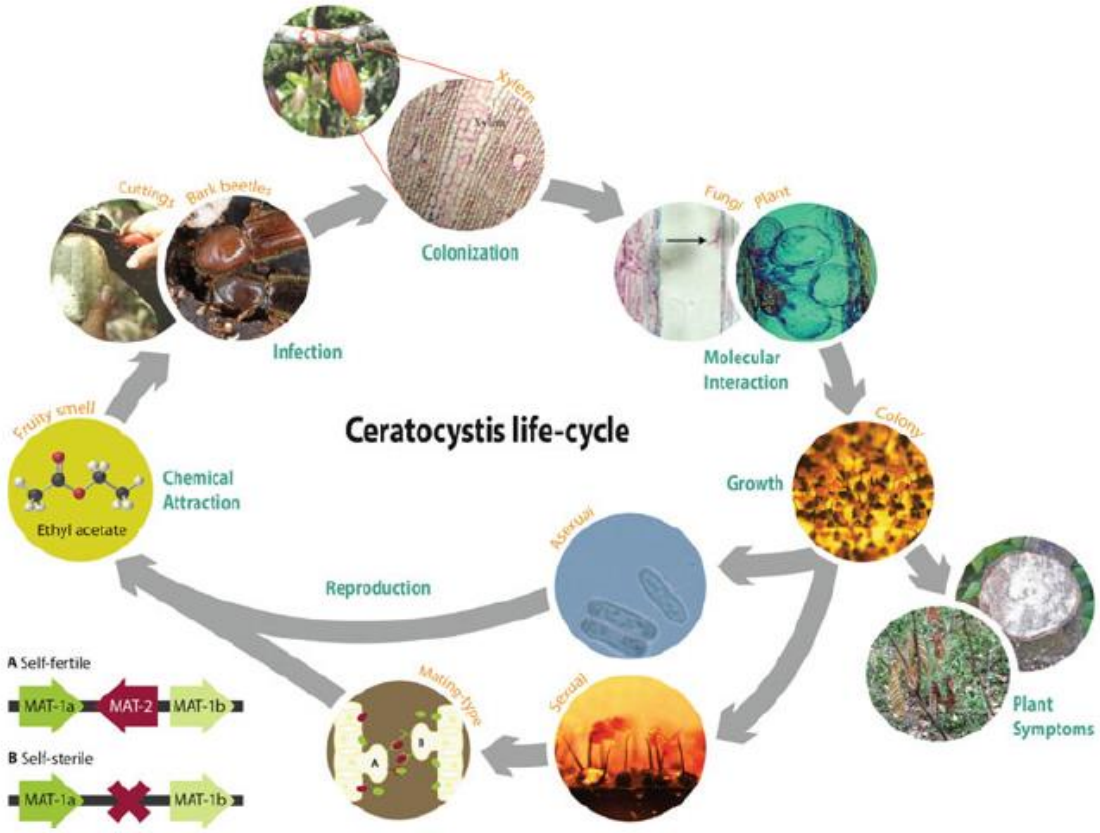

Figure 2. Ciclo de vida de *Ceratocystis* representado pelo ciclo de *C. cacaofunesta* nos hospedeiros de cacau (CABRERA *et al.*, 2016).

Independentemente do mecanismo usado pelos fungos causadores de murcha vascular para entrar no hospedeiro, eles colonizam e proliferam os vasos do xilema e quando a planta morre, os fungos se movimentam para outros tecidos e esporulam na superfície da planta começando um novo ciclo de infecção (SAIN; REP, 2015). O período de colonização pode ser longo, cerca de 15-18 meses sem apresentar sintomas visíveis, mas quando os primeiros sintomas aparecem, a planta pode morrer após 10-30 dias (PALADINES-REZABALA, *et al*., 2022). Os primeiros sintomas associados a esta doença é a perda da coloração da folhagem, seguida do escurecimento do caule, ramos e murcha das folhas que leva a morte súbita da planta (MOLANO, 2019).

De acordo com Sain & Rep (2015), os patógenos, incluindo os fungos da murcha vascular, secretam proteínas durante a colonização para estabelecer uma interação bem-sucedida

entre patógeno e hospedeiro. Estas proteínas secretadas pelos patógenos nas plantas são conhecidas como proteínas efetoras e são usadas para obter vantagens de virulência, um papel importante destas proteínas é suprimir a imunidade das plantas e contribuir para o desenvolvimento das infecções patogênicas (WANG, *et al*., 2022).

2.2 Mecanismos de interação planta-patógeno

Os patógenos vegetais são um dos principais fatores que dificultam a criação de culturas comerciais (ZHANG, *et al*., 2022). Estes patógenos geralmente usam armas moleculares para infectar com sucesso seus hospedeiros, secretando um grande portfólio de diversas proteínas e enzimas (BECKERSON, *et al*., 2019). Micro-organismos patogênicos, incluindo omicetos, fungos e bactérias, precisam nutrientes das células hospedeiras para sobreviver e se reproduzir, por outro lado, as plantas hospedeiras empregam várias estratégias de defesa para inibir o crescimento de patógenos (ZHANG, *et al*., 2022).

Segundo Ngou *et al*. (2021), as plantas detectam os agentes patogênicos e ativam a imunidade utilizando dois tipos muito diferentes de receptores. Os receptores de superfície celular (Receptores de reconhecimento de padrão PRRs) principalmente kinases tipo receptor (RRLK) e proteínas tipo receptor (RLP) as quais conseguem detectar moléculas derivadas de agentes patogênicos (Padrões moleculares associados a patógenos PAMPs) como a flagelina flg22, o fator de alongamento EF-Tu, alguns lipopolissacarídeos e oligossacarídeos em bactérias, quitina e glucano derivado da parede celular em fungos, e glutamina aminotransferase e β-glucano em oomicetos, que se acumulam fora das células vegetais, desencadeando uma imunidade ativada por PAMPs conhecida como PTI (NGOU, *et al.,* 2022, ZHANG, *et al*., 2022).

A PTI é suprimida por proteínas efetoras derivadas de agentes patogênicos, que são injetadas nas células hospedeiras para permitir o crescimento dos patógenos. Ela contribui para a defesa do hospedeiro contra patógenos não-adaptados e, quando estes agentes patogênicos encontram-se adaptados ao hospedeiro, esses patógenos secretam uma miríade de efetores em as células vegetais, alguns dos quais são detectados por proteínas de resistência intracelular (R), levando à imunidade acionada por efetores (ETI) de uma forma específica a cada raça patogênica (YU, *et al*., 2017, NGOU, *et al*., 2021).

Estudos mostram que os efetores patogênicos são fatores importantes no processo de Infecção nas plantas e conseguem atuar de múltiplas formas sobre diferentes alvos, suprimindo a imunidade das plantas, manipulando a fisiologia vegetal e sendo reconhecidos pelos mecanismos de defesa do hospedeiro, promovendo assim a infestação, expansão e colonização de patógenos (ZHANG, *et al*., 2022). Uma definição generalizada de efetores é "proteínas e moléculas pequenas que alteram as estruturas e funções das células hospedeiras, facilitando assim a colonização de patógenos" (TARIQJAVEED *et al*., 2021; ZHANG *et al*., 2022).

De acordo com Zhan *et al.* (2022), os patógenos vegetais produzem diferentes estratégias de infestação de acordo com seus diferentes modos nutricionais. Baseado na aquisição de nutrientes, os patógenos fúngicos são assim classificados: os biotróficos que precisam dos tecidos vivos para a absorção de nutrientes durante a colonização do hospedeiro; os necrotróficos que matam ativamente as células hospedeiras durante a colonização para prosperar nos tecidos mortos; e os hemibiotróficos que adotam a fase biotrófica inicial (que requer tecidos vivos do hospedeiro para sobrevivência), e mais tarde passam para a fase necrotrófica na qual conseguem matar os tecidos infectados para obter nutrientes (PRADHAN, *et al*., 2021).

Embora os patógenos fúngicos adotem estilos de vida diferentes para atacar as plantas hospedeiras, todos eles compartilham uma estratégia essencial para a colonização das plantas, que é a secreção de efetores de virulência nas células hospedeiras (TARIQJAVEED, *et al*., 2021). Para obter os nutrientes e estabelecer a patogenicidade, os fungos seguem uma série de etapas que podem ser decisivas para a infecção: (1) inicialmente aderem se à superfície da planta; (2) os esporos germinam nesta área e formam-se estruturas infecciosas que (3) permitem a penetração do hospedeiro através da quebra das barreiras externas (cutícula e parede celular), culminando na (4) colonização do tecido do hospedeiro (BARBOSA *et al*., 2018).

Com base na localização subcelular, os efetores fúngicos são classificados em apoplásticos e citoplasmáticos. Os efetores apoplásticos funcionam em espaços extracelulares relacionados com a quebra das barreiras físicas e químicas do hospedeiro conseguindo interromper a primeira linha de defesa das plantas. Os efetores citoplasmáticos, também denominados intracelulares, atuam dentro das células hospedeiras tendo como alvo componentes imunes importantes na membrana plasmática, citoplasma (incluindo organelas) e núcleo. Eles

conseguem suprimir a imunidade do hospedeiro e reprogramar sua fisiologia promovendo assim a colonização dos patógenos (TARIQJAVEED, *et al*., 2021, ZHANG, *et al*., 2022).

Em geral, os efetores fúngicos são altamente polimórficos, devido à sua rápida adaptação ao hospedeiro (BARBOSA, *et al*., 2018), a maioria possui um peptídeo de sinal para secreção, nenhum domínio transmembrana, nenhuma semelhança com outros domínios protéicos conhecidos, são bastante pequenas em tamanho, ricas em cisteína e, na maioria das vezes, específicas da espécie (DE MANDAL; JEON, 2022). Essas características tornam possível encontrar alguns padrões estruturais e/ou funcionais em comum dentre os efetores, permitindo determiná-los.

A previsão computacional de efetores fúngicos é difícil porque geralmente não compartilham semelhanças de sequência ou motivos de sequência conservados. Sperschneider *et al*. (2016) desenvolveram um software que permite prever efetores a partir de proteínas secretadas pelo patógeno, o EffectorP 3.0 consegue prever se uma proteína secretada é um efetor apoplástico, um efetor citoplasmático ou um não-efetor (SPERSCHNEIDER; DODDS, 2022). EffectorP é um programa pioneiro na aplicação de machine learning para identificação de efetores fúngicos, com especificidade e sensibilidade acima de a 80%, e é um programa promissor para tais predições (BARBOSA *et al.*, 2018).

2.3 Sequenciamento de genomas

Os avanços nas técnicas de sequenciamento genômico associado à sua redução de custos possibilitaram o sequenciamento de diversos microrganismos de interesse biotecnológico. Com o sequenciamento de genomas completos de fungos fitopatogênicos e o uso de ferramentas bioinformáticas, é possível prever genes relacionados à patogenicidade e, assim, melhorar o entendimento do processo de infecção no qual está envolvido, também tem sido possível prever candidatos a efetores em uma imensa gama de microrganismos (BARBOSA *et al*., 2018).

Segundo Malheiro (2015), os estudos de genômica, transcriptômica, proteômica e metabolômica, por sua complexidade, juntos contribuem para o entendimento global do organismo. Ao longo dos anos, muitos estudos foram realizados para analisar as interações

planta-patógeno, e o grande número de biomoléculas envolvidas nessa interação que a torna muito complexa e dinâmica.

Em uma análise global mitocondrial de *C. cacaofunesta*, usando metagenômica, transcriptômica e proteômica, foi possível identificar proteínas relacionadas à virulência, como transportadores dependentes de ATP, aldeído desidrogenase, desidrogenase alternativa (NADH), oxidase alternativa (AOX), transportador mitocondrial FOW1, glutationa S-transferase e proteína relacionada à hipóxia, onde as enzimas NADH e AOX fazem parte dos componentes alternativos da cadeia respiratória e estão relacionadas à patogenicidade fúngica (AMBROSIO *et al*., 2013).

