**VALÉRIA CRISTINA HOLTMAN**

# IDENTIFICATION AND CHARACTERIZATION OF CFEM PROTEINS FROM *Phakopsora pachyrhizi,* THE ASIAN SOYBEAN RUST FUNGUS

Dissertation submitted to the Applied Biochemistry Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

Adviser: Luciano Gomes Fietto

Co-adviser: Sérgio H. Brommonschenkel

**VIÇOSA - MINAS GERAIS**
**2023**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa – Campus Viçosa**

T
H758i
2023

Holtman, Valéria Cristina, 1996-
Identificação e caracterização de proteínas CFEM do fungo *Phakopsora pachyrhizi*, agente etiológico da ferrugem asiática da soja / Valéria Cristina Holtman. - Viçosa, MG, 2023.
1 dissertação eletrônico (72 f.): il. (algumas color.).

Texto em inglês.
Orientador: Luciano Gomes Fietto
Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Bioquímica e Biologia Molecular, 2023.
Referências bibliográficas: f. 62-72.
DOI: https://doi.org/10.47328/ufvbbt.2023.168
Modo de acesso: World Wide Web.

1. Soja - Doenças e pragas; 2. *Phakopsora pachyrhizi*; 3. Ferrugem da soja (Doenças) - Identificação; I. Fietto, Luciano Gomes II. Universidade Federal de Viçosa.
Departamento de Bioquímica e Biologia Molecular. Programa de Pós-Graduação em Bioquímica Aplicada III. Título

CDD 22. ed. 634.3494

Bibliotecário(a) responsável: ALICE REGINA PINTO PIRES CRB-6/2523

**VALÉRIA CRISTINA HOLTMAN**

# IDENTIFICATION AND CHARACTERIZATION OF CFEM PROTEINS FROM *Phakopsora pachyrhizi,* THE ASIAN SOYBEAN RUST FUNGUS

Dissertation submitted to the Applied Biochemistry Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

APPROVED: February 14, 2023.

Assent:

Documento assinado digitalmente
gov.br
VALERIA CRISTINA HOLTMAN
Data: 11/04/2023 10:08:10-0300
Verifique em https://validar.iti.gov.br

Valéria Cristina Holtman
Author

Documento assinado digitalmente
gov.br
LUCIANO GOMES FIETTO
Data: 12/04/2023 15:02:51-0300
Verifique em https://validar.iti.gov.br

Luciano Gomes Fietto
Adviser

## ACKNOWLEDGEMENTS

To the Federal University of Viçosa, for the opportunity to complete the postgraduate course.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

## ABSTRACT

HOLTMAN, Valéria Cristina, M.Sc., Universidade Federal de Viçosa, February, 2023. **Identification and characterization of CFEM Proteins from *Phakopsora pachyrhizi*, the Asian soybean rust fungus**. Adviser: Luciano Gomes Fietto. Co-adviser: Sérgio Herminio Brommosnchenkel.

Asian soybean rust (ASR) caused by the fungus *Phakopsora pachyrhizi* is one of the main fungal diseases of soybean, which can cause losses of up to 95% of production under favorable environmental conditions and abundant inoculum. This pathogen is an obligate, biotrophic parasite that during the infection process secretes several effector proteins that suppress plant defense responses and promotes parasitism. One of the characteristics of fungal and oomycete effector proteins is the high content of cysteine residues (>3%). CFEM proteins (Common in Several Fungal Extracellular Membrane Proteins) are proteins secreted by different fungi species with a protein domain with 8 cysteine residues in conserved positions and diverse cellular functions. The best characterized CFEM proteins are Csa2, Rbt5, and Pga7 from the fungus *Candida albicans*, that are involved in a cascade to capture and absorb heme-iron from the mammalian host. The objectives of this study were to identify and characterize all CFEM proteins present in the repertoire of *P. pachyrhizi* fungus and to understand the biological function of these proteins as membrane component, immunity suppressor or iron capture in the pathogen life cycle. *In silico* analysis showed that *P. pachyrhizi* has 10 CFEM proteins (PpCFEM1-PpCFEM10), six of which are predicted to be secreted soluble proteins (PpCFEM1, 5, 7, 8, 9, 10) and four (PpCFEM2, 3, 4, 6) predicted as extracellular proteins with Glycosylphosphatidylinositol (GPI) docking site. Seven of the CFEM protein-coding genes (*PpCFEM1-7*) were expressed in the period 0 to 96 hours post-inoculation; *PpCFEM2*, *5*, and *6* showed higher expression in the time of zero hours post-inoculation (hpi), and *PpCFEM3* and *4* are expressed at 24 hpi. PpCFEM1, 3, 5, and 7 were identified in the nucleus and cytoplasm in subcellular localization assays, PpCFEM2 and PpCFEM6 proteins were identified in the cell membrane, and PpCFEM4 showed no defined subcellular localization. PpCFEM1, 4, 5, and 7 were able to suppress PTI (immunity triggered by recognition of PAPMPs), whereas only the PpCFEM4 protein was also able to suppress ETI (immunity triggered by recognition of effectors) in *Nicotiana benthamiana* leaves. The PpCFEM proteins

did not present homology with the *C. albicans* CFEM proteins that participate in the heme-iron capture and absorption mechanism, and tertiary structure conformation analyses indicated that the PpCFEM proteins did not present the necessary structure for the heme-iron coordination as present in *C. albicans* CFEM proteins. Our findings suggest that the PpCFEM1, 4, 5, and 7 proteins can present effector molecule functions, and PpCFEM2, 3, and 6 proteins could present structural functions in the cellular process that occurs at the beginning of the infectious process. CFEM-based heme-iron capturing and absorption mechanism in *P. pachyrhizi* was not observed.

Keywords: CFEM proteins. *Phakopsora pachyrhizi*. Functional characterization

## RESUMO

HOLTMAN, Valéria Cristina, M.Sc., Universidade Federal de Viçosa, fevereiro de 2023. **Identificação e caracterização de proteínas CFEM do fungo *Phakopsora pachyrhizi*, agente etiológico da ferrugem asiática da soja**. Orientador: Luciano Gomes Fietto. Coorientador: Sérgio Herminio Brommonschenkel.

A ferrugem asiática da soja (FAS) causada pelo fungo *Phakopsora pachyrhizi* é uma das principais doenças fúngicas da cultura da soja, podendo causar perdas de até 95% da produção em condições de ambiente favorável e inóculo abundante. Este patógeno é um parasita obrigatório, biotrófico, que durante o processo de infecção, secreta diversas proteínas efetoras que suprimem as respostas de defesa da planta e promovem o parasitismo. Uma das características de proteínas efetoras de fungos e oomicetos é o elevado conteúdo de resíduos de cisteína (>3%). As proteínas CFEM (Common in Several Fungal Extracellular Membrane Proteins) são proteínas secretadas por diferentes espécies de fungos que possuem um domínio proteico com 8 resíduos de cisteína em posições conservadas, e apresentam funções celulares diversas. As proteínas CFEM melhor caracterizadas são Csa2, Rbt5 e Pga7 do fungo *Candida albicans* envolvidas em uma cascata de captura e absorção de moléculas de ferro-heme a partir do hospedeiro. Os objetivos deste estudo foram identificar e caracterizar todas as proteínas CFEM presentes no repertório do fungo *P. pachyrhizi* e entender a função biológica dessas proteínas como componente de membrana, supressor de imunidade ou captura de ferro no ciclo de vida do patógeno. Análises *in sílico* revelaram que *P. pachyrhizi* possui 10 proteínas CFEM (PpCFEM1-PpCFEM10), sendo seis delas preditas como proteínas secretadas solúveis (PpCFEM1, 5, 7, 8, 9, 10) e quatro (PpCFEM2, 3, 4, 6) preditas como proteínas extracelulares com predição de sítio de adição de âncora de glicosilfosfatidilinositol (GPI). Sete dos genes codificadores de proteínas CFEM (*PpCFEM1-7*) foram expressos no período de 0 a 96 horas pós-inoculação, *PpCFEM2, 5* e *6* apresentam maior expressão no tempo de zero horas após inoculação (hpi) e *PpCFEM3* e *4* são expressos em 24 hpi. As proteínas PpCFEM1, 3, 5, e 7 foram identificadas no núcleo e citoplasma celular em ensaios de localização subcelular, as proteínas PpCFEM2 e PpCFEM6 foram identificadas na membrana celular, e PpCFEM4 não apresentou localização subcelular definida. PpCFEM1, 4, 5 e 7 foram capazes de suprimir PTI (imunidade

desencadeada pelo reconhecimento de PAMPs), e apenas a proteína PpCFEM4 foi capaz de suprimir ETI (imunidade desencadeada pelo reconhecimento de efetores) em folhas de *Nicotiana benthamiana*. A expressão e função dos genes Pp-CFEM 8 a 10 não foi analisada. As proteínas PpCFEM não apresentaram homologia com as proteínas CFEM de *C. albicans* que participam do mecanismo de captura e absorção do ferro heme, e análises de conformação de estrutura terciária indicaram que as proteínas PpCFEM não apresentaram a estrutura necessária para a coordenação ferro-heme como presentes em proteínas CFEM de *C. albicans*. Os resultados sugerem que as proteínas PpCFEM1, 4, 5 e 7 apresentam funções de moléculas efetoras, e as proteínas PpCFEM2, 3 e 6 podem apresentar funções estruturais em processos celulares que ocorre no início do processo infeccioso.

Palavras-chave: Proteínas CFEM. *Phakopsora pachyrhizi.* Caracterização funcional.

**SUMMARY**

## 1. INTRODUCTION

The fungus *Phakopsora pachyrhizi* Sydow & P. Sydow, first described in 1902 in Japan (Hennings, 1903), belongs to the phylum Basidiomycota, order Uredinales, and class Urediniomycetes. It can infect more than 95 plant species, especially plants of the *Fabaceae* family, and is the causal agent of Asian soybean rust (ASR), the main disease of soybean (*Glycine max* (L.) Merr.) worldwide (Goellner et al., 2010). Under favorable environmental conditions and without effective control measurements, the ASR can cause losses of up to 95% of the production (Godoy et al. 2016).

The infection begins with the germination of urediniospores under favorable environmental conditions on the field and is completed with the urediniospores released 7-9 days after the establishment of the infection (Bromfield, 1984; Koch & Hoppe, 1983). During the host infection, biotrophic fungi such as *P. pachyrhizi* utilizes specialized feeding structures called *haustoria* to colonize their hosts. The haustoria play an important role in parasitism acting as a platform for the uptake of nutrients, such as sugars and amino acids, and the production and secretion of effector molecules that modulate the host immune system and metabolism (Mapuranga et al., 2022). The rust fungi haustoria-host cell interface is composed of an extrahaustorial matrix (EHMA) that is delimited by the host extrahaustorial membrane (EHM) and the haustorial cell wall (HW). Effector molecules are secreted by the pathogen in the EHMA and are transported into the host cell through unknown mechanisms. Once in the cytoplasm, the effectors can target different membranes, host organelles, or the nucleus, manipulating the host's defense and metabolism pathways to favor its parasitism (Catanzariti et al., 2007; Mapuranga et al., 2022) and survival. Haustoria transcriptome identification and characterization studies are essential for the elucidation and understanding of host-pathogen interactions, and the development of novel pathogen control strategies (Link et al., 2014).

Plants have evolved a complex innate immune system responsible for surveillance, perception, and defense against biotic threats. This sophisticated immune system prevents the growth and development of detrimental microorganisms via two interconnected defense strategies based on pathogen perception (Jones & Dangl, 2006; Dangl et al., 2013). The first strategy utilizes the recognition of pathogen-associated molecular patterns (PAMPs) through plant cell surface-anchored pattern recognition receptors (PRRs) to induce a set of responses collectively referred to as

pattern−triggered immunity (PTI). In the second strategy occurs the recognition of microbial effectors, virulence factors that suppress PTI and promote parasitism, through resistance (R) proteins that initiate the effector-triggered immunity response (ETI). The activation of these immune responses triggers a cascade of complex signaling events, leading to the suppression of the pathogen attack (Jones & Dangl, 2006; Zhang et al., 2022).

Effector molecules are proteins, metabolites, toxins, small peptides, or small RNAs that alter the structure and function of the host cells to favor infection or colonization by the microorganism. First described as ‘avirulence factors’ for their ability to trigger a hypersensitive response in resistant plants, were later found to contribute to virulence in susceptible plants (host plants that lack effective resistance genes). Thus, the term effector began to be adopted to compensate for the conceptual limitation of the term avirulence, since to define the virulence of the protein it is necessary to define the host genotype beforehand to know whether the interaction will be compatible or not (Todd et al., 2022). Due to the selective pressure exerted by the hosts to recognize effector proteins, these tend to evolve quickly and do not usually show signs of homology with other proteins. The lack of homology makes it difficult to predict whether a protein is an effector or not (Todd et al., 2022), but there is a consensus on some common characteristics: effector proteins generally are small secreted proteins (< 250 amino acids), with high content of cysteine residues (> 3%), without transmembrane domains, and with gene expression restricted to certain periods of the infectious cycle, and possesses taxonomic specificity (Lorrain et al., 2015; Sperschneider et al., 2016; de Carvalho et al., 2017; Sperschneider et al., 2018).

The recently characterized *P. pachyrhizi* isolates K8108, MT2006, and UFV02 genomes have a predicted secretome of 2.183, 2.027, and 2.125 secreted proteins, respectively. About 40% of these proteins are predicted to be candidate effector proteins (Gupta et al., 2022) however, only two candidate effector proteins had the effector function established. PpEC23 protein of *P. pachyrhizi*, a small cysteine-rich protein effector that targets the cell nucleus and it is able to suppress the host immune system by interaction with the soybean transcription factor *Gm*SPL12l (SQUAMOSA Promoter-binding-Like protein 12-like) (Qi et al., 2016). Another effector recently described was the Phapa_731740 protein that contains a calcium-binding epidermal growth factor (EGF) domain able to suppress soybean PTI response by interacting

with the pathogenesis-related protein GmβGLU (glucan endo-1,3-β-glucosidase) (Bueno et al., 2022).

Some oomycete and fungal effector proteins exhibit conserved motifs and domains such as RXLR (Arg-Xaa-Leu-Arg) and RXLR-like motifs, lysin motif (LysM), or CFEM (Common in Several Fungal Extracellular Membrane Proteins, Pfam code - PF05730) domains located at their N- or C- terminal regions (Liu et al., 2019). This domain is composed of a conserved motif containing a consensus sequence of eight cysteines with the sequence $PxC[A/G]x_2Cx_{8-12}Cx_{1-3}[x/T]Dx_{2-5}CxCx_{9-14}Cx_{3-4}Cx_{15-16}$, where x is any residue (Zhang et al., 2015). Candidate effector proteins with the CFEM domain have been reported exclusively in fungal species, suggesting that they are essential for host-pathogen interaction.

The first identified and characterized CFEM protein ACI1, from *Magnaporthe oryzae*, interacts with an adenylate cyclase (MAC1) that has key roles in the formation, growth, and development of appressorium structure (Deng & Dean, 2008). The CFEM-GPCR (G-protein coupled receptor) proteins WISH and Pth11 from this fungus have also important roles in virulence and infection development (DeZwaan et al., 1999; Sabnam & Barman, 2017). Five CFEM candidate effector proteins from the maize anthracnose fungus *Colletotrichum graminicola* were shown to suppress the Bcl-2 associated X protein (BAX)-induced programmed cell death in *Nicotiana benthamiana* leaves (Gong et al., 2019). VdSCP76 and VdSCP77 candidate effector proteins from *Verticillium dahliae* with CFEM domain can suppress the host PTI defense mechanism (Wang et al., 2022). The effector proteins CFEM1, CFEMn1, and CFEM5 from *Fusarium graminearum* can suppress the host ETI defense mechanism by interacting with the ZmLRR5 and ZmWAK17ET secreted proteins from maize plants (Zuo et al., 2022). So far, the repertoire of effectors containing the CFEM domain from *P. pachyrhizi* fungus has not yet been functionally characterized.

CFEM proteins can also have functions not related to parasitism. For example, CCW14 from *Saccharomyces cerevisiae* and CfmA-C from *Aspergillus fumigatus* are involved in the maintenance of fungi cell wall stability (Moukadiri et al., 1997; Mrša & Tanner, 1999; Vaknin et al., 2013) and Csa2, Rbt5, and Pga7, CFEM proteins from *Candida albicans* fungus are important for capture and absorption of heme-iron molecules (Weissman & Kornitzer, 2004; Nasser et al., 2016; Kuznets et al., 2014).

The pathogen's ability to capture and assimilate nutrients from its host is essential for pathogenicity. Trace nutrients, including iron, zinc, and manganese, are actively withheld from invading pathogens by mammalian cells in a process called nutritional immunity. Therefore, successful mammalian pathogenic species must have evolved specialized mechanisms to circumvent these host defense mechanisms and cause disease (Crawford & Wilson, 2015). The human pathogen *C. albicans* possesses a cascade mechanism of binding, absorption, and internalization of heme-iron mediated by three CFEM proteins: Csa2, Rbt5, and Pga7 (Weissman & Kornitzer, 2004; Nasser et al., 2016; Kuznets et al., 2014). *Cryptococcus neoformans* Cig1 cell surface protein is a haem-binding protein, regulated by iron availability. Despite having similar function, *C. neoformans* Cig1 does not present a phylogenetic relationship with the CFEM heme-iron binding family of *C. albicans*. This suggests that the ability to exploit mammalian heme as an iron source is an independently evolved trait (Crawford & Wilson, 2015). The mechanism by which obligate plant parasites such as rust pathogens assimilate iron from their hosts is still completely unknown.

In preliminary studies of transient expression in *N. benthamiana* plants using the *Agrobacterium tumenfaciens* infiltration method executed in our laboratory, one protein homologous of PP-CSEP-33, a CFEM protein described by Kunjeti et al. (2016), was able to induce a bleaching reaction on the leaves. PP-CSEP-33 was also observed in the chloroplasts in the subcellular localization assay. Since iron is a central component of photosynthesis, and 90% of the iron in plant leaves is stored in the chloroplast where photosynthesis takes place (Akmakjian et al., 2021), these preliminary results raised the hypothesis that PP-CSEP-33 could be an iron-binding CFEM protein, and the bleaching phenotype could be the result of iron depletion caused by iron capture.

Therefore, the objectives of this study were to identify and characterize all CFEM proteins present in the repertoire of *P. pachyrhizi* fungus and to understand the biological function of these proteins as membrane components, immunity suppressors, or iron capture in the pathogen life cycle.

# 2. LITERATURE REVIEW

## *2.1. Phakopsora pachyrhizi*

The fungus *Phakopsora pachyrhizi* Sydow & P. Sydow, first described in 1902 in Japan (Hennings, 1903), belongs to the phylum Basidiomycota, order Uredinales, and class Urediniomycetes. It can infect more than 95 plant species, especially plants of the Fabaceae family. In soybean (*Glycine max* (L.) Merr.) *P. pachyrhizi* causes Asian soybean rust (ASR), the main fungal disease of this crop worldwide. (Goellner et al., 2010).

The losses caused by ASR can range from 5-95% of the production, depending on the management systems employed and environmental conditions (Godoy et al. 2016). The host's infection process begins with the germination of the spores and formation of the germinative tube when favorable conditions of humidity and temperature are present, followed by the formation of an appressorium which performs the rupture of the leaf epidermis and permits the fungus' entry on the leaf via a penetration hypha. When inside the leaf the fungus develops haustoria, a structure responsible for capturing and absorbing nutrients from the host. The development of new hyphae and haustoria structures proceeds and 7 - 9 days after infection, occurs the rupture of uredia and the release of the asexual spores, called urediniospores (Bromfield, 1984, Koch & Hoppe, 1983). The urediniospores produced are spread in the field by wind, reaching long distances and leading to rapid disease progression, making disease control difficult under inoculum pressure and favorable environmental conditions (Goellner et al., 2010). Due to its exclusively biotrophic behavior, the pathogen survives between crops in the field on soybeans plants or alternative hosts (Bromfield, 1984).

## 2.2. Plant-pathogen interaction and plant defense system

Plants do not have an adaptive defense system or mobile defense cells as found in animals, but they present an innate defense system capable of recognizing and controlling pathogens (Cook et al., 2015). The first defense barrier of plants is performed by an innate immunity mechanism known as PAMP-triggered immunity (PTI), in which conserved molecular patterns of pathogens or microorganisms (PAMP/MAMPs) are recognized by pattern recognition receptors (PRRs) located on the surface of the plant cells (Schwessinger & Zipfel, 2008). To evade PTI, pathogens evolved effector molecules that are secreted into the apoplastic environment and

translocated into the cytoplasmic environment to suppress PTI, causing a susceptibility reaction also known as effector-triggered susceptibility (ETS). These molecules can, in turn, be recognized by host resistance (R) proteins, leading to the activation of the second plant defense barrier known as effector-triggered immunity (ETI). Activation of the ETI mechanism is characterized by the occurrence of a hypersensitivity reaction (HR), a cell death of the infected cells, and other defense responses that restrict the pathogen development (Jones & Dangl, 2006, Dangl et al., 2013).

Effector molecules do not present a general motif signature and possess a high rate of evolution and wide range of functions, making difficult the prediction and study of these molecules. Several algorithms were developed to facilitate the discovery and characterization of new effector molecules. These effector prediction algorithms perform searches in genomic and transcriptome databases for proteins taking into account the main characteristics of effector proteins characterized so far such as reduced size (< 250 amino acids), high content of cysteine residues (> 3%), low molecular weight, presence of secretion peptide signal, gene expression in certain periods of the infectious cycle, and taxonomic specificity (Lorrain et al., 2015, Sperschneider et al., 2016, de Carvalho et al., 2017, Sperschneider et al., 2018). Some fungal effector proteins can also contain conserved motifs, such as RXLR and RXLR-like motifs, lysin motif (LysM), or CFEM domains located at their N- or C- terminal regions (Liu et al., 2019).

### 2.3. CFEM proteins

CFEM proteins (Common in Several Fungal Extracellular Membrane Proteins) are proteins found exclusively in fungal species, exhibiting a characteristic domain of approximately 60 amino acid residues containing 8 cysteine residues in conserved positions (Kulkarni et al. 2003). This CFEM domain can be summarized as follows: PxC[A/G]x2Cx8–12Cx1–3[x/T]Dx2–5CxCx9–14Cx3–4Cx15–16C, where x can be any amino acid residue (Zhang et al., 2015).

The first identified and characterized CFEM protein was ACI1 from *Magnaporthe oryzae*, it was able to interact with an adenylate cyclase (MAC1) acting in the formation, growth, and development of appressorium structure (Deng & Dean, 2008). Two others CFEM proteins also from *M. oryzae*, WISH and Pth11, show a G-

protein coupled receptor, (GPCR) with important roles in virulence and infection development (DeZwaan et al., 1999; Sabnam & Barman, 2017).

Present in species of the Basidiomycota and Ascomycota phyla, the evolutionary origin of the CFEM domain is remote from the common ancestor. In phylogenetic studies carried out with 22 fungal species, it was observed that the number of proteins with CFEM domain varies between one protein from *Saccharomyces cerevisiae* (Moukadiri et al., 1997; Mrša & Tanner, 1999) to 20 proteins in *Magnaporthe grisea* (Zhang et al., 2015). The number of CFEM genes in pathogenic fungi is significantly higher than in non-pathogenic fungi (Zhang et al., 2015) and this variation may be related to the function that CFEM proteins play in the organism (cell wall maintenance, virulence, pathogenicity, immune response, and/ or iron uptake) (Table 1). Curiously, the CFEM functions may also vary among fungi species, and one given species can have a CFEM protein repertoire with diverse functions.

