



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE BIOCIÊNCIAS
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

JOSÉ DIOGO CAVALCANTI FERREIRA

**ABORDAGENS BIOTECNOLÓGICAS PARA OBTENÇÃO DE LINHAGENS DE
FEIJÃO-CAUPI RESISTENTES A FITOPATÓGENOS**

Recife
2020

JOSÉ DIOGO CAVALCANTI FERREIRA

**ABORDAGENS BIOTECNOLÓGICAS PARA OBTENÇÃO DE LINHAGENS DE
FEIJÃO-CAUPI RESISTENTES A FITOPATÓGENOS**

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

Área de concentração: Biotecnologia

Orientadora: Prof^a. Dr^a. Ana Maria Benko-Iseppon

Coorientadores: Dr. Antônio Félix da Costa.
Dr^a. Roberta Lane de Oliveira Silva

Recife

2020

Catálogo na fonte:
Bibliotecária Claudina Queiroz, CRB4/1752

Ferreira, José Diogo Cavalcanti

Abordagens biotecnológicas para obtenção de linhagens de feijão-caupi resistentes a fitopatógenos / José Diogo Cavalcanti Ferreira - 2020.

131 folhas: il., fig., tab.

Orientadora: Ana Maria Benko-Iseppon

Coorientadores: Antônio Félix da Costa

Roberta Lane de Oliveira Silva

Dissertação (mestrado) – Universidade Federal de Pernambuco.
Centro de Biociências. Programa de Pós-Graduação em Genética.
Recife, 2020.

Inclui referências e apêndice.

1. Feijão-caupi 2. Transgenia 3. Imunidade inata

I. Benko-Iseppon, Ana Maria (Orientadora) II. Costa, Antônio Félix da
(Coorientador) III. Silva, Roberta Lane de Oliveira (Coorientadora)
IV. Título

582.74 CDD (22.ed.)

UFPE/CB-2020-196

JOSÉ DIOGO CAVALCANTI FERREIRA

**ABORDAGENS BIOTECNOLÓGICAS PARA OBTENÇÃO DE LINHAGENS DE
FEIJÃO-CAUPI RESISTENTES A FITOPATÓGENOS**

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

Aprovado em: 27/02/2020.

BANCA EXAMINADORA

Dra. Ana Maria Benko Iseppon
Universidade Federal de Pernambuco

Dr. Tercílio Calsa Júnior
Universidade Federal de Pernambuco

Dr. João Pacífico Bezerra Neto
Universidade Federal de Pernambuco

Dr. Francisco José Lima Aragão
Empresa Brasileira de Pesquisa Agropecuária

Dr. José Ribamar Costa Ferreira Neto
Universidade Federal de Pernambuco

AGRADECIMENTOS

Agradeço primeiramente a minha mãe, Maria Marceli Carneiro Cavalcanti, e ao meu pai, José Ferreira Gomes Neto, pelo carinho, dedicação, incentivo e todo o imenso esforço que realizaram para minha formação educacional básica, que foi o pilar para a formação conquistada até aqui. A meus irmãos Diego e Débora Cavalcanti, pelo apoio incondicional a tudo que necessitei e necessito.

Um agradecimento especial à minha esposa, Jéssica Xavier Batista, pela confiança, estímulo e por aturar meus estresses, principalmente nas proximidades da qualificação, entrega e defesa da tese. O seu suporte foi fundamental para poder chegar a esse momento especial. Além de sempre que possível ter ajudado a semear, regar e colher meus feijões.

A toda minha família, que por ser muito grande vou citar os meus avós paternos (Vovô Ferreira e Vovó Lilia) como pilares da nossa estrutura familiar fundamental em toda minha formação. A partir deles agradeço especialmente a toda sua prole (tios, tias e primos), fazendo a mesma consideração a meus avós maternos (Vovô Pedro e Vovó Mariazinha) assim como toda sua prole. Um agradecimento especial a Amanda Ferreira, que além de todo carinho me ajudou bastante com as imagens.

Aos amigos do LGBV, que compartilham a rotina cansativa de um laboratório de biologia molecular, sempre incentivando e não permitindo nos momentos de resultados negativos (que não são poucos) deixar a peteca cair. Vou citar alguns nomes especiais: Flávia, Carol, Artemisa, Mitalle, Neto, Carlos, Lívia, Marx, Manu, Jéssica Bárbara, Jéssica Barboza, Roberta, Vanessa, Jaysa, Rodrigo, Bruna, Sheyla e João. A meu amigo Geraldo Moura, um grande incentivador da minha formação acadêmica, assim como os amigos Reginaldo Fonseca e Jonas Marôpo, pelo apoio e incentivo de sempre. A Hayana Azevedo por ter me ingressado na biologia molecular, além da amizade inestimável que temos.

Ao Francisco Aragão, que se não fosse a abertura das portas do seu laboratório seria impossível o desenvolvimento do nosso trabalho, além de todo carinho, atenção e ensinamentos que me promove mesmo após retorno de Brasília-DF. Em especial para os amigos do LEG Embrapa Cenargen que me acolheram

como uma família, fundamental durante o período de estadia em Brasília. Queria destacar Pedro Berbert, pela grande amizade e auxílio na pesquisa; Heitor, pelo incentivo e discussão vigorosa das nossas hipóteses; Jéssica Carnijo e Elsa, pelo apoio técnico em várias ocasiões; Gláucia que além da parte técnica, demonstrou um carinho fora do normal e ao amigo Cristobal, pelas conversas descontraídas. A Giovanni Viana, Natália, Tomas, Taís, Emanuel, Kenny e Lúdia que fizeram do LEG uma rotina mais feliz e um ambiente mais harmonioso.

Um agradecimento especial ao Dr. Félix, que não mede esforços e sempre está à disposição para o desenvolvimento da nossa pesquisa, assim como as professoras Valesca Pandolfi e Ana Christina pelo apoio, carinho, incentivo e auxílio desse trabalho. As agências de fomento que custearam essa pesquisa, com destaque para a Facepe e a CAPES (que concederam a bolsa durante o sanduíche), assim como o CNPq e a CAPES (Edital Biologia Computacional) que financiaram a pesquisa. Ao IFPE Campus Pesqueira, que me concedeu dois anos e meio de licença remunerada, o que foi fundamental para o avanço da minha formação acadêmica, assim como ao Programa de Pós-Graduação em Genética da UFPE pelo apoio, estrutura e conhecimento gerado ao longo do doutorado.

E por último, agradeço ao carinho e consideração a pessoa especial que é minha orientadora Ana Benko. Ela é a idealizadora da consolidação dessa linha de pesquisa em nosso laboratório, fazendo o possível e impossível, nesses tempos sombrios de financiamento da ciência, para o desenvolvimento da nossa pesquisa. Meu muito obrigado!

RESUMO

O feijão-caupi é uma leguminosa caracterizada por produzir grãos altamente nutritivos, com elevado teor de proteínas, vitaminas e sais minerais. Entretanto, existem perdas significativas na sua produção devido a estresses ambientais e fatores bióticos. Assim, o presente estudo objetivou a utilização da biotecnologia na aquisição de resistência do feijão-caupi frente a fitopatógenos. Plantas transgênicas de feijão-caupi foram obtidas via biobalística, expressando uma defensina de *Momordica charantia* (Curcubitaceae) com eficiência de transformação de 0,7%, confirmada por PCR e *Southern Blot*. Quatro linhagens tiveram seus níveis de expressão das defensinas detectados por RT-qPCR, das quais três tiveram níveis bem superiores comparativamente a uma das linhagens, o que mostra possíveis efeitos epigenéticos devido à inserção aleatória do transgene no genoma hospedeiro. Através de bioensaio foi constatada atividade antifúngica contra *Macrophomina phaseolina*. Também foi realizada uma edição do fator de iniciação da tradução eIF(iso)4E de feijão-caupi através da tecnologia de CRISPR/Cas9. A transformação foi via *Agrobacterium tumefaciens*, sendo uma inovação do protocolo. A aplicação da técnica objetivou o nocaute deste gene, visto que ele interage com vírus da família Potyviridae, fazendo com que as plantas se tornem suscetíveis. Após a aplicação da técnica observou-se que a edição foi do tipo substituição de base, a qual substitui um aminoácido por outro da mesma natureza química. Os resultados mostram a possibilidade de desenvolvimento de uma linhagem superior de feijão-caupi resistente à *M. phaseolina*.

Palavras-chave: Transgenia. *Pathogenesis Related*. Imunidade inata. CABMV. Melão-de-são-caetano.

ABSTRACT

Cowpea is a legume characterized by producing highly nutritious grains, with a high content of proteins, vitamins, and minerals. However, there are significant losses in its production due to abiotic and biotic stresses. Thus, the present study aimed to use biotechnology in the acquisition of cowpea resistance against phytopathogens. Transgenic cowpea plants were obtained via biobalistics, expressing a *Momordica charantia* (Cucurbitaceae) defensin with an efficiency of 0.7%, confirmed by PCR and Southern Blot. Four strains had the expression level of defensins detected by RT-qPCR, of which three had higher levels compared to one of the strains, which shows possible epigenetic effects due to the random insertion of the transgene in the host genome. Through a bioassay, antifungal activity against *Macrophomina phaseolina* was found. Besides, an edition of the translation initiation factor eIF(iso)4E of cowpea was also carried out using the CRISPR/Cas9 technology. The transformation was via *Agrobacterium tumefaciens*, which was an innovation of the protocol. The application of the technique aimed at the knock-out of this gene, since it interacts with viruses of the Potyviridae family, turning the plants susceptible. After the application of the technique, it was observed that the edition resulted in a base substitution, which replaced one amino acid by another of the same chemical nature. The results show the possibility of developing a superior strain of cowpea resistant to *M. phaseolina*.

Key words: Genetic transformation. Pathogenesis Related. Innate immunity. CABMV. Bitter melon.

LISTA DE FIGURAS

Revisão de Literatura

- Figura 1 –** Processo de transferência horizontal do DNA entre *Agrobacterium tumefaciens* e uma célula vegetal. Através de sinalizadores como compostos fenólicos ocorre a estimulação dos genes de virulência do plasmídeo Ti, o qual promove a expressão do T-DNA, que é transferido da bactéria até o núcleo da célula vegetal, onde ocorrerá a integração do T-DNA no genoma da planta hospedeira. Fonte: Hwangb *et al.* (2017)..... 26
- Figura 2 –** Métodos de reparo do DNA. Após a clivagem do DNA, pelo método NHEJ (*non-homologous end joining*) a dupla fita clivada tem suas extremidades unidas de forma passível de erro (*error-prone*) como, por exemplo, a origem de *indels*. Já pelo método HDR (*homology directed repair*) é fornecido um *donor* DNA, que possui homologia com as extremidades da dupla-fita clivada, mais uma região intermediária para a inserção programada da sequência em questão. Fonte: Sander e Joung (2013)..... 32
- Figura 3 –** Estrutura do ZFNs, com a formação da estrutura dimérica. A região ZFP, corresponde ao domínio de ligação ao DNA, onde cada *zinc-finger* reconhece três pb, e encontra-se ligada a proteína de clivagem do DNA (nuclease FokI). Na imagem é apresentado o consenso do *zinc-finger*, cuja letra X representa qualquer aminoácido. A estrutura dimérica leva à formação de uma região espaçadora (5-7 pb), porém aumenta a especificidade da região alvo. Fonte: Kim e Kim (2014)..... 33
- Figura 4 –** Estrutura do TALENs, cuja região com domínios de ligação ao DNA (*TALE repeat domains*) é flanqueada pelos domínios N e C-terminais. No *TALE repeat domains* cada combinação (seja HD, NN, NG e NI) reconhece uma base específica e corresponde aos aminoácidos da posição 12-13 de uma sequência de 33-35 aminoácidos. Esta estrutura de reconhecimento e ligação ao DNA encontra-se ligada à proteína de clivagem do DNA (nuclease FokI).

	Fonte: Joung e Sander (2013).....	34
Figura 5 –	Esquema do funcionamento do CRISPR natural das bactérias. Na etapa 1 chamada de aquisição, é incorporado ao sistema um DNA exógeno (DNA de bacteriófago ou plasmídeo) localizado entre as sequências palindrômicas curtas repetidas. As etapas 2 e 3 correspondem ao processamento do RNA. Na 2 ocorre a transcrição do pre-CRISPR RNA e do tracrRNA. Na etapa 3 ocorre a interação do pre-crRNA com o tracrRNA, e destes com a RNase III e a Cas9, onde a RNase III promove o corte na região palindrômica curta repetida, originando o CRISPR RNA ou crRNA. Por fim, na etapa 4 (interferência) ocorre a interação do crRNA com a região homóloga do DNA viral, ocorrendo o corte da dupla fita do DNA viral pela ação da Cas9. Adaptado de Singh <i>et al.</i> (2017) e Belhaj <i>et al.</i> (2015), com modificações pelo autor.....	36
Figura 6 –	Esquema do processo de início da tradução em eucariotos. Inicialmente ocorre a interação do eIF4E com o CAP da extremidade 5' do RNA mensageiro (RNAm), o qual recruta o eIF4A e eIF4G. Posteriormente ocorre a interação do eIF4G com o PABP, proteína que se liga à cauda poli-A da extremidade 3' do RNAm. Por fim, ocorre a interação da subunidade menor do ribossomo, com o GTP, Met-tRNA _i , eIF2 e eIF3, sendo todos estes componentes recrutados pelo eIF4G, que interage com o eIF3. Fonte: Montero <i>et al.</i> , 2015.....	39
	Expressão de uma defensina de Melão-de-São-Caetano em feijão-caupi transgênico promovendo resistência a fitopatógeno fúngico	
Figura 1 –	Vetor pAHASMCDEF1 utilizado na transformação vegetal, com destaque para os sítios de restrição NcoI, SacI e NotI, utilizados na construção do vetor, e SphI utilizado na digestão do DNA genômico para o Southern Blot.....	46
Figura 2 –	Gel de agarose 1,2% com os produtos da PCR com primers específicos para o gene <i>Mcdef1</i> que amplifica um fragmente de 405 pb. L – Ladder 1 Kb; B – Branco (sem DNA); C: Controle negativo (planta não transformada); P: Controle positivo correspondente ao	

	plasmídeo pAHASMCDEF1.....	53
Figura 3 –	Padrão da hibridização via Southern Blotting das sete linhagens que transmitiram o transgene para a progênie. L – Ladder 1 Kb; C: Controle negativo (tipo selvagem, planta não transformada); Linhagens transformadas 67, 115, 122, ,133 e 175; P: Plasmídeo pAHASMCDEF1 (Controle positivo).....	56
Figura 4 –	Análise da abundância relativa do transcrito do gene <i>McDef1</i> via RT-qPCR em quatro linhagens de feijão-caupi transgênico (67, 73, 115 e 175) na geração T3.....	56
Figura 5 –	Imagem representativa do experimento de inoculação do Potyvirus CABMV. Na imagem A, as plantas com uma semana após inoculação do vírus, na imagem B, duas semanas após inoculação. As letras C no interior dos vasos, mostram as plantas-controle (não-transformadas, tipo selvagem) inoculadas.....	57
Figura 6 –	Imagem representativa do experimento de inoculação do Potyvirus CABMV e CPSMV. Na imagem A, as plantas com 17 dias de idade, após sete dias da inoculação, com os sintomas do CABMV, como mosaico intenso do limbo foliar e distorção das folhas. Na imagem B e C, tem um paralelo de uma mesma linhagem de plantas transformadas inoculadas com o CABMV (B) e com o CPSMV (C) na mesma data, evidenciando a velocidade e severidade do vírus do mosaico severo, que provoca a morte precoce comparativamente ao potyvirus.....	58
Figura 7 –	Imagem representativa do experimento de inoculação do fungo <i>R. solani</i> mostrando quatro linhagens transformadas (67, 73, 115 e 175), além da planta controle não inoculada (C). A imagem foi realizada cinco dias após a semeadura com inoculação nas linhagens transformadas e sem inoculação (C).....	58
Figura 8 –	Imagem representativa do experimento de inoculação do fungo <i>M. phaseolina</i> mostrando quatro linhagens transformadas (67, 73, 115 e 175), além da planta controle não inoculada (C). A imagem foi realizada cinco dias após a semeadura com inoculação nas linhagens transformadas e sem inoculação (C).....	59

Figura 9 – Imagem representativa do experimento de inoculação do fungo *M. phaseolnai* mostrando os sintomas das plantas que foram infectadas. Pela imagem A é perceptível que a maior parte das plantas transformadas com a defensina não foi infectada. Na imagem B, verifica-se com maior proximidade os sintomas (setas vermelhas), como escurecimento do caule, seguido de necrose, tombamento e morte das plantas..... 60

**Edição do genoma de feijão-caupi através do Sistema
CRISPR/Cas9 direcionada ao gene eIF(iso)4E visando aquisição
de imunidade/resistência ao CABMV**

Figura 1 – Árvore fenética das 42 sequências de eIF4E e eIF(iso)4E utilizadas no alinhamento, gerado pelo método de neighbor-joining, sendo o teste de confiabilidade dos nós realizado pelo método de bootstrap com 1000 replicações. As proteínas de eIF4E e eIF(iso)4E de *V. unguiculata* encontram-se destacadas em negrito..... 77

Figura 2 – Modelos tridimensionais dos fatores de iniciação eIF4E de *Vigna unguiculata*. (A) potencial eletrostático de ± 5 kT / e de FAS2 realizado no PyMOL plotado na superfície acessível por solvente para os modelos tridimensionais teóricos das cultivares de Boca Negra, BR14 Mulato e IT85F-2687, com cargas neutras, negativas e positivas em branco, vermelho e azul, respectivamente. (B) alinhamento estrutural dos modelos teóricos depois da simulação em dinâmica molecular dos fatores de iniciação eIF4E das cultivares Boca Negra (em verde), BR14 Mulato (em ciano) e IT85F-2687 (em magenta). (C) potencial eletrostático apenas realizado no PyMOL dos resíduos de aminoácidos modificados (Pro13_Arg13, Gly54_Arg54 e Ala55_Val55)..... 78

Figura 3 – Confirmação do gRNA desenhado e selecionado a partir do RGENOME através do programa CRISPR Optimal Target Finder. A partir deste programa foi realizado o desenho do oligo antisense, sendo manualmente adicionado o sítio de ligação (em vermelho na extremidade 5' do oligo senso e antisense) gerado pelo corte com a enzima de restrição AarI..... 81

- Figura 4 –** Vetor pGEM-T sgRNA digerido com PmeI (imagem A), cujo fragmento de 393 pb (poço 1) correspondente ao sgRNA completo foi isolado e purificado com kit da Promega, sendo quantificado com o marcador *Low DNA Mass Ladder* da Invitrogen (imagem B poço M), cuja intensidade está mais próxima ao fragmento de 400 pb correspondente a uma concentração de 40 ng de DNA..... 81
- Figura 5 –** Confirmação do vetor final pAHASCas9_sgRNA-eIF(iso)4E através de digestão enzimática. Inicialmente os plasmídeos isolados de seis colônias foram digeridos com NcoI e XbaI que geram quatro fragmentos (8947, 4568, 1683 e 1067 pb), sendo que só a colônia do poço 2 teve o padrão esperado (imagem A). Posteriormente este vetor do poço 2 da imagem A foi digerido com as enzimas PvuI que geraram fragmentos de 10417 pb e 5848 pb (poço 1 imagem B), PvuII que geram fragmentos de 9676, 3770, 1643 e 880 pb (poço 2 imagem B) e NotI que geram fragmentos de 13443, 1532 e 1290 pb (poço 3 imagem B)..... 82
- Figura 6 –** Gel de agarose 1,5% do produto da amplificação das possíveis plantas transformadas com o par de *primers* gRNA-speI / gRNA-KpnI com fragmento de 684 pb. L – Ladder 1 Kb; B – Branco (sem DNA); C – Controle negativo (planta não transformada); P – Controle positivo correspondente ao plasmídeo pAHASMCDEF1; 1 – planta transformada 1; 2 –planta transformada 2; 3 – planta transformada 3.L – Ladder 1 Kb; B – Branco (sem DNA); C – Controle negativo (planta não transformada); P – Controle positivo correspondente ao plasmídeo pAHASMCDEF1; 1 – planta transformada 1; 2 – planta transformada 2; 3 – planta transformada 3. 83
- Figura 7 –** Alinhamento do produto do sequenciamento do gene eIF(iso)4E da planta controle (cv. Boca Negra não transformada) e das três plantas transformadas PCR positivas para verificação da ocorrência da edição. Verifica-se uma mutação do tipo substituição de base (uma guanina na planta controle por uma citosina nas plantas transformadas / editadas – vide caixa destacada em vermelho). A mutação ocorreu duas bases antes do RNA guia, destacado pelo preenchimento laranja / amarelo..... 83

LISTA DE TABELAS

Expressão de uma defensina de Melão-de-São-Caetano em feijão-caupi transgênico promovendo resistência a fitopatógeno fúngico

- Tabela 1** – Pares de *primers* utilizados para amplificação de regiões do DNA exógeno inserido no genoma das plantas transformadas de feijão-caupi, assim como para PCR quantitativa em tempo real (RTqPCR)..... 49
- Tabela 2** – Tabela com dados da progênie das plantas transformantes no avanço da geração T₀ para a T₁, com os valores do teste do qui-quadrado (X²) com índice de confiabilidade de 95%. Para a linhagem 67 foi utilizado no teste a proporção 15:1 e nas demais 3:1. Para seguir o padrão Mendeliano o valor do teste tem que ser menor que 5%. T₁ + representa os indivíduos testados positivos para PCR com os primers que amplificam a defensina, sendo plantas transformadas, e T₁ – representam as plantas testadas negativas para PCR sendo não transformadas..... 54
- Tabela 3** – Tabela representativa da análise da progênie das plantas transformantes no avanço de uma geração (T₁ → T₂), com os valores do teste do X² e índice de confiabilidade de 95% com correção de Yates, observando-se o padrão de herança Mendeliano de 3:1 para todos os tratamentos. Cada linhagem (67, 73, 115, 122, 133 e 169) segue de um ponto e o indivíduo T₁ que avançou a geração para a T₂..... 55

Edição do genoma de feijão-caupi através do Sistema CRISPR/Cas9 direcionada ao gene *eIF(iso)4E* visando aquisição de imunidade/resistência ao CABMV

- Tabela 1** – Pares de *primers* utilizados para amplificação de regiões do gene *eIF4E*, *eIF(iso)4E1* e *eIF(iso)4E2* a partir do DNA genômico de feijão-caupi..... 70
- Tabela 2** – Três melhores gRNAs a partir do gene *eIF4E* do feijão-caupi baseado no *out-of-frame score*..... 80
- Tabela 3** – Três melhores gRNAs a partir do gene *eIF(iso)4E* do feijão-caupi baseado no *out-of-frame score*..... 80

LISTA DE ABREVIATURAS

At-UBP	<i>Arabidopsis thaliana</i> - Promotor da Ubiquitina 6
BAP	Benzilaminopurina
CENARGEN	Centro Nacional de Recursos Genéticos e Biotecnologia
CTAB	Cetyl-trimethyl-amoniumbromide
CRISPR	<i>Clustered Regularly Interspaced Short Palindromic Repeats</i>
Cas9	<i>CRISPR associated protein 9</i>
Cv.	Cultivar
DNA	<i>Desoxyribonucleic Acid</i> ; Ácido Desoxirribonucleico
DBS	<i>Double-Stranded Breaks</i>
Embrapa	Empresa Brasileira de Pesquisa Agropecuária
gRNA	<i>Guide RNA</i>
CABMV	<i>Cowpea Aphid-Borne Mosaic Vírus</i>
CPSMV	<i>Cowpea Severe Mosaic Virus</i>
IPA	Instituto Agrônômico de Pernambuco
LB	Luria Bertani
PCR	<i>Polymerase Chain Reaction</i>
RT-PCR	<i>Reverse Transcription Polymerase Chain Reaction</i>
RT-qPCR	<i>Reverse Transcription Polymerase Chain Reaction Quantitative real time</i>
T-DNA	<i>Transfer DNA</i>
Meio MS	<i>Murashige and Skoog Basal Medium</i>
NHEJ	<i>Non-Homologous End Joining</i>
IBA	Indole-3-Butyric Acid
PAM	<i>Protospacer Adjacent Motif</i>
p.s.i.	<i>Pound Force Per Square Inch</i> , Libra Força Por Polegada Quadrada
Rpm	Rotação por minuto
sgRNA	<i>single guide RNA</i>
TDZ	<i>Tidiazuron</i>
TALENs	<i>Transcription Activator-Like Effector Nucleases</i>
ZFN	<i>Zinc-Finger Nucleases</i>

SUMÁRIO

1	INTRODUÇÃO.....	17
1.1	OBJETIVOS.....	20
1.1.1	Objetivo Geral	20
1.1.2	Objetivos Específicos.....	20
2	REVISÃO DA LITERATURA.....	21
2.1	FEIJÃO-CAUPI (<i>VIGNA UNGUICULATA</i>).....	21
2.2	MELHORAMENTO TRADICIONAL DA DEFESA CONTRA FITOPATÓGENO.....	21
2.2.1	Impactos do CABMV e CPSMV em feijão-caupi e <i>status quo</i> do melhoramento visando a resistência.....	21
2.2.2	Impactos dos fungos fitopatogênicos em feijão-caupi e <i>status quo</i> do melhoramento visando resistência.....	24
2.3	TRANSFORMAÇÃO GENÉTICA DE PLANTAS.....	25
2.3.1	Transformação via <i>Agrobacterium tumefaciens</i>.....	25
2.3.2	Transformação via Biobalística.....	26
2.4	DEFENSINAS VEGETAIS.....	27
2.5	PLANTAS TRANSGÊNICAS EXPRESSANDO DEFENSINAS EXÓGENAS.....	28
2.6	MÉTODOS DE EDIÇÃO CONTROLADA DO GENOMA ATRAVÉS DE NUCLEASES PROGRAMÁVEIS.....	30
2.6.1	Ação das nucleases programáveis e método de reparo do DNA.....	31
2.6.2	ZFN.....	32
2.6.3	TALENS.....	33
2.6.4	CRISPR.....	35
2.7	GENES EIF4E E EIF(ISO)4E: FUNÇÕES NOS VEGETAIS E USO BIOTECNOLÓGICO.....	38
2.8	BIOTECNOLOGIA APLICADA PARA OBTENÇÃO DE CULTIVARES SUPERIORES EM FEIJÃO-CAUPI.....	40
3	EXPRESSÃO DE UMA DEFENSINA DE MELÃO-DE-SÃO- CAETANO EM FEIJÃO-CAUPI TRANSGÊNICO PROMOVENDO RESISTÊNCIA A FITOPATÓGENO FÚNGICO..	42

4	EDIÇÃO DO GENOMA DE FEIJÃO-CAUPI ATRAVÉS DO SISTEMA CRISPR/CAS9 DIRECIONADA AO GENE EIF(ISO)4E VISANDO AQUISIÇÃO DE IMUNIDADE/RESISTÊNCIA AO CABMV.....	67
5	DISCUSSÃO GERAL.....	99
6	CONCLUSÕES.....	101
	REFERÊNCIAS.....	102
	APÊNDICE – ARTIGO PUBLICADO NA REVISTA BIOINFORMATICS AND BIOLOGY INSIGHTS.....	110

1 INTRODUÇÃO

O feijão-caupi é uma excelente fonte de proteínas (cerca de 24% de sua composição), carboidratos (70%), vitaminas e sais minerais, além de possuir baixos teores de lipídios (Frota *et al.*, 2008). De acordo com Silva e Freire-Filho (1998), essa leguminosa é a principal fonte de proteínas de origem vegetal no Nordeste Brasileiro, compreendendo a principal fonte de proteínas nas regiões pobres do continente africano, devido ao alto preço da proteína animal em relação à vegetal.

As doenças figuram como os principais fatores de redução da produtividade do feijão-caupi. Fungos e vírus são responsáveis por 68,7% dos ataques de fitopatógenos nessa cultura (Obi e Barriuso-Vargas, 2014). Dentre as viroses, destaca-se o mosaico do feijão transmitido por *afídeo* (CABMV), cuja transmissão pode ocorrer através de sementes contaminadas ou por diferentes espécies de pulgão. O CABMV promove um mosaico intenso do limbo foliar, distorção das folhas e, em casos mais extremos, redução do crescimento da planta, o que pode acarretar perdas de até 65% na produtividade (Néya *et al.*, 2015). Dentre as infecções fúngicas, a *Macrophomina phaseolina* provoca lesões no caule, necrose e podridão, provocando um completo amarelecimento da planta, morte das sementes ou tombamento das plantas, possuindo uma alta frequência na cultura do feijão-caupi (Mohanapriya *et al.*, 2017).

Diante desse panorama, o melhoramento genético visando à obtenção de linhagens elite resistentes a fitopatógenos é fundamental para o crescimento da produtividade dessa cultura. Porém, o melhoramento convencional frequentemente enfrenta dificuldades para a introgressão de genes exógenos, devido à incompatibilidade de cruzamento do feijão-caupi com outras espécies do gênero *Vigna*, e à dificuldade de obtenção de uma linhagem elite a partir de banco de germoplasma das linhagens silvestres (Fang *et al.*, 2007). Nesse contexto, a engenharia genética torna-se uma ferramenta fundamental no melhoramento genético dessa espécie (Popelka *et al.*, 2006; Ji *et al.*, 2015).

Nos últimos anos, alguns trabalhos propiciaram o desenvolvimento de indivíduos transgênicos de feijão-caupi com transmissão estável do transgene para as gerações seguintes, seguindo o padrão Mendeliano (Popelka *et al.* 2006; Ivo *et al.* 2008; Bakshi *et al.* 2011; Mishra *et al.* 2014). Nesse cenário, merece destaque a transformação via biobalística, através do sistema de aceleração de micropartículas

revestidas com o transgene, propiciou a obtenção de linhagens transformadas de diversas cultivares da cultura (Ivo *et al.*, 2008; Citadin *et al.*, 2013; Cruz e Aragão, 2014).

Um dos mecanismos de defesa dos vegetais é baseado na imunidade inata, na qual a planta reconhece substâncias conservadas em patógenos e induz a produção de compostos com atividade antimicrobiana, dentre eles as defensinas (Burketová *et al.*, 2015). Elas são classificadas na superfamília PR-12, sendo formada por oito resíduos de cisteínas, formando quatro pontes dissulfeto e apresentam atividades antimicrobianas (antifúngica, antibacteriana e antiviral) e contra estresses abióticos (Lacerda *et al.*, 2014). Nesse contexto, o melão-de-são-caetano (*Momordica charantia*, Cucurbitaceae) destaca-se pelo seu potencial antimicrobiano (Saravani *et al.*, 2019), anticancerígeno (Ray *et al.*, 2010; Günes *et al.*, 2019), inclusive sendo doadora de gene para atividade antifúngica em arroz transgênico (Zhang *et al.*, 2019). Portanto, o banco de dados transcriptômicos de melão-de-são-caetano pode ter um potencial na identificação de novas defensinas com atividade antimicrobiana, aplicáveis para a engenharia genética.

Além da transformação genética, uma nova estratégia pode ser utilizada na busca pela resistência a patógeno, compreendendo a chamada “edição gênica” pelo sistema CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*), o qual compreende um componente natural do sistema imunológico bacteriano, atuando contra invasão de DNA exógeno de vírus ou plasmídeos (Barrangou *et al.*, 2007). Esse sistema tem sido utilizado como uma ferramenta poderosa na edição de genoma de microrganismos, animais e plantas (Belhaj *et al.*, 2015; Singh *et al.*, 2017).

O gene eIF4E é um fator de iniciação da tradução, sendo o responsável pelo reconhecimento do CAP ligado à extremidade 5' do RNA mensageiro, promovendo o recrutamento dos demais fatores traducionais e do ribossomo (Sesma *et al.*, 2017). Sabe-se que os vírus de RNA de fita simples, como os *Potyvirus*, possuem uma proteína denominada VPg que simula o CAP e utiliza o eIF4E para a tradução das suas próprias proteínas (Chandrasekaran *et al.*, 2016). Pesquisa recente com a utilização do sistema CRISPR/Cas9 mostrou que o nocaute do gene *eIF(iso)4E* em *Arabidopsis thaliana* conferiu imunidade ao *Turnip mosaic virus* (TuMV), não afetando o crescimento e a morfologia do vegetal, pois a presença do parálogo selvagem *eIF4E* permitiu a ocorrência normal dos processos traducionais na planta

(Pyott *et al.*, 2016). Semelhante ao trabalho supracitado, Chandrasekaran *et al.* (2016) verificaram que existem dois parálogos em pepino (*eIF4E* e *eIF(iso)4E*), e realizaram o nocaute do gene *eIF4E* através do sistema CRISPR/Cas9. As plantas mutantes homozigotas apresentaram imunidade ao *Ipomovirus* CVYV e elevado grau de resistência a dois *potyvirus* (ZYMV e PRSV-W) (Chandrasekaran *et al.*, 2016).

Em vista do exposto, o presente trabalho visou aplicar metodologias de modificação genética (transformação vegetal e edição gênica), de modo a conferir resistência a fitopatógenos do feijão-caupi (CABMV, CPSMV, *Rizoctonia solani* e *Macrophomina phaseolina*), desta forma auxiliando na fitossanidade e propiciando aumento na produtividade agrícola da cultura em regiões semiáridas.

1.1 OBJETIVOS

1.1.1 Objetivo geral

- ✓ Obter uma linhagem de feijão-caupi (*Vigna unguiculata*) com resistência a um fitopatógeno viral ou fúngico a partir de inferências biotecnológicas, com ênfase para a transformação e a edição gênica (técnica de CRISPR/Cas9).

1.1.2 Objetivos específicos

- ✓ Isolar, caracterizar e selecionar ao menos um gene candidato do grupo das defensinas a partir da biblioteca de genes diferencialmente expressos disponíveis nos transcriptomas de semente de *Momordica charantia*, através de ferramentas de bioinformática, considerando seu potencial antimicrobiano.
- ✓ Construir o vetor e transformar feijão-caupi com o gene da defensina de *M. charantia*.
- ✓ Realizar bioensaios das plantas de feijão-caupi transformadas com a defensina contra fitopatógenos do feijão-caupi.
- ✓ Desenhar primers e sequenciar os genes *eIF4E* e *eIF(iso)4E* da cultivar Boca Negra de feijão-caupi, avaliando sua estrutura em nível genômico.
- ✓ Desenhar o gRNA a partir do resultado do sequenciamento dos genes *eIF4E* e possíveis isoformas.
- ✓ Transformar embriões de feijão-caupi via *Agrobacterium* com vetor de CRISPR/Cas9 para edição do *eIF4E* e *eIF(iso)4E*, selecionando plantas modificadas.
- ✓ Promover o desafio das plantas editadas via CRISPR/Cas9 com o CABMV.

2 REVISÃO DA LITERATURA

2.1 FEIJÃO-CAUPI (*VIGNA UNGUICULATA*)

O feijão-caupi, feijão fradinho ou macassar [*Vigna unguiculata* (L.) Walp.] é uma leguminosa da família Fabaceae. É uma espécie herbácea anual, autógama, caracterizada por apresentar raízes profundas (Verdcourt, 1970), além de possuir diversos genes e mecanismos moleculares que favorecem a tolerância à seca (Chamarthi *et al.*, 2019). Seus grãos são altamente nutritivos, possuindo alto teor de proteínas (19% a 24%), carboidratos (70%), além de excelente fonte de vitaminas do complexo B e lipídios insaturados (Frota *et al.*, 2008). Nas regiões mais pobres da África, América Central e Nordeste Brasileiro o feijão-caupi representa a principal fonte de proteína da dieta diária, tendo em vista o baixo custo quando comparado à proteína de origem animal (Silva e Freire-Filho, 1998; Hall, 2012).

Em relação à safra de 2018/19 a cultura do feijão-caupi no Brasil ocupou uma área de 1.276.200 hectares (ha) com uma produção de 637.700 t e produtividade média de 377,7 kg/ha. A maior produtora foi a região Nordeste, seguida pelo Centro-oeste e o Norte, com cerca de 64,2%, 24,4% e 10% da produção brasileira, respectivamente. A área total de cultivo no Brasil é de 1.276.200,1 ha, sendo o Nordeste responsável por 82% desta área, seguido do Centro-Oeste com 11,2 e o Norte 5,4%. A produtividade da região Nordeste foi 4,4 vezes menor que a da região Centro-oeste (CONAB, 2019). Vale ressaltar o crescimento da produção na região centro-oeste entre 2009 e 2019, período em que passou de 8,14% para 24,4% (Freira-Filho *et al.*, 2011; CONAB, 2019). Isso se deve ao fato do grão ter um preço bem atrativo (comparativamente a outros feijões), ao uso de agricultura mecanizada e ao regime pluviométrico adequado.

2.2 MELHORAMENTO TRADICIONAL DA DEFESA CONTRA FITOPATÓGENOS

2.2.1 Impactos dos vírus CABMV e CPSMV em feijão-caupi e *status quo* do melhoramento visando à resistência

Os fatores mais limitantes da produção do feijão-caupi no Brasil são as viroses, com destaque para o Cowpea severe mosaic virus (CPSMV) (família Comoviridae) e o Cowpea aphid-borne mosaic virus (CABMV) (família Potyviridae),

devido à severidade e à ampla ocorrência desses fitopatógenos, respectivamente (Barros *et al.*, 2013).

O CPSMV, ou vírus do mosaico severo do caupi, reduz a capacidade germinativa das sementes das plantas infectadas, além de provocar escurecimento dos grãos que adquirem uma coloração amarronzada. A doença é extremamente severa na planta infectada, levando a uma drástica redução do crescimento devido à necrose dos brotos foliares e florais, além de provocar mosaico de coloração amarelada nas folhas estabelecidas. O vírus é facilmente transmitido mecanicamente, sendo naturalmente transmitido por algumas espécies de coleópteros, não sendo transmitido por sementes (Lima *et al.*, 2005).

Por sua vez, o CABMV ou vírus do mosaico do caupi transmitido por afídeos, é disseminado por diferentes espécies de pulgões, principalmente *Aphis craccivora* Koch. Destaca-se pela ampla ocorrência e por provocar mosaico intenso do limbo foliar, distorção das folhas e, em casos mais extremos, redução do crescimento da planta. Também é transmitido através das sementes das plantas infectadas, o que propicia a disseminação da doença para regiões livres da virose (Barros *et al.*, 2013; Sousa, 2013).

O melhoramento clássico é fundamental na obtenção de variedades de plantas produtivas com características agronômicas superiores, sendo o pilar do avanço da agricultura desde a domesticação das plantas. Em relação ao melhoramento convencional do feijão-caupi para obtenção de linhagens resistentes ou imunes ao CABMV e CPSMV, os resultados ainda são pouco satisfatórios. A maioria dos trabalhos publicados avaliou diversos genótipos em relação à resistência ou à susceptibilidade ao CABMV ou ainda reportam a realização de cruzamentos entre cultivares resistentes e susceptíveis para análise da proporção mendeliana e detecção do modo de herança da resistência, ou seja, se é monogênica ou poligênica, assim como se é recessiva ou dominante (Bashir e Hampton, 1996; Lima *et al.*, 2011; Orawu *et al.*, 2013; Dhanasekar e Reddy, 2015; Antoine *et al.*, 2016; Ayeleke *et al.*, 2016).

Considerando-se o padrão de herança da resistência ao potyvirus, os diversos trabalhos exibem resultados distintos. Orawu *et al.* (2013) realizaram uma análise da resistência e susceptibilidade de 54 genótipos de feijão-caupi, promovendo cruzamentos entre linhagens resistentes e susceptíveis. Os autores detectaram que quatro genótipos resistentes tiveram como resultado do cruzamento

em F₂ a proporção de 1 R : 3 S (resistente / susceptível), o que mostra o efeito de um único gene recessivo responsável pela resistência. Outras oito linhagens tiveram resultados distintos, indicando que mais de um par de genes governaria a resistência ao CABMV.

Antoine *et al.* (2016) verificaram o padrão de herança de duas variedades resistentes ao CABMV de Burkina Faso, cruzando-as com cultivares susceptíveis comumente utilizadas pelos agricultores locais. Como resultado foi verificado que ambas tiveram resultados na F₂ de 15 R : 1 S, demonstrando que dois pares de genes dominantes determinam a resistência a essa potyvirose. Porém, quando realizado o cruzamento entre as duas variedades resistentes, a F₂ apresentou proporção de 15 R : 1 S, o que demonstra que os genes que governam a resistência ao CABMV de uma cultivar podem ser diferentes daqueles da outra.

Leão *et al.* (2016) realizaram cruzamentos de linhagens de feijão-caupi com resistência a duas viroses (CABMV e CPSMV) com cultivares com características agronômicas superiores (sementes brancas, rugosas, com pequeno anel e hilo, sem nenhum halo e peso de 100 sementes de 25 a 30 g), sendo obtidas linhagens resistentes às duas viroses, embora as demais características agronômicas não tenham sido as desejadas.

Em relação à busca de resistência ao CPSMV em feijão-caupi, o panorama é semelhante ao CABMV. Lima *et al.* (2011) testaram 33 genótipos contra quatro isolados de CPSMV, dois de CABMV e um de CMV, visto que infecções sinérgicas são detectadas frequentemente no cultivo de feijão-caupi no Brasil e nos EUA. Desses, apenas um genótipo foi imune a todos os sorotipos de CABMV e CPSMV, embora nenhum deles tenha apresentado imunidade às três viroses. Alguns resultados mostraram que determinados genótipos são imunes a algum(ns) sorotipos de CABMV ou CPSMV, mas susceptíveis a outro(s), o que demonstra que mais de um par de genes governa a resistência a essas viroses. Segundo Umaharan *et al.* (1997), três pares de genes recessivos governam a resistência ao CPSMV. Tal resultado foi obtido a partir de cruzamentos entre linhagens resistentes e susceptíveis de feijão-caupi, onde a progênie apresentou a proporção de 63 susceptíveis para uma resistente.

Trabalho semelhante ao anterior foi realizado por Barros *et al.* (2013), porém visando à obtenção de uma linhagem resistente às duas viroses (CABMV e CPSMV). Através de oito cruzamentos com linhagens resistentes e susceptíveis foi

obtida uma progênie com a proporção de 15 susceptíveis para cada resistente aos dois vírus, o que indica a ocorrência de dois pares de genes recessivos governando a resistência a ambas as viroses.

As experiências com o melhoramento convencional mostram um gargalo na obtenção de linhagens resistentes ao CABMV e CPSMV, especialmente na obtenção de cultivares resistentes à virose e com características agronômicas superiores (Leão *et al.*, 2016). Também vale destacar que o feijão-caupi pode ser acometido simultaneamente por múltiplas viroses, tornando o trabalho dos melhoristas na busca por múltiplas fontes de resistência ainda mais desafiador (Palanga *et al.*, 2016). Diante desse panorama, a utilização de técnicas modernas como o auxílio da engenharia genética, pode ser fundamental para a obtenção de uma linhagem elite resistente às viroses em feijão-caupi.

2.2.2. Impactos dos fungos fitopatogênicos em feijão-caupi e *status quo* do melhoramento visando à resistência

Em relação ao desenvolvimento de linhagens de feijão-caupi resistentes a fungos fitopatogênicos, os resultados são menos conclusivos do que os disponíveis para viroses. As principais doenças que acometem a cultura são o oídio, a podridão cinzenta do caule e a mela (Pereira *et al.*, 2013; Kankam *et al.*, 2018; Shoaib *et al.*, 2018).

O fungo *Oidium sp.* provoca manchas branco-acinzentadas nas folhas, caule e vagens, provocando sua queda, sendo frequentemente associado a regime de sequeiro ou irrigação por gotejamento. O controle é realizado através do manejo e forma de cultivo da cultura (Pereira *et al.*, 2013)

A podridão cinzenta do caule é ocasionada pela *Macrophomina phaseolina*, que – além de provocar lesões no caule, necrose e podridão – leva a um completo amarelecimento da planta, podendo ocorrer a morte das sementes ou o tombamento das plantas (Athayde Sobrinho, 2005). Essa doença é de alta frequência na cultura de feijão-caupi, visto que a maior parte das plantações é em regime de sequeiro, o que favorece o desenvolvimento e a proliferação do fungo (Mohanapriya *et al.*, 2017). Alguns trabalhos vêm desvendando fontes de resistência a isolados dos fungos, verificando-se a existência de possíveis QTLs e marcadores moleculares

para auxílio no desenvolvimento de linhagens resistentes (Muchero *et al.*, 2011; Gomes-Silva *et al.*, 2017).

A mela é provocada pelo fungo *Rhizoctonia solani*, acomete drasticamente a germinação, porém quando a planta acometida já emergiu, provoca lesões foliares como manchas de aparência aquosa, bordas endurecidas que evoluem para necrose (Nechet e Vieira, 2007). As ações para combate à doença basicamente se dão na busca por cultivares com algum nível de resistência, assim como por controle químico e biológico (Nechet e Vieira, 2007; Ramusi *et al.*, 2017; Tenório *et al.*, 2019).

2.3 TRANSFORMAÇÃO GENÉTICA DE PLANTAS

A transformação genética compreende a transferência de genes utilizando técnicas de engenharia genética. A transformação de plantas começou a ser executada no início da década de 80. As técnicas mais utilizadas nos dias atuais compreendem a transformação indireta via *Agrobacterium tumefaciens* e a transformação direta via biobalística (Zanettini e Pasquali, 2004).

2.3.1 Transformação via *Agrobacterium tumefaciens*

A transformação via *A. tumefaciens* compreende o primeiro mecanismo conhecido de transferência horizontal de genes em plantas. Isso se deve à capacidade que essa bactéria possui em transferir seu plasmídeo para o genoma da planta hospedeira, promovendo a formação de tumores. Para a transformação genética de plantas, plasmídeos nativos dessa bactéria têm sido modificados, com a retirada dos genes responsáveis pela formação dos tumores sem afetar os genes responsáveis pela inserção do T-DNA no genoma da planta hospedeira. Para transformação por esse método, é necessário o cocultivo do explante com a bactéria tanto em meio líquido, quanto em meio semissólido, sendo realizados ferimentos nos explantes para aumentar o potencial de penetrabilidade das bactérias nas células vegetais (Zanettini e Pasquali, 2004).

Alguns fatores são fundamentais para a eficiência da transformação via *A. tumefaciens*, como: o explante utilizado; a forma de ferimento (por exemplo agulha, bisturi e/ou sonicação); o uso de compostos como a acetoseringona, que ativa os genes de virulência presentes no vetor e permite uma maior probabilidade de inserção do T-DNA; o uso de compostos de Tiol, que favorece uma maior

penetrabilidade das bactérias. Após os ajustes de parâmetros, o método de transformação em geral é eficiente, ocorrendo a inserção de uma ou poucas cópias do transgene (Figura 1). Apesar do índice de sucesso, o processo envolve muitas etapas na cultura de tecidos, sendo imprescindível que a planta hospedeira seja susceptível à agrobactéria (Popelka *et al.*, 2006; Solleti *et al.*, 2008b; Bakshi *et al.*, 2011; Bett *et al.*, 2019).

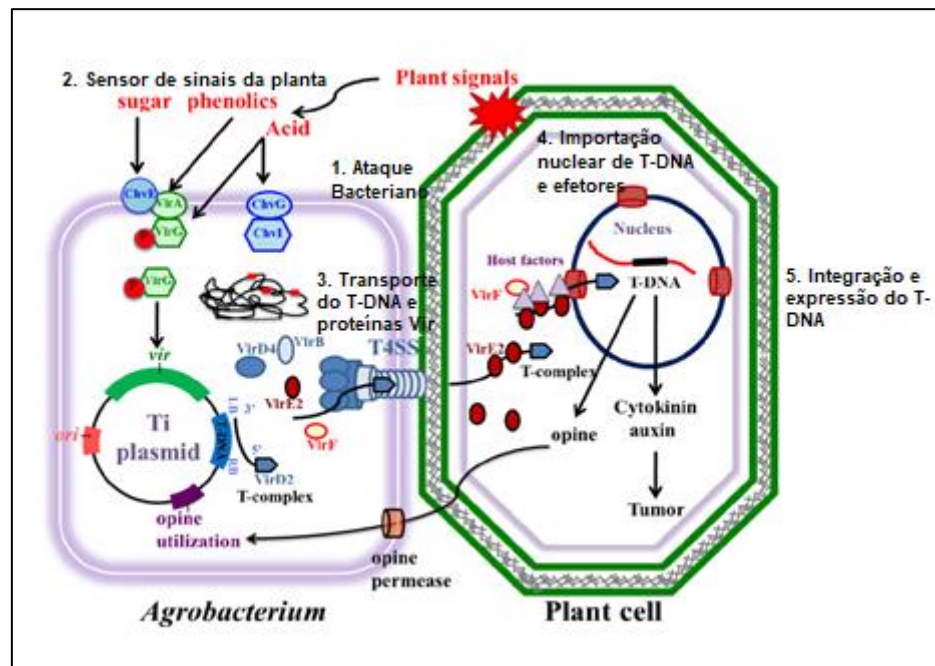


Figura 1: Processo de transferência horizontal do DNA entre *Agrobacterium tumefaciens* e uma célula vegetal. Através de sinalizadores como compostos fenólicos ocorre a estimulação dos genes de virulência do plasmídeo Ti, o qual promove a expressão do T-DNA, que é transferido da bactéria até o núcleo da célula vegetal, onde ocorrerá a integração do T-DNA no genoma da planta hospedeira. Fonte: Hwangb *et al.* (2017).

2.3.2 Transformação via Biobalística

A biobalística corresponde à inserção direta de um DNA exógeno no organismo-alvo, com a utilização de um equipamento denominado Gene Gun. A estrutura desse equipamento corresponde a: uma câmara de alta pressão, gerada pelo gás hélio, visto que é inerte e, portanto, não promoverá alterações no desenvolvimento dos explantes após a cultura de tecidos; a câmara de vácuo, local onde são inseridos os explantes a serem bombardeados. Na parte terminal da câmara pressurizada com gás hélio é inserida a membrana de ruptura, que no

momento do “tiro” será perfurada permitindo desencadear uma onda de choque que irá mover as partículas recobertas com o plasmídeo. Poucos milímetros abaixo da membrana de ruptura é inserida uma membrana de carregamento que, como o próprio nome informa, irá carregar as micropartículas de ouro ou tungstênio revestidas com o DNA a ser inserido no organismo-alvo. Alguns milímetros abaixo da membrana de carregamento é colocada uma tela de retenção, visando reter a membrana de carregamento após o “tiro”, permitindo apenas a passagem das micropartículas juntamente com o DNA até os explantes, colocados de 5 a 20 cm abaixo, na câmara de vácuo (Zanettini e Pasquali, 2004; Rech *et al.*, 2008).

Alguns fatores são fundamentais para a eficiência da transformação via biobalística, como: a distância entre a membrana de carregamento e a tela de retenção; a distância entre a tela de retenção e os explantes; a pressão do vácuo, para reduzir a resistência do ar e permitir a velocidade ótima do deslocamento das micropartículas; a pressão do gás hélio, bem como o explante a ser utilizado. Ajustados os parâmetros ideais, o processo de biobalística tem sido considerado um processo rápido, comparativamente às demais técnicas para a transformação genética do feijão-caupi, envolvendo um número menor de etapas de cultura de tecidos. Outra vantagem se deve ao fato de que pode ser executado em qualquer espécie vegetal, sem nenhuma restrição. Porém, os equipamentos e insumos são custosos e pode ocorrer com mais frequência a inserção de múltiplas cópias ou fragmentos dos transgenes (Aragão *et al.*, 1996; Ivo *et al.*, 2008; Rech *et al.*, 2008; Abdallah *et al.*, 2010).

2.4. DEFENSINAS VEGETAIS

As defensinas são pequenos peptídeos antimicrobianos com cerca de 50 aminoácidos, caracterizadas pela conservação de oito resíduos de cisteína, que promovem quatro ligações dissulfeto. Elas possuem três folhas-beta e uma alfa-hélice, peso molecular de cerca de 5 KDa e são classificadas na família PR-12, ou seja, no grupo das proteínas relacionadas à patogênese (*Pathogenesis-Related – PR*) (Lacerda *et al.*, 2014). Sua atividade mais reportada é antifúngica, porém existem defensinas com atividades antibacteriana e inseticida (Gao *et al.*, 2000; Choi *et al.*, 2009; Darwish *et al.*, 2014), ou ainda mais raramente antiviral e protozoocida (Brogden, 2005).

Defensinas possuem um peptídeo sinal que direciona sua secreção para o meio extracelular, porém algumas apresentam um pró-peptídeo C-terminal (CTPP) responsável pela exportação da defensina para o vacúolo central, o que reduz sua fitotoxicidade no ambiente celular (Gaspar *et al.*, 2014). As defensinas atuam promovendo a degradação da membrana plasmática dos fungos, formando poros, ou aumentando a permeabilidade da membrana a compostos iônicos. Ambas as ações culminam com a necrose e/ou morte do patógeno (Stotz *et al.*, 2009). Mais recentemente, alguns trabalhos vêm demonstrando outras atividades da defensina que não a atividade antimicrobiana, como aumento da tolerância ao déficit hídrico, frio, embebição e déficit de oxigênio (Weerawanich *et al.*, 2018; Kumar *et al.*, 2019),

2.5 PLANTAS TRANSGÊNICAS EXPRESSANDO DEFENSINAS EXÓGENAS

Devido ao seu potencial antimicrobiano, diversos trabalhos com a superexpressão de um peptídeo antimicrobiano em plantas transgênicas foram publicados nos últimos anos. O primeiro transgênico expressando uma defensina foi gerado em experimento conduzido por Gao e colaboradores (2000). A batata transgênica portadora do gene de defensina *alfAFP*, de alfafa, teve o nível de infecção por colônias do fungo *Verticillium dahliae* reduzido em seis vezes comparativamente ao controle não transformado. As plantas transformadas não tiveram diferenças fenotípicas e de taxa de crescimento quando comparadas ao controle sem a presença do patógeno. Os resultados obtidos em casa de vegetação foram consistentes com os resultados das análises em campo, o que reforçou a atividade *in vivo* da defensina analisada (Gao *et al.*, 2000).

Em 2002, Kazanki e colaboradores desenvolveram uma planta transgênica de arroz expressando uma defensina de wasabi (*Eutrema japonicum* Matsum., Brassicaceae). Porém, os níveis de resistência ao fungo *Magnaporthe grisea* (*blast fungus*) dos diversos transformantes de arroz obtidos foram bastante variáveis, com plantas transgênicas sendo tão resistentes quanto o controle resistente, enquanto outras linhagens transgênicas foram tão susceptíveis quanto o controle não transgênico. Esta variação dos níveis de resistência foi justificada pelos efeitos epigenéticos da posição de inserção do gene no genoma, associados a eventos de silenciamento gênico ou ainda a possíveis fatores pós-transcricionais.