Conforme descrito por Malheiro (2015), em outro estudo proteico realizado em *C. cacaofunesta* com *Teobroma cacao*, foram identificadas três proteínas secretadas pelo fungo, sendo uma delas glicoamilase e outras duas proteínas as quais são inibidores de tripsina e uma cerato- platanina. A cerato-platanina faz parte de uma classe importante de proteínas conhecidas como efetores, que são moléculas produzidas por patógenos durante a infecção. Estas proteínas modulam a imunidade inata da planta, promovendo supressão ou indução da defesa das plantas.

A análise *in silico* de cacau infectado com *C. cacaofunesta* identificou outras proteínas relacionadas à patogênese (PR) produzidas pela planta como uma resposta de defesa que pode danificar diretamente o fungo ou proteger a célula vegetal de agentes patogênicos. Dentre elas, foram identificadas as osmotinas e as taumatinas que pertencem à família das proteínas relacionadas à patogênese do grupo 5 (PR5), que estão envolvidas na resistência sistêmica adquirida e induzida. Membros da família PR5 têm demonstrado ação antifúngica, inclusive contra fungos que atacam o cacau. Estas proteínas também podem promover a liberação de fitoalexinas elicitoras, bem como induzir a síntese de compostos fenólicos (FERREIRA *et al*., 2020).

Além disso, resultado de estudos enfatizados em patógenos fúngicos demostram que apenas alguns deles são capazes de colonizar o sistema vascular das plantas e causar murcha. Algumas das espécies causadoras de murchas em plantas incluem *Ceratocystis*, e várias espécies pertencentes aos gêneros *Fusarium, Verticillium e Ophistoma,* destacando-se as seguintes: *Fusarium oxysporum (Fo), Verticillium dahliae (Vd) V. albo-atrum (Vaa) e Ophiostoma ulmi (Ou*). Estes fungos são patógenos do solo ou saprofíticos, e geralmente penetram através da raiz

para colonizar, eles obstruem o xilema e podem penetrar diretamente a parede celular da planta ou por meio de feridas usando vetores (SAIN; REP, 2015).

Estudos de sequenciamento genômico e proteico das cepas de *Fo, Vd, Vaa* e *Ou* e alguns de seus hospedeiros como o tomate, arabidopsis, algodão e banana fornecem também grande informação, revelando genes expressados e proteínas secretadas durante a interação planta-patógeno (MORA-OCAMPO, 2016). De acordo com Malheiro (2015), as proteínas secretadas (secretoma) por fungos fitopatogênicos são a chave para conhecer a interação planta-patógeno uma vez que podem alterar o metabolismo do hospedeiro, fornecendo ao fungo um ambiente adequado para a conclusão do seu ciclo de vida.

Alguns das análises realizadas em *Fusarium oxysporum sp lycopersici (FOL),* revelaram que este patógeno codifica 126 proteínas pequenas, ricas em cisteína e outras proteínas potencialmente secretadas. Essas pequenas proteínas ricas em cisteína como as Six secretadas por *Fo* são efetores necessários para a virulência completa e podem estar envolvidos na supressão das respostas de defesa em tomate (MORA-OCAMPO, 2016). Estas proteínas Six também contribuem para o murchamento das plantas através da colonização fúngica (SRINIVAS *et al*., 2019). Além disso, *Fo, Vd e Vaa* também secretam proteínas tipo Nep1 que podem ser citotóxicas e induzir a necrose celular em alguns hospedeiros (SAIN; REP, 2015).

Na mesma linha, Srinivas *et al*. (2019) descreve que a maioria destes fungos para superar as barreiras químicas das plantas, produzem muitas enzimas como a endopoligalacturonase (PG), exopoligalacturonase (PGX), tomatinase (TOM), metaloprotease (Mep) e serino protease (Sep) produzidas pela FOL que também contribuem para a patogenicidade. Outras enzimas chamadas proteases secretadas por estes patógenos clivam quitinases que são usadas pela planta como defesa para degradar a parede celular do fungo. Alguns exemplos dessas proteases são as Mep1 e Sep1, identificadas em Fol, as quais clivam quitinases do tomate, assim como também a enzima tomatinase secretada por *Fol e Vaa*, as quais degradam a saponina chamada alpha-tomatina encontrada no tomate (MORA-OCAMPO, 2016).

Entretanto, os estudos a nível molecular sobre o fungo *C. cacaofunesta* e sua interação com o cacau são poucos. A fim de compreender os mecanismos subjacentes às interações entre *C. cacaofunesta e T. cacao* na Murcha de Ceratocystis do cacaueiro, este trabalho apresenta uma abordagem comparativa dos diferentes genomas de *Ceratocystis e C. cacaofunesta*, que se

tornarão uma ferramenta chave para elucidar processos adaptativos desses patógenos de plantas, permitindo-nos entender melhor sua base molecular, reconhecimento de efetores e patogenicidade.

3. A repertoire of candidate effector proteins of the fungus *Ceratocystis cacaofunesta*

Artigo submetido para publicação no periódico "Scientific Reports", do grupo Nature.

Gabriela N. Ramos-Lizardo[1], Jonathan J. Mucherino[1] M., Eric R. G. R. Aguiar[1], Carlos P. Pirovani[1], Ronan X. Corrêa[1*]

[1]Departamento de Ciências Biológicas (DCB), Centro de Biotecnologia e Genética (CBG), Universidade Estadual de Santa Cruz (UESC), Ilhéus, BA, 45662-900, Brasil.

E-mail: ronanxc@uesc.br*

ABSTRACT. The genus *Ceratocystis* includes many phytopathogenic fungi that affect different plant species. One of these is *Ceratocystis cacaofunesta*, which is pathogenic to the cocoa tree and causes *Ceratocystis* wilt, a lethal disease for the crop. However, little is known about how these pathogenic fungi interact with their hosts. Knowledge and identification of effector proteins are essential to understanding this pathosystem. In this study, we annotated the *C. cacaofunesta*

genome once more. We analyzed the secretome and effectorome through *in silico* analysis, using the characteristics of the peptides as to the presence of signal peptide for secretion, absence of transmembrane domain, small size, and richness of cysteine residues. We identified 160 effector candidates in the *C. cacaofunesta* genome that could be classified as cytoplasmic (102) or apoplastic (58). Only 45 of these proteins displayed information associated with functional characteristics. Our database was created using a robust bioinformatics strategy, followed by manual curation, generating information on the functional proteins involved in plant interactions, including CAZymes, hydrolases, linases, and oxidoreductases. Comparison with other effector proteins from *Sordariomycetes* species revealed five groups of protein sequences homologous to *C. cacaofunesta.* These data provide a valuable resource for studying the infection mechanisms of these pathogens in their hosts.

Keywords: pathogenicity factors, phytopathogens, Vascular wilt, effectors, *Theobroma cacao.*

INTRODUCTION

Ceratocystis is one of many genera in the *Ceratocystidaceae* family (*Microascales, Sordariomycetes, Ascomycota*)[1]. The different species of this genus cause cankers and wilts in many economically important plant species[2]. For example, *Ceratocystis platani* causes severe wilt in plane (*Platanus*) trees in Europe, *Ceratocystis manginecans* produces mango tree wilt, and *Ceratocystis Fimbriata* sensu stricto is a sweet potato pathogen[3]. The species *Ceratocystis cacaofunesta* is specific to cocoa trees and causes wilt associated with cocoa tree mortality in

Central and South America[4]. Ceratocystis wilt of cacao was first described in 1918 in Ecuador and later found in Aragua, Venezuela, in the 1950s[5]. In Brazil, this pathogen was initially reported in the Amazon region. In the 1990s, *C. cacaofunesta* was introduced into the Southern Region of Bahia, one of Brazil's main cocoa-producing areas[2]. This disease has been responsible for losses of 20 to 30% in cocoa production in the state of Bahia[6].

As a necrotic fungus, *C. cacaofunesta* causes cellular death during colonization and obstructs the transport of water and nutrients in plants, turning them yellow and then brown before they wither and die[7]. Its reproduction can occur asexually, through vegetative propagation and conidia formation, as well as sexually[8]. Plant infection mainly occurs through injuries, such as cuts, incurred by tools during agricultural practices and crop management, e.g., thinning and pruning, and by the attack of Coleoptera, e.g., *Xyleborus sp. (Coleoptera-Scolytidae)*[5]. After the onset of symptoms, the disease is difficult to control since the pathogen quickly and irreversibly infects and colonizes the vascular system[9]. Using fungicides, phytosanitary techniques, and selecting resistant cocoa tree varieties have been considered the best strategies to control Ceratocystis wilt. However, it is difficult to control this disease due to the short period between the appearance of symptoms and the death of the plant[10]. Therefore, analyzing characteristics associated with pathogenicity and the evolution of pathogens is crucial to developing effective strategies for managing fungal diseases[11].

Phytopathogenic fungi mainly interact with their hosts through the secretion of protein effectors to promote pathogenesis[12]. These effectors are small molecules associated with an organism that alter the structure and function of the host cell[13]. These changes facilitate infection (virulence factors or toxins) and induce plant defenses (avirulence factors: Avr)[14]. Criteria for defining candidate-secreted effector proteins (CSEP) include fungal proteins with a signal peptide

for secretion, no transmembrane domain, no similarity to other obvious protein domains, small size, cysteine-rich, and mainly species-specific[15]. CSEPs can be classified according to their mode of release within the host. Thus, effectors are called extracellular if they are secreted into the apoplast or xylem of the host plant and cytoplasmic if they are translocated into host cells[16]. The prediction of these proteins is generally based on *in silico* methods[17].

Advances in DNA sequencing techniques and falling costs have fueled the publication of many fungal genomes, including many phytopathogenic fungi[11]. Analyzing these genomes allows the identification of effector proteins in different species of pathogenic fungi[18]. This enables us to understand their function, which is very important to develop technological strategies to reduce losses caused by plant diseases[17]. Carrying out *in silico* analyzes and functional characterization of these proteins is the first step to identifying colonization mechanisms in different host species, contributing to our knowledge of the effectors' biology and specific ways of action[14]. The present study provides a database of the protein effectorome of *Ceratocystis sp.*, enabling us to understand the fundamental role of these effectors in developing diseases. These findings may facilitate understanding the host-pathogen interactions involved in *Ceratocystis* wilt disease in cocoa, the underlying molecular mechanism, and the different infection strategies among different strains of *Ceratocystis.*

MATERIALS AND METHODS

Genome sequences

For this study, we selected the following five *Ceratocystis* genomes available in the National Center for Biotechnology Information's (NCBI) GenBank (**Supplementary table 11)**. The genome of *C. albifundus* strain CMW17620, GCA_000813685.1, sequenced by the University of Pretoria[19]; the genome of *C. cacaofunesta* strain C1593, GCA_002776505.1, sequenced by Universidade Estadual de Campinas[8]; the genome of *C. fimbriata* strain CBS 114723, GCA_000389695.3, sequenced by the Instituto de Biotecnologia Forestal y Agrícola FABI – FABI[3]; the genome of *C. manguinecans* strain CMW17570, GCA_000712455.1, also sequenced by the Instituto de Biotecnologia Forestal y Agrícola[20]; and the genome of *C. platani* strain CFO, GCA_000978885.1, sequenced by the University of Neuchâtel in 2015.

Different bioinformatics tools were used sequentially and complementarily in the multiple steps of *in silico* genomic analysis (Figure 1). The quality and integrity of the genomes were evaluated using Benchmarking Universal Single-Copy Orthologs version 5. 3. 2 (BUSCO v5. 3. 2.). For this purpose, BUSCO was configured using the Sordariomycetes lineage[21].