The best-characterized CFEM proteins are the *C. albicans* proteins Rbt5, Csa2, and Pga7 involved in the capture and absorption of extracellular heme-iron (Weissman & Kornitzer, 2004; Nasser et al., 2016; Kuznets et al., 2014). Other two CFEM proteins from *C. albicans*, Csa1 and Pga10*,* were demonstrated to be involved in biofilm formation (Perez et al., 2006, 2011; Weissman & Kornitzer, 2004). CCW14 of *S. cerevisiae* is responsible for cell wall integrity maintenance (Moukadiri et al., 1997; Mrša & Tanner, 1999), a function also attributed to the homologous protein GCCCW14 from *Candida glabrata* (Srivastava et al., 2014). BcCFEM1 protein, with a putative GPI-anchored site, from the necrotrophic *Botrytis cinerea,* is associated with conidial production, virulence, and stress tolerance (Zhu et al., 2017).

Table 1 - CFEM (Common in Several Fungal Extracellular Membrane Proteins) proteins of different fungus species functionally characterized.

| CFEM protein | Organism | Function | References |
|---|---|---|---|
| ACI1 | *Magnaporthe oryzae* | Required for normal appressorium formation | Deng & Dean, 2008 |
| PHT11 | *Magnaporthe oryzae* | Involved in the development of the appressoria, appressoria-like structures, and pathogenicity. | DeZwaan et al., 1999 |
| WISH | *Magnaporthe oryzae* | Involved in the infection, surface sensing, and hyper-conidiation | Sabnam & Barman, 2017 |
| Pga10; Csa1 | *Candida albicans* | Involved in the biofilm formation process | Perez et al., 2006, 2011 |
| Csa2; Pga7; Rbt5 | *Candida albicans* | Involved in the heme uptake system | Perez et al., 2006, 2011; Weissman and Kornitzer, 2004; Weissman et al., 2008. |
| CFEM2; CFEM3; CFEM6 | *Candida parapsilosis* | Involved in the heme uptake system | Ding et al., 2011; Weissman and Kornitzer, 2004; Weissman et al., 2008. |
| CGCCW14 | *Candida glabrata* | Required in the maintenance of intracellular iron content, adherence, and virulence. | Srivastava et al., 2014 |
| CCW14 | *Saccharomyces cerevisiae* | Involved in cell-wall biogenesis and maintenance of the cell-wall integrity and stability | Moukadiri et al., 1997; Mrša & Tanner, 1999 |
| VdSCP76; VdSCP77 | *Verticillium dahliae* | Immunity suppression | Wang et al., 2022 |
| VdSCP33, VdSCP41, VdSCP43, VdSCP72, VdSCP99, VdSCP116, VdSCP120 | *Verticillium dahliae* | Possibly involved in an iron uptake system | Wang et al., 2022 |
| CfmA-C | *Aspergillus fumigatus* | Involved in the stabilization of the cell wall | Vaknin et al., 2014 |
| PTTG_08198 | *Puccinia triticina* | Candidate effector | Zhao et al., 2020 |
| BcCFEM1 | *Botrytis cinerea* | Contributes to virulence, conidial production, and stress tolerance | Zhu et al., 2017 |
| Bcin07g03260 | *Botrytis cinerea* | Involved in plant pathogenicity and conidial germination | Arya et al., 2020 |
| A0A0F2MDI4 | *Sporothrix schenckii* | Involved in interaction with the host (biofilm formation) | Saucedo-Campa et al., 2022 |
| StCFEM12 | *Setosphaeria turcica* | Candidate effector | Wang et al., 2021 |
| FgCFEM11; FgCFEM23; CFEM1; CFEMn1; CFEM5; CFEM8 | *Fusarium graminearum* | Candidate effector | Chen et al., 2021; Zuo et al., 2022 |
| SS1G_03611; SS1G_13935 | *Sclerotinia sclerotiorum* | Candidate effector | Guyon et al., 2014 |

Effector functions have been attributed to CFEM proteins from pathogenic fungi such as VdSCP76 and VdSCP77 from *Verticillium dahlia* (Wang et al., 2022); PTTG_08198 from *Puccinia triticina* (Zhao et al., 2020); FgCFEM11, FsCFEM23, CFEM1, CFEMn1, CFEM5, and CFEM8 from *Fusarium graminearum* (Chen et al., 2021; Zuo et al., 2022); SS1G_03611 and SS1G13935 from *Sclerotinia sclerotiorum* (Guyon et al., 2014). This prediction of CFEM proteins as candidate effectors proteins occurs due to their high content of cysteine residues, a characteristic feature of apoplastic and cytoplasmatic effectors characterized so far. Wang et al. (2021), recently, also performed the characterization of 13 CFEM proteins from *Setosphaeria turcica* by bioinformatics analyses. Six proteins were subjected to functional assays of ETI suppression, and one candidate effector protein tested, StCFEM12 was able to prevent the formation of HR induced by BAX/INF1 inoculations. This ability allowed Want et al. (2021) to classify StCFEM12 as an ETI suppressor protein. In *P. pachyrhizi*, to date, only the CESP-33 effector canidate protein with CFEM domain was reported (Kunjeti et al., 2016), but no functional characterization of this protein was carried out so far.

Among the CFEM proteins already characterized, only the protein Csa2 of *C. albicans* had the three-dimensional conformation resolved by X-ray crystallography. The three-dimensional structure of the CFEM domain deposited in the Protein Data Bank (PDB) under the identification 4Y7S (https://www.rcsb.org/structure/4Y7S ) (Dvir & Kornitzer, 2018) indicated a unique helical-basket structure formed by six α-helices stabilized by four disulfide bonds (Figure 1).

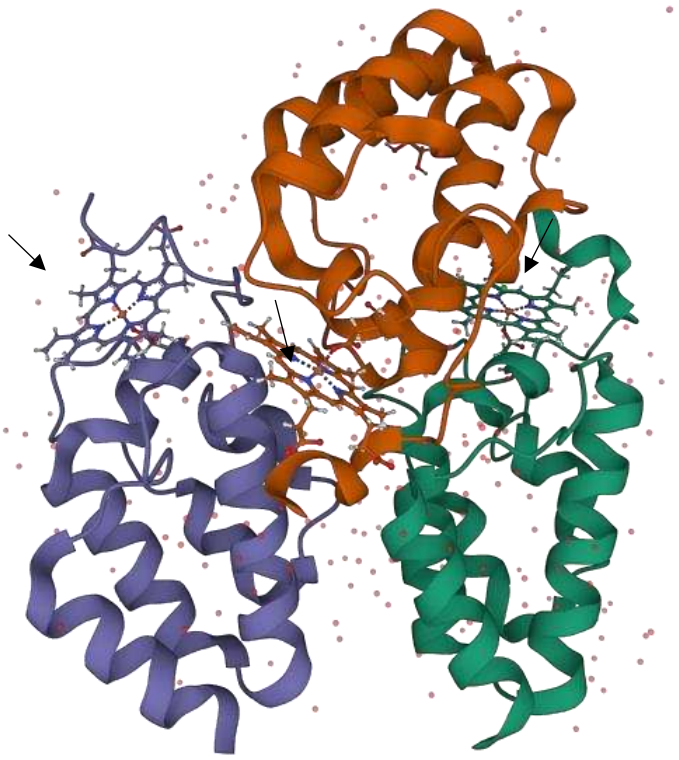


Figure 1: Three-dimensional conformation of Surface antigen protein 2 (Csa2) of *C. albicans* a member of the Common in Fungal Extracellular Membranes (CFEM) protein superfamily, obtained by X-ray crystallography. The structure shows an asymmetric homomeric trimer with extensive interaction between its monomers (CFEM domains) with six α-helices arranged in a compact helical-basket shape in blue, red, and green. The haem groups complexed with iron molecules (small orange spheres) are shown by black arrows. Structures reproduced from Protein Data Bank (PDB ID: 4Y7S; RCSB PDB - 4Y7S: Crystal Structure of the CFEM protein Csa2).

Alignments between CFEM protein sequences indicate that CFEM proteins that participate in heme-iron molecule capture contain an aspartic acid residue in a conserved position between cysteine 3 and 4, usually the 7$^{th}$ amino acid after 3rd cysteine (Dvir & Kornitzer, 2018). A hydrophobic platform formed on the top of the helical-basket structure plus the aspartic acid residue is responsible for the heme-iron coordination (Figure 2) (Nasser et al., 2016).

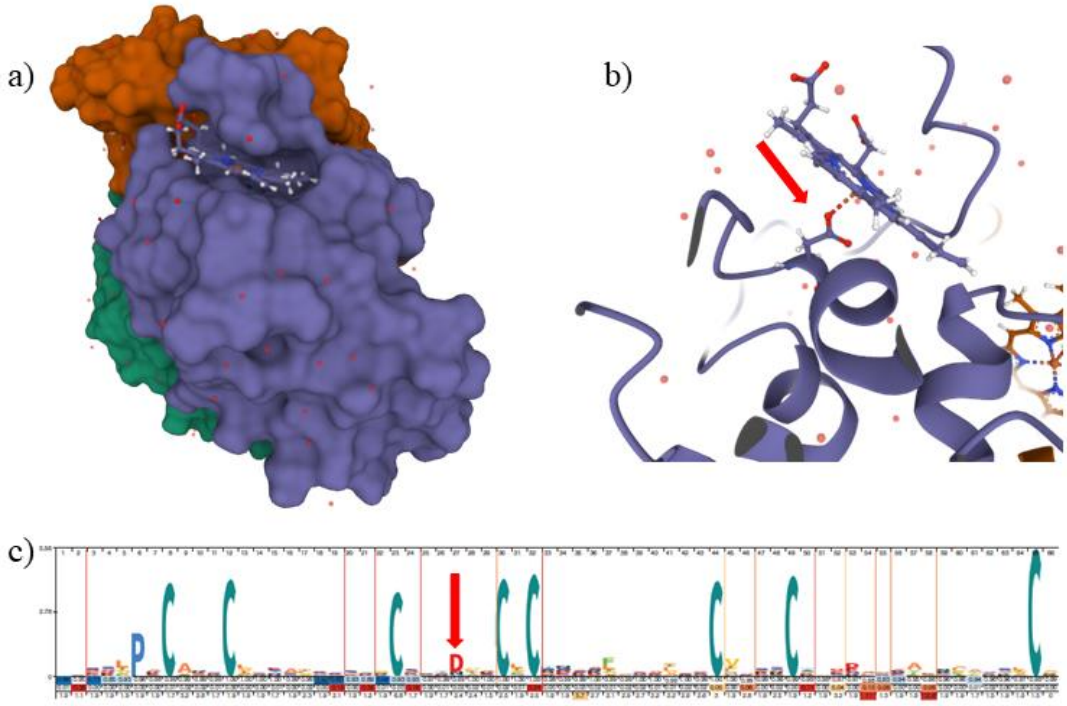



Figure 2: Schematic representation of the CFEM domain heme-iron coordination. a) Hydrophobic platform where occurs the coordination of heme-iron molecules (PDB ID: 4Y7S). b) Zoom view of the aspartic acid residue involved in the heme-iron coordination (red arrow) (PDB ID: 4Y7S). c) CFEM domain logo representation displaying the 8 cysteine residues in conserved positions, and the aspartic acid residue, usually the 7$^{th}$ amino acid after 3rd cysteine, necessary for heme-iron coordination (red arrow) (Kornitzer and Roy, 2020).

### 2.4. Nutrient uptake by rust fungi

Rust fungi are obligate biotrophic parasites that obtain nutrients, such as sugars and amino acids, necessary for their development from the host through a feeding structure known as haustoria. Dikaryotic haustoria develop from an external haustorial mother cell (HMC) with a slender neck penetrating the host cell and a haustorial body forming distally to the neck. With the formation of the haustorial body, an interface zone of separation between the plasma membranes of the host and parasite is established consisting of the cell wall and extrahaustorial matrix (EHMA) of the fungus (Figure 3). The EHMA resembles an amorphous mixture of components, mainly carbohydrates and proteins, partly of fungal, but primarily of plant origin, representing a “trading place” for the exchange of nutrients and information between host and fungus (Voegele & Mendgen, 2033; Catanzariti et al., 2007; Mapuranga et al., 2022).

Rust fungi make use of several strategies to cover their nutritional demands. The uptake of carbohydrates, for example, occurs from the EHMA into the haustoria by a proton gradient established by an $H^+$-ATPase in the haustorial plasm membrane (HM)

(Figure 3). The amino acid uptake is realized via amino acid transporters (AAT1 and AAT2), also located in the HM (Mapuranga et al., 2022) (Figure 3). In the context of the EHMA interface, it would be advantageous for the host, through nutritional immunity, to limit the bioavailability of compounds like iron (Fe) to pathogens, given the metal's importance for the maintenance of vital cellular processes. In general, little is known about how these pathogens obtain micronutrients from the host.

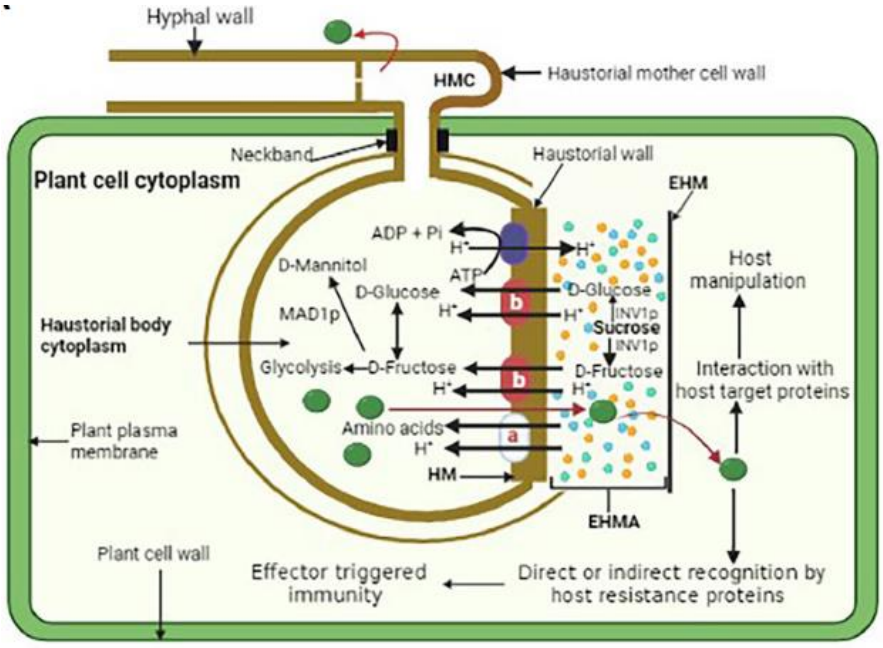



Figure 3: General representation of a rust fungus haustoria and a proton symport model for the active uptake of amino acids and hexoses and redistribution from an infected leaf cell into a rust haustorium. EHMA, extrahaustorial matrix; EHM, extrahaustorial membrane; HM, haustorial plasma membrane. Haustorial plasma membrane H1-ATPase drives the uptake of nutrients by supplying protons. a—amino acid transporters AAT1 and AAT2; b—hexose transporter HXT1 (Image reproduced from Mapuranga et al., 2022).

### 2.5. Iron metabolism and nutritional immunity

Iron (Fe) is one of the most important essential micronutrients in the cellular metabolism of organisms, its ferrous ($Fe^{+2}$) and ferric ($Fe^{+3}$) oxidized forms are involved in several processes such as enzymatic cofactor of different enzymes, oxygen transport, and cellular respiration (Nasser et al., 2016). In plant cells, the iron cofactors are presented as heme, Iron-Sulfur (FeS) clusters, and di-iron centers. Despite being necessary for different cellular compartments, 80% - 90% of the iron cofactors are in the chloroplasts, participating in the photosynthesis process, and in mitochondria. The vacuoles and the protein ferritin are the major storage points of iron (Fe) in the plant cell (Connorton et al., 2017; Kobayashi & Nishizawa, 2012).

The plants are the entry point of the iron in the animal food chain and the absorption of the compound by the plants occurs in the roots by two different strategies. The nongraminaceous plants use the strategy I, absorbing the $Fe^{+2}$ ion, while the graminaceous plants use the strategy II, being able to absorb $Fe^{+3}$ ion. The $Fe^{+3}$ oxidative state, which presents low solubility, is the main form of the compound present in nature. So, in strategy I, the proton pumps acidify the soil to increase the solubility of the compound, while $Fe^{+3}$ is reduced to $Fe^{+2}$ by Ferric Reduction Oxidase 2 (FRO2) and transported by Iron-Regulated Transporter 1 (IRT1) across the cell membrane. In strategy II, phytosiderophores like deoxymugineic acid (MAs) are exported by the transporter of mugineic acid 1 (TOM1) to the extracellular space to chelate $Fe^{+3}$; the complex MA-$Fe^{3+}$ is imported to the cell by the oligopeptide phytosiderophore transporter YS1 (yellow stripe1) (Connorton et al., 2017; Kobayashi & Nishizawa, 2012).

Once in the plant, the iron needs to be transported to the different tissues. During the symplastic transport and cell entry, the iron is in the reduced to $Fe^{+2}$ form, while in the xylem the iron is present in the oxidized form $Fe^{+3}$. To prevent cell damage due to the high oxidation-reduction capacity of free iron, the compound must be complexed to a chelator. The main plant iron chelators are citrate, nicotianamine, and MAs (Connorton et al., 2017; Kobayashi & Nishizawa, 2012). Iron chelation prevents cell damage and reduces the compound's bioavailability, making it difficult for pathogenic microorganisms to capture and use the compound, which limits the pathogen’s development. This process of limiting nutrients available to pathogens is known as nutritional immunity (Weinberg, 1975).

To circumvent the nutritional immunity mechanisms of the hosts, different pathogens have developed mechanisms of capture and absorption of iron complexes from their hosts. Pathogenic microorganisms secrete siderophores, which are proteins capable of binding reversibly to iron, participating in mechanisms of capture and absorption of iron. The iron can be captured in its elemental form ($Fe^{2+}$ or $Fe^{3+}$) or when in a complex with other proteins such as protoporphyrin IX, known as heme. Heme is an organic molecule present in all living organisms, and in decaying organic matter (Kornitzer & Roy, 2020). The HasA proteins from *Serratia marcescens*, the first hemophore to be described at the molecular level, (Letoffe et al., 1994), and IsdX from *Bacillus anthracis* (Ekworomadu et al., 2012) are examples of siderophore-like proteins

that play a role in iron capture and absorption function in these organisms (Kornitzer, 2009; Dvir & Kornitzer, 2018).

*C. albicans* is a commensal fungus capable of causing skin, nail, and mucosal infections in humans., that belongs to the phylum *Ascomycota*, class *Saccharomycetes*, order *Saccharomycetales*, and family *Cryptococcaceae*. In immunosuppressed patients, the fungus can spread systematically and cause death (Kuznets et al., 2014). As it presents a threat to human health, the fungus *C. albicans* has been extensively studied, and different mechanisms of micronutrient capture and absorption have already been characterized (Kuznets et al., 2014). Among them is the mechanism of heme-$Fe^{3+}$ capture and absorption involving the CFEM proteins Rbt5, Csa2, and Pga7. These three proteins form a cascade of capture and transfer of heme-$Fe^{3+}$ molecules (Figure 4). The Csa2 protein, secreted in the extracellular space, captures heme-$Fe^{+3}$ molecules from the host's hemoglobin, albumin, and from free heme. Rbt5 protein fixed to the cell wall by a Glycosylphosphatidylinositol (GPI) anchor, captures heme molecules in the vicinity of the cell, in addition to receiving heme molecules captured by the Csa2 protein. When reversibly bound to Rbt5, the heme-$Fe^{3+}$ molecule is then transferred to the Pga7 protein which is anchored to the cytoplasmic membrane by a GPI anchor. The mechanism of internalization and processing of heme-iron molecules was not yet elucidated, but there is evidence of the transfer of the heme-iron molecule from the Pga7 protein to another membrane protein that would help the endocytosis of the heme-$Fe^{3+}$ molecule by the ESCRT (Endosomal Sorting Complex Required for Transport) complexes I, II, and III. When inside the cell, a vacuolar mechanism releases the iron from the protoporphyrin IX (Dvir & Kornitzer, 2018; Kornitzer & Roy, 2020).

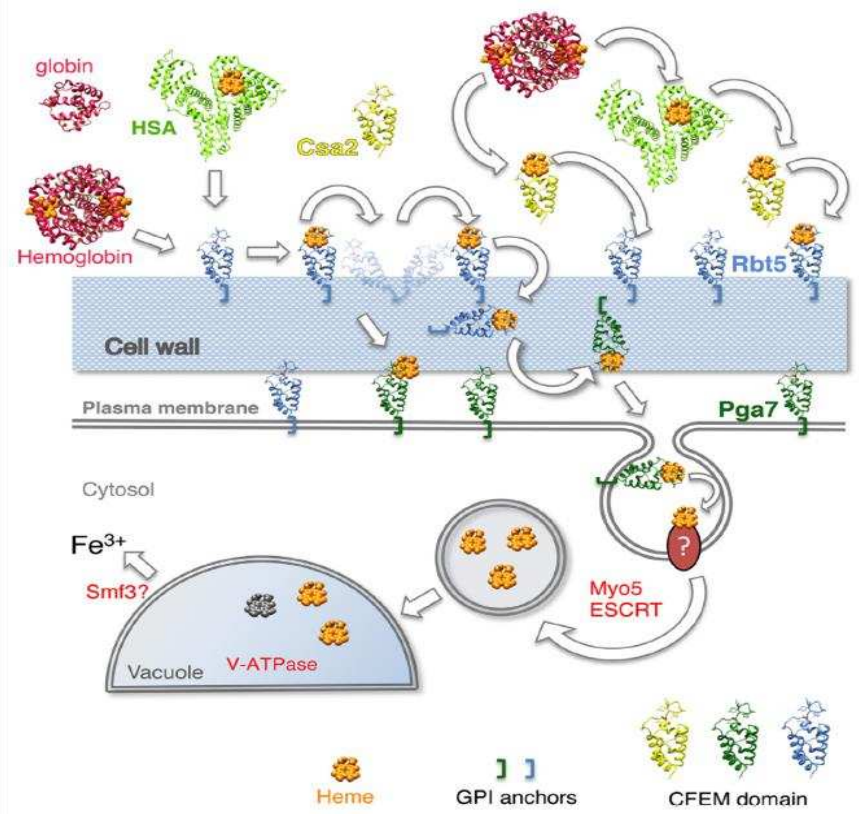



Figure 4: Schematic representation of the *C. albicans* heme-$Fe^{3+}$ capture mechanism. Three *C. albicans* proteins are necessary for the heme-$Fe^{3+}$ mechanism: Cas2, Rbt5, and Pga7. Csa2 protein is secreted in the extracellular space and captures heme-$Fe^{+3}$ molecules from hemoglobin, albumin, and free heme in the host. Rbt5 protein associated with the cell wall captures heme molecules in the vicinity of the cell and receives heme molecules captured by the Csa2 protein. Pga7 protein which is anchored to the cytoplasmic membrane by a GPI anchor receives Heme-$Fe^{3+}$ that is reversibly bound to Rbt5 and transferred to another unknown membrane protein (represented by a question mark). This transference would help in the endocytosis of the heme-Fe3+iron molecule by the ESCRT (Endosomal Sorting Complex Required for Transport). Inside the cell, a vacuolar mechanism releases the $Fe^{3+}$ from the protoporphyrin IX (Image reproduced from Kornitzer & Roy, 2020).