Chen e colaboradores (2006) usaram três construções de vetores para transformação de plantas de tomate. Uma com o gene *alfAFP* (defensina de alfafa),

outra com o gene *GLU* (glucanase de tabaco), bem como a construção com os dois genes combinados (alfAFP-GLU). Os pesquisadores observaram que o nível de resistência à bactéria *Ralstonia solanacearum* das plantas transgênicas carregando apenas um gene de defensina foram semelhantes, porém as plantas transgênicas com os dois genes tiveram uma resistência mais acentuada, o que indica um efeito sinérgico da coexpressão desses genes.

A transformação de tabaco e amendoim com o gene *BjD* (defensina de mostarda) resultou em níveis significativamente elevados de resistência das plantas transgênicas aos fungos *Phytophthora parasitica* e *Fusarium moniliforme*; enquanto as plantas de amendoim se tornaram resistentes aos fungos *Pheoisariopsis personata* e *Cercospora arachidicola* (Anuradha *et al.*, 2008).

Em 2010, Abdallah e colaboradores desenvolveram um tomate transgênico, portador do gene *Msdef1* (defensina de *Medicago sativa*), com níveis de resistência elevados a *Fusarium oxysporum*. Nos bioensaios de inoculação, as plantas transgênicas chegaram a uma taxa de 70% sem apresentar sintomas do patógeno enquanto apenas 7% das plantas-controle (tomate não transformados) foram assintomáticas.

Gaspar e colaboradores (2014) desenvolveram plantas de algodão cv. Coker transgênicas portando o gene *NaD1* responsável pela síntese de uma defensina da classe II de *Nicotiana glauca*, sendo o primeiro transgênico com defensinas dessa classe. O grupo realizou, além de testes em casa de vegetação, três anos de bioensaios em campo. Foi verificado que os níveis de resistência das plantas transgênicas de algodão foram duas a três vezes maiores em relação à mesma cultivar não transformada submetida à inoculação do fungo patogênico *F. oxysporum* e 10% maiores nas transgênicas em relação à resistência ao fungo *V. dahliae*. O trabalho também mostrou que a produtividade das plantas transformadas com uma defensina, dirigida por um promotor constitutivo, foi semelhante àquela das não transformadas na ausência de patógeno, possivelmente devido ao armazenamento da defensina da classe II no vacúolo, o que promoveu uma menor fitotoxicidade.

Em arroz transgênico, foi demonstrado que a defensina NmDef02 foi eficiente contra o fungo *Sarocladium oryzae* (Pérez-Bernal *et al.*, 2017). Esta mesma defensina foi testada em tabaco e batata transgênicos visando resistência contra fungos (Portieles *et al.*, 2010). Thao e colaboradores (2017) transformaram tabaco com um gene codificante da defensina oriunda de *Vigna radiata* (VrDEF1) que está

relacionada à inibição da alfa-amilase de caruncho. Através de ensaio enzimático da alfa-amilase com proteína total das sementes das plantas transgênicas e não-transgênicas (controle) foi verificada a redução da atividade da enzima a 18%, tornando-a uma excelente candidata para uso biotecnológico relacionado à resistência a insetos.

A defensina J1-1 derivada de pimenta foi superexpressa em tabaco transgênico, a qual foi dirigida por um promotor específico de raiz (pPRP3) ou constitutivo (p35S). Foi observado um aumento da resistência a *Phytophthora parasítica* var. *nicotianae* nas plantas transgênicas com a defensina dirigida por qualquer um dos promotores. A superexpressão da defensina exógena promoveu um aumento da expressão de genes endógenos das classes PR2 e PR10, o que demonstra além do potencial que a defensina exógena tem em aumentar a resistência a fitopatógeno, ocorre uma alteração das vias de sinalização com mobilização do sistema de defesa da planta (Lee *et al.*, 2018). Portanto, o desenvolvimento inédito de um feijão-caupi transgênico superexpressando uma defensina, pode ser a chave para obtenção de resistência a fitopatógenos desta cultura.

2.6 MÉTODOS DE EDIÇÃO CONTROLADA DE GENOMA ATRAVÉS DE NUCLEASES PROGRAMÁVEIS

A modificação do genoma é realizada há mais de três décadas, assim como as técnicas de mutagênese e transgenia (Citadin *et al.*, 2013). Porém, através destas técnicas não é possível controlar o local exato da mutação ou inserção do transgene, sendo um evento aleatório. No caso da transgenia, pode haver insucesso da superexpressão do transgene, caso o mesmo se integre a uma região metilada do DNA, por exemplo (Cruz e Aragão, 2014). Através das nucleases programadas é possível realizar edições nas regiões-alvo específicas, realizando mutações ou inserções de sequências de forma direcionada, com exata precisão da região do genoma onde se pretende inserir a nova sequência (Kim e Kim, 2014). As três técnicas de nucleases programadas para edição controlada do genoma compreendem: ZFNs (Zinc-Finger Nucleases), TALENs (*transcription activator-like effector nucleases*) e CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) associado a nuclease Cas9 (Sun e Zhao, 2013; Kim and Kim, 2014).

Através destas técnicas é possível realizar mutações em sequências-alvo específicas, promovendo o nocaute de um gene específico para auxiliar no entendimento de sua função, visando ao melhoramento genético (Shan *et al.*, 2013; Chandrasekaran *et al.*, 2016), bem como para o reparo de um gene defeituoso (Kim e Kim, 2014) ou a modulação da expressão com modificações mais finas nas sequências gênicas (Liao *et al.*, 2017). É possível realizar a inserção programada de uma sequência nucleotídica numa região exata do genoma (Shi *et al.*, 2017) ou a indução de rearranjos cromossômicos, como inversão e translocação (Kim e Kim, 2014).

2.6.1 Ação das nucleases programáveis e métodos de reparo do DNA

Para que técnicas baseadas em nucleases programáveis sejam bem sucedidas, é realizado um desenho experimental para que ocorra o reconhecimento de uma região alvo específica do genoma, a qual será clivada pela nuclease programada (Kim e Kim, 2014).

Após a clivagem da dupla fita, há a ativação do sistema de reparo do DNA, seja pelo método NHEJ (*non-homologous end joining*) ou HDR (*homology directed repair*) (Figura 2). O método mais comum utilizado no reparo do DNA é o NHEJ, no qual ocorre a união direta das extremidades da dupla fita cortadas, o que promove inserções ou deleções alterando a matriz de leitura do gene que foi cortado, culminando com o nocaute (*knockout*) do gene (Shan *et al.*, 2013; Chandrasekaran *et al.*, 2016; Nekrasov *et al.*, 2017). Já pelo método HDR, o reparo da dupla fita cortada é realizado com uma sequência que serve como molde (*Donor DNA*), reparando a região da dupla fita do DNA que foi clivada, permitindo a inserção planejada e perfeita da sequência nucleotídica desejada (Belhaj *et al.*, 2015, Shi *et al.*, 2017). Para isto, o *Donor DNA* deve ter a região programada para inserção flanqueada por regiões homólogas às extremidades do DNA cortado pela nuclease (Shi *et al.*, 2017).

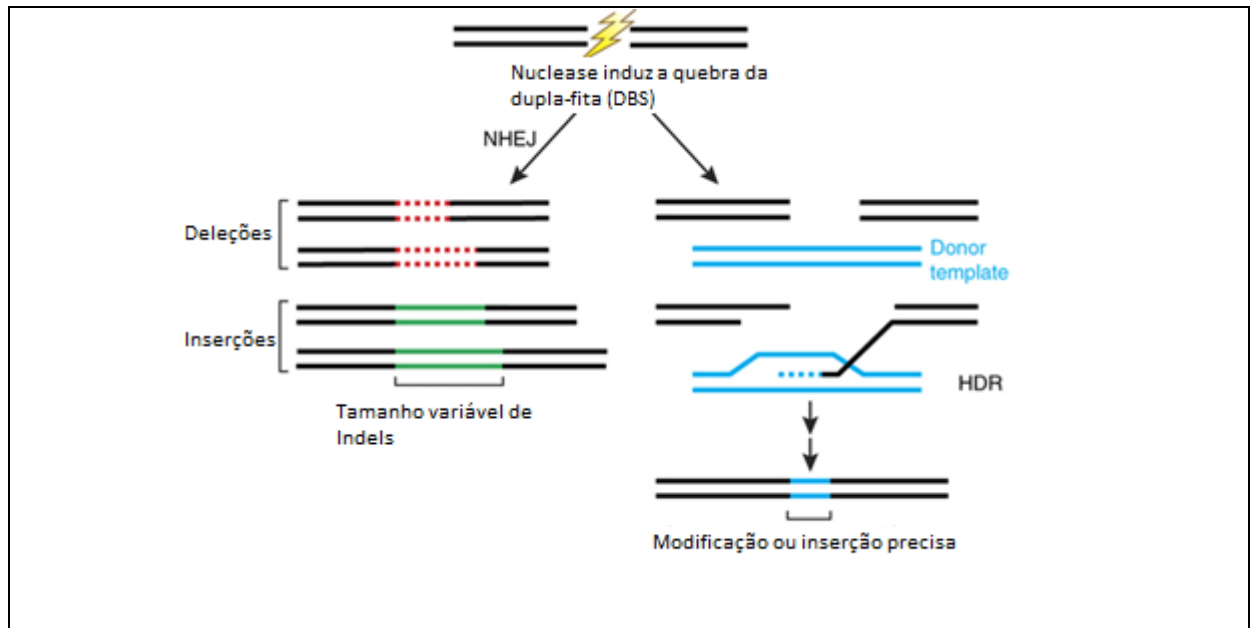


Figura 2: Métodos de reparo do DNA. Após a clivagem do DNA, pelo método NHEJ (*non-homologous end joining*) a dupla fita clivada tem suas extremidades unidas de forma passível de erro (*error-prone*) como, por exemplo, a origem de *indels*. Já pelo método HDR (*homology directed repair*) é fornecido um *donor* DNA, que possui homologia com as extremidades da dupla-fita clivada, mais uma região intermediária para a inserção programada da sequência em questão. Fonte: Sander e Joung (2013).

2.6.2 ZFN

Os ZFNs (*Zinc-Finger Nucleases*) são proteínas, cujas estruturas são formadas por dois domínios. Um domínio da proteína “dedo de zinco” (ZFP, *Zinc-Finger protein*) corresponde à região que reconhece e se liga à sequência específica do DNA conectada ao domínio da nuclease FokI, responsável pela quebra da dupla-fita de DNA. Cada *zinc-finger* reconhece três pares de bases da sequência-alvo, sendo cada ZFN formado por 3-6 *zinc-fingers*, o que permite o reconhecimento de uma região-alvo de 9-18 pb. Para uma maior especificidade dos ZFNs, foram desenvolvidas nucleases FokI que atuam apenas na forma dimérica. Portanto, dois ZFNs são necessários para edição planejada do genoma, reduzindo a frequência de *off-target* devido à duplicação da sequência de reconhecimento, aumentando a especificidade. Nesta interação dos dois ZFNs, há um espaçador de 5-7 pb para interação com uma sequência de 18 a 36 pb (Figura 3). Cada *zinc-finger* reconhece três bases, sendo necessária a existência de 64 *zinc-finger* para reconhecimento de

qualquer sequência. A limitação desta técnica, porém, é justamente devida à não existência de todas as combinações, além da alta taxa de *off-target*, devido ao pequeno tamanho da região alvo de reconhecimento do DNA e especificidade do reconhecimento do *zinc-finger*, além da elevada citotoxicidade (Kim e Kim, 2014).

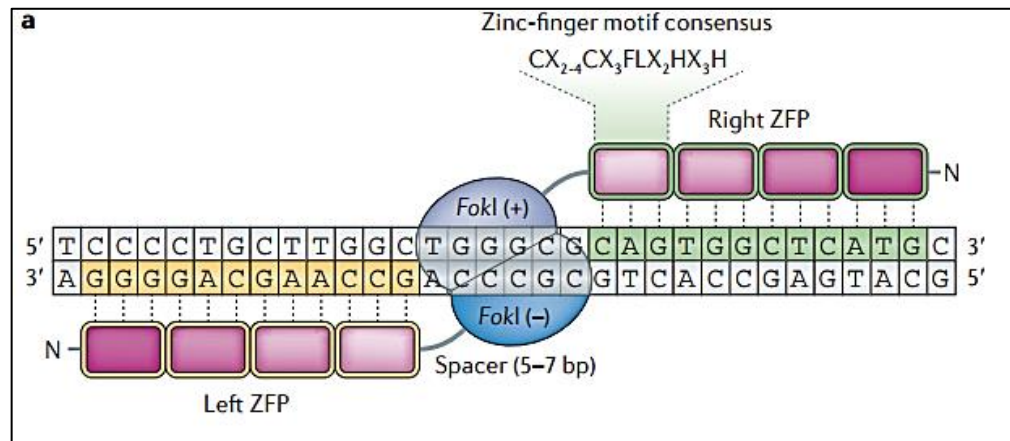


Figura 3: Estrutura dos ZFNs, com a formação da estrutura dimérica. A região ZFP, corresponde ao domínio de ligação ao DNA, onde cada *zinc-finger* reconhece três pb, e encontra-se ligada a proteína de clivagem do DNA (nuclease FokI). Na imagem é apresentado o consenso do *zinc-finger*, cuja letra X representa qualquer aminoácido. A estrutura dimérica leva à formação de uma região espaçadora (5-7 pb), porém aumenta a especificidade da região alvo. Fonte: Kim e Kim (2014).

2.6.3 TALENS

Os TALENs (*transcription activator-like effector nucleases*) são estruturalmente semelhantes aos ZFNs (Kim e Kim, 2014). A estrutura dos TALENs corresponde ao domínio da nuclease FokI que cliva o DNA, sendo ligada a uma proteína com um domínio de ligação de regiões específicas do DNA, podendo ser projetadas para a interação com qualquer sequência (Joung e Sander, 2013). A estrutura da região de ligação ao DNA é composta por regiões repetidas altamente conservadas (*TALE repeat domains*), sendo secretadas por bactérias do gênero *Xanthomonas*. O *TALE repeat domain* é flanqueado de cada lado por um domínio (N- e C-terminal), cujo complexo encontra-se ligado à nuclease FokI, conforme Figura 4 (Sun e Zhao, 2013). O TALEN é formado por estruturas de repetições em tandem, compostas por 33-35 aminoácidos, sendo que cada uma reconhece uma única base. A especificidade com que cada domínio repetido reconhece à sua base específica é determinada pelos aminoácidos das posições 12 e 13, que são

2.6.4 CRISPR

O sistema CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) é formado por repetições palindrômicas curtas, interespaçadas e regularmente agrupadas estando associado à proteína Cas (CRISPR-Cas). O CRISPR/Cas é um componente natural do sistema imunológico bacteriano, atuando contra a invasão de DNA exógeno de vírus ou plasmídeos (Barrangou *et al.*, 2007). Esse sistema tem sido utilizado como uma ferramenta poderosa na edição de genomas de microrganismos, animais e plantas (Belhaj *et al.*, 2015; Singh *et al.*, 2017). O sistema CRISPR-Cas9 possui, além das sequências curtas palindrômicas repetidas, uma sequência que transcreve o *trans-acting CRISPR RNA* (tracrRNA) e um locus do gene que sintetiza a proteína Cas9. A tracrRNA possui em torno de 42 nucleotídeos, tendo uma sequência homóloga às regiões palindrômicas repetidas e forma dois *hairpins* (ou grampos), sendo responsável pela ligação à proteína Cas9 (Barrangou *et al.*, 2007).

Existem três tipos de sistema CRISPR: I, II e III. O tipo II é o mais compreendido e utilizado na edição de genomas, no qual o CRISPR está associado à proteína Cas9. A ação do CRISPR-Cas9 é dividida em três passos: aquisição, processamento do RNA e interferência. Durante a aquisição, o fragmento do DNA invasor é reconhecido, capturado e integrado entre as sequências palindrômicas repetidas do CRISPR (etapa 1, Figura 5). Esse DNA exógeno é conhecido como espaçador (*spacer*) ou protoespaçador (*protospacer*). Posteriormente, o CRISPR realiza a transcrição do protoespaçador (DNA ou RNA exógeno integrado ao sistema) juntamente com as sequências palindrômicas repetidas originando o pré-CRISPR RNA (pre-crRNA) (etapa 2, Figura 5). A seguir ocorre a associação do pré-crRNA com o tracrRNA. Essa interação ocorre devido ao tracrRNA possuir uma sequência complementar às sequências palindrômicas repetidas presentes no pré-crRNA, o que permite a associação do pré-crRNA com a tracrRNA (pré-crRNA:tracrRNA). Posteriormente esse complexo interage com a proteína Cas9 e a RNase III, a qual cliva o pré-crRNA:tracrRNA no CRISPR RNA ativo ou crRNA (Singh *et al.*, 2017) (etapa 3, Figura 5).

Após o processamento do RNA, inicia-se a terceira e última fase que é a interferência, onde o complexo Cas9, crRNA e tracrRNA é direcionado para interagir com o DNA exógeno, com base no pareamento de bases com o crRNA. Após o

pareamento do crRNA com o DNA alvo, ocorre a separação da dupla fita de DNA, e a proteína Cas9 promove o corte da dupla fita de DNA (DBS, *double-stranded breaks*) três pares de bases antes do PAM (*protospacer adjacent motif*), promovendo a impossibilidade da replicação do DNA viral, o que determina a imunidade ao fago (Barrangou *et al.*, 2007) (etapa 4, Figura 5).

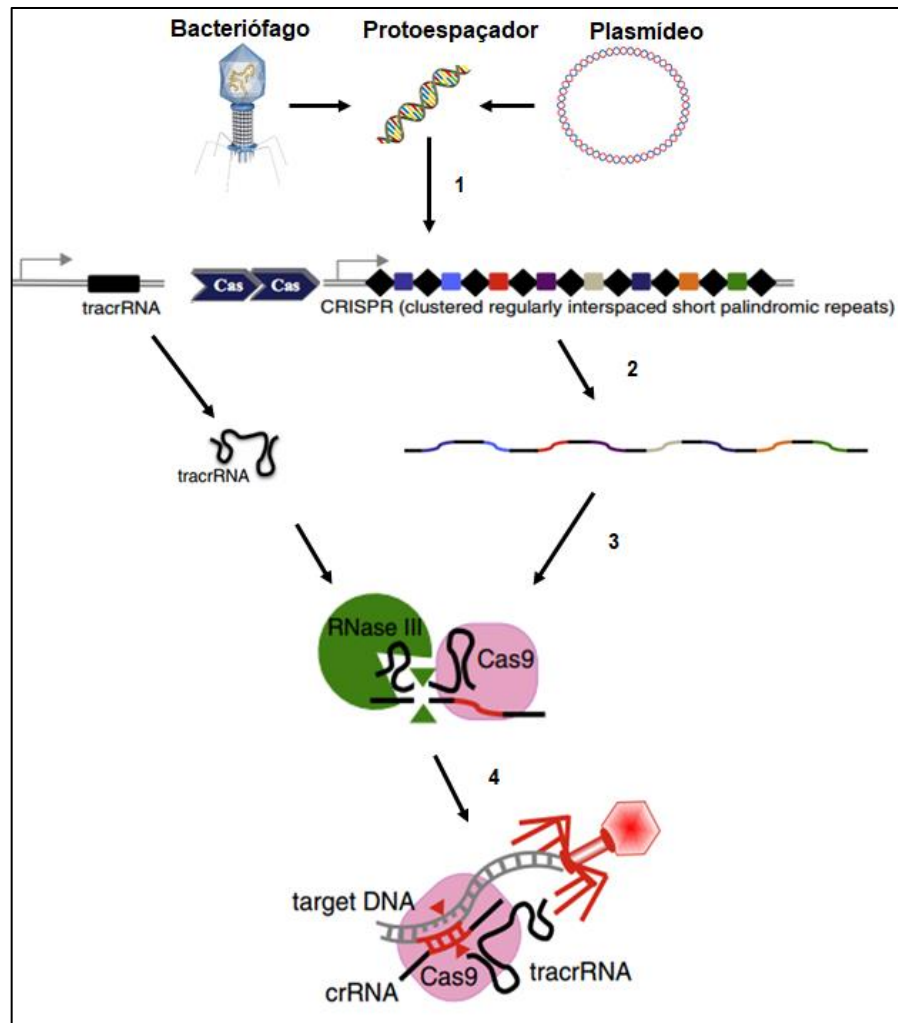


Figura 5: Esquema do funcionamento do CRISPR natural das bactérias. Na etapa 1 chamada de aquisição, é incorporado ao sistema um DNA exógeno (DNA de bacteriófago ou plasmídeo) localizado entre as sequências palindrômicas curtas repetidas. As etapas 2 e 3 correspondem ao processamento do RNA. Na 2 ocorre a transcrição do pre-CRISPR RNA e do tracrRNA. Na etapa 3 ocorre a interação do pre-crRNA com o tracrRNA, e destes com a RNase III e a Cas9, onde a RNase III promove o corte na região palindrômica curta repetida, originando o CRISPR RNA ou crRNA. Por fim, na etapa 4 (interferência) ocorre a interação do crRNA com a região homóloga do DNA viral, ocorrendo o corte da dupla fita do DNA viral pela

ação da Cas9. Adaptado de Singh *et al.* (2017) e Belhaj *et al.* (2015), com modificações pelo autor.

O PAM é uma sequência de três nucleotídeos localizada na extremidade 3' do DNA alvo, que em geral corresponde à sequência NGG, não presente no crRNA (Pyott *et al.*, 2016). Essa sequência PAM é fundamental para o sistema CRISPR-Cas9 diferenciar sequências de DNA endógenas e exógenas. Caso a sequência do crRNA não seja homóloga ao DNA exógeno ou este último não possua o PAM, não ocorre a clivagem do DNA e, conseqüentemente a bactéria não terá resistência ao vírus ou plasmídeo (Singh *et al.*, 2017).

Em relação aos outros dois tipos de CRISPR, a principal diferença está em relação às etapas do processamento do RNA e da interferência. No CRISPR tipo I quem realiza o processamento do pré-crRNA são as proteínas Cas6e e Cas6f. Após a formação do crRNA, ocorre a associação com a Cas3, que promoverá o corte da dupla fita no ato da interferência. Por sua vez, o CRISPR tipo III tem a Cas6 como processadora do pré-crRNA, sendo que após o processamento, forma-se o complexo do crRNA com proteínas Csm/Cmr, promovendo a interferência. Porém, diferentemente dos outros dois tipos, este sistema não requer o PAM e portanto é inespecífico e gera edições off-target (Singh *et al.*, 2017).

A utilização do sistema CRISPR/Cas9 em plantas baseia-se no fato de que a região do DNA exógeno é incorporada ao sistema e transcrita na forma de um *single guide* RNA ou sgRNA, que equivale à fusão do crRNA com o tracrRNA. Isso permite que o sistema seja desenhado para alvos específicos, como para a utilização na resistência a um vírus específico de uma determinada espécie vegetal, bem como para edição de outras regiões do genoma. A região guia do sgRNA compreende uma sequência de 20 a 25 nucleotídeos complementar à sequência alvo desejada, levando em conta que após o ponto de homologia é fundamental a existência da sequência PAM no alvo desejado. Isto permite que para cada edição do genoma, seja necessária apenas a alteração da sequência de 20-25 nucleotídeos do sgRNA. Atualmente, existem diversas ferramentas computacionais para realizar o desenho do sgRNA, embora o mesmo possa ser desenhado manualmente (Belhaj *et al.*, 2015; Singh *et al.*, 2017).

O sistema CRISPR/Cas9 está se mostrando como um dos mais eficientes, práticos e de fácil manipulação, uma vez que para edição do genoma só é

necessário alterar a sequência de 20 nucleotídeos do sgRNA. Isso torna este método ainda mais promissor, quando comparado aos outros dois métodos tais como TALENs e ZFNs, anteriormente mais utilizados na edição do genoma. Isso se deve ao fato desses dois sistemas necessitarem do desenho de uma proteína que age em sítios-alvo específicos do DNA promovendo sua edição, necessitando da construção de uma proteína específica para cada região específica do DNA a ser editada, o que torna o processo laborioso, de elevado custo financeiro, além de demandar elevado tempo experimental (Hsu *et al.*, 2013; Shan *et al.*, 2013; Shan *et al.*, 2014).

Quando se objetiva o nocaute de um gene da própria planta ou de um patógeno viral por CRISPR/Cas9, é realizado o desenho de um sgRNA específico para clivagem do gene-alvo, com sistema de reparo após a quebra da dupla fita pelo método de NHEJ, promovendo *indels* que resultam em uma alteração da matriz de leitura do gene (Shan *et al.*, 2014; Ji *et al.*, 2015). Quando o objetivo é adicionar uma sequência específica no genoma da planta ou realizar a substituição de uma ou mais bases nitrogenadas na sequência específica do genoma vegetal, é desenhado um sgRNA com um guia para a clivagem da região desejada, além de fornecer uma sequência de DNA com homologia para as regiões que flanqueiam o ponto de corte, juntamente com uma sequência específica para inserção ou substituição das bases pelo método de reparo por homologia direta (HDR), promovendo uma alteração da sequência genômica desejada (Shan *et al.*, 2014; Shi *et al.*, 2017).

2.7 GENES eIF4E E eIF(iso)4E: FUNÇÕES NOS VEGETAIS E USO BIOTECNOLÓGICO

Os genes da classe *eIF4E* (*eukaryotic translation Initiation Factor 4E*) compreendem fatores de iniciação da tradução, que atuam ligando-se ao RNA mensageiro (RNAm) e, posteriormente, permitindo a associação do RNAm à maquinaria dos ribossomos (Sesma *et al.*, 2017). O processo traducional é dividido em três etapas: iniciação, alongação e terminação, sendo a etapa de iniciação a mais regulada e aquela onde a proteína eIF4E atua. Existem três complexos fundamentais para a realização desta etapa: o **complexo eIF4F**, que é composto pela proteína eIF4E, e as helicases eIF4A e eIF4G; o **complexo 43S**, sendo formado pela subunidade menor (40S) do ribossomo, com a proteína eIF3 e o

complexo ternário, que é formado pela eIF2, GTP e o RNA transportador inicial com o aminoácido metionina (Met-tRNAi) (Montero *et al.*, 2015).

O início da tradução ocorre pela interação da proteína eIF4E com a região cap localizada na extremidade 5' do RNAm. Após esta interação, a eIF4E recruta a eIF4G e a eIF4A, formando o complexo eIF4F. Este complexo, através do eIF4G se liga ao PABP (*poly-A-binding protein*), proteína que se liga à cauda poli-A presente na extremidade 3' do RNAm de eucariotos (Bastet *et al.*, 2017). Em seguida o complexo eIF4F recruta o complexo 43S através da ligação do eIF4G com o eIF3. A seguir a subunidade menor transporta o GTP, Met-tRNAi e eIF2 até o códon de iniciação, no qual ocorre a associação da subunidade maior, originando o ribossomo 80S, com a liberação do eIF2 com o GDP. O GDP é transformado em GTP pelo fator eIF2B, ficando disponível para uma nova etapa de iniciação da tradução (Figura 6) (Montero *et al.*, 2015).

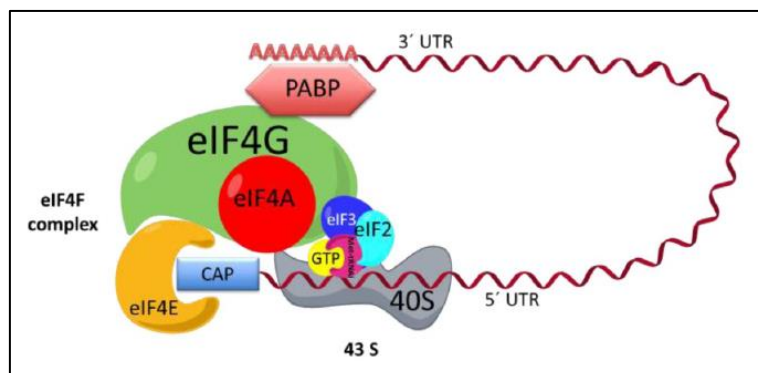


Figura 6: Esquema do processo de início da tradução em eucariotos. Inicialmente ocorre a interação do eIF4E com o CAP da extremidade 5' do RNA mensageiro (RNAm), o qual recruta o eIF4A e eIF4G. Posteriormente ocorre a interação do eIF4G com o PABP, proteína que se liga à cauda poli-A da extremidade 3' do RNAm. Por fim, ocorre a interação da subunidade menor do ribossomo, com o GTP, Met-tRNAi, eIF2 e eIF3, sendo todos estes componentes recrutados pelo eIF4G, que interage com o eIF3. Fonte: Montero *et al.*, 2015.

No curso do processo evolutivo dos vegetais, os genes *eIF4E* sofreram rearranjos (como duplicações), mutações e translocações, o que permitiu o surgimento de parálogos e/ou isoformas dos *eIF4E*, que têm funções redundantes. Estas proteínas possuem estruturas terciárias bastante conservadas, sendo

formadas por oito folhas-beta antiparalelas sustentadas por três alfa-hélices longas (Marcotrigiano *et al.*, 1997).

Estudos têm demonstrado que os vírus de RNA de fita simples que acometem diversas doenças em plantas, em especial o grupo do *Potyvirus*, utilizam os *eIF4E* para a tradução do seu RNA. Eles possuem uma proteína VPg (*viral protein genome-linked*), que recruta a proteína eIF4E da célula hospedeira, utilizando-a para o seu processo traducional (Chandrasekaran *et al.*, 2016). No entanto, muitos deles possuem especificidade a apenas um dos parálogos dos *eIF4E* no seu processo traducional, sendo verificado que a presença do parálogo *eif4e* mutante (ou nocauteado) no genoma da planta pode impedir a interação com o vírus e, consequentemente impedir que complete seu ciclo, gerando resistência ou até imunidade da planta ao vírus (Chandrasekaran *et al.*, 2016; Pyott *et al.*, 2016; Bastet *et al.*, 2017).

2.8 BIOTECNOLOGIA APLICADA PARA OBTENÇÃO DE CULTIVARES SUPERIORES EM FEIJÃO-CAUPI

A aplicação de técnica de transgenia em feijão-caupi se deu em meados dos anos 80 (Garcia *et al.*, 1987). Porém, o desenvolvimento de linhagens com características superiores e transmissão do transgene seguindo o padrão mendeliano só foi possível no século XX, com destaque para Soleti e colaboradores (2008a) com desenvolvimento de plantas com ação inibitória da alfa-amilase 1. Mais tarde Higgins e colaboradores (2010) desenvolveram o feijão-caupi Bt com o gene *Cry1Ab* obtendo linhagem resistente ao inseto *Maruca vitrata*, sendo realizados diversos testes de campo no continente africano, com liberação para comercialização em 2019 (Nigerian Biosafety Management Agency). Este foi o primeiro feijão-caupi geneticamente modificado liberado para consumo humano (ISAAA).

Em 2013, Citadin e colaboradores desenvolveram feijão-caupi GMO com tolerância ao herbicida *Imazapyr*. Cruz e Aragão (2014) desenvolveram um feijão-caupi transgênico resistente aos vírus CABMV e CPSMV através do uso da tecnologia do RNA de interferência. No mesmo ano, Mishra e colaboradores (2014) desenvolveram um feijão-caupi GMO tolerante à seca através da inserção do gene *VrNHX1* de *Vigna radiata*. Por sua vez, Bett e colaboradores (2017) também

desenvolveram feijão-caupi transgênico Bt e Kumar e colaboradores (2017) desenvolveram um feijão-caupi transgênico com um RNAi relacionado à resistência ao *Mungbean yellow mosaic India virus* (MYMIV).

Com relação ao uso de técnicas de edição de genoma, existe apenas um registro de uso de CRISPR em feijão-caupi, onde através de *Agrobacterium rizhogenes* foi realizado o nocaute do gene SYMRK (*symbiosis receptor-like kinase*) em raízes da planta, o que impediu a formação de nódulos. O principal objetivo do trabalho foi verificar a eficiência do sistema CRISPR/Cas9 em feijão-caupi (Ji *et al.*, 2019). No entanto, plantas com o genoma editado ainda não foram relatadas.

3 CAPÍTULO I

Manuscrito de artigo
Em preparação para submissão à revista *Plant Cell Reports*
(Springer, Qualis B1)
<https://www.springer.com/journal/299>

Expressão de uma defensina de Melão-de-São-Caetano em feijão-caupi transgênico promovendo resistência a fitopatógeno fúngico

José Diogo Cavalcanti Ferreira^{1,2}, Francisco José Lima Aragão³, Luciana Gonçalves de Oliveira⁴, Antônio Félix da Costa⁴ and Ana Maria Benko-Iseppon²

¹Instituto Federal de Pernambuco, Br-232 Km 208, CEP 55200-000, Pesqueira, PE, Brasil;

²Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Departamento de Genética, Av. Prof. Moraes Rêgo, 1235, CEP 50670-420, Recife, PE, Brasil;

³Empresa Brasileira de Pesquisa Agropecuária, Embrapa Cenargen, Parque Estação Biológica – PqEB s/nº, Av. W3 Norte (final), CEP 70770-917, Brasília, DF, Brasil.

⁴Instituto Agrônomo de Pernambuco, Av. General San Martin, 1371, CEP: 50761-000, Recife, PE, Brasil

Autor correspondente: Ana Maria Benko-Iseppon, Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Av. Prof. Moraes Rego 1235, CEP 50.670-420, Recife, PE, Brasil. Tel. +55 81-2126-8520; Email: ana.iseppon@gmail.com

Resumo

O feijão-caupi é uma leguminosa caracterizada por sua rusticidade e por apresentar grãos com alto teor de proteínas. Sua produção e produtividade são reduzidas drasticamente pela ação de vírus e fungos. As defensinas são peptídeos antimicrobianos já utilizados para transformar diversas culturas vegetais com evidências de seu potencial *in vivo* na defesa vegetal contra diversos tipos de patógenos. O objetivo deste trabalho foi transformar o feijão-caupi com uma defensina de melão-de-são-caetano e testar o aumento da resistência das plantas transformadas a vírus e fungos fitopatogênicos. Após 18 eventos de bombardeamento com 1403 embriões, 16 plantas foram positivas para o marcador de seleção e 10 foram positivas tanto para o marcador de seleção, quanto para a inserção da defensina via PCR. As plantas PCR positivas foram analisadas pela técnica de Southern blot, confirmando a eficiência de 0,7% da transformação, corroborando com trabalhos realizados empregando metodologia semelhante. As plantas apresentaram um alto nível de expressão da defensina, mostrando o potencial do promotor CaMV35S por qPCR. Os ensaios com os vírus (CABMV e CPSMV) e com *Rhizoctonia solani* não mostraram nenhum incremento quanto à resistência aos patógenos das plantas transgênicas quando comparadas ao tipo selvagem. Por sua vez, os bioensaios com *Macrophomina phaseolina* demonstraram um aumento significativo da resistência ao fungo, comprovada pelo índice de germinação (85% controle não-inoculado e 67,5% plantas transgênicas) e ausência de sintomas nas plantas transformadas, quando comparadas ao tipo selvagem. A atividade antifúngica da defensina de *Momordica charantia* ratifica a principal atividade deste grupo de peptídeos antimicrobianos, bem como o potencial de desenvolvimento de uma linhagem superior de feijão-caupi resistente a esse patógeno.

Palavras-chave: AMP, PR-12, *Vigna unguiculata*, transgenia, cultura de tecidos.

1. Introdução

O feijão-caupi, também conhecido no Brasil como feijão macassar, fradinho ou feijão-de-corda (entre outros nomes populares), é uma leguminosa com centro de origem e diversificação no continente africano. Este continente figura como o principal produtor mundial, responsável por 64% da produção dos grãos, sendo o país de maior produção mundial a Nigéria (FAO, 2019). Fora do continente africano, o Brasil se destaca como sendo o segundo maior produtor mundial, tendo a região Nordeste como destaque, responsável por 85% da área de plantação e 53% da sua produção (CONAB, 2019).

As doenças são os principais fatores de redução da produtividade do feijão-caupi, provocando em algumas situações a devastação completa dos plantios. Destacam-se fungos e vírus, os quais respondem por 68,7% dos ataques de fitopatógenos nesta cultura (Obi e Barriuso-Vargas, 2014). Dentre a viroses, destaca-se o CABMV, pela ampla ocorrência no cultivo do feijão-caupi (Néya et al., 2015) e o CPSMV, pela severidade (Barros et al., 2013). Em relação aos fungos, destaca-se a *Macrophomina phaseolina*, que provoca lesões no caule, necrose e podridão, provocando um completo amarelecimento da planta, morte das sementes ou tombamento das plantas (Athayde Sobrinho, 2005). A frequência dessa doença é alta na cultura do feijão-caupi, pois o tipo de plantio favorece o desenvolvimento e proliferação do fungo, uma vez que a maior parte ocorre sob regime de sequeiro (Mohanapriya et al., 2017). O melhoramento genético convencional realizou alguns avanços, porém demanda muita mão-de-obra, recursos e tempo (Leão et al., 2016; Gomes-Silva et al., 2017). Dessa forma, a biotecnologia juntamente com a tecnologia do DNA recombinante representa uma solução eficaz e possibilita a introgressão de genes ausentes no banco de germoplasma (Fang et al., 2007).

As defensinas são peptídeos antimicrobianos classificados na superfamília PR-12 (pathogenesis-related 12), sendo relacionadas à imunidade inata dos vegetais. Quanto à sua atividade antimicrobiana, a ação fungicida tem sido a mais reportada, pois atua promovendo o aumento da permeabilidade da membrana celular dos fungos, o que provoca a necrose e interrupção do desenvolvimento do patógeno (Stotz *et al.*, 2009; Lacerda *et al.*, 2014). Porém há relatos de ações antivirais (Sher *et al.*, 2019), cujas defensinas se ligam as proteínas de revestimento do vírus promovendo a perda de atividade biológica viral (Wei *et al.*, 2019).

Há registros de diversas culturas transgênicas, como banana, mamão, batata, pimenta, tabaco, arroz, amendoim e melão, expressando defensinas. Dentre os resultados obtidos, o destaque foi a ação antifúngica contra uma série de fungos de espécies distintas, além de registros de outras atividades como antibacteriana, inseticida e antiviral (Wang *et al.*, 1999; Gao *et al.*, 2000; Brodgen, 2005; Anuradha *et al.*, 2008; Choi *et al.*, 2009; Ghag *et al.*, 2014). Em relação à transformação genética do feijão-caupi, experimentos foram iniciados com a obtenção de calos embrionários transformados, realizados por Garcia e colaboradores (1986). A partir deste marco até o presente momento, cerca de três dezenas de trabalhos foram publicados com a transformação da cultura supracitada (Citadin *et al.*, 2011). Dentre estes, destaca-se a obtenção do primeiro feijão-caupi transgênico via agrobactéria (*Agrobacterium tumefaciens*), o qual transmitiu o transgene para sua prole seguindo o padrão Mendeliano (Popelka *et al.*, 2006). Em relação à biobalística, um protocolo simples, rápido e eficiente, foi desenvolvido por Ivo *et al.* (2008), que foi posteriormente aplicado para obtenção de feijão-caupi tolerante ao herbicida Imazapyr (Citadin *et al.*, 2013) e resistente a vírus (Cruz e Aragão, 2014). O presente trabalho objetivou o desenvolvimento de plantas de feijão-caupi

transgênico superexpressando uma defensina de *Momordica charantia*, conhecida popularmente por Melão-de-São-Caetano, visando à resistência a fitopatógenos da cultura em tela.

2. Materiais e Métodos

2.1. Transformação genética vegetal via Biobalística

2.1.1. Preparo do meristema

Foram utilizados eixos embrionários de feijão-caupi cv. Boca Negra como explante. As sementes foram desinfestadas e embebidas em água destilada estéril por 16 a 18 h. A seguir, os embriões foram excisados dos cotilédones, e tiveram suas folhas primárias removidas para exposição do meristema (Aragão *et al.*, 1996). Essa cultivar foi escolhida por ter as folhas primordiais mais curtas, o que facilita a exposição do meristema (Aragão, pers. comm.), além de ser suscetível aos patógenos CABMV, CPSMV, *R. solani* e *M. phaseolina*.

2.1.2. Preparo das micropartículas

O preparo das micropartículas foi realizado segundo a metodologia de Rech e colaboradores (2008). Micropartículas de tungstênio (60 mg) foram adicionadas a 1 mL de álcool 70% e vigorosamente agitadas em vórtex por 15 min. Posteriormente, foram centrifugadas a 3.000 rpm por 5 min. O sobrenadante resultante da centrifugação foi descartado com cuidado para preservação do sedimento formado. Foi adicionado 1 mL de água destilada autoclavada para diluição, seguida de agitação no vórtex por 15 min, e então submetida à centrifugação nas mesmas condições anteriores. Houve novo descarte do sobrenadante e o passo foi repetido mais duas vezes. Por fim, as micropartículas foram ressuspensas em 1 mL de

Glicerol 50%. Todo procedimento foi realizado em câmara de fluxo laminar horizontal.

2.1.3. Construção do vetor binário e revestimento das micropartículas pelo DNA

O vetor pAHASMCDEF1 foi construído a partir da inserção do gene *Mcdef1* (defensina de *Momordica charantia*) nos sítios NcoI, presente na extremidade final do promotor 35S do CaMV, e SacI presente na extremidade inicial do Nos terminador, do vetor pBSK. Posteriormente, o cassete de expressão (promotor-*Mcdef1*-terminador) foi isolado através do sítio NotI, sendo inserido no vetor pAC321 (Rech *et al.*, 2008) formando o vetor final pAHASMCDEF1, que possui o gene *Atahas* como marcador de seleção, o qual promove tolerância ao herbicida sistêmico *Imazapyr* (Figura 1).

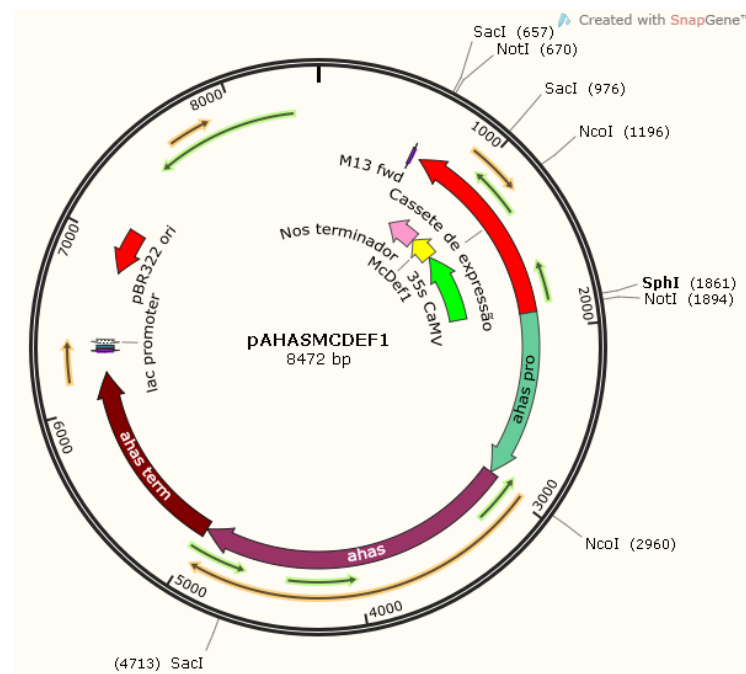


Figura 1: Vetor pAHASMCDEF1 utilizado na transformação vegetal, com destaque para os sítios de restrição NcoI, SacI e NotI, utilizados na construção do vetor, e SphI utilizado na digestão do DNA genômico para o Southern Blot.

O revestimento das micropartículas seguiu o protocolo de Rech *et al.* (2008). As partículas de tungstênio foram colocadas no sonicador por 10 min, seguido de agitação em vórtex durante 5 min. Posteriormente, a uma alíquota de 50 µL (retirada das micropartículas previamente purificadas) foram adicionadas 5 µg do vetor plasmidial, 50 µL de CaCl₂ a 100 mM e 20 µL de espermidina a 0,1 M. A mistura foi submetida ao vórtex por 10 min, e centrifugada a 6.000 rpm por 15 s. Após descarte do sobrenadante com o auxílio de micropipeta, foram acrescentados 150 µL de etanol absoluto ao pellet. O pellet foi ressuspendido com leves batidas no fundo do tubo, realizando-se uma centrifugação a 6.000 rpm por 15 s. Esta lavagem foi repetida mais duas vezes. Após o descarte do sobrenadante na última etapa da lavagem, foram adicionados 24 µL de etanol absoluto. Alíquotas de 3,2 µL foram distribuídas na região central de cada membrana de carregamento previamente posicionada em um suporte de membrana. Os discos das membranas carreadoras contendo as micropartículas revestidas com o DNA foram imediatamente transferidos para uma placa de Petri, que foi inserida num dessecador contendo sílica gel por 10 min. Após a evaporação do etanol, as micropartículas revestidas ficaram prontas para o bombardeamento.

2.1.4. Transformação das plantas

Eixos embrionários com meristema exposto foram dispostos em placa de Petri contendo meio de cultura apropriado para bombardeamento (Meio MS meia força). Os eixos foram colocados em posições equidistantes na placa formando um círculo com diâmetro de um centímetro com a região do meristema apical posicionada para o centro. Tanto as micropartículas revestidas com DNA, quanto os eixos foram colocados no gene gun de acordo com Ivo e colaboradores (2008).

2.1.5. Cultivo, seleção das plantas transgênicas e aclimação

Imediatamente após o bombardeamento, os eixos bombardeados foram transferidos para meio de cultura MS (Murashige e Skoog, 1962) de seleção e alongamento, contendo o fitormônio BAP (6-benzilaminopurina) (1 mg/L) e o herbicida Imazapyr (250 nM), capaz de se translocar sistemicamente até atingir o meristema apical da planta. Os exemplares foram mantidos a 28°C com fotoperíodo de 16 h. Após cerca de um mês, aquelas que passaram pela pressão de seleção e tiveram a formação das primeiras folhas, foram transferidas para meio com carvão ativado sem fitormônio, que além de realizar uma purificação dos vestígios do herbicida, promovem um melhor enraizamento pela ausência do BAP. Neste meio, as plântulas passaram de uma a duas semanas. As plantas enraizadas foram transferidas para copos plásticos descartáveis contendo vermiculita e terra (proporção 1:1), sendo cobertas por um saco plástico transparente. Após uma semana, esse saco foi furado diariamente para aclimação, pois desta forma a planta se adaptou à umidade externa, com posterior remoção do saco. Neste momento, foi retirada uma folha da planta presumivelmente transformada para análises moleculares.

2.2. Extração de DNA genômico

O DNA genômico foi extraído de folhas jovens de feijão-caupi cv. Boca Negra seguindo o método de Doyle e Doyle (1987). Aproximadamente 150 mg de tecido foliar de cada planta foi macerado em tubos de 1,5 mL com 400 µL do tampão CTAB de extração. A mistura foi incubada a 60°C por 20 min. Em seguida foram adicionados 400 µL de clorofórmio/álcool isoamílico (24:1), sendo a amostra centrifugada a 14.000 rpm a 25°C por 5 min. O sobrenadante foi resgatado e transferido para novo tubo de 1,5 mL, sendo a seguir adicionado igual volume de

isopropanol ao volume do material vegetal coletado (cerca de 200 μ L) a fim de precipitar o DNA. Então, foi realizada uma nova centrifugação a 14.000 rpm a 25°C por 10 min, sendo o sobrenadante descartado e o pellet lavado com 200 μ L de etanol 70% por 5 min. Após uma última centrifugação a 14.000 rpm a 25°C por 5 min, o álcool foi descartado e o DNA foi colocado para secar por 10 min no speed-vac. Finalmente o pellet foi ressuscitado em 50 μ L de TE, sendo o DNA estocado a 20°C negativos.

2.3. Amplificação por PCR da defensina

A reação de amplificação foi realizada com uso de 15 ng de DNA genômico, tampão de PCR 10x, 2 mM de $MgCl_2$, 0,6 mM do mix de dNTP, 2 mM de cada primer (Tabela I), 1 U de Taq DNA polymerase, sendo o volume final ajustado para 20 μ L. As amplificações tiveram como programa da PCR uma desnaturação inicial por 5 minutos a 95°C, seguida por 35 ciclos, incluindo 1 min de desnaturação a 95°C, 1 min de anelamento a 56°C, extensão a 72°C por 1 min e uma extensão final a 72°C por 7 min. Os produtos da amplificação foram separados por eletroforese em gel de agarose (1%), corados com GelRed e visualizados em transiluminador.

Tabela I: Pares de *primers* utilizados para amplificação de regiões do DNA exógeno inserido no genoma das plantas transformadas de feijão-caupi, assim como para PCR quantitativa em tempo real (RTqPCR).

Primers	Sequências (5'→3')	Fragmento esperado	Região amplificada
AHAS1F AHAS1R	CATTGCTTCTGCTATCGACACG CCTCGGGATTTGATTTTTGGTCCT	685 pb	Fragmento do promotor do gene <i>Atahas</i>
MCDEF1F MCDEF1R	ATCATCGCAAGACCGGCA CCACTATCCTTCGCAAGAC	405 pb	Gene completo <i>McDef1</i> com fragmento do promotor e terminador
MCDEF2F MCDEF2R	TCAAACGCAGTGCTTTGTGCA TGCTTGTA CTGATTGTTGGCACTG	197 pb	Fragmento do gene <i>Mcdef1</i>

2.4. Southern blotting

A extração do DNA genômico foi realizada seguindo o protocolo de Doyle e Doyle (1987) com pequenas modificações a partir de 3 g de folhas jovens de plantas com 30 dias após a semeadura. Um total de 30 µg de DNA de cada linhagem, além do controle negativo e vetor foram digeridos com a enzima de restrição SphI, que possui um único sítio de restrição no vetor (Figura 1). As amostras foram carregadas em gel de agarose 1,5% com corrida a 20 V overnight. As amostras de DNA digeridas foram transferidas do gel de agarose para membrana de nitrocelulose, sendo fixadas através de luz UV de alta intensidade com o uso de um *crosslinker* da marca Stratagene. A hibridização foi realizada com a sonda referente a uma região do promotor do DNA 35S que dirige a defensina, marcada radioativamente com $\alpha^{32}\text{P}$ dCTP (1.13×10^{14} Bq mol⁻¹) utilizando primers randômicos de acordo com o fabricante (Amersham Pharmacia Biotech). Os fragmentos de DNA foram visualizados e registrados em fotodocumentador (Biorad Gel Doc™).

2.5. Extração de RNA, síntese de cDNA e RT-qPCR

A extração do RNA das folhas do feijão-caupi transformado e tipo selvagem foi realizada em plantas com 30 dias após a semeadura em triplicata biológica. A integração e expressão do transgene foi testada em planta PCR positivas de quatro linhagens transgênicas na geração T₃. O método de extração utilizado é o do kit SV Total RNA Isolation System da Promega, conforme o protocolo do fabricante, seguido de tratamento com DNase I (*DNase Treatment and Removal* - Ambion), cuja qualidade e quantidade do RNA foram verificadas em eletroforese em gel de agarose 1,5% e em fluorímetro (Qubit), respectivamente. Alíquotas de 1 µg de RNA de cada amostra foram convertidas em cDNA utilizando o *Improm-II Reverse Transcription Systems* (Promega), seguindo as recomendações do fabricante. O par de *primers* para o gene *Mcdef1* (MCDEF2F e MCDEF2R) foi desenhado através do

programa SnapGene software (GSL Biotech; available at snapgene.com). As reações de RT-qPCR foram realizadas utilizando o equipamento CFX96 Touch™ Real-Time PCR Detection System (fabricante Bio-Rad), sob as seguintes condições: 95 °C por 2 min para desnaturação inicial, seguido de 40 ciclos de 95 °C por 15 s; 1 min à 60 °C para anelamento. A curva de *melting* foi gerada variando a temperatura de amplificação de 65 - 95°C. Todas as reações foram realizadas em triplicata (réplicas biológicas e técnicas) (Bustin et al., 2009). A eficiência de amplificação (E) foi determinada a partir de uma curva padrão gerada por diluições seriadas de cDNA (1/10, 1/100, 1/1000) para cada *primer*, em triplicata, usando a equação: $E = 10^{(-1/\text{slope of the standard curve})}$ (Rasmussen, 2001), cujo intervalo de eficiência foi de 95% a 105%. A quantificação relativa foi realizada pelos valores de *fold change* utilizando como genes de referência o UBQ10 (Ubiquitina), Beta-Tubulina e o F-BOX (dados não publicados) analisada pelo software REST (Relative Expression Software Tool) versão 2.0.13 (Pfaffl et al., 2002). Foi realizado um teste não paramétrico (ANOVA) para verificar se existem diferenças no perfil de expressão do gene codificante da defensina entre as quatro linhagens. Os testes de ANOVA foram seguidos pelo "post-hoc" de Bonferroni. A significância estatística, para todos os testes, foi estabelecida para $p < 0,05$.

2.6. Bioensaios com CABMV e CPSMV

Os bioensaios com os vírus CABMV e CPSMV foram realizados a partir da inoculação mecânica dos vírus em plantas positivas (transformadas) de forma individualizada. Plantas com cerca de 10 dias após a semeadura, foram previamente injuriadas com o abrasivo carborundum de modo a facilitar a penetração do vírus. O inóculo foi preparado a partir de folhas sintomáticas de plantas previamente infectadas, sendo maceradas com pistilo em almofariz com tampão fosfato de sódio 0,01M pH 7,0, na proporção de 1,0 g de tecido foliar para 9,0 mL de tampão. A inoculação dos folíolos foi realizada com o pistilo e ocorreu no final da tarde, período

de menor temperatura na casa de vegetação (25° a 28°C) (Oliveira *et al.*, 2012). Foram utilizados de cinco a sete vasos com três plantas cada, com pelo menos 15 plantas transgênicas de cada linhagem inoculada para cada vírus. Paralelamente, foram inoculadas seis plantas-controle (não transformadas, tipo selvagem) distribuídas em dois vasos, assim como um vaso com plantas-controle e um vaso de plantas transformadas como testemunhas sem inoculação do patógeno.

2.7. Bioensaio com *Rhizoctonia solani*

Os isolados de *R. solani* foram cultivados em meio BDA (Batata Dextrose Ágar) com cloranfenicol (0,25 g/L) a 28°C por 5 dias. Posteriormente, foram retirados discos de micélios com 0,5 cm³, que foram colocados juntamente com as sementes de plantas transformadas em covas de sacos contendo areia lavada previamente autoclavada, seguindo o protocolo de Lima *et al.* (2013). Foram distribuídas quatro covas em cada saco, sendo utilizados cinco sacos para cada linhagem de plantas transgênicas. Como testemunha, foi realizada uma semeadura de plantas-controle (tipo selvagem) em solo estéril e outra em solo contendo o inóculo. As plantas foram analisadas diariamente até completarem 30 dias após a semeadura.