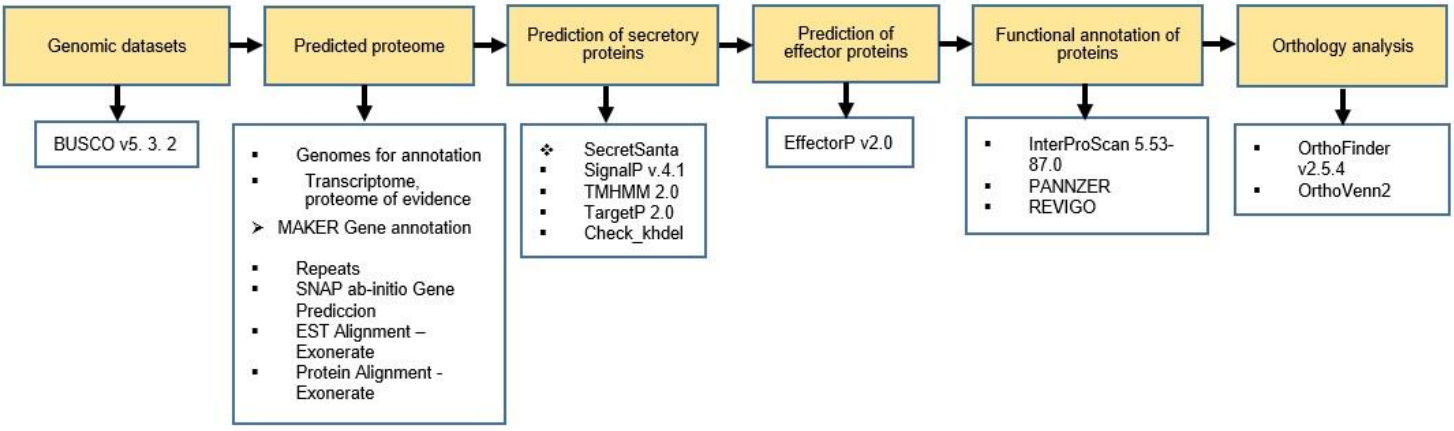

Figure 1. The workflow shows each of the steps for genome annotation and *in-silico* analysis of the secretome and effectorome of *Ceratocystis* species and the different bioinformatics tools used.

Structural annotation

The annotated genome of the species *C. cacaofunesta* is available via the NCBI[2.] ˙Still, it was annotated once more to obtain data that would meet the purposes of the present study, along with the species *C. albifundus* and *C. manguinecas*, which still lacked structural annotation (**Supplementary table 12)**. The genome assemblies of these three species were annotated using the MAKER 3.01.04 annotation pipeline[22]. A combination of *ab initio* and homology-based methods (transcriptome data and homology with known proteomes) was used for gene prediction. Before gene prediction, all sequences were masked with RepeatMasker[23]. For the *ab initio* prediction, two interactions were made using SNAP through the Hidden Markov Model (HMM), which is generally used for predicting protein-coding genes [24] [25]. For homology-based gene prediction, Exonerate[26] was used with the est2genome and protein2genome alignment models; for the est2genome model, evidence transcriptomes of the *C. platani* GCA_000978885.1 and *C. fimbriata* GCA_000389695.3 were used. For the protein2genome model, which permits the alignment of a protein sequence with the genomic DNA, the known protein sequences of each species were used, with the reference proteome identified by the ID: UP000034841 used for the *C. platani* species, and the proteome with the ID: UP000222788 used for *C. fimbriate*, both available on the Uniprot database.

Secretome and effectorome prediction

Secreted *in silico* proteins were characterized using the proteomes resulting from the structural annotation of the species *C. albifundus*, *C. cacaofunesta*, and *C. manguinecas*, as well as the reference proteomes of the species *C. platani* and *C. fimbriata*. This characterization was

performed using the SecretSanta package, which implements functions that permit prediction and analysis of the secretome using the R interface [27 28]. Secretome prediction requires multiple steps. One of these steps is the identification of the signal peptide, which was performed in the present study with the SignalP v.4.1 software that detects the presence of the SP signal peptide from amino acid sequences[29]. The TMHMM 2.0 software was used in the second step to identify sequences without TM transmembrane domains, using the SP output sequence file[30]. Subsequently, prediction of the extracellular location was performed using the TargetP 2.0 software. To achieve this, the output of the TM sequences was the input file for the TargetP. Finally, the TargetP 2.0 output sequence was compiled and scanned to identify signals of endoplasmic reticulum (ER) retention using the Check_khdel software[27]. The EffectorP v2.0 software was used to distinguish between cytoplasmic and apoplastic effectors[31].

Functional annotation

Functional characterization of effector proteins was performed using InterProScan 5.53-87.0, which analyzes the function of proteins, providing an integrative classification of protein sequences into families, as well as identifying functionally essential domains and conserved sites[32]. PANNZER software was used to predict the functional description, and gene ontology (GO) classes [33 34]. REVIGO was used to analyze the GO terms. It consists of a web server that summarizes long and unintelligible lists of GO terms by finding a representative subset of the terms, using a simple clustering algorithm based on measures of semantic similarity[35]. The simRel parameters were used for this, providing a measure of functional similarity to compare two GO terms with each other. In this case, the terms with the highest "singularity", the negative

or average similarity of a term with all other terms with a cut-off value of 0.7, were prioritized for p-values.

Analysis of orthologous gene families in Sordariomycetes

Comparison of orthology was performed by OrthoFinder v2.5.4, which is a fast, accurate, and comprehensive platform for genomic comparison. This platform makes it possible to find orthogroups and orthologs, infer rooted gene trees for all orthogroups, and identify all gene duplication events in those gene trees[36]. The reference proteomes of *C. platani* (ID: UP000034841) and *C. fimbriata* (ID; UP000222788) were used for this analysis, as well as the proteomes resulting from the structural annotation of *C. cacaofunesta, C. manguinecans,* and *C. albifundus*. Additionally, the proteomes of five related *Sordariomycetes* species were used: *Verticillium dahliae* (ID: UP000001611), *Verticillium albo-atrum* (ID: UP000008698), *Fusarium oxysporum* strain Fo5176 (ID: UP000002489), *Fusarium verticillioides* 7600 (ID: UP000009096), and *Fusarium graminearum* PH-1 (ID: 070000) (**Supplementary table 11)**. Comparison of the orthology of the *Ceratocystis* species was performed using OrthoVenn, which is a powerful web-based platform for the comparison and analysis of orthologous groups. These analyzes used proteins considered to be effector proteins[37].

RESULTS

A total of five *Ceratocystis* species were selected, considering the availability of the assembled genome and assessment of the genome assembly quality metric characteristics (**Table 1**). In the case of selected species with more than one assembled genome, only one of these was

chosen according to the level of integrity that the genome presented, always prioritizing the most complete available. Those selected were *C. albifundus* GCA_000813685.1, *C. cacaofunesta* GCA_002776505.1, *C. fimbriata* GCA_000389695.3, and *C. manguinecas* GCA_000712455.1, at the scaffold level, as well as *C. platani* GCA_000978885.1, at the contigs level.

Table 1. Comparative analysis of metrics referring to the genome and proteome of a lineage of each of the following species: *Ceratocystis albifundus, C. cacaofunesta, C. fimbriata, C. manguinecans,* and *C. platani*.

Species→	*C. albifundus*	*C. cacaofunesta*	*C. fimbriata*	*C. manguinecans*	*C. platani*
Strain→	CMW17620	C1593	CBS114723	CMW17570	CFO
Assembled genome size (Mb)	26.88	30.48	30.16	31.71	29.18
Number of Scaffolds	1,405	603	399	980	1,213
N50	20,627	54,222	23,089	30,399	77,580
Number of Contigs	2,894	1,442	2,524	2,295	1,213
GC content (%)	48.6	48.1	47.6	47.9	48.2
Predicted proteome	7,619	7,879	7,266	7,563	5,622

Quality analysis for each of the genomes was performed using the BUSCO software (**Table 2**). In this analysis, it was possible to observe genomic completeness of 94.3% for *C. cacaofunesta,* 94.3% for *C. fimbriata*, 94.1% for *C. manguinecans*, 93.7% for *C. albifundus*, and 94.2% for *C. platani.* Structural annotation of the genomes performed through the MAKER2 pipeline allowed us to predict 7879, 7619, and 7563 proteins for *C. cacaofunesta, C. albifundus*, and *C. manguinecans*, respectively (**Table 1**).

Table 2. Quality analysis of *Ceratocystis sp.* genomes; *C. albifundus, C. cacaofunesta, C. fimbriata, C. manguinecans,* and *C. platani*, results of the BUSCO categories (Complete (C), complete and single copy (S), complete and duplicated (D), fragmented (F), missing (M), n: gene number).

Species	BUSCO notation assessment result
C. albifundus	C:93.7% [S:93.6%, D:0.1%], F:0.4%, M:5.9%, n:3817
C. cacaofunesta	C:94.3% [S:94.2%, D:0.1%], F:0.3%, M:5.4%, n:3817
C. fimbriata	C:94.3% [S:94.1%, D:0.2%], F:0.2%, M:5.5%, n:3817
C. manguinecans	C:94.1% [S:93.9%, D:0.2%], F:0.4%, M:5.5%, n:3817
C. platani	C:94.2% [S:94.0%, D:0.2%], F:0.3%, M:5.5%, n:3817

Secretome and effectorome

The secretome of each *Ceratocystis* species was predicted from the proteome using the SecretSanta pipeline, which uses a series of combined software to predict and classify the secreted proteins. For this purpose, all 7619 proteins from *C. albifundus*, 7879 from *C. cacaofunesta*, 7266 from *C. fimbriata*, 7563 from *C. manguinecans* and 5622 from *C. platani*, were examined. Those that fit into one of the following four categories were considered to be secreted proteins: I, secreted proteins containing signal peptides; II, secreted proteins containing a signal peptide and lacking a transmembrane domain; III, secreted proteins containing a signal peptide, without a transmembrane domain and with a subcellular location of the secretory pathway; IV, secreted proteins containing a signal peptide without a transmembrane domain, with the subcellular location of the secretory pathway and without ER retention signal (**Figure 2**).

In total, our Bioinformatics strategy predicted 399 secreted proteins for *C. fimbriata*, 452 for *C. platani*, 451 for *C. cacaofunesta*, 351 for *C. albifundos* and 549 for *C. manguinecans*, representing 5.49%, 8.04%, 5.72%, 4.61% and 7.26% of all detected proteins, respectively (**Supplementary table 1)**. The effector proteins were identified using EffectorP by recovering

the secreted proteins that had been cut during one step of SecretSanta to obtain the complete sequences. Finally, the following quantities were predicted for each species: a total of 120 effectors belonging to *C. fimbriata*, consisting of 72 cytoplasmic and 48 apoplastic effectors; 144 effectors for *C. platani*, 85 cytoplasmic and 59 apoplastic; 160 effectors for *C. cacaofunesta*, with 102 cytoplasmic and 58 apoplastic effectors; 117 for *C. albifundus*, of which 79 were predicted as cytoplasmic effectors and 38 as apoplastic effectors, and 193 for *C. manguinecans*, with 126 cytoplasmic and 67 apoplastic effectors (**Figure 2, Supplementary table 2**).

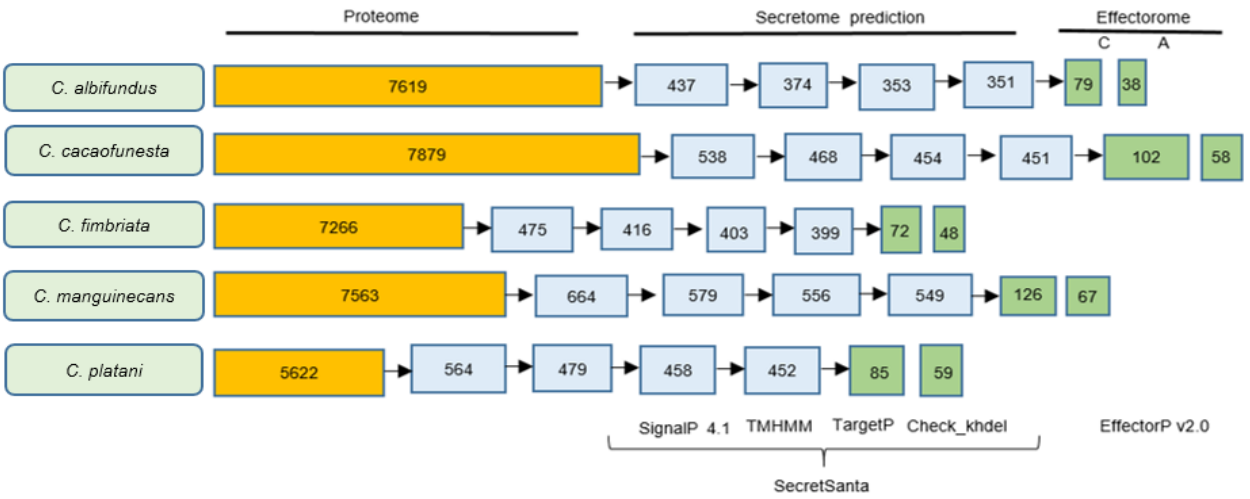

Figure 2. Pipeline for *in silico* characterization of secreted effector proteins. The figure shows the proteome, secretome, and effectorome of *Ceratocystis sp*. The first column contains the *Ceratocystis* species; *Ceratocystis albifundus, C. cacaofunesta, C. fimbriata, C. manguinecans,* and *C. platani*. Column two shows the amount of the proteome of each species, column three the amount of sequences that contain a signal peptide, column four shows the number of sequences with a signal peptide and without transmembrane domains, column five shows the number of sequences with a signal peptide, without transmembrane domains and with the subcellular location of the secretory pathway, column six shows the number of sequences with a signal peptide, without transmembrane domains, with the subcellular location of the secretory pathway and no retention signals, being the total secretome of

the different species and column seven contains the candidate effector proteins identified by the letters (C) cytoplasmic and (A) apoplastic.