Besides the *C. albicans* heme-iron capturing and absorption mechanism other fungi species had developed other mechanisms for host iron capturing. *Cryptococcus neoformans,* an environmental basidiomycete, captures and absorbs heme-iron capture using the secreted protein Cig1 without a CFEM domain. However, similar to CFEM heme-iron capturing mechanisms, the ESCRT complexes are needed for endocytosis, and processing of heme-iron molecules (Kornitzer & Roy, 2020). *Schizosaccharomyces pombe* fungus possesses two heme-iron capture mechanisms. The first mechanism is executed by StuI, a GPI-anchored membrane protein with low homology with CFEM proteins of *C. albicans*. This protein contains a partial CFEM domain formed by seven cysteine residues and the conserved aspartic acid residue,

and it is capable to capture and internalize the heme-iron through endocytosis by the ESCRT complexes. The second one is performed by Str3, a membrane transporter capable of lower-affinity heme-iron binding and direct internalization to the cytoplasm (Labbé et al., 2020; Kornitzer & Roy, 2020).

In preliminary studies of transient expression in *N. benthamiana* leaves using the *Agrobacterium tumenfaciens* infiltration method, PP-CSEP-33 a CFEM protein described by Kunjeti et al. (2016), was able to induce a bleaching reaction. PP-CSEP-33 was also observed in the chloroplasts in the subcellular localization assay. Since 90% of the iron in plant leaves is stored in chloroplast where photosynthesis takes place, this bleaching phenotype could be associated with the possible iron capture mediated by PP-CSEP-33, raising the hypothesis that *P. pachyrhizi* could have an iron capture mechanism mediated by CFEM proteins homologous to that one described in *C. albicans.* Therefore, the objectives of this study were to identify and characterize all CFEM proteins present in the repertoire of *P. pachyrhizi* fungus and to understand the biological function of these proteins as membrane component, immunity suppressor or iron capture in the pathogen life cycle.

# 3. MATERIAL AND METHODS

## 3.1. Identification of *P. pachyrhizi* CFEM proteins (PpCFEM)

Identification of CFEM proteins was performed utilizing the predicted proteome of *P. pachyrhizi* isolate K8108 available on the MycoCosm website at the Joint Genome Institute (https://mycocosm.jgi.doe.gov/PhapaK8108/PhapaK8108.home.html). Due to the cysteine residues spacing position conservation and a high divergence between other amino acids residues, the proteins were detected using the CFEM Pfam code - PF05730 (https://pfam.xfam.org/family/PF05730). Gene sequences were manually annotated by comparing the prediction results of different gene predictors tools available in the MycoCosm portal with the RNA coverage data available. The exon-intron structure, copy number, and polymorphisms among copies were also manually analyzed. The sequence comparisons were performed utilizing the MEGA v.7.02 software (Kumar et al., 2015).

## 3.2. In silico characterization of CFEM protein sequences

PpCFEM candidate proteins were submitted to the protein family classification tool - InterProt (https://www.ebi.ac.uk/interpro/ ) (Blum et al., 2020) for confirmation of the CFEM domain prediction. For identification of secretion signal were used the SignalP 6.0 server (http://www.cbs.dtu.dk/services/SignalP/) and PHOBIUS program (https://phobius.sbc.su.se/) (Käll et al., 2007). GPI anchors were predicted in PredGPI server (http://gpcr2.biocomp.unibo.it/predgpi/) (Gíslason et al., 2021) and NetGPI prediction tool (https://services.healthtech.dtu.dk/service.php?NetGPI). The identification of transmembrane domains was also performed in PHOBIUS program (https://phobius.sbc.su.se/) and PROTTER tool (http://wlab.ethz.ch/protter/start/) (Omasits et al., 2014). For predicting the subcellular localization of proteins were used on line predictors DeepLOC-1.0 (http://www.cbs.dtu.dk/services/DeepLoc/) and WoLF PSORT (https://wolfpsort.hgc.jp/ ). Due to the high content of cysteine residues in the CFEM proteins, the EffectorP 3.0 software (https://effectorp.csiro.au/ ) was also used to predict candidates effectors proteins among the *P. pachyrhizi* proteins with CFEM domains. The pairwise comparations were performed in MEGA-X software by p-distance method.

The evolutionary relationships and a possible role association among the predicted PpCFEM proteins and proteins of other fungi exhibiting the CFEM domain were performed using phylogenetic analysis. For this, the MAFFT v.7 software (Katoh

et al., 2019) was used for sequence alignments and the evolutionary tree was constructed using the maximum-likelihood approach, with edge-linked partition model and branch supports with the ultrafast bootstrap determined in IQ-TREE (Chernomor et al., 2016; Nguyen et al., 2015; Hoang et al., 2018). The consensus trees were visualized and rooted in midpoints using iTOL v.5 (Letunic and Bork, 2021).

The tridimensional structure prediction of PpCFEM proteins was modeled in the AlphaFold program (Jumper et al., 2021), and the comparisons among the tridimensional structures were performed in UCSF Chimera v.1.16 (Pettersen et al., 2004).

### 3.3. Gene expression assay

Soybean plants cv. Conquista in the V2 stage was sprayed with a suspension ($5x10^5$ uredospore $mL^{-1}$) of the *P. pachyrhizi* isolate PPUFV02, collected in Viçosa - Minas Gerais - Brazil. The plants were kept in a dark chamber at 25ºC for 24 hours and then transferred to a growth chamber at 22 ºC with a photoperiod of 12/12 hours. Soybean leaf samples were collected in liquid nitrogen at 0, 24, 48, 72, and 96 hours after inoculation, and stored in an ultra-freezer at -80ºC until processing.

Total RNA was purified using the Concert™ Plant RNA Reagent (Thermo Fisher), following the manufacturer's instructions. RNA samples were treated with RNase-free DNase I (Qiagen, Valencia, CA, USA) and purified with RNeasy CleanUp Kit (Qiagen). Reverse transcription (RT) was performed using the SuperScript First-Strand Synthesis kit for RT-PCR (Invitrogen), according to the manufacturer's protocol, using ten micrograms of total RNA and Oligo primer (dT20), followed by qPCR with gene-specific primers (Table 2).

The RT-qPCR reactions were carried out in the 7500 Real-Time PCR Systems (Applied Biosystems) equipment, using the SYBR® Green PCR Master Mix (Applied Biosystems). Cycling conditions were 95°C for 10 min, 40 cycles of 15s at 95°C, and 1 min at 60°C, followed by a dissociation curve (melting) step to verify the specificity of the amplification. All genes were analyzed in three biological replicates and three technical repetitions for each time tested. The α-tubulin gene was used as a reference gene to normalize real-time qPCR data (Kunjeti et al., 2016). Relative quantification was calculated using the comparative cycle threshold method (ΔΔCt) (Livak & Schmittgen, 2001) and statistical comparisons were done using the student’s t-test.

Table 2 - Oligonucleotides to quantify the expression of *PpCFEM 1-10* genes at 0, 24, 48, 72 and 96 hours after inoculation with the PPUFV02 *P. pachyrhizi* isolate using RT-qPCR.

| Primer | 5'-3' Sequence |
|---|---|
| qPCR_PpCFEM1_F | GAGTCAGCAGGACGAAGTAAG |
| qPCR_PpCFEM1_R | ATTTGGGAGCTACACAAACAAC |
| qPCR_PpCFEM2_F | GGCGAACACGATCACTAAC |
| qPCR_PpCFEM2_R | TAGGCCCAATAAAGCGTATAAC |
| qPCR_PpCFEM3_F | GCGCAAGAGACTGCTATAC |
| qPCR_PpCFEM3_R | ACTCATCCCTCCTGCATAA |
| qPCR_PpCFEM4_F | GAGTCTGCGAAACCTCAATTA |
| qPCR_PpCFEM4_R | GTAAGGGTCGAGAAGCTAGA |
| qPCR_PpCFEM5_F | CCTGTTACGTGGGATCTTTG |
| qPCR_PpCFEM5_R | CTTGGGCCTGAGTCATAAAC |
| qPCR_PpCFEM6_F | GGCGGAACCCAGTTTTTGA |
| qPCR_PpCFEM6_R | TTGGGCACATTAATTTGAAAGGT |
| qPCR_PpCFEM7_F | GAGATAAATGCAAACCTGCAGATAA |
| qPCR_PpCFEM7_R | ACTGCGTCTTTGTTGGTAGAA |
| qPCR_PpCFEM8_F | GCCTCAGAGACTACTGGTGTAT |
| qPCR_PpCFEM8_R | CTTTCGGCGTGGTATCATCTT |
| qPCR_PpCFEM9_F | CTTGCACCTCACTCTAAATCTC |
| qPCR_PpCFEM9_R | TCAGCAAGGTTCGGTTCTA |
| qPCR_PpCFEM10_F | TTGCACTGCCGAGGATTT |
| qPCR_PpCFEM10_R | GCTGAATCGCACCTACTATGAA |

### 3.4. Gene cloning

For subcellular localization and PTI/ETI suppression assays, the coding region of the *PpCFEM* genes was amplified from cDNA, prepared as mentioned for the RT-qPCR assays, using specific primers for each gene and assay (Table 3). Amplification reactions were prepared to a final volume of 50 ul containing 25 ng of cDNA; 0.3 µM of each primer (Table 3); dNTP mixture (200 µM of each dNTP); 2.5 mM $MgCl_2$; 2X SuperFi Buffer; 2.5 U of Platinum™ SuperFi™ DNA Polymerase (Invitrogen™). The PCR reactions were performed in a Veriti™ Thermal Cycler 96-well thermocycler (Applied Biosystems) with the following cycling conditions: initial denaturation at 98 °C for 30s; 35 cycles of denaturation at 98°C for 10s, annealing at the optimal temperature of each primer pair for 10s, extension at 72°C for 1 min/kb. PCR products were purified using the QIAquick® Gel Extraction Kit (Qiagen) and the DNA amount was quantified in NanoDrop 1000 equipment (Thermo Scientific).

Table 3 - Oligonucleotides used to *PpCFEM* genes amplification for subcellular localization and PTI/ ETI assays.

| Primer | 5'-3' sequence | Goal |
|---|---|---|
| PK7FWG2_F | CTTCCTCTATATAAGGAAGTTCATTTCATTTGG | PCR cloning confirmation |
| PK7FWG2_R | CAGCTCCTCGCCCTTGCTC | PCR cloning confirmation |
| PEDV6_F | CGCCAGGGTTTTCCCAGTCACGAC | PCR cloning confirmation |
| PEDV6_R | AGCGGATAACAATTTCACACAGGA | PCR cloning confirmation |
| PpCFEM1_TOPO-F | CACCATGCTGAGTCAGCAGGACGAAG | Subcelular localization/ PTI/ ETI |
| PpCFEM1_TOPO-R | CTGGTTACAAACAGTTCTTGAATAAC | Subcelular localization |
| PpCFEM1_TOPO-R_ST | TCACTGGTTACAAACAGTTCTTGAAT | Subcelular localization/ PTI/ ETI |
| PpCFEM2_TOPO-F | CACCATGGATTCATTAACTAAAAGTATTCATTC | Subcelular localization/ PTI/ ETI |
| PpCFEM2_TOPO-R | GGAAAGTACTTCAGATAGGCCC | Subcelular localization |
| PpCFEM2_TOPO-R_ST | CTAGGAAAGTACTTCAGATAGGCCC | Subcelular localization/ PTI/ ETI |
| PpCFEM3_TOPO-F | CACCATGGATTCATTCGACAAAAATATCTATAC | Subcelular localization/ PTI/ ETI |
| PpCFEM3_TOPO-R | GGAAAGATCTACAGAGAGGTGAAATAAAT | Subcelular localization |
| PpCFEM3_TOPO-R_ST | CTAGGAAAGATCTACAGAGAGGTGAAATAAAAT | Subcelular localization/ PTI/ ETI |
| PpCFEM4_TOPO-F | CACCATGGCAGTGGGTCTAGGGAATATTC | Subcelular localization/ PTI/ ETI |
| PpCFEM4_TOPO-R | ACACGAATAGCATTGCAAAGCTGC | Subcelular localization |
| PpCFEM4_TOPO-R_ST | TTAACACGAATAGCATTGCAAAGCTGC | Subcelular localization/ PTI/ ETI |
| PpCFEM5_TOPO-F | CACCATGACTTCAGATATTAAAACTACAGATG | Subcelular localization/ PTI/ ETI |
| PpCFEM5_TOPO-R | GTGTTTTGGAATGGGGTTGGTG | Subcelular localization |
| PpCFEM5_TOPO-R_ST | TTAGTGTTTTGGAATGGGGTTGG | Subcelular localization/ PTI/ ETI |
| PpCFEM6_TOPO-F | CACCATGCAGGATCTGGCCGGACTACCTAAC | Subcelular localization/ PTI/ ETI |
| PpCFEM6_TOPO-R | CTGTCCTGGCTGCAGTCTTCATG | Subcelular localization |
| PpCFEM6_TOPO-R_ST | CTGTCCTGGCTGCAGTCTTCATGTGA | Subcelular localization/ PTI/ ETI |
| PpCFEM7_TOPO-F | CACCATGGCGGACGGACAAAAAC | Subcelular localization/ PTI/ ETI |
| PpCFEM7_TOPO-R | TCTGCGGCCAGGAAGAGTTAC | Subcelular localization |
| PpCFEM7_TOPO-R_ST | CTATCTGCGGCCAGGAAGAGTTAC | Subcelular localization/ PTI/ ETI |
| PpCFEM8_TOPO-F | CACCATGAACGGTGAACAACCTAACGAC | Subcelular localization/ PTI/ ETI |
| PpCFEM8_TOPO-R | ATTTATCTTTTTAATGTTAGACCTTCTTC | Subcelular localization |
| PpCFEM8_TOPO-R_ST | TTAATTTATCTTTTTAATGTTAGACCTTCTTC | Subcelular localization/ PTI/ ETI |
| CFEM9_TOPO-F | CACCATGGCAGAAAAAAAAGGTTTGGGG | Subcelular localization/ PTI/ ETI |
| CFEM9_TOPO-R | ATCTATCTTCACTGTATTGGTCTC | Subcelular localization |
| CFEM9_TOPO-R_ST | TTAATCTATCTTCACTGTATTGGTCTC | Subcelular localization/ PTI/ ETI |
| CFEM010_TOPO-F | CACCATGTTCATTAGCTATTGTTCAGCAG | Subcelular localization/ PTI/ ETI |
| CFEM010_TOPO-R | TCGTTTTCTGATAGATCTTCTCTTATG | Subcelular localization |
| CFEM010_TOPO-R_ST | CTATCGTTTTCTGATAGATCTTCTCTTATG | Subcelular localization/ PTI/ ETI |

Purified PCR products were ligated into the pENTR D-TOPO vector (Invitrogen), following the manufacturers' recommendation, and then transformed into chemically competent *Escherichia coli* DH5α cells. The transformed cells were selected on Luria Bertani (LB) medium (5g $L^{-1}$ yeast extract; 10g $L^{-1}$ bacteriological peptone; 10g $L^{-1}$ NaCl; 15g $L^{-1}$ agar) with kanamycin (50 mg $mL^{-1}$) and kept overnight at 37ºC. PCR-confirmed transformants were grown in 5 mL of liquid LB medium with kanamycin (50 mg $mL^{-1}$) in a shaker incubator at 180 rpm for 16 hours at 37 ºC. Plasmidial DNA extractions were performed with the NucleoSpin® Plasmid kit (MACHEREY-NAGEL), following the manufacturer's recommendations.

Confirmation of the in-frame cloning in the target vectors was made by sequencing using the universal primers (M13 F/R) and the BigDye™ Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, USA), with the following conditions: 25 cycles of denaturation at 96°C for 15s, annealing at 50°C for 15s, extension at 60°C for 4 min. The reactions were purified with the BigDye® XTerminator™ kit (Applied Biosystems, USA), following the manufacturer's protocol, and then analyzed in the SeqStudio™ Genetic Analyzer capillary electrophoresis equipment (Applied Biosystems, USA). The sequences were analyzed in MEGA v.7.02 software (Kumar et al., 2015).

### 3.5. Subcellular localization assays

The *PpCFEM* genes cloned in the pENTR D-TOPO vector (Invitrogen) were transferred to pK7FWG2 and pK7WGF2 vectors, using the Cloning Gateway® LR Clonase™ II Enzyme Mix kit (Invitrogen), following the manufacturer's recommendations and transformed in *E. coli* DH5α cells. Plasmids containing *PpCFEM* genes cloned in-frame, confirmed by sequencing as previously described, were transferred into chemically competent *Agrobacterium tumefaciens* GV3101 cells. The transformation reactions were plated in a solid LB medium with spectinomycin (100 µg $mL^{-1}$), gentamicin (50 µg $mL^{-1}$), and rifampicin (100 µg $mL^{-1}$), and kept for 48 hours at 28°C.

For subcellular localization assays, *A. tumefaciens* GV3101 carrying the pK7FWG2::PpCFEMs and *A. tumefaciens* GV3101 containing proteins with fusion fluorescent proteins used to mark the different cellular compartments (Table 4) were striated on plates of solid LB medium, containing the above-mentioned antibiotics, and kept for 48 hours at 28ºC. An isolated colony from each transformant was grown in 10 mL of liquid LB medium, containing the selection antibiotics at 28 °C in a shaker incubator with shaking at 180 rpm for 16 hours. The cultures were centrifuged for 10 minutes at 5000 g, the supernatant was discarded, and the pellets were resuspended in infiltration solution (10 mM MES, 10 mM $MgCl_2$, 200 µM acetosyringone (pH 5.6)). After a second wash, the bacterial suspension concentrations were adjusted to $OD_{600}$ = 0.4. Agrobacterium suspensions were prepared, containing equivalent concentrations of CFEM proteins and marker proteins in a final concentration of $OD_{600}$ = 0.2, and were incubated at room temperature for 3 hours before the infiltration of 5-week-old *Nicotiana benthamiana* TSL plants.

Table 4 - Subcellular compartment marker proteins constructs used in this study.

| Construct | Description |
|---|---|
| pMP90::WWP1-mCherry | Nuclear marker - *Arabidopsis thaliana* AtWWP1 protein (Calil et al., 2018) |
| pMP90::vac-mcherry | Vacuole marker - *A. thaliana* c-TIP protein (Saito et al., 2002) |
| pMP90::mp-mCherry | Plasma membrane marker – *A. thaliana* AtPIP2A protein (Cutler et al., 2000) |
| pMP90::pt-mCherry | Plastid marker - Partial sequence of Rubisco protein from N. tabacum (Dabney-Smith C et al., 1999) |

Three independent *N. benthamiana* plants were infiltrated with the suspension of *A. tumefaciens* carrying each PpCFEM gene individually with the marker constructs. Infiltration was performed with 1 ml syringes without needles. After infiltration, the plants were kept at 22 °C for 72 hours and then analyzed under a confocal laser scanning microscope (LSM 510 META, Carl Zeiss, Oberkochen, Germany). GFP was excited at 488 nm and then captured at 550 nm, and mCherry was excited at 554 nm and captured at 560-615 nm. The images were processed with the software ZEN2.3 (Carl Zeiss).

### 3.6. ETI and PTI suppression assays

PpCFEM genes cloned in the pENTR D-TOPO vector (Invitrogen) were transferred to the pEDV6 vector (effector detector vector) (Sohn, et al., 2007, Badel et al., 2013; Fabro et al., 2011) using the Gateway® LR Clonase™ II Enzyme Mix kit (Invitrogen), and transformed into *E. coli* DH5α cells. The transformation reactions were plated in a solid LB medium with selection for the antibiotic gentamicin (25 μg $mL^{-1}$) and kept overnight at 37ºC. Confirmation of transformants and plasmid DNA extraction was carried out as previously described. The pEDV6 plasmids containing the *PpCFEM* genes (pEDV::PpCFEM) were transferred from *E. coli* DH5α to *Pseudomonas fluorescens EtHAn*, which carries a functional Type Three Secretion System (TTSS), using the standard triparental mating system and *E. coli* HB101 (pRK2013) as a helper strain. The transformation reactions were plated in a solid LB medium with selection for the antibiotic gentamicin (25 μg $mL^{-1}$) and chloramphenicol (30 μg $mL^{-1}$) and kept at 28 °C for 48 hours. Confirmation of transformants and plasmid DNA extraction was carried out as previously described.

For PTI and ETI suppression assays, *P. fluorencens EtHAn* containing the *CFEM* genes (PfECFEM), and *Pseudomonas syringae pv garcae* (Psgc) 1202 were striated on plates of solid LB medium, containing gentamicin (25 μg $mL^{-1}$) and/or

chloramphenicol (30 µg $mL^{-1}$)] and [rifampicin (100 µg $mL^{-1}$), and kept for 48 hours at 28ºC. An isolated colony from each transformant was grown in 10 mL of liquid LB medium, containing the selection antibiotics at 28 °C in a shaker incubator with shaking at 180 rpm for 16 hours. The cultures were centrifuged for 10 minutes at 5000 g, the supernatant was discarded, and the pellets were resuspended in infiltration solution ($MgCl_2$ 10 mM). For PTI assays the bacterial suspensions (PfECFEM) had their concentrations adjusted to $OD_{600}$ = 0.2 and Psgc $OD_{600}$ = 0.025, suspensions were used to infiltrate *N. benthamiana* leaves. Twenty independent leaves of *N. benthamiana* plants were infiltrated in different positions at each repetition, and after seven hours the Psgc culture was infiltrated, to maintain an area of intersection between the infiltrated regions. PpCFEM proteins that maintained the hypersensitivity reaction (HR) in the intersection region in more than 60% of the repetitions after 48 hours were considered PTI suppressors.

For ETI assays, *N. benthamiana* plants were infiltrated with the same PfECFEM cultures, diluted in infiltration solution with concentrations adjusted to $OD_{600}$ = 0.15, and bacterial suspensions were mixed with *P. syringae EtHAn* suspension containing the AvrB gene (an HR inducer) at concentration $OD_{600}$ = 0.1, previously cultured in LB medium with chloramphenicol (30 µg $mL^{-1}$) and Kanamycin (50 µg $mL^{-1}$). Twenty independent leaves of *N. benthamiana* were infiltrated in different positions at each repetition. PpCFEM proteins that maintained the infiltrated region without inducing HR in more than 60% of the replicates after 48 hours were considered ETI suppressors.

For both assays, 5-week-old *N. benthamiana* plants were used, the plants were grown and kept in growth chambers with a photoperiod of 12/12 hours and 22°C, after the inoculations.