2.8. Bioensaio com *Macrophomina phaseolina*

Os isolados de *M. phaseolina* foram cultivados em meio BDA (Batata Dextrose Ágar) com cloranfenicol (0,25 g/L) a 28°C por cinco dias, juntamente com sementes de sorgo, seguindo o protocolo de Lima *et al.* (2013) com modificações. Posteriormente, um grão de sorgo repleto de micélios de *M. phaseolina* foi colocado em cada cova contendo areia lavada previamente autoclavada. Em cada saco foram semeadas quatro sementes de plantas transformadas. Foram utilizados cinco sacos por linhagem transgênica. Plantas-controle (tipo selvagem) foram semeadas em solo

contendo o patógeno, assim como em solo estéril. As plantas foram analisadas diariamente até completarem 30 dias após a semeadura.

3. RESULTADOS

3.1. Transformação de feijão-caupi via biobalística

Foram realizados 18 eventos de bombardeamento com o vetor pAHASMCDEF1 (Figura 1), totalizando 1403 embriões bombardeados, dos quais 289 originaram plantas possivelmente transformadas. Foi realizada a extração de DNA das 289 plantas pelo método de Doyle e Doyle (1987), bem como a PCR com primers específicos para o gene de interesse (defensina *Mcdef1*) e o gene marcador de seleção (*Ahas*).

Das 289 plantas, 16 foram PCR positivas para o marcador de seleção, sendo 10 positivas também para a presença da sequência do gene que codifica a defensina (Figura 2). A taxa de escape foi cerca de 94%, possivelmente devido à qualidade, fabricante e pureza do Imazapyr comercial utilizado, sendo essa taxa reduzida em experimentos posteriores utilizando Imazapyr de uso laboratorial.

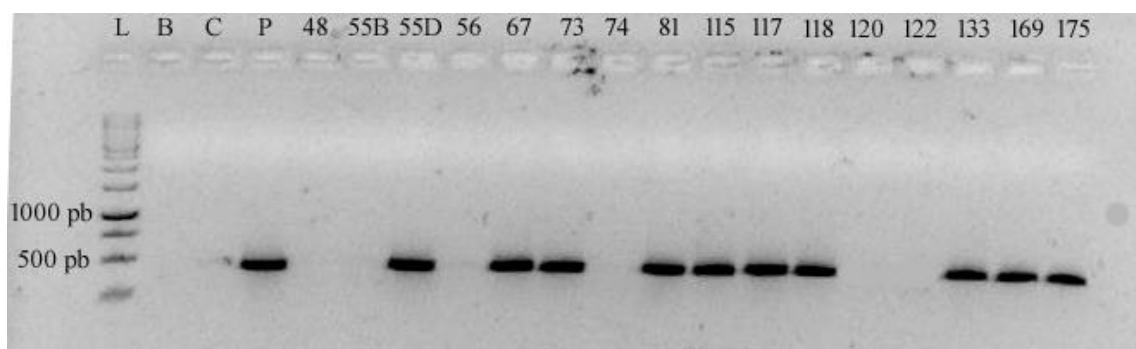


Figura 2: Gel de agarose 1,2% com os produtos da PCR com primers específicos para o gene *Mcdef1* que amplifica um fragmente de 405 pb. L – Ladder 1 Kb; B – Branco (sem DNA); C: Controle negativo (planta não transformada); P: Controle positivo correspondente ao plasmídeo pAHASMCDEF1.

3.2. Análise da progênie

A avaliação da transmissão do transgene para as próximas gerações foi realizada através de PCR direcionada ao gene *Mcdef1*. Dentre as 10 plantas positivas para os dois transgenes (*Atahas* e *Mcdef1*), apenas sete os transmitiram para a progênie T_1 , porém apenas as linhagens 122 e 133 seguiram o padrão mendeliano com índice de confiabilidade de 95% para o padrão 3:1 correspondendo à inserção de um locus. Por sua vez, a linhagem 67 seguiu o padrão 15:1 para inserção de dois loci (Tabela II), enquanto as demais não apresentaram herança Mendeliana possivelmente devido ao fato de que da geração T_0 apresentar características quiméricas, requerendo mais análises da progênie na etapa de avanço da geração T_1 para T_2 .

Tabela II. Tabela com dados da progênie das plantas transformantes no avanço da geração T_0 para a T_1 , com os valores do teste do qui-quadrado (X^2) com índice de confiabilidade de 95%. Para a linhagem 67 foi utilizado no teste a proporção 15:1 e nas demais 3:1. Para seguir o padrão Mendeliano o valor do teste tem que ser menor que 5%. T_1+ representa os indivíduos testados positivos para PCR com os primers que amplificam a defensina, sendo plantas transformadas, e T_1- representam as plantas testadas negativas para PCR sendo não transformadas.

T_0	T_1+	T_1-	Teste X^2	P
67	25	1	0,78	61,26%
73	11	16	16,9	0%
115	10	14	14,2	0,02%
122	17	9	1,22	25,75%
133	18	8	0,46	49,69%
169	9	20	29	0%
175	11	16	16,9	0%

Após avanço da geração T_0 para T_1 , foi realizado o avanço da geração T_1 para a T_2 , tanto para obtenção de linhagens homozigotas, quanto para seleção de plantas positivas para bioensaios com inoculação do patógeno. Os resultados apresentados na tabela sugerem que os indivíduos 67.8, 115.4 e 133.2 são

homozigotos. As linhagens para as quais não foram obtidas plantas homozigotas (73, 122, 169 e 175), continuarão avançando gerações, para a obtenção de plantas homozigotas para o gene *Mcdef1* (Tabela III).

Tabela III. Tabela representativa da análise da progênie das plantas transformantes no avanço de uma geração (T1 → T2), com os valores do teste do χ^2 e índice de confiabilidade de 95% com correção de Yates, observando-se o padrão de herança Mendeliano de 3:1 para todos os tratamentos. Cada linhagem (67, 73, 115, 122, 133 e 169) segue de um ponto e o indivíduo T₁ que avançou a geração para a T₂.

T1	T2+	T2-	Teste X2	P
67.1	3	1	0,33	100,00%
67.2	3	1	0,33	100,00%
67.3	19	4	0,97	39,94%
67.4	11	7	1,56	17,36%
67.5	6	0	2,89	15,73%
67.6	13	1	2,95	12,28%
67.7	13	1	2,95	12,28%
67.8	15	0	5,76	2,53%

T1	T2+	T2-	Teste X2	P
73.1	1	0	2,33	56,37%
73.2	7	5	1,44	18,24%
73.3	3	9	14,78	0,01%
73.4	12	13	9,05	0,18%
73.5	15	11	3,74	4,15%
73.6	9	15	17,06	0,00%

T1	T2+	T2-	Teste X2	P
115.1	4	1	0,47	79,63%
115.2	7	2	0,26	84,74%
115.3	1	0	2,33	56,37%
115.4	13	0	5,1	3,74%
115.5	5	10	12,87	0,02%
115.6	14	3	0,76	48,38%
115.7	3	2	0,47	43,86%

T1	T2+	T2-	Teste X2	P
122.1	2	3	2,6	7,07%
122.2	3	5	5,17	1,43%

T1	T2+	T2-	Teste X2	P
133.1	7	1	1,17	41,42%
133.2	11	0	4,45	5,55%
133.3	5	5	2,8	6,79%
133.4	1	5	9,56	0,10%
133.5	1	1	0,67	41,42%
133.6	1	1	0,67	41,42%

T1	T2+	T2-	Teste X2	P
169.1	9	3	0,11	100,00%
169.2	9	3	0,11	100,00%
169.3	3	8	12,21	0,03%
169.4	9	4	0,18	63,10%
169.5	2	8	14,8	0,01%
169.6	1	3	4,33	2,09%
169.7	1	5	9,55	0,10%
169.8	7	5	1,44	18,24%

T1	T2+	T2-	Teste X2	P
175.1	8	8	4,75	2,09%

3.3. Southern Blotting

Foram selecionadas cinco plantas que transmitiram o transgene para a progênie, sendo testadas novamente para a presença do mesmo pela técnica de Southern Blotting, cuja sonda utilizada e marcada radiotivamente com P₃₂ foi uma sequência do promotor CaMV35S que dirige a defensina *McDef1*. As linhagens testadas foram marcadas com a sonda, sendo reconfirmadas como positivas para o transgene (Figura 3).

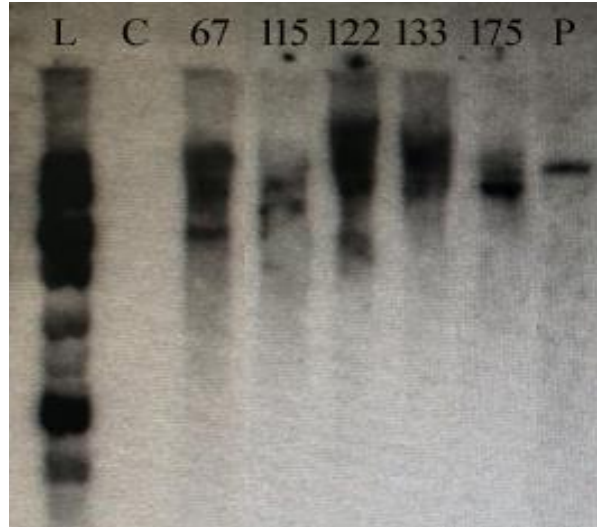


Figura 3: Padrão da hibridização via Southern Blotting das sete linhagens que transmitiram o transgene para a progênie. L – Ladder 1 Kb; C: Controle negativo (tipo selvagem, planta não transformada); Linhagens transformadas 67, 115, 122, 133 e 175; P: Plasmídeo pAHASMCDEF1 (Controle positivo).

3.4. Expressão Gênica da defensina

A análise do RT-qPCR mostrou diferentes padrões de expressão do *gene Mcdef1* entre as linhagens transformadas. Três linhagens apresentaram padrão de expressão bastante elevado (linhagens 67, 73 e 175), e a linhagem 115 apresentou um padrão de expressão inferior, quando comparada às demais linhagens (Figura 4). Todas as quatro linhagens tiveram perfil de expressão estatisticamente distintos.

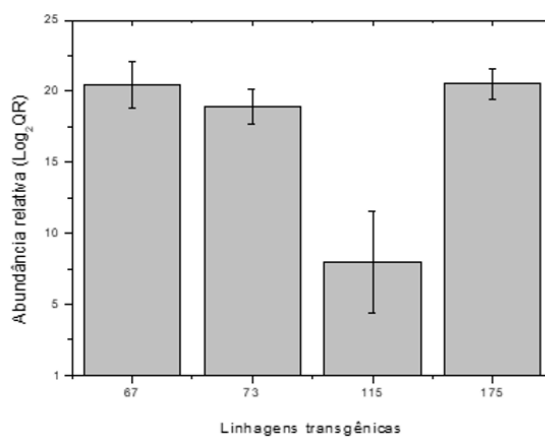


Figura 4: Análise da abundância relativa do transcrito do gene *McDef1* via RT-qPCR em quatro linhagens de feijão-caupi transgênico (67, 73, 115 e 175) na geração T3.

3.5. Bioensaios com os vírus CABMV e CPSMV

Após a inoculação dos vírus isoladamente em cada linhagem de plantas transgênicas 10 dias após a semeadura, as plantas foram monitoradas diariamente para verificação da evolução dos sintomas. Em relação ao potyvirus CABMV, entre duas a três semanas todas as plantas inoculadas morreram (Figura 5), sendo que os sintomas iniciais apareceram por volta de uma semana após inoculação (Figura 6). Não houve diferença entre as plantas transformadas e tipo selvagem inoculadas.

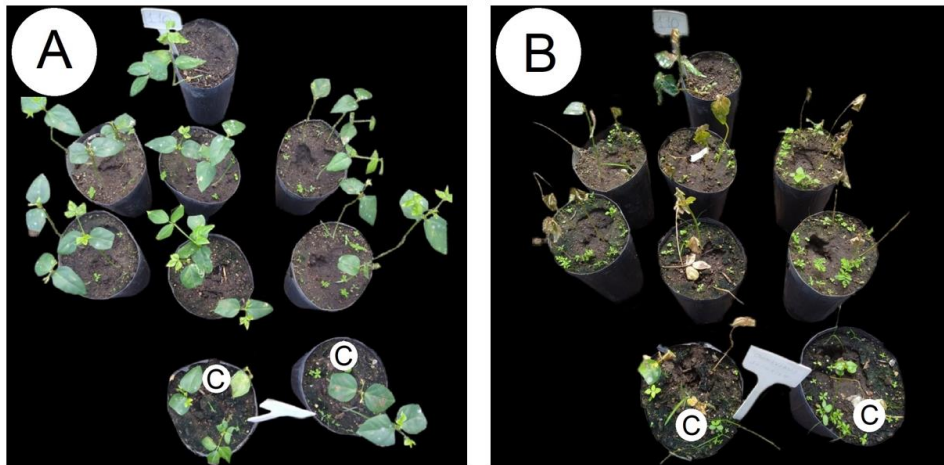


Figura 5: Imagem representativa do experimento de inoculação do Potyvirus CABMV. Na imagem A, as plantas com uma semana após inoculação do vírus, na imagem B, duas semanas após inoculação. As letras C no interior dos vasos, mostram as plantas-controle (não-transformadas, tipo selvagem) inoculadas.

Em relação ao vírus do mosaico severo do caupi (CPSMV), os resultados foram semelhantes, porém o desenvolvimento dos sintomas, bem como a morte das plantas inoculadas ocorreu significativamente mais cedo comparativamente ao potyvirus, sendo uma semana a menos o tempo para provocar a letalidade das plantas submetidas à inoculação (Figura 6). Não houve diferença entre as plantas transformadas e tipo selvagem inoculadas.

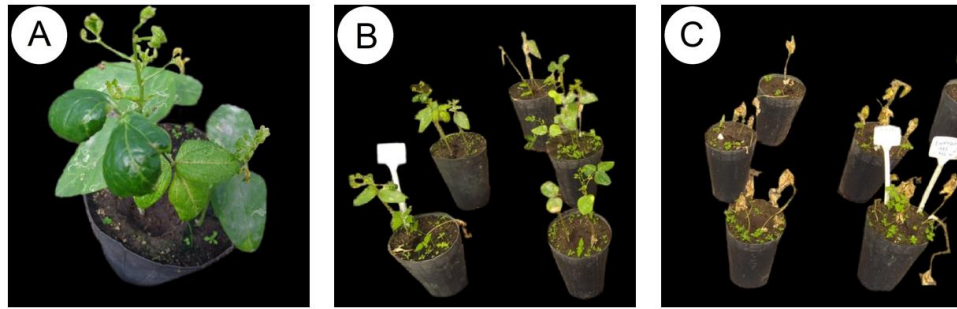


Figura 6: Imagem representativa do experimento de inoculação do Potyvirus CABMV e CPSMV. Na imagem A, as plantas com 17 dias de idade, após sete dias da inoculação, com os sintomas do CABMV, como mosaico intenso do limbo foliar e distorção das folhas. Na imagem B e C, tem um paralelo de uma mesma linhagem de plantas transformadas inoculadas com o CABMV (B) e com o CPSMV (C) na mesma data, evidenciando a velocidade e severidade do vírus do mosaico severo, que provoca a morte precoce comparativamente ao potyvirus.

3.6. Bioensaio com *Rhizoctonia solani*

Em relação aos ensaios com *R. solani*, os resultados não foram satisfatórios, pois o índice de germinação das plantas inoculadas foi muito baixo e as poucas plantas que germinaram (18,75%), com menos de duas semanas tiveram necrose do caule e morreram (Figura 7).

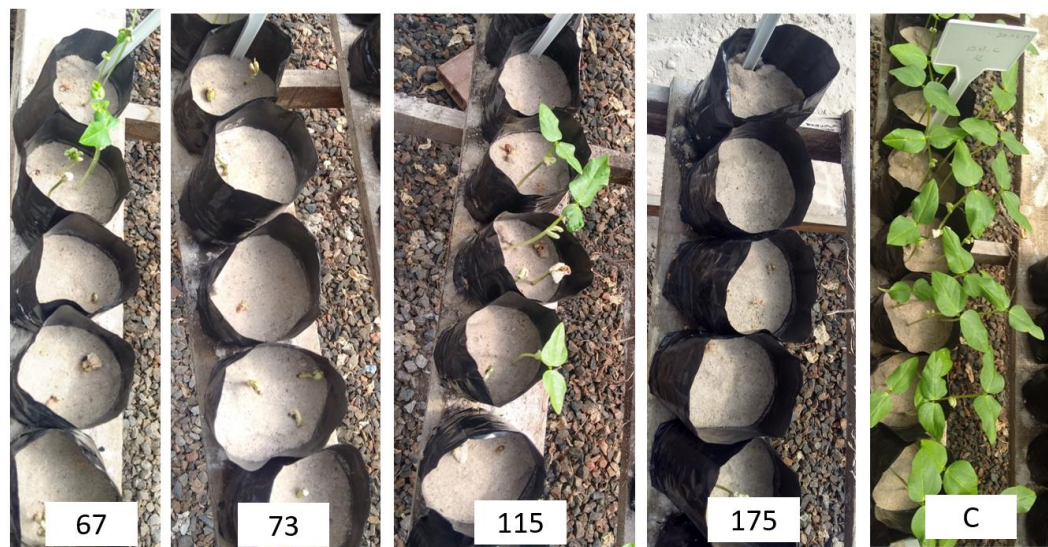


Figura 7: Imagem representativa do experimento de inoculação do fungo *R. solani* mostrando quatro linhagens transformadas (67, 73, 115 e 175), além da planta controle não inoculada (C). A imagem foi realizada cinco dias após a semeadura com inoculação nas linhagens transformadas e sem inoculação (C).

3.7. Bioensaio com *Macrophomina phaseolina*

Os bioensaios com *M. phaseolina* foram realizados paralelamente com *R. solani*. O índice de germinação foi satisfatório comparativamente (67,5%) ao controle não inoculado (85%) (Figura 8). Além disso, os sintomas mancha marrom e necrose um pouco acima da base do caule, com posterior morte da planta, só foram visualizados em 30% das plantas transformadas e inoculadas (Figura 9). A partir deste *screening*, há um forte indicativo que pelo menos parte das plantas transformadas com a defensina tenham resistência / imunidade à *M. phaseolina*. Vale ressaltar que a linhagem 115 foi a que teve menor índice de germinação (50%), comparativamente com as linhagens 67 (75%), 73 (60%) e 175 (85%), e também foi a que teve o índice de expressão da defensina 50% inferior ao das demais linhagens.



Figura 8: Imagem representativa do experimento de inoculação do fungo *M. phaseolina* mostrando quatro linhagens transformadas (67, 73, 115 e 175), além da planta controle não inoculada (C). A imagem foi realizada cinco dias após a semeadura com inoculação nas linhagens transformadas e sem inoculação (C).

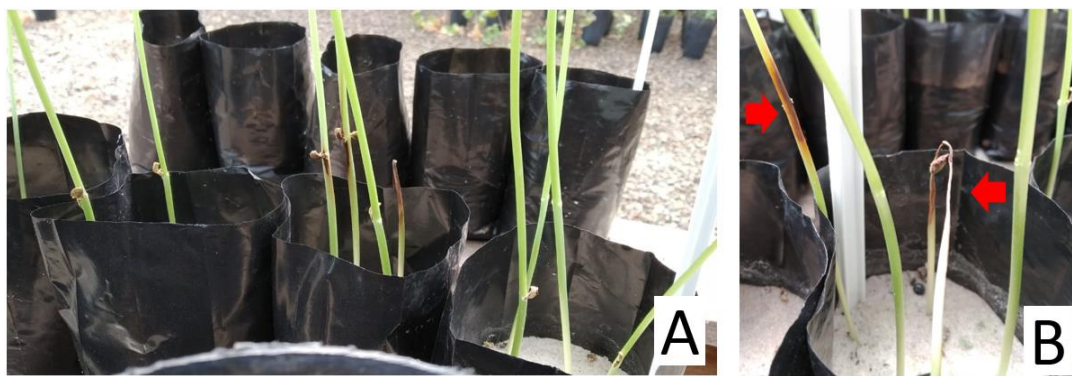


Figura 9: Imagem representativa do experimento de inoculação do fungo *M. phaseolina* mostrando os sintomas das plantas que foram infectadas. Pela imagem A é perceptível que a maior parte das plantas transformadas com a defensina não foi infectada. Na imagem B, verifica-se com maior proximidade os sintomas (setas vermelhas), como escurecimento do caule, seguido de necrose, tombamento e morte das plantas.

4. DISCUSSÃO

O processo de transformação genética de feijão-caupi não é trivial, caracterizado pela sua recalcitrância na cultura de tecidos *in vitro* (Citadin *et al.*, 2011). No presente trabalho, a eficiência da transformação considerando-se apenas o transgene *Atahas* foi de 1,14%. Por sua vez, quando se considera tanto o *Atahas* quanto a defensina a eficiência foi de 0,71%. Portanto, nossos resultados foram comparáveis aos de trabalhos prévios com feijão-caupi utilizando metodologia semelhante, como no caso de Ivo *et al.* (2008), no qual houve eficiência de 0,90% e do estudo de Cruz e Aragão (2014) com 1,04%.

Além da limitação na eficiência, a transformação genética também é afetada pela geração de quimeras num primeiro momento, além de muitas vezes não haver transmissão do transgene para a progênie. Além disso, havendo transmissão essa pode não seguir um padrão mendeliano, como já relatado para feijão-caupi (Muthukumar *et al.*, 1996; Kononowicz *et al.*, 1996; Ikea *et al.*, 2003; Raveendar e Ignacimuthu, 2010; Okeyo-Ikawa *et al.*, 2016).

O nível de expressão gênica comprova o sucesso da transformação e a atividade do promotor constitutivo CaMV35S, porém a variação dos níveis de expressão gênica é comum em eventos transgênicos distintos, dada a aleatoriedade da inserção do transgene (Eissa *et al.*, 2017). Inclusive os transgenes podem sofrer silenciamento ao longo das gerações devido à metilação do promotor (Howarth *et al.*, 2005; Melander *et al.*, 2006; Eissa *et al.*, 2017).

Em relação aos bioensaios com os dois principais vírus que acometem a cultura do feijão-caupi (CPSMV e CABMV), os resultados preliminares (T1) indicam que não houve nenhuma alteração da suscetibilidade / resistência aos patógenos. A maioria das defensinas tem como principal ação a antifúngica, embora existam registros de atividade antibacterianas e inseticidas com raros registros de atividade antiviral (Khan *et al.*, 2019; Sathoff e Samac, 2018). As defensinas apresentam também atividade imunomoduladora, induzindo cascatas de sinalização e resposta hormonal, além da expressão de outras famílias PR (Piereck *et al.*, in prep.). Desta forma, a ideia de realizar esse bioensaio, foi devido ao fato de que defensinas são agentes indutores da imunidade inata em vegetais, consequentemente a sua superexpressão poderia promover um aumento dos mecanismos de defesa globais nas plantas transformantes, promovendo um amplo espectro de ação antimicrobiana (Khan *et al.*, 2019), o que aparentemente não aconteceu.

O presente trabalho demonstra o potencial antifúngico da defensina *McDef1* de *M. charantia*, uma vez que mais de 90% das publicações de plantas transgênicas superexpressando defensinas apresentam aumento da resistência ou imunidade a fungos (Khan *et al.*, 2019). Apenas três trabalhos mostram defensinas com atividade inseticida (Choi *et al.*, 2009; Thao *et al.*, 2017; Vi *et al.*, 2019) e dois com atividade antibacteriana (Chen *et al.*, 2006; Sjahril *et al.*, 2006). Em relação às plantas

transformadas com defensinas com atividades antifúngicas, a maior parte dos trabalhos demonstram ações contra espécies dos gêneros *Phytophthora* (Zhu *et al.*, 2007; Anuradha *et al.*, 2008; Lee *et al.*, 2018) e *Fusarium* (Ghag *et al.*, 2012; Gaspar *et al.*, 2014; Ghag *et al.*, 2014), assim como contra *Verticillium dahliae* (Gao *et al.*, 2000; Gaspar *et al.*, 2014). Porém, até o presente momento não há relatos sobre plantas transformadas superexpressando uma defensina com atividade contra *M. phaseolina*. Vale destacar que a infecção fúngica supracitada também provoca danos consideráveis em outras culturas, acometendo mais de 500 plantas hospedeiras, sendo que as leguminosas tem uma maior suscetibilidade (Lodha e Mawar, 2019). Especialmente na soja esse fungo dificulta o sistema de plantio direto, decorrendo em um aumento de custos devido ao controle químico da doença (Khan *et al.*, 2019). Portanto, este mesmo vetor pode ser utilizado para obtenção de outras culturas transgênica expressando a defensina de *M. charantia* através da mesma metodologia.

5. CONCLUSÃO

As plantas transgênicas superexpressam uma defensina de *M. charantia* que se mostrou eficaz contra fungo da espécie *M. phaseolina*, corroborando com a principal atividade destas moléculas. Esses achados mostram o potencial do desenvolvimento de uma linhagem elite de feijão-caupi resistente a esse patógeno de ampla distribuição, bem como a possibilidade do uso biotecnológico do *Mcdef1* em outras culturas comprometidas pela podridão cinzenta do caule.

6. REFERÊNCIAS BIBLIOGRÁFICAS

Anuradha TS, Divya EK, Jami ESK, Kirti EPB (2008) Transgenic tobacco and peanut plants expressing a mustard defensin show resistance to fungal pathogens. *Plant Cell Rep* 27: 1777–1786.

- Aragão FJL *et al.* (1996) Inheritance of foreign genes in transgenic bean (*Phaseolus vulgaris* L.) co-transformed via particle bombardment. *Theor Appl Genet* 93: 142–150.
- Athayde Sobrinho C (2005) Patossistema caupi X *Macrophomina phaseolina*: método de detecção em sementes, esporulação e controle do patógeno. Tese (Doutorado em Agronomia) - Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, 150 f.
- Brogden KA (2005) Antimicrobial Peptides: Pore Formers Or Metabolic Inhibitors In Bacteria? *Nat Rev Microbiol* 3: 238–250.
- Bustin SA *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin chemistry* 55(4): 611-622.
- Chen SC, Liu AR, Zou ZR (2006) Overexpression of glucanase gene and defensin gene in transgenic tomato enhances resistance to *Ralstonia solanacearum*. *Russia J Plant Physiol* 53: 756–763.
- Choi MS, Kim YH, Park HM, Seo BY, Jung JK, Kim ST, Kim MC, Shin DB, Yun HT, Choi IS, Kim C, Lee JY (2009) Expression of BrD1, a Plant Defensin from *Brassica rapa*, Confers Resistance against Brown Planthopper (*Nilaparvata lugens*) in Transgenic Rices. *Mol Cells* 28:131-137.
- Citadin CT, Ibrahim AB, Aragão FJL (2011) Genetic engineering in cowpea (*Vigna unguiculata*): History, status and prospects. *GM Crops* 2(3): 144-149.
- Citadin CT, Cruz ARR, Aragão FJL (2013) Development of transgenic imazapyr-tolerant cowpea (*Vigna unguiculata*). *Plant Cell Rep* 3: 537–543.
- CONAB. Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira. V. 6 - SAFRA 2018/19 - N. 11 - Décimo primeiro levantamento | AGOSTO 2019. ISSN 2318-6852 Acompanhamento da Safra Brasileira de Grãos, v. 11 Safra 2018/19 - Décimo Primeiro levantamento, Brasília, p. 1-45, agosto 2019.
- Cruz ARR and Aragão FJL (2014) RNAi-based enhanced resistance to Cowpea Severe Mosaic Virus and Cowpea Aphid-Borne Mosaic Virus in transgenic cowpea. *Plant Pathol* 63: 831–837.
- Doyle JJ and Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.
- Eissa HF, Hassanien SE, Ramadan AM, El-Shamy MM, Saleh OM, Shokry AM, *et al.* (2017) Developing transgenic wheat to encounter rusts and powdery mildew by overexpressing barley chi26 gene for fungal resistance. *Plant Methods* 13(1).
- Fang J, Chao CT, Roberts PA, Ehlers JD (2007) Genetic diversity of cowpea [*Vigna unguiculata* (L.) Walp.] in four West African and USA breeding programs as determined by AFLP analysis. *Genet Resour Crop Evol* 54: 1197–1209.
- Food and Agriculture Organization, Statistics Division (FAOSTAT), <http://faostat3.fao.org/browse/Q/QC/E>. (Julho 20, 2019).
- Gao Ai-Guo, Hakimi SM, Mittanck CA, Wu Y, Woerner BM, Stark DM, Shah DM, Liang J, Rommenset CMT (2000) Fungal pathogen protection in potato by expression of a plant defensin peptide. *Nat Biotechnol* 18: 1307–1310.

- Garcia JA, Hille J, Goldbach R (1986) Transformation of cowpea *Vigna unguiculata* cells with an antibiotic resistance gene using a Ti-plasmid-derived vector. *Plant Sci* 44: 37-46.
- Gaspar YM, Mckenna JÁ, McGinness BS, Hinch J, Poon S, Connelly AA, Anderson MA, Heath RL (2014) Field resistance to *Fusarium oxysporum* and *Verticillium* in transgenic cotton expressing the plant defensin NaD1. *J Exp Bot* 65(6): 1541–1550.
- Ghag SB, Shekhawat UKS, Ganapathi TR (2012) Petunia Floral Defensins with Unique Prodomains as Novel Candidates for Development of *Fusarium* Wilt Resistance in Transgenic Banana Plants. *PLoS ONE*, 7: e39557.
- Ghag SB, Shekhawat UKS, Ganapathi TR (2014) Transgenic banana plants expressing a *Stellaria media* defensin gene (Sm-AMP-D1) demonstrate improved resistance to *Fusarium oxysporum*. *Plant Cell Tiss Organ Cult* 119: 247–255.
- Gomes-Silva F, Almeida CMA, Silva AG, Leão MPC, Silva KP, Oliveira LG, Silva MV, Costa AF, Lima VLM (2017) Genetic Diversity of Isolates of *Macrophomina* phase. *J Agric Sci* 9(11).
- Howarth JR, Jacquet JN, Doherty A, Jones HD, Cannell ME (2005) Molecular genetic analysis of silencing in two lines of *Triticum aestivum* transformed with the reporter gene construct pAHC25. *Ann Appl Biol* 146(3): 311–320.
- Ikea J, Ingelbrecht I, Uwaifo A, Thotttappilly G (2003) Stable gene transformation in cowpea (*Vigna unguiculata* L. Walp.) using particle gun method. *Afr J Biotechnol* 2: 211-218.
- Ivo NL, Nascimento CP, Vieira LS, Campos FA, Aragão FJ (2008) Biolistic-mediated genetic transformation of cowpea (*Vigna unguiculata*) and stable Mendelian inheritance of transgenes. *Plant Cell Rep* 27(9):1475-83.
- Khan MMM, Firdous H, Tahir Z, Zahid K, Haider MA, Qadus A, Sultan A, Hussan MU, Mushtaq MS (2019) Response of commercial and candidate lines of soybean (*Glycine max* L.) against root rot disease and its chemicals management. *Discovery Agriculture* 5: 29-40.
- Khan RS, Iqbal A, Malak R, Shehryar K, Attia S, Ahmed T, Khan MA, Arif M, Mii M. (2019) Plant defensins: types, mechanism of action and prospects of genetic engineering for enhanced disease resistance in plants. *Biotech* 9(5).
- Kononowicz AK, Cheah KT, Narasimhan ML, Murdock LL, Shade RE, Chrispeels MJ, Filippone E, Monti LM, Bressan RA, Hasegawa PM (1997) Development of transformation system for cowpea (*Vigna unguiculata* L. Walp). *Adv Cowpea Res* 1: 361-371.
- Lacerda AF, Vasconcelos EAR, Pelegrini PB, Sa MFG (2014) Antifungal defensins and their role in plant defense. *Front Microbiol* 5:116.
- Leão US, Eiras M, Freire Filho FR, Nogueira MSR, Rodrigues LK, Ribeiro VQ, Lopes ACA (2016) Selection of phenotypic traits and resistance to Cowpea Severe Mosaic Virus and Cowpea Aphid-Borne Mosaic Virus in cowpea [*Vigna unguiculata* (L.) Walp.] seeds with rugose white coat. *Aust J Crop Sci* 10(4):470-481.

- Lee HH, Kim JS, Hoang QTN, Kim JI, Kim YS (2018) Root-specific expression of defensin in transgenic tobacco results in enhanced resistance against *Phytophthora parasitica* var. *nicotianae*. *Eur J Plant Pathol* 151(3): 811-823.
- Lima LRL, Noronha MA, Silva KJDE, Lopes ACA, Rocha MM (2013) Influência de metodologia de inoculação e da concentração de inóculo de *Macrophomina phaseolina* no desenvolvimento da podridão-cinzenta-do-caule em feijão-caupi. In: III Congresso Nacional de Feijão-caupi, 2013, Recife.
- Lodha S and Mawar R (2019) Population dynamics of *Macrophomina phaseolina* in relation to disease management: A review. *J Phytopathol* 00: 1-17.
- Melander M, Kamnert I, Happstadius I, Liljeroth E, Bryngelsson T (2006) Stability of transgene integration and expression in subsequent generations of doubled haploid oilseed rape transformed with chitinase and β -1,3-glucanase genes in a double-gene construct. *Plant Cell Rep* 25(9): 942–952.
- Mohanapriya R, Naveenkumar R, Balabaskar P (2017) Survey, Virulence and Pathogenicity of Root Rot Incidence of Cowpea in Selected Districts of Tamilnadu caused by *Macrophomina phaseolina* (Tassi.) Goid. *Int J Curr Microbiol and App Sci* 6 (3): 694-705.
- Muthukumar B, Mariamma M, Veluthambi K, Gnanam A (1996) Genetic transformation of cotyledon explants of cowpea (*Vigna unguiculata* L. Walp) using *Agrobacterium tumefaciens*. *Plant Cell Rep* 15: 980-985.
- Obi VI and Barriuso-Vargas JJ (2014) Situation of Biofungicides Reconnaissance, a Case of Anthracnose Disease of Cowpea. *American J Plant Sci* 5: 1202-1211.
- Okeyo-Ikawa R, Amugune NO, Njoroge NC, Asami P, Holton T (2016) In planta seed transformation of Kenyan cowpeas (*Vigna unguiculata*) with P5CS gene via *Agrobacterium tumefaciens*. *J Agric Biotech Sustain Dev* 8(5): 32-45.
- Oliveira CRR, Freire-Filho FR, Nogueira MSR, Barros GB, Eiras M, Ribeiro VQ, Lopes ACA (2012) Reação de genótipos de feijão-caupi revela resistência às coinfeções pelo Cucumber mosaic virus, Cowpea aphid-borne mosaic virus e Cowpea severe mosaic virus. *Bragantia. Campinas* 71(1): 59-66.
- Pfaffl MW, Horgan GW and Dempfle L (2002) Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucl Acids Res* 30:1-10.
- Popelka JC, Gollasch S, Moore A, Molvig L, Higgins TJV (2006) Genetic transformation of cowpea (*Vigna unguiculata* L.) and stable transmission of the transgenes to progeny. *Plant Cell Rep* 25: 304-312.
- Rasmussen, R (2001) Quantification on the LightCycler. Book: Rapid Cycle Real-Time PCR: Methods and Applications 21-34.
- Raveendar S and Ignacimuthu S (2010) Improved *Agrobacterium* mediated transformation in cowpea *Vigna unguiculata* L. Walp. *Asian J Plant Sci* 9: 256-263.
- Rech EL, Vianna GR, Aragão FJ (2008) High-efficiency transformation by biolistics of soybean, common bean and cotton transgenic plants. *Nat Protocol* 3: 410-8.
- Sathoff AE and Samac D (2019) Antibacterial Activity of Plant Defensins. *Mol Plant-Microbe Int* 32(5): 507-514.

- Sher Khan R, Iqbal A, Malak R, Shehryar K, Attia S, Ahmed T, Ali Khan M, Arif M, Mii Mv (2019) Plant defensins: types, mechanism of action and prospects of genetic engineering for enhanced disease resistance in plants. *Biotech* 9(5): 192.
- Sjahril R, Chin DP, Khan RS, Yamamura S, Nakamura I, Amemiva Y (2006) Transgenic *Phalaenopsis* plants with resistance to *Erwinia carotovora* produced by introducing wasabi defensin gene using *Agrobacterium* method. *Plant Biotech* 23: 191-194.
- Stotz HU, Thomson JG, Wang Y (2009) Plant defensins Defense, development and application. *Plant Signali Behav* 4(11): 1010-1012.
- Thao HT, Lan NTN, Tuong HM, Thanh NVT, Son LV, Mau CH (2017) Expression analysis of recombinant *Vigna radiata* plant defensin 1 protein in transgenic tobacco plants. *J App Biol Biotech* 5(1): 70-75.
- Vi TXT, Nguyen TNL, Pham TTN, Nguyen HQ, Nguyen THY, Tu QT, Le VS, Chu HM (2019) Overexpression of the ZmDEF1 gene increases the resistance to weevil larvae in transgenic maize seeds. *Mol Biol Rep* 46: 2177.
- Wang YP, Nowak G, Culley D, Hadwiger LA, Fristensky B (1999) Constitutive expression of pea defensin gene DRR206 confers resistance to blackleg (*Leptosphaeria maculans*) disease in transgenic canola (*Brassica napus*). *Mol Plant Microbe Interact* 12: 410–418.
- Wei H, Movahedi A, Xu C, Sun W, Li L, Li D, Zhuge Q (2019) Characterization, expression profiling, and functional analysis of a *Populus trichocarpa* defensin gene and its potential as an anti-*Agrobacterium* rooting medium additive. *Sci Rep* 9, 15359.
- Zhu YJ, Agbayani R, Moore PH (2007) Ectopic expression of *Dahlia merckii* defensin DmAMP1 improves papaya resistance to *Phytophthora palmivora* by reducing pathogen vigor. *Planta* 226: 87–97.

4 CAPÍTULO II

Manuscrito de artigo
Em preparação para submissão à revista *Genetics and Molecular Biology*
(Qualis B1)
<https://www.gmb.org.br/>

Edição do genoma de feijão-caupi através do Sistema CRISPR/Cas9 direcionada ao gene eIF(iso)4E visando aquisição de imunidade/resistência ao CABMV

José Diogo Cavalcanti Ferreira^{1,2}, Francisco José Lima Aragão³, Carlos André dos Santos-Silva² and Ana Maria Benko-Iseppon²

¹Instituto Federal de Pernambuco, Br-232 Km 208, CEP 55200-000, Pesqueira, PE, Brasil;

²Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Departamento de Genética, Av. Prof. Moraes Rêgo 1235, CEP 50670-420, Recife, PE, Brasil;

³Empresa Brasileira de Pesquisa Agropecuária, Embrapa Cenargen, Parque Estação Biológica – PqEB s/nº, Av. W3 Norte (final), Brasília, DF, Brasil, CEP 70770-917

Autor correspondente: Ana Maria Benko-Iseppon, Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Av. Prof. Moraes Rego, 1235, CEP 50.670-420, Recife, PE, Brasil. Tel. +55 81-2126-8520; Email: ana.iseppon@gmail.com

Resumo

O *Cowpea aphid born mosaic virus* (CABMV) é um *Potyvirus* que acomete o feijão-caupi causando grandes prejuízos. Os *Potyvirus* dependem dos fatores de tradução (eIF4E e/ou eIF(iso)4E) da planta hospedeira para sua sobrevivência. Sabe-se que a proteína viral VPg precisa interagir com o eIF4E para replicar-se e infectar sistemicamente a planta. As mutações de perda-de-função em componentes do complexo de tradução estão associadas com resistência a vários *Potyvirus*. Este trabalho teve como objetivo avaliar dados genômicos e transcriptômicos do feijão-caupi quanto à presença e estrutura de membros da família *eIF4E*, assim como o nocaute destes genes visando à obtenção da resistência do feijão-caupi ao CABMV. O sequenciamento do eIF4E e eIF(iso)4E de três cultivares de feijão-caupi, Boca Negra (suscetível), BR14 Mulato (Suscetível) e IT85F-2687 (resistente) mostrou polimorfismos associados à substituição de aminoácidos de naturezas distintas. No eIF4E, duas mutações (P68R e G108R) observadas já são relatadas na literatura como associadas à resistência recessiva em pimenta. Os dados da modelagem comparativa e simulação da dinâmica dessas proteínas mostraram que na cv. resistente há alteração do loop e de cargas superficiais que possivelmente promovem alteração ou incompatibilidade na interação com o vírus. Em relação ao CRISPR/Cas9 foram obtidas três plantas PCR positivas para nocaute do eIF(iso)4E, sendo observada a ocorrência de uma mutação envolvendo a substituição de dois pares de bases antes da sequência utilizada como gRNA. O local e o tipo de edição ainda não foram relatados na literatura, pois o sítio-alvo de clivagem da Cas9 localiza-se no gRNA três pares de base anterior ao NGG, bem como após a clivagem o mecanismo de reparo promove mutações do tipo inserção ou deleção de bases.

Palavras-chave: *Vigna unguiculata*, Potyvirus, fitopatógeno, edição gênica, cultura de tecidos.

1. Introdução

O feijão-caupi, também conhecido como feijão-de-corda, macassar ou feijão-verde, é uma excelente fonte de proteínas (cerca de 24% de sua composição), carboidratos (70%), vitaminas e sais minerais, além de possuir baixo teor de lipídios (2%) (Frota *et al.*, 2008). Compreende a principal fonte de proteína de origem vegetal em regiões pobres da África e da América do Sul, devido à indisponibilidade e ao alto preço da proteína animal em relação à vegetal (Diouf, 2011).

As doenças são os principais fatores de redução da produtividade do feijão-caupi no Brasil. Fungos e vírus são responsáveis por 68,7% dos ataques de fitopatógenos nesta cultura (Obi e Barriuso-Vargas, 2014). Dentre as viroses, destacam-se os vírus *Cowpea aphid-borne mosaic virus* (CABMV) (família Potyviridae), *Cucumber mosaic virus* (CMV) (família Bromoviridae) e o *Cowpea severe mosaic virus* (CPSMV) (família Comoviridae). O CABMV, cuja transmissão pode ocorrer através de sementes contaminadas ou por diferentes espécies de pulgão, é caracterizado pela ampla ocorrência no Brasil, cuja sintomatologia corresponde a mosaico intenso do limbo foliar, distorção das folhas e, em casos mais extremos, redução do crescimento da planta, o que pode acarretar perdas de até 65% na produtividade (Néya *et al.*, 2015). Já o vírus do mosaico severo (CPSMV) destaca-se pela severidade, podendo provocar perda total da produção desta cultura (Barros *et al.*, 2013).

As fontes de resistência às viroses ainda são restritas e os trabalhos basicamente tentam desvendar quantos genes governam a resistência para tentar realizar um melhoramento clássico, porém linhagem elite com as características agronomicamente úteis e com resistência as viroses ainda não se encontra no mercado (Leão *et al.*, 2016).

Os fatores de iniciação da tradução – denominados eIFs – se ligam ao RNAm e permitem a associação do mesmo com a maquinaria dos ribossomos (Sesma *et al.*, 2017). Os fatores eIF4E possuem especificidade para ligação ao *cap* presente na região 5' do RNAm, sendo fundamental na interação do RNAm com os demais eIFs (Bastet *et al.*, 2017). No curso do processo evolutivo dos vegetais, os genes *eIF4E* sofreram rearranjos como duplicações, mutações e translocações, propiciando o surgimento de parálogos e/ou isoformas, as quais têm funções

redundantes. Estes genes são de cópias únicas no genoma do feijão-caupi, sendo facilmente transmitido suas mutações para linhagens comerciais.

Estudos têm demonstrado que os vírus de RNA fita simples que acometem diversas doenças em plantas, em especial aquelas provocadas por membros do grupo do *Potyvirus*, utilizam os *eIF4E* para a tradução do seu RNA. Porém, muitos deles utilizam apenas um dos parálogos dos *eIF4E* no seu processo traducional, sendo verificado que a presença de um parálogo *eif4e* (ou sua isoforma) mutante (com perda de função) pode acarretar imunidade ou resistência do vegetal ao vírus, devido à ausência de interação do gene mutante com o RNA viral (Chandrasekaran *et al.*, 2016; Pyott *et al.*, 2016; Bastet *et al.*, 2017).

Pesquisa recente com a utilização do sistema CRISPR/Cas9 mostrou que o nocaute do gene *eIF(iso)4E* em *Arabidopsis thaliana* conferiu imunidade ao Turnip mosaic virus (TuMV), não afetando o crescimento e a morfologia do vegetal, pois a presença de uma única cópia do parálogo selvagem *eIF4E* permitiu a ocorrência normal dos processos traducionais na planta (Pyott *et al.*, 2016). De forma semelhante ao trabalho supracitado, Chandrasekaran e colaboradores (2016) verificaram que existem dois parálogos em pepino (*eIF4E* e *eIF(iso)4E*), tendo realizado o nocaute do gene *eIF4E* através do sistema CRISPR/Cas9. As plantas mutantes homozigotas apresentaram imunidade ao *ipomovirus* CVYV e elevado grau de resistência a dois *potyvirus* (ZYMV e PRSV-W) (Chandrasekaran *et al.*, 2016). Em relação ao uso da técnica de CRISPR/Cas9 em feijão-caupi, Ji e colaboradores (2019) realizaram o nocaute do gene responsável pela formação dos nódulos de raízes a partir da transformação de raízes via *Agrobacterium rhizogenes*.

Diante do exposto, o presente trabalho objetivou (1) avaliar dados genômicos e transcriptômicos do feijão-caupi quanto à presença e estrutura de membros da família *eIF4E*; (2) Desenvolver plantas editadas de feijão-caupi, com o nocaute do gene *eIF4E* ou sua isoforma *eIF(iso)4E* visando à aquisição de imunidade/resistência aos vírus patogênicos de RNA fita-simples, como o Cucumber mosaic virus (CMV) (família Bromoviridae), o Cowpea severe mosaic virus (CPSMV) (família Comoviridae) e o Cowpea aphid-borne mosaic virus (CABMV) (família Potyviridae).

2. Metodologia

2.1. Gene-Alvo e desenho dos primers

O gene *eIF4E* e sua isoforma *eIF(iso)4E* de *Arabidopsis thaliana* (NP_193538.1 e At5G35620) foi utilizado como sequência-sonda, sendo alinhado contra o genoma de *Vigna unguiculata* (Lonardi *et al.*, 2019) tendo como resultado a obtenção de três genes, o *eIF4E* (Vigun06g182700.1), a isoforma *eIF(iso)4E* (Vigun07g011000.1) e uma segunda isoforma, denominada *eIF(iso)4E2* (Vigun04g112700.1).

Foram desenhados quatro primers para o gene *eIF4E*, com os quais foram realizadas três combinações distintas (Tabela 1): uma para amplificação do fragmento completo do gene; outra para o início do gene; e a última para a região terminal do mesmo. A mesma abordagem foi realizada para o gene *eIF(iso)4E*. Uma segunda isoforma que foi minerada via blast, também foi sequenciada, sendo denominada *eIF(iso)4E2*.

Tabela 1: Pares de *primers* utilizados para amplificação de regiões do gene *eIF4E*, *eIF(iso)4E1* e *eIF(iso)4E2* a partir do DNA genômico de feijão-caupi.

Primers	Sequências (5'→3')	Fragmento esperado	Região amplificada
Vu_eIF4E_1F	CGAAATGCACTCATCTATTTGC	1242 pb	Gene eIF4E completo
Vu_eIF4E_1R	CCCTCCAAACTTACAGAAACCA		
Vu_eIF4E_1F	CGAAATGCACTCATCTATTTGC	700 pb	Região inicial do gene eIF4E
Vu_eIF4E_2R	TGGTATCACAAGCAGAGGGTAA		
Vu_eIF4E_2F	CAGGAAATGAACAAATCTGAAAGG	338 pb	Região final do gene eIF4E
Vu_eIF4E_1R	CCCTCCAAACTTACAGAAACCA		
Vu_eIF(iso)4E_1F	TACATTCCTCTTAGCGTTGCTG	1073 pb	Gene eIF(iso)4E completo
Vu_eIF(iso)4E_1R	CCTCAGCATCCTCAAATTGCT		
Vu_eIF(iso)4E_1F	TACATTCCTCTTAGCGTTGCTG	327 pb	Região inicial do gene eIF4E
Vu_eIF(iso)4E_2R	TCCAGATTGCATGTTCAATTC		
Vu_eIF(iso)4E_2F	TGGGAGAAATGACTTTGATGTG	351 pb	Região final do gene eIF4E
Vu_eIF(iso)4E_1R	CCTCAGCATCCTCAAATTGCT		
Vu_eIF(iso)4E2_1F	GGATTAGGGTTTGGGAAGCA	1604 pb	Gene eIF(iso)4E2 completo
Vu_eIF(iso)4E2_1R	AAACTATTTGATGCAAGCACGA		
Vu_eIF(iso)4E2_1F	GGATTAGGGTTTGGGAAGCA	483 pb	Região inicial do gene eIF(iso)4E2
Vu_eIF(iso)4E2_2R	ACGGATGAGGCTATGAGTCAA		
Vu_eIF(iso)4E2_2F	TGTTTGGCTCTATCCTTTCTCC	702 pb	Região final do gene eIF4E2
Vu_eIF(iso)4E2_1R	AAACTATTTGATGCAAGCACGA		

A reação de PCR para amplificação específica foi ajustada para um volume final de 20 µL (contendo 10 ng do DNA), incluindo: 6 mmol MgCl₂, 200 mmol dNTPs, 3 pmol de cada *primer*, 1.0 U Taq polymerase em 1/10 volume de buffer 10x. Os

parâmetros da PCR foram desnaturação inicial a 95°C durante 5 minutos, seguido de 35 ciclos de desnaturação a 95°C por 1 minuto, anelamento a 55°C por 1 minuto e extensão a 72° por 1 minuto, com extensão final a 72°C durante 7 minutos. As reações foram realizadas a partir do DNA da cultivar Boca Negra extraído pelo método de Doyle e Doyle (1987). Esta cultivar foi utilizada no estudo devido à sua capacidade regenerativa (Aragão, pers. comm.), além de ser suscetível as viroses CABMV e CPSMV.

Os fragmentos amplificados foram isolados do gel de agarose com auxílio de um bisturi sob visualização em transiluminador, sendo purificados com o kit Illustra GFX (GE Healthcare). Na sequência, foram sequenciados através do método de Sanger (BigDye v3.1 e 3730XL DNA analyzer).

2.2. Extração de RNA, síntese de cDNA, RT-PCR e sequenciamento

A extração do RNA das folhas do feijão-caupi das cultivares Boca Negra (suscetível ao CABMV), BR14 Mulato (suscetível ao CABMV) e IT85F-2687 (resistente ao CABMV) foram realizadas em plantas com 30 dias após a semeadura. A extração foi realizada a partir do kit da Promega SV Total RNA Isolation System da Promega, conforme o protocolo do fabricante, seguido de tratamento com DNase I (*DNase Treatment and Removal* - Ambion), cuja qualidade e quantidade do RNA foram verificadas em eletroforese em gel de agarose 1,5% e em fluorímetro (Qubit), respectivamente. Alíquotas de 1 µg de RNA de cada amostra foram convertidas em cDNA utilizando o *Improm-II Reverse Transcription Systems* (Promega), seguindo as recomendações do fabricante. Os pares de *primers* para amplificação da sequência codificante completa do gene *VueIF4E* (ATGGTTGTGGAAGATTCACAAA e TCATATCACGTATTTATTTTATAGCACCC) e *VueIF(iso)4E* (ATGGCAACAAGCGAGG e TTATACGGTGTATCGACCCTTTG) foram desenhados através do programa SnapGene software (GSL Biotech; available at snapgene.com). As reações de RT-PCR foram realizadas utilizando o equipamento Techne TC-412, sob as seguintes condições: 95 °C por 7 min para desnaturação inicial, seguido de 35 ciclos de 95 °C por 1 min; 1 min à 60 °C para anelamento; 1 min à 72 °C para extensão, com extensão final a 72 °C por 10 min. Os *amplicons* foram isolados do gel e clonados no vetor pGEM-T easy conforme recomendações do fabricante, sendo realizado o sequenciamento (BigDye v3.1 e 3730XL DNA analyzer).

2.3. Alinhamento Múltiplo e Árvore Fenética

Foram utilizadas 42 sequências no alinhamento, sendo 21 de eIF4E e 21 da isoforma (Tabela Suplementar 1). O alinhamento foi realizado a partir do domínio conservado eIF4E completo com a ferramenta Clustal W (Thompson *et al.*, 1994), sendo submetido a uma análise de distância método de Neighbor-Joining (NJ), modelo de Dayhoff com bootstrap de 1000 replicações (MEGA 6.0; <http://www.megasoftware.net>) (Tamura *et al.*, 2003).

2.4. Modelagem da proteína eIF4E e eIF(iso)4E

Para modelagem tridimensional, foi utilizado o banco de dados de proteínas PDB, a fim de encontrar o melhor *template* para a construção dos modelos comparativos, onde foram selecionadas fator de iniciação 4E levando em consideração maior identidade de sequência e a cobertura. Então, utilizando o Modeller 9.23 (Webb e Sali, 2014), mil modelos foram construídos, e o modelo final foi escolhido de acordo com os valores de *discrete optimized protein energy* (DOPE) que acessa a energia do modelo e indica a estrutura mais provável, com base em um estado de referência aprimorado que corresponde a átomos que não interagem em uma esfera homogênea com o raio dependente de uma estrutura nativa da amostra. Por fim, para validação dos modelos ProSa II (Wiederstein E Sippl, 2007) e PROCHECK (Laskowski *et al.*, 1993) foram utilizados para analisar a qualidade do dobramento e a estereoquímica, respectivamente.

2.5. Simulação em dinâmica molecular

As simulações foram conduzidas com o pacote GROMACS 2019.4, as proteínas foram colocadas no centro de uma caixa cúbica de dimensões 5 x 5 x 5 nm e solvatadas usando o modelo de água SPC (Berendsen *et al.*, 1981). Os sistemas foram solvatados em solução de NaCl em concentração fisiológica (0,15 M) substituindo moléculas de soluto por íons Na⁺ e Cl⁻, em seguida a energia foi minimizada. A temperatura foi mantida a 300 K no conjunto NVT com restrição de átomos de soluto na posição inicial. O método LINCS (Hess *et al.*, 1997) foi usado para restringir ligações envolvendo átomos de hidrogênio. A integração foi realizada pelo algoritmo leapfrog (Hockney, 1970), usando uma etapa de tempo de integração de 2 fs. Os sistemas foram inicialmente otimizados em energia usando 50.000 etapas do algoritmo de descida mais íngreme. A simulação atomística foi realizada usando o campo de força GROMOS 53A6 (Oostenbrink *et al.*, 2004) e condições de

contorno periódicas nas direções x, y e z. Em seguida, a dinâmica molecular foi realizada sem restrições com pressão e temperatura constantes de 1 atm e 300 K respectivamente em um tempo de 100 ns.