Functional annotation and classification of effector proteins

Effector proteins identified with functional annotation were separated according to the biological processes, molecular function, and cellular component in which they are involved. Only some of the total number of effector proteins present in the different species have known functions, presenting 56 proteins for *C. platani*, 50 for *C. manguinecans*, 46 for *C. fimbriata*, 45 for *C. cacaofunesta* and 36 for *C. albifundus* **(Figure 3. a)**. In *C. cacaofunesta* 46 GO terms were related to biological processes, 44 in *C. manguinecans*, 44 in *C. fimbriata*, 38 in *C. platani* and 20 in *C. albifundus* **(Figure 3. b)**.

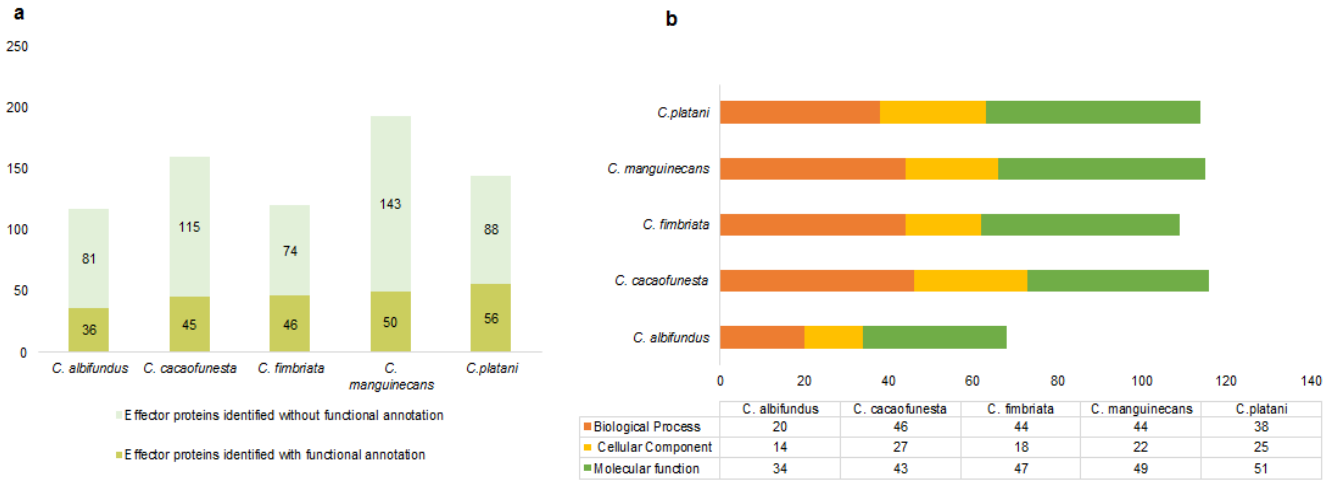

Figure 3. a) The bar graph shows in green color the total number of effector proteins identified with functional annotation in each *Ceratocystis* species and in blue color the total number of effector proteins that do not have functional annotation. **b)** The bar graph shows the total GO terms of each *Ceratocystis* species, divided according to the biological processes, molecular function, and cellular component in which they are involved.

The GO analyzes showed 13 GO terms common to the five *Ceratocystis* species in the "Biological Process" category, of which the most abundant terms were related to carbohydrate metabolic processes, proteolysis, lipid metabolic process, protein folding and cell wall organization. 11 GO terms were exclusive to *C. cacaofunesta*, with the highest values presented for the lipid catabolic process, glycerolipid metabolic process, endonucleolytic hydrolysis of the phosphodiester RNA bond **(Figure 4)**, 9 GO terms to *C. fimbriate*, with the highest values in the biosynthetic process of carbohydrates derivatives, cellular protein metabolic process, RNA biosynthetic process, and glycoprotein metabolic process, 4 GO terms for *C. manguinecans* involved in rRNA processing, RNA catabolic process, autolysis, and large ribosomal subunit assembly, 4 GO terms in *C. albifundus* related to the metabolic process of organic substances, the metabolic process of nitrogen compounds, the metabolic process of amines and peptidyl-tyrosine dephosphorylation, and 1 GO term specific to *C. platani* related to the insertion of a tail-anchored membrane protein in the ER membrane (**Supplementary table 3**).

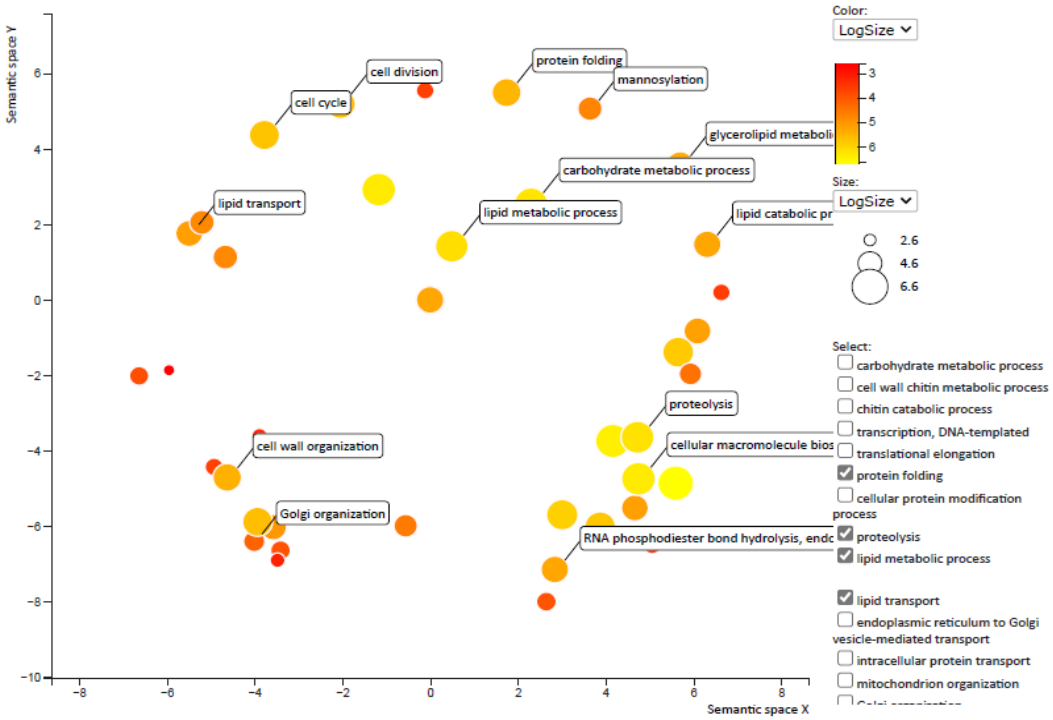

Figure 4. The scatter plot shows the GO representatives of the "biological process" category common to the five *Ceratocystis* species and the GO terms specific to the species *C. cacaofunesta*, in a two-dimensional space derived from the application of the multidimensional scale to a matrix of semantic similarities of the GO terms. The color of the bubble indicates the p-value given for each GO term (caption in the upper right corner); the size indicates the frequency of the GO term in the underlying GO database.

In the "Molecular Function" category, 51 GO terms were identified in *C. platani* related to molecular functions, 49 in *C. manguinecans*, 47 in *C. fimbriata*, 43 in *C. cacaofunesta*, and 34 in *C. albifundus* **(Figure 3. b)**. Of the GO terms found, 22 display a function in the five species of *Ceratocystis*, the GO terms that presented the highest values in the five species were related to hydrolase activity, catalytic activity, protein binding, zinc ion binding, and ATP binding. Other GO terms showed specific activity in some species, 5 GO terms were found in *C. cacaofunesta* belonging to 4 effector proteins related to nuclease activity, RNA-DNA hybrid ribonuclease, peroxidase activity, and lipid binding and transfer activity **(Figure 5).** Another 5 GO terms in *C. albifundus* showed activity with the proteins tyrosine phosphatase, N-acetyltransferase,

phosphoprotein phosphatase, primary amine oxidase, and quinone binding. 3 GO terms found in *C. fimbriata* are related to oxidoreductase activity, GTP binding, and amidophosphoribosyltransferase activity, 2 in *C. platan*i are for carbon lyase activity, and for sulfide and alkaline phosphatase activity, and 1 *C. manguinecans* specific GO term is for 3'-5' exonuclease activity (**Supplementary table 4**).

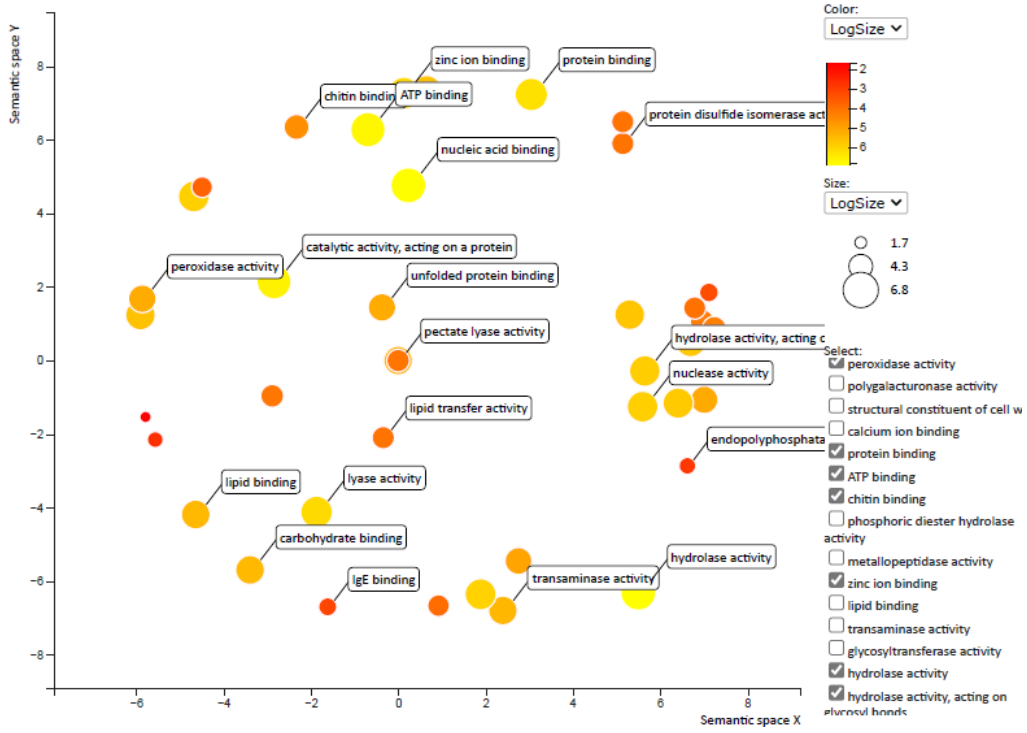

Figure 5. The scatter plot shows the GO representatives of the "molecular function" category common to the five species of *Ceratocystis* and the GO terms specific to the species *C. cacaofunesta*, in a two-dimensional space derived from the application of the multidimensional scale to a matrix of semantic similarities of the GO terms. The color of the bubble indicates the p-value given for each GO term (caption in the upper right corner); the size indicates the frequency of the GO term in the underlying GO database.