# 4. RESULTS

## 4.1. Bioinformatic analyses reveal the presence of ten CFEM genes in the *P. pachyrhizi* isolate K8108 genome.

The search on the *P. pachyrhizi* isolate K8108 proteome revealed the presence of 10 proteins exhibiting the CFEM domain. These proteins were named PpCFEM1 up to PpCFEM10 (Table 5). Due to the lack of manual annotation of the genes encoding these PpCFEM proteins, a manual annotation process was performed (Table 5) and the gene structure analysis was based on the RNA coverage available at Mycocosm (Figure 5). The coding region of the *PpCFEM* genes varied from 294 nucleotides (*PpCFEM1*) to 1275 nucleotides (*PpCFEM10*) (Table 5). The exon number varied from 3 (*PpCFEM 1, 7, 8, 9*) to 7 (*PpCFEM4*) (Table 5).

The *P. pachyrhizi* fungus has a dikaryotic genome, for this reason, the existence of different alleles and copy numbers of the *PpCFEM* genes were also analyzed (Table 6). All *PpCFEM* are single copy with 2 alleles each. Among the ten *PpCFEM* genes analyzed, *PpCFEM2*, *PpCFEM4*, *PpCFEM7*, and *PpCFEM9* did not show any SNP variation between alleles resulting in amino acid substitutions, while for *PpCFEM1* gene the SNPs G59A and G179A that leads to amino acid substitutions in position N20S (in signal peptide) and S60N (in CFEM domain region) was observed, respectively (Table 5). In *PpCFEM3* two SNPs G58T and A214G, lead to V20L (in signal peptide) and N72D (in CFEM domain) substitutions. For *PpCFEM5* only one SNP (T431C) causing the L144S substitution in the transmembrane domain was observed (Table 5). *PpCFEM6* also showed only one SNP (G151A) in the alleles leading to A51T substitution inside the CFEM domain. *PpCFEM8* gene presents the SNPs A451T and A556G that leads to substitutions T151S and T186A, and *PpCFEM10* presents the SNP A40C that causes I14L substitution in the transmembrane domain.

Table 5 – Characterizations of genes encoding CFEM proteins identified in the *P. pachyrhizi* isolate K8108 proteome available in Join Genome Institute (JGI) Portal.

| JGI protein model identification | JGI protein ID | Project ID | ORF (bp) | #Exon | #Intron | #Alleles | #Copy | SNPs [a] | Substitution ty | Amino acid Substitution [b] |
|---|---|---|---|---|---|---|---|---|---|---|
| fgenesh1_kg.tig00005426_2099_#_509_#_evgvelvOases_81_Loc13092t3_150403 | 8946493 | *PpCFEM1* | 294 | 3 | 2 | 2 | 1 | G59A / G179A | NS | N20S / S60N |
| fgenesh1_kg.334_#_407_#_TRINITY_DN35891_c0_g2_i21 | 7098847 | *PpCFEM2* | 480 | 4 | 3 | 2 | 1 | - [c] | S | - |
| CE1747391_9878 | 1747392 | *PpCFEM3* | 480 | 4 | 3 | 2 | 1 | G58T / A214G | NS | V20L / N72D |
| fgenesh1_kg.tig00003678_1472_#_1488_#_evgsoapSoap_81_scaffold5012_65944 | 8587998 | *PpCFEM4* | 738 | 7 | 6 | 2 | 1 | - | S | - |
| fgenesh1_kg.tig00008861_3133_#_551_#_evgsoapSoap_121_scaffold3511_68401 | 9511845 | *PpCFEM5* | 603 | 4 | 3 | 2 | 1 | T431C | NS | L144S |
| fgenesh1_kg.tig00000285_115_#_1237_#_evgsoapSoap_41_scaffold133758_57417 | 7539766 | *PpCFEM6* | 486 | 4 | 3 | 2 | 1 | G151A | NS | A51T |
| CE2073106_2215 | 2073107 | *PpCFEM7* | 468 | 3 | 2 | 2 | 1 | - | S | - |
| estExt_fgenesh1_pg.C_tig00002954_11870013 | 11627847 | *PpCFEM8* | 600 | 3 | 2 | 2 | 1 | A451T / A556G | NS NS | T151S / T186A |
| CE266267_596 | 266268 | *PpCFEM9* | 666 | 3 | 2 | 2 | 1 | - | S | - |
| fgenesh1_kg.tig00013277_4280_#_2102_#_evgvelvOases_111_Loc7481t1_21427 | 10051878 | *PpCFEM10* | 1275 | 4 | 3 | 2 | 1 | A40C | NS | I14L |

a: Nucleotide substitutions in *PpCFEM* alleles, resulting in amino acid substitutions.
b: Amino acids substitutions in PpCFEM proteins caused. Nucleotide substitutions.
a/b: Analyses performed manually in the MEGA v.7.02 software (Kumar et al., 2015).
c: Absence of nucleotide differences between *PpCFEM* alleles leading to amino acid substitutions.

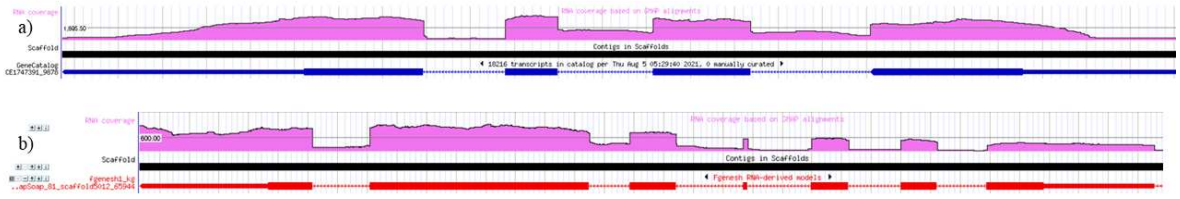



Figure 5: Examples of the manual annotation process of the *PpCFEM* genes at Mycocosm portal in JGI. The gene prediction (blue for the primary allele/ red for the secondary) was analyzed against the RNA coverage (pink). The thicker blue and red bars correspond to exons while the thinner blue and red lines are introns. The presence of exons and introns in RNA is confirmed by the pink valleys and depressions, respectively. a) RNA coverage data for the *PpCFEM3* gene showing the presence of four exons and three introns. b) RNA coverage data for the *PpCFEM4* gene exhibiting seven exons and five introns predicted and manually validated.

### 4.2. Sequence analyses of PpCFEM proteins show the conservation of the eight cysteine residues in the CFEM domain and the presence of the asparagine residue needed for the metal-iron coordination

The presence of the CFEM domain was analyzed and confirmed in all PpCFEM proteins by the family classification tool in the InterProt platform. The PpCFEM10 was also classified as a member of the Auxiliary Activity Family 9 (AA9) of CAZy (IPR005103). The protein sequences were then aligned using ClustalW, with the characterized CFEM proteins Csa2, Rbt5, and Pga7 (GenBank accession XP_715426.1, XP_713317.1, XP_713316.1) from *C. albicans* as reference sequences. The main characteristic of the CFEM domain is the position conservation of eight cysteine residues, which were observed in the alignment of all PpCFEM proteins (Figure 6a). PpCFEMs 2, 3, 5, 9, and 10 showed additional cysteine residues in the protein sequence, which ranged from 9-15 residues in total (Table 6). The acid aspartic residue needed for the metal-iron coordination in the *C. albicans* proteins was observed in eight of the ten PpCFEM proteins, not being present in PpCFEM1 and PpCFEM9 protein (Figure 6a), suggesting a possible association of these PpCFEM proteins with iron via the heme-iron complex.

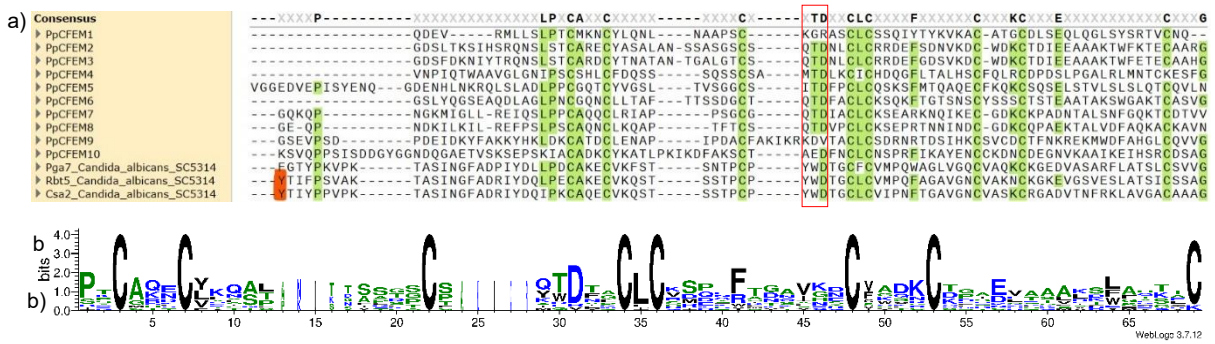



Figure 6: Multi-sequence alignment of CFEM domain from PpCFEM proteins. a) PpCFEM domain alignment against *C. albicans* CFEM domain proteins. The alignment was performed using ClustalW, and the conserved amino acids are highlighted in green, including the eight cysteines. The red box indicates the acid aspartic residue needed for the metal-iron coordination in the *C. albicans* proteins and in red the tyrosine residues observed in CFEM proteins from *C. albicans*. b) PpCFEM domain cysteine residues position conservation, created by WebLogo version 3.7.12.

### 4.3. PpCFEM proteins exhibit secretion signal and may be extracellular or associated with membranes proteins according to silico analysis

The signal peptide for secretion was predicted at the N-terminal portion of nine PpCFEM (PpCFEM1-9) (Table 6), except for PpCFEM10. The GPI anchor site was predicted only for PpCFEM2, PpCFEM3, PpCFEM4, and PpCFEM6. Only three *P. pachyrhizi* CFEM proteins (PpCFEM2, PpCFEM3, and PpCFEM10) exhibited the presence of one transmembrane domain predicted by PHOBIUS and PROTTER tools.

Before the biological assays of subcellular localization, the possible cellular targets of each PpCFEM protein were predicted. PpCFEM proteins were predicted to have an extracellular localization based on DeepLoc (Table 7), though, PpCFEM2, PpCFEM3, PpCFEM4, and PpFEM6 also presented the indication of cell membrane localization. WoLF PSORT predictions classified the PpCFEM proteins with extracellular localization but also suggested that PpCFEM3 could also have localization in the plastids, PpCFEM5 in the cytoplasm, and PpCFEM9 in the Golgi apparatus (Table 8).

Table 6 – Predicted characteristics of PpCFEM proteins.

| CFEM | Protein size (aa) | SignalP 6.0[1] | SP-Phobius[2] | TM-Phobius[3] | POTTER[4] | PredGPI[5] | NETGpi[6] | #Cysteine residues |
|---|---|---|---|---|---|---|---|---|
| PpCFEM1 | 97 | Y (0.999) | Y (1-26) | N | N | N | N | 8 |
| PpCFEM2 | 159 | Y (0.999) | Y (1-21) | Y (135-154) | Y (135-154) | Y (129) | Y (0.545) | 9 |
| PpCFEM3 | 159 | Y (0.999) | Y (1-21) | Y (135-156) | Y (135-156) | Y (129) | Y (0.363) | 9 |
| PpCFEM4 | 245 | Y (0.964) | Y (1-27) | N | N | Y (225) | Y (0.432) | 8 |
| PpCFEM5 | 201 | N (0.677) | Y (1-23) | N | N | N | N | 9 |
| PpCFEM6 | 161 | Y (0.999) | Y (1-25) | N | N | Y (138) | Y (0.525) | 8 |
| PpCFEM7 | 155 | Y (0.999) | Y (1-25) | N | N | N | N | 8 |
| PpCFEM8 | 199 | Y (0.999) | Y (1-29) | N | N | N | N | 8 |
| PpCFEM9 | 221 | N (0.612) | Y (1-27) | N | N | N | N | 10 |
| PpCFEM10 | 424 | N (0.639) | N | Y (7-27) | Y (7-27) | N | N | 15 |

1/2: Signal peptide presence prediction: Y (yes) N (no).
3/4: Number of transmembrane domains prediction.
5/6: GPI Anchor presence prediction: Y (yes) N (no).

Table 7 - PpCFEM proteins localization prediction by DeepLoc

| Protein_ID | Localization | Signals | Cytoplasm | Nucleus | Extrac. | Cell membrane | Mitoch. | Plastid | End. reticulum | Lysosome/ Vacuole | Golgi apparatus | Peroxisome |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PpCFEM1 | Extracellular | Signal peptide | 0.1044 | 0.0572 | 0.9732 | 0.067 | 0.0718 | 0.0027 | 0.0851 | 0.1095 | 0.068 | 0.005 |
| PpCFEM2 | Extracellular | - | 0.0908 | 0.0662 | 0.7544 | 0.5283 | 0.0918 | 0.0042 | 0.1529 | 0.3078 | 0.2427 | 0.0371 |
| PpCFEM3 | Extracellular | - | 0.1183 | 0.0861 | 0.6862 | 0.5398 | 0.0835 | 0.0021 | 0.191 | 0.4584 | 0.3771 | 0.0436 |
| PpCFEM4 | Extracellular | - | 0.093 | 0.0821 | 0.6076 | 0.5121 | 0.0625 | 0.0029 | 0.1548 | 0.1068 | 0.1347 | 0.0103 |
| PpCFEM5 | Extracellular | Signal peptide | 0.1738 | 0.0809 | 0.897 | 0.3925 | 0.0852 | 0.0065 | 0.2499 | 0.3121 | 0.2334 | 0.0135 |
| PpCFEM6 | Extracellular | - | 0.0886 | 0.0994 | 0.6888 | 0.4045 | 0.1128 | 0.0026 | 0.1456 | 0.2338 | 0.165 | 0.0219 |
| PpCFEM7 | Extracellular | Signal peptide | 0.1317 | 0.0774 | 0.9606 | 0.1031 | 0.0865 | 0.0043 | 0.2194 | 0.152 | 0.1546 | 0.008 |
| PpCFEM8 | Extracellular | Signal peptide | 0.1326 | 0.0785 | 0.9655 | 0.1531 | 0.1116 | 0.0066 | 0.1836 | 0.1157 | 0.1728 | 0.0097 |
| PpCFEM9 | Extracellular | Signal peptide | 0.1299 | 0.0703 | 0.967 | 0.1288 | 0.0616 | 0.0207 | 0.1734 | 0.1065 | 0.106 | 0.0091 |
| PpCFEM10 | Extracellular | Signal peptide | 0.1729 | 0.1076 | 0.9495 | 0.1135 | 0.1332 | 0.0041 | 0.3141 | 0.1468 | 0.3336 | 0.0102 |

The numbers correspond to the probability from 0 to 1 of protein localization in subcellular compartments. Only values above or 0.4 were considered, marked with red square.

Table 8 – Subcellular localization of PpCFEM proteins predicted by WoLF PSORT tool.

| Protein_ID | Localization [a] |
|---|---|
| PpCFEM1 | Extracellular: 23, mitochondria: 2, plastid: 1, endoplasmic reticulum.: 1 |
| PpCFEM2 | Extracellular: 24, plastid: 2 |
| PpCFEM3 | Extracellular: 18, plastid: 6, endoplasmic reticulum: 3 |
| PpCFEM4 | Extracellular: 26 |
| PpCFEM5 | Extracellular: 17, cytoplasm: 3, mitochondria: 2, peroxisome: 2, endoplasmic reticulum: 2, nucleus: 1 |
| PpCFEM6 | Extracellular: 26 |
| PpCFEM7 | Extracellular: 26 |
| PpCFEM8 | Extracellular: 20, Golgi apparatus: 3, peroxisome: 2, endoplasmic reticulum: 1 |
| PpCFEM9 | Extracellular: 10, Golgi apparatus: 6, vacuole: 5, endoplasmic reticulum: 3, mitochondria: 2, cytoplasm: 1 |
| PpCFEM10 | Extracellular: 20, vacuole: 3, Golgi apparatus: 2, cytoplasm: 1 |

a: Protein localization prediction based on amino acid sequence and sorting signals presence by comparing the query protein with the 27 nearest proteins in the WoLF PSORT database. The numbers correspond to the number of similar proteins and their subcellular localization.

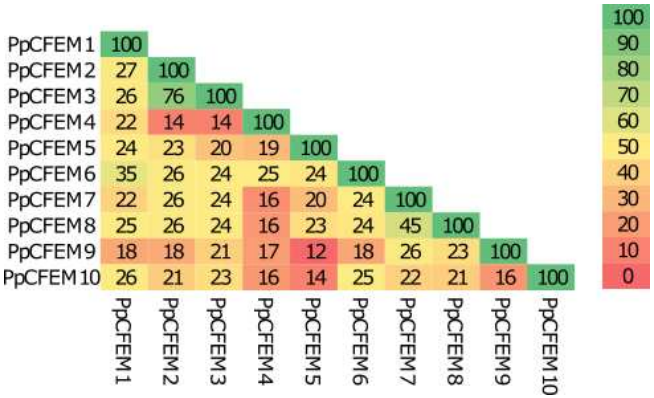
### 4.4. Pairwise comparison between protein sequences containing CFEM domain reveals low inter- and intra-species identity

Sequence comparations among the PpCFEM proteins revealed identity among the sequences (Figure 7a) ranging from 12-76%. PpCFEM2 and PpCFEM3 showed the highest identity percentage (76%), which corroborates with the results of the predictions, indicating a possible homology relation between these proteins. PpCFEM7 and PpCFEM8 presented 45% of identity, while the identity values among the other PpCFEM proteins ranged from 12-27%, with the lowest value corresponding to the identity between PpCFEM5 and PpCFEM9. In phylogenetic analyses, reconstructed using only PpCFEM proteins sequences, two clades are observed (Figure 7b) one formed by the proteins with GPI anchor, transmembrane domain, and cell membrane localization prediction (PpCFEM1, 2, 3, 4, 6, and 10 and the soluble PpCFEM5 protein) (Clade 1-in red), and the other formed exclusively by the extracellular soluble proteins PpCFEM7, 8, and 9 (Figure 7b) (Clade 2-in yellow).

All PpCFEMs were also compared with the CFEMs of other fungi and was expected, the greatest identity was observed with CFEM proteins of the *Puccinia triticina* (identity values ranging from 9 to 41%) (Figure 8), another rust species. The lowest identity values when compared with *P. triticina* CFEMs were observed to PpCFEM 9 and 10, around 9-12%. A phylogenetic tree was constructed with the same CFEM sequences used in pairwise comparations and was observed the clustering in three clades (Figure 9). PpCFEM 1, 7, 8, 9, and 10 were clustered in clade 1, with CFEM proteins of the *Candida* species, suggesting greater phylogenetic proximity among them. For PpCFEM1 and PpCFEM10 was observed a specific clustering with

CFEM18 and CFEM8 of *Fusarium graminearum* (Figure 9), respectively. PpCFEM 2, 3, 4, 5, and 6 clustered with CFEMs from *P. triticina* with supports from 50-100% (Figure 9) in clade 2. There was no grouping of PpCFEMs with CFEM proteins from cluster 3.

a)




b)

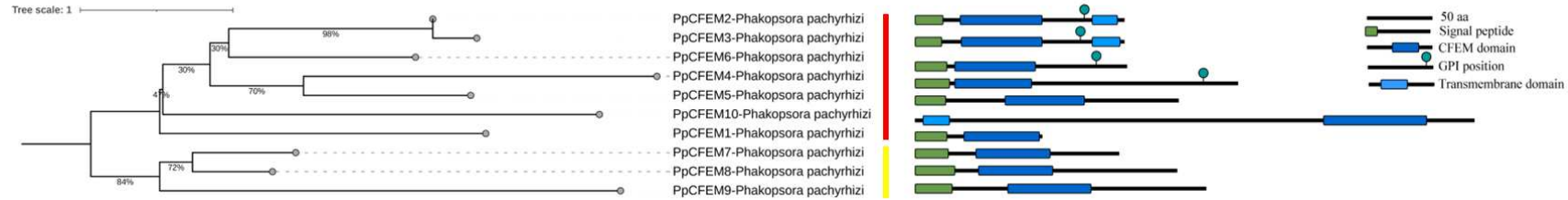



Figure 7: Pairwise identity comparison and phylogenetic analyses of PpCFEM proteins. a) Plot representing the percent amino acid sequence identities among PpCFEM proteins of *P. pachyrhizi* in the range of 0-100% identity. b) Phylogenetic tree constructed based on alignment performed in MAFFT v.7 software with the evolutionary tree constructed using the maximum-likelihood approach. In nodes are presented the bootstrap value determined in IQ-TREE. The consensus trees were visualized and rooted in midpoints using iTOL v.5 (Letunic and Bork, 2021). In red is represented clade 1 and in yellow clade 2. The presence of signal peptide, GPI anchor, and transmembrane domain can be visualized together with the phylogenetic. CFEM domains are observed in the N-terminal position in all PpCFEM proteins except in PpCFEM10.

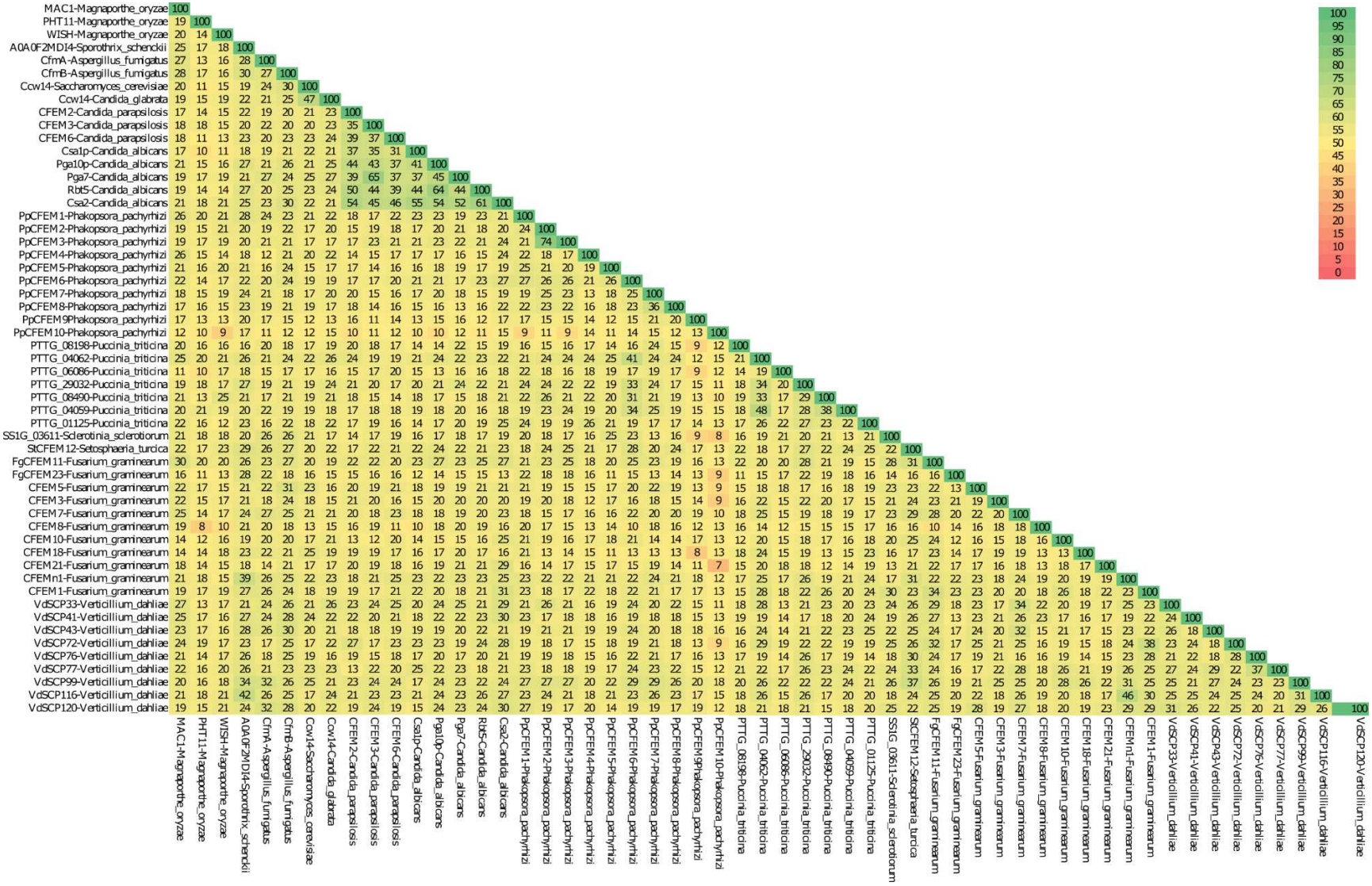



Figure 8: Pairwise identity comparison and phylogenetic analyses of CFEM proteins. Plot representing the percent amino acid sequence identities among CFEM proteins of different fungi species the range of 0-100% identity.