2.6. Desenho do gRNA e construção do vetor CRISPR/Cas9

Com base na sequência genômica do gene *eIF4E* e do seu parálogo *eIF(iso)4E* amplificados da cultivar Boca Negra foram desenhados os RNAs guia (gRNA) tendo como alvo o primeiro exon de cada um dos genes, pois uma mutação do tipo *indels* nesta região aumentaria a probabilidade de *frameshift* no início da proteína com possível nocaute do gene. Este desenho foi realizado através das ferramentas computacionais RGENOME (Bae *et al.*, 2014), sendo confirmado pelo CRISPR Optimal Target Finder (Gratz *et al.*, 2014) e CRISPR Design (<https://github.com/sanjanalab/GUIDES>), tendo como sequência PAM o NGG e tamanho do gRNA de 20 nucleotídeos. Vale ressaltar que o programa RGENOME recomenda a utilização de sequências com valor de out-of-frame score a partir de 66%. Após o desenho dos gRNAs, os mesmos foram submetidos ao programa CRISPR Optimal Target Finder (Gratz *et al.*, 2014) para desenho do reverso complementar, sendo realizada manualmente a adição dos sítios de ligação gerados pelo corte com a enzima de restrição *AarI* para clonagem do gRNA no vetor pGEM-T sgRNA (Figura Suplementar 1). Posteriormente, a partir do sítio *PmeI* o sgRNA completo e clonado com o guia foi transferido para o vetor final pAC-AHASCas9sgRNA_eIF(iso)4E (Figura Suplementar 2). O vetor tem o gene *AtAhas* mutante de *Arabidopsis thaliana* que confere resistência ao herbicida sistêmico Imazapir. Assim como o gene codificante da proteína Cas9 e o sgRNA dirigido pelo promotor Ubiquitina 6 de *A. thaliana*.

O gRNA e seu reverso complementar foram ligados com o kit Oligo Annealing da Invitrogen, seguindo a instrução do fabricante. Para confirmação da ligação, tanto o produto da ligação como o oligo linear foram submetidos a uma corrida em gel de agarose a 2%, 70 V durante 20 minutos. Posteriormente os oligos ligados foram fosforilados com o kit da Promega seguindo as instruções do fabricante. Após a clonagem do gRNA no sítio de restrição supracitado, a integração do mesmo ao vetor foi confirmada através do sequenciamento da região através dos primers universais SP6 e T7.

Para confirmação da construção dos vetores foram utilizados os pares de primers: (1) ACT2 / Cas9 336c que amplifica um fragmento de 1.127 pb correspondente a um fragmento do gene da Cas9; (2) gRNA-speI / gRNA-KpnI para amplificar o fragmento de 684 pb correspondente ao sgRNA completo juntamente com seu promotor; (3) gRNA-KpnI / Vu_eIF(iso)4E_F que amplifica um fragmento de 303 pb, cujo primer Vu_eIF(iso)4E_F corresponde ao gRNA; (4) gRNA-speI / Vu_eIF(iso)4E_R que amplifica um fragmento de 409 pb e o primer Vu_eIF(iso)4E_R corresponde à sequência complementar do gRNA.

Todas as etapas da construção dos vetores foram utilizadas bactérias das linhagens DH10b de *Escherichia coli*, cujos vetores foram isolados por miniprep a partir de colônias isoladas e crescidas em meio LB líquido.

2.7. Transformação de feijão-caupi

A transformação do feijão-caupi via *Agrobacterium tumefaciens* foi realizada a partir de adaptações do protocolo de Soleti *et al.* (2008a) e Bakshi *et al.* (2011). As sementes de feijão-caupi foram desinfestadas a partir de lavagem com álcool etílico 70% por um minuto, seguido pela lavagem com água sanitária 2% por 20 minutos com agitação ocasional. Posteriormente, as sementes foram lavadas cinco vezes com água destilada estéril e ficaram secando por 30 minutos em câmara de fluxo.

As sementes foram colocadas para germinar em meio MSB5 (meio MS salt suplementado de vitamina B5, sacarose 3%, TDZ 10 µM, ágar e pH 5,8) por quatro dias a 28 °C com fotoperíodo de 16 h. Após a germinação, foram extraídos os nós cotiledonares a partir do corte do epicótilo e hipocótilo, sendo o explante ferido através de cinco furos realizados com agulha estéril, imediatamente antes da cocultura líquida.

Agrobacterium tumefaciens da linhagem EHA105 foi transformada por eletroporação com o vetor pAHASCas9_sgRNA-eIF(iso)4E. Após a confirmação das bactérias transformadas, as mesmas foram colocadas para crescer em meio AB sólido por 48 h no escuro a 28 °C. As bactérias foram raspadas do meio de cultura e colocadas em meio de cocultura líquida (MS salt suplementado de vitamina B5, BAP 1 µM, sacarose 3%, acetoseringona 100 µM, pH 5,5) cuja OD600 foi ajustada para 0,8. Após 30 minutos da adição das bactérias no meio de cocultura líquido, os explantes foram inseridos no meio, e cocultivados durante 30 minutos. Posteriormente, os explantes foram secos e colocados na cocultura sólida durante

três dias no escuro a 22 °C. O meio de cocultura sólido apresentou composição semelhante ao meio de cocultura líquido, com a diferença da suplementação de ágar e pH 5,2.

Após o cocultivo sólido, os explantes foram colocados em meio de seleção (Meio MS salt suplementado de vitamina B5, sacarose 3%, BAP 5 µM, cinetina 0,5 µM, cefotaxima 200 mg/L, timetina 100 mg/L, Imazapyr 500 nM, ágar e pH 5,7), sendo o mesmo substituído a cada duas semanas. Os brotos desenvolvidos foram destacados e colocados em meio de alongamento (Meio MS salt suplementado de vitamina B5, sacarose 3%, BAP 2,5 µM, cinetina 0,5 µM, cefotaxima 200 mg/L, timetina 100 mg/L, Imazapyr 50 nM, ágar e pH 5,7). Após cerca de um mês, os brotos alongados foram colocados em meio de enraizamento (Meio MS salt suplementado de vitamina B5, sacarose 3%, IBA 2,5 µM, cefotaxima 200 mg/L, timetina 100 mg/L, ágar e pH 5,7).

As plantas enraizadas foram aclimatadas e transferidas para solo com vermiculita na proporção 1:1, coletando-se uma folha para extração do DNA através do método de Doyle e Doyle (1987).

As plantas transformadas, cuja confirmação foi realizada pela amplificação da região alvo, também foram submetidas ao sequenciamento pela metodologia de Sanger, para verificação da edição do gene-alvo.

As modificações realizadas nos protocolos de transformação seguida nesta metodologia foram: 1) esterilização das sementes com o uso de hipoclorito de sódio em comparação ao cloreto de mercúrio; 2) o crescimento bacteriano foi realizado em meio AB sólido em contrapartida ao líquido; 3) a cultivar utilizada; 4) o marcador de seleção foi o gene *Atahas* com concentração de 500 µM do herbicida Imazapyr no meio de seleção e reduzido para 50 µM no meio de alongação e enraizamento, cujas metodologias seguidas utilizaram os antibióticos geneticina e/ou canamicina.

3 Resultados

3.1. Gene-Alvo

Os três genes foram amplificados e sequenciados, sendo todos os primers desenhados confirmados como funcionais (Figura Suplementar 3).

O gene *eIF(iso)4E* apresentou similaridade de 99% com a sequência genômica disponível no *phytozome*. Em relação ao *eIF(iso)4E2*, ela apresentava duas ORFs e

através de RT-PCR do cDNA de folha foi verificado que esta variante alélica do gene não se expressa nesse tecido (Figura Suplementar 4), não sendo utilizada para análises posteriores.

3.2. RT-PCR e Sequenciamento

A partir dos resultados dos sequenciamentos das três cultivares de feijão-caupi (Boca Negra, BR14 Mulato e IT85F-2687) foram verificadas algumas mutações do tipo substituição de base que promovem alterações de aminoácidos na proteína. Em relação a proteína eIF4E, ocorreu a substituição de cinco aminoácidos (Figura Suplementar 5): a posição 68 ocorreu a substituição de um resíduo de prolina nas cultivares suscetíveis (Boca Negra e BR14 Mulato) para arginina da resistente (IT85F-2687); na posição 75 ocorreu a substituição da alanina na Boca Negra pelo ácido aspártico nas cultivares BR14 Mulato e IT85F-2687; na posição 108 ocorreu a substituição de uma glicina nas linhagens suscetíveis para uma arginina da resistente; na posição 109 ocorreu a substituição de uma alanina (suscetíveis) por uma valina (resistente); e na posição 174 ocorreu a substituição da tirosina (Boca Negra) por asparagina (BR14 Mulato e IT85F-2687). Em relação à proteína eIF(iso)4E ocorreu apenas um SNP (*Single Nucleotide Polymorphism*) com a substituição do aminoácido ácido aspártico, nas cultivares suscetíveis, pela tirosina na linhagem resistente (Figura Suplementar 6).

3.3. Alinhamento Múltiplo e Árvore Fenética

O alinhamento múltiplo mostrou uma conservação dos seis resíduos de triptofano, além do resíduo de histidina no início da proteína e os dois resíduos de glicina na porção final do domínio (Figura Suplementar 7).

Em relação à árvore gerada pelo método de Neighbor Joining (análise fenética; Figura 1), foram observados dois grupamentos, um com a forma padrão e outro com a isoforma, com valor de bootstrap de 100% para a base do ramo compreendendo o grupo eIF4E (ramo superior), embora internamente baixos valores (até 5) tenham sido indicados. Em relação à proteína eIF4E de *V. unguiculata*, a mesma se agrupou com outras leguminosas como *P. vulgaris* homólogo 1 e 2, *G. max* homólogo 1 e 2, *Arachis hypogaea* e *V. radiata* com valor de bootstrap de 83%

Por sua vez, a análise não indicou nenhum valor de bootstrap na base do grupo eIF(iso)4E, indicando maior diversificação entre os grupamentos. A proteína eIF(iso)4E de feijão-caupi agrupou-se com sequências de outras leguminosas (*Phaseolus coccineus*, *P. vulgaris*, *Glycine max* e *Vigna radiata*) com valor de bootstrap 92%.

Quando adicionada a eIF(iso)4E2 na árvore fenética, a mesma ficou de fora dos dois grupos (dados não mostrados).

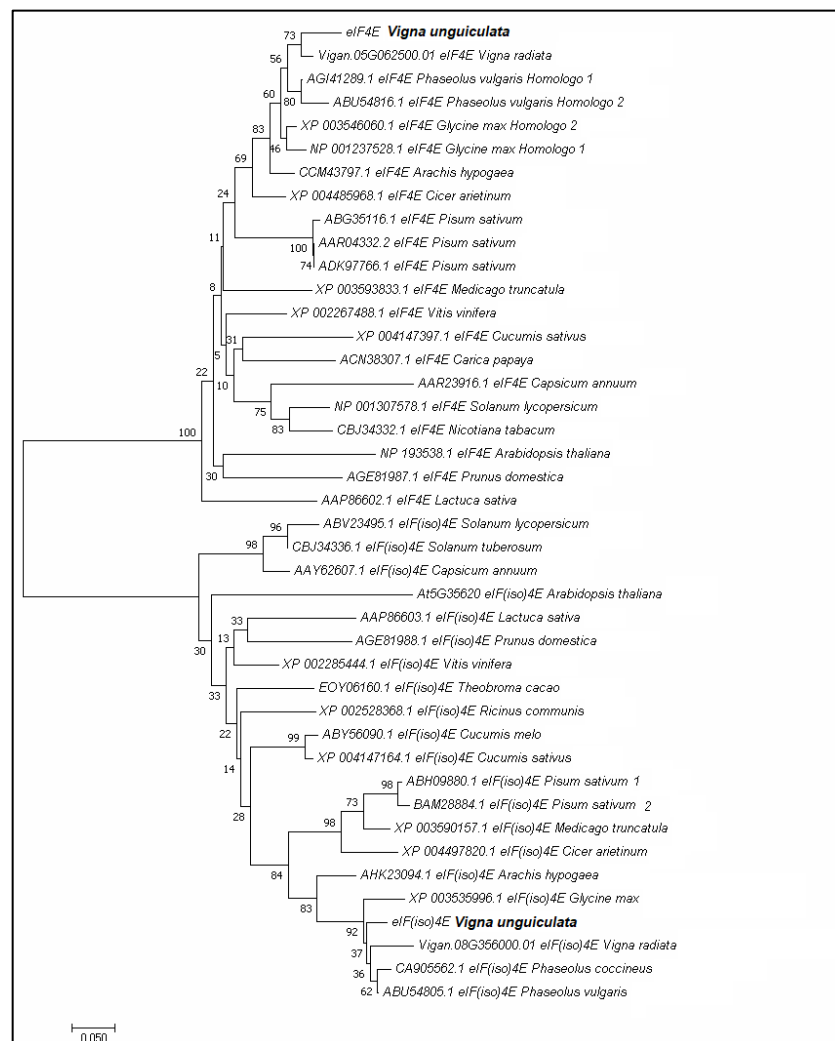


Figura 1: Árvore fenética das 42 sequências de eIF4E e eIF(iso)4E utilizadas no alinhamento, gerado pelo método de neighbor-joining, sendo o teste de confiabilidade dos nós realizado pelo método de bootstrap com 1000 replicações. As proteínas de eIF4E e eIF(iso)4E de *V. unguiculata* encontram-se destacadas em negrito.

3.4. Modelagem

A modelagem das sequências foi baseada em estruturas resolvidas disponíveis no *Protein Data Bank* (PDB) e mostrou a estrutura tridimensional da proteína eIF4E de feijão-caupi, formada por três alfa-hélices grandes e oito folhas-beta. A isoforma eIF(iso)4E de feijão-caupi foi um pouco menor, tendo uma folha-beta a menos (Figura Suplementar 8).

Em relação à comparação dos eIF4E (Figura 2) das cultivares Boca Negra, BR14 Mulato e IT85F-2687, o melhor modelo selecionado foi um fator de iniciação 4E de *Pisum sativum* (código PDB 2WMC), compartilhando 87,90%, 87,26% e 85,35% de identidade, respectivamente. Após análise estereoquímica dos aminoácidos, foi possível selecionar um modelo para cada cultivar que apresentava 90% dos aminoácidos em posições favoráveis e o Z-score indicou que todos os modelos estavam dentro da faixa de pontuações normalmente encontrado para proteínas de tamanho semelhante (Figura Suplementar 9).

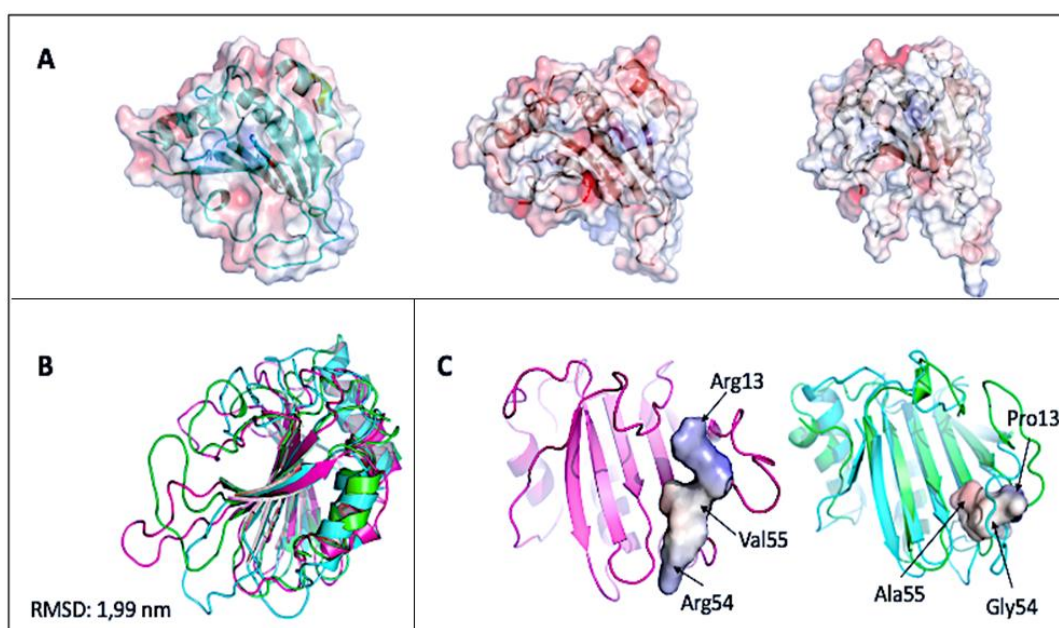


Figura 2: Modelos tridimensionais dos fatores de iniciação eIF4E de *Vigna unguiculata*. **(A)** potencial eletrostático de ± 5 kT / e de FAS2 realizado no PyMOL plotado na superfície acessível por solvente para os modelos tridimensionais teóricos das cultivares de Boca Negra, BR14 Mulato e IT85F-2687, com cargas neutras, negativas e positivas em branco, vermelho e azul, respectivamente. **(B)** alinhamento estrutural dos modelos teóricos depois da simulação em dinâmica molecular dos fatores de iniciação eIF4E das cultivares Boca Negra (em verde), BR14 Mulato (em ciano) e IT85F-2687 (em magenta). **(C)** potencial eletrostático apenas realizado no PyMOL dos resíduos de aminoácidos modificados (Pro13_Arg13, Gly54_Arg54 e Ala55_Val55).

Observa-se que as regiões com estrutura secundária permaneceram conservadas após a simulação em dinâmica molecular (Figura 2B). No entanto, as regiões de loop foram mais flexíveis, dificultando o alinhamento dessas regiões. Ao observar a carga residual (Figura 2A), foi possível perceber que o modelo Boca Negra e BR14 Mulato apresentam mais regiões intercaladas por superfície de carga negativa quando comparado com o modelo de IT85F-2687. Ao observar especificamente os três resíduos que diferem entre as cultivares sensíveis e a cultivar resistente (Figura 2C), é possível perceber que a superfície de contato foi alterada.

3.5. Simulação em dinâmica molecular

A fim de observar as possíveis modificações que o modelo poderia sofrer, foram avaliados os gráficos de *root mean square deviation* (RMSD), *root mean square fluctuation* (RMSF), o raio de giro, além da estrutura secundária (Figura Suplementar 10).

Com a análise dos gráficos foi possível observar que todos os modelos saíram da conformação original (Figura Suplementar 10A), contudo se mostraram estáveis em maior parte da simulação em dinâmica molecular. Quando foi analisada a flutuação dos aminoácidos (Figura Suplementar 10C) foi possível perceber quatro picos que mais flutuaram, sendo dois nas regiões das extremidades, com maiores picos para o modelo da cultivar Boca Negra, e outros dois picos maiores localizados entre os aminoácidos 13 – 23 e 67 – 75, sendo este último com maiores flutuações para o modelo da cultivar IT85F-2687. A Figura Suplementar 10C revelou que essas regiões as quais tiveram maior mobilidade dinâmica dos átomos estavam localizadas ou nas extremidades ou em regiões sem estrutura secundária (loop), que normalmente possui um grau de flexibilidade maior para flutuação.

O gráfico do raio de giro (Figura Suplementar 10B), comprova que as estruturas se mantiveram compactas corroborando a hipótese de que as regiões com estrutura secundária se mantiveram rígidas enquanto que as outras regiões sem estruturas secundárias foram mais flexíveis.

3.6. Desenho do gRNA e construção do vetor CRISPR/Cas9

Através da sequência do primeiro exon foram obtidos 46 gRNAs para o gene eIF4E e 22 gRNAs a partir do eIF(iso)4E, com 20 nucleotídeos cada, contendo a presença de um NGG na porção 3' do DNA alvo. Os três melhores gRNAs de cada

gene foram selecionados para construção do vetor, sendo selecionado o terceiro melhor gRNA tanto do eIF4E, quanto da isoforma, baseado no *out-of-frame score*, por possuírem sítio de restrição na região de corte da Cas9, podendo ser utilizado como marcador para selecionar as plantas editadas (Tabelas 2 e 3). Outro fator favorável a esta escolha referiu-se ao fato de não haver *off-target* baseado no genoma disponível para feijão-caupi (Lonardi *et al.*, 2019), assim como de espécies próximas do gênero *Vigna* como *V. radiata* (Kang *et al.*, 2014) e *V. angularis* (Sakai *et al.*, 2016).

Tabela 2: Três melhores gRNAs a partir do gene *eIF4E* do feijão-caupi baseado no *out-of-frame score*.

eIF4E			
gRNA 5' → 3' (com NGG)	Out-of-frame Score	Conteúdo GC (%)	Sítio de restrição
GTGGGGGTTGCGGAGGAGTGCGG	84,8	70	Não tem
GGGTTGCGGAGGAGTGCGGACGG	81,7	70	Acil
AGGTCCAGGAATTCTCGAGAGGG	78,7	50	XhoI

Tabela 3: Três melhores gRNAs a partir do gene *eIF(iso)4E* do feijão-caupi baseado no *out-of-frame score*.

eIF(iso)4E			
gRNA 5' → 3' (com NGG)	Out-of-frame Score	Conteúdo GC (%)	Sítio de restrição
AGCACAAGCTGGAGCGAAAATGG	78,7	50	Não tem
CCTAAGCAGGGTGCTGCTTGGGG	74,3	60	TseI / ApeKI
TTGGGGAACCTCTCTGCGCAAGG	73,8	60	FspI / HhaI

A partir dos dois gRNAs selecionados, foi desenhado o seu reverso complementar, adicionando-se manualmente os sítios de ligação da enzima AaRI. Para o gRNA selecionado foi adicionada a sequência ATTG na extremidade 5', e no reverso complementar foi adicionada a sequência AAAC na extremidade 5' (Figura 3).

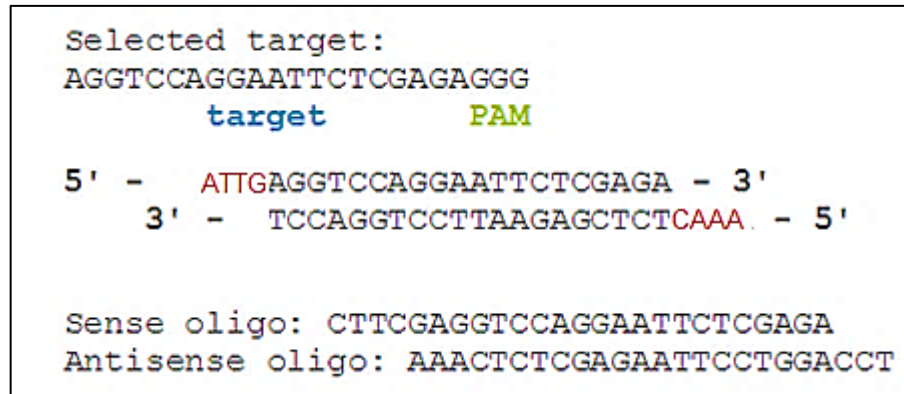


Figura 3: Confirmação do gRNA desenhado e selecionado a partir do RGENOME através do programa CRISPR Optimal Target Finder. A partir deste programa foi realizado o desenho do oligo antisenso, sendo manualmente adicionado o sítio de ligação (em vermelho na extremidade 5' do oligo senso e antisenso) gerado pelo corte com a enzima de restrição AarI.

O gRNA e seu reverso complementar foram ligados e confirmada a ligação através da corrida no gel de agarose confirmando a ligação pela diferença de tamanho das bandas visualizadas em transiluminador (Figura Suplementar 11). Posteriormente, os oligos ligados foram fosforilados e realizada a reação de clonagem dos oligos ligados com o vetor digerido originando os vetores pGEM-T sgRNA_eIF4E e pGEM-T sgRNA_eIF(iso)4E (Figura suplementar 1). A presença do gRNA foi confirmada pelo sequenciamento do vetor (Figura Suplementar 12).

Os vetores pGEM-T sgRNA_eIF4E e pGEM-T sgRNA_eIF(iso)4E foram digeridos com a enzima PmeI, cujo fragmento de 393 pb (correspondente ao sgRNA completo) foi isolado e purificado (Figura 4).

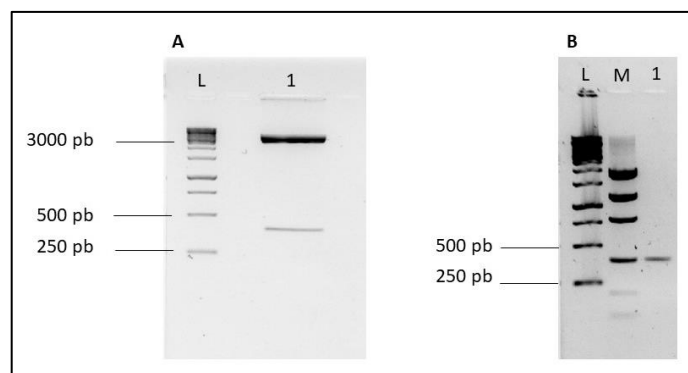


Figura 4: Vetor pGEM-T sgRNA digerido com PmeI (imagem A), cujo fragmento de 393 pb (poço 1) correspondente ao sgRNA completo foi isolado e purificado com kit da Promega, sendo quantificado com o marcador *Low DNA Mass Ladder* da Invitrogen (imagem B poço M), cuja intensidade está mais próxima ao fragmento de 400 pb correspondente a uma concentração de 40 ng de DNA.

O sgRNA foi a seguir ligado ao vetor pAHASCas9 previamente digerido com a PmeI, para obtenção dos vetores finais pAHASCas9_sgRNA-eIF4E e pAHASCas9_sgRNA-eIF(iso)4E (Figura Suplementar 2).

Os vetores isolados de seis colônias de bactérias foram inicialmente digeridos com a enzima NcoI e XbaI, que geraram quatro fragmentos (8947, 4568, 1683 e 1067 pb) realizando corte nas três sequências principais dos vetores (no gene *Ahas*, utilizado como seleção, no Cas9 e no sgRNA) (Figura 5A). A colônia que teve o padrão esperado foi digerida separadamente com as enzimas PvuI, PvuII e NotI para maior confiabilidade da confirmação do vetor final (Figura 5B). Por fim, os vetores que passaram por todas as digestões foram confirmados por PCR e sequenciamento (Figura Suplementar 13).

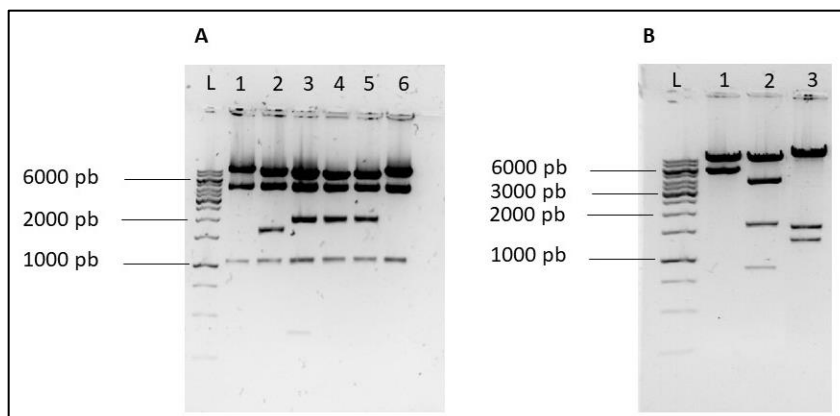


Figura 5: Confirmação do vetor final pAHASCas9_sgRNA-eIF(iso)4E através de digestão enzimática. Inicialmente os plasmídeos isolados de seis colônias foram digeridos com NcoI e XbaI que geram quatro fragmentos (8947, 4568, 1683 e 1067 pb), sendo que só a colônia do poço 2 teve o padrão esperado (imagem A). Posteriormente este vetor do poço 2 da imagem A foi digerido com as enzimas PvuI que geraram fragmentos de 10417 pb e 5848 pb (poço 1 imagem B), PvuII que geram fragmentos de 9676, 3770, 1643 e 880 pb (poço 2 imagem B) e NotI que geram fragmentos de 13443, 1532 e 1290 pb (poço 3 imagem B).

Um total de 350 explantes foi utilizado em quatro eventos de transformação via *A. tumefaciens*, sendo regeneradas três plantas. Duas destas três plantas foram oriundas do mesmo explante. Após a PCR as três plantas transformadas foram confirmadas, contendo o fragmento do vetor referente ao sgRNA (Figura 6). A eficiência de transformação, calculada pela razão do número de plantas transformadas pelo número de explantes foi de 0,86%.

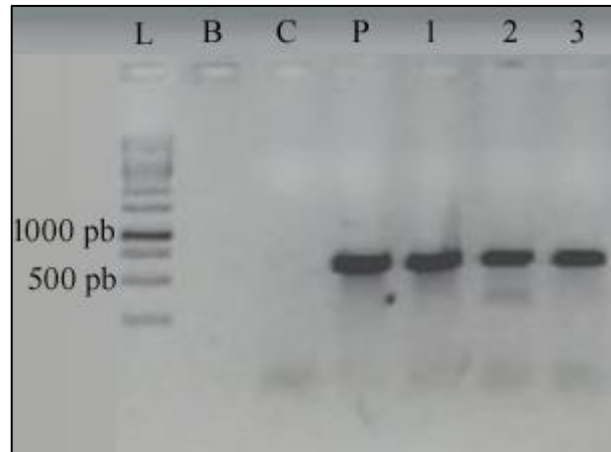


Figura 6: Gel de agarose 1,5% do produto da amplificação das possíveis plantas transformadas com o par de primers gRNA-speI / gRNA-KpnI com fragmento de 684 pb. L – Ladder 1 Kb; B – Branco (sem DNA); C – Controle negativo (planta não transformada); P – Controle positivo correspondente ao plasmídeo pAHASMCDEF1; 1 – planta transformada 1; 2 –planta transformada 2; 3 – planta transformada 3. L – Ladder 1 Kb; B – Branco (sem DNA); C – Controle negativo (planta não transformada); P – Controle positivo correspondente ao plasmídeo pAHASMCDEF1; 1 – planta transformada 1; 2 – planta transformada 2; 3 – planta transformada 3.

As três plantas confirmadas por PCR como transformadas com o vetor de CRISPR foram submetidas ao sequenciamento do gene-alvo da edição (eIF(iso)4E). Foi verificada a substituição de uma guanina por uma citosina, na segunda base anterior à região do RNA guia (Figura 7). Essa mutação promove a substituição de um aminoácido alanina por uma prolina nas plantas editadas, mantendo o quadro de leitura da proteína.



Figura 7: Alinhamento do produto do sequenciamento do gene eIF(iso)4E da planta controle (cv. Boca Negra não transformada) e das três plantas transformadas PCR positivas para verificação da ocorrência da edição. Verifica-se uma mutação do tipo substituição de base (uma guanina na planta controle por uma citosina nas plantas transformadas / editadas – vide seta amarela). A mutação ocorreu duas bases antes do RNA guia, destacado pelas letras vermelhas.

4 Discussão

As sequências dos genes e proteínas eIF4E e eIF(iso)4E formaram dois agrupamentos, sendo que as sequências de feijão-caupi ficaram posicionadas em agrupamentos separados juntamente com sequências de outras leguminosas, demonstrando uma relativa conservação dessas proteínas entre as Fabaceae. A conservação e separação das proteínas eIF4E e eIF(iso)4E tem sido observada inclusive com a introdução nos alinhamentos de grupos taxonômicos distantes como humanos, ratos, insetos, leveduras e plantas (Marcotrigiano *et al.*, 1997; Xu *et al.*, 2017).

Em relação à modelagem, a principal diferença entre as duas proteínas de feijão-caupi deve-se a da presença de uma folha-beta adicional na forma padrão, o que corrobora com as estruturas de diversas proteínas eIF4E descritas por Marcotrigiano e colaboradores (1997).

Após a obtenção do vetor, o processo de transformação teve uma eficiência de 0,86%, portanto abaixo dos trabalhos de Soleti e colaboradores (2008b) e Bakshi e colaboradores (2011), com 3,09% e 1,64% de plantas de feijão-caupi transformadas, respectivamente. A comparação com estes dois trabalhos foi devido ao fato de que o protocolo utilizado corresponde a uma adaptação dos trabalhos citados, com modificações. Vale salientar que foram realizadas diversas mudanças como uso de uma estirpe diferente de *A. tumefaciens*. Ainda, o modo de crescimento da bactéria no presente trabalho ocorreu em meio sólido, diferentemente dos referidos estudos onde foi usado meio líquido. Finalmente, a cultivar de feijão-caupi utilizada foi diferente, além de algumas pequenas modificações na composição dos meios de cultivo e do marcador de seleção utilizado (gene *Atahas*) que confere resistência ao herbicida Imazapyr, enquanto nos outros dois trabalhos foi utilizado o gene *nptII*, que confere resistência a antibióticos. Todas essas mudanças podem afetar a taxa de transformação, uma vez que o processo é genótipo dependente, assim como depende dos protocolos e composição do meio de cultura de tecidos (Popelka *et al.*, 2006; Citadin *et al.*, 2011; Bett *et al.*, 2019).

A interação das proteínas virais depende de sua carga superficial, essa carga eletrostática determina a mobilidade da partícula em um campo elétrico e, portanto, governa seu comportamento, que desempenha um papel importante na interação

vírus-hospedeiro (Michen e Graule 2010). Quando ocorre mudança na carga superficial, essa mudança pode interferir na ligação da proteína do hospedeiro com a proteína viral. A substituição de aminoácidos apolares (Prolina e Glicina) por aminoácidos carregados positivamente (Arginina), como no caso da proteína do fator de iniciação eIF4E na cultivar IT85F-2687 fez com que a superfície de contato eletrostática mudasse e a tornasse menos aniônica (Figura 2.A), ao invés de uma superfície compacta e hidrofóbica agora, tem-se uma superfície expansiva e hidrofílica. Essa mudança no potencial eletrostático (Figura 2.C) pode ser uma das hipóteses para a não interação da proteína viral com o fator de iniciação eIF4E em IT85F-2687, desta forma, a planta torna-se resistente pela ausência de ligação entre essas proteínas. As mutações que promoveram a substituição de aminoácidos entre as linhagens suscetíveis comparativamente à resistente já foram descritas na literatura para outras espécies (Nicaise *et al.*, 2003; Yeam *et al.*, 2007; Charron *et al.*, 2008). A mutação G108R acarreta resistência recessiva ao *potyvirus* (*Capsicum-Tobacco etch virus* - TEV) em pimenta, impedindo a interação do eIF4E com a VPg viral (Yeam *et al.*, 2007). As mutações nas posições 68, 75 e 109 também já foi relatada e estão relacionadas à co-evolução do eIF4E x VPg viral (Nicaise *et al.*, 2003; Charron *et al.*, 2008). Trabalho recente mostrou que o nocaute do gene eIF4E em tomate promove resistência ao *Potyvirus pepper veinal mottle virus* em tomate (Mourya *et al.*, 2020). Em relação à resistência a *Potyvirus* correlacionada com os fatores de iniciação da tradução 4E em leguminosas, Naderpour e colaboradores (2010) detectaram que o locus relacionado à resistência ao *Bean Common Mosaic Virus* (BCMV) corresponde ao gene *eIF4E* em feijão comum. Xu e colaboradores (2017) verificaram em amendoim que o *Peanut Stripe Virus* (PStV) interagem tanto com o eIF4E, quanto com sua isoforma. Contudo, técnicas *in silico* como *docking* molecular podem fornecer indícios da região de ligação entre essas proteínas, que poderiam ser validadas pelo sistema de duplo híbrido em *Saccharomyces cerevisiae*.

Com relação à edição do genoma do feijão-caupi, obteve-se um resultado inesperado, pois o ponto alvo da ação da Cas9 estava direcionado para três pares de bases antes do NGG, local que inclui a ancoragem do RNA guia (Belhaj *et al.*, 2015). Após a clivagem da dupla fita do DNA, entra em ação o mecanismo de reparo promovendo mutações do tipo inserção ou deleção (Chandrasekaran *et al.*, 2016; Pyott *et al.*, 2016; Gomez *et al.*, 2018). Entretanto, no presente estudo, foi verificada

uma mutação do tipo substituição de uma base, dois pares de bases anterior à sequência do RNA guia, algo ainda não relatado nos experimentos de CRISPR/Cas9, tanto em relação ao tipo de mutação quanto ao local. Essa mutação promove apenas uma substituição do aminoácido alanina por uma prolina, ambos possuindo a mesma natureza química, pois são apolares. Como não houve alteração na matriz de leitura da proteína, possivelmente o seu funcionamento, assim como a possível interação com a VPg viral, continuará intacto.

O presente trabalho desenvolveu um feijão-caupi transgênico via *A. tumefaciens* com uma metodologia nova a partir de diversas modificações de protocolos já existentes, assim como promoveu a edição do genoma, porém sem a efetividade esperada em relação ao objetivo, pois provavelmente não conseguiu realizar o nocaute do gene-alvo. No entanto, nas plantas transformadas portadora do marcador de seleção (gene *Ahas*), possivelmente o cassete continuará a realizar o processo de edição, o que provavelmente aumentará a eficiência e os tipos de mutações possibilitando o nocaute do gene alvo, situação já descrita para soja (Bao *et al.*, 2019).

5 Agradecimentos

Os autores agradecem a Michael Timko por ceder gentilmente as sequências de DNA do de fatores de iniciação do feijão-caupi antes mesmo da publicação do genoma. A Pedro Berbert pelo cuidado com as plantas transformadas e a Heitor Vicente Rosa pelas discussões calorosas quanto às hipóteses do trabalho. Assim como agradecem as agências de fomento Facepe, Capes e CNPq que propiciaram o desenvolvimento deste projeto.

6 Referências Bibliográficas

- | | |
|---|---|
| <p>Bae S, Park J, Kim J-S (2014) Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. <i>Bioinformatics</i> 30: 1473-1475.</p> <p>Bao A, Chen H, Chen L, Chen S, Hao Q, Guo W, Qiu D, Shan Z, Yang Z, Yuan S, Zhang C, Zhang X, Liu B,</p> | <p>Kong F, Li X, Zhou X, Tran LSP, Cao D (2019) CRISPR/Cas9-mediated targeted mutagenesis of GmSPL9 genes alters plant architecture in soybean. <i>BMC Plant Biol</i> 19, 131.</p> <p>Barros GB, Nogueira MSR, Oliveira CRR, Freire-Filho FR, Ribeiro VQ, Veiga CFM, Briosi PST, Eiras M (2013) Obtenção de plantas de feijão-caupi</p> |
|---|---|

- resistentes ao Cowpea severe mosaic virus e ao Cowpea aphid-borne mosaic virus. *Summa Phytopathol* 39(2): 130-136.
- Bakshi S, Sadhukhan A, Mishra S, Sahoo L (2011) Improved *Agrobacterium*-mediated transformation of cowpea via sonication and vacuum infiltration. *Plant Cell Rep* 30: 2281–2292.
- Bastet A, Robaglia C, Gallois J (2017) eIF4E Resistance: Natural Variation Should Guide Gene Editing. *Trends Plant Sci* 22(5): 411-419.
- Belhaj K, Chaparro-Garcia A, Kamoun S, Patron NJ, Nekrasov V (2015) Editing plant genomes with CRISPR/Cas9. *Current Opin Biotech* 32: 76-84.
- Berendsen HJ, Postma JP, van Gunsteren WF, Hermans J (1981) Interaction Models for Water in Relation to Protein Hydration, In *Intermolecular Forces*, edited by Bernard Pullman, 14: 331–42. Dordrecht: Springer Netherlands.
- Bett B, Gollasch S, Moore A, Harding R and Higgins TJV (2019) An Improved Transformation System for Cowpea (*Vigna unguiculata* L. Walp) via Sonication and a Kanamycin-Geneticin Selection Regime. *Front Plant Sci* 10:219.
- Chandrasekaran J, Brumin M, Wolf D, Leibman D, Klap C, Pearlsman M, Sherman A, Arazi T, Gal-On A (2016) Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol Plant Pathol* 17(7): 1140–1153.
- Charron C, Nicolai M, Gallois JL, Robaglia C, Moury B, Palloix A, Caranta C (2008) Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg. *Plant J* 54: 56–68.
- Citadin CT, Ibrahim AB, Aragão FJL (2011) Genetic engineering in cowpea (*Vigna unguiculata*): History, status and prospects. *GM Crops* 2(3): 144-149.
- Diouf D (2011) Recent advances in cowpea [*Vigna unguiculata* (L.) Walp.] “omics” research for genetic improvement. *African J Biotechnol* 10: 2803–2810.
- Doyle JJ and Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.
- Eswar N, Webb B, Marti-Renom MM, Eramian D, Shen MY, Pieper U, Sali A (2007) Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* 50(1): 2.9.1–2.9.31.
- Frota KGM, Soares RAM, Arêas JAG (2008) Composição química do feijão caupi (*Vigna unguiculata* L. Walp.), cultivar BRS-Milênio. *Ciênc Tecnol Aliment* 28: 470-476.
- Gomez MA, Lin ZD, Moll T, Chauhan RD, Hayden L, Renninger K, Beyene G, Taylor NJ, Carrington JC, Staskawicz BJ, Bart RS (2019) Simultaneous CRISPR/Cas9-mediated editing of cassava *eIF4E* isoforms *nCBP-1* and *nCBP-2* reduces cassava brown streak disease symptom severity and incidence. *Plant Biotechnol J* 17: 421-434.
- Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, O'Connor-Giles KM (2014) Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics* 196(4):961-71.
- Hess B, Bekker H, Berendsen HJ, Fraaije JG (1997) LINCS: A Linear Constraint Solver for Molecular Simulations. *J Comput Chem* 18: 1463–72.
- Hockney RW (1970) *Methods in Computational Physics*. Alder B, 136–211.
- Ji J, Zhang C, Sun Z, Wang L, Duanmu D, Fan Q (2019) Genome Editing in Cowpea *Vigna unguiculata* Using CRISPR-Cas9. *Int J Mol Sci* 20(10).
- Kang YJ, Kim SK, Kim MY, Lestari P, Kim KH, Ha BK, et al. (2014) Genome sequence of mungbean and insights

- into evolution within *Vigna* species. *Nat Commun* 5: 5443.
- Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26(2): 283–291.
- Leão US, Eiras M, Freire Filho FR, Nogueira MSR, Rodrigues LK, Ribeiro VQ, Lopes ACA (2016) Selection of phenotypic traits and resistance to Cowpea severe mosaic virus and Cowpea aphidborne mosaic virus in cowpea [*Vigna unguiculata* (L.) Walp.] seeds with rugose white coat. *Aust J Crop Sci* 10(4):470-481.
- Lonardi S, Muñoz-Amatriaín M, Liang Q, Shu S, Wanamaker SI, Lo S, et al. (2019) The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant J* 98(5): 767-782.
- Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK (1997) Cocystal Structure of the Messenger RNA 59 Cap-Binding Protein (eIF4E) Bound to 7-methyl-GDP. *Cell* 89: 951–961.
- Michen B, Graule T (2010) Isoelectric points of viruses. *J appl microbiol* 109: 388-397.
- Mourya B, Lebaron C, Szadkowski M, Khalifa MB, Girardot G, Bi BAB, Koné D, Nitiema LW, Fakhfakh H, Gallois JL (2020) Knock-out mutation of eukaryotic initiation factor 4E2 (eIF4E2) confers resistance to *Pepper Veinal Mottle Virus* in tomato. *Virology* 539: 11-17.
- Naderpour M, Lund OS, Larsen R, Johansen E (2010) Potyviral resistance derived from cultivars of *Phaseolus vulgaris* carrying bc-3 is associated with the homozygotic presence of a mutated eIF4E allele. *Mol Plant Pathol* 11(2):255-63.
- Néya BJ, Zida PE, Sérémé D, Lund OS, Traoré O (2015) Evaluation of Yield Losses Caused by *Cowpea Aphid-Borne Mosaic Virus* (CABMV) in 21 Cowpea (*Vigna unguiculata* (L.) Walp.) Varieties in Burkina Faso. *PJBS* 18(7): 304-313.
- Nicaise V, German-Retana S, Sanjuán R, Dubrana MP, Mazier M, Maisonneuve B, Candresse T, Caranta C, LeGall O (2003) The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the Potyvirus *Lettuce Mosaic Virus*. *Plant Physiol* 132(3): 1272–1282.
- Obi VI, Barriuso-Vargas JJ (2014) Situation of Biofungicides Reconnaissance, a Case of Anthracnose Disease of Cowpea. *Am J Plant Sci* 5: 1202-1211.
- Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J comput chem*, 25: 1656-1676.
- Popelka JC, Gollasch S, Moore A, Molvig L and Higgins TJV (2006) Genetic transformation of cowpea (*Vigna unguiculata* L.) and stable transmission of the transgenes to progeny. *Plant Cell Rep* 25: 304-312.
- Pyott DE, Sheehan E, Molnar A (2016) Engineering of CRISPR/Cas9-mediated potyvirus resistance in transgene-free Arabidopsis plants. *Molecular Plant Pathol* 17(8): 1276–1288.
- Sakai H, Naito K, Takahashi Y, Sato T, Yamamoto T, Muto I, Itoh T, Tomooka N (2016) The *Vigna* Genome Server, 'VigGS': A Genomic Knowledge Base of the Genus *Vigna* Based on High-Quality, Annotated Genome Sequence of the Azuki Bean, *Vigna angularis* (Willd.) Ohwi & Ohashi. *Plant Cell Physiol* 57(1):e2.
- Sesma A, Castresana C, Castellano MM (2017) Regulation of Translation by TOR, eIF4E and eIF2 α in Plants: Current Knowledge, Challenges and Future Perspectives. *Front Plant Sci* 8644: 1-7.
- Silva SMS, Freire-Filho FR (1998) Proteína de feijão-caupi (*Vigna unguiculata* (L.) Walp.):

- Caracterização e aplicação nutricional. Embrapa Documentos 44: 5-20.
- Solleti SK, Bakshi S and Sahoo L (2008a) Additional virulence genes in conjunction with efficient selection scheme, and compatible culture regime enhance recovery of stable transgenic plants in cowpea via *Agrobacterium tumefaciens*-mediated transformation. J Biotechnol 135: 97–104.
- Solleti SK, Bakshi S, Purkayastha J, Panda SK, Sahoo L (2008b) Transgenic cowpea (*Vigna unguiculata*) seeds expressing a bean α -amylase inhibitor 1 confer resistance to storage pests, bruchid beetles. Plant Cell Rep 27: 1841-1850.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30(12): 2725-9.
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673-80.
- Webb B, Sali A (2014) Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinforma 47: 5.6.1-5.6.32.
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35: W407–W410.
- Xu M, Xie H, Wu J, Xie L, Yang J, Chi Y (2017) Translation Initiation Factor eIF4E and eIFiso4E Are Both Required for Peanut stripe virus Infection in Peanut (*Arachis hypogaea* L.). Front Microbiol 8: 338.
- Yeaman I, Cavatorta JR, Ripoll DR, Kang BC, Jahn MM (2007) Functional Dissection of Naturally Occurring Amino Acid Substitutions in eIF4E That Confers Recessive Potyvirus Resistance in Plants. Plant Cell 19 (9): 2913-2928.

ANEXOS

Tabela Suplementar 1: Nomes, espécies e número de acesso dos genes *eIF4E* e *eIF(iso)4E* utilizados na Figura 2 e Anexo II. O banco de acesso as sequências foi o NCBI.

Nomes	Espécies	Número de acesso
AtelF4E	<i>Arabidopsis thaliana</i>	NP_193538.1
AhelF4E	<i>Arachis hypogaea</i> L.	CCM43797.1
CaelF4E	<i>Capsicum annuum</i>	AAR23916.1
CpelF4E	<i>Carica papaya</i>	ACN38307
CaelF4E	<i>Cicer arietinum</i>	XP_004485968
CseIF4E	<i>Cucumis sativus</i>	XP_004147397.1
GmelF4E1	<i>Glycine max</i>	NP_001237528
GmelF4E2	<i>Glycine max</i>	XP_003546060
LseIF4E	<i>Lactuca sativa</i>	AAP86602.1
MtelF4E	<i>Medicago truncatula</i>	XP_003593833
NtelF4E	<i>Nicotiana tabacum</i>	CBJ34332
PvelF4E2	<i>Phaseolus vulgaris</i>	ABU54816.1
PvelF4E1	<i>Phaseolus vulgaris</i>	AGI41289.1
PseIF4E1	<i>Pisum sativum</i>	ABG35116
PseIF4E2	<i>Pisum sativum</i>	ADK97766
PseIF4E_P1	<i>Pisum sativum</i>	AAR04332.2
PdelF4E	<i>Prunus domestica</i>	AGE81987
SleIF4E	<i>Solanum lycopersicum</i>	NP_001307578.1
VvelF4E	<i>Vitis vinifera</i>	XP_002267488
VrelF4E	<i>Vigna radiata</i>	Vigan.05G062500.01
AtelF(iso)4E	<i>Arabidopsis thaliana</i>	At5G35620
AhelF(iso)4E	<i>Arachis hypogaea</i> L.	AHK23094.1
CaelF(iso)4E	<i>Capsicum annuum</i>	AAY62607
CaelF(iso)4E	<i>Cicer arietinum</i>	XP_004497820
CmelF(iso)4E	<i>Cucumis melo</i>	ABY56090
CseIF(iso)4E	<i>Cucumis sativus</i>	XP_004147164.1
GmelF(iso)4E	<i>Glycine max</i>	XP_003535996
LseIF(iso)4E	<i>Lactuca sativa</i>	AAP86603
MtelF(iso)4E	<i>Medicago truncatula</i>	XP_003590157
PcelF(iso)4E2	<i>Phaseolus coccineus</i>	CA905562.1
PvelF(iso)4E	<i>Phaseolus vulgaris</i>	ABU54805
PseIF(iso)4E1	<i>Pisum sativum</i>	ABH09880.1
PseIF(iso)4E2	<i>Pisum sativum</i>	BAM28884.1
PdelF(iso)4E	<i>Prunus domestica</i>	AGE81988
VrelF(iso)4E1	<i>Vigna radiata</i>	Vigan.08G356000.01
RcelF(iso)4E	<i>Ricinus communis</i>	XP_002528368
SleIF(iso)4E	<i>Solanum lycopersicum</i>	ABV23495.1
StelF(iso)4E	<i>Solanum tuberosum</i>	CBJ34336
TcelF(iso)4E	<i>Theobroma cacao</i>	EOY06160.1
VvelF(iso)4E	<i>Vitis vinifera</i>	XP_002285444

Figura Suplementar 1. Vetor pGEM-T sgRNA_eIF(iso)4E, sendo utilizado como intermediário para clonagem do gRNA, cujo isolamento do sgRNA completo foi realizado a partir dos sítios de restrição da PmeI.

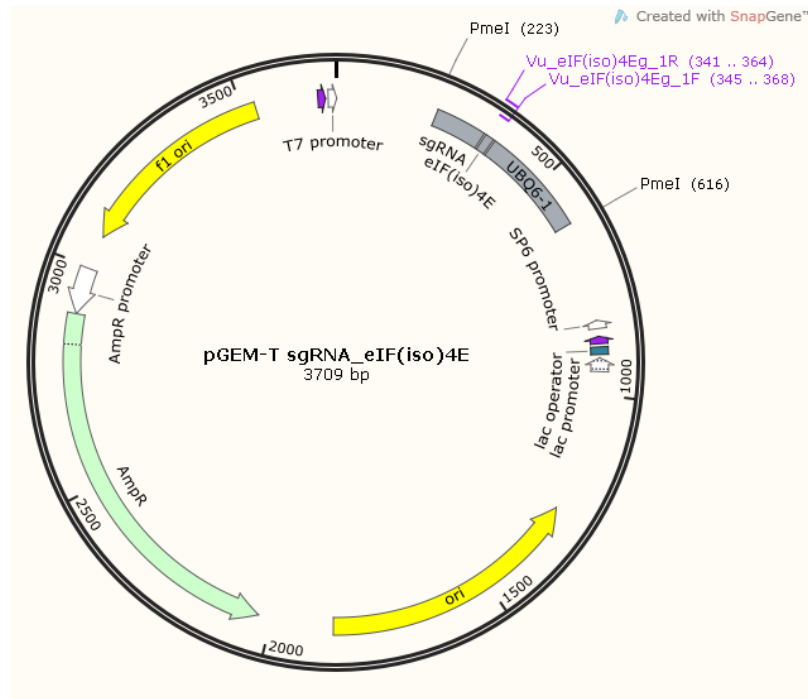


Figura Suplementar 2: Vetor pAHASCas9_sgRNA_eIF(iso)4E com sítio de restrição AarI para clonagem do gRNA utilizado na edição do genoma a ser escolhido. Neste caso é necessária a introdução de uma guanina antes do gRNA selecionado para dirigir a expressão do guia.

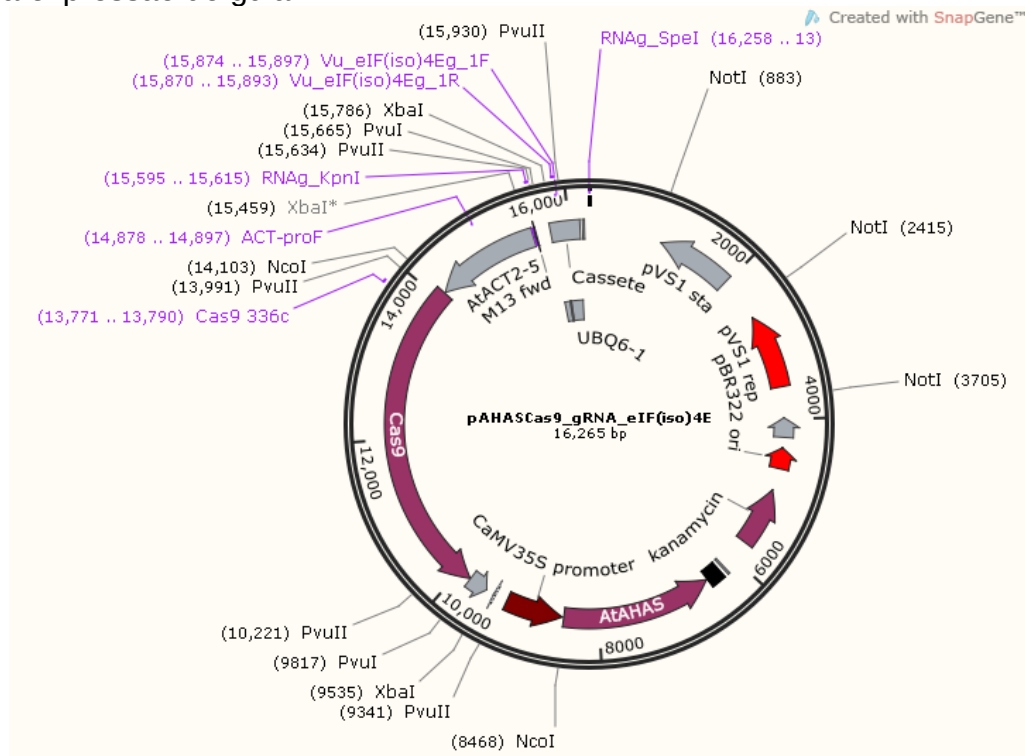


Figura Suplementar 3: Eletroforese em gel de agarose 1% dos *amplicons* dos genes: A - eIF4E (1: Par de primers 1F com 1R; 2: 1F com 2R; 3: 2F com 1R); B - eIF(iso)4E1 (1: Par de primers 1F com 1R; 2: 1F com 2R; 3: 2F com 1R); C - eIF(iso)4E2 (1: Par de primers 1F com 1R; 2: 1F com 2R; 3: 2F com 1R). O primeiro fragmento corresponde à sequência genômica completa, o segundo à região inicial do gene e o terceiro à porção final, para todos os géis. As sequências dos primers, assim como o tamanho dos *amplicons* encontram-se na Tabela 1. L – Ladder de 100 pb.