The GO results for the "Cellular Components" category showed 27 GO terms in *C. cacaofunesta*, 25 in *C. platani*, 22 in *C. manguinecans,* 18 in *C. fimbriata* and 14 in *C. albifundus* **(Figure 3. b).** In this category, 7 GO terms were found in all five species of *Ceratocystis*, related to the integral component of the membrane, extracellular region, large cytosolic ribosomal subunit, fungal cell wall, and the anchored component of the membrane. The species *C. cacaofunesta* had 3 specific GO terms for the mitochondrial inner membrane, respirasome, and the ERMES complex **(Figure 6)**, *C. albifundus* had 3 GO terms for the cytoplasmic vesicle membrane, Golgi membrane, and cell surface, and *C. manguinecans* and *C. platani* had 1 GO term each, related to the ribonuclease H2 complex and the GET complex, respectively (**Supplementary table 5**).

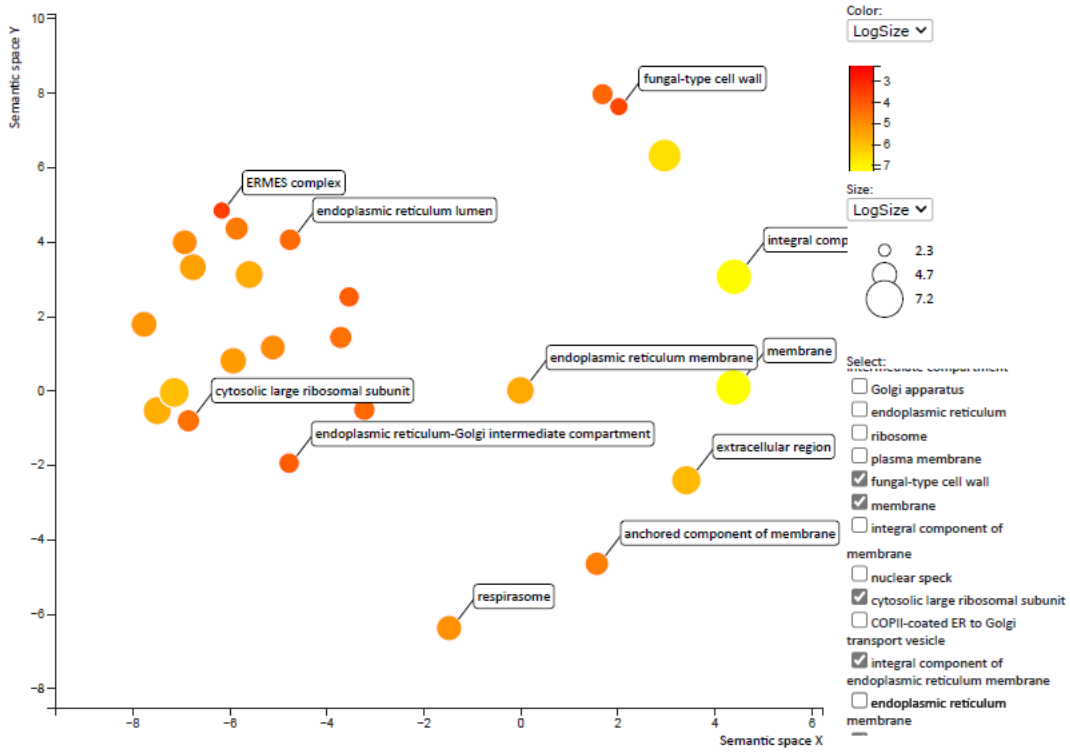

Figure 6. The scatter plot shows the GO representatives of the "cellular component" category common to the five species of *Ceratocystis* and the GO terms specific to the species *C. cacaofunesta*, in a two-dimensional space derived from the application of the multidimensional scale to a matrix of the semantic similarities of the GO terms. The color of the bubble indicates the p-value given for each GO term (caption in the upper right corner); the size indicates the frequency of the GO term in the underlying GO database.

Orthology analysis

OrthoFinder analysis of the different selected *Sordariomycetes* fungi enabled identification of groups of orthologous genes among the proteins of the ten species of *Sordariomycetes.* The proteomes of these ten species were assigned to a total of 14605 orthogroups, with the *Ceratocystis* species providing a smaller amount than the other proteomes (**Figure 7. a**). Comparison of the orthology of *Ceratocystis* effector proteins with the other *Sordariomycetes* species showed that only 207 of the total number of orthogroups found encode effector proteins within the five species of *Ceratocystis*; of these, 60 orthogroups were common

to all ten species of *Sordariomycetes*, with a total of 950 effector proteins. Only four of the 60 orthogroups found showed functional annotation against the Swiss-Prot database. One of these is known as Gibberella pigment protein 9 and demonstrates cellular component function acting in the extracellular region. Another demonstrates a cellular component function acting in the extracellular region and the host cell nucleus, presenting a molecular function, of endohydrolytic activity of (1->4)-beta-D-xylosidic bonds in xylans, and biological process, of effector-mediated modulation of the host's innate immune response by symbiont and xylan catalytic process.

The third orthogroup showed a protein related to autophagy 27, with a cellular component function in the Golgi membrane, cytoplasmic vesicular membrane, and mitochondrial membrane, related to biological processes such as protein transport and autophagy. The fourth showed an endoplasmic reticulum protein 25, also with a cellular component function in the Golgi membrane, endoplasmic reticulum membrane, and integral membrane component, performing biological processes of protein transport and vesicle-mediated transport. Another 8 orthogroups with a total of 84 proteins were found in all species except for *Verticillium alboatrum*. Likewise, 3 orthogroups (28 proteins) were found in most species, with the exception of *C. platani,* with one of these showing functional annotation related to cellular component function in the nucleus and molecular function related to nucleic acid binding and exonuclease activity acting on rRNA processing. Another 2 orthogroups with a total of 20 proteins were found in all species except *C. albifundus*, without any functional annotation (**Figure 7. b, Supplementary table 6**).

The comparative analysis between the species also identified 110 orthogroups present only in the five species of *Ceratocystis*; 38 of these were common to all five species, with a total of 342 effector proteins, 13 were found in all of the species, except *C. albifundus*; another 13 were present in *C. albifundus, C. manguinecans, C. cacaofunesta,* and *C. fimbriata,* 6 were found

in all species with the exception of *C. fimbriata*, 7 in *C. cacaofunesta* and *C. manguinecans*, 7 in *C. cacaofunesta*, *C. manguinecans* and *C. platani* , 7 in *C. cacaofunesta*, *C. manguinecans*, and *C. fimbriata*, 4 in *C. cacaofunesta*, and *C. albifundus*, 2 in *C. albifundus*, *C. cacaofunesta*, and *C. fimbriata*, 2 in *C. manguinecans* and *C. platani*, and 2 others present in all species except *C. cacaofunesta*. One orthogroup was found exclusively in the species *C. cacaofunesta*, 1 was present in *C. albifundus* and *C. fimbriata*, 1 in *C. cacaofunesta*, *C. fimbriata* and *C. platani*, and 1 in *C. albifundus*, *C. fimbriata* and *C. manguinecans* (**Figure 7. b**).

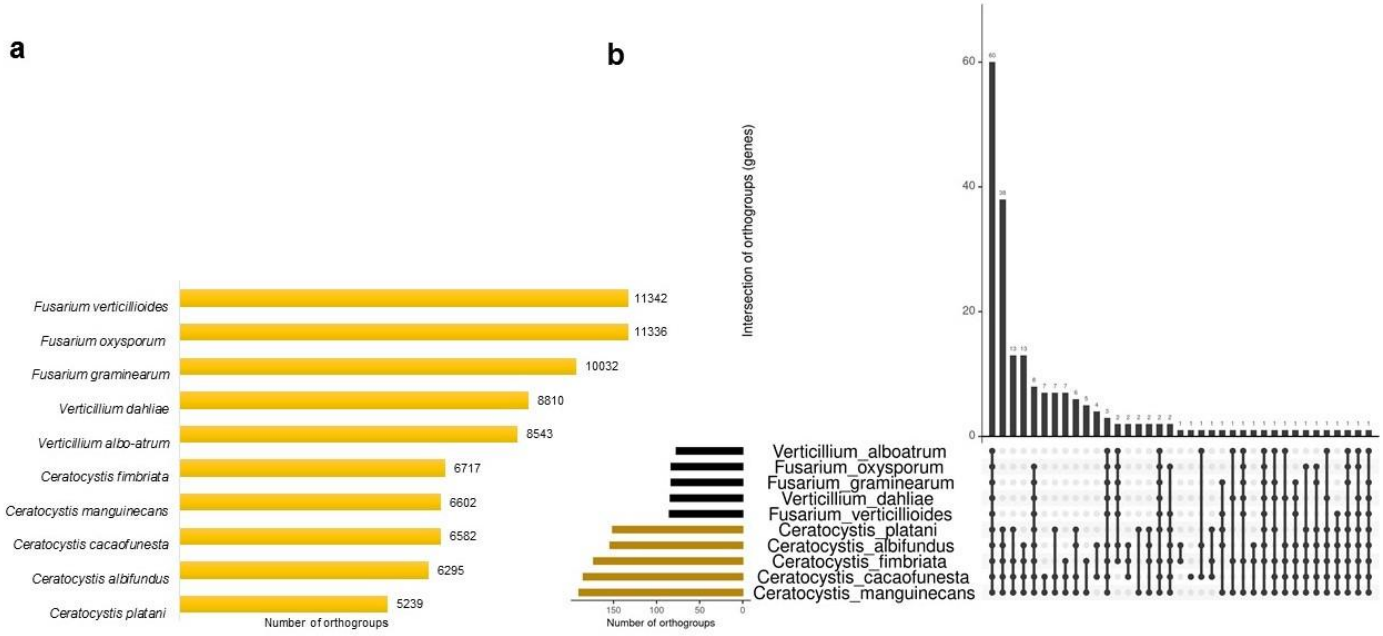

Figure 7. ***a*)** *Sordariomycete*s fungus species compared in this analysis, the graphic shows the species name and the total number of orthogroups in each proteome. ***b*)** Comparative genomic analysis of five target species of *Ceratocystis* and five other species of *Sordariomycetes; Verticillium dahlia*, *Verticillium albo-atrum*, *Fusarium oxysporum*, *Fusarium verticillioides*, *Fusarium graminearum*. The Upset plot of the protein cluster analysis shows in the bars on the upper side the number of orthogroups shared by the species highlighted in the black dots found on the lower side.

Comparison of the orthology of effector proteins secreted by *Ceratocystis* species was performed using Orthovenn, and sequence similarity was calculated with an electronic cut-off

value of 1e-2 and an inflation value of 1.5. The analysis shows 199 groups in total, among which 161 orthogroups showed more than one copy per species and 38 groups were single-copy genes. Among the 161 orthogroups found, 41 were shared by all five *Ceratocystis* species, with a total of 213 proteins. Of these 41 orthogroups, 12 had functional annotation (**Figure 8, Supplementary table 7**), related to the cellular component category with a function in the extracellular region, the category of molecular function related to ionic bonding, and the category of biological processes related to functions involved in the metabolic processing of carbohydrates, polysaccharides, macromolecules, peptides, cellular metabolic processes, and other processes involved in protein translation and folding, considered as the main effectors of *Ceratocystis*. Two exclusive groups were found for *C. cacaofunesta* with 4 proteins, 2 groups for *C. manguinecans* with 7 proteins, and one exclusive group for *C. platani* with 2 proteins, however, none of these presented functional annotation. The groups with single-copy genes showed 32 for *C. albifundus*, 27 for *C. manguinecan*s, 25 for *C. cacaofunesta*, 15 for *C. fimbriata*, and 15 for *C. platani* (**Supplementary table 8 and 9**).