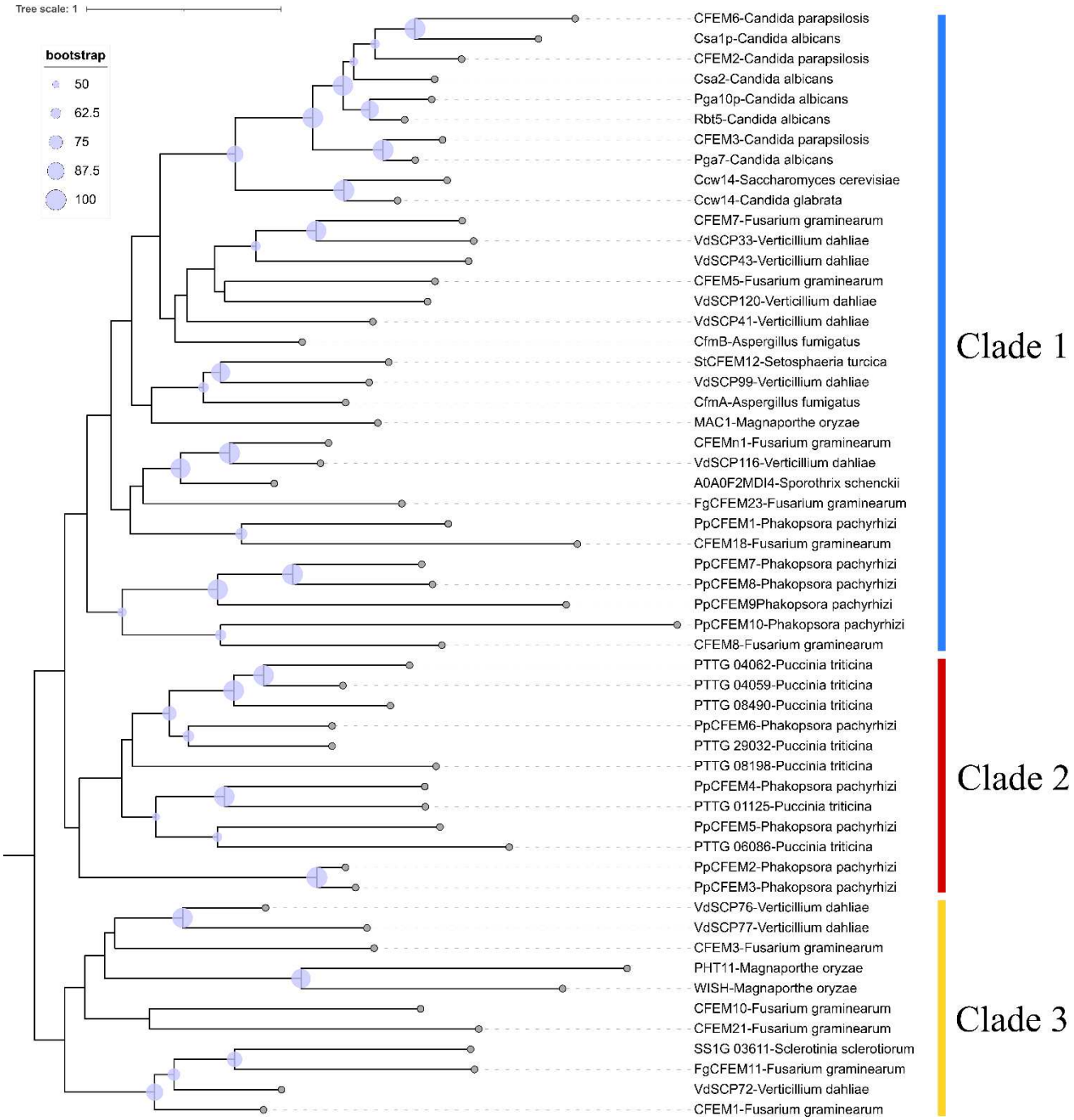



Figure 9: Phylogenetic analyses of PpCFEM proteins and other fungal CFEM proteins. Phylogenetic tree constructed based on alignment performed in MAFFT v.7 software with the evolutionary tree constructed using the maximum-likelihood approach. In nodes are presented the bootstrap value determined in IQ-TREE ranging from 50-100%. The consensus trees were visualized and rooted in midpoints using iTOL v.5 (Letunic and Bork, 2021). Three clades are observed: in blue is represented clade 1, in red the clade 2, and in yellow the clade 3.

### 4.5. Tertiary structures show that PpCFEMs exhibit helical-basket structure, but do not form the pocket structure

PpCFEM tertiary structure was modeled in the AlphaFold program and the results revealed that the CFEM domain of the *P. pachyrhizi* proteins were shaped in a similar helical-basket structure as observed in the Csa2 CFEM protein from *C. albicans* (Figure 10). The six alpha-helices typical from CFEM domains were visualized in all PpCFEMs, except in PpCFEM 9 (Figure 10), and in the PpCFEM 1 and 9 was not possible to find the aspartic acid residue associated with heme-iron coordination (Figure 10). Despite the presence of the conserved aspartic acid residue in the PpCFEM2, 3, 4, 5, 6, 7, 8, and 10 proteins sequences, none of the 10 PpCFEM proteins have the ‘hand-like’ structure formed with tyrosine residue coordination, necessary for the pocket structure formation in the heme-iron bounding as observed for Csa2 CFEM protein from *C. albicans* (Figure 6a and 10). These results suggested that the mechanism for obtaining iron by *P. pachyrhizi* is not in association with the heme group as observed for *C. albicans*.

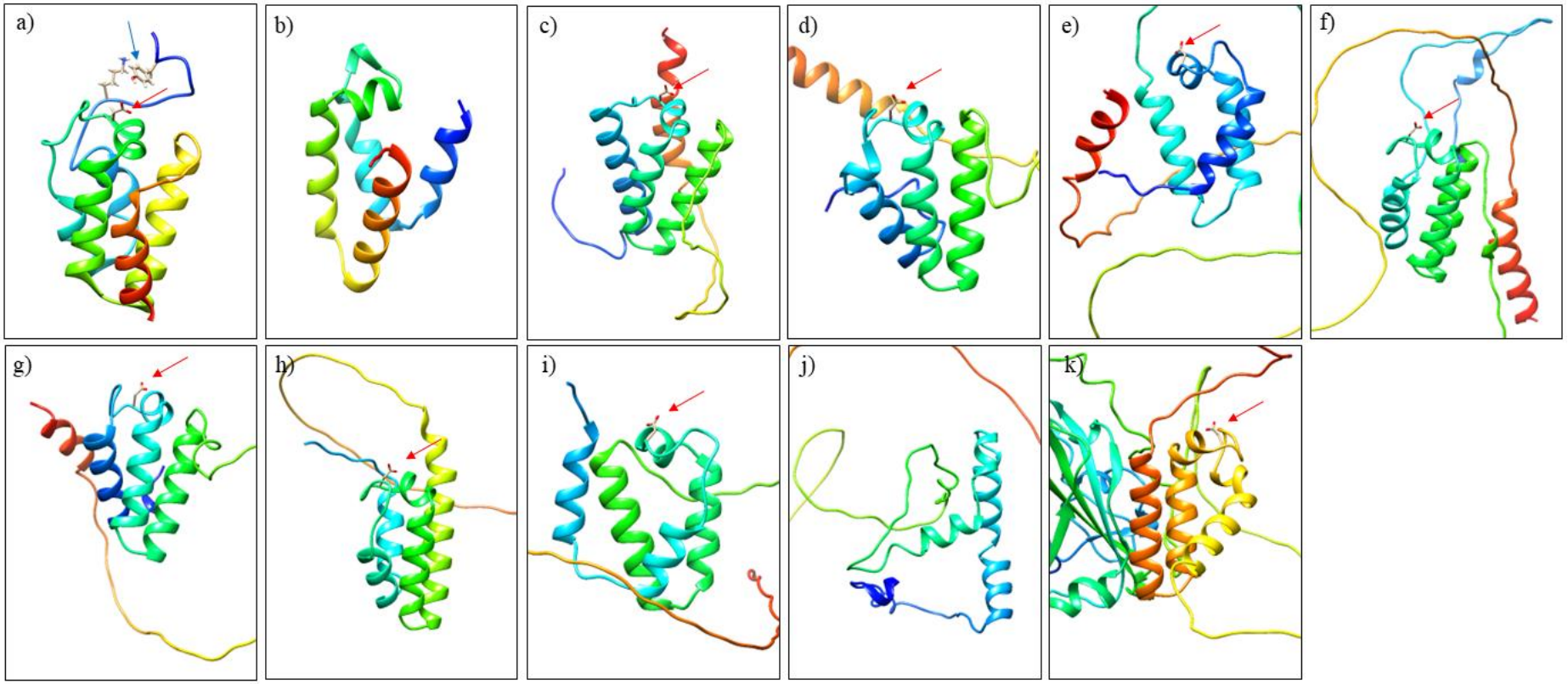



Figure 10: Structural comparison between CFEM protein Csa2 X-ray resolved structure (PDB: 4Y7S) and PpCFEM protein tertiary structure modeled by the AphaFold program. a) CSA2 CFEM protein from *C. albicans* with essential heme-iron coordination amino acids (aspartic acid and tyrosine) in red. b) PpCFEM1. c) PpCFEM2. d) PpCFEM3. e) PpCFEM4. f) PpCFEM5. g) PpCFEM6. h) PpCFEM7. i) PpCFEM8. j) PpCFEM9. k) PpCFEM10. The structures show six typical alpha-helices of CFEM proteins with the colors progressing from blue to red corresponding to N- and C-terminal portions, respectively. Red arrows indicate the aspartic acid amino acid residue needed for metal-iron coordination and not observed in PpCFEM 1 and 9. The blue arrow shows the tyrosine amino acid residue present exclusively in CSA2 CFEM.

### 4.6. *PpCFEM* genes are differentially expressed during the soybean infection process.

To confirm that all *PpCFEM* predicted *in silico* were transcribed during the period of time analyzed, a differential expression analysis was performed using RT-qPCR. Although ten *Pp*CFEM genes we predicted by *in-silico* identification analysis, only seven of them (*PpCFEM1 – 7)* were amplified from the cDNA samples prepared for temporal expression analysis and cloning process. *PpCFEM9* amplification was obtained from genomic DNA (gDNA) samples but not from cDNA samples, demonstrating that this gene is present in gDNA, but is not expressed in the evaluated times. *PpCFEM8* and *PpCFEM10* were not amplified either from gDNA or from cDNA using different pairs of oligonucleotides (Figure 11).

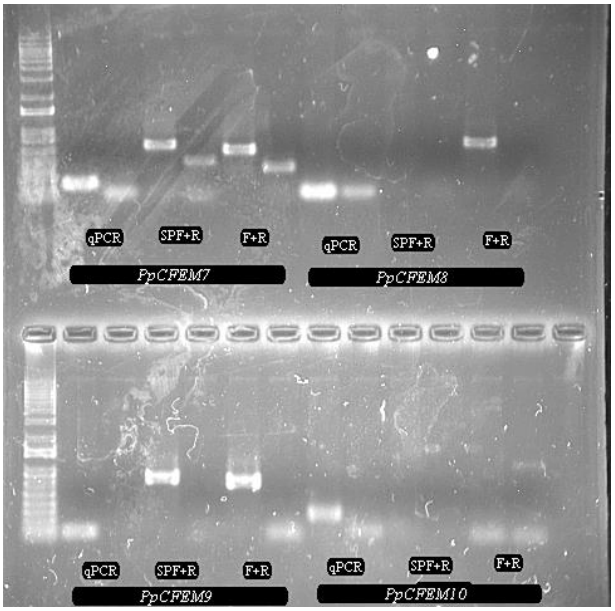



Figure 11: Amplification of *PpCFEMs* by polymerase chain reaction (PCR). PCR amplifications were performed for PpCFEM7-10. The figure demonstrates the amplification with qPCR primers with expected sizes of 100, 129, 109, 138 bp for PpCFEM7, 8, 9, and 10, respectively, SpF+R primers with expected sizes of 468, 600, 666, 1275 bp for PpCFEM7, 8, 9, and 10, respectively and F+R primers with expected sizes of 393, 513, 585, 1200 bp forPpCFEM7, 8, 9 and 10, respectively. Genomic DNA amplification in even channels, and cDNA amplification in odd channels. *PpCFEM7* and *PpCFEM8* are displayed in the first line and *PpCFEM9* and *PpCFEM10* are displayed in the second line.

The expression pattern of seven *Pp*CFEM genes (*PpCFEM1 – 7*) was evaluated in the period of 0 up to 96 hours post-inoculation (hpi) (Figure 12). *PpCFEM1* presented the lowest gene expression in the evaluated time intervals. *PpCFEM2*, *PpCFEM5*, and *PpCFEM6* had the highest levels of gene expression at the 0 hpi, period preceding the

establishment of infection in the host, with a significant reduction in the level expression throughout the infectious process (24-96 hpi). *PpCFEM6* had the highest level of gene expression among the *PpCFEM* analyzed, with expression up to 25 times greater than the endogenous gene. *PpCFEM3, 4,* and *7* were expressed only at 24 hpi, showing lower relative level of expression (below 1) in the other evaluated time intervals (Figure 12).

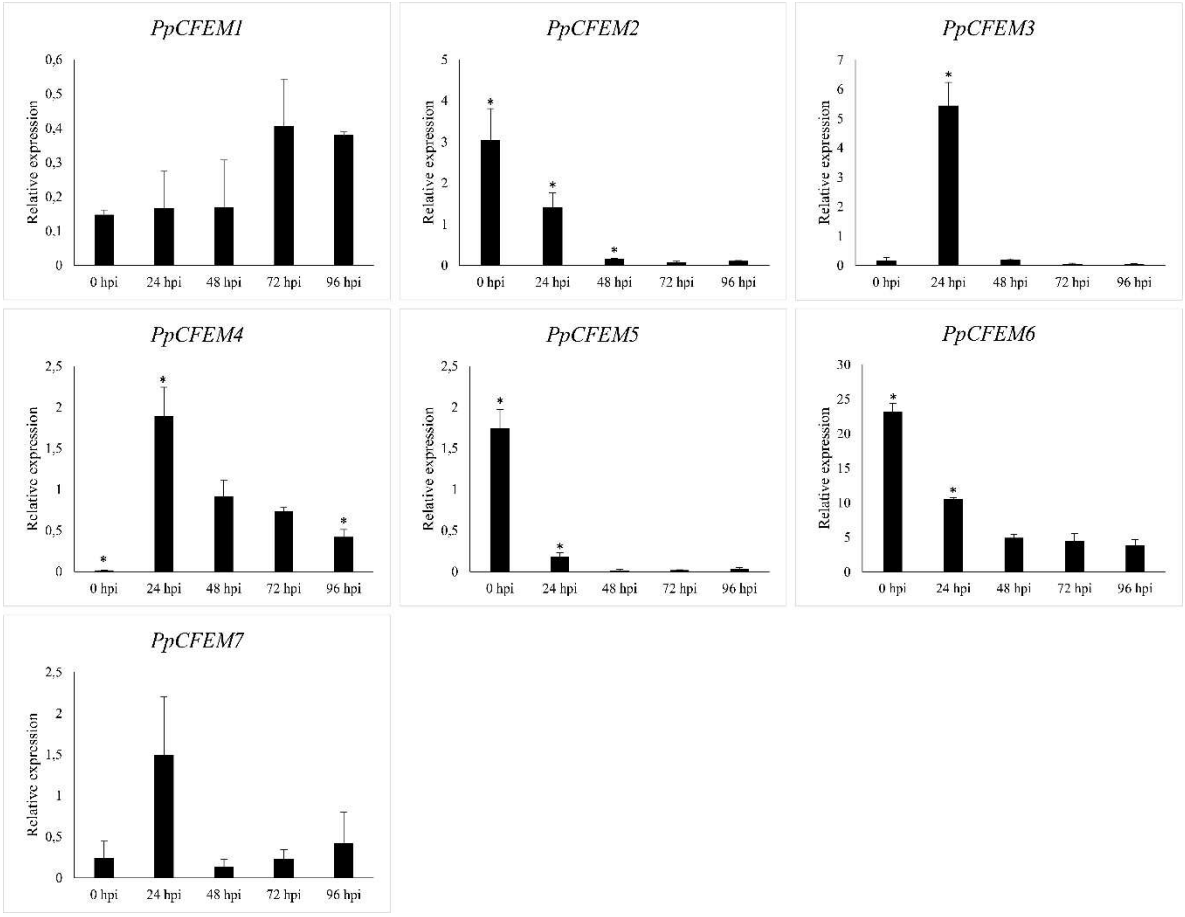



Figure 12: Expression analysis of *PpCFEM* genes in the soybean infection process. Transcript accumulation was determined for *PpCFEM 1-7* by real-time quantitative polymerase chain reaction (RT-qPCR) in samples obtained from infected leaves at 0, 24, 48, 72, and 96 hours post inoculation (hpi). Points represent mean of relative expression of the *PpCFEM* normalized using the *α-tubulin* gene as a reference (Kunjeti *et al*., 2016). All analyses were done using three biological and three technical repetitions/time-points were compared using Student's t-test at $p<0.05$.

### 4.7. Subcellular localization of PpCFEM proteins

To confirm the *in-silico* predictions of subcellular localizations, the coding regions without the region encoding predicted signal peptide, of each one of the seven *PpCFEM* (1-7) were fused to GFP, the in vectors pK7FWG2 (C-Terminal GFP fusion) and pK7WGF2 (N-terminal GFP fusion), and transiently expressed in *N. benthamiana*

leaves together with marker proteins of cellular compartments (nucleus, vacuole, plasmatic membrane or plastid) fused with an m-cherry fluorescent protein (Figure13-19).

The analyses showed that PpCFEM1, 3, 5, and 7 (Figures 13, 15, 17, 19) displayed localization in the nucleus and cytoplasm. PpCFEM2 and PpCFEM6 (Figures 14, 18) were observed in the cytoplasm and plasmatic membrane. while PpCFEM4 did not show any specific subcellular localization, with punctual structures observed all over the cells (Figure 16).

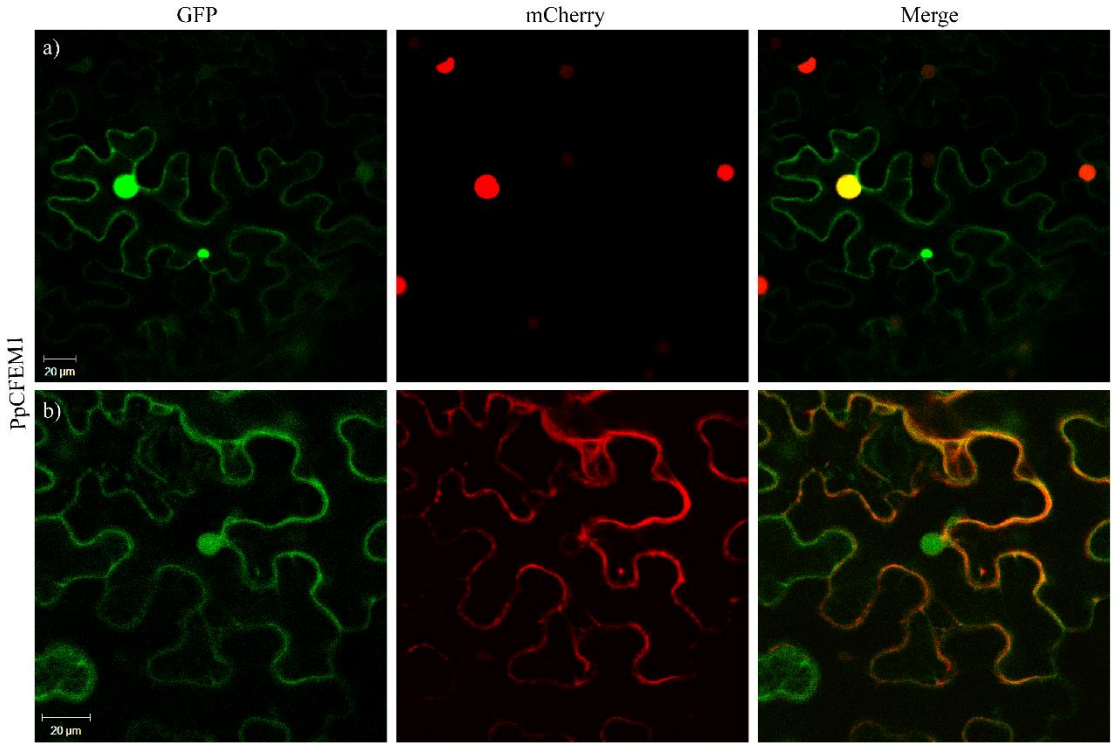



Figure 13: PpCFEM1 protein subcellular localization in *N. benthaniana* leaves. a) Co-infiltration of pK7FWG2::PpCFEM1 + pMP90::WWP1 (Nuclear marker); b) Co-infiltration of pK7FWG2::PpCFEM1 + pMP90::*vac* (Vacuole marker).