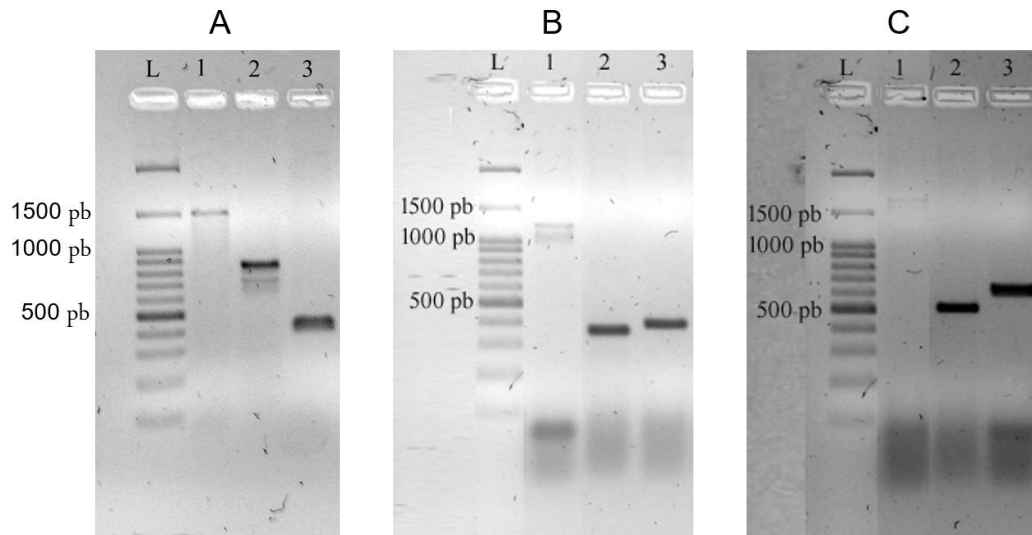


Figura Suplementar 4: Eletroforese em gel de agarose 1,2% dos *amplicons* da RT-PCR a partir do cDNA da cv. Boca Negra de *V. unguiculata* dos genes: 1 - eIF4E (com os primers ATCACGGAGATGAAATCTGTGG e GTCACCTGAGCAGCTTCATTT, que amplificam um fragmento de 103 pb); 2 - eIF(iso)4E (com os primers AACAAGCGAGGAAGTGGTTG e CCTGCTTAGGTTTGGATTGATT, que amplificam um fragmento de 110 pb); 3 - eIF(iso)4E2 (com os primers AGAGAACAACAGCGCCAATAAC e TCGTTCTTCGGTTTCTTTGCT, que amplificam um fragmento de 106 pb). L – Ladder 1 Kb.

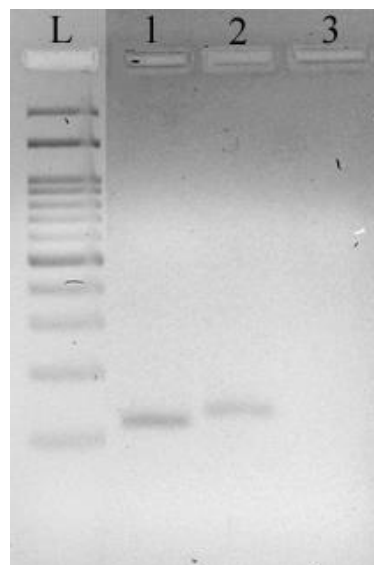


Figura Suplementar 5: Sequenciamento, tradução e alinhamento das *cds* do gene eIF4E em três cultivares de feijão-caupi: Boca Negra (suscetível ao CABMV); BR14 Mulato (suscetível) e IT85F-2687 (resistente). A seta amarela destaca a mutação P68R e a vermelha a mutação G108R, ambas são descritas na literatura na aquisição de resistência à *Potyvirus*.

Vu_eIF4E_Boca_Negra	MVVEDSQKSTITDDQNPSRVNDDEDLEDGEILEDADDAASAASKPPSALLRNPHPLENS	60
Vu_eIF4E_Br14_Mulato	MVVEDSQKSTITDDQNPSRVNDDEDLEDGEILEDADDAASAASKPPSALLRNPHPLENS	60
Vu_eIF4E_IT85F-2687	MVVEDSQKSTITDDQNPSRVNDDEDLEDGEILEDADDAASAASKPPSALLRNPHPLENS	60

Vu_eIF4E_Boca_Negra	WTFWFDNPSAKSKQAEWGSSIRPIYTFATVEEFWSIYNNIHHPKSLGVGADFHCIFYKIE	120
Vu_eIF4E_Br14_Mulato	WTFWFDNPSAKSKQDEWGSSIRPIYTFATVEEFWSIYNNIHHPKSLGVGADFHCIFYKIE	120
Vu_eIF4E_IT85F-2687	WTFWFDNRSKSKQDEWGSSIRPIYTFATVEEFWSIYNNIHHPKSLGVVRDFHCIFYKIE	120

Vu_eIF4E_Boca_Negra	PKWEDPICANGGKWTMTFQRGKSDTSWLYTLLAMIGEQQFDHGDEICGAVNVNRNRQDKIS	180
Vu_eIF4E_Br14_Mulato	PKWEDPICANGGKWTMTFQRGKSDTSWLYTLLAMIGEQQFDHGDEICGAVNVNRYRQDKIS	180
Vu_eIF4E_IT85F-2687	PKWEDPICANGGKWTMTFQRGKSDTSWLYTLLAMIGEQQFDHGDEICGAVNVNRNRQDKIS	180

Vu_eIF4E_Boca_Negra	IWTKNASNEAAQMSIGKQWKEFLDYNETIGFIFHEDAKKHERGAKNKYV	229
Vu_eIF4E_Br14_Mulato	IWTKNASNEAAQMSIGKQWKEFLDYNETIGFIFHEDAKKHERGAKNKYV	229
Vu_eIF4E_IT85F-2687	IWTKNASNEAAQMSIGKQWKEFLDYNETIGFIFHEDAKKHERGAKNKYV	229

Figura Suplementar 6: Sequenciamento, tradução e alinhamento das *cds* do gene eIF(iso)4E em três cultivares de feijão-caupi: Boca Negra (suscetível ao CABMV); BR14 Mulato (suscetível) e IT85F-2687 (resistente). A seta verde mostra a única mutação ocorrida no gene que provoca a alteração D31Y das linhagens suscetíveis para a resistente em relação ao *potyvirus* CABMV.

Vu_eIF(iso)4E_Boca_Negra	MATSDEVVAAAPAPEAGLKHKLERKWTFWCDNQSKPKQGAAGTSLRKVYTFDVTVEEFC	60
Vu_eIF(iso)4E_BR14_Mulato	MATSDEVVAAAPAPEAGLKHKLERKWTFWCDNQSKPKQGAAGTSLRKVYTFDVTVEEFC	60
Vu_eIF(iso)4E_IT85F-2687	MATSDEVVAAAPAPEAGLKHKLERKWTFWCYNQSKPKQGAAGTSLRKVYTFDVTVEEFC	60

Vu_eIF(iso)4E_Boca_Negra	LYHEIFKPSKLQINADFHLFKTGVEPKWEDPECANGGKWTVTNIGRKTNLNFWLETLM	120
Vu_eIF(iso)4E_BR14_Mulato	LYHEIFKPSKLQINADFHLFKTGVEPKWEDPECANGGKWTVTNIGRKTNLNFWLETLM	120
Vu_eIF(iso)4E_IT85F-2687	LYHEIFKPSKLQINADFHLFKTGVEPKWEDPECANGGKWTVTNIGRKTNLNFWLETLM	120

Vu_eIF(iso)4E_Boca_Negra	ALIGEQFEDAEDICGVVASVRQWQDKLSLWTKTAANEAQMSIGRKWKEIIDVNDKITYN	180
Vu_eIF(iso)4E_BR14_Mulato	ALIGEQFEDAEDICGVVASVRQWQDKLSLWTKTAANEAQMSIGRKWKEIIDVNDKITYN	180
Vu_eIF(iso)4E_IT85F-2687	ALIGEQFEDAEDICGVVASVRQWQDKLSLWTKTAANEAQMSIGRKWKEIIDVNDKITYN	180

Vu_eIF(iso)4E_Boca_Negra	FHDDSRTKGATKGRYTV	197
Vu_eIF(iso)4E_BR14_Mulato	FHDDSRTKGATKGRYTV	197
Vu_eIF(iso)4E_IT85F-2687	FHDDSRTKGATKGRYTV	197

Figura Suplementar 7. Alinhamento das 42 sequências das proteínas eIF4E e eIF(iso)4E realizado com o auxílio do MEGA 6.0. As colunas destacadas em cinza representam os aminoácidos conservados em todas as sequências. As proteínas de feijão-caupi encontram-se destacadas em negrito. O restante da sequência de aminoácidos destas proteínas encontra-se na próxima imagem.

VuelF4E	H	P	L	E	N	S	W	T	F	W	F	D	N	R	S	A	K	S	K	Q	D	E	W	G	S	S	I	R	P	I	Y	T	F	A	T	V	E	E	F	W	S	I	Y	N	N	I	H	H	P	S	K	L	G	V	R	V	D	F	H	C	F	K	Y	K	I	E	P	K	W	E	D	P	I	C	A	N	G	G	K	W	T	M	T	F	Q	-	-	R	G	K	S	
VuelF(iso)4E	H	K	L	E	R	K	W	T	F	W	C	D	N	Q	S	K	P	K	Q	G	A	A	W	G	T	S	L	R	K	V	Y	T	F	D	T	V	E	E	F	W	C	L	Y	H	E	I	F	K	P	S	K	L	Q	I	N	A	D	F	H	L	F	K	T	G	V	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	V	T	S	N	I	G	R	K	A	N	L
StelF(iso)4E	H	K	L	E	R	K	W	T	F	W	F	D	N	Q	S	K	P	K	Q	G	V	A	W	G	S	S	L	R	K	A	Y	T	F	E	T	V	E	E	F	W	S	L	Y	D	Q	I	F	K	P	S	K	V	T	N	A	D	F	H	L	F	R	A	G	I	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	A	T	S	S	-	-	R	K	A	N	L	
CaelF(iso)4E	H	K	L	E	R	K	W	T	F	W	F	D	N	Q	S	K	P	K	Q	G	A	A	W	G	S	S	L	K	K	A	Y	T	F	D	T	V	E	E	F	W	S	L	Y	D	Q	I	F	K	P	S	K	L	T	V	N	A	D	F	H	L	F	K	A	G	I	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	V	T	S	S	-	-	R	K	A	N	L
CareF(iso)4E	H	K	L	E	R	R	W	T	F	W	F	D	N	Q	S	K	P	K	Q	G	A	A	W	G	T	S	L	R	K	V	Y	T	F	E	T	V	E	E	F	W	C	L	H	D	Q	I	F	K	P	G	K	L	P	G	N	A	D	F	H	L	F	K	D	G	V	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	L	T	I	S	-	-	R	K	P	T	L
CmelF(iso)4E	H	K	L	E	R	K	W	T	F	W	F	D	N	Q	S	R	P	K	Q	G	A	A	W	G	T	S	L	R	K	V	Y	T	F	E	T	V	E	E	F	W	C	L	Y	D	Q	I	F	K	P	S	K	L	P	A	N	A	D	F	H	L	F	K	T	G	V	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	V	T	S	S	-	-	R	K	A	N	L
LzelF(iso)4E	H	R	L	D	K	K	W	T	F	W	F	D	N	Q	S	K	P	K	Q	G	A	G	W	G	N	N	L	R	N	G	Y	T	F	D	T	V	E	E	F	W	C	L	Y	D	Q	V	F	K	P	S	K	L	P	G	N	A	D	F	H	L	F	K	A	G	I	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	V	T	S	S	-	-	R	K	A	N	L
MtelF(iso)4E	H	K	L	E	R	R	W	T	F	W	Y	D	N	Q	S	K	P	K	Q	G	A	A	W	G	T	S	L	R	K	V	Y	S	F	D	T	V	E	E	F	W	C	L	H	D	Q	I	F	K	P	S	K	L	P	S	N	A	D	F	H	L	F	K	D	G	I	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	L	T	S	K	-	-	S	K	G	N	L
PdelF(iso)4E	H	K	L	E	R	K	W	T	F	W	F	D	N	Q	S	K	P	K	Q	G	A	A	W	G	S	S	L	R	K	A	Y	T	F	E	T	V	E	E	F	W	C	L	Y	D	Q	V	F	K	P	S	K	F	P	P	N	A	D	F	H	L	F	R	A	G	V	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	V	T	S	R	-	-	S	K	A	S	L
RcelF(iso)4E	H	K	L	E	R	K	W	T	F	W	F	D	N	Q	S	K	P	K	Q	G	A	A	W	G	T	S	L	R	K	V	Y	T	F	D	T	V	E	E	F	W	C	L	Y	D	Q	I	F	K	P	S	K	L	P	A	S	A	D	F	H	L	F	K	A	G	V	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	S	V	T	C	G	-	-	R	K	A	L	L
StelF(iso)4E	H	K	L	E	R	K	W	T	F	W	F	D	N	Q	S	K	P	K	Q	G	A	A	W	G	S	S	L	R	K	A	Y	T	F	E	T	V	E	E	F	W	S	L	Y	D	Q	I	F	K	P	S	K	L	T	V	N	A	D	F	H	L	F	K	A	G	I	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	A	T	S	S	-	-	R	K	A	N	L
TcelF(iso)4E	H	K	L	E	R	K	W	T	F	W	F	D	N	Q	S	K	P	K	Q	G	A	A	W	G	T	S	L	R	K	V	Y	T	F	D	T	V	E	E	F	W	C	L	Y	D	Q	I	F	K	P	S	K	L	P	G	N	A	D	F	H	L	F	K	A	G	I	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	S	V	T	I	N	-	-	R	K	P	I	L
VvelF(iso)4E	H	R	L	E	R	K	W	T	F	W	F	D	N	Q	S	K	P	K	Q	G	A	A	W	G	T	S	L	R	K	A	Y	T	F	E	T	V	E	E	F	W	C	L	Y	D	Q	I	F	K	P	S	K	L	P	A	N	A	D	F	H	L	F	K	A	G	V	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	V	A	S	S	-	-	S	K	A	S	L
AtelF(iso)4E	H	K	L	E	R	K	W	S	F	W	F	D	N	Q	S	-	-	K	G	A	A	W	G	A	S	L	R	K	A	Y	T	F	D	T	V	E	D	F	W	G	L	H	E	T	I	F	Q	T	S	K	L	T	A	N	A	E	I	H	L	F	K	A	G	V	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	W	V	T	A	N	R	K	E	A	L		
CselF(iso)4E	H	K	L	E	R	K	W	T	F	W	F	D	N	Q	S	R	P	K	Q	G	A	A	W	G	T	S	L	R	K	V	Y	T	F	D	T	V	E	E	F	W	C	L	Y	D	Q	I	F	K	P	S	K	L	P	A	N	A	D	F	H	L	F	K	T	G	V	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	V	T	S	S	-	-	R	K	A	N	L
PselF(iso)4E_1	H	K	L	E	R	R	W	T	F	W	F	D	N	Q	S	K	P	K	Q	G	A	A	W	G	T	T	L	R	K	V	Y	S	F	D	T	V	E	E	F	W	L	H	D	Q	I	F	K	P	S	K	L	P	G	N	A	D	F	H	L	F	K	D	G	V	E	P	K	W	E	D	P	L	C	A	S	G	G	K	W	T	L	T	S	K	-	-	G	K	G	N	L	
PselF(iso)4E_2	H	K	L	E	R	R	W	T	F	W	F	D	N	Q	S	K	P	K	Q	G	A	A	W	G	T	T	L	R	K	V	Y	S	F	D	T	V	E	E	F	W	L	H	D	Q	I	F	K	P	S	K	L	P	G	N	A	D	F	H	L	F	K	D	G	V	E	P	K	W	E	D	P	L	C	A	S	G	G	K	W	T	L	T	S	K	-	-	G	K	G	N	L	
PcelF(iso)4E	H	K	L	E	R	K	W	T	F	W	C	D	N	Q	S	K	P	K	Q	G	A	A	W	G	T	S	L	R	K	V	Y	T	F	D	T	V	E	E	F	W	C	L	Y	D	Q	I	F	K	P	G	K	L	Q	N	N	A	D	F	H	L	F	K	T	G	V	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	V	T	S	N	I	G	R	K	A	N	L
PvelF(iso)4E	H	K	L	E	R	K	W	T	F	W	C	D	N	Q	S	K	P	K	Q	G	A	A	W	G	T	S	L	R	K	V	Y	T	F	D	T	V	E	E	F	W	C	L	Y	D	Q	I	F	K	P	S	K	L	Q	N	N	A	D	F	H	L	F	K	T	G	V	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	V	T	S	N	I	G	R	K	A	N	L
GmelF(iso)4E	H	K	L	E	R	K	W	T	F	W	C	D	N	Q	S	K	P	K	Q	G	A	A	W	G	T	S	L	R	K	V	Y	T	F	D	T	V	E	E	F	W	C	L	Y	D	Q	V	F	K	P	S	K	L	Q	I	N	A	D	F	H	L	F	K	T	G	I	D	P	K	W	E	D	P	E	C	A	N	G	G	K	W	S	I	T	S	N	S	G	R	K	A	N	L
AhelF(iso)4E	H	K	L	E	R	K	W	T	F	W	F	D	N	Q	S	K	P	K	Q	G	A	A	W	G	T	S	L	R	E	V	Y	T	F	D	T	V	E	E	F	W	C	L	Y	D	Q	V	F	K	P	S	K	L	P	G	N	A	D	F	H	L	F	K	T	G	I	E	P	K	W	E	D	P	E	C	A	K	G	G	K	W	T	V	T	S	N	-	-	R	K	A	N	L
VrelF(iso)4E	H	K	L	E	R	K	W	T	F	W	C	D	N	Q	S	K	P	K	Q	G	A	A	W	G	T	S	L	R	K	V	Y	T	F	D	T	V	E	E	F	W	C	L	Y	D	Q	I	F	K	P	K	D	L	Q	N	N	A	D	F	H	L	F	K	T	G	I	E	P	K	W	E	D	P	E	C	A	N	G	G	K	W	T	V	T	S	N	I	G	R	K	D	N	L
VrelF4E	H	P	L	E	N	S	W	T	F	W	F	D	S	P	S	A	K	S	K	Q	A	A	W	G	S	S	I	R	P	I	Y	T	F	A	T	V	E	E	F	W	S	I	Y	N	N	I	H	H	P	S	K	L	G	V	G	A	D	F	H	C	F	K	Y	K	I	E	P	K	W	E	D	P	I	C	A	N	G	G	K	W	T	M	T	F	Q	-	-	-	R	N	K	S
AtelF4E	H	P	L	E	H	S	W	T	F	W	F	D	N	P	A	V	K	S	K	Q	T	S	W	G	S	S	L	R	P	V	T	F	S	T	V	E	E	F	W	S	L	Y	N	N	M	K	H	P	S	K	L	A	H	G	A	D	F	Y	C	F	K	H	I	E	P	K	W	E	D	P	I	C	A	N	G	G	K	W	T	M	T	F	P	-	-	-	K	E	K	S		
PvelF4E_1	H	P	L	E	N	S	W	T	F	W	F	D	N	P	S	A	K	S	K	Q	A	A	W	G	S	S	I	R	P	I	Y	T	F	S	T	V	E	E	F	W	S	I	Y	N	N	I	H	H	P	S	K	L	G	V	G	A	D	F	H	C	F	K	H	K	I	E	P	K	W	E	D	P	I	C	A	N	G	G	K													

VuelF4E	DTSWLYTLLAMIGE QFDHGDEICGAVVNVR--NRQDKISIWTKNASNEAAQMSIGKQWKFEFLDYNETIGFIFHEDAKKHERGAKNKYVI
VuelF(iso)4E	ENMWLETLMALIGE QFEDAEDICGVVASVR--QWQDKLSLWTKTAANEAAQMSIGRKWKEIIDVNDKITYNFHDDSRTKGATKGRYTV-
SlelF(iso)4E	ETMWLETLMALVGE QFDESEDICGVVASVR--RSQDKLSLWTKTATNEAAQMGIGRKWKEIIDA-EKISYSFHHDDSKRERSAKSRYTV-
Caelf(iso)4E	ETMWLETLMALVGE QFDDSEDICGVVASVR--RSQDKLSLWTKTATNEAAQMGIGRKWKEIIDT-EKISYSFHHDDSKRERSAKSRYTV-
Carelf(iso)4E	DAMWLETLMALIGE QFGDSEDICGVVASVR--QWQDKLSLWTKTAANESVQMSIGRKWKEIIESNOKMTYNFHHEDAK-NRGAKARYTV-
Cmelf(iso)4E	DNMWLETLMALIGE QFEESDEICGVVASVR--QRQDKLALWTKTATNEAAQMSIGRKWKEIIDVNDKISFSFHHEDLRREKSAKARYSV-
Lzelf(iso)4E	ETMWLETLMALIGE QFDEADEICGVVASVR--QRQDKLSLWTKNAANEAAQMSIGRKWKEIIDVTDKITYNFHDDSKTRTS-KGRYSV-
Mtelf(iso)4E	DTMWLETLMALIGE QFGDTEICGVVASVR--QWQDKLSLWTKTAANESNQMSIGRKWKEIIDVSDKMTYNFHHEDAK-TRGAKARYTV-
Pdelf(iso)4E	DTMWLETLMALIGE QFDEADEICGVVASVR--QRQDKLALWTRNAANEAAQMGIGRKWKEIIDVTDKITYSFHHDDSKRERSAKPRYNV-
Rcelf(iso)4E	DTMWLETLMALIGE QFDDSDICGVVASVR--PRGDKLALWTRTATNEAAQMGVGGKWKEIIDVTDKITYSFHHDDSRERSVKSRYNV-
Stelf(iso)4E	ETMWLETLMALVGE QFDESEDICGVVASVR--RSQDKLSLWTKTATNEAAQMGIGRKWKEIIDT-EKISYSFHHDDSKRERSAKSRYTV-
Tcelf(iso)4E	DNMWLETLMALIGE QFDEADEICGVVASVR--QRQDKLALWTKTATNEAAQMGIGRKWKEIIDFNDKITYSFHHDDSRERSAKGRYNV-
VvelF(iso)4E	DTMWLETLMALIGE QFDEADEICGVVASVR--QRQDKLALWTKTATNEAAQMSIGRKWKEIIDVTDKITYSFHHDDSRERSVKVRYNV-
Atelf(iso)4E	DKGWLETLMALIGE QFDEADEICGVVASVRPQSKQDKLSLWTRTKSNEAVLMGIGKKWKEILDVTDKITFNNHDDSRRSRTV-----
Cself(iso)4E	DNMWLETLMALIGE QFEESDEICGVVASVR--QRQDKLALWTKTATNEAAQMSIGRKWKEIIDVTDKISFSFHHEDLRREKSAKARYSV-
Pself(iso)4E_1	DTMWLETLMALIGE QFGDSEDICGVVASVR--QWQDKLSLWTKTAANESVQMSIGRKWKEIIDVSDKMTYNFHHEDAK-TRGAKARYTV-
Pself(iso)4E_2	DTMWLETLMALIGE QFGDSEDICGVVASVR--QWQDKLSLWTKTAANESVQMSIGRKWKEIIDVSDKMTYNFHHEDAK-TRGAKARYTV-
Pcelf(iso)4E	DNMWLETLMALIGE QFEDAEDIC-----
PvelF(iso)4E	DNMWLETLMALIGE QFEDAEDICGVVASVR--QWQDKLSLWTKTAANEAAQMSIGRKWKEIIDVNDKITYNFHDDSRTKGATKGRYTV-
Gmelf(iso)4E	DNMWLETLMALIGE QFEDAEDICGVVASVR--QWQDKLSLWTKTAANEAAQMSIGRKWKEIIDVNDKITYNFHDDSRSGATKGRYTV-
Ahelf(iso)4E	DNMWLETLMALIGE QFDDAEDICGVVASVR--QRQDKLSLWTKTAANEATHMGIGRKWKEIIDVTDKIITYNFHDDSRTRSSKSRYSV--
VrelF(iso)4E	EKMWLETLMALIGE QFEDAEDICGVVASVR--QWQDKLSLWTKTAANEAAQMSIGRKWKEIIDVNDKITYNFHDDSRTKGATKGRYTV-
VrelF4E	DTSWLYTLLAMIGE QFDHGDEICGAVVNVR--NRQDKISIWTKNASNEAAQMSIGKQWKFEFLDYNETIGFIFHEDAMKHERSAKNKYVI
Atelf4E	DKSWLYTLLALIGE QFDHGDEICGAVVNIR--GKQERISIWTKNASNEAAQVSGKQWKFEFLDYNNISIGFIFHEDAKKLDRNAKNAYTA
PvelF4E_1	DTSWLYTLLAMIGE QFDYGDEICGAVVNVR--NRQDKISIWTKNASNEAAQMSIGKQWKFEFLDYNEPIGFIFHEDAKKHERSAKNKYVV
PvelF4E_2	DTSWLYTLLAMIGE QFDYGDEICGAVVNVR--NRQDKISIWTKNASNEAAQMSIGKQWKFEFLDYNEPIGFIFHEDAKKHER-----
Pself4E	DTSWLYTLLAMIGE QFDHGDEICGAVVNVG--GRAEKISIWTKNASNEAAQVSGKQWKFEFLDYNETMGFIFHDDARKLDRNAKNKYVV
Gmelf4E_2	DTSWLYTLLAMIGE QFDHGDEICGAVVNVR--SRQEKIALWTKNASNEAAQVSGKQWKFEFLDYNDTIGFIFHEDAKKLDRGAKNKYVV
Gmelf4E_1	DTSWLYTLLAMIGE QFDHGDEICGAVVNVR--SRQEKIALWTKNASNEAAQVSGKQWKFEFLDYNDTIGFIFHEDAKKLDRGAKNKYVV
Ahelf4E	DTSWLYTLLGMIGE QFDHGDEICGAVVNVR--NRQEKIALWTKNASNEAAQVSGKQWKFEFLDYNDTIGFIFHEDAKKHDRAAKNKYVI
Cself4E	DNGWLYTLLAMIGE QFDCGDEICGAVVNVR--SGQDKISIWTKNASNEAAQASIGKQWKFEFLDYNESIGFIFHDDAKKFDRAKNKYMV
Lself4E	DTCWLYTLLAMIGE QFDHGDDICGAVVNVR--ARQEKIALWTKNANESAQLSIGKQWKFEFLDYNDTIGFIFHEDAKTLDRSAKNKYTV
Caelf4E	DTSWLYTLLAMIGHQFDHEDEICGAVVSVR--GKGKISLWTKNAANETAQVSGKQWKQFLDYSDSVGFIHDDAKRLDRNAKNRYTV
Pself4E	DTSWLYTLLAMIGE QFDHGDEICGAVVNVR--GRAEKISIWTKNASNEAAQVSGKQWKFEFLDYNETMGFIFHDDARKLDRNAKNKYVV
Mtelf4E	DTSWLYTLLAMIGE QFDHGDEICGAVVNVR--SRAEKISIWTKNAANEAAQVSGKQWKFEFLDYNETIGFIFHDDAKKLDRAAKNKYVV
SlelF4E	DTCWLYTLLALIGE QFDYGDEICGAVINVR--VRQEKIALWTRNAANETAQVSGKQWKFEFLDYNDTIGFIFHDDAKKLDRAAKNRYSV
VvelF4E	DTCWLYTLLAMIGE QFDHGDEICGAVINVR--ARQEKIALWTKNASNEAAQLSIGKQWKFEFLDYSESIGFIFHEDAKKLDRAAKNRYTI
Pself4E	DTSWLYTLLAMIGE QFDHGDEICGAVVNVR--GRAEKISIWTKNASNEAAQVSGKQWKFEFLDYNETMGFIFHDDARKLDRNAKNKYVV
Pdelf4E	DTSWLYTLLGMIGE QFDHGDEICGAVVNVR--NRQEKISIWTKNAINEAAQLSIGKQWKGLDYNETIGFIFHEDAMRHERSAKNKYVV
Ntelf4E	DTCWLYTLLAMIGE QFDCGDEICGAVINVR--VRQEKIALWTRNAANETAQVSGKQWKFEFLDYNDSVGFIHDDAKKLDRAAKNRYSV
Caelf4E	DTSWLYTLLAMIGE QFDHGDEICGAVVNVR--NRQEKISIWTKNASNEAAQVSGKQWKFEFLDYNETIGFIFHEDAKKLDRNAKNKYVV
Cpelf4E	DTCWLYTLLAMIGE QFDHGDEICGVVNVR--GRQEKIALWTKNAANEAAQMSIGKQWKFEFLDYNDTMGFIFHEDAKKLERAANKRYSV

Figura Suplementar 8: Modelos *in silico* das proteínas eIF4E (lado esquerdo) e eIF(iso)4E (lado direito) gerados a partir do Modeller 9.10. Conforme esperado, as proteínas possuem três alfa-hélices grandes (violeta) e no mínimo sete folhas-betas grandes (em amarelo).

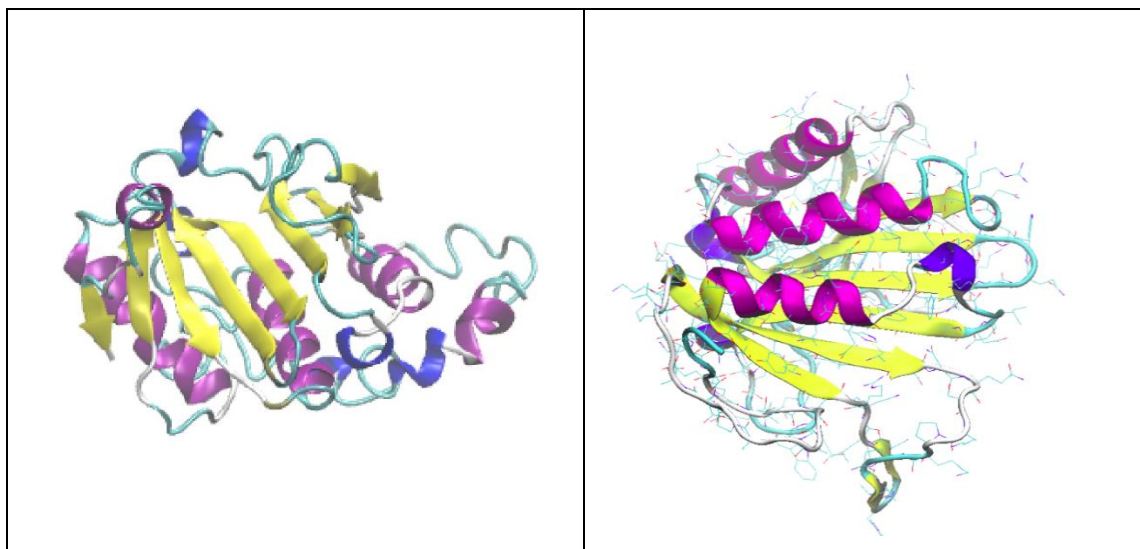


Figura Suplementar 9: Figuras referentes a validação dos modelos teóricos dos fatores de iniciação 4E de *Vigna unguiculata*. No quadro superior estão as figuras representando o gráfico de Ramachandran na avaliação estereoquímica dos modelos, já no quadrante inferior as figuras relacionadas com o Z-score que avalia a qualidade do dobramento referente ao banco de proteínas de tamanho similares.

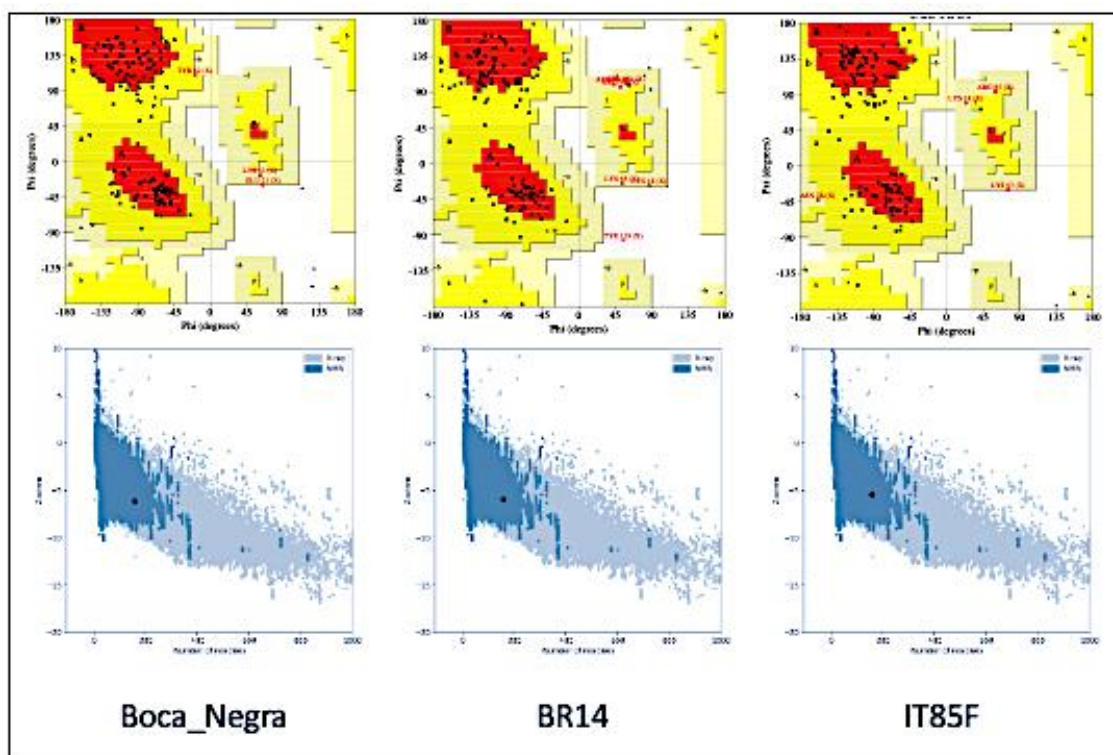


Figura Suplementar 10: Figuras relacionados à simulação em dinâmica molecular dos modelos tridimensionais teóricos dos fatores de iniciação 4E de *Vigna unguiculata*. Em **A**- o RMSD do modelo com o tempo de 100 ns dos três modelos: Boca Negra (em preto), BR14 Mulato (em azul) e IT85F-2687 (em vermelho). Em **B**, o raio de giro dos três modelos Boca Negra (em preto), BR14 Mulato (em vermelho) e IT85F-2687 (em verde) Em **C**- o RMSF dos modelos representados pela flutuação de aminoácidos pelo tempo em 100 ns, ao lado as figuras correspondentes ao *b-factor* ou espalhamento dos átomos, dos três modelos relacionados, quanto mais próximo do vermelho, maior mobilidade dinâmica dos átomos e mais próximo do azul, menor mobilidade dinâmica dos átomos.

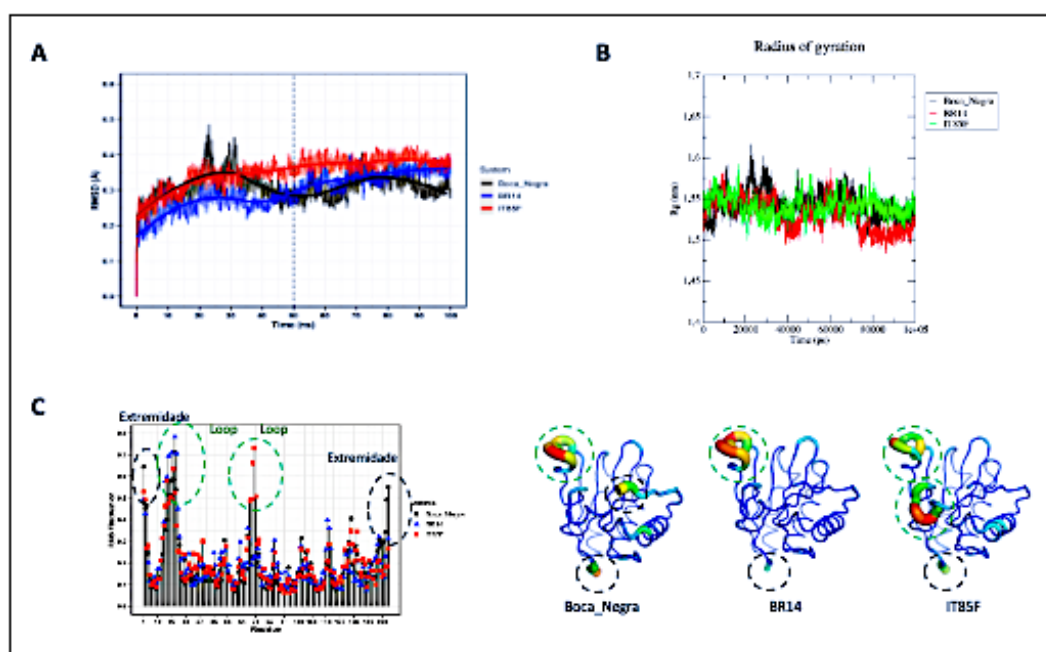


Figura Suplementar 11. Gel de agarose a 2% para confirmação da ligação dos oligos pela diferença de tamanho da sequência ligada comparativamente à linear. No poço 1 os oligos ligados referentes ao gRNA tendo como alvo o gene *eIF4E*, o poço 2 os oligos ligados do gRNA do *eIF(iso)4E* e no poço 3 o oligo linear.

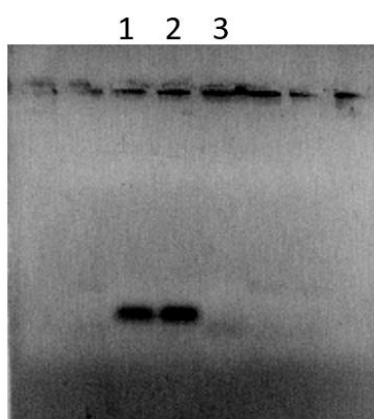
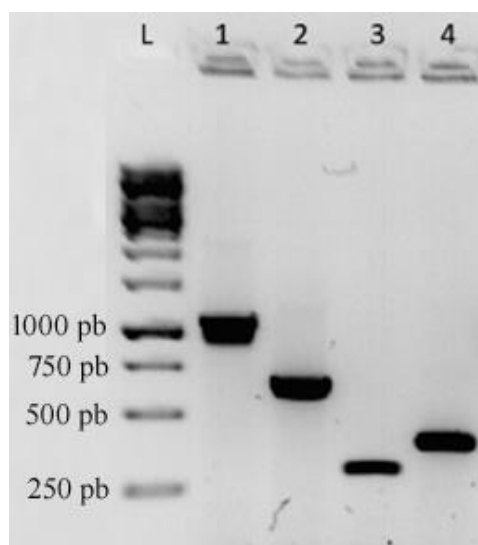


Figura Suplementar 12. Resultado do sequenciamento com o primer SP6 para confirmação da integração do gRNA do eIF(iso)4E.

>pGEM-T sgRNA_eIF(iso)4E

```
AGAAAATGCATCACGCGTTGGGAGCTCTCCCATATGGTTCGACCTGCAGGCGGCCGCGAATTC
ACAGTGATTACTAGTGTGGTTGGCATGCACATACAAATGGACGAACGGATAAACCTTTTCACG
CCCTTTTAAATATCCGATTATTCTAATAAACGCTCTTTTCTCTTAGGTTTACCCGCCAATATATC
CTGTCAAACACTGATAGTTTAAACGTCATTGTTTTAGTTCCACCACGATTATATTTGAAATTTAC
GTGAGTGTGAGTGAGACTTGACATAAGAAAATAAAATCTTTAGTTGGGAAAAAATTCAATAATAT
AAATGGGCTTGAGAAGGAAGCGAGGGATAGGCCTTTTTCTAAAATAGGCCCATTTAAGCTATT
AACAACTTTCAAAAGTACCACAGCGCTTAGGTAAAGAAAGCAGCTGAGTTTATATATGGTTAGA
GTCAAGTAGTGATTGTTGGGGAACCTCTCTGCGCAGTTTTAGAGCTAGAAATAGCAAGTTAA
AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTTTTCTAGACC
CAGCTTTCTTGACAAAGTTGGCATTAGTTTAACTGAAGGCGGGAAACGACAATCTGATCCA
AGCTCAAGCTGCTCTAGTCCCCCACCCTCGGGGCTGGGAAATTGGTGGGGAACGGGCCA
CCCGGTTGCGGGCCCTCCTCCGCTAATGCCGCCCGCCGGGGCGAAAGGGGGGAAGGTGGC
TGCCAGGGCCGACTAGGCCGGGGGGCACCCCATCCGAATTCCCCCGCGGGCCCCGCCATG
GGCCGGCCCCGGGGGGGCAGGCCGGCCGATCGGGGGCCCCCATTCCGGCCCCCCCCGCGGG
GGGCGGCCCGCCAAGTAACCGCCCCGCCCGGGGGCCCGGTACGGGTTTCCACCCGCCCG
GCCCCGGGGGAACGGGGGGGCCAAAACCGCCTGGGGGCGGGCTTACCCCCAAAACCAAGGA
GGGGCCGGGCGAGGGGGCGAAGGGGAGCCCTGGGCCTTCCCCCTTTGGGGAGGGCCCCAG
GGCCCCGGGGCCCCGGATTCAAGCCCGGCCCGGCAAGGCCGGGGCCCCCACCAGCCCCC
CCCCGGGCACCCCGGGCCATTCCCCTTTTCCCAAGGGGAAAATTCCCGGGGTTCCGCCCCCT
GGGCCCCGGGGACCAAAAAGGGGGGAAACCCCGGGAAAACGGGGGCTTAAAAACGGGGCC
CAAATGGCCCAGGGGGGGTTCCACAGGCAACAAAAGGGGTCAACCCGCCCCCCCCCGCCCC
TGGGGGGGGATCCCCAGAGCACACAGGGGGGGGGG
```

Figura Suplementar 13. Confirmação do vetor final pAHASCas9_sgRNA-eIF(iso)4E por PCR. Poço 1 primers ACT2 / Cas9 336c que amplifica um fragmento de 1127 pb. Poço 2, amplificação com o par de primer gRNA-speI / gRNA-KpnI com fragmento de 684 pb. Poço 3, amplificação com par de primer gRNA-KpnI / Vu_eIF(iso)4E_F com fragmento de 303 pb. No poço 4, amplificação com o par de primers gRNA-speI / Vu_eIF(iso)4E_R gerando um fragmento de 409 pb.



5 DISCUSSÃO GERAL

O processo de transformação de feijão-caupi ainda não é trivial, sendo a eficiência ainda um grande entrave para obtenção de linhagens elites (Citadin *et al.*, 2011). Em relação às leguminosas, existe uma dificuldade maior relacionada à recalcitrância destes vegetais na cultura *in vitro* (Mandal *et al.*, 2020). Em relação ao feijão-caupi, a eficiência da transformação por biobalística é cerca de 1% (Ivo *et al.*, 2008; Cruz e Aragão, 2014), resultado semelhante obtido no presente trabalho. Em relação à transformação via *Agrobacterium tumefaciens* em feijão-caupi, trata-se da abordagem com o maior número de trabalhos publicados, onde há registro de uma eficiência superior a 3% (Kumar *et al.*, 2017).

A maioria dos trabalhos de transformação do feijão-caupi buscam um aumento da eficiência do processo da transformação (Bett *et al.*, 2019), havendo alguns poucos trabalhos buscando o desenvolvimento de linhagens elites para resistência a insetos (Solleti *et al.*, 2008b; Higgns *et al.*, 2010; Bakshi *et al.*, 2011), vírus (Cruz e Aragão, 2014; Kumar *et al.*, 2017) e tolerância à seca (Mishra *et al.*, 2014). Em relação à edição genômica por CRISPR/Cas9, até o presente há apenas uma publicação envolvendo edição do genoma de feijão-caupi em raízes via *A. rhizogenes* visando ao nocaute do gene que promove a nodulação das raízes, para avaliar a eficiência da técnica (Ji *et al.*, 2019).

Na presente tese foram desenvolvidas linhagens de feijão-caupi transgênicas com uma defensina que tem atividade antifúngica contra a *Macrophomina phaseolina*, bem como plantas de feijão-caupi com edição de seu genoma envolvendo a sequência de interesse, mas sem sucesso no nocaute do gene eIF(iso)4E, o que pode ocorrer nas próximas gerações, visto que o cassete contendo os genes da Cas9 e sgRNA continuam integrado no genoma das

plantas transformantes, permitindo a ocorrência de novos processos de edição assim como possivelmente irá aumentar a eficiência da edição (Bao *et al.*, 2019).

A expressão de transgenes tem demonstrado uma grande variação em diferentes eventos de transformação (Eissa *et al.*, 2017). Nesse contexto, o feijão-caupi com inserção da defensina do melão-de-são-caetano mostrou padrões distintos em relação às quatro linhagens testadas por qRT-PCR, onde um dos eventos de transformação mostrou uma expressão estatisticamente inferior às outras três testadas. O nível de expressão das defensinas foi o dobro da expressão das chitinases de Eissa e colaboradores (2017), o que pode ser devido ao efeito de posição no genoma, considerando-se variações na posição onde o transgene se insere (Howarth *et al.*, 2005). Outro fator envolve o promotor utilizado, uma vez que a defensina foi dirigida pelo promotor CaMV35S e a chitinase pela ubiquitina do milho (Howarth *et al.*, 2005).

As defensinas mostram como principal atividade a ação antimicrobiana contra fungos e bactérias (Gaspar *et al.*, 2014), porém existem poucos registros de atividade antiviral (Brodgen, 2005) e de aumento de tolerância a estresses abióticos (Kumar *et al.*, 2019). A defensina de melão-de-são-caetano mostrou atividade antifúngica contra a podridão cinzenta do caule, porém não teve atividade antiviral, considerando-se as estirpes testadas no presente trabalho.

6 CONCLUSÕES

O processo de transformação de várias leguminosas como feijão-caupi, feijão comum e amendoim, ainda está longe de ser uma prática corriqueira de laboratório, porém foi possível realizar a transformação de feijão-caupi utilizando as duas principais ferramentas da transgenia: a transformação via biobalística e via *Agrobacterium*.

O potencial da defensina de *M. charantia* foi evidenciado no bioensaio com o fungo *M. phaseolina*. Este é o primeiro registro da transformação genética do feijão-caupi com uma defensina, assim como a demonstração da atividade *in vivo* de uma defensina contra a podridão cinzenta do caule.

A edição do genoma do feijão-caupi foi realizada a partir da transformação de embriões com vetor de CRISPR/Cas9. Embora o gene-alvo tenha sido modificado, o tipo de mutação (substituição de bases) e o local da mutação (anterior ao gRNA) ocorreu fora da posição esperada. Vale destacar que o protocolo utilizado na transformação com o vetor de CRISPR/Cas9 foi via *Agrobacterium tumefaciens*, correspondendo a uma adaptação com várias diferenças em relação aos dois protocolos utilizados como base do processo.

Foram identificados polimorfismos na sequência codificante do fator de iniciação eIF4E, quando comparadas três cultivares analisadas. A análise dos produtos gênicos e a modelagem de proteínas indicam que o acesso resistente IT85F-2687 apresenta a variante mais divergente, possivelmente responsável pela resistência, o que demanda validação experimental.

REFERÊNCIAS

- Abdallah, N.A., Shah, D., Abbas, D., and Madkour, M. (2010) Stable integration and expression of a plant defensin in tomato confers resistance to *Fusarium* wilt. *GM Crops* 1: 344–350.
- Antoine, B., La Salle, T.J.B., Zakaria, D., Zakaria, K., Leandre, P., and Mahamadou, S. (2016) Inheritance and the Allelic Relationship of Resistance to *Cowpea Aphid Borne Mosaic Virus* (CABMV) in Two Cowpea Genotypes, K VX640 and K VX396-4-5-2D, in Burkina Faso. *Int J Curr Microbiol App Sci* 5(8): 285-292.
- Anuradha, T.S., Divya, E.K., Jami, E.S.K., and Kirti, E.P.B. (2008) Transgenic tobacco and peanut plants expressing a mustard defensin show resistance to fungal pathogens. *Plant Cell Rep* 27: 1777–1786.
- Aragão, F.J.L., Barros, L.M.G., Brasileiro, A.C.M., Ribeiro, S.G., Smith, F.D., Sanford, J.C., Faria, J.C., and Rech, E.L. (1996) Inheritance of foreign genes in transgenic bean (*Phaseolus vulgaris* L.) co-transformed via particle bombardment. *Theor Appl Genet* 93: 142–150.
- Athayde Sobrinho, C. (2005) Patossistema caupi X *Macrophomina phaseolina*: método de detecção em sementes, esporulação e controle do patógeno. Tese (Doutorado em Agronomia) - Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, 150 f.
- Ayeleke, D.A., Salaudeen, M.T., Adama, C.J., Abdullahi, A.A., and Ibrahim, A.D. (2016) Applications of area under disease progress curves in assessing resistance to cowpea aphid-borne mosaic virus infection in groundnut. *Int J Adv Biotechnol Res* 7(1): 88 - 97.
- Bakshi, S., Sadhukhan, A., Mishra, S., and Sahoo, L. (2011) Improved *Agrobacterium*-mediated transformation of cowpea via sonication and vacuum infiltration. *Plant Cell Rep* 30: 2281–2292.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007) CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* 315: 1709-1712.
- Barros, G.B., Nogueira, M.S.R., Oliveira, C.R.R., Freire-Filho, F.R., Ribeiro, V.Q., Veiga, C.F.M., Brioso, P.S.T., and Eiras, M. (2013) Obtaining cowpea plants resistant to *Cowpea Severe Mosaic Virus* and *Cowpea Aphid-Borne Mosaic Virus*. *Summa Phytopathol* 39(2).
- Bashir, M.I., and Hampton, R.O. (1996) Sources of genetic resistance in cowpea (*Vigna unguiculata* (L.) Walp.) to cowpea aphid-borne mosaic potyvirus. *European J Plant Pathol* 102: 411-419.
- Bastet, A., Robaglia, C., and Gallois, J. (2017) eIF4E Resistance: Natural Variation Should Guide Gene Editing. *Trends Plant Sci* 22(5): 411-419.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N.J., and Nekrasov, V. (2015) Editing plant genomes with CRISPR/Cas9. *Curr Opin Biotechnol* 32: 76-84.

Bett, B., Gollasch, S., Moore, A., Harding, R., and Higgins, T.J.V. (2019) An Improved Transformation System for Cowpea (*Vigna unguiculata* L. Walp) via Sonication and a Kanamycin-Geneticin Selection Regime. *Front Plant Sci* 10: 219.

Brogden, K.A. (2005) Antimicrobial Peptides: pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol* 3: 238–250.

Burketová, L., Trdá, L., Ott, P., and Valentová, O. (2015) Bio-based resistance inducers for sustainable plant protection against pathogens. *Biotechnol Adv* 33: 994-1004.

Chamarthi, S.K., Belko, N., Togola, A., Fatokun, C.A., and Boukar, O. (2019) Genomics-Assisted Breeding for Drought Tolerance in Cowpea. In: Rajpal V., Sehgal D., Kumar A., Raina S. (eds) *Genomics Assisted Breeding of Crops for Abiotic Stress Tolerance*, Vol. II. Sustainable Development and Biodiversity, vol 21. Springer, Cham.

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., Sherman, A., Arazi, T., and Gal-On, A. (2016) Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol Plant Pathol* 17(7): 1140–1153.

Chen, S.C., Liu, A.R., and Zou, Z.R. (2006) Overexpression of glucanase gene and defensin gene in transgenic tomato enhances resistance to *Ralstonia solanacearum*. *Russia J Plant Physiol* 53: 756–763.

Choi, M.S., Kim, Y.H., Park, H.M., Seo, B.Y., Jung, J.K., Kim, S.T., Kim, M.C., Shin, D.B., Yun, H.T., Choi, I.S., Kim, C., and Lee, J.Y. (2009) Expression of BrD1, a Plant Defensin from *Brassica rapa*, Confers Resistance against Brown Planthopper (*Nilaparvata lugens*) in Transgenic Rices. *Mol Cells* 28:131-137.

Citadin, C.T., Ibrahim, A.B., and Aragão, F.J.L. (2011) Genetic engineering in cowpea (*Vigna unguiculata*): History, status and prospects. *GM Crops* 2(3): 144-149.

Citadin, C.T., Cruz, A.R.R., and Aragão, F.J.L. (2013) Development of transgenic imazapyr-tolerant cowpea (*Vigna unguiculata*). *Plant Cell Rep* 3: 537–543.

Conab. ACOMPANHAMENTO DA SAFRA BRASILEIRA. V. 6 - SAFRA 2018/19 - N. 11 - Décimo primeiro levantamento | AGOSTO 2019. ISSN 2318-6852 Acomp. safra bras. grãos, v. 11 Safra 2018/19 - Décimo Primeiro levantamento, Brasília, p. 1-45, agosto 2019.

Cruz, A.R.R., and Aragão, F.J.L. (2014) RNAi-based enhanced resistance to Cowpea Severe Mosaic Virus and Cowpea Aphid-Borne Mosaic Virus in transgenic cowpea. *Plant Pathol* 63: 831–837.