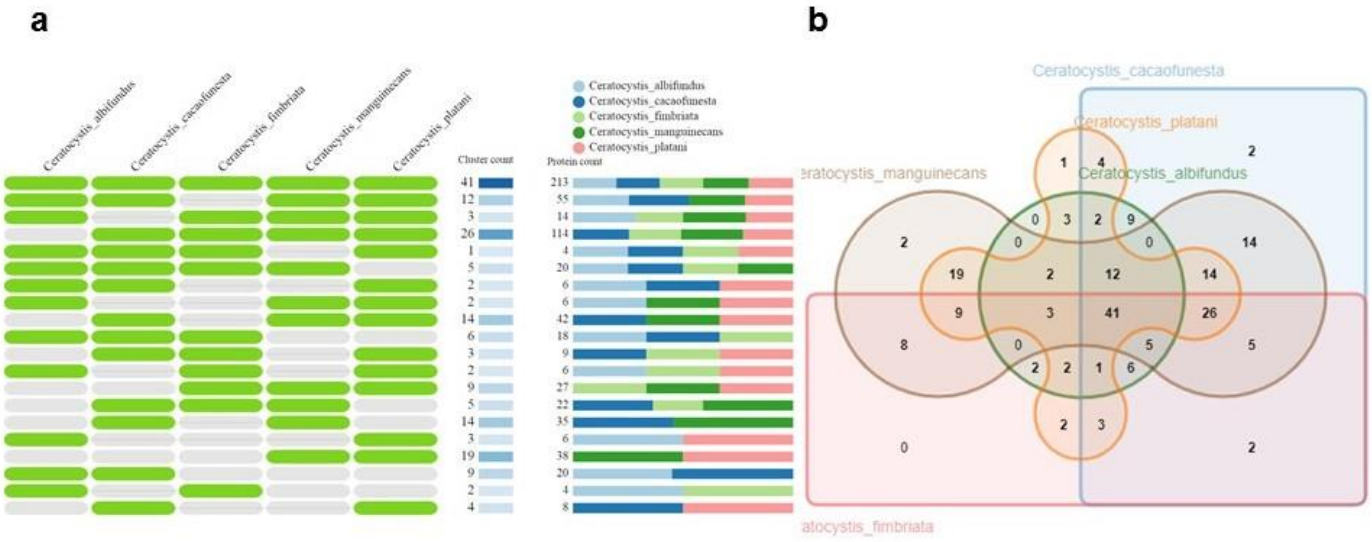

Figure 8. ***a)*** The overlap shows in the green columns the name of the species and in the lines the orthologous groups, the blue column shows the total number of orthologous groups contained in each row and the final column shows the total number of proteins for each set of orthologous groups. ***b)*** The Venn diagram shows shared and exclusive orthogroups among the five species of Ceratocystis.

DISCUSSION

Ceratocystis wilt infection is classified as an emerging disease due to the sudden increase in the number of plant species and geographic areas affected[8]. Like most known pathogenic microorganisms, vascular wilt fungi secrete proteins during host colonization[38]. Whole-genome sequencing analyses, as well as proteomics and bioinformatics analyses, have identified many of these proteins, including small proteins, usually rich in cysteine, necrosis-inducing proteins, and enzymes [39] [40]. Through analysis of the genome and expression of different species of *Ceratocystis*, it is possible to provide information about the molecular mechanisms of this important pathogen. The availability of the genomes of *C. platani, C. manguinecans, C.*

cacaofunesta, C. fimbriata, and *C. albifundus*, allowed us to investigate the candidate effector repertoire of each species and make an orthological comparison between them. So far, only a small number of genome studies on *Ceratocystis* species have been based on their pathogenicity [41] [42] [8].

Comparisons of the genome size of *C. cacaofunesta* and other *Ceratocystis* species showed that they have very similar genome sequences and genetic content, despite being pathogens of different hosts. Studies show that the genome of *C. cacaofunesta*, like the other genomes of *Ceratocystis sp.*, is small and has a low genetic content compared to other filamentous fungi [3] [8]. Our results for the functional annotation of different genomes revealed that proteins from *Ceratocystis* species show common functions. According to Molano et al (2018), this could be related to the short evolutionary distance between them. Usually, most pathogenic fungi produce secondary metabolites that induce plant cell death and secrete different phytotoxins that can disrupt host cells and induce the release of nutrients to facilitate colonization of host tissues[43]. In the GO category associated with the biological processes of the five species of *Ceratocystis*, it was possible to observe a greater expression in processes related to metabolic pathways, and secondary and primary metabolites, as well as encoding of hydrolytic enzymes, which play a crucial role in allowing fungal pathogens to establish successful infections[44].

Among the molecular functions of vascular wilt fungi are the secretion of different types of proteins and enzymes, which are used to manipulate their hosts and increase susceptibility to disease. These enzymes can destroy the plant's physical and chemical barriers, as well as induce defense responses. One of these physical barriers is the plant cell wall, which can be broken down by cell wall degrading enzymes (CWDE)[38]. The main components of plant cell wall polysaccharides are cellulose, hemicelluloses, and pectins, and the fungal enzymes involved in

the degradation of these plant polysaccharides are attributed to families of glycoside hydrolases, carbohydrate esterases, and polysaccharide lyases[45]. In the present study, it was possible to observe the presence of carbohydrate-active enzymes (CAZymes) produced by the pathogens in the genomes of the five *Ceratocystis* species analyzed. These play a central role in the synthesis and breakdown of the plant cell wall, becoming of great importance in fungal pathogens for successful penetration and infection in their hosts[46]. This result was consistent with the results obtained by Molano et al. (2018), who observed the presence of CAZymes enzyme activity in *Ceratocystis spp.*, but in small quantities compared to other *Sordariomycetes*, including non-pathogenic species[8].

Hydrolase enzymes were also found among the different genomes. The presence of hydrolases in the secretome of other pathogens has been associated with the degradation of plant cell wall polymers, permitting the penetration of fungi into host tissues, in addition to providing them with a source of water and nutrients[14]. Another enzyme involved in cell wall degradation, specifically plant pectin degradation, is the enzyme lyase. It was found in small amounts in the *Ceratocystis* genomes compared to other vascular wilt fungi and root pathogens, such as V*erticillium albo-atrum, Verticillium dahlia, N. haematococca,* and *Fusarium oxysporum*, where most tend to have more pectinases, which may be related to blockage or collapse of vascular bundles during disease development [46].

In other plant pathogenic fungi, oxidoreductases have been observed to participate in the elimination of reactive oxygen species and in plant pathogenicity [47]. The results showed genes in each genome that encode enzymes related to extracellular oxidative processes. The presence of these oxidoreductases in an oxidative reaction may play a crucial role in the degradation of fungal cellulose [48]. Regarding plant lignin degradation, the main lignin-degrading organisms are white

and brown rot fungi, but other organisms with this ability have been identified. These include the pathogen *Fusarium oxysporum*, considered a soft-rot fungus, which contains genes encoding lignin-degrading enzymes in low quantities compared to other rot-causing pathogens [49]. In the case of *Ceratocystis*, the only genome that showed the presence of the peroxidase enzyme involved in lignin degradation was the genome of *C. cacaofunesta.*

Effector proteins are one of several important virulence factors required for fungal colonization of plants and are produced and secreted by plant pathogens. Many of these proteins are translocated to the apoplast or cytoplasm, where they alter the host's defense responses to allow colonization by the pathogen [50]. The genome of *C. cacaofunesta* has a wide variety of proteins with effector characteristics. Within this category, we find proteins that have already been studied, such as the allergen Arg, and cyanovirin, which can provoke plant responses, and proteins possibly involved in the resistance to oxidative stress generated by the host[8]. Studies show the presence of cerato-platanin (CPPs) as effector proteins secreted by phytopathogenic fungi. These fungal proteins were found in some of the *Ceratocystis* species, such as *C. platani, C. fimbriata* and *C. cacaofunesta*, causing nutrient leakage and cell death in cocoa [51 52 8]. In other fungi such as *Botrytis cinerea* and *Sclerotinia sclerotiorum*, proteins of the cerato-platanin family were found to act as elicitors, inducing hypersensitive responses in plants. These responses are beneficial for the virulence of necrotrophic fungi [53]. Our *in silico* study shows the presence of similar sequences with the CPP effector protein present in the proteomes of each of the five *Ceratocystis* species. An alignment using Diamond was necessary for its identification since our methodology was not able to detect its presence in the secretome and the effectorome.

Our results show that most of the effectors found in the five species of *Ceratocystis* included genes with a common function shared between the strains. Others were strain-specific,

as observed by Molano et al. (2018), which explains their variation in pathogenicity/virulence. Many of these small-secreted proteins were predicted to be effectors, most of them correspond to uncharacterized proteins and therefore could be new species- or genus-specific effectors. The OrthoFinder results showed clusters of molecules with similar sequences, previously described as small-sized pathogenic/effector proteins, with less than 300 amino acids. These included the effector Vd424Y, belonging to the glycosidic hydrolase 11 family, which is secreted by *Verticillium dahliae* and located in the nucleus, contributing to the process of virulence and induction of cellular death in plants [54].

After opening up to proteins above the size of 300 amino acids, it was possible to observe other proteins considered to be effectors. This included identification of the autophagy-related protein 27, which has 348 aa and is known as a signaling effector of phosphatidylinositol 3-phosphate kinase VPS34, necessary for adequate vegetative growth, asexual/sexual reproduction, and total virulence in *Fusarium graminearum*. This protein is involved in the biosynthesis of deoxynivalenol, which is an important determinant of virulence [55]. A protein with 407 aa known to accumulate aurofusarin S biosynthesis was also identified, containing the fasciclin domain associated with fungal pigment biosynthesis in *Fusarium graminearum* [56]. This protein is part of one of the effector proteins found, characterized by being the main homologous sequences in the five *Ceratocystis* species. Other predicted proteins that were specifically identified within *Ceratocystis* species using OrthoVenn displayed "gene of unknown function" characteristics or were homologous with proteins larger than 300 aa, belonging to fungi that cause disease in humans. Other orthologous gene groups determined using OrthoFinder and OrthoVenn do not have predicted functions or GO annotations. Thus, the identification of these proteins provides

additional support and insight into possible infection mechanisms of *C. cacaofunesta* in relation to other fungi.

CONCLUSIONS

This study provides a repertoire of effector proteins generated from the genome of *C. cacaofunesta* and other publicly available *Ceratocystis* species, achieved through structural annotation of the genome and prediction of the secretome and effectorome. Our database provides information on functional proteins involved in plant interactions present in the five *Ceratocystis* species, including CAZymes, hydrolases, linases, and oxidoreductases. Some groups of effector proteins previously described in different *Sordariomycetes* species were also revealed in the present work, showing homology with *Ceratocystis* sequences. The identification of these proteins as the effector with Glyco_hydro_11 domain present in other vascular wilt-causing species such as Verticillium dahliae and Fusarium graminearum has been considered an important effector directed to the host nucleus, causing cell death and efficiently activating plant immunity, the presence of these similar sequences contributes to a better understanding of *C. cacaofunesta'*s infection processes, and generates a valuable resource for future studies of effector function. It also opens the possibility of identifying complementary R genes in the host that can be used to develop resistant varieties or cultivars.

Data availability

Supplementary file 12 contains the accession numbers of the genomes and proteomes analyzed in this article available in the NCBI and UniProt databases. All data generated during this study are included in this published article.