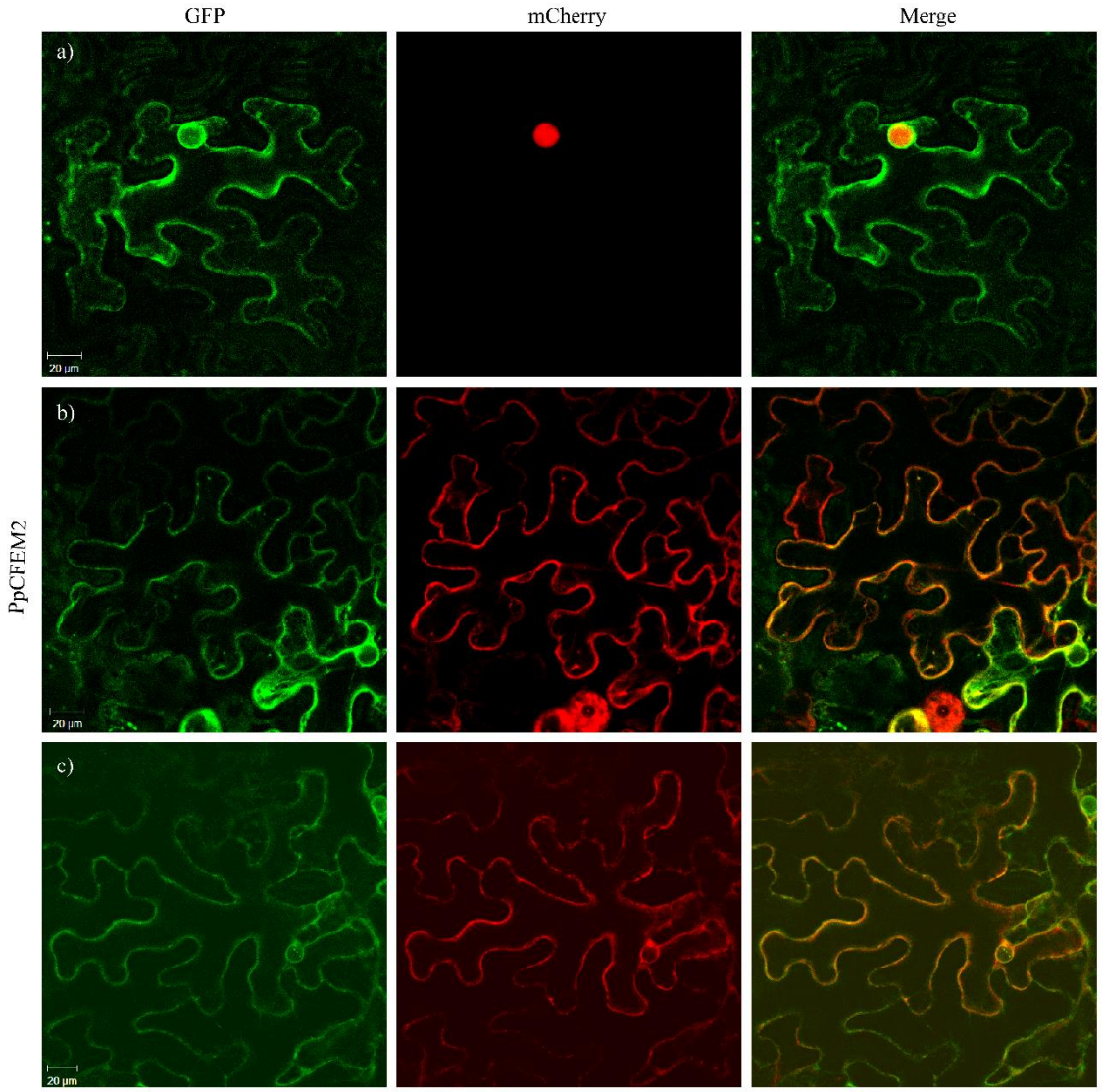



Figure 14: PpCFEM2 protein subcellular localization in *N. benthaniana* leaves. a) Co-infiltration of pK7FWG2::PpCFEM2 + pMP90::WWP1 (Nuclear marker); b) Co-infiltration of pK7FWG2::PpCFEM2 + pMP90::*vac* (Vacuole marker); c) Co-infiltration of pK7FWG2::PpCFEM2 + pMP90::*mp* (Membrane marker).

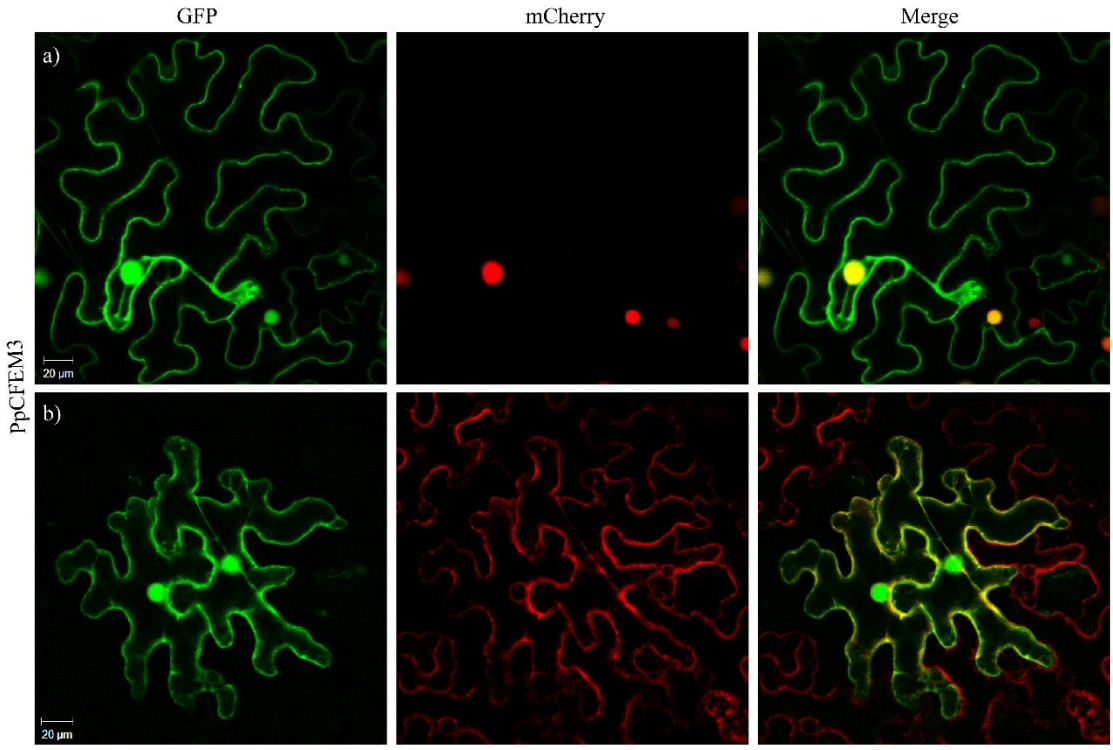



Figure 15: PpCFEM3 protein subcellular localization in *N. benthaniana* leaves. a) Co-infiltration of pK7FWG2::PpCFEM3 + pMP90::WWP1 (Nuclear marker); b) Co-infiltration of pK7FWG2::PpCFEM3 + pMP90::*vac* (Vacuole marker).

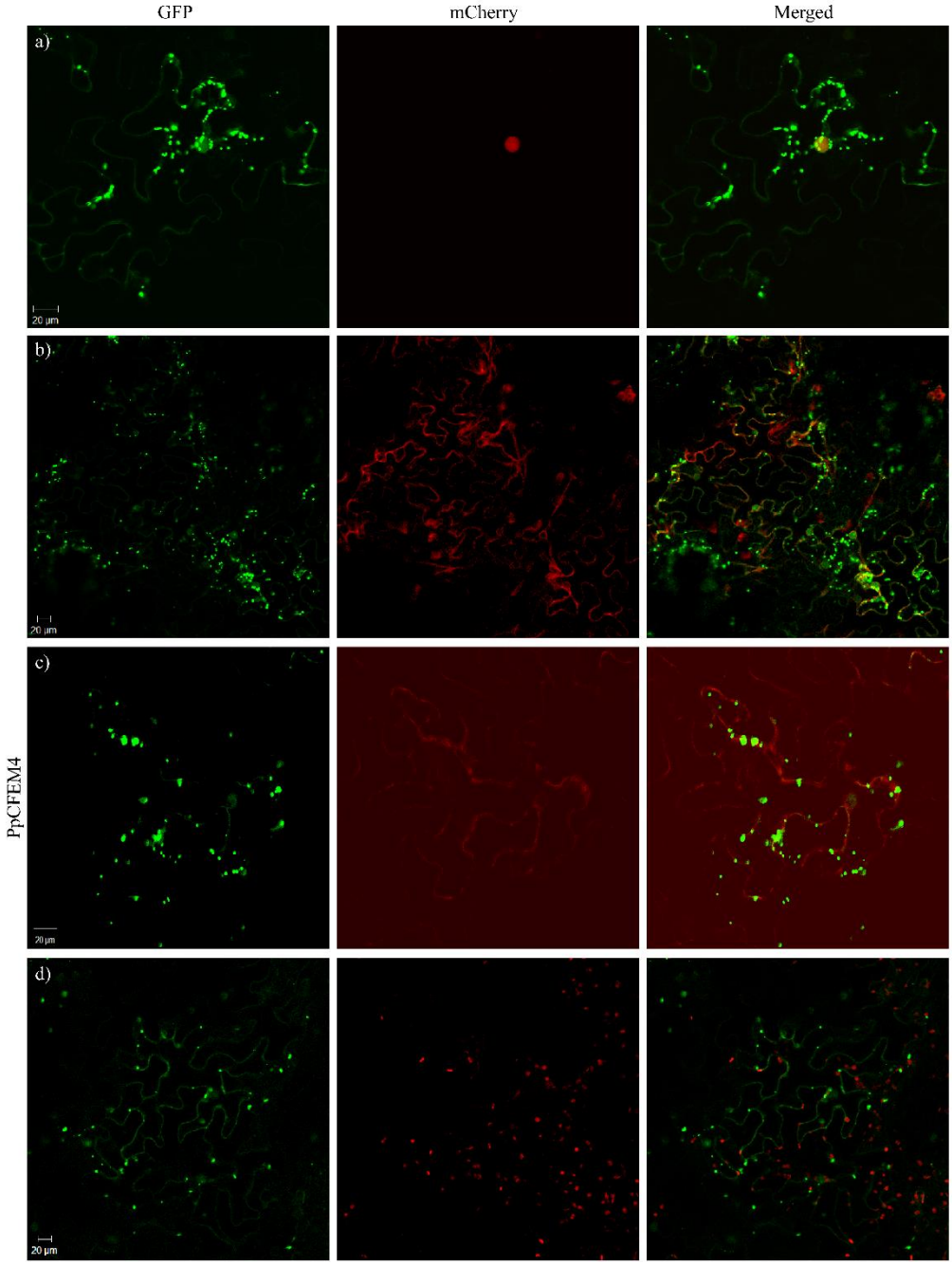



Figure 16: PpCFEM4 protein subcellular localization in *N. benthaniana* leaves. a) Co-infiltration of pK7WGF2::PpCFEM4 + pMP90::WWP1 (Nuclear marker); b) Co-infiltration of pK7WGF2::PpCFEM4 + pMP90::*vac* (Vacuole marker); c) Co-infiltration of pK7WGF2::PpCFEM4 + pMP90::*mp* (Membrane marker). d) Co-infiltration of pK7WGF2::PpCFEM4 + pMP90::*pt* (Plastid marker).

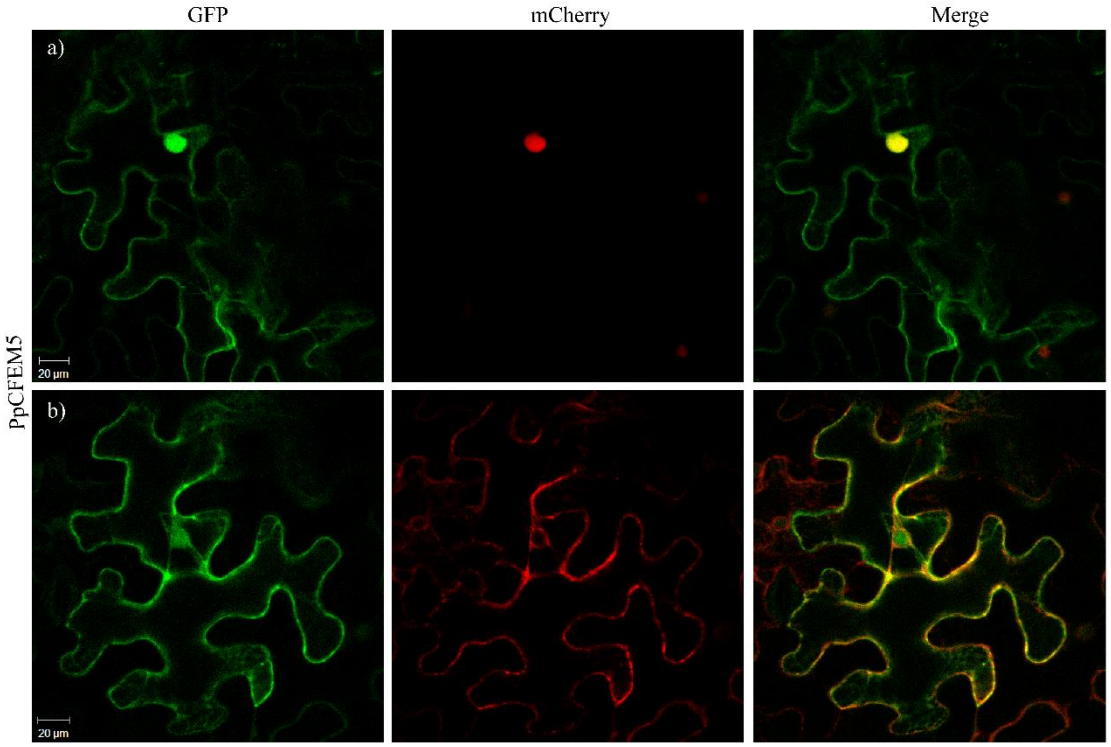



Figure 17: PpCFEM5 protein subcellular localization in *N. benthaniana* leaves. a) Co-infiltration of pK7WGF2::PpCFEM5 + pMP90::WWP1 (Nuclear marker); b) Co-infiltration of pK7WGF2::PpCFEM5 + pMP90::*vac* (Vacuole marker).

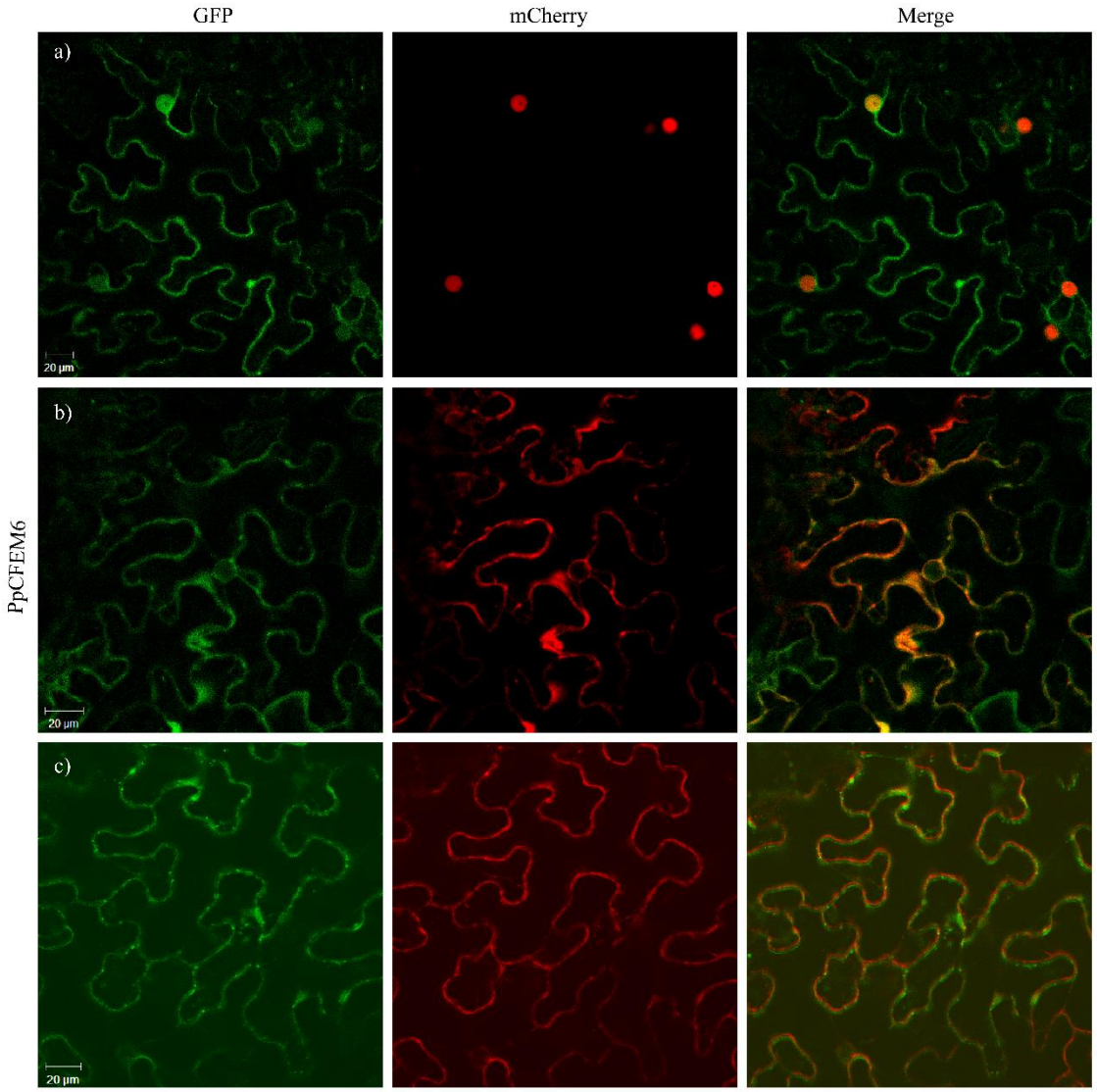



Figure 18: PpCFEM6 protein subcellular localization in *N. benthaniana* leaves. a) Co-infiltration of pK7FWG2::PpCFEM6 + pMP90::WWP1 (Nuclear marker); b) Co-infiltration of pK7FWG2::PpCFEM6 + pMP90::*vac* (Vacuole marker); c) Co-infiltration of pK7FWG2::PpCFEM6 + pMP90::*mp* (Membrane marker).

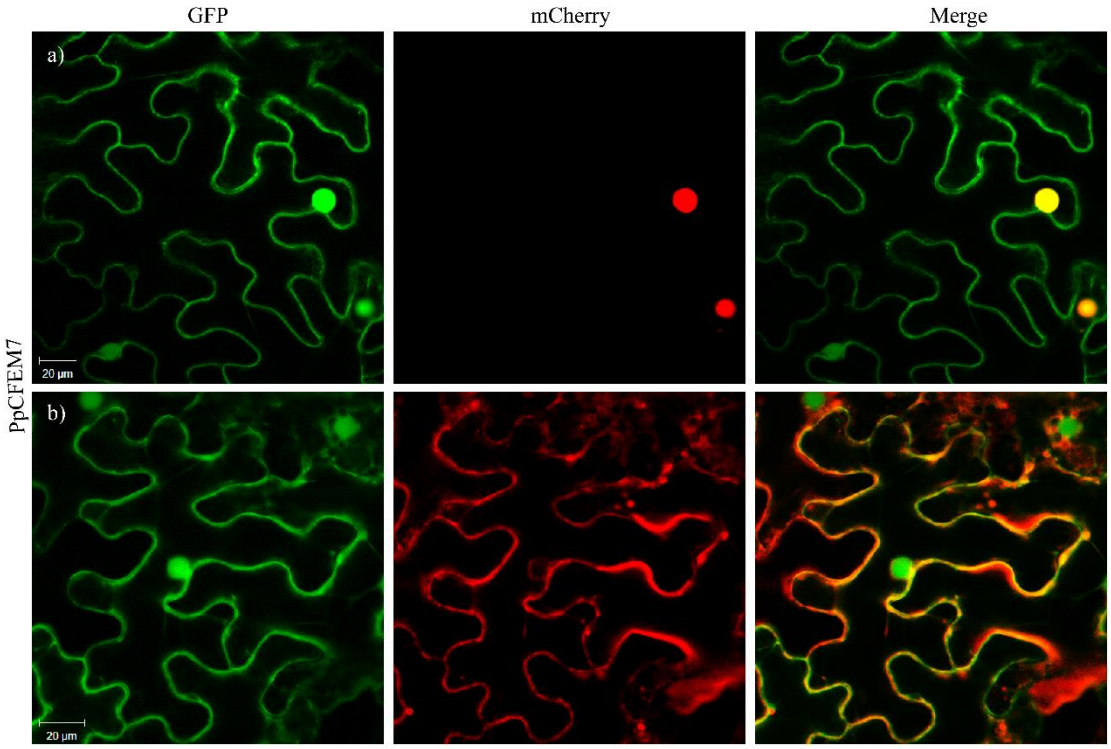



Figure 19: PpCFEM7 protein subcellular localization in *N. benthaniana* leaves. a) Co-infiltration of pK7FWG2::PpCFEM7 + pMP90::WWP1 (Nuclear marker); b) Co-infiltration of pK7FWG2::PpCFEM7 + pMP90::*vac* (Vacuole marker).

### 4.8. PpCFEM proteins are able to suppress PTI and ETI in *N. benthamiana* leaves by PpCFEM proteins

PpCFEM1, PpCFEM4, PpCFEM5, and PpCFEM7 proteins were capable of PTI suppression (Figure 20), while only the PpCFEM4 protein was capable of ETI suppression in *N. benthamiana* leaves (Figure 21). These results indicate that these five proteins could be classified as candidate effector proteins of *P. pachyrhizi* (Table 10).

Table 10 - Suppression of PTI and ETI by Pp-CFEMs in *Nicotiana benthamiana* plants.

| PpCFEM | PTI suppression [a] | ETI suppression [b] | EffectorP prediction [c] |
|---|---|---|---|
| PpCFEM1 | + | - | Apoplastic effector |
| PpCFEM2 | - | - | Apoplastic effector |
| PpCFEM3 | - | - | Apoplastic effector |
| PpCFEM4 | + | + | Non-effector |
| PpCFEM5 | + | - | Apoplastic effector |
| PpCFEM6 | - | - | Apoplastic effector |
| PpCFEM7 | + | - | Cytoplasmatic / Apoplastic effector |

a: The proteins that maintain hypersensitivity reaction (HR) in the intersection area in more than 60% of the repetitions after 48 hours post-inoculation were considered capable of PTI suppression (+).
b: The proteins that maintain the inoculation area without hypersensitivity reaction (HR) in more than 60% of the repetitions after 48 hours post-inoculation were considered capable of ETI suppression (+).
C: The EffectorP predictor was used to predict the PpCFEM1-7 as effector proteins.