- Darwish, N.A., Khan, R.S., Ntui, V.O., Nakamura, I., Mii, M. (2014) Generation of selectable marker-free transgenic eggplant resistant to *Alternaria solani* using the R/RS site-specific recombination system. *Plant Cell Rep*, 33: 411-421.
- Dhanasekar, P., and Reddy, K.S. (2015) Serological Screening of Cowpea Genotypes for Resistance against *Cowpea Aphid Borne Mosaic Virus* Using DAS-ELISA. *Asian J Plant Pathol* 9(2): 83-90.
- Eissa, H.F., Hassanien, S.E., Ramadan, A.M., El-Shamy, M.M., Saleh, O.M., Shokry, A.M., ... and Bahieldin, A. (2017) Developing transgenic wheat to encounter rusts and powdery mildew by overexpressing barley chi26 gene for fungal resistance. *Plant Methods* 13(1).
- Fang, J., Chao, C.T., Roberts, P.A., and Ehlers, J.D. (2007) Genetic diversity of cowpea [*Vigna unguiculata* (L.) Walp.] in four West African and USA breeding programs as determined by AFLP analysis. *Genet Resour Crop Evol* 54: 1197–1209.
- Freire-Filho, F.R., Ribeiro, V.Q., Rocha, M.M., Silva, K.J.D., Nogueira, M.S.R., and Rodrigues, E.V. (2011) Produção, melhoramento genético e potencialidades do feijão-caupi no Brasil. In: *Reunião de Biofortificação*, 4, Teresina: Embrapa Meio-Norte.
- Frota, K.G.M., Soares, R.A.M., and Arêas, J.A.G. (2008) Composição química do feijão caupi (*Vigna unguiculata* L. Walp), cultivar BRS-Milênio. *Ciênc Tecnol Aliment* 28: 470-476.
- Gao, Ai-Guo., Hakimi, S.M., Mittanck, C.A., Wu, Y., Woerner, B.M., Stark, D.M., Shah, D.M., Liang, J., and Rommenset, C.M.T. (2000) Fungal pathogen protection in potato by expression of a plant defensin peptide. *Nat Biotechnol* 18: 1307–1310.
- Garcia, J.A., Hille, J., and Goldbach, R. (1987) Transformation of cowpea *Vigna unguiculata* with a full-length DNA copy of cowpea mosaic virus M-RNA. *Plant Sci* 48: 89-98.
- Gaspar, Y.M., Mckenna, J.Á., McGinness, B.S., Hinch, J., Poon, S., Connelly, A.A., Anderson, M.A., and Heath, R.L. (2014) Field resistance to *Fusarium oxysporum* and *Verticillium* in transgenic cotton expressing the plant defensin NaD1. *J Exp Bot* 65(6): 1541–1550.
- Gomes-Silva, F., Almeida, C.M.A., Silva, A.G., Leão, M.P.C., Silva, K.P., Oliveira, L.G., Silva, M.V., Costa, A.F., and Lima, V.L.M. (2017) Genetic Diversity of Isolates of *Macrophomina* phase. *J Agric Sci* 9(11).
- Güneş, H., Alper, M., and Çelikoğlu, N. (2019) Anticancer effect of the fruit and seed extracts of *Momordica charantia* L. (Cucurbitaceae) on human cancer cell lines. *Trop J Pharm Res* 18 (10): 2057-2065.
- Hall, A.E. (2012) Phenotyping Cowpeas for Adaptation to Drought. *Front Physiol* 3: 155.

Higgins, T.J.V., Gollasch, S., Molvig, L., Moore, A., Popelka, C., Watkins, P., ... and Murdocket, L. (2010) Genetic transformation of cowpea for protection against bruchids and caterpillars. In: Fifth World Cowpea Research Conference, 50. IITA, Sally-Senegal. 2010.

Howarth, J.R., Jacquet, J.N., Doherty, A., Jones, H.D., and Cannell, M.E. (2005) Molecular genetic analysis of silencing in two lines of *Triticum aestivum* transformed with the reporter gene construct pAHC25. *Ann Appl Biol* 146(3): 311–320.

Hsu, P.D., Scott, D.A., Weinstein, J.Á., Ran, F.A., Konermann, S., ... and Zhang, F. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31(9): 827-832.

Hwang, H.H., Yu, M., and Lai, E.M. (2017) *Agrobacterium* mediated plant transformation: biology and applications. *Arabidopsis book* 15:e0186.

Ivo, N.L., Nascimento, C.P., Vieira, L.S., Campos, F.A., and Aragão, F.J.L. (2008) Biolistic-mediated genetic transformation of cowpea (*Vigna unguiculata*) and stable Mendelian inheritance of transgenes. *Plant Cell Rep* 27(9):1475-83.

Ji, X., Zhang, H., Zhang, Y., Wang, Y., and Gao, C. (2015) Establishing a CRISPR–Cas-like immune system conferring DNA virus resistance in plants. *Nat Plants* 1: 1-4.

Ji, J., Zhang, C., Sun, Z., Wang, L., Duanmu, D., Fan, Q. (2019) Genome Editing in Cowpea *Vigna unguiculata* Using CRISPR-Cas9. *Int J Mol Sci* 20(10).

Joung, J.K., and Sander, J.D. (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Bio* 14: 49-55.

Kankam, F., Sowley, E.N.K., and Ibrahim, A. (2018) Evaluation of Cowpea (*Vigna unguiculata*) Genotypes for Resistance to Web Blight Caused by *Rhizoctonia solani*. *Asian J Crop Sci* 2(3): 1-7.

Kanzaki, H., Nirasawa, S., Saitoh, H., Ito, M., Nishihara, M., Terauchi, R., and Nakamura, I. (2002) Overexpression of the wasabi defensin gene confers enhanced resistance to blast fungus (*Magnaporthe grisea*) in transgenic rice. *Theor Appl Genet* 105: 809–814.

Kim, H., and Kim, J.S. (2014) A guide to genome engineering with programmable nucleases. *Nat Rev Genet* 15: 321-334.

Kumar, S., Tanti, B., Patil, B.L., Mukherjee, S.K., and Sahoo, L. (2017) RNAi-derived transgenic resistance to Mungbean yellow mosaic India virus in cowpea. *PLoS ONE* 12(10): e0186786.

Kumar, M., Yusuf, M.A., Yadav, P., Narayan, S., and Kumar, M. (2019) Overexpression of chickpea defensin gene confers tolerance to water-deficit stress in *Arabidopsis thaliana*. *Front Plant Sci* 10(290).

- Lacerda, A.F., Vasconcelos, E.A.R., Pelegrini, P.B., and Sa, M.F.G. (2014) Antifungal defensins and their role in plant defense. *Front Microbiol* 5:116.
- Leão, U.S., Eiras, M., Freire-Filho, F.R., Nogueira, M.S.R., Rodrigues, L.K., Ribeiro, V.Q., and Lopes, A.C.A. (2016) Selection of phenotypic traits and resistance to *Cowpea Severe Mosaic Virus* and *Cowpea Aphidborne Mosaic Virus* In *Cowpea* [*Vigna unguiculata* (L.) Walp.] seeds with rugose white coat. *Aust J Crop Sci* 10(4):470-481.
- Lee, H.H., Kim, J.S., Hoang, Q.T.N., Kim, J.I., and Kim, Y.S. (2018) Root-specific expression of defensin in transgenic tobacco results in enhanced resistance against *Phytophthora parasitica* var. *nicotianae*. *Eur J Plant Pathol* 151: 811.
- Liao, H.K., Hatanaka, F., Araoka, T., Reddy, P., Wu, M.Z., Sui, Y., ... and Izpisua Belmonte, J.C. (2017) In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans-epigenetic Modulation. *Cell* 171(7): 1495-1507.
- Lima, J.A.A., Sittolin, I.M., and Lima, R.C.A. (2005) Diagnose e estratégias de controle de doenças ocasionadas por vírus. In: Freire-Filho FR, Lima JAA, Ribeiro VQ (Eds.). *Feijão-caupi: avanços tecnológicos*. Brasília, DF: Embrapa Informação Tecnológica; Teresina: Embrapa Meio-Norte 403-459.
- Lima, J.A.A., Silva, A.K.F., Aragão, M.L., Ferreira, N.R.A., and Teófilo, E.M. (2011) Simple and multiple resistances to viruses in cowpea genotypes. *Pesq agropec bras* 46(11): 1432-1438.
- Mandal, D., Srivastava, D., and Sinharoy, S. (2020) Optimization of Hairy Root Transformation for the Functional Genomics in Chickpea: A Platform for Nodule Developmental Studies. In: Jain M., Garg R. (eds) *Legume Genomics. Methods in Molecular Biology* 2107.
- Marcotrigiano, J., Gingras, A.C., Sonenberg, N., and Burley, S.K. (1997) Cocystal Structure of the Messenger RNA 5'Cap-Binding Protein (eIF4E) Bound to 7-methyl-GDP. *Cell* 89: 951–961.
- Mishra, S., Behura, R., Awasthi, J.P., Dey, M., Sahoo, D., Bhowmik, S.S.D., Panda, S.K., and Sahoo, L. (2014) Ectopic overexpression of a mungbean vacuolar Na⁺/H⁺ antiporter gene (VrNHX1) leads to increased salinity stress tolerance in transgenic *Vigna unguiculata* L. Walp. *Mol Breeding* 34: 1345–1359.
- Mohanapriya, R., Naveenkumar, R., and Balabaskar, P. (2017) Survey, Virulence and Pathogenicity of Root Rot Incidence of Cowpea in Selected Districts of Tamilnadu caused by *Macrophomina phaseolina* (Tassi.) Goid. *Int J Curr Microbiol App Sci* 6 (3): 694-705.
- Montero, H., García-Román, R., Mora, S.I. (2015) eIF4E as a Control Target for Viruses. *Viruses* 7: 739-750.
- Muchero, W., Ehlers, J.D., Timothy, J.C., and Roberts, P.A. (2011) Genic SNP markers and legume synteny reveal candidate genes underlying QTL for

Macrophomina phaseolina resistance and maturity in cowpea [*Vigna unguiculata* (L.) Walp.]. *BMC Genomics* 12: 8.

Nechet, K.L., and Vieira, B.A.H. (2007) Reação de cultivares de feijão-caupi à mela (*Rhizoctonia solani*) em Roraima. *Fitopatol Bras* 32:424-428.

Nekrasov, V., Wang, C., Win, J., Lanz, C., Weige, D., and Kamoun, S. Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci Rep* 7(82): 1-6.

Néya, B.J., Zida, P.E., Sérémé, D., Lund, O.S., and Traoré, O. (2015) Evaluation of Yield Losses Caused by *Cowpea Aphid-Borne Mosaic Virus* (CABMV) in 21 Cowpea (*Vigna unguiculata* (L.) Walp.) Varieties in Burkina Faso. *PJBS* 18 (7): 304-313.

Obi, V.I., and Barriuso-Vargas, J.J. (2014) Situation of Biofungicides Reconnaissance, a Case of Anthracnose Disease of Cowpea. *American J Plant Sci* 5: 1202-1211.

Orawu, M., Melis, R., Laing, M., and Derera, J. (2013) Genetic inheritance of resistance to *Cowpea Aphid-Borne Mosaic Virus* in cowpea. *Euphytica* 189: 191–201.

Palanga, E., Filloux, D., Martin, D.P., Fernandez, E., Gargani, D., Ferdinand, R., ... and Roumagnac, P. (2016) Metagenomic-Based Screening and Molecular Characterization of Cowpea-Infecting Viruses in Burkina Faso. *PLoS ONE* 11(10): e0165188.

Pereira, D.L., Evangelista-Júnior, W.S., Souza, L.P., Laranjeiras-Júnior, A.L.R., and Leopoldino-Neto, A. (2013) Incidência De Oídio Na Cultura Do Feijão-Caupi Sob Diferentes Tipos De Irrigação. In: *III Conac – Congresso Nacional de feijão-caupi, 3, Recife-PE*.

Pérez-Bernal, M., Delgado, M., Cruz, A., Abreu, D., Valdivia, O., and Armas, R. (2017) Marker-free Transgenic Rice Lines with a Defensin Gene are Potentially Active against Phytopathogenic Fungus *Sarocladium oryzae*. *Acta Phytopathol Entomol Hung* 52(2): 135–144.

Popelka, J.C., Gollasch, S., Moore, A., Molvig, L., and Higgins, T.J.V. (2006) Genetic transformation of cowpea (*Vigna unguiculata* L.) and stable transmission of the transgenes to progeny. *Plant Cell Rep* 25: 304-312.

Portieles, R., Ayra, C., Gonzalez, E., Gallo, A., Rodriguez, R., Chacón, O., ... and Hidalgo, O.B. (2010) NmDef02, a novel antimicrobial gene isolated from *Nicotiana megalosiphon* confers high-level pathogen resistance under greenhouse and field conditions. *Plant Biotechnol J* 8: 678–690.

Pyott, D.E., Sheehan, E., and Molnar, A. (2016) Engineering of CRISPR/Cas9-mediated potyvirus resistance in transgene-free Arabidopsis plants. *Mol Plant Pathol* 17(8): 1276–1288.

Ramusi, T.M., van der Waals, J.E., Labuschagne, N., and Aveling, T.A.S. (2017) Evaluation of mefenoxam and fludioxonil for control of *Rhizoctonia solani*, *Pythium ultimum* and *Fusarium solani* on cowpea. J S Afr J Plant & Soil (34)1: 27-33.

Ray, R.B., Raychoudhuri, A., Steele, R., Nerurkar, P. (2010) Bitter Melon (*Momordica charantia*) Extract Inhibits Breast Cancer Cell Proliferation by Modulating Cell Cycle Regulatory Genes and Promotes Apoptosis. Cancer Res March 70(5): 1925-1931.

Rech, E.L., Vianna, G.R., and Aragão, F.J.L. (2008) High-efficiency transformation by biolistics of soybean, common bean and cotton transgenic plants. Nat Protoc 3: 410-418.

Sander, J.D., and Joung, J.K. (2013) CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32: 347–355.

Saravani, K., Javadian, E., Mohsenbeygi, M., and Hassanshahian, M. (2019) Study the Antimicrobial Effects of *Momordica charantia* on Pathogenic Bacteria. J Med Bacteriol 8(5, 6): 1-7.

Sesma, A., Castresana, C., and Castellano, M.M. (2017) Regulation of Translation by TOR, eIF4E and eIF2 α in Plants: Current Knowledge, Challenges and Future Perspectives. Front Plant Sci 8(644): 1-7.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., ... and Gao, C. (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol 31(8).

Shan, Q., Wang, Y., Li, J., and Gao, C. (2014) Genome editing in rice and wheat using the CRISPR/Cas system. Nat Protoc 9(10): 2395-2410.

Shi, J., Gao, H., Wang, H., Lafitte, H.R., Archibald, R.L., Yang, M., Hakimi, S.M., Mo, H., Habben, J.E. (2017) ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. Plant Biotechnol J 15: 207–216.

Shoaib, A., Munir, M., Javaid, A., Awan, Z.A., and Rafiq, M. (2018) Anti-mycotic potential of *Trichoderma* spp. and leaf biomass of *Azadirachta indica* against the charcoal rot pathogen, *Macrophomina phaseolina* (Tassi) Goid in cowpea. Egypt J Biol Pest Control 28 (26).

Silva, S.M.S., and Freire-Filho, F.R. (1998) Proteína de feijão-caupi (*Vigna unguiculata* (L.) Walp.): Caracterização e aplicação nutricional. Embrapa Documentos 44: 5-20.

Singh, V., Braddick, D., Dhar, P.K. (2017) Exploring the potential of genome editing CRISPR-Cas9 technology. Gene 599: 1–18.

Solleti, S.K., Bakshi, S., and Sahoo, L. (2008a) Additional virulence genes in conjunction with efficient selection scheme, and compatible culture regime

enhance recovery of stable transgenic plants in cowpea via *Agrobacterium tumefaciens*-mediated transformation. J Biotechnol 135: 97–104.

Solleti, S.K., Bakshi, S., and Sahoo, L. (2008b) Additional virulence genes in conjunction with efficient selection scheme, and compatible culture regime enhance recovery of stable transgenic plants in cowpea via *Agrobacterium tumefaciens*-mediated transformation. J Biotechnol 135: 97–104.

Sousa, A.R.R.C. (2013) Transformação genética de feijão-caupi (*Vigna unguiculata* L. Walp) visando à introdução de genes de resistência a viroses. Tese (Doutorado em Agronomia) - Universidade de Brasília, Brasília, Faculdade de Agronomia e Medicina Veterinária, 102 f.

Stotz, H.U., Thomson, J.G., and Wang, Y. (2009) Plant defensins Defense, development and application. Plant Signal Behav 4(11): 1010-1012.

Sun, N., and Zhao, H. (2013) Transcription Activator-Like Effector Nucleases (TALENs): A Highly Efficient and Versatile Tool for Genome Editing. Biotechnol Bioeng 110(7).

Tenório, D.A., Medeiros, E.V., Lima, C.S., Silva, J.M., Barros, J.A., Neves, R.P., and Laranjeira, D. (2019) Biological control of *Rhizoctonia solani* in cowpea plants using yeast. Trop plant pathol (2019) 44: 113.

Thao, H.T., Lan, N.T.N., Tuong, H.M., Thanh, N.V.T., Son, L.V., and Mau, C.H. (2017) Expression analysis of recombinant *Vigna radiata* plant defensin 1 protein in transgenic tobacco plants. J App Biol Biotech 5(01): 070-075.

Umaharan, P., Ariyanayagam, R.P., and Haque, S.Q. (1997) Resistance to Cowpea Severe Mosaic Virus, determined by three dosage dependent genes in *Vigna unguiculata* L.Walp. Euphytica 95: 49–55.

Verdcourt, B. (1970) Studies in the leguminosae: papilionoideae for the 'Flora of tropical East Africa'. Kew Bulletin 24: 507-569.

Weerawanich, K., Webster, G., Ma, J.K-C., Phoolcharoen, W., Sirikantaramas, S. (2018) Gene expression analysis, subcellular localization, and in planta antimicrobial activity of rice (*Oryza sativa* L.) defensin 7 and 8. Plant Physiol Biochem 124: 160-166.

Zanettini, M.H.B., and Pasquali, G. (2004) Plantas transgênicas. In: LUIS, M. I. R. (org.). Genômica. Atheneu, São Paulo, cap. 35, p. 721-736.

Zhang, C., Huang, M., Sang, X., Li, P., Ling, Y., Zhao, F., Du, D., ... and He, G. (2019) Association between sheath blight resistance and chitinase activity in transgenic rice plants expressing McCHIT1 from bitter melon. Transgenic Res 28: 381.

Plant Antimicrobial Peptides: State of the Art, In Silico Prediction and Perspectives in the Omics Era

Carlos André dos Santos-Silva^{1*}, Luisa Zupin^{2*} ,
Marx Oliveira-Lima¹, Lívia Maria Batista Vilela¹,
João Pacifico Bezerra-Neto¹, José Ribamar Ferreira-Neto¹,
José Diogo Cavalcanti Ferreira^{1,3}, Roberta Lane de Oliveira-Silva¹ ,
Carolline de Jesús Pires¹, Flavia Figueira Aburjaile¹ ,
Marianne Firmino de Oliveira¹, Ederson Akio Kido¹,
Sergio Crovella^{2,4} and Ana Maria Benko-Iseppon¹

¹Departamento de Genética, Universidade Federal de Pernambuco, Recife, Brazil. ²Genetic Immunology laboratory, Institute for Maternal and Child Health—IRCCS, Burlo Garofolo, Trieste, Italy. ³Departamento de Genética, Instituto Federal de Pernambuco, Pesqueira, Brazil.

⁴Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy.

Bioinformatics and Biology Insights
Volume 14: 1–22
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1177932220952739



ABSTRACT: Even before the perception or interaction with pathogens, plants rely on constitutively guardian molecules, often specific to tissue or stage, with further expression after contact with the pathogen. These guardians include small molecules as antimicrobial peptides (AMPs), generally cysteine-rich, functioning to prevent pathogen establishment. Some of these AMPs are shared among eukaryotes (eg, defensins and cyclotides), others are plant specific (eg, snakins), while some are specific to certain plant families (such as heveins). When compared with other organisms, plants tend to present a higher amount of AMP isoforms due to gene duplications or polyploidy, an occurrence possibly also associated with the sessile habit of plants, which prevents them from evading biotic and environmental stresses. Therefore, plants arise as a rich resource for new AMPs. As these molecules are difficult to retrieve from databases using simple sequence alignments, a description of their characteristics and in silico (bioinformatics) approaches used to retrieve them is provided, considering resources and databases available. The possibilities and applications based on tools versus database approaches are considerable and have been so far underestimated.

KEYWORDS: Defensin, lipid transfer protein, hevein, cyclotide, snakin, knotin, macadamia β -barrelins, impatiens-like, puroindoline, thaumatin

RECEIVED: July 14, 2020. **ACCEPTED:** July 30, 2020.

TYPE: Review

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors are very grateful to CAPES (Coordination for the Improvement of Higher Education Personnel, BioComputational Program), CNPq (Brazilian National Council for Scientific and Technological Development), and FACEPE (Fundação de Amparo à Ciência e Tecnologia de Pernambuco) for valuable financial support. The project is also supported by the Interreg Italia- Slovenia, ISE-EMH 07/2019 and RC 03/20 from IRCCS Burlo Garofolo/Italian Ministry of Health.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHORS: Luisa Zupin, Institute for Maternal and Child Health—IRCCS, Burlo Garofolo, via dell'Istria 65/1, 34137, Trieste, Italy. Email: luisa.zupin@burlo.trieste.it Ana Maria Benko-Iseppon, Departamento de Genética, Universidade Federal de Pernambuco, Av. Prof. Moraes Rego 1235, 50.670-423, Recife, Pernambuco, Brazil. Email: ana.iseppon@ufpe.br

Introduction

Proteins and peptides play different roles depending on their amino acids (aa) constitution, which may vary from tens to thousands residues.¹ Peptides are conventionally understood as having less than 50 aa.² Proteins, on the contrary, would be any molecule presenting higher amino acid content and both—proteins and peptides—present a plethora of variations in plants. Despite that, plant proteomes have been much more studied than peptidomes. It is well-known that the biochemical machinery necessary for the synthesis and metabolism of peptides is present in every living organism. From the variations of this machinery, a wide structural and functional diversity of peptides was generated, justifying the growing interest in their study.

In eukaryotes, peptides are prevalent in intercellular communication, performing as hormones, growth factors, and

neuropeptides, but they are also present in the defense system.³ Besides plants and animals, several pathogenic microorganisms, peptides can serve as classical virulence factors, which disrupt the epithelial barrier, damage cells, and activate or modulate host immune responses. An example of this performance is represented by Candidalysin,⁴ a fungal cytolytic peptide toxin found in the pathogenic fungus *Candida albicans* that damages epithelial membranes, triggers a response signaling pathway, and activates epithelial immunity. There are also reports of defense-related fungal peptides. For example, the Copsin, a peptide-based fungal antibiotic recently identified in the fungus *Coprinopsis cinerea*⁵ kills bacteria by inhibiting their cell wall synthesis. Regarding bacterial peptides, certain species from the gastrointestinal microbial community can release low-molecular-weight peptides, able to trigger immune responses.⁶ There are additionally peptides that act like bacterial “hormones” that allow bacterial communities to organize multicellular behavior such as biofilm formation.⁷ Some peptides are known for their medical importance, as defensins that

*Carlos André dos Santos Silva and Luisa Zupin contributed equally to the work.



present antibacterial, antiviral, and antifungal activities. For example, human alpha- and beta-defensins present in the saliva may potentially impede virus replication, including SARS-CoV-2,⁸ besides other roles as protection against intestinal inflammation (colitis).⁹

Considering the roles of plant peptides, they can also be multifunctional, and have been classified into 2 main categories¹⁰ (Supplementary Figure S1): (1) Peptides with no bioactivity, primarily resulting from the degradation of proteins by proteolytic enzymes, aiming at their recycling, and (2) bioactive peptides, which are encrypted in the structure of the parent proteins and are released mainly by enzymatic processes.

The first group is innocuous regarding signaling, regulatory functions, and bioactivity. So far, it has been reported that some of them may play a significant role in nitrogen mobilization across cellular membranes.¹¹ The second group of bioactive peptides has a substantial impact on the plant cell physiology. Some peptides of this group can act in the plant growth regulation (through cell-to-cell signaling), endurance against pathogens and pests by acting as toxins or elicitors, or even detoxification of heavy metals by ion-sequestration.

Comprising bioactive peptides, additional subcategorization has been proposed regarding their function. Tavormina et al¹² divided them into 3 groups (Supplementary Figure S1) based on the type of precursor:

- Derived from functional precursors: originated from a functional precursor protein;
- Derived from nonfunctional precursors: originated from a longer precursor that has no known biological function (as a preprotein, proprotein, or preproprotein);
- Not derived from a precursor protein: some sORFs (small Open Read Frames; usually <100 codons) are considered to represent a potential new source of functional peptides (known as “short peptides encoded by sORFs”).

A more intuitive classification of bioactive peptides was further proposed by Farrokhi et al¹⁰ according to their intracellular role (Supplementary Figure S1):

- Phytohormone peptides: the characteristic feature of these peptides is the regulation of fundamental plant physiological processes. They can be classified into those with (1) signaling roles in non-defense functions or those with (2) signaling roles in plant defense. Concerning the first group (Supplementary Figure S1), the peptide CLE25 (CLAVATA3/EMBRYO-SURROUNDING REGION-RELATED 25) is one of the representatives. This peptide transmits water-deficiency signals through vascular tissues in *Arabidopsis thaliana*, affecting abscisic acid biosynthesis and stomatal control of transpiration in association with BAM (BARELY ANY MERISTEM)

receptors in leaves.¹³ Another example is the PLS (POLARIS) peptide that acts during early embryogenesis but later activates auxin synthesis, also affecting cytokine synthesis and ethylene response.¹⁴ Regarding the second group, it includes peptides with signaling roles in plant defense, comprising at least 4 subgroups, including SYST (systemin) (Supplementary Figure S1). The SYST peptides were identified in Solanaceae members, like tomato and potato¹⁵ (acting on the signaling response to herbivory). The SYST leads to the production of a plant protease inhibitor that suppresses insect's proteases.¹⁶ Stratmann¹⁷ suggested that in plants, SYSTs act to stimulate the jasmonic acid signaling cascade within vascular tissues to induce a systemic wound response.

- Defense peptides or antimicrobial peptides (AMPs): to be fitted into this class, a plant peptide must fulfill some specific biochemical and genetic prerequisites. Regarding biochemical features, *in vitro* antimicrobial activity is required. Concerning the genetic condition, the gene encoding the peptide should be induced in the presence of infectious agents.¹⁸ In practice, this last requirement is not ever fulfilled as some AMPs are tissue-specific and are considered as part of the plant innate immunity, while other isoforms of the same class appear induced after pathogen inoculation.¹⁹

Plant AMPs are the central focus of the present review, comprising information on their structural features (at genomic, gene, and protein levels), resources, and bioinformatic tools available, besides the proposition of an annotation routine. Their biotechnological potential is also highlighted in the generation of both transgenic plants resistant to pathogens, and new drugs or bioactive compounds.

Overall Features of Plant AMPs

Antimicrobial peptides are ubiquitous host defense weapons against microbial pathogens. The overall plant AMP characterization regards the following variables (Figure 1): electrical charge, hydrophilicity, secondary and 3-dimensional (3D) structures, and the abundance or spatial pattern of cysteine residues.²⁰ These features are primarily related to their defensive role(s) as membrane-active antifungal, antibacterial, or antiviral peptides.

Regarding the nucleotide sequence, plant AMPs are hyper-variable and this genetic variability is considered crucial to provide diversity and the ability to recognize different targets. For their charges, AMPs can be classified as cationic or anionic (Figure 1). Most plant AMPs have positive charges, which is a fundamental feature for the interaction with the membrane lipids of pathogens.²¹ Concerning hydrophilicity, AMPs are generally amphipathic, that is, they exhibit molecular conformation with both hydrophilic and hydrophobic domains.²²

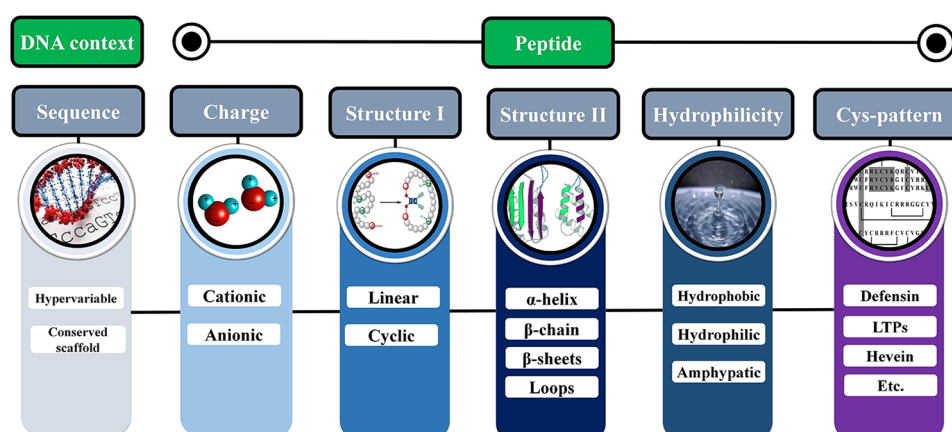


Figure 1. Plant antimicrobial peptide features considering DNA sequence level, protein structure, and physicochemical properties. LTP indicates lipid transfer protein.

With respect to their 3D structure, AMPs can be either linear or cyclic (Figure 1). Some linear AMPs adopt an amphipathic α -helical conformation, whereas non- α -helical linear peptides generally show 1 or 2 predominant amino acids.²³ In turn, cyclic AMPs, including cysteine-containing peptides, can be divided into 2 subgroups based on the presence of a single or multiple disulfide bonds. A peculiar feature of these peptides is a cationic and amphipathic character, what improves their functioning as membrane-permeabilizing agents.²³

Considering the secondary structures, AMPs may exhibit α -helices, β -chains, β -pleated sheets, and loops (Figure 1). Wang²⁴ classified plant AMPs into 4 families (α , β , $\alpha\beta$, and non- $\alpha\beta$), based on the protein classification of Murzin et al,²⁵ with some modifications. Antimicrobial peptides of the “ α ” family present α -helical structures,¹ whereas AMPs from the “ β ” family contains β -sheet structures usually stabilized by disulfide bonds.^{26,27} Some plant AMPs show an α -hairpinin motif formed by antiparallel α -helices that are stabilized by 2 disulfide bridges.²⁸ Such AMPs present a higher resistance to enzymatic, chemical, or thermal degradation.²⁹ Antimicrobial peptides from the “ $\alpha\beta$ ” family having both “ α ” and “ β ” structures are also stabilized by disulfide bridges. An example of AMP presenting “ $\alpha\beta$ ” structures are defensins, usually with a cysteine-stabilized $\alpha\beta$ motif (CS $\alpha\beta$), an α -helix, and a triple-stranded antiparallel β sheet stabilized mostly by 4 disulfide bonds.³⁰ Finally, AMPs that do not belong to the “ $\alpha\beta$ ” group exhibit no clearly defined “ α ” or “ β ” structures.²⁶

Plant AMPs are also classified into families considering protein sequence similarity, cysteine motifs, and distinctive patterns of disulfide bonds, which determine the folding of the tertiary structure.³¹ Therefore, plant AMPs are commonly grouped as thionins, defensins, heveins, knottins (linear and cyclic), lipid transfer proteins (LTP), snakins, and cyclo-tides.^{27,31} These AMP categories will be detailed in the next

sections, together with other groups here considered (Impatiens-like, Macadamia [β -barrelins], Puroindoline (PIN), and Thaumatin-like protein [TLP]) and the recently described α -hairpinin AMPs. The description includes comments on their structure, pattern for regular expression (REGEX) analysis (when available), functions, tissue-specificity, and scientific data availability.

Thionin

Thionins are composed by 45 to 48 amino acid residues with a molecular weight around 5 kDa, considering the mature peptide. They are synthesized with a signal peptide together with the mature thionin and the so-called acidic domain.³² To date, there is no experimental information available about possible functions of the acidic domain, even though it is clearly not dispensable as shown by the high conservation of the cysteine residues.³³

The thionin superfamily comprises 2 distinct groups of plant peptides α/β -thionins and γ -thionins with distinguished structural features.³⁴ The α/β thionins have homologous amino acid sequences and similar structures.³⁵ Besides, they are rich in arginine, lysine, and cysteine.³⁶ In turn, γ -thionins have a greater similarity with defensins, and some authors classify them within this group.³⁷ However, compared with the defensins, they present a longer conserved amino acid sequence.³¹

Regarding the cysteine motif, it can be divided into 2 subgroups, one with 8 residues connected by 4 disulfide bonds called 8C and the other with 6 residues connected by 3 disulfide bonds called 6C.³⁸ The general designation of thionins has been proposed as a family of homologous peptides that includes purothionins. The first plant thionin was isolated in 1942 from wheat flour and labeled as purothionin.³⁹ Since then, homologues from various taxa have been also identified, like

viscotoxins (*Viscum album*) and crambins (*Crambe abyssinica*).⁴⁰ They have also been isolated from different plant tissues like seeds, leaves, and roots.^{41,42}

Thionins have been tested for different elicitors: Gram-positive^{43,44} or Gram-negative bacteria,^{45,46} yeast,^{38,43} insect larvae,⁴⁷ nematode,³³ and inhibitory proteinase.⁴⁸ Thionins are hydrophobic in nature, interact with hydrophobic residues, and lyse bacteria cell membrane. Their toxicity is due to an electrostatic interaction with the negatively charged membrane phospholipids, followed by either pore formation or a specific interaction with a membrane.³⁸ It has been reported that they are able to inhibit other enzymes possibly through covalent attachment mediated by the formation of disulfide bonds, as previously observed for other thionin/enzyme combinations.⁴⁸

Thionin representatives with known 3D structures determined by X-ray crystallography are crambin (PDB ID: 1CRN), α 1- and β -purothionins (PDB ID: 2PHN and 1BHP), β -hordothionin (PDB ID: 1WUW), and viscotoxin-A3 (PDB ID: 1OKH). The first to be determined was the mixed form of crambin.^{35,49} It showed a distinct capital Γ shape with the N terminus forming the first strand in a β -sheet. The architecture of this sheet is additionally strengthened by 2 disulfide bonds.⁵⁰ After a short stretch of extended conformation, there is a helix-turn-helix motif. In crambin, there is a single disulfide involved in stabilizing the helix-to-helix contacts. At the center of this motif, there is a crucial Arg10 that forms 5 hydrogen bonds to tie together the first strand, the first helix, and the C terminus.⁵⁰

Defensin

The first plant defensins were isolated from wheat⁵¹ and barley grains,⁵² initially called γ -hordothionins. Due to some similarities in cysteine content and molecular weight, they were classified as γ -thionins. Later, the term " γ -thionin" was replaced by "defensin" based on the higher number of primary and tertiary structures of these proteins and also on their antifungal activities more related to insect and mammalian defensins than to plant thionins.⁵³ Plant defensins belong to a diverse protein superfamily called cis-defensin⁵⁴ and exhibit cationic charge, consisting of 45 to 54 aa with 2 to 4 disulfide bonds.^{53,55} Plant defensins share similar tertiary structures and typically exhibit a triple-stranded antiparallel β sheet, enveloped by an α -helix and confined by intramolecular disulfide bonds¹ (Figure 2A). This motif is called cysteine-stabilized $\alpha\beta$ (CS $\alpha\beta$).⁵⁶ The CS $\alpha\beta$ defensins were classified into 3 groups based on their sequence, structure, and functional similarity.

Defensins are known for their antimicrobial activity at low micromolar concentrations against Gram-positive and Gram-negative bacteria,⁵⁷ fungi,⁵⁸ viruses, and protozoa.⁵⁹ In addition, they present protein inhibition, insecticidal, and antiproliferative activity, acting as an ion-channel blocker, being also associated with the inhibition of pathogen protein

synthesis.⁶⁰ Instead, plant defensins act in the regulation of signal transduction pathways and induce inflammatory processes, in addition to wound healing, proliferation control, and chemotaxis.⁶¹

In general, plant defensins do not present high toxicity to human cells, having *in vivo* efficacy records, with relevant therapeutic potential, and can be applied in treatments associated with traditional medicine.⁶² Cools et al⁶³ reported that a peptide derived from a plant defensin (HsAFP1) acted synergistically with caspofungin (an antimycotic) (*in vivo* and *in vitro*) against the formation of *Candida albicans* biofilm on polystyrene and catheter substrate, indicating that the HsAFP1 variant presented a strong antifungal potential in the proposed treatment.

Other biotechnological applications of defensins are described, as in the case of EcgDf1, which was isolated from a legume (*Erythrina crista-galli*), heterologously expressed in *Escherichia coli* and purified. EcgDf1 inhibited the growth of various plant and human pathogens (such as *Candida albicans* and *Aspergillus niger* and the plant pathogens *Clavibacter michiganensis* ssp. *michiganensis*, *Penicillium expansum*, *Botrytis cinerea* and *Alternaria alternata*).⁶⁴ Due to these features, EcgDf1 is a candidate for the development of antimicrobial products for both agriculture and medicine.⁶⁴

Lipid transfer protein

Non-specific lipid transfer proteins (ns-LTPs) were first isolated from potato tubers⁶⁵ and are actually identified in diverse terrestrial plant species. They comprise a large gene family, are abundantly expressed in most tissues, but absent in most basal plant groups as chlorophyte and charophyte green algae.⁶⁶ They generally include an N-terminal signal peptide that directs the protein to the apoplastic space.⁶⁷ Some LTPs have a C-terminal sequence that allows their post-translational modification with a glycosylphosphatidylinositol molecule, facilitating the integration of LTP on the extracellular side of the plasma membrane.

The ns-LTPs are small proteins which were thus named because of their function of transferring lipids between the different membranes carrying lipids (non-specifically, the list includes phospholipids, fatty acids, their acylCoAs, or sterols). They consist of approximately 100 aa and are relatively larger in size than other AMPs, such as defensins.

Depending on their sizes, LTPs may be classified into 2 subfamilies: LTP1 and LTP2, with relative molecular weight of 9 and 7 kDa, respectively.^{68,69} The limited sequence conservation turned this classification inadequate. Thus, a modified and expanded classification system was proposed, presenting 5 main types (LTP1, LTP2, LTPc, LTPd, and LTPg) and 5 additional types with a smaller number of members (LTPe, LTPf, LTPh, LTPj, and LTPk).⁶⁶ The new classification system is not based on molecular size but rather on (1) the position of a conserved intron, (2) the identity of the amino acid sequence, and

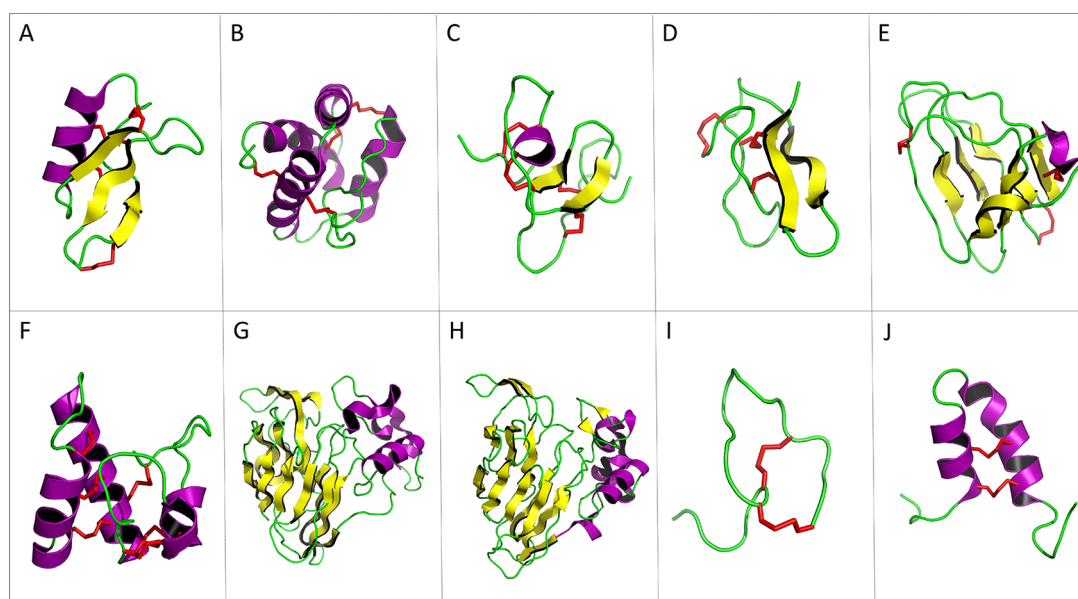


Figure 2. Three-dimensional structure of plant antimicrobial peptide representatives. Yellow arrows correspond to β -sheets; α -helices are represented in purple and disulfide bonds in red. (A) Defensin NaD1 of *Nicotiana glauca* (PDB ID: 1MR4). (B) Lipid transfer protein TaLTP1.1 isolated from wheat (PDB ID: 1GH1). (C) Hevein of *Hevea brasiliensis* (PDB ID: 1HEV). (D) Knottin Ep-AMP1 of *Echinopsis pachanoi* (PDB ID: 2MFS). (E) MIAMP1 peptide of *Macadamia integrifolia* (PDB ID: 1C01). (F) Snakin-1 of *Solanum tuberosum* (PDB ID: 5E5Q). (G) Thaumatin from *Thaumatococcus daniellii* (1RQW). (H) Zeamatin of *Zea mays* (1DU5, chain A). (I). Theoretical 3-dimensional model of *Impatiens balsamina* IbAMP1. (J) Helical hairpin structure of the novel antimicrobial peptide EcAMP1 from seeds of barnyard grass (*Echinochloa crus-galli*) (PDB ID: 2L2R). PDB indicates protein data bank.

(3) the spacing between the cysteine (Cys) residues (Supplementary Figure S2). Although this latter classification system is the most recent, the conventional classification of LTP1 and LTP2 types has been maintained by most working groups.

Lipid transfer protein nomenclature has been confusing and without consistent guidelines or standards. There are several examples where specific LTPs receive different names in different scientific articles. The lack of a robust terminology sometimes turns it quite difficult, extremely time-consuming, and frustrating to compare LTPs with different roles/functions.⁶⁷ Therefore, an additional nomenclature was also proposed by Salminen et al,⁶⁷ naming LTPs as follows: AtLTP1.3, OsLTP2.4, HvLTPc6, PpLTPd5, and TaLTPg7, with the first 2 letters indicating the plant species (eg, At = *Arabidopsis thaliana*, Pp = *Physcomitrella patens*); LTP1, LTP2, and LTPc indicating the type; while the last digit (here 3-7) regards the specific number given to each gene or protein within a given LTP type. For the sake of clarity, the authors recommend the inclusion of a point between the type specification (LTP1 and LTP2) and the gene number. For LTPc, LTPd, LTPg, and other types of LTP defined with a letter, the punctuation mark was not recommended. This latter classification system is

currently recommended as it comprises several features of LTPs and is more robust than the previous classification systems.

Lipid transfer proteins are small cysteine-rich proteins, having 4 to 5 helices in their tertiary structure (Figure 2B), which is stabilized by several hydrogen bonds. Such a folding gives LTPs a hydrophobic cavity to bind the lipids through hydrophobic interactions. This structure is stabilized by 4 disulfide bridges formed by 8 conserved cysteines, similar to defensins, although bound by cysteines in different positions. The disulfide bridges promote LTP folding into a very compact structure, which is extremely stable at different temperatures and denaturing agents.⁷⁰⁻⁷² These foldings provide a different specificity of lipid binding at the LTP binding site, where the LTP2 structure is relatively more flexible and present a lower lipid specificity when compared with LTP1.³⁴

The first 3D structure of an LTP was established for TaLTP1.1 based on 2D and 3D data of 1H-NMR, purified from wheat (*Triticum aestivum*) seeds in aqueous solution.^{73,74} Currently, several 3D structures of LTPs have been determined, either by nuclear magnetic resonance (NMR) or X-ray crystallography; in their free, unbound form or in a complex with ligands.

Hevein

The heveins were first identified in 1960 in the rubber-tree (*Hevea brasiliensis*), but its sequence was determined later, whereas a similarity was detected to the chitin-binding domain of an agglutinin isolated from *Urtica dioica* (L.).⁷⁵ with 8 cysteine residues forming a typical Cys motif.⁷⁶ The primary structure of the hevein consists of 29 to 45 aa, positively charged, with abundant glycine (6) and cysteine (8-10) residues,⁷⁶ and aromatic residues.^{31,77} The chitin-binding domain is a determinant component in the identification of hevein-like peptides whose binding site is represented by the amino acid sequence SXFGY/SXYGY, where X regards any amino acid.^{76,78} Most heveins have a coil- β 1- β 2-coil- β 3 structure that occurs by variations with the secondary structural motif in the presence of turns in 2 long coils in the β 3 chain.³¹ Antiparallel β chains form the central β sheet of the hevein motif with 2 long coils stabilized by disulfide bonds (Figure 2C).

Although the presence of chitin has not been identified in plants, there are chitin-like structures present in proteins that exhibit a strong affinity to this polysaccharide isolated from different plant sources.⁷⁹

The presence of 3 aromatic amino acids in the chitin-binding domain favors chitin binding by providing stability to the hydrophobic group C-H and the π electron system through van der Waals forces, as well as the hydrogen bonds between serine and N-acetylglucosamine (GlcNAc) present in the chitin structure.^{76,77} This domain is commonly found in chitinases of classes I to V, in addition to other plant antimicrobial proteins, such as lectins and PR-4 (Pathogenesis-Related protein 4) members.^{80,81} It may also occur in other proteins that bind to polysaccharide chitin,⁸⁰ such as the antimicrobial proteins AC-AMP1 and AC-AMP2 of *Amaranthus caudatus* (Amaranthaceae) seeds which are homologous to hevein but lack the C-terminal glycosylated region.⁸² Plant chitinases (class I) have the hevein-like domains, called HLDs. Due to the similar structural epitopes between chitinases and heveins, they are responsible for the cross reactive syndrome (latex-fruit syndrome).^{83,84}

Among the several classes of proteins mentioned, the proteins with a high degree of similarity to hevein are chitinases I and IV.⁷⁶ Chitinases are known to play an essential role in plant defense against pathogens,⁸⁵ also inhibiting *in vitro* fungal growth,⁸⁶ especially when combined with β -1,3-glucanases.⁸⁷ It also interferes with the growth of hyphae, resulting in abnormal ramification, delay, and swelling in their stretching.⁸¹ However, it has been shown that heveins have a higher inhibitory potential than chitinases and that their antifungal effect is not related only to the presence of chitinases⁸⁸; Pn-AMP1 and Pn-AMP2 AMPs with hevein domains have potent antifungal activities against a broad spectrum of fungi, including those without chitin in their cell walls.^{88,89} Modes of action of chitinases usually include degradation and disruption of the

fungal cell wall and plasma membrane due to its hydrolytic action, causing extravasation of plasma particles.^{21,89} Therefore, heveins have good antifungal activity, and only a few are active against bacteria, most of them with low activity.

Another role of hevein chitinases regards the antagonistic effect in triggering the aggregation of rubber particles in the latex extraction process in rubber trees. Unlike heveins, other chitinases inhibit rubber particle aggregation. However, its action in conjunction with other proteins (β -1,3-glucanase) increases the effect of β -1,3-glucanase on rubber particle aggregation.⁹⁰ A study by Shi et al⁹¹ found that the interaction of the protein network related to the antipathogenic activity released by luteoids (lysosomal microvacuole in latex) is essential in closing laticiferous cells (cells that produce and store latex), not only providing a physical barrier, but a biochemical barrier used by laticiferous cells affected by pathogen invasion.

Knottins (cysteine-knot peptides)

Knottins are part of the cysteine-rich peptides (CRPs) superfamily, sharing the Cysteine-knot motif and therefore resembling other families as defensins, heveins, and cyclotides.⁹² Their structure was initially identified by crystallography of carboxypeptides isolated from potato, showing the Cysteine-knot motif with 39 aa and 6 cysteine residues.⁹³ They are also called "Cysteine-knot peptides," "inhibitor Cysteine-knot peptides," or even "Cysteine-knot miniproteins" because their mature peptide presents less than 50 aa, forming 3 interconnected disulfide bonds in the Cysteine-knot motif, characterizing a particular scaffold.⁹² This conformation confers thermal stability at high temperatures. For example, the cysteine-stabilized β -sheet (CSB) motif derived from knottins presents stability at approximately 100°C with only 2 disulfide bonds.⁹⁴ The knottins may have linear or cyclic conformation. However, both exhibit connectivity between the cysteines at positions 1-4C, 2-5C, and 3-6C, forming a ring at the last bridge⁹² (Figure 2D).

Knottins have different functions, such as signaling molecules,⁹⁵ response against biotic and abiotic stresses,⁹⁶ root growth,⁹⁷ symbiotic interactions as well as antimicrobial activity against bacteria,⁹⁸ fungi,⁹⁹ virus,¹⁰⁰ and insecticidal activity,¹⁰¹ among others. Knottins antimicrobial activity has been attributed to the action of functional components of the plasma membrane, leading to alterations of lipids, ion flux, and exposed charge.⁹⁹ The accumulation of peptides on the surface of the membrane results in the weakening of the pathogen membrane,¹⁰² resulting in transient and toroidal perforations.⁹⁹

Macadamia (β -barrelins)

In the course of a large-scale survey to identify novel AMPs from Australian plants,^{103,104} an AMP with no sequence homology was purified. Its complementary DNA (cDNA) was

cloned from *Macadamia integrifolia* (Proteaceae) seeds, containing the complete peptide coding region. The peptide was named MiAMP1, being highly basic with an estimated isoelectric point (pI) of 10 and a mass of 8 kDa.

The MiAMP1 is 102 aa long, including a 26 aa signal peptide in the N-terminal region, bound to a 76 aa mature region with 6 cysteine residues.¹⁰⁵ Its 3D structure was determined using NMR spectroscopy,¹⁰⁴ revealing a unique conformation among plant AMPs, with 8 beta-strands arranged in 2 Greek key motifs, forming a Greek key beta-barrel (Figure 2E). Due to its particularities, MiAMP1 was classified as a new structural family of plant AMPs, and the name β -barrelins was proposed for this class.¹⁰⁴ This structural fold resembles a superfamily of proteins called γ -crystallin-like characterized by the precursors $\beta\gamma$ -crystallin.¹⁰⁶ This family includes AMPs from other organisms, for example, WmKT, a toxin produced by the wild yeast *Williopsis mraki*.¹⁰⁷

The MiAMP1 exhibited in vitro antimicrobial activity against various phytopathogenic fungi, oomycetes, and gram-positive bacteria¹⁰³ with a concentration range of 0.2 to 2 μ M generally required for a 50% growth inhibition (IC50). In addition, the transient expression of MiAMP1 in canola (*Brassica napus*) provided resistance against blackleg disease caused by the fungus *Leptosphaeria maculans*,¹⁰⁸ turning MiAMP1 potentially useful for genetic engineering aiming at disease resistance in crop plants.

There are few scientific publications with Macadamia-like peptides, maybe because they prevail in primitive plant groups (eg, lycophytes, gymnosperms to early angiosperms as *Amborella* and *Papaver*), being apparently absent in derived angiosperms (eg, Asteridae, including Brassicaceae as *Arabidopsis thaliana*). On the contrary, they have been identified in some monocots (as *Zantedeschia*, *Zea*, and *Sorghum*).¹⁰⁹ In fact, peptides similar to MiAMP1 appear to play a role in the defense against pathogens in gymnosperms, including species of economic importance (as *Pinus* and *Picea*) thus deserving attention for their biotechnological potential.¹⁰⁹

Impatiens-like (Ib-AMPs)

Four closely related AMPs (Ib-AMP1, Ib-AMP2, Ib-AMP3, and Ib-AMP4) were isolated from seeds of *Impatiens balsamina* (Balsaminaceae) with antimicrobial activity against a variety of fungi and bacteria, and low toxicity to human cells in culture. These AMPs are the smallest isolated from plants to date, consisting of only 20 aa in length. The Ib-AMPs are highly basic and contain 4 cysteine residues that form 2 disulfide bonds. Interestingly, they have no significant homology with other AMPs available in public databases. Sequencing of cDNAs isolated from *I. balsamina* revealed that all 4 peptides are encoded within a single transcript. Concerning the predicted precursor of Ib-AMP protein, it consists of a

pre-peptide followed by 6 mature peptide domains, each one of them flanked by propeptide domains ranging from 16 to 35 aa in length (Supplementary Figure S3). This primary structure with repeated domains of alternating basic peptides and acid propeptide domains has, to date, not been reported in other plant species.¹¹⁰

Patel et al¹¹¹ conducted an experiment to purify Ib-AMP1 from seeds of *Impatiens balsamina*. After purification, this peptide had its secondary structure tested by Circular Dichroism (CD). The results revealed a peptide that may include a β -turn but do not show evidences for either helical or β -sheet structure over a range of temperature and pH. Structural information from 2D 1H-NMR was obtained in the form of proton-proton internuclear distances inferred from nuclear overhauser enhancements (NOEs) and dihedral angle restraints from spin-spin coupling constants, which were used for distance geometry calculations. Owing to the difficulty in obtaining the correct disulfide connectivity by chemical methods, the authors had built and performed 3 separate calculations: (1) a model with no disulfides; (2) another with predicted disulfide bonds; and (3) a model with alternative connectivity disulfide, as assigned from the Nuclear Overhauser Effect spectroscopy (NOESY) NMR spectra. As a result, 2 hydrophilic patches were observed at opposite ends and opposite sides of the models, whereas in between them a large hydrophobic patch was identified. However, the study did not conclude which of the 3 models would be the most likely representative of Ib-AMP1, reporting only that cysteines are necessary for maintaining the structure.

Based on the experiment performed by Patel et al,¹¹¹ the present work built 3 different models: *Model 1*: without disulfide bonds, and the other 2 models with different disulfide connections—*Model 2*: NMR prediction by Patel et al¹¹¹ 6-Cys;16-Cys and 7-Cys;20-Cys, and *Model 3*: Disulfide Bond partner prediction by DiANNA 7-Cys;16-Cys and 6-Cys;20-Cys. Calculations have shown that although the peptide is small, the cysteines constrain part of it to adopt a well-defined main chain conformation. From residue 4 to 20 (except 11), the main chain is well-defined, whereas residues 1 to 3 in the N-terminal region present few restrictions and appear to be more flexible (Supplementary Figure S4). Analyzing the RMSD (root mean square deviation), we observed that all the models lost the initial conformation and, among them, Model 3 was the most stable. Models 1 and 2 showed a similar pattern (Supplementary Figure S5), as in the models of Patel et al,¹¹¹ although Model 1 was the most flexible.

Little is known about *Impatiens*-like AMPs mode of action. Lee et al¹¹² investigated the antifungal mechanism of Ib-AMP1 noting that when oxidized (bound by disulfide bridges), there occurs a 4-fold increase in antifungal activity against *Aspergillus flavus* and *Candida albicans*, as compared with reduced Ib-AMP1 (without disulfide bridges). Confocal microscopy analyses have shown that Ib-AMP1 can either bind to the cell

surface or penetrate cell membranes, indicating an antifungal activity by inhibiting a distinct cellular process, rather than ion channel or membrane pore formation. Fan et al¹¹³ reported the Ib-AMP4 antimicrobial activity dependent of β -sheet configuration to enable insertion into the lipid membrane, thus killing the bacteria through a non-lytic mechanism.¹¹⁴

Current approaches aim to make changes in Ib-AMP to improve its antimicrobial activity. As an example, synthetic variants of Ib-AMP1 were fully active against yeasts and fungi, where the replacement of amino acid residues by arginine or tryptophan improved more than twice the antifungal activity.¹¹⁵ Another study involving AMP modification generated a synthetic peptide without the disulfide bridges (ie, a linear analog of Ib-AMP1), which showed an antimicrobial specificity 3.7 to 4.8 times higher than the wild-type Ib-AMP1.¹¹⁶

Puroindoline

Puroindolines are small basic proteins that contain a single domain rich in tryptophan. These proteins were isolated from wheat endosperm, have a molecular mass around 13 kDa, and a calculated isoelectric point higher than 10. At least 2 main isoforms (called PIN-a and PIN-b) are known, which are encoded by *Pina-D1* and *Pinb-D1* genes, respectively. These genes share 70.2% identical coding regions but exhibit only 53% identity in the 3' untranslated region.¹¹⁷

Both PIN-a and PIN-b contain a structure with 10 conserved cysteine residues and a tertiary structure similar to LTPs, consisting of 4 α -helices separated by loops of varying lengths, with the tertiary structure joined by 5 disulfide bonds, 4 of which identical to ns-LTPs.¹¹⁷

The conformation of the 2 PIN isoforms was studied by infrared and Raman spectroscopy. Both PIN-a and PIN-b have similar secondary structures comprising approximately 30% helices, 30% β -sheets, and 40% non-ordered structures at pH 7. It has been proposed that the folding of both PINs is highly dependent on the pH of the medium. The reduction of the disulfide bridges results in a decrease of PINs solubility in water and to an increment of the β -sheet content by about 15% at the expense of the α -helix content.¹¹⁸ No high-resolution structure for any of the PIN isoforms is available, bringing challenges to understanding the function of their hydrophobic regions, with some evidence coming only from partially homolog peptides.¹¹⁷ However, Wilkinson et al¹¹⁹ proposed a theoretical model for several sequences of this AMP.