REFERENCES

1. Kanzi, A.M. *et al.* Phylogenomic incongruence in *Ceratocystis*: a clue to speciation. *BMC Genomics.* **21,** 362; 10.1186/s 12864-020-6772-0 (2020).
2. Ambrosio, A.B. *et al.* Global analyses of *Ceratocystis cacaofunesta* mitochondria: from genome to proteome**.** *BMC Genomics.* **14,** 91; 10.1186/1471-2164-14-91 (2013).
3. Wilken, P.M., Steenkamp E.T., Wingfield M.J., de Beer ZW, & Wingfield BD. Draft nuclear genome sequence for the plant pathogen, *Ceratocystis fimbriata*. *IMA Fungus.* **4,** 357–358; 10.5598/imafungus.2013.04.02.14 (2013).
4. Mbenoun, M. *et al.* Diversity and pathogenicity of the Ceratocystidaceae associated with cacao agroforests in Cameroon. *Plant Pathol*. **65,** 64-78; 10.1111/ppa.12400 (2016).
5. Magalhães, L.A., Magalhães, D.M., Luz, E.D., & Sodré, G.A. Diversidade patogênica de isolados de *Ceratocystis cacaofunesta* na região cacaueira da Bahia. *Agrotrópica*. **31,** (1): 27 – 36; 10.21757/0103-3816.2019v31n1p27-36 (2019).
6. Neves dos Santos, F., Magalhães, D.M.A., Luz, E.D.M.N., Eberlin, M.N. & Simionato, A.V.C. Metabolite mass spectrometry profiling of cacao genotypes reveals contrasting resistances to Ceratocystis cacaofunesta phytopathogen. *Wiley Analytical Science, electroforesis*. **42,** 2519-2527; 10.1002/elps.202100097 (2021).

7. Delgado-Ospina, J., Molina-Hernández, J.B., Chaves-López, C., Romanazzi, G. & Paparella, A. The Role of Fungi in the Cocoa Production Chain and the Challenge of Climate Change. *Journal of Fungi*. **7,** (3) 202; 10.3390/jof7030202 (2021)

8. Molano, E.P.L. *et al*. *Ceratocystis cacaofunesta* genome analysis reveals a large expansion of extracellular phosphatidylinositol-specific phospholipase-C genes (PI-PLC). *BMC Genomics.* **19,** (1):58; 10.1186/s12864-018-4440-4 (2018).

9. Dos Santos, E., Magalhães, D.M.A., Lopes, U.V. & Luz, E.D.M.N. Selection of cacao trees resistant to *Ceratocystis* wilt by inoculation in leaf discs and field. *Trop. plant pathol.* **46,** 536–544; 10.1007/s40858-021-00452-2 (2021).

10. Da Cruz, M. B. *et al*. Interference of aqueous and ethanolic solutions of Adiantum latifolium Lam. (Pteridaceae) leaves on in vitro *Ceratocystis cacaofunesta* mycelial growth. *Scientific communication, plant pathology, Arq. Inst. Biol*. **86,** 10.1590/1808-1657000192019 (2019).

11. Severn-Ellis, A. A. *et al*. Genome analysis of the broad host range necrotroph nalanthamala psidii highlights genes associated with virulence. *Frontiers in Plant Science*. **13,** 10.3389/fpls.2022.811152 (2022).

12. Wu, Y., Xie, L., Jiang, Y. & Li, T. Prediction of effector proteins and their implications in pathogenicity of phytopathogenic filamentous fungi: A review. *International Journal of Biological Macromolecules.* **206,** 188-202; 10.1016/j.ijbiomac.2022.02.133 (2022).

13. Ellis, J. G., Rafiqi, M., Gan, P., Chakrabarti, A. & Dodds, P. N. Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. *Current Opinion in Plant Biology.* **12,** (4) 399-405; 10.1016/j.pbi.2009.05.004 (2009).

14. Barbosa, C. S. *et al.* Genome sequence and effectorome of *Moniliophthora perniciosa* and *Moniliophthora roreri* subpopulations. *BMC Genómica.* **19,** 509; 10.1186/s12864-018-4875-7 (2018).

15. Humira, S., Rupesh K. D. & Richard R. B. Computational Prediction of Effector Proteins in Fungi: Opportunities and Challenges. *Frontiers in Plant Science.* **7,** 10.3389/fpls.2016.00126 (2016).

16. Huang, Z. *et al.* Predication of the effector proteins secreted by *Fusarium sacchari* using genomic analysis and heterogenous expression. *Journal of Fungi.* **8(1),** 59; 10.3390/jof8010059 (2022).

17. Jones, D. A., Bertazzoni, S., Turo, C. J., Syme, R. A. & Hane, J. K. Bioinformatic prediction of plant–pathogenicity effector proteins of fungi. *Current Opinion in Microbiology*. **46,** 43-49; 10.1016/j.mib.2018.01.017 (2018).

18. Mena, E., Garaycochea, S., Stewart, S., Montesano, M. & De León, P. I. Comparative genomics of plant pathogenic *Diaporthe* species and transcriptomics of *Diaporthe caulivora* during host infection reveal insights into pathogenic strategies of the genus. *BMC Genomics* **23, 175**; 10.1186/s12864-022-08413-y (2022).

19. Van der Nest M. A. *et al.* Draft genomes of Amanita jacksonii, *Ceratocystis albifundus, Fusarium circinatum, Huntiella omanensis, Leptographium procerum, Rutstroemia sydowiana y Sclerotinia echinophila*. *IMA Fungus.* **5, 472–486**; 10.5598/imafungus.2014.05.02.11 (2014).

20. Van der Nest M. A. *et al.* Draft genome sequences of *Diplodia sapinea*, *Ceratocystis manginecans*, and *Ceratocystis moniliformis*. *IMA Fungus.* **5, 135-140**; 10.5598/imafungus.2014.05.01.13 (2014).

21. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* **31(19),** 3210–3212; 10.1093/bioinformatics/btv351 (2015).

22. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics.* **12,** 491; 10.1186/1471-2105-12-491 (2011).

23. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Currrent Protocols Bioinformatics.* **25,** 4.10.1-4.10.14; 10.1002/0471250953.bi0410s25 (2009).

24. Bromberg, Y. & Rost, B. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res*. **35 (11),** 3823-3835; 10.1093/nar/gkm238 (2007).

25. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics.* **5,** 59 ;10.1186/1471-2105-5-59 (2004).

26. Slater, G.S.C.& Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformática.* **6,** 31 ; 10.1186/1471-2105-6-31 (2005).

27. Gogleva, A., Drost, H. G. & Schornack, S. SecretSanta: flexible pipelines for functional secretome prediction. *Bioinformatics.* **34(13),** 2295–2296; 10.1093/bioinformatics/bty088 (2018).

28. Giorgi, F.M., Ceraolo, C. & Mercatelli, D. The R Language: An Engine for Bioinformatics and Data Science. *Life (Basel, Switzerland).* **12(5),** 648; 10.3390/life12050648 (2022).

29. Nielsen, H. Predicting Secretory Proteins with SignalP. *Methods in molecular biology.* **1611,** 59–73; 10.1007/978-1-4939-7015-5_6 (2017).

30. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*. **305(3),** 567-580; 10.1006/jmbi.2000.4315 (2001).

31. Sperschneider, J. & Dodds, P.N. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. *Mol Plant Microbe Interact*. **35(2),** 146-156; 10.1094/MPMI-08-21-0201-R (2022).

32. Blum, M. *et al.* The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research.* **49(D1),** D344–D354; 10.1093/nar/gkaa977 (2021).

33. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic acids research*. **47(D1),** D330-D338; doi: 10.1093/nar/gky1055 (2019).

34. Toronen, P. & Holm, L. PANNZER - a practical tool for protein function prediction. *Protein science: a publication of the Protein Society*. **31 (1),** 118– 128; 10.1002/pro.4193 (2022).

35. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One.* **6(7),** e21800; 10.1371/journal.pone.0021800 (2011).

36. Emms, D.M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20,** 238; 10.1186/s13059-019-1832-y (2019).

37. Xu, L. *et al.* OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res*. **47(W1),** W52-W58; 10.1093/nar/gkz333 (2019).

38. De Sain, M. & Rep, M. The Role of Pathogen-Secreted Proteins in Fungal Vascular Wilt Diseases. *Int. J. Mol. Sci.* **16,** 23970-23993. https://doi.org/10.3390/ijms161023970 (2015).

39. Brown, N.A., Antoniw, J. & Hammond-Kosack K. E. The Predicted Secretome of the Plant Pathogenic Fungus *Fusarium graminearum*: A Refined Comparative Analysis. *Plos one*. **7(4),** e33731; 10.1371/journal.pone.0033731 (2012).

40. Yan, L. *et al.* Genome sequencing and comparative genomic analysis of highly and weakly aggressive strains of *Sclerotium rolfsii*, the causal agent of peanut stem rot. *BMC Genomics*. **22,** 276; 10.1186/s12864-021-07534-0 (2021).

41. Fourie, A. *et al.* QTL mapping of mycelial growth and aggressiveness to distinct hosts in *Ceratocystis* pathogens. *Fungal Genetics and Biology*. **131,** 103242; 10.1016/j.fgb.2019.103242 (2019).

42. Fourie, A. *et al.* Genome comparisons suggest an association between *Ceratocystis* host adaptations and effector clusters in unique transposable element families. *Fungal Genetics and Biology.* **143,** 103433; 10.1016/j.fgb.2020.103433 (2020).

43. Stergiopoulos, I., Collemare, J., Mehrabi R. & De Wit, P.J. Phytotoxic secondary metabolites and peptides produced by plant pathogenic Dothideomycete fungi. *FEMS Microbiol Rev.* **37,** 67-93; 10.1111/j.1574-6976.2012.00349.x (2013).

44. Nagel, J.H., Wingfield, M.J. & Slippers, B. Increased abundance of secreted hydrolytic enzymes and secondary metabolite gene clusters define the genomes of latent plant pathogens in the *Botryosphaeriaceae*. *BMC Genomics*. **22,** 589; 10.1186/s12864-021-07902-w (2021).

45. Van den Brink, J., & de Vries, R.P. Fungal enzyme sets for plant polysaccharide degradation. *Appl Microbiol Biotechnol.* **91,** 1477–1492; 10.1007/s00253-011-3473-2 (2011).

46. Zhao, Z., Liu, H., Wang, C.& Xu, J. R. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics.* **14,** 274; 10.1186/1471-2164-14-274 (2013).

47. Kim, K. T. *et al*. Kingdom-Wide Analysis of Fungal Small Secreted Proteins (SSPs) Reveals their Potential Role in Host Association. *Frontiers in Plant Science.* **7,** 10.3389/fpls.2016.00186 (2016).

48. Takeda, K. *et al.* Characterization of a Novel PQQ-Dependent Quinohemoprotein Pyranose Dehydrogenase from Coprinopsis cinerea Classified into Auxiliary Activities Family 12 in Carbohydrate-Active Enzymes. *Plos one.* **10(2),** e0115722; 10.1371/journal.pone.0115722 (2015).

49. Kumar, A., Kameshwar, S. & Qin, W. Comparative study of genome-wide plant biomass-degrading CAZymes in white rot, brown rot and soft rot fungi**.** *Mycology*. 9(2), 93-105; 10.1080/21501203.2017.1419296 (2018).

50. Boufleur, T.R. *et al.* Identification and comparison of *Colletotrichum* secreted effector candidates reveal two independent lineages pathogenic to soybean. *Pathogens*. **10(11),** 1520; 10.3390/pathogens10111520 (2021).

51. Meinhardt, L.W. *et al.* Genome and secretome analysis of the hemibiotrophic fungal pathogen, Moniliophthora roreri, which causes frosty pod rot disease of cacao: mechanisms of the biotrophic and necrotrophic phases. *BMC Genomics*. **15,** 164; 10.1186/1471-2164-15-164 (2014).

52. Baccelli I, Lombardi L, Luti S, Bernardi R, Picciarelli P, et al. Cerato-Platanin Induces Resistance in *Arabidopsis* Leaves through Stomatal Perception, Overexpression of Salicylic Acid- and Ethylene-Signalling Genes and Camalexin Biosynthesis. *Plos one.* **9(6),** e100959; 10.1371/journal.pone.0100959 (2014).

53. Tanaka, S. & Kahmann, R. Cell wall–associated effectors of plant-colonizing fungi. *Mycologia.*113(2), 247-260; 10.1080/00275514.2020.1831293 (2021).

54. Liu, L. *et al. Verticillium dahlia* secreted protein Vd424Y is required for full virulence, targets the nucleus of plant cells, and induces cell death. *Molecular Plant Pathology.* **22,** 1109–1120; 10.1111/mpp.13100 (2021).

55. Lv, W. *et al.* Genome-wide functional analysis reveals that autophagy is necessary for growth, sporulation, deoxynivalenol production and virulence in *Fusarium graminearum*. Sci Rep. **7,** 11062; 10.1038/s41598-017-11640-z (2017).