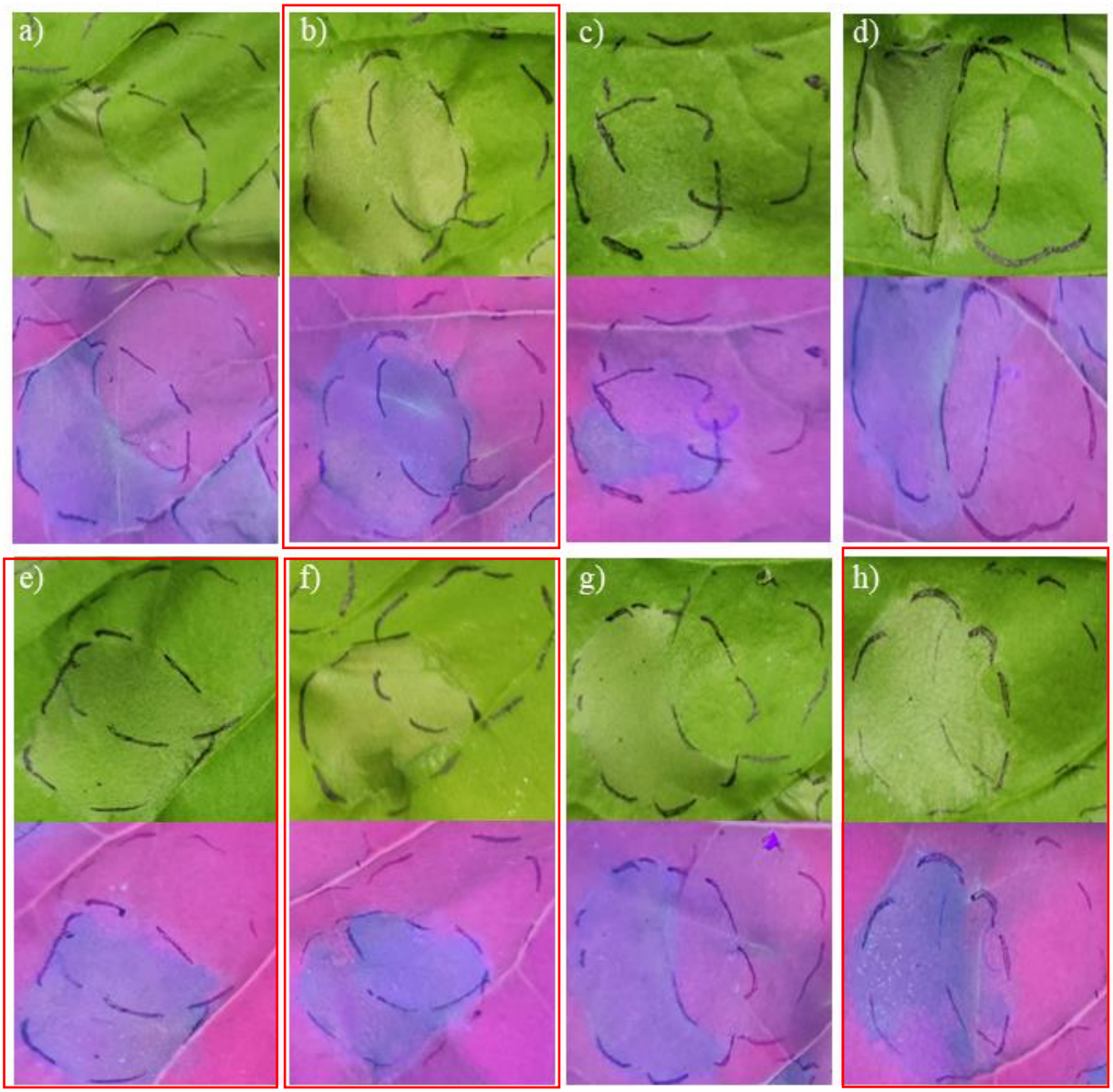



Figure 20: PTI suppression patterns of seven PpCFEM proteins in *N. benthamiana* leaves. Bacterial suspensions of *Pseudomonas fluorescens* EtHAn carrying pEDV6-empty vector and *P. fluorescens* EtHAn expressing each pEDV6::PpCFEM construction at $OD_{600}$: 0.2 were pre-infiltrated and after seven hours a suspension of *P. syringae* pv. *garcae* at OD was also infiltrated keeping an overlapping area. a) negative control – pEDV6 empty vector. b) pEDV6::PpCFEM1. c) pEDV6::PpCFEM2. d) pEDV6::PpCFEM3. e) pEDV6::PpCFEM4. f) pEDV6::PpCFEM5. g) pEDV6::PpCFEM6. h) pEDV6::PpCFEM7. Pictures were taken 48 hours post *Psg* inoculation under bright light (above) and long-wave UV light (down). The images show representative inoculations from independent experiments.

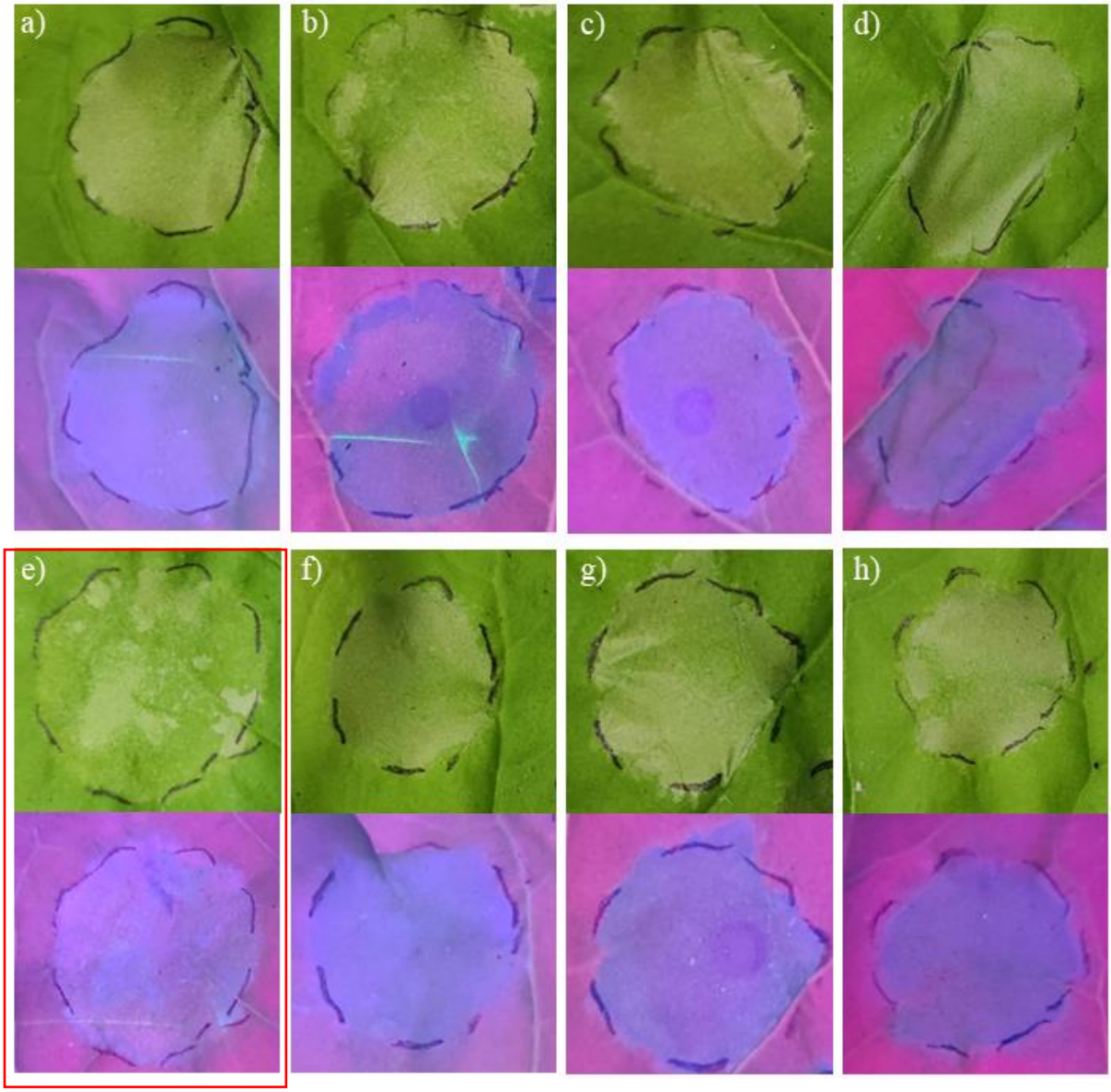



Figure 21: ETI suppression patterns of the sevenPpCFEM1-7 proteins in *N. benthamiana* leaves. Suspensions of *P. fluorescens* EtHAn expressing pVSP6::*AvrB* and *P. fluorescens* EtHAn expressing pEDV6::PpCFEM constructions (AvrB + PpCFEM) were co-infiltrated at a 1:1 ratio to a final of $OD_{600}$: 0.15 (black dotted circle) in *N. benthamiana* leaves. a) negative control – pEDV6 empty vector. b) pEDV6::PpCFEM1. c) pEDV6::PpCFEM2. d) pEDV6::PpCFEM3. e) pEDV6::PpCFEM4. f) pEDV6::PpCFEM5. g) pEDV6::PpCFEM6. h) pEDV6::PpCFEM7. Pictures were taken 48 hours post-inoculation under bright light (above) and long-wave UV light (down). The images are representative inoculations of independent experiments.

## 5. DISCUSSION

*In silico* analysis demonstrated that *P. pachyrhizi* has 10 genes encoding CFEM proteins (*PpCFEM1-PpCFEM10*). This number is close to the number reported to *Puccinia graminis f. sp. tritici* isolate 21-0 haplotype A, another rust fungus, that has 8 predicted genes encoding CFEM proteins (Li et al., 2017). The number of CFEM genes presented by a fungus species can be related to virulence and pathogenicity, *F. graminearum* causal agent of the fusarium blight disease, for example, contains 23 CFEM genes (Cheng et al., 2021), while *S. cerevisiae* a non-pathogenic fungus presents only one CFEM gene (CCW14) (Zhang et al., 2015). Similar variation also is observed among rust species.

All PpCFEM are single copy genes. Since *P. pachyrhizi* is a dikaryotic fungi allelic variations were also identified for *PpCFEM1, 3, 5, 6, 8, and 10* genes. These allelic forms possess single nucleotide variations (SNPs) that lead to amino acid substitutions, which can lead to protein function/recognition pattern alteration. However, no significative alterations in *P. pachyrhizi* alleles predictions were observed, and for this reason, just one allele for each PpCFEM gene was further analyzed in this study.

As observed for other fungi, the proteins encoded by *PpCFEM* shows a wide variation in size. The predicted protein size for the ten PpCFEM varied from 97 (PpCFEM1) up to 424 (PpCFEM10) amino acids (aa). *F. graminearum* fungus, has 23 CFEM proteins with lengths ranging from 95 up to 864 aa (Cheng et al., 2021). The size of CFEM proteins from *C. graminicola* varies from 74 up to 538 aa (Dong et al., 2019). This variation may reflect the different cellular function of CFEM proteins.

The CFEM domain was present in the N-terminal region of nine PpCFEM proteins (PpCFEM1-9) and in the C-terminal region of only one (PpCFEM10), this differential localization could be functionally associated. PpCFEM10 is the only *P. pachyrhizi* CFEM protein classified as a protein family member, the AA9 family. The “Auxiliary Activity” (AA) families represent a redox enzymes group, which act with glycosyl hydrolases (GH), polysaccharide lyases, and carbohydrates, reducing the plant cell wall recalcitrance. There are 13 classes of AA, and four of them are classified as fungal lytic polysaccharide monooxygenases (LPMO) including AA9, which are copper-dependent enzymes from fungi and acting depolymerization of polysaccharide

(Monclaro et al., 2016). It was already demonstrated that this group of proteins presents a weak endoglucanase activity and acts as an enhancer for other cellulolytic enzymes (Bauer et al., 2006). The prediction of the CFEM domain in the C-terminal region of PpCFEM10, and the prediction of a transmembrane domain in the N-terminal region indicates that this protein may have a cellular function distantly related to the function of the other PpCFEM. However, gene amplification from gDNA and cDNA samples for *PpCFEM8, 9*, and *10* was not successful preventing further functional analysis. We confirmed the prediction of PpCFEM*8, PpCFEM9, and PpCFEM10* in the genomes of *P. pachyrhizi* isolates K8108, UFV02, and MT2006 available at JGI. The failure of cDNA amplification may reflect the period of time in which the gene is expressed, perhaps occurring at shorter intervals than the evaluated times or being expressed after 96 hours.

All PpCFEMs showed the eight cysteine residues typical of the CFEM domain, but PpCFEM4 presents one less amino acid residue in the interval between the seventh and eighth cysteine. This modification affects the three-dimensional structure of the protein (figure 10) by changing the position of the sixth alpha-helix structure. Five CFEM proteins from *F. graminearum* (CFEM 8, 9, 15, 18, and 21), of a total of 23, that did not possess the eight cysteine residues were classified as CFEM proteins by Cheng et al. (2021), what may inflate the number of CFEM proteins reported for this species. In addition to, these authors did not perform tridimensional structure prediction and functional analyses to verify the effects of the lack of cysteine residues on the CFEM domain. Wang et al. (2022), demonstrate that the substitution of one cysteine residue on the CFEM domain of the VdSCP76 and VdSCP77 CFEM proteins eliminated their suppression immunity function, demonstrating the essential function of the disulfide bonds in the stabilization and function of CFEM proteins.

The presence of signal peptide was predicted for nine PpCFEM proteins (PpCFEM1-9) and cellular localization predictions suggested that all the nine PpCFEM proteins are secreted proteins. In addition, PpCFEM2, 3, and 10 were predicted to be associated with the cell membrane by a transmembrane domain, and PpCFEM4 and PpCFEM6 were associated with the cell membrane by a GPI anchor. In phylogenetic analysis, realized exclusively with PpCFEM proteins, resulted in clustering correlated with their structural characteristics, and presence or absence of transmembrane domains/ GPI anchor. The same was observed for the *C. graminicola* fungus, where

the 24 CFEM proteins were grouped in two different clades based on the presence or absence of transmembrane domains (Dong et al., 2019).

The *in silico* subcellular localization predictions were not supported by the biological assays. Only PpCFEM2 and PpCFEM6 protein localization in the membrane was confirmed. Additional experiments are necessary to confirm PpCFEM3 and PpCFEM4 subcellular localization. PpCFEM3 was observed in the nucleus and cell cytoplasm, while PpCFEM4 did not present a precise cellular localization, with punctual aggregates been observed on the cell space. It has been observed that carboxy terminus (C′) or amino terminus (N′) GFP fusion can affect protein folding, and post-translational modifications, interfering with protein delivery and localization (Weill et al., 2019). Weill et al. (2018), demonstrated that for 10% of *S. cerevisiae* proteins analyzed the C-terminus or N-terminus GFP fusion generated different cellular localization patterns. Therefore, more precise localization of PpCFEM will require both GFP fusions analysis since, except for PpCFEM4 and PpCFEM5, all PpCFEM proteins were only analyzed with the C-terminus GFP fusion.

Some PpCFEM showed plant immunity suppression activity. PpCFEM1, 5, and 7 were capable of PTI suppression. In phylogenetic tree analyses, PpCFEM1 clustered in clade 1 with the FgCFEM18 of *F. graminearum* (Chen et al., 2021), indicating that these proteins may have similar function. PpCFEM1 exhibited a very low expression, the smallest among the PpCFEM analyzed, a similar characteristic observed for FgCFEM18. However, FgCFEM18 lacks three cysteine residues in the CFEM domain. Like PpCFEM1, PgCFEM18 was also predicted as an effector by the EffectorP predictor and exhibited a signal peptide, suggesting that both proteins are secreted extracellular proteins.

PpCFEM5 clustered together with the protein PTTG_06086 from *P. triticina* (clade 2) which was also predicted to be an effector protein (Zhao et al., 2020). PTTG_06086 was not able to suppress BAX-induced HR (ETI suppression) in tobacco plants, but the PTI suppression was not tested. PpCFEM5 was specifically expressed at the beginning of the infection process (0 hpi) when the host cell entry process occurs and exhibits nuclear and cytoplasmatic localization, reinforcing its probable function as and effector at this stage of infection. PpCFEM7 clustered with others PpCFEMs (PpCFEM8, 9, and 10, not characterized), in clade 1, together with CFEMs proteins

from yeast-like fungi without any immunity suppression activity reported. PpCFEM7 occurred the expression induction at 24 hpi, a time interval that coincides with the formation of the first haustoria and infection establishment suggestion a key role at this stage.

PpCFEM4 protein was capable of PTI and ETI suppression mechanisms induced by *Psgc* and *P. fluorencens EtHAn AvrBb* in *N. benthamiana* leaves, respectively. In phylogenetic analyses PpCFEM4 clustered in clade 2 with PTTG_01125 a CFEM protein from *P. triticina* (Zhao et al., 2020). Like PpCFEM4, PTTG_01125 was also classified as a non-effector protein by EffectorP (Table 10). Both proteins exhibit a signal peptide in the N-terminal portion and an extracellular portion, in which the CFEM domain is located. PpCFEM4 did not show a defined subcellular localization been observed to be associated with punctual structures dispersed over the cell. However, PpCFEM4 also showed specific gene expression at 24 hpi, coinciding with the haustoria formation period. Therefore, it is possible that this protein is secreted by a non-conventional pathway to be internalized into the plant cell and to perform its immunity suppression activity.

Although genes *PpCFEM2*, *3*, and *6* showed the highest gene expression at 0 - 24 hpi and they were not able to suppress plant immunity. It is possible that the encoded proteins have a structural function at this infection stage. The CFEM protein Bcin07g03260 from *B. cinerea*, that is expressed at 16 hpi is involved in conidial germination (Arya et al., 2020). PpCFEM 2 and 3 show the greatest sequence similarity to each other among the PpCFEMs and may have related functions. PpCFEM2, 3, and 6 clustered in clade 2 with CFEM proteins from *P. triticina*, indicating similar functions. PpCFEM6 which was considered an apoplastic effector by Effector predictor, clustered with PTTG_29032, which was considered a non-effector candidate because it was incapable of suppressing HR induced by BAX gene in *N. benthamiana* plants (Zhao et al., 2020). Both proteins were predicted to be extracellular proteins with a signal peptide.

PpCFEM6 protein corresponds to the PP-CSEP-33 described by Kunjeti et al. (2016) as a candidate effector protein. We could not confirm the bleaching phenotype initially observed in the *N. benthamiana* transient assays and do not have the “hand-like” structure essential for the heme-iron stabilization as presented in the CFEM Csa2 protein, excluding any the possibility to function in an iron-capture cascade.

Similar to PpCFEM6, PpCFEM2, 3, 4, 5, 7, 8, and 10 have the aspartic acid residue needed for metal-iron coordination. However, the AlphaFold protein modeling results indicated that none of the PpCFEM proteins present the “hand-like” structure essential for the heme-iron stabilization as presented in the CFEM Csa2 protein (Nasser et al., 2016). It is possible that *P. pachyrhizi* utilizes alternatives mechanisms to uptake iron from soybean plants. The first would be based on in the obtention of $Fe^{3+}$ form from the host by Iron-Regulated Transporter 1 (IRT1) and the second one that is based on the phytosiderophores exportation, by transporter of mugineic acid 1 (TOM1) to the extracellular space to chelate $Fe^{3+}$; a complex mugenic acid (MA) is formed and MA-$Fe^{3+}$ is imported by transporter YS1. Preliminary search results showed that K8101 isolate of *P. pachyrhizi* possesses a homologue protein of IRT1 classified as a $Fe^{2+/}Zn^{2+}$ regulated transporter (ID 7409316) and three proteins were also predicted as TOM1 proteins (ID 7395818, 3121029 and 7507346), suggesting that this mechanism deserves further investigation.

## 6. CONCLUSIONS

This study allows a better understanding of the CFEM proteins of the fungus *P. pachyrhizi* isolate K8108. Our findings demonstrate the presence of ten CFEM proteins in the *P. pachyrhizi* protein repertoire, and functional analyses indicate that the PpCFEM1, 4, 5, and 7 proteins can present effector molecule functions, and PpCFEM2, 3, and 6 proteins could present structural functions in the cellular process that occurs at the beginning of the infectious process.

More studies are needed for the functional characterization of the PpCFEM functions in the host-pathogen interaction scenario. *In silico* analyses indicate that none of the PpCFEM proteins characterized seems to have a heme-iron capture function as observed in *C.* albicans.

## 7. REFERENCES

AKMAKJIAN, G. Z.; RIAZ, N.; GUERINOT, M. L. Photoprotection during iron deficiency is mediated by the bHLH transcription factors PYE and ILR3. **PNAS**, v. 118, n. 40, 2021 e2024918118.

ARYA, G. C.; SRIVASTAVA, D. A.; PANDARANAYAKA, E. P. J.; MANASHEROVA, E.; PRUSKY, D. B.; ELAD, Y.; FRENKEL, O.; DVIR, H.; HAREL, A. Characterization of the Role of a Non-GPCR Membrane-Bound CFEM Protein in the Pathogenicity and Germination of *Botrytis cinerea*. **Microorganisms**, v, 8, 2020. doi:10.3390/microorganisms8071043.

BLUM, M.; CHANG, H.; CHUGURANSKY, S.; GREGO, T.; KANDASAAMY, S.; MITCHELL, A.; NUKA, G.; PAYSAN-LAFOSSE, T.; QURESHI, M. et al. The InterPro protein families and domains database: 20 years on. **Nucleic Acids Research**. 2020, (doi: 10.1093/nar/gkaa977)

BROMFIELD, K. R. Soybean Rust. St. Paul, MN: **American Phytopathological Society**. v. 65, 1984.

BUENO, T. V.; FONTES, P. P.; ABE, V. Y.; UTIYAMA, A. S.; SENRA, R. L.; OLIVEIRA, L. S.; SANTOS, A. B.; FERREIRA, E. G. C.; DARBEN, L. M.; OLIVEIRA, A. B.; ABDELNNOR, R. V.; WHITHAM, S. A.; FIETTO, L. G.; MARCELINO-GUIMARÃES, F. C. A *Phakopsora pachyrhizi* Effector Suppresses PAMP-Triggered Immunity and Interacts with a Soybean Glucan Endo-1,3-β-Glucosidase to Promote Virulence. **MPMI,** v. 35, n. 9, p. 779–790, 2022. https://doi.org/10.1094/MPMI-12-21-0301-R

BAUER, S.; VASU, P.; PERSSON, S.; MORT, A. J.; SOMERVILLE, C. R. Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. **Proc. Natl. Acad. Sci. U.S.A.** 103, 11417-22, 2006.

CADIEUX, B. et al. The mannoprotein Cig1 supports iron acquisition from heme and virulence in the pathogenic fungus Cryptococcus neoformans. **Journal of the Infectious Diseases**, v. 207, p. 1339–1347, 2013.

CAI, N.; LIU, R.; YAN, D.; ZHANG, N.; ZHU, K.; ZHANG, D.; NONG, X.; TU, X.; ZHANG, Z.; WANG, G. Bioinformatics Analysis and Functional Characterization of the CFEM Proteins of Metarhizium anisopliae. **Journal of Fungi**, v.8, v.661, 2022 .https://doi.org/10.3390/jof8070661

CALIL, I. P.; QUADROS, I. P. S.; ARAUJO, T. C.; DUARTE, C. E. M.; GOUVEIA-MAGESTE, B. C.; SILVA, J. C. F.; et al. A WW Domain-Containing Protein Forms Immune Nuclear Bodies against Begomoviruses. **Molecular plant**, v. 11, p.1449-1465, 2018.

CATANZARITI, Ann-M.; DODDS, P. N.; ELLIS, J. G. Avirulence proteins fromhaustoria-forming pathogens. **FEMS Microbiol Lett**, n. 269, p. 181–188, 2007.