Puroindolines are proposed to be functional components of wheat grain hardness loci, control core texture, besides antifungal activity.¹²⁰⁻¹²³ Although the biological function of PINs is unknown, their involvement in lipid binding has been proposed. While LTPs bind to hydrophobic molecules in a large cavity, PINs interact only with lipid aggregates, that is, micelles or liposomes, through a single stretch of tryptophan residues. This stretch of tryptophan residues is especially significant in

the main form, PIN-a (WRWWKWWK), while it is truncated in the smaller form, PIN-b (WPTWWK).¹²⁴⁻¹²⁶

Puroindolines form protein aggregates in the presence of membrane lipids, and the organization of such aggregates is controlled by the lipid structure. In the absence of lipids, these proteins may aggregate, but there is no accurate information on the relationship between aggregation and interaction with lipids. The antimicrobial activity of PINs is targeted to cell membranes. Charnet et al¹²⁷ indicated that PIN is capable of forming ion channels in artificial and biological membranes that exhibit some selectivity over monovalent cations. The stress and Ca^{2+} ions modulate the formation and/or opening of channels. Puroindolines may also be membranotoxins, which may play a role in the plant defense mechanism against microbial pathogens.

Morris¹²⁸ reported that the PIN-a and PIN-b act through similar but somewhat different modes, which may involve "membrane binding, membrane disruption and ion channel formation" or "intracellular nucleic acid binding and metabolic disruption." Natural and synthetic mutants have allowed the identification of PINs as key elements for antimicrobial activity.

Snakin

Snakins are CRPs first identified in potato (*Solanum tuberosum*).^{129,130} Due to their sequence similarity to GASA (Gibberellic Acid Stimulated in *Arabidopsis*) proteins, the snakins were classified as members of the snakin/GASA family.¹³¹ The genes that encode these peptides have (1) a signal sequence of approximately 28 aa, (2) a variable region, and (3) a mature peptide of approximately 60 residues, with 12 highly conserved cysteine residues. These cysteine residues maintain the 3D structure of the peptide through disulfide bonds, besides providing stability to the molecule when the plant is under stress.^{129,130,132,133} (Figure 2F; Supplementary Figure S6).

Snakins may be expressed in different parts of the plant, like stem, leaves, flowers, seeds, and roots,¹³⁴⁻¹³⁷ both constitutive or induced by biotic or abiotic stresses. In vitro activity was observed against a variety of fungi, bacteria, and nematodes, acting as a destabilizer of the plasma membrane.^{129,138,139} Moreover, they were reported as essential agents in biological processes such as cell division, elongation, cell growth, flowering, embryogenesis, and signaling pathways.¹⁴⁰⁻¹⁴³

Alpha-hairpinins

As reported by Nolde et al,¹⁴⁴ alpha-hairpin emerged as a new AMP with unusual motif configuration. These peptides prevail in plants and their structure was resolved based on NMR data obtained from the EcAMP-1 peptide isolated from Barnyard grass seeds (*Echinoa crus-galli*).¹⁴⁴ Some α -hairpinins comprise trypsin inhibitors with helical hairpin structure and this group

was recently proposed as a new plant AMP family.¹⁴⁵ Similar to other AMPs, the amino acid sequences of α -hairpinins are variable. They share the conserved cysteine motif (CX3CX1-15CX3C) that form a helix-loop-helix fold and may have 2 disulfide bridges C1-C4 and C2-C3.¹⁴⁶ Its structural stability is maintained by forming hydrogen bonds, so that the side chains have a relatively stable spatial orientation.¹⁴⁷

As reviewed by Slavokhotova et al,¹⁴⁸ members of α -hairpin family have been described in both mono and dicot groups, including species as *Echinochloa crus-galli* and *Zea mays* (both Poaceae, monocot), *Fagopyrum esculentum* (Polygonaceae, eudicot), and *Stellaria media* (Caryophyllaceae, eudicot). Several transcripts with α -hairpinin motif exhibit similarities to snakin/GASA genes and are sometimes positioned within this family.

Although the α -hairpinins structure has been published, its mechanism of action is still not resolved (Figure 2J, PDB ID: 2L2R). However, studies indicate they present a potential DNA binding capacity.¹⁴⁹

Cyclotide

The term cyclotide was created at the end of the past century to designate a family of plant peptides with approximately 30 aa in size and a structural motif called cyclic cysteine knot (CCK).¹⁵⁰ This motif is composed by a head-to-tail cyclization that is stabilized by a knotted arrangement of disulfide bridges, with 6 conserved cysteines, connected as follows: C1-2, C3-6, C4-5.¹⁵¹ Cyclotides are generally divided into 2 subfamilies, Möbius and Bracelets, based on structural aspects. In addition to CCKs, 2 loops (between C1-2 and C4-5) have high similarity between both subfamilies, while the other 2 loops (between C2-3 and C3-4) exhibit some conservation within the subfamilies^{152,153} (Supplementary Figure S7).

To date, several cyclotides were identified in eudicot families such as Rubiaceae,¹⁵⁴ Violaceae,¹⁵⁵ Fabaceae,¹⁵⁶ and Solanaceae,¹⁵⁷ in addition to some monocots of Poaceae family.¹⁵⁸ In general, cyclotides may act in defense against a range of agents like insects, helminths, or mollusks. In addition, they can also act as ecobolic (inducer of uterus contractions),¹⁵⁴ antibacterial,¹⁵⁹ anti-HIV,¹⁰⁰ and anticancer factors.¹⁶⁰ All these characteristics added to the stability conferred by the CCK motif turn these peptides into excellent candidates for drug development.^{161,162}

Thaumatococcus-like protein

Thaumatins or TLPs belong to the PR-5 (Pathogen-related protein) family and received this name due to its first isolation from the fruit of *Thaumatococcus daniellii* (Maranthaceae) from West Africa.¹⁶³ Thaumatococcus-like proteins are abundant in the plant kingdom,¹⁶⁴ being found in angiosperms, gymnosperms, and bryophytes,¹⁶³ being also identified in other organisms, including fungi,^{165,166} insects,¹⁶⁷ and nematodes.¹⁶⁸

Thaumatococcus-like proteins are known for their antifungal activity, either by permeating fungal membranes¹⁶⁹ or by binding and hydrolyzing β -1,3-glucans.^{170,171} In addition, they may act by inhibiting fungal enzymes, such as xylanases,¹⁷² α -amylases, or trypsin.¹⁷³ Besides, the expression of TLPs is regulated in response to some stress factors, such as drought,¹⁷⁴ injuries,¹⁷⁵ freezing,¹⁷⁶ and infection by fungi^{177,178} viruses, and bacteria.¹⁷⁹

As to the TLP structure, this protein presents characteristic thaumatococcus signature (PS00316): G-x-[GF]-x-C-x-T-[GA]-D-C-x(1,2)-[QG]-x(2,3)-C.^{180,181} Most of the TLPs have molecular mass ranging from 21 to 26 kDa,¹⁶³ possessing 16 conserved cysteine residues (Supplementary Figure S8) involved in the formation of 8 disulfide bonds,¹⁸² which help in the stability of the molecule, allowing a correct folding even under extreme conditions of temperature and pH.¹⁸³ Thaumatococcus-like proteins also contain a signal peptide at the N-terminal, which is responsible for targeting the mature protein to a particular secretory pathway.¹⁶³ The tertiary structure presents 3 distinct domains, which are conserved and form the central cleft, responsible for the enzymatic activity of the protein, being located between domains I and II.¹⁸⁴ This central cleft may be of an acidic, neutral, or basic nature depending on the binding of the different linkers/receptors. All plant TLPs with antifungal activity have an acidic cleft known as motif REDDD due to 5 highly conserved amino acid residues (arginine, glutamic acid, and 3 aspartic acid; Supplementary Figure S8), being very relevant for specific receptor binding and antifungal activity.^{169,185,186}

Crystallized structures were determined for some plant TLPs, such as thaumatococcus¹⁸⁷ (Figure 2G), zeamatin¹⁶⁹ (Figure 2H), tobacco PR-5d¹⁸⁵ and osmotin,¹⁸⁶ the cherry allergen PruAv2,¹⁸⁸ and banana allergen Ba-TLP,¹⁸⁴ among other TLPs.

Some TLPs are known as small TLPs (sTLPs) due to the deletion of peptides in one of their domains, culminating in the absence of the typical central cleft. These sTLPs exhibit only 10-conserved cysteine residues, forming 5 disulfide bonds, resulting in a molecular weight of approximately 16 to 17 kDa. They have been described in monocots, conifers, and fungi, so far.^{163,189,190} Other TLPs exhibit an extracellular TLP domain and an intracellular kinase domain, being known as PR5K (PR5-like receptor kinases)¹⁹¹ and are present in both monocots and dicots. For example, *Arabidopsis* contains 3 PR5K genes, while rice has only 1.¹⁶³

Bioinformatic Tools and Databases for Plant AMPs

Databases

With the rapid growth in the number of available sequences, it is unfeasible to handle such amount of data manually. Thus, AMP sequences (as well as their biological information) have been deposited in large general databases, such as UniProt and TrEMBL, which contain sequences of multiple origins.^{192,193} In this sense, the construction of databases that deal specifically with AMPs was an important step to organize the data.

During the past decade, several databases were built to support the deposition, consultation, and mining of AMPs. Thus, these databases can be classified into 2 groups: general and specific.¹⁹⁴ The specific databases can be divided into 2 subgroups: those containing only 1 specific group (defensins or cyclotides) and those containing data from a supergroup of peptides (plant, animal, or cyclic peptides) (Supplementary Table 1). In general, both types of databases share some characteristics such as the way that the data are available or the tools to analyze AMPs.

The Collection of Antimicrobial Peptides (CAMPR3) is a database that comprises experimentally validated peptides, sequences experimentally deduced and still those with patent data, besides putative data based on similarity.¹⁹⁵⁻¹⁹⁷ The current version includes structures and signatures specific to families of prokaryotic and eukaryotic AMPs.¹⁹⁷ The platform also includes some tools for AMP prediction.

The antimicrobial peptide database (APD)¹⁹⁸ collects mature AMPs from natural sources, ranging from protozoa to bacteria, archaea, fungi, plants, and animals, including humans. AMPs encoded by genes that undergo post-translational modifications are also part of the scope, besides some peptides synthesized by multienzyme systems. The APD provides interactive interfaces for peptide research, prediction, and design, statistical data for a specific group, or for all peptides available in the database.

The LAMP (Database Linking Antimicrobial Peptides) comprises natural and synthetic AMPs, which can be separated into 3 groups: experimentally validated, predicted, and patented. Their data were primarily collected from the scientific literature, including UniProt and other AMP-related databases.¹⁹⁹

The Database of Antimicrobial Activity and Structure of Peptides (DBAASP)²⁰⁰ contains information about AMPs from different origins (synthetic or non-synthetic) and complexity levels (monomers and dimers) that were retrieved from PubMed using the following keywords: antimicrobial, antibacterial, antifungal, antiviral, antitumor, anticancer, and antiparasitic peptides. This database is manually curated and provides information about peptides that have specific targets validated experimentally. It also includes information on chemical structure, post-translational modifications, modifications in the N/C terminal amino acids, antimicrobial activities, cell target and experimental conditions in which a given activity was observed, besides information about the hemolytic and cytotoxic activities of the peptides.²⁰⁰ Due to the diversity of AMPs and the need to accommodate the most representative subclasses, several databases were established, focusing on specific types, sources, or features. There are several ways to classify AMPs, and they can range from biological sources such as bacterial AMPs (bacteriocins), plants, animals, and so on; biological activity: antibacterial, antiviral, antifungal, and insecticide; and based on molecular properties, pattern of covalent bonds, 3D structure and molecular targets.^{201,202}

The “Defensins Knowledgebase” is a database with manual curation and focused exclusively on defensins. This database contains information about sequence, structure, and activity, with a web-based interface providing access to information and enabling text-based search. In addition, the site presents information on patents, grants, laboratories, researchers, clinical studies, and commercial entities.^{203,205}

The CyBase is a database dedicated to the study of sequences and 3D structures of cyclized proteins and their synthetic variants, including tools for the analysis of mass spectral fingerprints of cyclic peptides, also assisting in the discovery of new circular proteins.²⁰⁵

The PhytAMP is a database designed to be solely dedicated to plant AMPs based on information collected from the UniProt database and from the scientific literature through PubMed.²⁰⁶

PlantPepDB is a database with manual curation of plant-derived peptides, mostly experimentally validated at the protein level. It includes data on the physical-chemical properties and tertiary structure of AMPs, also useful to identify their therapeutic potential. Different search options for simple and advanced compositing are provided for users to perform dynamic search and retrieve the desired data. Overall, PlantPepDB is the first database that comprises detailed analysis and comprehensive information on phyto-peptides from a wide functional range.²⁰⁷

Biological data banks (DBs) are organized collections of data of diverse nature that can be retrieved using different inputs. The management of this information is done through various software and hardware resources, whose retrieval and organization can be performed in a quick and efficient way.²⁰⁸ Considering biological data, information can be classified into (1) primary (sequences), (2) secondary (structure, expression, metabolic pathways, types of drugs, etc), and (3) specialized, for example, containing information on a species or on a class of protein.²⁰⁹ Within this third group, some references to AMPs can be mentioned, such as CAMPR3¹⁹⁶ and APD¹⁹⁸ that compile sequence data and structure retrieved from diverse sources, and also the Defensin knowledgebase²⁰³ and the CyBase²⁰⁵ which are dedicated to specific classes of peptides (defensins and cyclotides, respectively), in addition to PhytAMP,²⁰⁶ a specific database of plant AMPs (Supplementary Table 2).

Retrieving and annotating sequences from databases

The first step to infer the function of a given sequence (annotation) is to retrieve it in databases. For this purpose, 3 approaches have been used mostly: (1) local alignments, especially by using Basic Local Alignment Search Tool (BLAST)²¹⁰ and FASTA²¹¹; by searching for specific patterns using (2) REGEX or (3) Hidden Markov Model (HMM).¹⁹⁴

The first approach has been widely used, since most of the information is available in databases as sequences, together

with tools to align them, whereas the BLAST is the primary tool for doing so.²¹² This tool splits the sequence into small pieces (words), comparing it with the database. However, this approach has a limitation. Small motifs may not be significantly aligned as they comprise small portions of the sequences that can be smaller than 20% of the total size.^{31,194} Due to the high variability of AMPs, only few highly conserved sequences can be identified using this type of inference. To reduce the effects of local alignment limitations, other strategies based on the search for specific patterns were introduced, such as REGEX²¹³ (Supplementary Table 1) and HMM.²¹⁴

The REGEX is a precise way of describing a pattern in a string where each REGEX position must be set, although ambiguous characters (or wildcards) can also be used. For example, if we want to find a match for both amino acid sequences CAIESSK and WAIESK, we can use the following expression: [CW]AIES{1,2}K, this expression would find a pattern starting with the letter “C” or “W,” followed by an “A,” an “I,” and an “E,” 1 or 2 “S,” and ending with a “K.”

The HMMs are well-known for their effectiveness in modeling the correlations between adjacent symbols, domains, or events, and they have been extensively used in various fields of biological analysis, including pairwise and multiple sequence alignment, base-calling, gene prediction, modeling DNA sequencing errors, protein secondary structure prediction, non-coding RNA (ncRNA) identification, protein and RNA structural alignments, acceleration of RNA folding and alignment, fast noncoding RNA annotation, and many others. Using HMM, a statistic profile is included in the model, which is calculated from a sequence alignment, and a score is determined site-to-site, with conserved and variable positions defined a priori.^{194,215}

Predicting antimicrobial activity

The design of new AMPs led to the development of methods for the discovery of new peptides, thus allowing new experiments to be done by researchers. In this sense, the new challenge lies in the construction of new prediction models capable of discovering peptides with desired activities.

The APD DB has established a prediction interface based on some parameters defined by the entire set of peptides available in this database. These values are calculated from natural AMPs to consider features like length, net charge, hydrophobicity, amino acid composition, and so on. If we take as an example the net load, the AMPs deposited in the APD range from −12 to +30. This is the first parameter incorporated into the prediction algorithm. However, most AMPs have a net load ranging from −5 to +10, which then becomes the alternative prediction condition. Therefore, the same method is applied to the remaining parameters. The prediction in APD is performed in 3 main steps. First, the sequence parameters will be calculated and compared. If defined as an AMP, the peptide

can then be classified into 3 groups: (1) rich in given amino acids, (2) stabilized by disulfide, and (3) linear bridges. Finally, sequence alignments will be conducted to find 5 peptides of higher similarity.^{198,216,217}

The advent of machine learning (ML) methods has promoted new possibilities for drug discovery. In ML inferences, both a positive and a negative dataset are usually required to train the predictive models. The positive data, in this case, regard preferably experimentally validated AMPs that can be collected in databases, whereas negative data are randomly selected protein sequences that do not have AMP characteristics.^{197,218} Machine learning methods based on support vector machine (SVM), random forest (RF), and neural networks (NN) have been the most widely used. SVM is a specific type of supervised method of ML, aiming to classify data points by maximizing the margin between classes in a high-dimensional space.^{219,220} Random forest is a non-parametric tree-based approach that combines the ideas of adaptive neighbors with bagging for efficient adaptive data inference. Neural networks is an information processing paradigm inspired by how a biological nerve system process information. It is composed of highly interconnected processing elements (neurons or nodes) working together to solve specific problems.^{221–223}

Evaluating proteomic data

Regarding the use of AMPs in peptide therapeutics, as an alternative to antimicrobial treatment, new efficient and specific antimicrobials are demanded. As aforementioned, AMPs are naturally occurring across all classes of life, presenting high active potential as therapeutic agents against various kinds of bacteria.²²⁴ The identification of novel AMPs in databases is primarily dependent on knowing about specific AMPs together with a sufficient sequence similarity.²²⁵ However, orthologs may be divergent in sequence, mainly because they are under strong positive selection for variation in many taxa,²²⁶ leading to remarkably lower similarity, even in closely related species. In this scenario, where alignment tools present limited use, 1 strategy to identify AMPs is related to proteomic approaches.

Proteins and peptides are biomolecules responsible for various biochemical events in living organisms, from formation and composition to regulation and functioning. Thus, understanding of the expression, function, and regulation of the proteins encoded by an organism is fundamental, leading to the so-called “Proteomic Era.” The term “proteome” was first used by Marc Wilkins in 1994 and it represents the set of proteins encoded by the genome of a biological system (cell, tissue, organ, biological fluid, or organism) at a specific time under certain conditions.²²⁷ Protein extraction, purification, and identification methods have significantly advanced our capacity to elucidate many biological questions using proteomic approaches.^{228,229} Due to the wide diversity of proteomic analysis, methods makes the choice of the correct

approach dependent on the type of material and compounds to be analyzed.^{213,230} Two main tools are used to isolate proteins: (1) the 2-dimensional electrophoresis (2-DE) associated with mass spectrometry (MS) and (2) liquid chromatography associated with MS, each one with its own limitations.^{230–232} Obtaining native proteins is a challenge in proteomics or peptidomics, due to high protein complexity in samples, as the occurrence of post-translational modifications. Alternative strategies applied to extraction, purification, biochemical, and functional analyses of these molecules have been proposed, favoring access to structural and functional information of hard-to-reach proteins and peptides.²³³

Based on 2D gel, Al Akeel et al²³⁴ evaluated 14 spots obtained from seeds of *Foeniculum vulgare* (Apiaceae) aiming at proteomic analyses and isolation of small peptides. Extracted proteins were subjected to 3 kDa dialysis, and separation was carried out by DEAE-ion exchange chromatography while further proteins were identified by 2D gel electrophoresis. One of its spots showed high antibacterial activity against *Pseudomonas aeruginosa*, pointing to promising antibacterial effects, but requiring further research to authenticate the role of the anticipated proteins. For AMPs, 2DE is challenging due to the low concentration of the peptide molecules captured by this approach, their small sizes, and their ionic features (strongly cationic). In addition, the limited number of available specific databases and high variability turn their identification through proteolysis techniques and mass spectrometry, matrix-assisted laser desorption/ionization (MALDI-MS) difficult. In addition, the partial hydrophobicity characteristics and surface charges facilitate peptide molecular associations, making analysis difficult by any known proteomic approaches.²³² In addition, peptides are most often cleaved from larger precursors by various releasing or processing enzymes.²³⁵ Furthermore, profiles generated do not represent integral proteome, as 2DE has limitations to detect proteins with low concentration, values of extreme molecular masses, pIs, and hydrophobic proteins, including those of membranes.²³⁶ Due to these limitations, multidimensional liquid chromatography–high-performance liquid chromatography (MDLC-HPLC) has been successfully employed as an alternative to 2D gels. Techniques and equipments for the newly developed separation and detection of proteins and peptides, such as nano-HPLC and multidimensional HPLC, have improved proteomics evaluation.²³⁷ Molecular mass values obtained are used in computational searches in which they are compared with in silico digestion results of proteins in databases. In silico approaches, usually by the action of trypsin as a proteolytic agent, may generate a set of unique peptides whose masses are determined by MS.^{238,239} These methodologies are widely adopted for large-scale identification of peptide from MS/MS spectra.²⁴⁰ Theoretical spectra are generated using fragmentation patterns known for specific series of amino acids. The first 2 widely used search engines in database searching were SEQUEST²⁴¹ and MASCOT (Matrix

Science, Boston, MA; www.matrixscience.com).²⁴² They rank peptide matches based on a cross-correlation to match the hypothetical spectra to the experimental one.

MASCOT is widely used for peptidomics and proteomics analysis, including AMP identification in many organisms, or to evaluate the antibacterial efficacy of new AMPs. Evaluating new AMP against multidrug-resistant (MDR) *Salmonella enterica*, Tsai et al²⁴³ used 2D gel electrophoresis and liquid chromatography–electrospray ionization–quadrupole-time-of-flight tandem MS to determine the protein profiles. The protein identification was performed using the MASCOT with trypsin as cutting enzyme, whereas NCBI nr protein was set as a reference database. The methodology used in this study indicated that the novel AMP might serve as a potential candidate for drug development against MDR strains, confirming the usability of MASCOT. In a similar way, Umadevi et al²⁴⁴ described the AMP profile of black pepper (*Piper nigrum* L.) and their expression on *Phytophthora* infection using label-free quantitative proteomics strategy. For protein/peptide identification, MS/MS data were searched against the APD database²⁴⁵ using an in-house MASCOT server, established full tryptic peptides with a maximum of 3 missed cleavage sites and carbamidomethyl on cysteine, besides an oxidized methionine included as variable modifications. The APD database was used for AMP signature identification,²⁴⁵ together with PhytAMP²⁰⁶ and CAMPR3.¹⁹⁷ To enrich the characterization parameters, isoelectric point, aliphatic index, and grand average of hydropathy were also used²⁴⁶ (GRAVY) (using ProtParam tool), besides the net charge from PhytAMP database. Based on label-free proteomics strategy, they established for the first time the black pepper peptidomics associated with the innate immunity against *Phytophthora*, evidencing the usability of proteomics/peptidomics data for AMP characterization in any taxa, including plant AMPs, aiming the exploitation of these peptides as next-generation molecules against pathogens.²⁴⁴

Other tools use database searching algorithms, such as X!TANDEM,²⁴⁷ Open mass spectrometry search algorithm (OMSSA),²⁴⁸ ProBID,²⁴⁹ RADARS,²⁵⁰ and so on. These search engines are based on database search but use different scoring schemes to determine the top hit for a peptide match. General information on database search engines, their algorithms, and scoring schemes were reviewed by Nesvizhskii et al.²⁵¹ Despite its efficient ability to identify peptides, database searching presents several drawbacks, like false positive identifications due to overly noisy spectra and lower quality peptides score (related to the short size of peptides). So, the identification is strongly influenced by the amount of protein in the sample, the degree of post-translational modification, the quality of automatic searches, and the presence of the protein in the databases.^{252,253} In this scenario, the knowledge about the genome from a specific organism is important to allow the identification of the exact pattern of a given peptide. If an organism has no sequenced genome, it is not searchable

using these methods.^{235,240} Once the sequences are obtained, bioinformatic tools can be used to predict peptides structure and estimate bioactive peptides.²⁵⁴

More recently, an interactive and free web software platform, MixProTool, was developed, aiming to process multigroup proteomics data sets. This tool is compiled in R (www.r-project.org), providing integrated data analysis workflow for quality control assessment, statistics, gene ontology enrichment, and other facilities. The MixProTool is compatible with identification and quantification outputs from other programs, such as MaxQuant and MASCOT, where results may be visualized as vector graphs and tables for further analysis, in contrast to existing softwares, such as GiaPronto.²⁵⁵ According to the authors, the web tool can be conveniently operated, even by users without bioinformatics expertise, and it is beneficial for mining the most relevant features among different samples.²⁴

Antimicrobial peptide modeling

The central tenet of structural biology is that structure determines function. For proteins, it is often said the “function follows form” and “form defines function.” Therefore, to understand protein function in detail at the molecular level, it is mandatory to know its tertiary structure.²⁵⁶ Experimental techniques for determining structures, such as X-ray crystallography, NMR, electron paramagnetic resonance, and electron microscopy, require significant effort and investments.²⁵⁷

All methods mentioned have their own limitations, and the gap between the number of known proteins and the number of known structures is still substantial. Thus, there is a need for computational framework methods to predict protein structures based on the knowledge of the sequence.²⁵⁶ In addition, in recent years, there has been impressive progress in the development of algorithms for protein folding that may aid in the prediction of protein structures from amino acid sequence information.²⁵⁸

Historically, the prediction of a protein structure has been classified into 3 categories: comparative modeling, threading, and ab initio. The first 2 approaches construct protein models by aligning the query sequences with already solved model structures. If the models are absent in the Protein Data Bank (PDB), the models must be constructed from scratch, that is, by ab initio modeling, considered the most challenging way to predict protein structures.²⁵⁶

In the case of comparative modeling methods, when inserting a target sequence, the programs identify evolutionarily related models of solved structures based on their sequence or profile comparison, thus constructing structure models supported by these previously resolved models.²⁵⁹ This approach comprises 4 main steps: (1) fold assignment, which identifies similarity between the target and the structure of the solved model; (2) alignment of the target sequence to the model; (3) generation of a model based on alignment with the chosen

template; and (4) analysis of errors considering the generated model.²⁶⁰

There are several servers and computer models that automate the comparative modeling process, with SWISS-MODEL and MODELER figuring as the most used.^{261,262} Although automation makes comparative modeling accessible to experts and beginners, some adjustments are still needed in most cases to maximize model accuracy, especially in the case of more complex proteins.²⁶² Therefore, some caution must be taken regarding the generated models, considering the resolution and quality of the model used, as well as homology between the model and the protein of interest.

Threading modeling methods are based on the observation that known protein structures appear to comprise a limited set of stable folds, and those similarity elements are often found in evolutionarily distant or unrelated proteins. The most used servers based on this approach are MUSTER,²⁶³ SPARKS-X,²⁶⁴ RaptorX,²⁵⁹ ProSa-Web,²⁶⁵ and most notably the I-TASSER.²⁶⁶ In some cases, the incorporation of structural information to combine the sequence used in the search with possible models allows the detection of similarity in the fold, even in the absence of an explicit evolutionary relation.

The prediction of structures from known protein models is, at first sight, a more straightforward task than the prediction of protein structures from available sequences. Therefore, when no solved model is available, another approach is recommended, namely, the ab initio modeling. This method is intended to predict the structure only from the sequence information, without any direct assistance from previously known structures. The ab initio modeling aims to predict the best model, based on the minimum energy for a potential energy function by sampling the potential energy surface using various searchable information.^{267,268} Such approaches turn it challenging to produce high-resolution modeling, essential for determining the native protein folding and its biochemical interpretation. On the contrary, later resolved structures and comparisons with previously predicted proteins point to a higher successful modeling generated by ab initio methods than those generated by pure energy minimization methods, classical or even pure methods.²⁵⁶

Among the most used servers and programs for ab initio modeling, we highlight the ROSETTA,²⁵⁷ QUARK,²⁶⁹ and TOUCHSTONE II.²⁶⁷ The accuracy of the models calculated by many of these methods is evaluated by CAMEO (Continuous Automated Model Evaluation)²⁷⁰ and by CASP (Critical Assessment of protein Structure Prediction).²⁵⁸ Probably the first reasonably accurate ab initio model was built in CASP4. Since then, sustained progress was achieved in ab initio prediction, but mainly for small proteins (120 residues or less). In CASP11, for the first time, a novel 256-residue protein with a sequence identity with known structures lower than 5% was constructed with high precision for sequences of this size.²⁷¹

In CASP12, a significant improvement was reported in 4 areas: contact prediction, free modeling, template-based modeling, and estimating the accuracy of models. The authors report that this improvement is due to the accuracy of modeling and alignment methods, as well as increased data availability for both sequence and structure.²⁵⁸

Due to the number of AMPs deposited in the PDB (to date approximately 1099 structures), comparative modeling is the most used. However, when it comes to de novo peptide design, the most recommended choice would be *ab initio*²⁷² or a hybrid approach that uses more than 1 modeling method.²⁷³

Molecular dynamics simulation

After the generation of a model, the AMP stability should be evaluated using molecular dynamics (MD). Molecular dynamics comprises the application of computational simulations that predict the changes in the positions and velocities of the constituent atoms of a system under given time and condition. This calculation is done through a classical approximation of empirical parameters, called “force field.”²⁷⁴ If, on one hand, this approximation makes the dynamics of a system containing thousands of atoms numerically accessible, it obviously limits the nature of the processes that can be observed during the simulations. No quantum effect is visualized in a MD simulation; just as no chemical bond is broken, no interactions occur between orbitals, resonance, polarization, or charge transfer effects.²⁷⁵ However, the molecules go beyond a static system. Thus, MD is a computational technique that can be used for predicting or refining structures, dynamics of molecular complexes, drug development, and action of molecular biological systems.²⁷⁶ Molecular dynamics simulation is widely used for protein research, aiming to extract information about the physical properties of individual proteins. The results of such simulations are then compared with experimental results. As these experiments are generally carried out in solvents, it is necessary to simulate molecular systems of protein in water. These simulations have a variety of applications, such as determining the folding of a structure to a native structure and analyzing the dynamic stability of this structure.²⁷⁷

The use of MD to simulate protein folding processes is one of the most challenging applications and should be relatively long (in the order of microseconds to milliseconds) to allow observing a single fold event. In addition, the force field used must correctly describe the relative energies of a wide variety of shapes, including unfolding and poorly folded shapes that may occur during the simulation.²⁷⁵ The considerable application potential led to the implementation of MD simulation in many software packages, including GROMACS,^{278–280} AMBER,²⁸¹ NAMD,²⁸² CHARMM,²⁸³ LAMMPS,²⁸⁴ and Desmond.²⁸⁵ In addition to the above mentioned, there are other simulation types available, such as the Monte Carlo Method, Stochastic Dynamics, and Brownian Dynamics.²⁸⁰

In the last decades, MD simulation has become a standard tool in theoretical studies of large biomolecular systems,

including DNA or proteins, in environments with near realistic solvents. Indeed, simulations have proven valuable in deciphering functional mechanisms of proteins and other biomolecules, in uncovering the structural basis for disease, and in the design and optimization of small molecules, peptides, and proteins.²⁸⁶ Historically, the computational complexity of this type of computation has been extremely high, and much research has focused on algorithms to achieve unique simulations that are as long or as large as possible.²⁷⁸

Plant–pathogen interaction and molecular docking

The interplay between a given pathogen (eg, virus, bacteria, fungus) must be studied through a holistic approach. Host–pathogen relationships are very complex and occur at diverse conceivable levels, including the cellular/molecular level of both, pathogen and host, under given environmental conditions. A most approximate understanding of these interactions at every level is the ultimate goal of “systems biology” (SB). It comprises a **holistic** approach, integrating distinct disciplines, as biology, computer science, engineering, bioinformatics, physics, and others to predict how a given system behaves under given conditions and what is the role of its parts. Systems biology stands out because it is capable of correlating omics data for the understanding of plant–pathogen interaction. The construction of a plant–pathogen interaction network includes the reconstruction of metabolic pathways of these organisms, identification of the degree of pathogenicity, besides the expression of genes and proteins from both plant and pathogen. The networks can be classified into 5 types: (1) regulatory; (2) metabolic; (3) protein–protein interaction; (4) signaling and regulatory; and (5) signaling, regulatory, and metabolic.²⁸⁷ Each of these networks can be plotted according to computational approaches.

Also, further studies are required to contemplate the construction of evolutionary *in silico* models and the characterization of these molecular targets *in vitro*.^{288,289} Studies of protein–protein interactions to understand the regulatory process are essential²⁹⁰ and new computational methods are necessary for this purpose with more optimized algorithms, also to remove potential false positives. Thus, in-depth studies on the orientation of molecules and their linkages to the formation of a stable complex are of great importance for understanding plant–pathogen studies and also to develop new drugs.²⁹¹

Molecular docking

The understanding of the regulatory principles by which protein receptors recognize, interact, and associate with molecular substrates or inhibitors is of paramount importance to generate new therapeutic strategies.²⁹² In modern drug discovery, docking plays an important role in predicting the orientation of the binder when it is attached to a protein receptor or enzyme, using forms and electrostatic interactions, van der Waals,

Coulombic, and hydrogen bond as parameters to quantify or predict a given interaction.^{293,294} Molecular docking aims at exploring the predominant mode(s) of binding of a molecule (protein or ligand) when it binds to a protein with a known 3D structure based on a scoring function that has 3 main functions: the first is to determine the binding mode and the binding site of a protein, the second is to predict the absolute binding affinity between protein and ligand (or other protein) in lead optimization, and the third is virtual screening, which can identify potential drug leads for a given protein target by searching a large ligand or protein in database.²⁹⁵

Protein-protein interactions are essential for cellular and immune function. In many cases, due to the absence of an experimentally determined structure of the complex, these interactions must be modeled to obtain an understanding about their structure and molecular basis.²⁹⁶ Few studies on plant-pathogen interactions include docking approaches and most studies focus on drug development for medical purposes. Drug research based on structure is a powerful technique for the rapid identification of small molecules against the 3D structure of available macromolecular targets, usually by X-ray crystallography, NMR structures, or homology models.

Due to abundant information on protein sequences and structures, the structural information on specific proteins and their interactions have become crucial for current pharmacological research.²⁹⁷ Even in the absence of knowledge about the binding site and limited backbone movements, a variety of algorithms have been developed for docking over the past 2 decades. Although the ZDOCK,²⁹⁶ the rDOCK,²⁹⁸ and the HEX²⁹⁹ have provided results with high coupling precision, the complexes provided are not very useful for designing inhibitors for protein interfaces due to constraints on rigid body docking.²⁹⁴ In this context, more flexible approaches have been developed which generally examine very limited conformations compared with rigid body methods. These docking methods predict that binding is more likely to occur in broad surface regions and then defines the sites in complex structures of high affinity.³⁰⁰ The best example is the HADDOCK software,²⁹⁷ which has been successful in solving a large number of precise models for protein-protein complexes. A good example of its use is the study of the complex formed between plectasin, a member of the innate immune system, and a precursor lipid of bacterial cell wall II. The study identified the residues involved in the binding site between the 2 proteins, providing valuable information for planning new antibiotics.³⁰¹

However, the absolute energies associated with intermolecular interaction are not estimated with satisfactory accuracy by the current algorithms. Some significant issues as solvent effects, entropic effects, and receptor flexibility still need to be addressed. However, some methods, such as MOE-Dock,³⁰² GOLD,³⁰³ Glide,³⁰⁴ FlexX,³⁰⁵ and Surflex³⁰⁶ which deal with lateral chain flexibility, have proven to be effective and adequate in most cases. Realistic interactions between small

molecules and receptors still depend on experimental wet-lab validation.^{294,307}

Despite the current difficulties, there is a growing interest in the mechanisms and prediction of small molecules such as peptides, as they bind to proteins in a highly selective and conserved manner, being promising as new medicinal and biological agents.³⁰⁸ While both “small molecule docking methods” and “custom protocols” can be used, short peptides are challenging targets because of their high torsional flexibility.³⁰⁷ Protein-peptide docking is generally more challenging than those related to other small molecules, and a variety of methods have been applied so far. However, few of these approaches have been published in a way that can be reproduced with ease.³⁰⁹⁻³¹¹ Although it is difficult to use peptide docking, a recent focus of basic and pharmacological research has used computational tools with modified peptides to predict the selective disruption of protein-protein interactions. These studies are based on the involvement of some critical amino acid residues that contribute most to the binding affinity of a given interaction, also called hot-spots.^{312,313}

Despite the number of docking programs, existing algorithms still demand improvements. However, approaches are being developed to improve all issues related to punctuation, protein flexibility, interaction with plain water, among other issues.³¹⁴ In this context, the CAPRI (Critical Assessment of Predicted Interactions) is a community that provides a quality assessment of different docking approaches. It started in 2001 and since then has aided the development and improvement of the methodologies applied for docking.³¹⁵ An evaluation was carried out for CAPRI in 2016, resulting in an improvement in the integration of different modeling tools with docking procedures, as well as the use of more sophisticated evolutionary information to classify models. However, adequate modeling of conformational flexibility in interacting proteins remains an essential demand with a crucial need for improvement.³¹⁴ Different docking programs are currently available,²⁹⁴ and new alternatives continue to appear. Some of these alternatives will disappear, just as others will become the top choices among field users.

Molecular docking technique is not often used for AMPs, due to its standard mechanism of action based on the classical association with the external membrane of the pathogen. Despite that, some AMPs have the ability to bind other proteins and/or enzymes, a feature still scarcely studied. In such cases, molecular docking can be useful. An example of success is the study performed by Melo et al,⁴⁷ where they showed the specific binding of a trypsin to a cowpea (*Vigna unguiculata*) thionine, revealing that this interaction occurs in a canonical manner with Lys11, located in an extended exposed loop. Therefore, further application of docking may bring new evidences about the antimicrobial mechanisms revealing other molecular targets of interest.

It is clear that the combination of data bank information with bioinformatic tools (especially those allowing the identification of patterns, rather than sequence order) is able to revolutionize the identification of AMPs and prediction of their

activity. The data may come from genomic, transcriptomic, or proteomic databases, or a combination of different information sources (eg, genomic and transcriptomics, transcriptomics and proteomics).

Supplementary Figure S9 brings a schematic flowchart describing the steps for mining, annotation, and structural/functional analysis of AMPs, in addition to some wet-lab analyses that can be integrated to assess/confirm candidate AMPs.

Similar bioinformatic approaches have been actually used to identify potential peptide candidates with anti-SARS-CoV-2 activity, especially those potentially able to interact with the spike protein and proteases involved in viral penetration.^{316,317}

Concluding Remarks and Perspectives

As emphasized, plant AMPs show greater diversity and abundance, when compared with other kingdoms. It can be speculated that plants shelter many yet undescribed AMP classes, given their vast abundance and isoform diversity.

The genomic and peptidic structure of AMPs can be variable, with few key residues conserved, which turns their identification, classification, and comparison challenging even in the omics age. Nevertheless, advances in the generation of new bioinformatics tools and specialized databases have led to new and more efficient approaches for both the identification of primary sequences and molecular modeling, besides the analysis of the stability of the generated models.

Despite the large availability of omics data and bioinformatics tools, most new plant peptides have been discovered by wet-lab approaches regarding single candidates. High throughput in silico methods have the potential to transform this scenario, revealing many new candidates, including some new or “non-canonical” peptides. It may be also speculated that a myriad of new peptides may exist considering even smaller peptides, still less considered and more difficult to identify. Finally, in silico approaches shall in future studies be mandatory to define the design of wet-lab studies, turning the identification more efficient and requiring reasonably less time to track, identify, and confirm new candidate AMPs.

Considering the actual pandemic scenario of COVID-19, plant AMPs may be regarded as an important source of antiviral drug candidates, especially considering that some AMP categories present not only antiviral effects but also a wide spectrum antimicrobial activity, act as anti-inflammatory, and also induce the immune response.

Acknowledgements

The authors are very grateful to CAPES (Coordination for the Improvement of Higher Education Personnel, BioComputational Program), CNPq (Brazilian National Council for Scientific and Technological Development), and FACEPE (Fundação de Amparo à Ciência e Tecnologia de Pernambuco) for fellowships.

The project is supported by the Interreg Italia- Slovenia, ISE-EMH 07/2019 and RC 03/20 from IRCCS Burlo Garofolo/ Italian Ministry of Health

Author Contributions

CASS performed the literature review and wrote the manuscript. LZ, MOL, LMBV, JPBN, JRFN, JDCE, RLOS, CJP, FFA, and MFO wrote specific chapters, EAK and SC critically revised the text and included relevant suggestions. AMBI conceived the review, wrote the introduction and concluding remarks, besides critically revising the manuscript. All authors have read the manuscript and agree to its content.

ORCID iDs

Luisa Zupin  <https://orcid.org/0000-0001-5886-9129>

Roberta Lane de Oliveira-Silva  <https://orcid.org/0000-0003-1876-0094>

Flavia Figueira Aburjaile  <https://orcid.org/0000-0002-1067-1882>

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

- Whisstock JC, Lesk AM. Prediction of protein function from protein sequence and structure. *Q Rev Biophys.* 2003;36:307-340. doi:10.1017/S0033583503003901.
- Marmioli N, Maestri E. Plant peptides in defense and signaling. *Peptides.* 2014;56:30-44. doi:10.1016/j.peptides.2014.03.013.
- Germain H, Chevalier E, Matton DP. Plant bioactive peptides: an expanding class of signaling molecules. *Can J Bot.* 2006;84:1-19. doi:10.1139/b05-162.
- Moyes DL, Wilson D, Richardson JP, et al. Candidalysin is a fungal peptide toxin critical for mucosal infection. *Nature.* 2016;532:64-68. doi:10.1038/nature17625.
- Essig A, Hofmann D, Münch D, et al. Copsin, a novel peptide-based fungal antibiotic interfering with the peptidoglycan synthesis. *J Biol Chem.* 2014;289:34953-34964. doi:10.1074/jbc.M114.599878.
- Singh V, Singh K, Amdekar S, et al. Innate and specific gut-associated immunity and microbial interference. *FEMS Immunol Med Microbiol.* 2009;55:6-12. doi:10.1111/j.1574-695X.2008.00497.x.
- Flaherty RA, Freed SD, Lee SW. The wide world of ribosomally encoded bacterial peptides. *PLoS Pathog.* 2014;10:e1004221. doi:10.1371/journal.ppat.1004221.
- Baghizadeh Fini M. Oral saliva and COVID-19. *Oral Oncol.* 2020;108:104821.
- Koeninger L, Armbruster NS, Brinch KS, et al. Human β -defensin 2 mediated immune modulation as treatment for experimental colitis. *Front Immunol.* 2020;11:93.
- Farrokhi N, Whitelegge JP, Brusslan JA. Plant peptides and peptidomics. *Plant Biotechnol J.* 2008;6:105-134. doi:10.1111/j.1467-7652.2007.00315.x.
- Higgins CF, Payne JW. Plant peptides. In: Boulter, D, Parthier, B, eds. *Nucleic Acids and Proteins in Plants I*. Berlin, Germany: Springer; 1982:438-458. doi:10.1007/978-3-642-68237-7_13.
- Tavormina P, De Coninck B, Nikonov N, De Smet I, Cammue BP. The plant peptidome: an expanding repertoire of structural features and biological functions. *Plant Cell.* 2015;27:2095-2118. doi:10.1105/tpc.15.00440.
- Takahashi F, Suzuki T, Osakabe Y, et al. A small peptide modulates stomatal control via abscisic acid in long-distance signalling. *Nature.* 2018;556:235-238. doi:10.1038/s41586-018-0009-2.
- Liu J, Mehdi S, Topping J, Friml J, Lindsey K. Interaction of PLS and PIN and hormonal crosstalk in Arabidopsis root development. *Front Plant Sci.* 2013;4:75. doi:10.3389/fpls.2013.00075.
- Pearce G, Bhattacharya R, Chen Y-C. Peptide signals for plant defense display a more universal role. *Plant Signal Behav.* 2008;3:1091-1092. doi:10.4161/psb.3.12.6907.
- Ryan CA. Protease inhibitors in plants: genes for improving defenses against insects and pathogens. *Annu Rev Phytopathol.* 1990;28:425-449. doi:10.1146/annurev.py.28.090190.002233.

17. Stratmann J. Long distance run in the wound response—jasmonic acid is pulling ahead. *Trends Plant Sci.* 2003;8:247-250. doi:10.1016/S1360-1385(03)00106-7.
18. García-Olmedo F, Molina A, Alamillo JM, Rodríguez-Palenzuela P. Plant defense peptides. *Pept Sci.* 1998;47:479-491.
19. Benko-Iseppon AM, Lins Galdino S, Calsa T Jr, et al. Overview on plant antimicrobial peptides. *Curr Protein Pept Sci.* 2010;11:181-188. doi:10.2174/138920310791112075.
20. Benko-Iseppon AM, Crovella S. Ethnobotanical bioprospection of candidates for potential antimicrobial drugs from Brazilian plants: state of art and perspectives. *Curr Protein Pept Sci.* 2010;11:189-194.
21. Kaas Q, Westermann J-C, Craik DJ. Conopeptide characterization and classifications: an analysis using ConoServer. *Toxicon.* 2010;55:1491-1509. doi:10.1016/j.toxicon.2010.03.002.
22. Yu J, Kan Y, Rapp M, et al. Adaptive hydrophobic and hydrophilic interactions of mussel foot proteins with organic thin films. *Proc Natl Acad Sci USA.* 2013;110:15680-15685. doi:10.1073/pnas.1315015110.
23. Zanetti M. Cathelicidins, multifunctional peptides of the innate immunity. *J Leukoc Biol.* 2004;75:39-48. doi:10.1189/jlb.0403147.
24. Wang G, ed. *Antimicrobial Peptides: Discovery, Design and Novel Therapeutic Strategies*. Wallingford, UK: CABI; 2010. doi:10.1079/9781845936570.0000.
25. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol.* 1995;247:536-540. doi:10.1016/S0022-2836(05)80134-2.
26. Broekaert WF, Cammue BPA, De Bolle MFC, et al. Antimicrobial peptides from plants. *Crit Rev Plant Sci.* 1997;16:297-323.
27. Wang CK, Wacklin HP, Craik DJ. Cyclotides insert into lipid bilayers to form membrane pores and destabilize the membrane through hydrophobic and phosphoethanolamine-specific interactions. *J Biol Chem.* 2012;287:43884-43898. doi:10.1074/jbc.M112.421198.
28. Terras FR, Schoofs HM, De Bolle MF, et al. Analysis of two novel classes of plant antifungal proteins from radish (*Raphanus sativus* L.) seeds. *J Biol Chem.* 1992;267:15301-15309.
29. Vriens K, Cammue B, Thevissen K. Antifungal plant defensins: mechanisms of action and production. *Molecules.* 2014;19:12280-12303. doi:10.3390/molecules190812280.
30. Bruix M, González C, Santoro J, et al. 1 H-NMR studies on the structure of a new thionin from barley endosperm: structure of a new thionin. *Biopolymers.* 1995;36:751-763. doi:10.1002/bip.360360608.
31. Tam J, Wang S, Wong K, Tan W. Antimicrobial peptides from plants. *Pharmaceuticals.* 2015;8:711-757. doi:10.3390/ph8040711.
32. Bohlmann H, Apel K, Garcia-Olmedo F. Thionins. *Plant Mol Biol Rep.* 1994;12:S75.
33. Almaghrabi B, Ali MA, Zahoor A, Shah KH, Bohlmann H. Arabidopsis thionin-like genes are involved in resistance against the beet-cyst nematode (*Heterodera schachtii*). *Plant Physiol Biochem.* 2019;140:55-67.
34. Goyal RK, Mattoo AK. Plant antimicrobial peptides. In: Epanand RM, ed. *Host Defense Peptides and Their Potential as Therapeutic Agents*. Cham, Switzerland: Springer International Publishing; 2016:111-136. doi:10.1007/978-3-319-32949-9_5.
35. Stec B. Plant thionins—the structural perspective. *Cell Mol Life Sci.* 2006;63:1370-1385.
36. Nawrot R, Barylski J, Nowicki G, Broniarczyk J, Buchwald W, Goździcka-Józefiak A. Plant antimicrobial peptides. *Folia Microbiol (Praba).* 2014;59:181-196. doi:10.1007/s12223-013-0280-4.
37. De Coninck B, De Smet I. Plant peptides—taking them to the next level. *J Exp Bot.* 2016;67:4791-4795.
38. Yamuna S, Kumar V, Subramaniam D, Kumaravel K. Antimicrobial peptides from plants and their mode of action. *Am Int J Res Sci Technol Eng Math.* 2019;105:265-269.
39. Balls AK, Harris TH. The inhibitory effect of a protamine from wheat flour on the fermentation of wheat mashes. *Cereal Chem.* 1944;21:74-79.
40. Garcia-Olmedo F, Carmona MJ, Lopez-Fando JJ, et al. Characterization and analysis of thionin genes. In: Boller T, Meins F, eds. *Genes Involved in Plant Defense*. Vienna, Austria: Springer; 1992:283-302.
41. Gausing K. Thionin genes specifically expressed in barley leaves. *Planta.* 1987;171:241-246.
42. Iqbal A, Khan RS, Shehryar K, et al. Antimicrobial peptides as effective tools for enhanced disease resistance in plants. *Plant Cell Tissue Organ Cult.* 2019;139:1-15.
43. Franco OL, Murad AM, Leite JR, Mendes PAM, Prates MV, Bloch C. Identification of a cowpea γ -thionin with bactericidal activity. *FEBS J.* 2006;273:3489-3497.
44. Vasilchenko AS, Smirnov AN, Zavriev SK, Grishin EV, Vasilchenko AV, Rogozhin EA. Novel thionins from black seed (*Nigella sativa* L.) demonstrate antimicrobial activity. *Int J Pept Res Ther.* 2017;23:171-180.
45. Vila-Perelló M, Sánchez-Vallet A, García-Olmedo F, Molina A, Andreu D. Synthetic and structural studies on *Pyricularia pubera* thionin: a single-residue mutation enhances activity against Gram-negative bacteria. *FEBS Lett.* 2003;536:215-219.
46. Choi Y, Choi YD, Lee JS. Antimicrobial activity of γ -thionin-like soybean SE60 in *E. coli* and tobacco plants. *Biochem Biophys Res Commun.* 2008;375:230-234.
47. Melo FR, Rigden DJ, Franco OL, et al. Inhibition of trypsin by cowpea thionin: characterization, molecular modeling, and docking. *Proteins.* 2002;48:311-319.
48. Kramer KJ, Klassen LW, Jones BL, Speirs RD, Kammer AE. Toxicity of purothionin and its homologues to the tobacco hornworm, *Manduca sexta* (L.) (Lepidoptera: sphingidae). *Toxicol Appl Pharmacol.* 1979;48:179-183.
49. Wada K, Ozaki Y, Matsubara H, Yoshizumi H. Studies on purothionin by chemical modifications. *J Biochem (Tokyo).* 1982;91:257-263.
50. Stec B, Zhou R, Teeter MM. Full-matrix refinement of the protein crambin at 0.83 Å and 130 K. *Acta Crystallogr D Biol Crystallogr.* 1995;51:663-681.
51. Colilla FJ, Rocher A, Mendez E. γ -purothionins: amino acid sequence of two polypeptides of a new family of thionins from wheat endosperm. *FEBS Lett.* 1990;270:191-194.
52. Mendez E, Moreno A, Colilla F, et al. Primary structure and inhibition of protein synthesis in eukaryotic cell-free system of a novel thionin, gamma-hordothionin, from barley endosperm. *Eur J Biochem.* 1990;194:533-539. doi:10.1111/j.1432-1033.1990.tb15649.x.
53. Broekaert WF, Terras FRG, Cammue BPA, Osborn RW. Plant defensins: novel antimicrobial peptides as components of the host defense system. *Plant Physiol.* 1995;108:1353-1358.
54. Parisi K, Shafee TMA, Quimbar P, van der Weerden NL, Bleackley MR, Anderson MA. The evolution, function and mechanisms of action for plant defensins. *Semin Cell Dev Biol.* 2019;88:107-118. doi:10.1016/j.semdb.2018.02.004.
55. Pellegrini PB, Franco OL. Plant γ -thionins: novel insights on the mechanism of action of a multi-functional class of defense proteins. *Int J Biochem Cell Biol.* 2005;37:2239-2253. doi:10.1016/j.biocel.2005.06.011.
56. Zhao B-C, Lin H-C, Yang D, Ye X, Li Z-G. Disulfide bridges in defensins. *Curr Top Med Chem.* 2015;16:206-219. doi:10.2174/15680266150701115911.
57. Kraszewska J, Beckett MC, James TC, Bond U. Comparative analysis of the antimicrobial activities of plant defensin-like and ultrashort peptides against food-spoiling bacteria. *Appl Environ Microbiol.* 2016;82:4288-4298. doi:10.1128/AEM.00558-16.
58. Oddepally R, Guruprasad L. Isolation, purification, and characterization of a stable defensin-like antifungal peptide from *Trigonella foenum-graecum* (fenugreek) seeds. *Biochemistry (Moscow).* 2015;80:332-342. doi:10.1134/S0006297915030086.
59. Brogden KA. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol.* 2005;3:238-250. doi:10.1038/nrmicro1098.
60. Carvalho Ade O, Gomes VM. Plant defensins—prospects for the biological functions and biotechnological properties. *Peptides.* 2009;30:1007-1020. doi:10.1016/j.peptides.2009.01.018.
61. Shi J. Defensins and Paneth cells in inflammatory bowel disease. *Inflamm Bowel Dis.* 2007;13:1284-1292. doi:10.1002/ibd.20197.
62. Sher Khan R, Iqbal A, Malak R, et al. Plant defensins: types, mechanism of action and prospects of genetic engineering for enhanced disease resistance in plants. *3 Biotech.* 2019;9:192. doi:10.1007/s13205-019-1725-5.
63. Cools TL, Struyfs C, Drijfhout JW, et al. A linear 19-mer plant defensin-derived peptide acts synergistically with caspofungin against *Candida albicans* biofilms. *Front Microbiol.* 2017;8:2051. doi:10.3389/fmicb.2017.02051.
64. Rodríguez-Decuadro S, Dans PD, Borba MA, Benko-Iseppon AM, Cecchetto G. Gene isolation and structural characterization of a legume tree defensin with a broad spectrum of antimicrobial activity. *Planta.* 2019;250:1757-1772. doi:10.1007/s00425-019-03260-w.
65. Mazliak P, Douady D, Demandre C, Kader JC. *Recent Advances in the Chemistry and Biochemistry of Plant Lipids* (ed Gailliard, T, Mercer, EJ). New York, NY: Academic Press; 1975.
66. Edstam MM, Viitanen L, Salminen TA, Edqvist J. Evolutionary history of the non-specific lipid transfer proteins. *Mol Plant.* 2011;4:947-964. doi:10.1093/mp/ssl019.
67. Salminen TA, Blomqvist K, Edqvist J. Lipid transfer proteins: classification, nomenclature, structure, and function. *Planta.* 2016;244:971-997. doi:10.1007/s00425-016-2585-4.
68. Kader JC. Lipid-transfer proteins in plants. *Annu Rev Plant Physiol Plant Mol Biol.* 1996;47:627-654.
69. Castro MS, Gerhardt IR, Orrù S, Pucci P, Bloch C. Purification and characterization of a small (7.3 kDa) putative lipid transfer protein from maize seeds. *J Chromatogr B.* 2003;794:109-114. doi:10.1016/S1570-0232(03)00423-9.
70. Lindorff-Larsen K, Winther JR. Surprisingly high stability of barley lipid transfer protein, LTP1, towards denaturant, heat and proteases. *FEBS Lett.* 2001;488:145-148. doi:10.1016/S0014-5793(00)02424-8.
71. Berecz B, Mills ENC, Tamás L, Láng F, Shewry PR, Mackie AR. Structural stability and surface activity of sunflower 2S albumins and nonspecific lipid transfer protein. *J Agric Food Chem.* 2010;58:6490-6497. doi:10.1021/jf100554d.