56. Frandsen, R. J. N. *et al.* Two novel classes of enzymes are required for the biosynthesis of aurofusarin in *Fusarium graminearum*. *Journal of Biological Chemistry.* **286(12),** 10419-10428; 10.1074/jbc. M110.179853 (2011).

Author Contribution Statement

G.N.R.L: Conceptualization, methodology, research, software, data acquisition and processing, writing: original draft. R.X.C: Conceptualization, methodology, writing: proofreading and editing. J.J.M: Methodology, research, software, data acquisition and processing, proofreading, and editing. C.P.P: Conceptualization, methodology, review, and editing. E.R.G.R: Conceptualization, methodology, review, and editing. All authors have read and approved the final version of the manuscript.

Author Information

Authors and affiliations

Departamento de Ciências Biológicas (DCB), Centro de Biotecnologia e Genética (CBG), Universidade Estadual de Santa Cruz (UESC), Ilhéus, BA, 45662-900, Brasil.

Gabriela N. Ramos-Lizardo, Jonathan J. Mucherino, Eric R. G. R. Aguiar, Carlos P. Pirovani e Ronan X. Corrêa.

Author for correspondence

Ronan X. Corrêa, ronanxc@uesc.br

Ethical Statements

Conflict of interest

The authors have no conflicts of interest to declare relevant to this article's content.

4. Conclusões gerais e perspectivas

Este estudo fornece a anotação estrutural do genoma de três espécies do gênero *Ceratocystis; C. albifundus, C. cacaofunesta e C. manguinecans*, uma listagem do repertório de proteínas secretadas candidatas a efetoras de *C. cacaofunesta* e outras espécies de *Ceratocystis*, assim como um repertório de proteínas efetoras, identificando aquelas comuns entre as espécies de *Ceratocystis* deste estudo e aquelas exclusivas a cada espécie.

Nossa base de dados proporciona informação de proteínas funcionais envolvidas nas interações das plantas, incluindo CAZymes, hidrolases, linases e oxidoredutases, presentes nas cinco espécies de *Ceratocystis*. A identificação de proteínas com sequências homólogas como os efetores apresentados, encontrados em outras espécies causadoras de murcha vascular como *Verticillium dahliae e Fusarium graminearum*, contribui a uma melhor compreensão dos processos de infecção por *C. cacaofunesta.*

A presença de sequências similares e o repertório de proteínas efetoras secretadas candidatas de *C. cacaofunesta*, gera um recurso valioso para estudos futuros da função do efetor, abre a possibilidade de identificar genes R complementares no hospedeiro que podem ser usados para desenvolver variedades ou cultivares resistentes. As perspectivas de pesquisa que se seguem ao presente trabalho são realizar estudos do transcriptoma mediante a avaliação da expressão dos efetores candidatos, preditos computacionalmente, e fazer uma caracterização funcional tendo como prioridade aquelas proteínas secretadas candidatas a efetores exclusivos e aquelas proteínas que careciam de função conhecida.

5. Referências complementares

AMBROSIO, A.B.*et al.* Global analyses of *Ceratocystis cacaofunesta* mitochondria: from genome to proteome**. BMC Genomics**.**14,** 91; (2013). doi.org/10.1186/1471-2164-14-91

BARBOSA, S. *et al*. Genome sequence and effectorome of *Moniliophthora perniciosa and Moniliophthora roreri* subpopulations. **BMC Genomics.** 2018 Jul 3;19(1):509. http:// doi.org/10.1186/s12864-018-4875-7.

BECKERSON, W. C. *et al.* "Cause and Effectors: Whole-Genome Comparisons Reveal Shared but Rapidly Evolving Effector Sets among Host-Specific Plant-Castrating Fungi." **MBio vol.** 10,6 e02391-19. 5 nov. 2019. https://doi.org 10.1128/mBio.02391-19

BORGES, R. C. F. **Aspectos epidemiológicos e análise transcriptômica da interação teca (*Tectona grandis) -Ceratocystis fimbriata* e caracterização de novos patógenos afetando o cultivo de teca no Brasil**. 2018. iv, 209 f., il. Tese (Doutorado em Fitopatologia). Universidade de Brasília, Brasília, 2018.

CABRERA, O.G. *et al*. *Ceratocystis* Wilt Pathogens: History and Biology—Highlighting *C. cacaofunesta*, the Causal Agent of Wilt Disease of Cacao. In: Bailey, B., Meinhardt, L. (eds) Cacao Diseases. **Springer, Cham.** doi.org/10.1007/978-3-319-24789-2_12. (2016).

DELGADO-OSPINA J. *et al.* The Role of Fungi in the Cocoa Production Chain and the Challenge of Climate Change. **Journal of Fungi.** 2021; 7(3):202. doi.org/10.3390/jof7030202

DHARMARAJ K. *et al.* A New Real-Time PCR Assay for Detecting Fungi in Genus *Ceratocystis* . **Plant Disease**. 106:2, 661-668. https://doi.org/10.1094/PDIS-08-21-1639-RE. (2022).

FERREIRA, R. *et al*. Identification in sílico and confirmation via RT-qPCR of candidate resistance genes linked to *Ceratcystis* wilt in cacao. **Agrotrópica** 32(1): 21-30. 2020. doi.org/10.21757/0103-3816.2020v32n1p21-30

FERNANDES. L.S. *et.al*. Mapping of a Major QTL for *Ceratocystis* Wilt Disease in an F1 Population of *Theobroma cacao*. **Front. Plant Sci**. 14 February 2018. https://doi.org/10.3389/fpls.2018.00155

HOLANDA, S. M. E. *et al*. Aplicação de aprendizado de máquina profundo para detecção por imagens de doenças em frutos do cacaueiro. **International Journal of Development Research**. Vol. 11, Issue, 05, pp. 47378-47384, May, 2021. https://doi.org/10.37118/ijdr.22009.05.2021

INTERNATIONAL COCOA ORGANIZATION **(ICCO)** Quarterly Bulletin of Cocoa Statistics. Available online at: https://www.icco.org/about-us/icconews/ 372-may 2017-quarterly-bulletin-of-cocoa-statistics.html.

KANZI, A.M. *et al.* Phylogenomic incongruence in *Ceratocystis* : a clue to speciation.**BMC Genomics**. 21**,** 362; (2020). doi.org/10.1186/s 12864-020-6772-0

MAGALHÃES, L.A. *et al.* Diversidade patogênica de isolados de *Ceratocystis cacaofunesta* na região cacaueira da Bahia. *Agrotrópica*. **31,** (1):27–36; (2019). doi.org/10.21757/0103-3816.2019v31n1p27-36

MALHEIRO, D.L.V. **Análise do perfil proteômico e estudo de secretoma de *Ceratocystis cacaofunesta* induzido com fragmentos de ramo de *Theobroma cacao***. Dissertação (mestrado em Genética e Biologia Molecular) Universidade Estadual Santa Cruz. Ilhéus, Bahia, Brasil. 2015.

MODA, R.L.*et al.* Cenário econômico do mercado de cacau e chocolate: oportunidades para a cacauicultura brasileira. **Brazilian Journal of Development**. Vol 5, No 10 (2019) https://doi.org/10.34117/bjdv5n10-281

MOLANO, E. et al. *Ceratocystis cacaofunesta* genome analysis reveals a large expansion of extracellular phosphatidylinositol-specific phospholipase-C genes (PI-PLC). **BMC Genomics.** 2018 Jan 17;19(1):58. doi.org 10.1186/s12864-018-4440-4

MOLANO, E. P. L. **Análises do genoma e transcriptoma do fungo *Ceratocystis cacaofunesta*, visando a identificação de possíveis alvos para o controle da doença mal do facão**. Tese (Doutorado em Genética e Biologia Molecular) Universidade Estadual de Campinas, Instituto de Biologia. 2019.

MORA-OCAMPO, I. Y. **Padronização de análise proteômica e atividade enzimática em nível de xilema visando estudos de interação entre *Ceratocystis cacaofunesta* e clones de cacau.** Dissertação (mestrado em Genética e Biologia Molecular) Universidade Estadual Santa Cruz Ilhéus, Bahia, Brasil.76f.: Il 2016

NGOU B.P.M. *et al.* Thirty years of resistance: Zig-zag through the plant immune system, **The Plant Cell**, Volume 34, Issue 5, May 2022, Pages 1447–1478, https://doi.org/10.1093/plcell/koac041

PALADINES-REZABALA, A. *et al*. (2022). Advances in understanding of the interaction between *Ceratocystis cacaofunesta* and Xyleborus ferrugineus(Coleoptera: Curculionidae: Scolytinae) on cocoa trees. **Scientia Agropecuaria**. 13. 43-52. https://doi.org/10.17268/sci.agropecu.2022.004

PRADHAN, A.*et al*. Fungal effectors, the double edge sword of phytopathogens. **Current Genetics** vol.67, 27–40 (2021). https://doi.org/10.1007/s00294-020-01118-3

SAIN, M.; REP M. The Role of Pathogen-Secreted Proteins in Fungal Vascular Wilt Diseases. **International Journal of Molecular Sciences**. 2015; 16(10):23970-23993. https://doi.org/10.3390/ijms161023970

SANTOS, R. M. F. **Estudo genético, molecular e histopatológico da interação cacau-*Ceratocystis cacaofunesta***. Dissertação (mestrado em Genética e Biologia Molecular) Universidade Estadual Santa Cruz. Ilhéus, Bahia, Brasil. 2012

SRINIVAS, C. *et al*. *Fusarium oxysporum f. sp. lycopersici* causal agent of vascular wilt disease of tomato: Biology to diversity– A review. **Saudi Journal of Biological Sciences.** Volume 26, Issue 7, November 2019, Pages 1315-1324. https://doi.org/10.1016/j.sjbs.2019.06.002

SILVA, S. D.V.M. *et al*. Reação de genótipos de cacaueiros a isolados de *Ceratocystis cacaofunesta*. **Fitopatologia Brasileira** [online]. 2007, v. 32, n. 6 pp. 504-506. https://doi.org/10.1590/S0100-41582007000600009.

SIVIERO, A. *et al.* Doenças em cacaueiro e cupuaçuzeiro no Acre. **Agrotrópica** 34(2): 159-164. 2022. Centro de Pesquisas do Cacau, Ilhéus, Bahia, Brasil. https://doi.org/10.21757/0103-3816.2022v34n2p159-164

DE MANDAL, S.; JEON, J. Nuclear Effectors in Plant Pathogenic Fungi, **Mycobiology**, https://doi.org10.1080/12298093.2022.2118928. (2022)

SPERSCHNEIDER, J. *et al*. EffectorP: predicting fungal effector proteins from secretomes using machine learning. **New Phytologist** vol., 210: 743-761.https://doi.org/10.1111/nph.13794. (2016)

SPERSCHNEIDER, J.; DODDS, P.N. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. **Mol Plant Microbe Interact.** 35(2), 146-156; https://doi.org10.1094/MPMI-08-21-0201-R (2022).

TARIQJAVEED, M. *et al*. Versatile effectors of phytopathogenic fungi target host immunity. **Journal of integrative plant biology**,.63:1856–1873. (2021). https://doi.org10.1111/jipb.13162

WANG, Y. *et al.* Comparative genome analysis of plant ascomycete fungal pathogens with different lifestyles reveals distinctive virulence strategies. **BMC Genomics** 23, 34 (2022). https://doi.org/10.1186/s12864-021-08165-1

WILKEN, P.M. *et al.* Draft nuclear genome sequence for the plant pathogen, *Ceratocystis fimbriata*. ***IMA Fungus.* 4,** 357–358; https://doi.org10.5598/imafungus.2013.04.02.14 (2013).

YU, X. *et al*. From Chaos to Harmony: Responses and Signaling upon Microbial Pattern Recognition. **Annual review of phytopathology.** 2017 Aug 4; 55:109-137. https://doi.org10.1146/annurev-phyto-080516-035649.

ZHANG, S. *et al.* Action Mechanisms of Effectors in Plant-Pathogen Interaction. **Int. J. Mol. Sci**.**2022**, *23*, 6758. https://doi.org/10.3390/ijms23126758