CHEN, L.; WANG, H.; YANG, J.; YANG, X.; ZHANG, M.; ZHAO, Z.; FAN, Y.; WANG, C.; WANG, J. Bioinformatics and Transcriptome Analysis of CFEM Proteins in Fusarium graminearum. **Journal of Fungi**, 7, 871, 2021. https://doi.org/10.3390/jof7100871

CHERNOMOR, O.; VON HAESELER, A.; MINH, B. Q. Terrace aware data structure for phylogenomic inference from supermatrices. Syst. Biol., v. 65, p. 997-1008, 2016. https://doi.org/10.1093/sysbio/syw037

CHOI, W. and DEAN, R. A. The Adenylate Cyclase Gene MAC1 of *Magnaporthe grisea* Controls Appressorium Formation and Other Aspects of Growth and Development. **The Plant Cell**, v. 9, p. 1973-1983, 1997.

CONNORTON, J. M.; BALK, J.; RODRÍGUEZ-CELMA, J. Iron homeostasis in plants – a brief overview. **Metallomics**, v. 9, n. 813, 2017.

COOK, D. E.; MESARICH, C. H.; THOMMA, B. P. Understanding plant immunity as a surveillance system to detect invasion. ***Annual Review of Phytopathology***, v. 53, p. 541-563, 2015.

CRAWFORD, A. AND WILSON, D. Essential metals at the host–pathogen interface: nutritional immunity and micronutrient assimilation by human fungal pathogens. **FEMS Yeast Research**, v. 15, 2015.

CUOMO, C. A.; BAKKEREN, G.; KHALIL, H. B.; PANWAR, V.; JOLY, D.; LINNING, R.; SAKTHIKUMAR, S.; SONG, X.; ADICONIS, X.; FAN, L.; GOLDBERG, J. M.; LEVIN, J. Z.; YOUNG, S.; ZENG, Q.; ANIKSTER, Y.; BRUCE, M.; WANG, M.; YIN, C.; MCCALLUM, B.; SZABO, L. J.; HULBERT, S.; CHEN, X.; FELLERS, J. P. Comparative Analysis Highlights Variable Genome Content of Wheat Rusts and Divergence of the Mating Loci. **G3 (Bethesda)**, v. 9, n. 7, p. 361-376, 2017. doi: 10.1534/g3.116.032797

CUTLER, S.; EHRHARDT, D.; GRIFFITTS, J.; SOMERVILLE, C. Random GFP::cDNA fusions enable visualization of subcellular structures in cells of Arabidopsis at a high frequency. **Proceedings of the National Academy of Sciences**, v. 97, p. 3718-3723, 2000.

DABNEY-SMITH, C.; VAN DEN WIJNGAARD, P.W.; TREECE, Y.; VREDENBERG, W.J; BD, B. The C terminus of a chloroplast precursor modulates its interaction with the translocation apparatus and PIRAC. **Journal of Biological Chemistry**, n. 274, p. 32351-32359, 1999.

DANGL, J. L., HORVATH, D. M., STASKAWICZ, B. J. Pivoting the plant immune system from dissection to deployment. **Science**, v. 341, p. 746-751, 2013.

DEANG, J. and DEAN, R. A. Characterization of Adenylate Cyclase Interacting Protein ACI1 in the Rice Blast Fungus, *Magnaporthe oryzae*. **The Open Mycology Journal**, v. 2, p. 74-81, 2008.

de CARVALHO, M. C., COSTA NASCIMENTO, L., DARBEN, L. M., POLIZEL-PODANOSQUI, A. M., LOPES-CAITAR, V. S., QI, M., et al. Prediction of the in planta Phakopsora pachyrhizi secretome and potential effector families. Molecular plant pathology, v. 18, p. 363-377, 2007.

DEZWAAN, T. M.; CARROLL, A. M.; VALENT, B.; SWIGARD, J. A. *Magnaporthe grisea* Pth11p Is a Novel Plasma Membrane Protein That Mediates Appressorium Differentiation in Response to Inductive Substrate Cues. **The Plant Cell**, v. 11, p. 2013–2030, 199.

DING, C.; VIDANES, G. M.; MAGUIRE, S. L.; GUIDA, A.; SYNNOTT, J. M.; ANDES, R.; BUTLER, G. Conserved and Divergent Roles of Bcr1 and CFEM Proteins in *Candida parapsilosis* and *Candida albicans.* **PLoS ONE**, 6(12): e28151, 2011. doi:10.1371/journal.pone.0028151

DVIR, H., KORNITZER, D. CFEM protein Csa2. **Encyclopedia of Inorganic and Bioinorganic Chemistry**, Online © 2011–2018 John Wiley & Sons, Ltd. DOI: 10.1002/9781119951438.eibc2618

EKWOROMADU, M. T. et al. Differential function of lip residues in the mechanism and biology of an anthrax hemophore. **PLOS Pathogens**, v. 8, e1002559, 2012.

EMMS, D. M.; KELLY. S. SHOOT: phylogenetic gene search and ortholog inference. **Genome Biology**, v. 23, n. 85, 2022. https://doi.org/10.1186/s13059-022-02652-8

FABRO, G., STEINBRENNER, J., COATES, M., ISHAQUE, N., BAXTER, L., STUDHOLME, D. J., et al. Multiple candidate effectors from the oomycete pathogen Hyaloperonospora arabidopsidis suppress host plant immunity. **PLoS pathogens**, 7, e1002348, 2011.

GÍSLASON, M. H.; NIELSEN, H.; ARMENTEROS, J. J. A.; JOHANSEN, A. R. Prediction of GPI-anchored proteins with pointer neural networks. **Current Research in Biotechnology**, v. 3, p. 6-13, 2021.

GODOY, C. V.; SEIXAS, C. D. S.; SOARES, R. M.; MARCELINO-GUIMARÃES, F. C.; MEYER, M. C.; COSTAMILAN, L. M. Asian soybean rust in Brazil: past, present, and future. **Pesquisa Agropecuária Brasileira**, v. 51, p. 407–421, 2016. https://doi.org/10.1590/S0100-204X2016000500002

GOELLNER, K.; LOEHRER, M.; LANGENBACH, C.; CONRATH, U.; KOCH, E.; SCHAFFRATH, U. Phakopsora pachyrhizi, the causal agent of Asian soybean rust. **Molecular plant pathology**, v. 11, p. 169-177, 2010.

GONG, A.; KING, Z.; ZHANG, K.; TAN, Q.; WANG, G.; LIU, W. Bioinformatic analysis and functional characterization of the cfem proteins in maize anthracnose fungus *Colletotrichum graminicola*. **Journal of Integrative Agriculture**. v.18, n.0, p. 2-11, 2019.

GUPTA, Y. K.; MARCELINO-GUIMARÃES, F. C.; LORRAIN, C.; FARMER, A.; et al. The soybean rust pathogen Phakopsora pachyrhizi displays transposable element proliferation that correlates with broad host-range adaptation on legumes. **BioRxiv preprint**, doi: https://doi.org/10.1101/2022.06.13.495685;

GUYON, K.; BALAGUÉ, C.; ROBY, D.; RAFFAELE, S. Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen *Sclerotinia sclerotiorum*. BCM Genomics, 15, 336, 2014. http://www.biomedcentral.com/1471-2164/15/336

HENNINGS, P. **Einige neue japanische Uredinales**. Hedwigia, v. IV (Suppl.), p. 107-108, 1993.

HOANG, D. T.; CHERNOMOR, O.; VON HAESELER, A.; MINH, B. Q.; VINH, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol., v. 35, p. 518-522, 2018. https://doi.org/10.1093/molbev/msx281

JONES, J. D. and DANGL, J. L. The plant immune system. **Nature**, v. 444, p. 323-329, 2006.

JUMPER, J. et al. Highly accurate protein structure prediction with AlphaFold. **Nature**, v. 596, p. 583–589, 2021.

KÄLL, L.; KROGH, A.; SONNHAMMER, E. L. L. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. **Nucleic Acids Research**, v. 35, n. W429-32, 2007.

KATOH, K.; ROZEWICKI, J.; YAMADA, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. **Brief Bioinform**, v. 20, n. 4, p.1160-1166, 2019. doi: 10.1093/bib/bbx108. PMID: 28968734; PMCID: PMC6781576.

KOBAYASHI, T.; NISHIZAWA, N. K. Iron Uptake, Translocation, and Regulation in Higher Plants. **Annu. Rev. Plant. Biol**. v. 63, n. 131-52, 2012.

KOCH, E. F.; HOPPE, H. H. Light and electron microscopic studies on the development of soybean rust (Phakopsora pachyrhizi Syd.) in susceptible soybean leaves. **Phytopathology**, v. 106, p. 302-320, 1983.

KOU, Y.; TAN, Y. H.; RAMANUJAN, R.; NAQVI, N. I. Structure-function analyses of the Pth11 receptor reveal an important role for CFEM motif and redox regulation in rice blast. **New Phytologist**, v. 214, p. 330–342, 2017.

KULKARNI, R. D.; KELKAR, H. S.; DEAN, R. A. An eight-cysteine-containing CFEM domain unique to a group of fungal membrane proteins. **Trends in Biochemical Sciences**, v. 28, p. 118–121, 2003.

KUNJETI, S. G.; IYER, G.; JOHNSON, E.; LI, E.; BROGLIE, K. E.; RAUSCHER, G. et al. Identification of Phakopsora pachyrhizi Candidate Effectors with Virulence Activity in a Distantly Related Pathosystem. **Frontiers in plant science**, v. 7, n. 269, 2016.

KUZNETS, G.; VIGONSKY, E,; WEISSMAN, Z.; LALLI, D.; GILDOR, T. et al. A Relay Network of Extracellular Heme-Binding Proteins Drives *C. albicans* Iron Acquisition

from Hemoglobin. **PLOS Pathogens**, v. 10, n. 10, e1004407, 2014. doi:10.1371/journal.ppat.1004407

KORNITZER, D. Fungal mechanisms for host iron acquisition. **Current Opinion in Microbiology**, v. 12, p. 377–383, 2009.

KORNITZER, D. and ROY, U. Pathways of heme utilization in fungi. **BBA - Molecular Cell Research**, 2020. https://doi.org/10.1016/j.bbamcr.2020.118817

KUMAR, S., STECHER, G., TAMURA, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. **Molecular biology and evolution**, v.33, p.1870-1874, 2016.

LABBÉ, S.; MOURER, T.; BRAULT, A.; VAHSEN, T. Machinery for fungal heme acquisition. **Current Genetics**, 2017. https://doi.org/10.1007/s00294-020-01067-x

LETOFFE, S.; GHIGO, J. M.; WANDERSMAN, C. Iron acquisition from heme and hemoglobin by a Serratia marcescens extracellular protein. **National Academy of Sciences**, v. 91, p. 9876–9880, 1994.

LETUNIC, I.; BORK, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. **Nucleic Acids Res**., v. 49, 2021. https://doi.org/10.1093/nar/gkab301

LI, F.; UPADHYAYA, N. M.; SPERSCHNEIDER, J.; MATNY, O.; NGUYEN-PHUC, H.; MAGO, R.; RALEY, C.; MILLER, M. E.; SILVERSTEIN, K. A. T.; HENNINGSEN, E.; HIRSCH, C. D.; VISSER, B.; PRETORIUS, Z. A.; STEFFENSON, B. J.; SCHWESSINGER, B.; DODDS, P. N.; FIGUEROA, M. Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. **Nat Commun**., v. 7, n. 10, p. 5068, 2017. doi: 10.1038/s41467-019-12927-7

LINK, T. I.; LANG, P.; SCHEFFLER, B. E.; DUKE, M. V.; GRAHAM, M. A.; COOPER, B.; TUCKER, M. L.; VAN DE MORTEL, M.; VOEGELE, R. T.; MENDGEN, K.; BAUM, T. J.; WHITHAM, S. A. The haustorial transcriptomes of Uromyces appendiculatus and Phakopsora pachyrhizi and their candidate effector families. **Molecular Plant Pathology**, v. 15, n. 4, p. 379-393, 2014.

LIU, L.; XU, L.; JIA, Q.; PAN, R.; OELMÜLLER, R.; ZHANG, W.; WU, C. Arms race: diverse effector proteins with conserved motifs. **PLANT SIGNALING & BEHAVIOR**, v. 14, n. 2, 2019.

LIVAK, K. J. and SCHMITTGEN, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. **Methods**, v. 25, p. 402-408, 2001.

LORRAIN, C.; HECKER, A.; DUPLESSIS, S. Effector-Mining in the Poplar Rust Fungus Melampsora larici-populina Secretome. **Frontiers in plant science**, v. 6, n. 1051, 2015.

MAPURANGA, J.; ZHANG, N.; ZHANG, L.; CHANG, J.; YANG, W. Infection Strategies and Pathogenicity of Biotrophic Plant Fungal Pathogens. **Front. Microbiol**. 2022. 13:799396. doi: 10.3389/fmicb.2022.799396

MONCLARO, A. V.; FILHO, E. X. F. Fungal lytic polysaccharide monooxygenases from family AA9: Recent developments and application in lignocelullose breakdown. **Int J Biol Macromol**, v. 102, p. 771-778, 2017. doi:10.1016/j.ijbiomac.2017.04.077

NASSER, L.; WEISSMAN, Z.; PINSKY, M.; AMARTELY, H.; DVIR, Y.; KORNITZEIR, D. Structural basis of haem-iron acquisition by fungal pathogens. **Nature Microbiology**. v. 1. 2016.

NGUYEN, L. -T.; SCHMIDT, H. A.; VON HAESELER, A.; MINH, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. **Mol. Biol. Evol.**, v. 32, p. 268-274, 2015. https://doi.org/10.1093/molbev/msu300

OMASITS, U., AHRENS, C. H., MÜLLER, S., WOLLSCHEID, B. Protter: interactive protein feature visualization and integration with experimental proteomic data. **Bioinformatics**. 2014 Mar 15;30(6):884-6. doi:10.1093/bioinformatics/btt607.

PÉREZ, A.; PEDRÓS, B.; MURGUI, A.; CASANOVA, M.; LÓPEZ-RIBOT, J. L.; MERTÍNEZ, J. P. Biofilm formation by *Candida albicans* mutants for genes coding fungal proteins exhibiting the eight-cysteine-containing CFEM domain. **Federation of European Microbiological Societies**, v. 6, p. 1047-1084, 2006.

PÉREZ, A.; GORDON, R.; BLANES, R.; MURGUI, A.; CASANOVA, M.; MERTÍNEZ, J. P. Some biological features of *Candida albicans* mutants for genes coding fungal

proteins containing the CFEM domain. **Federation of European Microbiological Societies**, v. 11, p. 273-284, 2011.

PERSOONS, A.; MAUPETIT, A.; LOUET, C.; ANDRIEUX, A.; LIPZEN, A.; BARRY, K. W.; NA, H.; ADAM, C.; GRIGORIEV, I. V.; SEGURA, V.; DUPLESSIS, S.; FREY, P.; HALKETT, F.; DE MITA, S. Genomic Signatures of a Major Adaptive Event in the Pathogenic Fungus *Melampsora larici-populina*. **Genome Biol Evol.**, v.4, n. 14, 2022.

PETTERSEN, E. F.; GODDARD, T. D.; HUANG, C. C.; COUCH, G. S.; GREENBLATT, D. M.; MENG, E. C.; FERRIN, T. E. UCSF Chimera--a visualization system for exploratory research and analysis. **J Comput Chem**, v. 25, n. 13, p. 1605-1612, 2004.

PIERLEONI, A.; MARTELLI, P, L.; CASADIO, R. PredGPI: a GPI anchor predictor. **BMC Bioinformatics**, v. 9, n. 392, 2008.

QI, M.; LINK, TI.; MÜLLER, M.; HIRSCHBURGER, D.; PUDAKE, RN.; PEDLEY, KF. et al. A Small Cysteine-Rich Protein from the Asian Soybean Rust Fungus, Phakopsora pachyrhizi, Suppresses Plant Immunity. **PLoS Pathog**, v. 12, n. 9, 2016.

SABNAM, N. and BARMAN, S. R. WISH, a novel CFEM GPCR is indispensable for surface sensing, asexual and pathogenic differentiation in rice blast fungus. **Fungal Genetics and Biology**, v. 105, p. 37-51, 2017.

SAITO, C., UEDA, T., ABE, H., WADA, Y., KUROIWA, T., HISADA, A., et al. A complex and mobile structure forms a distinct subregion within the continuous vacuolar membrane in young cotyledons of Arabidopsis. **The Plant journal: for cell and molecular biology**, v. 29, p. 245-255, 2002.

SAUCEDO-CAMPA, D. O.; MARTÍNEZ-ROCHA, A. L.; RÍOS-CASTRO, E.; ALBA-FIERRO, C. A.; ESCOBEDO-BRETADO, M. A.; CUÉLLAR-CRUZ, M.; RUIZ-BACA, E. Proteomic Analysis of *Sporothrix schenckii* Exposed to Oxidative Stress Induced by Hydrogen Peroxide. **Pathogens**, 11, 230, 2022. https://doi.org/10.3390/pathogens11020230

SCHWESSINGER, B.; ZIPFEL, C. News from the frontline: recent insights into PAMP-triggered immunity in plants. **Current opinion in plant biology**, v. 11, p. 389-395, 2008.

SRIVASTAVA, V. K.; SUNEETHA, K. J.; KAUR, R. A systematic analysis reveals an essential role for high-affinity iron uptake system, haemolysin and CFEM domain-containing protein in iron homeostasis and virulence in *Candida glabrata*. **Biochemical Journal**, v. 463, p. 103-114, 2014.

SOHN, K. H., LEI, R., NEMRI, A., JONES, J. D. The downy mildew effector proteins ATR1 and ATR13 promote disease susceptibility in Arabidopsis thaliana. **The Plant cell**, v. 19, p. 4077-4090, 2007.

SPERSCHNEIDER, J.; DODDS, P. N.; GARDINER, D. M.; SINGH, K. B; TAYLOR, J. M. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. **Molecular plant pathology.** v. 19, p. 2094-2110, 2018.

SPERSCHNEIDER, J. et al. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. 2017

SPERSCHNEIDER, J.; GARDINER, D. M.; DODDS, P. N.; TINI, F.; COVARELLI, L.; SINGH, K. B.; et al. EffectorP: predicting fungal effector proteins from secretomes using machine learning. **The New phytologist**. v. 210, p. 743-761, 2016.

TODD, J.N.A.; CARREÓN-ANGUIANO, K.G.; ISLAS-FLORES, I.; CANTO-CANCHÉ, B. Fungal Effectoromics: AWorld in Constant Evolution. **Int. J. Mol. Sci**. 2022, 23,13433. https://doi.org/10.3390/ijms232113433

VAKNIN, Y.; SHACKCHAN, Y.; LEVDANSKY, E.; MOROZOV, M.; ROMANO, J.; OSHEROV, N. The three *Aspergillus fumigatus* CFEM-domain GPI-anchored proteins (CfmA-C) affect cell wall stability but do not play a role in fungal virulence. **Fungal Genetics and Biology**, v. 63, p. 55-64, 2014.

VOEGELE, R. T., and MENDGEN, K. Rust haustoria: nutrient uptake and beyond. **New Phytologist**, v. 159, p. 93–100, 2003.

WANG, J.; LONG, F.; ZHU, H.; ZHANG, Y.; WU, J.; SHEN, S.; DONG, J.; HAO, Z. Bioinformatic analysis and functional characterization of CFEM proteins in *Setosphaeria turcica*. **Journal of Integrative Agriculture**. v. 20. n.9. p. 2438–2449. 2021.

WANG, D., et al. MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. **Nucleic Acids Research.** v. 48, n. W1, p W140–W146, 2020

WANG, D., et al. Capsule network for protein post-translational modification site prediction, **Bioinformatics**. v. 35, n. 14, p. 2386-2394, 2019.

WANG, D., et al. MusiteDeep: a deep-learning framework for general and kinase-specific phosphorylation site prediction. **Bioinformatics**. v. 33, n. 24, p. 3909-3916, 2017.

WANG, J.; LONG, F.; ZHU, H.; ZHANG, Y.; WU, J.; SHEN, S.; DONG, J.; HAO Z. Bioinformatic analysis and functional characterization of CFEM proteins in *Setosphaeria turcica*. **Journal of Integrative Agriculture**, v. 20, n. 9, p. 2438-2449, 2021.

WANG, D.; ZHANG, D.; SONG, J.; LI, J.; WANG, J.; LI, R.; KLOSTERMAN, S. J.; KONG, Z.; LIN, F.; DAI, X.; SUBBARAO, K. V.; CHEN, J. *Verticillium dahliae* CFEM proteins manipulate host immunity and differentially contribute to virulence. **BMC Biology**, 2022. https://doi.org/10.1186/s12915-022-01254-x

WEILL, U.; YOFE, I.; SASS, E.; STYNEN, B.; DAVIDI, D.; et al. Genome-wide SWAp-Tag yeast libraries for proteome exploration. **Nature Methods**, v. 15, p. 617–622, 2018.

Weill, U.; Krieger, G.; Avihou, Z.; Milo, R.; Schuldiner, M.; Davidi, D. Assessment of GFP Tag Position on Protein Localization and Growth Fitness in Yeast. **Journal of Molecular Biology**, v. 431, p. 636-641, 2019.

WEINBERG. Nutritional Immunity. **The Journal of the American Medical Association**, v. 231, n. 1, 1975.

WEISSMAN, Z., KORNITZER, D. A family of Candida cell surface haem-binding proteins involved in haemin and hemoglobin-iron utilization. **Molecular Microbiology.** v. 53, n. 4, p. 1209–1220, 2004. doi:10.1111/j.1365-2958.2004.04199.x

WEISSMAN, Z.; SHEMER, R.; CONIBEAR, E.; KORNITZER, D. An endocytic mechanism for hemoglobin-iron acquisition in Candida albicans. **Molecular Microbiology**, v. 69, n. 1, p. 201-217, 2008.

WEISSMAN, Z.; PINSKY, M.; DONEGAN, R. K.; REDDI, A. R.; KORNITZER, D. Using genetically encoded heme sensors to probe the mechanisms of heme uptake and homeostasis in Candida albicans. **Cellular Microbiology**. 2021.https://doi.org/10.1111/cmi.13282

ZANG, Z. WU, Q. ZHANG, G. ZHU, Y. MURPHY, R. W. LIU, Z. L. ZOU, C. Systematic analyses reveal uniqueness and origin of the CFEM domain in fungi. **Scientific Reports**. 2015. DOI:10.1038/srep13032

ZHANG, S.; LI, C.; SI, J.; HAN, Z.; CHEN, D. Action Mechanisms of Effectors in Plant-Pathogen Interaction. **Int. J. Mol. Sci**. 2022, 23, 6758. https://doi.org/10.3390/ijms23126758

ZHAO, S.; SHANG, X.; BI, W.; YU, X.; LIU, D.; KANG, Z.; WANG, X.; WANG, X. Genome-Wide Identification of Effector Candidates with Conserved Motifs from the Wheat Leaf Rust Fungus *Puccinia triticina*. **Frontiers in Microbiology**. 11:1188. 2020. doi: 10.3389/fmicb.2020.01188

ZHU, W.; WEI, W.; WU, Y.; ZHOU, Y.; PENG, F.; ZHANG, S.; CHEN, P.; XU, X.; BcCFEM1, a CFEM Domain-Containing Protein with Putative GPI-Anchored Site, Is Involved in Pathogenicity, Conidial Production, and Stress Tolerance in Botrytis cinerea. **Frontiers in Microbiology.** 2017. doi: 10.3389/fmicb.2017