72. Edstam MM, Edqvist J. Involvement of GPI-anchored lipid transfer proteins in the development of seed coats and pollen in *Arabidopsis thaliana*. *Physiol Plant*. 2014;152:32-42. doi:10.1111/ppl.12156.
73. Simorre JP, Caille A, Marion D, Ptak M. Two- and three-dimensional proton NMR studies of a wheat phospholipid transfer protein: sequential resonance assignments and secondary structure. *Biochemistry*. 1991;30:11600-11608. doi:10.1021/bi00113a016.
74. Gincel E, Simorre J-P, Caille A, Marion D, Ptak M, Vovelle F. Three-dimensional structure in solution of a wheat lipid-transfer protein from multidimensional 1H-NMR data. A new folding for lipid carriers. *Eur J Biochem*. 1994;226:413-422. doi:10.1111/j.1432-1033.1994.tb20066.x.
75. Peumans WJ, De Ley M, Broekaert WF. An unusual lectin from stinging nettle (*Urtica dioica*) rhizomes. *FEBS Lett*. 1984;177:99-103. doi:10.1016/0014-5793(84)80989-8.
76. Slavokhotova AA, Shelenkov AA, Andreev YA, Odintsova TI. Hevein-like antimicrobial peptides of plants. *Biochemistry (Moscow)*. 2017;82:1659-1674. doi:10.1134/S0006297917130065.
77. Asensio JL, Cañada FJ, Siebert HC, et al. Structural basis for chitin recognition by defense proteins: GlcNAc residues are bound in a multivalent fashion by extended binding sites in hevein domains. *Chem Biol*. 2000;7:529-543. doi:10.1016/S1074-5521(00)00136-8.
78. Wong KH, Tan WL, Serra A, et al. Ginkgotides: proline-rich hevein-like peptides from gymnosperm *Ginkgo biloba*. *Front Plant Sci*. 2016;7:1639. doi:10.3389/fpls.2016.01639.
79. Raikhel NV, Lee HI, Broekaert WF. Structure and function of chitin-binding proteins. *Annu Rev Plant Physiol Plant Mol Biol*. 1993;44:591-615. doi:10.1146/annurev.pp.44.060193.003111.
80. Beintema JJ. Structural features of plant chitinases and chitin-binding proteins. *FEBS Lett*. 1994;350:159-163. doi:10.1016/0014-5793(94)00753-5.
81. Rogozhin EA, Slezina MP, Slavokhotova AA, et al. A novel antifungal peptide from leaves of the weed *Stellaria media* L. *Biochimie*. 2015;116:125-132. doi:10.1016/j.biochi.2015.07.014.
82. Broekaert WF, Marien W, Terras FRG, et al. Antimicrobial peptides from *Amaranthus caudatus* seeds with sequence homology to the cysteine/glycine-rich domain of chitin-binding proteins. *Biochemistry*. 1992;31:4308-4314. doi:10.1021/bi00132a023.
83. Volpicella M, Leoni C, Fanizza I, Placido A, Pastorello EA, Ceci LR. Overview of plant chitinases identified as food allergens. *J Agric Food Chem*. 2014;62:5734-5742. doi:10.1021/jf5007962.
84. Wagner S, Breiteneder H. The latex-fruit syndrome. *Biochem Soc Trans*. 2002;30:935-940. doi:10.1042/bst0300935.
85. Iseli B, Boller T, Neuhaus JM. The N-terminal cysteine-rich domain of tobacco class I chitinase is essential for chitin binding but not for catalytic or antifungal activity. *Plant Physiol*. 1993;103:221-226. doi:10.1104/pp.103.1.221.
86. Broekaert WF, Van Parijs J, Leyns F, Joos H, Peumans WJ. A chitin-binding lectin from stinging nettle rhizomes with antifungal properties. *Science*. 1989;245:1100-1102. doi:10.1126/science.245.4922.1100.
87. Leah R, Tommerup H, Svendsen I, Mundy J. Biochemical and molecular characterization of three barley seed proteins with antifungal properties. *J Biol Chem*. 1991;266:1564-1573.
88. Van Parijs J, Broekaert WF, Goldstein IJ, Peumans WJ. Hevein: an antifungal protein from rubber-tree (*Hevea brasiliensis*) latex. *Planta*. 1991;183:258-264. doi:10.1007/BF00197797.
89. Koo JC, Lee SY, Chun HJ, et al. Two hevein homologs isolated from the seed of *Pharbitis nil* L. exhibit potent antifungal activity. *Biochim Biophys Acta*. 1998;1382:80-90. doi:10.1016/S0167-4838(97)00148-9.
90. Wang X, Shi M, Wang D, et al. Comparative proteomics of primary and secondary luteoids reveals that chitinase and glucanase play a crucial combined role in rubber particle aggregation in *Hevea brasiliensis*. *J Proteome Res*. 2013;12:5146-5159. doi:10.1021/pr400378c.
91. Shi M, Li Y, Deng S, et al. The formation and accumulation of protein-networks by physical interactions in the rapid occlusion of laticifer cells in rubber tree undergoing successive mechanical wounding. *BMC Plant Biol*. 2019;19:8. doi:10.1186/s12870-018-1617-6.
92. Molesini B, Treggiari D, Dalbeni A, Minuz P, Pandolfini T. Plant cysteine-knot peptides: pharmacological perspectives: plant cysteine-knot proteins in pharmacology. *Br J Clin Pharmacol*. 2017;83:63-70. doi:10.1111/bcp.12932.
93. Rees DC, Lipscomb WN. Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *J Mol Biol*. 1982;160:475-498. doi:10.1016/0022-2836(82)90309-6.
94. Chiche L, Heitz A, Gelly J-C, et al. Squash inhibitors: from structural motifs to macrocyclic knottins. *Curr Protein Pept Sci*. 2004;5:341-349. doi:10.2174/1389203043379477.
95. Murphy E, Smith S, De Smet I. Small signaling peptides in *Arabidopsis* development: how cells communicate over a short distance. *Plant Cell*. 2012;24:3198-3217. doi:10.1105/tpc.112.099010.
96. Li G, Asiegbu FO. Use of Scots pine seedling roots as an experimental model to investigate gene expression during interaction with the conifer pathogen *Heterobasidion annosum* (P-type). *J Plant Res*. 2004;117:155-162. doi:10.1007/s10265-003-0140-4.
97. Iyer S, Acharya KR. Tying the knot: the cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines. *FEBS J*. 2011;278:4304-4322. doi:10.1111/j.1742-4658.2011.08350.x.
98. Aboye TL, Strömstedt AA, Gunasekera S, et al. A cactus-derived toxin-like cystine knot peptide with selective antimicrobial activity. *ChemBioChem*. 2015;16:1068-1077. doi:10.1002/cbic.201402704.
99. Göransson U, Burman R, Gunasekera S, Strömstedt AA, Rosengren KJ. Circular proteins from plants and fungi. *J Biol Chem*. 2012;287:27001-27006. doi:10.1074/jbc.R111.300129.
100. Gustafson KR, Sowder RC, Henderson LE, et al. Circulins A B. Novel human immunodeficiency virus (HIV)-inhibitory macrocyclic peptides from the tropical tree *Chassalia parvifolia*. *J Am Chem Soc*. 1994;116:9337-9338. doi:10.1021/ja00099a064.
101. Jennings CV, Rosengren KJ, Daly NL, et al. Isolation, solution structure, and insecticidal activity of kalata B2, a circular protein with a twist: do Möbius strips exist in nature? *Biochemistry*. 2005;44:851-860. doi:10.1021/bi047837h.
102. Burman R, Strömstedt AA, Malmsten M, Göransson U. Cyclotide-membrane interactions: defining factors of membrane binding, depletion and disruption. *Biochim Biophys Acta*. 2011;1808:2665-2673. doi:10.1016/j.bbame.2011.07.004.
103. Marcus JP, Goulter KC, Green JL, Harrison SJ, Manners JM. Purification, characterisation and cDNA cloning of an antimicrobial peptide from *Macadamia integrifolia*. *Eur J Biochem*. 1997;244:743-749. doi:10.1111/j.1432-1033.1997.00743.x.
104. McManus AM, Nielsen KJ, Marcus JP, et al. MiAMP1, a novel protein from *Macadamia integrifolia* adopts a Greek key β -barrel fold unique amongst plant antimicrobial proteins. *J Mol Biol*. 1999;293:629-638. doi:10.1006/jmbi.1999.3163.
105. Finkina EI, Melnikova DN, Bogdanov IV, Ovchinnikova TV. Peptides of the innate immune system of plants. Part II. Biosynthesis, biological functions, and possible practical applications. *Russ J Bioorganic Chem*. 2019;45:55-65. doi:10.1134/S1068162019020043.
106. Ohno A, Tate S, Seeram SS, et al. NMR structure of the *Streptomyces metalloproteinase* inhibitor, SMPI, isolated from *Streptomyces nigrescens* TK-23: another example of an ancestral β -crystallin precursor structure. *J Mol Biol*. 1998;282:421-433. doi:10.1006/jmbi.1998.2022.
107. Antuch W, Güntert P, Wüthrich K. Ancestral beta gamma-crystallin precursor structure in a yeast killer toxin. *Nat Struct Biol*. 1996;3:662-665.
108. Kazan K, Rusu A, Marcus JP, Goulter KC, Manners JM. Enhanced quantitative resistance to *Leptosphaeria maculans* conferred by expression of a novel antimicrobial peptide in canola (*Brassica napus* L.). *Mol Breed*. 2002;10:63-70. doi:10.1023/A:1020354809737.
109. Manners JM. Primitive defence: the MiAMP1 antimicrobial peptide family. *Plant Mol Biol Rep*. 2009;27:237-242. doi:10.1007/s11105-008-0083-y.
110. Tailor RH, Acland DP, Attenborough S, et al. A novel family of small cysteine-rich antimicrobial peptides from seed of *Impatiens balsamina* is derived from a single precursor protein. *J Biol Chem*. 1997;272:24480-24487. doi:10.1074/jbc.272.39.24480.
111. Patel SU, Osborn R, Rees S, Thornton JM. Structural studies of *Impatiens balsamina* antimicrobial protein (Ib-AMP1). *Biochemistry*. 1998;37:983-990. doi:10.1021/bi971747d.
112. Lee DG, Shin SY, Kim D-H, et al. Antifungal mechanism of a cysteine-rich antimicrobial peptide, Ib-AMP1, from *Impatiens balsamina* against *Candida albicans*. *Biotechnol Lett*. 1999;21:1047-1050. doi:10.1023/A:1005636610512.
113. Fan X, Xu W, Han J, Jiang X, Wink M, Wu G. Antimicrobial peptide hybrid fluorescent protein based sensor array discriminate ten most frequent clinic isolates. *Biochim Biophys Acta Gen Subj*. 2019;1863:1158-1166. doi:10.1016/j.bbagen.2019.04.010.
114. Fan X, Korytowski A, Makky A, Tanaka M, Wink M. Ib-AMP4 insertion causes surface rearrangement in the phospholipid bilayer of biomembranes: implications from quartz-crystal microbalance with dissipation. *Biochim Biophys Acta Biomembr*. 2018;1860:617-623. doi:10.1016/j.bbame.2017.10.025.
115. Thevissen K, François IE, Sijtsma L, et al. Antifungal activity of synthetic peptides derived from *Impatiens balsamina* antimicrobial peptides Ib-AMP1 and Ib-AMP4. *Peptides*. 2005;26:1113-1119. doi:10.1016/j.peptides.2005.01.008.
116. Wang P, Bang J-K, Kim HJ, Kim J-K, Kim Y, Shin SY. Antimicrobial specificity and mechanism of action of disulfide-removed linear analogs of the plant-derived Cys-rich antimicrobial peptide Ib-AMP1. *Peptides*. 2009;30:2144-2149. doi:10.1016/j.peptides.2009.09.020.
117. Gautier M-F, Aleman M-E, Guirao A, Marion D, Joudrier P. Triticum aestivum puroindolines, two basic cysteine-rich seed proteins: cDNA sequence

- analysis and developmental gene expression. *Plant Mol Biol.* 1994;25:43-57. doi:10.1007/BF00024197.
118. Le Bihan T, Blochet J-É, Désormeaux A, Marion D, Pézolet M. Determination of the secondary structure and conformation of puroindolines by infrared and Raman spectroscopy. *Biochemistry.* 1996;35:12712-12722. doi:10.1021/bi960869n.
 119. Wilkinson M, King R, Grimaldi R. Sequence diversity and identification of novel puroindoline and grain softness protein alleles in *Elymus*, *Agropyron* and related species. *Diversity.* 2018;10:114. doi:10.3390/d10040114.
 120. Giroux MJ, Sripo T, Gerhardt S, Sherwood J. Puroindolines: their role in grain hardness and plant defence. *Biotechnol Genet Eng Rev.* 2003;20:277-290. doi:10.1080/02648725.2003.10648047.
 121. Bhavé M, Morris CF. Molecular genetics of puroindolines and related genes: allelic diversity in wheat and other grasses. *Plant Mol Biol.* 2008;66:205-219. doi:10.1007/s11103-007-9263-7.
 122. Dhatwalia VK, Sati OP, Tripathi MK, Kumar A. Isolation, characterization and antimicrobial activity at diverse dilution of wheat puroindoline protein. *World J Agric Sci.* 2009;5:297-300.
 123. Zhang J, Martin JM, Balint-Kurti P, Huang L, Giroux MJ. The wheat puroindoline genes confer fungal resistance in transgenic corn: the puroindolines confer corn SLB resistance. *J Phytopathol.* 2011;159:188-190. doi:10.1111/j.1439-0434.2010.01744.x.
 124. Douliet J-P, Michon T, Elmorjani K, Marion D. Mini review: structure, biological and technological functions of lipid transfer proteins and indolines, the major lipid binding proteins from cereal kernels. *J Cereal Sci.* 2000;32:1-20. doi:10.1006/jcsc.2000.0315.
 125. Morris CF. Puroindolines: the molecular genetic basis of wheat grain hardness. *Plant Mol Biol.* 2002;48:633-647. doi:10.1023/A:1014837431178.
 126. Marion D, Bakan B, Elmorjani K. Plant lipid binding proteins: properties and applications. *Biotechnol Adv.* 2007;25:195-197. doi:10.1016/j.biotechadv.2006.11.003.
 127. Charnet P, Molle G, Marion D, Rousset M, Lullien-Pellerin V. Puroindolines form ion channels in biological membranes. *Biophys J.* 2003;84:2416-2426. doi:10.1016/S0006-3495(03)75046-2.
 128. Morris CF. The antimicrobial properties of the puroindolines, a review. *World J Microbiol Biotechnol.* 2019;35:86. doi:10.1007/s11274-019-2655-4.
 129. Segura A, Moreno M, Madueño F, Molina A, García-Olmedo F. Snakin-1, a peptide from potato that is active against plant pathogens. *Mol Plant Microbe Interact.* 1999;12:16-23. doi:10.1094/MPMI.1999.12.1.16.
 130. Berrocal-Lobo M, Segura A, Moreno M, Lopez G, García-Olmedo F, Molina A. Snakin-2, an antimicrobial peptide from potato whose gene is locally induced by wounding and responds to pathogen infection. *Plant Physiol.* 2002;128:951-961. doi:10.1104/pp.010685.
 131. Oliveira-Lima M, Benko-Iseppon A, Neto J, et al. Snakin: structure, roles and applications of a plant antimicrobial peptide. *Curr Protein Pept Sci.* 2017;18:368-374. doi:10.2174/1389203717666160619183140.
 132. Mao Z, Zheng J, Wang Y, et al. The new CaSn gene belonging to the snakin family induces resistance against root-knot nematode infection in pepper. *Phytoparasitica.* 2011;39:151-164. doi:10.1007/s12600-011-0149-5.
 133. Yeung H, Squire CJ, Yosaatmadja Y, et al. Radiation damage and racemic protein crystallography reveal the unique structure of the GASA/snakin protein superfamily. *Angew Chem Int Ed.* 2016;55:7930-7933. doi:10.1002/anie.201602719.
 134. Zhang S, Yang C, Peng J, Sun S, Wang X. GASA5, a regulator of flowering time and stem growth in *Arabidopsis thaliana*. *Plant Mol Biol.* 2009;69:745-759. doi:10.1007/s11103-009-9452-7.
 135. Almasia NI, Narhirañak V, Hopp HE, Vazquez-Rovere C. Isolation and characterization of the tissue and development-specific potato snakin-1 promoter inducible by temperature and wounding. *Electron J Biotechnol.* 2010;13:8-9. doi:10.2225/vol13-issue5-fulltext-12.
 136. Zimmermann R, Sakai H, Hochholdinger F. The gibberellic acid stimulated-like gene family in maize and its role in lateral root development. *Plant Physiol.* 2010;152:356-365. doi:10.1104/pp.109.149054.
 137. Guzmán-Rodríguez JJ, Ibarra-Laclette E, Herrera-Estrella L, et al. Analysis of expressed sequence tags (ESTs) from avocado seed (*Persea americana* var. *drymifolia*) reveals abundant expression of the gene encoding the antimicrobial peptide snakin. *Plant Physiol Biochem.* 2013;70:318-324. doi:10.1016/j.plaphy.2013.05.045.
 138. Faccio P, Vazquez-Rovere C, Hopp E, et al. Increased tolerance to wheat powdery mildew by heterologous constitutive expression of the *Solanum chacoense* snakin-1 gene. *Czech J Genet Plant Breed.* 2011;47:S135-S141. doi:10.17221/3268-CJPB.
 139. Herbel V, Schäfer H, Wink M. Recombinant production of snakin-2 (an antimicrobial peptide from tomato) in *E. coli* and analysis of its bioactivity. *Molecules.* 2015;20:14889-14901. doi:10.3390/molecules200814889.
 140. Kotilainen M. GEG participates in the regulation of cell and organ shape during corolla and carpel development in *Gerbera hybrida*. *Plant Cell.* 1999;11:1093-1104. doi:10.1105/tpc.11.6.1093.
 141. Furukawa T, Sakaguchi N, Shimada H. Two OsGASR genes, rice GAST homologue genes that are abundant in proliferating tissues, show different expression patterns in developing panicles. *Genes Genet Syst.* 2006;81:171-180. doi:10.1266/ggs.81.171.
 142. Roxrud I, Lid SE, Fletcher JC, Schmidt EDL, Opsahl-Sorteberg H-G. GASA4, one of the 14-member Arabidopsis GASA family of small polypeptides, regulates flowering and seed development. *Plant Cell Physiol.* 2007;48:471-483. doi:10.1093/pcp/pcm016.
 143. Lucau-Danila A, Laborde L, Legrand S, et al. Identification of novel genes potentially involved in somatic embryogenesis in chicory (*Cichorium intybus* L.). *BMC Plant Biol.* 2010;10:122. doi:10.1186/1471-2229-10-122.
 144. Nolde SB, Vassilevski AA, Rogozhin EA, et al. Disulfide-stabilized helical hairpin structure and activity of a novel antifungal peptide EcAMP1 from seeds of barnyard grass (*Echinochloa crus-galli*). *J Biol Chem.* 2011;286:25145-25153. doi:10.1074/jbc.M110.200378.
 145. Oparin PB, Mineev KS, Dunaevsky YE, et al. Buckwheat trypsin inhibitor with helical hairpin structure belongs to a new family of plant defence peptides. *Biochem J.* 2012;446:69-77. doi:10.1042/BJ20120548.
 146. Slavokhotova AA, Rogozhin EA, Musolyamov AK, et al. Novel antifungal α -hairpinin peptide from *Stellaria media* seeds: structure, biosynthesis, gene structure and evolution. *Plant Mol Biol.* 2014;84:189-202. doi:10.1007/s11103-013-0127-z.
 147. Wang X, Zhu G, Liang W, et al. Design, synthesis and docking of linear and hairpin-like α helix mimetics based on alkoxylated oligobenzamide. *ChemistrySelect.* 2019;4:6651-6655. doi:10.1002/slct.201900171.
 148. Slavokhotova AA, Shelenkov AA, Korostyleva TV, et al. Defense peptide repertoire of *Stellaria media* predicted by high throughput next generation sequencing. *Biochimie.* 2017;135:15-27. doi:10.1016/j.biochi.2016.12.017.
 149. Sousa D, Porto W, Silva M, da Silva T, Franco O. Influence of cysteine and tryptophan substitution on DNA-binding activity on maize α -hairpinin antimicrobial peptide. *Molecules.* 2016;21:1062. doi:10.3390/molecules21081062.
 150. Craik DJ, Daly NL, Bond T, Waine C. Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J Mol Biol.* 1999;294:1327-1336. doi:10.1006/jmbi.1999.3383.
 151. Lima S, Benko-Iseppon A, Neto J, et al. Plants defense-related cyclic peptides: diversity, structure and applications. *Curr Protein Pept Sci.* 2017;18:375-390. doi:10.2174/1389203717666160724194119.
 152. Weidmann J, Craik DJ. Discovery, structure, function, and applications of cyclotides: circular proteins from plants. *J Exp Bot.* 2016;67:4801-4812. doi:10.1093/jxb/erw210.
 153. Park S, Yoo K-O, Marcussen T, et al. Cyclotide evolution: insights from the analyses of their precursor sequences, structures and distribution in violets (*Viola*). *Front Plant Sci.* 2017;8:2058. doi:10.3389/fpls.2017.02058.
 154. Grain L. Isolation of oxytocic peptides from *Oldenlandia affinis* by solvent extraction of tetraphenylborate complexes and chromatography on sephadex LH-20. *Lloydia.* 1973;36:207-208.
 155. Claeson P, Göransson U, Johansson S, Luijendijk T, Bohlin L. Fractionation protocol for the isolation of polypeptides from plant biomass. *J Nat Prod.* 1998;61:77-81. doi:10.1021/np970342r.
 156. Poth AG, Colgrave ML, Philip R, et al. Discovery of cyclotides in the Fabaceae plant family provides new insights into the cyclization, evolution, and distribution of circular proteins. *ACS Chem Biol.* 2011;6:345-355. doi:10.1021/cb100388j.
 157. Poth AG, Mylne JS, Grassl J, et al. Cyclotides associate with leaf vasculature and are the products of a novel precursor in petunia (*Solanaceae*). *J Biol Chem.* 2012;287:27033-27046. doi:10.1074/jbc.M112.370841.
 158. Nguyen GKT, Zhang S, Nguyen NTK, et al. Discovery and characterization of novel cyclotides originated from chimeric precursors consisting of albumin-1 chain a and cyclotide domains in the Fabaceae family. *J Biol Chem.* 2011;286:24275-24287. doi:10.1074/jbc.M111.229922.
 159. Pranting M, Loov C, Burman R, Göransson U, Andersson DI. The cyclotide cycloviolacin O2 from *Viola odorata* has potent bactericidal activity against Gram-negative bacteria. *J Antimicrob Chemother.* 2010;65:1964-1971. doi:10.1093/jac/dkq220.
 160. Lindholm P, Göransson U, Johansson S, et al. Cyclotides: a novel type of cytotoxic agents. *Mol Cancer Ther.* 2002;1:365-369.
 161. Craik DJ, Clark RJ, Daly NL. Potential therapeutic applications of the cyclotides and related cystine knot mini-proteins. *Expert Opin Investig Drugs.* 2007;16:595-604. doi:10.1517/13543784.16.5.595.
 162. Northfield SE, Wang CK, Schroeder CI, et al. Disulfide-rich macrocyclic peptides as templates in drug design. *Eur J Med Chem.* 2014;77:248-257. doi:10.1016/j.ejmech.2014.03.011.
 163. Liu J-J, Sturrock R, Ekramoddoullah AKM. The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. *Plant Cell Rep.* 2010;29:419-436. doi:10.1007/s00299-010-0826-8.
 164. de Jesus-Pires C, Ferreira-Neto JRC, Pacifico Bezerra-Neto J, et al. Plant thaumatin-like proteins: function, evolution and biotechnological

- applications. *Curr Protein Pept Sci.* 2020;21:36–51. doi:10.2174/1389203720666190318164905.
165. Grenier J, Potvin C, Asselin A. Some fungi express beta-1,3-glucanases similar to thaumatin-like proteins. *Mycologia.* 2000;92:841–848. doi:10.2307/3761579.
 166. Sakamoto Y. *Lentinula edodes* tlg1 encodes a thaumatin-like protein that is involved in lentinan degradation and fruiting body senescence. *Plant Physiol.* 2006;141:793–801. doi:10.1104/pp.106.076679.
 167. Brandazza A, Angeli S, Tegoni M, Cambillau C, Pelosi P. Plant stress proteins of the thaumatin-like family discovered in animals. *FEBS Lett.* 2004;572:3–7. doi:10.1016/j.febslet.2004.07.003.
 168. Kitajima S, Sato F. Plant pathogenesis-related proteins: molecular mechanisms of gene expression and protein function. *J Biochem.* 1999;125:1–8. doi:10.1093/oxfordjournals.jbchem.a022244.
 169. Batalia MA, Monzingo AF, Ernst S, Roberts W, Robertus JD. The crystal structure of the antifungal protein zeamatin, a member of the thaumatin-like, PR-5 protein family. *Nat Struct Biol.* 1996;3:19–22. doi:10.1038/nsb0196-19.
 170. Trudel J, Grenier J, Potvin C, Asselin A. Several thaumatin-like proteins bind to β -1,3-glucans. *Plant Physiol.* 1998;118:1431–1438. doi:10.1104/pp.118.4.1431.
 171. Grenier J, Potvin C, Trudel J, Asselin A. Some thaumatin-like proteins hydrolyse polymeric β -1,3-glucans. *Plant J.* 1999;19:473–480. doi:10.1046/j.1365-3113.1999.00551.x.
 172. Fierens E, Rombouts S, Gebruers K, et al. TLXI, a novel type of xylanase inhibitor from wheat (*Triticum aestivum*) belonging to the thaumatin family. *Biochem J.* 2007;403:583–591. doi:10.1042/BJ20061291.
 173. Schimoler-O'Rourke R, Richardson M, Selitrennikoff CP. Zeamatin inhibits trypsin and alpha-amylase activities. *Appl Environ Microbiol.* 2001;67:2365–2366. doi:10.1128/AEM.67.5.2365-2366.2001.
 174. Jung YC, Lee HJ, Yum SS, et al. Drought-inducible—but ABA-independent—thaumatin-like protein from carrot (*Daucus carota* L.). *Plant Cell Rep.* 2005;24:366–373. doi:10.1007/s00299-005-0944-x.
 175. Ruperti B, Cattivelli L, Pagni S, Ramina A. Ethylene-responsive genes are differentially regulated during abscission, organ senescence and wounding in peach (*Prunus persica*). *J Exp Bot.* 2002;53:429–437. doi:10.1093/jxbbot/53.368.429.
 176. Hon WC, Griffith M, Mlynarz A, Kwok YC, Yang DSC. Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. *Plant Physiol.* 1995;109:879–889. doi:10.1104/pp.109.3.879.
 177. Kumar KRR, Kirti PB. Differential gene expression in *Arachis diogenes* upon interaction with peanut late leaf spot pathogen, *Phaeoisariopsis personata* and characterization of a pathogen induced cyclophilin. *Plant Mol Biol.* 2011;75:497–513. doi:10.1007/s11103-011-9747-3.
 178. Rudd JJ, Kanyuka K, Hassani-Pak K, et al. Transcriptome and metabolite profiling of the infection cycle of *Zymoseptoria tritici* on wheat reveals a biphasic interaction with plant immunity involving differential pathogen chromosomal contributions and a variation on the hemibiotrophic lifestyle definition. *Plant Physiol.* 2015;167:1158–1185. doi:10.1104/pp.114.255927.
 179. Breiteneder H, Radauer C. A classification of plant food allergens. *J Allergy Clin Immunol.* 2004;113:821–830. doi:10.1016/j.jaci.2004.01.779.
 180. Jami SK, Swathi Anuradha T, Guruprasad L, Kirti PB. Molecular, biochemical and structural characterization of osmotin-like protein from black nightshade (*Solanum nigrum*). *J Plant Physiol.* 2007;164:238–252. doi:10.1016/j.jplph.2006.01.006.
 181. Tachi H, Fukuda-Yamada K, Kojima T, Shiraiwa M, Takahara H. Molecular characterization of a novel soybean gene encoding a neutral PR-5 protein induced by high-salt stress. *Plant Physiol Biochem.* 2009;47:73–79. doi:10.1016/j.plaphy.2008.09.012.
 182. Breiteneder H. Thaumatin-like proteins—a new family of pollen and fruit allergens. *Allergy.* 2004;59:479–481. doi:10.1046/j.1398-9995.2003.00421.x.
 183. Fierens E, Gebruers K, Voet ARD, De Maeyer M, Courtin CM, Delcour JA. Biochemical and structural characterization of TLXI, the *Triticum aestivum* L. *J Enzyme Inhib Med Chem.* 2009;24:646–654. doi:10.1080/14756360802321831.
 184. Leone P, Menu-Bouaouiche L, Peumans WJ, et al. Resolution of the structure of the allergenic and antifungal banana fruit thaumatin-like protein at 1.7-Å. *Biochimie.* 2006;88:45–52. doi:10.1016/j.biochi.2005.07.001.
 185. Koiwa H, Kato H, Nakatsu T, Oda J, Yamada Y, Sato F. Crystal structure of tobacco PR-5d protein at 1.8 Å resolution reveals a conserved acidic cleft structure in antifungal thaumatin-like proteins. *J Mol Biol.* 1999;286:1137–1145. doi:10.1006/jmbi.1998.2540.
 186. Min K, Ha SC, Hasegawa PM, Bressan RA, Yun D-J, Kim KK. Crystal structure of osmotin, a plant antifungal protein. *Proteins.* 2003;54:170–173. doi:10.1002/prot.10571.
 187. Ogata CM, Gordon PF, de Vos AM, Kim S-H. Crystal structure of a sweet tasting protein thaumatin I, at 1.65 Å resolution. *J Mol Biol.* 1992;228:893–908. doi:10.1016/0022-2836(92)90873-I.
 188. Dall'Antonia Y, Pavkov T, Fuchs H, Breiteneder H, Keller W. Crystallization and preliminary structure determination of the plant food allergen Pru av 2. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2005;61:186–188. doi:10.1107/S1744309104033822.
 189. Oshero N, Mathew J, Romans A, May GS. Identification of conidia-enriched transcripts in *Aspergillus nidulans* using suppression subtractive hybridization. *Fungal Genet Biol.* 2002;37:197–204. doi:10.1016/S1087-1845(02)00502-9.
 190. Greenstein S, Shadkhan Y, Jadoun J, Sharon C, Markovich S, Oshero N. Analysis of the *Aspergillus nidulans* thaumatin-like cetA gene and evidence for transcriptional repression of pyr4 expression in the cetA-disrupted strain. *Fungal Genet Biol.* 2006;43:42–53. doi:10.1016/j.fgb.2005.10.001.
 191. Wang X, Zafian P, Choudhary M, Lawton M. The PR5K receptor protein kinase from *Arabidopsis thaliana* is structurally related to a family of plant defense proteins. *Proc Natl Acad Sci USA.* 1996;93:2598–2602. doi:10.1073/pnas.93.6.2598.
 192. Apweiler R. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 2004;32:115–119. doi:10.1093/nar/gkh131.
 193. Torrent M, Nogues MV, Boix E. Discovering new in silico tools for antimicrobial peptide prediction. *Curr Drug Targets.* 2012;13:1148–1157. doi:10.2174/138945012802002311.
 194. Porto WF, Pires AS, Franco OL. Computational tools for exploring sequence databases as a resource for antimicrobial peptides. *Biotechnol Adv.* 2017;35:337–349. doi:10.1016/j.biotechadv.2017.02.001.
 195. Thomas S, Karnik S, Barai RS, Jayaraman VK, Idicula-Thomas S. CAMP: a useful resource for research on antimicrobial peptides. *Nucleic Acids Res.* 2010;38:D774–D780. doi:10.1093/nar/gkp1021.
 196. Waghu FH, Gopi L, Barai RS, Ramteke P, Nizami B, Idicula-Thomas S. CAMP: collection of sequences and structures of antimicrobial peptides. *Nucleic Acids Res.* 2014;42:D1154–D1158. doi:10.1093/nar/gkt1157.
 197. Waghu FH, Barai RS, Gurung P, Idicula-Thomas S. CAMP3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res.* 2016;44:D1094–D1097. doi:10.1093/nar/gkv1051.
 198. Wang G, Li X, Wang Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* 2016;44:D1087–D1093. doi:10.1093/nar/gkv1278.
 199. Zhao X, Wu H, Lu H, Li G, Huang Q. LAMP: a database linking antimicrobial peptides. *PLoS ONE.* 2013;8:e66557. doi:10.1371/journal.pone.0066557.
 200. Gogoladze G, Grigolava M, Vishnepolsky B, et al. DBAASP: database of antimicrobial activity and structure of peptides. *FEMS Microbiol Lett.* 2014;357:63–68. doi:10.1111/1574-6968.12489.
 201. Jorge P, Lourenço A, Pereira MO. New trends in peptide-based anti-biofilm strategies: a review of recent achievements and bioinformatic approaches. *Bio-fouling.* 2012;28:1033–1061. doi:10.1080/08927014.2012.728210.
 202. Lee HT, Lee CC, Yang JR, Lai JZC, Chang KY. A large-scale structural classification of antimicrobial peptides. *Biomed Res Int.* 2015;2015:475062. doi:10.1155/2015/475062.
 203. Seebah S, Suresh A, Zhuo S, et al. Defensins knowledgebase: a manually curated database and information source focused on the defensins family of antimicrobial peptides. *Nucleic Acids Res.* 2007;35:D265–D268. doi:10.1093/nar/gkl866.
 204. Liu S, Fan L, Sun J, Lao X, Zheng H. Computational resources and tools for antimicrobial peptides. *J Pept Sci.* 2017;23:4–12. doi:10.1002/psc.2947.
 205. Mulvenna JP. CyBase: a database of cyclic protein sequence and structure. *Nucleic Acids Res.* 2006;34:D192–D194. doi:10.1093/nar/gkj005.
 206. Hammami R, Ben Hamida J, Vergoten G, Fliss I. PhytAMP: a database dedicated to antimicrobial plant peptides. *Nucleic Acids Res.* 2009;37:D963–D968. doi:10.1093/nar/gkn655.
 207. Das D, Jaiswal M, Khan FN, Ahamad S, Kumar S. PlantPepDB: a manually curated plant peptide database. *Sci Rep.* 2020;10:2194.
 208. Eltabakh MY, Ouzzani M, Aref WG. bdbms—a database management system for biological data. <https://arxiv.org/pdf/cs/0612127.pdf>. Published 2006.
 209. Rahman M, Shaheen T, Rahman M, Iqbal MA, Zafar Y. Bioinformatics: a way forward to explore “plant omics.” In: Abdurakhmonov, IY ed. *Bioinformatics: Updated Features and Applications*. London, England: IntechOpen; 2016. doi:10.5772/64043.
 210. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–410. doi:10.1016/S0022-2836(05)80360-2.
 211. Pearson WR. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* 1990;183:63–98. doi:10.1016/0076-6879(90)83007-V.
 212. Polyakov VO, Roytberg MA, Tumanyan VG. Comparative analysis of the quality of a global algorithm and a local algorithm for alignment of two sequences. *Algorithms Mol Biol.* 2011;6:25. doi:10.1186/1748-7188-6-25.
 213. Thompson K. Programming techniques: regular expression search algorithm. *Commun ACM.* 1968;11:419–422. doi:10.1145/363347.363387.

214. Eddy SR. Profile hidden Markov models. *Bioinformatics*. 1998;14:755-763. doi:10.1093/bioinformatics/14.9.755.
215. Yoon B-J. Hidden Markov models and their applications in biological sequence analysis. *Curr Genomics*. 2009;10:402-415. doi:10.2174/138920209789177575.
216. Zhou P, Yang C, Ren Y, Wang C, Tian F. What are the ideal properties for functional food peptides with antihypertensive effect? A computational peptidology approach. *Food Chem*. 2013;141:2967-2973. doi:10.1016/j.foodchem.2013.05.140.
217. Zhou P, Huang J, eds. *Computational Peptidology*. New York, NY: Springer; 2015. doi:10.1007/978-1-4939-2285-7.
218. Niarchou A, Alexandridou A, Athanasiadis E, Spyrou G. C-PAMP: large scale analysis and database construction containing high scoring computationally predicted antimicrobial peptides for all the available plant species. *PLoS ONE*. 2013;8:e79728. doi:10.1371/journal.pone.0079728.
219. Islam SMA, Sajed T, Kearney CM, et al. PredSTP: a highly accurate SVM based model to predict sequential cystine stabilized peptides. *BMC Bioinformatics*. 2015;16:210.
220. Islam SMA, Kearney CM, Baker EJ. Assigning biological function using hidden signatures in cystine-stabilized peptide sequences. *Sci Rep*. 2018;8:9049.
221. Lin Y, Jeon Y. Random forests and adaptive nearest neighbors. *J Am Stat Assoc*. 2006;101:578-590. doi:10.1198/016214505000001230.
222. Karasuyama M, Takeuchi I. Multiple incremental decremental learning of support vector machines. *IEEE Trans Neural Netw*. 2010;21:1048-1059. doi:10.1109/TNN.2010.2048039.
223. Ding S, Li H, Su C, Yu J, Jin F. Evolutionary artificial neural networks: a review. *Artif Intell Rev*. 2013;39:251-260. doi:10.1007/s10462-011-9270-6.
224. Dobsiaff K, Kreisig T, Berthold N, Hoffmann R, Zuchner T. Novel peptide-protein assay for identification of antimicrobial peptides by fluorescence quenching. *Anal Bioanal Chem*. 2012;403:2725-2731. doi:10.1007/s00216-012-6050-3.
225. Fedders H, Leippe M. A reverse search for antimicrobial peptides in *Ciona intestinalis*: identification of a gene family expressed in hemocytes and evaluation of activity. *Dev Comp Immunol*. 2008;32:286-298. doi:10.1016/j.dci.2007.06.003.
226. Tennessen JA. Positive selection drives a correlation between non-synonymous/synonymous divergence and functional divergence. *Bioinformatics*. 2008;24:1421-1425. doi:10.1093/bioinformatics/btn205.
227. Wilkins MR, Sanchez J-C, Gooley AA, et al. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev*. 1996;13:19-50. doi:10.1080/02648725.1996.10647923.
228. Wasinger VC, Corthals GL. Proteomic tools for biomedicine. *J Chromatogr B*. 2002;771:33-48. doi:10.1016/S1570-0232(02)00125-3.
229. Sheoran IS, Ross ARS, Olson DJH, Sawhney VK. Compatibility of plant protein extraction methods with mass spectrometry for proteome analysis. *Plant Sci*. 2009;176:99-104. doi:10.1016/j.plantsci.2008.09.015.
230. Muroi M, Osada H. Proteomic profiling for target identification of biologically active small molecules using 2D DIGE. In: Ziegler, S, Waldmann, H, eds. *Systems Chemical Biology*. New York, NY: Springer; 2019:127-139. doi:10.1007/978-1-4939-8891-4_7.
231. Cho WCS. Proteomics technologies and challenges. *Genomics Proteomics Bioinformatics*. 2007;5:77-85. doi:10.1016/S1672-0229(07)60018-7.
232. Baracat-Pereira MC, de Oliveira Barbosa M, Magalhães Júnior MJ, et al. Separomics applied to the proteomics and peptidomics of low-abundance proteins: choice of methods and challenges—a review. *Genet Mol Biol*. 2012;35:283-291. doi:10.1590/S1415-47572012000200009.
233. Kołodziejek I, van der Hoorn RA. Mining the active proteome in plant science and biotechnology. *Curr Opin Biotechnol*. 2010;21:225-233. doi:10.1016/j.copbio.2010.02.003.
234. Al Akeel R, Mateen A, Syed R, Alyousef A, Shaik M. Screening, purification and characterization of anionic antimicrobial proteins from *Foeniculum vulgare*. *Molecules*. 2017;22:602. doi:10.3390/molecules22040602.
235. Menschaert G, Vandekerckhove TTM, Baggerman G, Schoofs L, Luyten W, Crieckinge WV. Peptidomics coming of age: a review of contributions from a bioinformatics angle. *J Proteome Res*. 2010;9:2051-2061. doi:10.1021/pr900929m.
236. Nägele E, Vollmer M, Hörth P, Vad C. 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. *Expert Rev Proteomics*. 2004;1:37-46. doi:10.1586/14789450.1.1.37.
237. Mitulovic G, Mechtler K. HPLC techniques for proteomics analysis—a short overview of latest developments. *Brief Funct Genomic Proteomic*. 2006;5:249-260. doi:10.1093/bfgp/ell034.
238. Blueggel M, Chamrad D, Meyer H. Bioinformatics in proteomics. *Curr Pharm Biotechnol*. 2004;5:79-88. doi:10.2174/1389201043489648.
239. McHugh L, Arthur JW. Computational methods for protein identification from mass spectrometry data. *PLoS Comput Biol*. 2008;4:e12. doi:10.1371/journal.pcbi.0040012.
240. Hughes C, Ma B, Lajoie GA. De novo sequencing methods in proteomics. In: Hubbard, SJ, Jones, AR, eds. *Proteome Bioinformatics*. Totowa, NJ: Humana Press; 2010:105-121. doi:10.1007/978-1-60761-444-9_8.
241. Eng JK, Fischer B, Grossmann J, Maccoss MJ. A fast SEQUEST cross correlation algorithm. *J Proteome Res*. 2008;7:4598-4602. doi:10.1021/pr800420s.
242. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999;20:3551-3567.
243. Tsai WC, Zhuang ZJ, Lin CY, Chen WJ. Novel antimicrobial peptides with promising activity against multidrug resistant *Salmonella enterica* serovar Choleraesuis and its stress response mechanism. *J Appl Microbiol*. 2016;121:952-965. doi:10.1111/jam.13203.
244. Umadevi P, Soumya M, George JK, Anandaraj M. Proteomics assisted profiling of antimicrobial peptide signatures from black pepper (*Piper nigrum* L.). *Physiol Mol Biol Plants*. 2018;24:379-387. doi:10.1007/s12298-018-0524-5.
245. Wang Z. APD: the antimicrobial peptide database. *Nucleic Acids Res*. 2004;32:590D-592D.
246. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. In: Walker, JM, ed. *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press; 2005:571-607. doi:10.1385/1-59259-890-0.
247. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*. 2004;20:1466-1467. doi:10.1093/bioinformatics/bth092.
248. Geer LY, Markey SP, Kowalak JA, et al. Open mass spectrometry search algorithm. *J Proteome Res*. 2004;3:958-964. doi:10.1021/pr0499491.
249. Zhang N, Aebersold R, Schwikowski B. ProBID: a probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. *Proteomics*. 2002;2:1406-1412.
250. Field HI, Fenyö D, Beavis RC. RADARS, a bioinformatics solution that automates proteome mass spectral analysis, optimises protein identification, and archives data in a relational database. *Proteomics*. 2002;2:36-47.
251. Nesvizhskii AI, Vitek O, Aebersold R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods*. 2007;4:787-797. doi:10.1038/nmeth1088.
252. Gazzana G, Borlak J. Improved method for proteome mapping of the liver by 2-DE MALDI-TOF MS. *J Proteome Res*. 2007;6:3143-3151. doi:10.1021/pr070097p.
253. Pestana-Calsa MC, Ribeiro IL, Calsa T Jr. Bioinformatics-coupled molecular approaches for unravelling potential antimicrobial peptides coding genes in Brazilian native and crop plant species. *Curr Protein Pept Sci*. 2010;11:199-209. doi:10.2174/138920310791112138.
254. Tejano LA, Peralta JP, Yap EES, Panjaitan FCA, Chang Y-W. Prediction of bioactive peptides from *Chlorella sorokiniana* proteins using proteomic techniques in combination with bioinformatics analyses. *Int J Mol Sci*. 2019;20:1786. doi:10.3390/ijms20071786.
255. Weiner AK, Sidoli S, Diskin SJ, Garcia BA. Graphical interpretation and analysis of proteins and their ontologies (GiaPronto): a one-click graph visualization software for proteomics data sets. *Mol Cell Proteomics*. 2018;17:1426-1431. doi:10.1074/mcp.TIR117.000438.
256. Jothi A. Principles, challenges and advances in ab initio protein structure prediction. *Protein Pept Lett*. 2012;19:1194-1204. doi:10.2174/092986612803217015.
257. Kaufmann KW, Lemmon GH, DeLuca SL, Sheehan JH, Meiler J. Practically useful: what the ROSETTA protein modeling suite can do for you. *Biochemistry*. 2010;49:2987-2998. doi:10.1021/bi902153g.
258. Moutl J, Fidelis K, Kryshatfovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)—round XII. *Proteins*. 2018;86:7-15. doi:10.1002/prot.25415.
259. Källberg M, Wang H, Wang S, et al. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*. 2012;7:1511-1522. doi:10.1038/nprot.2012.085.
260. Marti-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Šali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*. 2000;29:291-325. doi:10.1146/annurev.biophys.29.1.291.
261. Biasini M, Bienert S, Waterhouse A, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. 2014;42:W252-W258. doi:10.1093/nar/gku340.
262. Webb B, Sali A. Comparative protein structure modeling using MODELLER: comparative protein structure modeling using modeller. In: Bateman, A, Pearson, WR, Stein, LD, Stormo, GD, Yates, JR eds. *Current Protocols in Bioinformatics*. Hoboken, NJ: John Wiley & Sons, Inc.; 2016:5.6.1-5.6.37. doi:10.1002/cpbi.3.
263. Wu S, Zhang Y. MUSTER: improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins*. 2008;72:547-556. doi:10.1002/prot.21945.
264. Yang Y, Faraggi E, Zhao H, Zhou Y. Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between

- predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics*. 2011;27:2076-2082. doi:10.1093/bioinformatics/btr350.
265. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 2007;35:W407-W410. doi:10.1093/nar/gkm290.
 266. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008;9:40. doi:10.1186/1471-2105-9-40.
 267. Zhang Y, Kolinski A, Skolnick J. TOUCHSTONE II: a new approach to ab initio protein structure prediction. *Biophys J*. 2003;85:1145-1164. doi:10.1016/S0006-3495(03)74551-2.
 268. Wu S, Skolnick J, Zhang Y. Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biol*. 2007;5:17. doi:10.1186/1741-7007-5-17.
 269. Zhang W, Yang J, He B, et al. Integration of QUARK and I-TASSER for ab initio protein structure prediction in CASP11: ab initio structure prediction in CASP11. *Proteins*. 2016;84:76-86.
 270. Haas J, Barbato A, Behringer D, et al. Continuous Automated Model Evaluation (CAMEO) complementing the critical assessment of structure prediction in CASP12. *Proteins*. 2018;86:387-398. doi:10.1002/prot.25431.
 271. Moulton J, Fidelis K, Kryshtalovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction: progress and new directions in round XI: progress in CASP XI. *Proteins*. 2016;84:4-14. doi:10.1002/prot.25064.
 272. Porto WF, Irazazabal L, Alves ESF, et al. In silico optimization of a guava antimicrobial peptide enables combinatorial exploration for peptide design. *Nat Commun*. 2018;9:1490.
 273. Song Y, DiMaio F, Wang RYR, et al. High-resolution comparative modeling with RosettaCM. *Structure*. 2013;21:1735-1742.
 274. Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc Natl Acad Sci USA*. 2018;115:E4758-E4766. doi:10.1073/pnas.1800690115.
 275. Freddolino PL, Harrison CB, Liu Y, Schulten K. Challenges in protein-folding simulations. *Nat Phys*. 2010;6:751-758. doi:10.1038/nphys1713.
 276. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Biol*. 2002;9:646-652. doi:10.1038/nsb0902-646.
 277. Mitsutake A, Takano H. Relaxation mode analysis for molecular dynamics simulations of proteins. *Biophys Rev*. 2018;10:375-389. doi:10.1007/s12551-018-0406-7.
 278. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*. 2008;4:435-447. doi:10.1021/ct700301q.
 279. Pronk S, Páll S, Schulz R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 2013;29:845-854. doi:10.1093/bioinformatics/btt055.
 280. Abraham MJ, Murtola T, Schulz R, et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1:19-25. doi:10.1016/j.softx.2015.06.001.
 281. Case DA, Cheatham TE 3rd, Darden T, et al. The Amber biomolecular simulation programs. *J Comput Chem*. 2005;26:1668-1688. doi:10.1002/jcc.20290.
 282. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005;26:1781-1802. doi:10.1002/jcc.20289.
 283. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem*. 1983;4:187-217. doi:10.1002/jcc.540040211.
 284. Plimpton S. Fast parallel algorithms for short-range molecular dynamics. *J Comput Phys*. 1995;117:1-19. doi:10.1006/jcph.1995.1039.
 285. Bowers KJ, Chow DE, Xu H, et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. Paper presented at: ACM/IEEE SC 2006 Conference (SC'06); November 11-17, 2006:43; Tampa, FL. doi:10.1109/SC.2006.54.
 286. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. *Neuron*. 2018;99:1129-1143. doi:10.1016/j.neuron.2018.08.011.
 287. Meylan S, Andrews IW, Collins JJ. Targeting antibiotic tolerance, pathogen by pathogen. *Cell*. 2018;172:1228-1238. doi:10.1016/j.cell.2018.01.037.
 288. Hay SI, Rao PC, Dolecek C, et al. Measuring and mapping the global burden of antimicrobial resistance. *BMC Med*. 2018;16:78. doi:10.1186/s12916-018-1073-z.
 289. Neu HC, Gootz TD. Antimicrobial chemotherapy. In: Baron, S, ed. *Medical Microbiology*. 4th ed. Galveston, TX: University of Texas Medical Branch at Galveston; 1996:1-10. <http://www.ncbi.nlm.nih.gov/books/NBK7986/>
 290. Hong W, Zeng J, Xie J. Antibiotic drugs targeting bacterial RNAs. *Acta Pharm Sin B*. 2014;4:258-265. doi:10.1016/j.apsb.2014.06.012.
 291. Cheng G, Dai M, Ahmed S, Hao H, Wang X, Yuan Z. Antimicrobial drugs in fighting against antimicrobial resistance. *Front Microbiol*. 2016;7:470. doi:10.3389/fmicb.2016.00470.
 292. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: current status and future challenges. *Proteins*. 2006;65:15-26.
 293. London N, Raveh B, Schueler-Furman O. Peptide docking and structure-based characterization of peptide binding: from knowledge to know-how. *Curr Opin Struct Biol*. 2013;23:894-902.
 294. Pagadala NS, Syed K, Tuszyński J. Software for molecular docking: a review. *Biophys Rev*. 2017;9:91-102. doi:10.1007/s12551-016-0247-1.
 295. Sun J, Deng Z, Yan A. Bacterial multidrug efflux pumps: mechanisms, physiology and pharmacological exploitations. *Biochem Biophys Res Commun*. 2014;453:254-267. doi:10.1016/j.bbrc.2014.05.090.
 296. Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics*. 2014;30:1771-1773.
 297. van Zundert GCP, Rodrigues JPGLM, Trellet M, et al. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J Mol Biol*. 2016;428:720-725. doi:10.1016/j.jmb.2015.09.014.
 298. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, et al. rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput Biol*. 2014;10:e1003571. doi:10.1371/journal.pcbi.1003571.
 299. Ghooorah AW, Devignes M-D, Smail-Tabbone M, Ritchie DW. Protein docking using case-based reasoning. *Proteins*. 2013;81:2150-2158. doi:10.1002/prot.24433.
 300. Andrusier N, Mashiach E, Nussinov R, Wolfson HJ. Principles of flexible protein-protein docking. *Proteins*. 2008;73:271-289. doi:10.1002/prot.22170.
 301. Schneider T, Kruse T, Wimmer R, et al. Plectasin, a fungal defensin, targets the bacterial cell wall precursor lipid II. *Science*. 2010;328:1168-1172. doi:10.1126/science.1185723.
 302. Corbeil CR, Williams CI, Labute P. Variability in docking success rates due to dataset preparation. *J Comput Aided Mol Des*. 2012;26:775-786. doi:10.1007/s10822-012-9570-1.
 303. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. *J Mol Bio*. 1997;267:727-748. doi:10.1006/jmbi.1996.0897.
 304. Friesner RA, Murphy RB, Repasky MP, et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem*. 2006;49:6177-6196. doi:10.1021/jm051256o.
 305. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: current status and future challenges. *Proteins*. 2006;65:15-26.
 306. Spitzer R, Jain AN. Surflex-dock: docking benchmarks and real-world application. *J Comput Aided Mol Des*. 2012;26:687-699. doi:10.1007/s10822-011-9533-y.
 307. Rentzsch R, Renard BY. Docking small peptides remains a great challenge: an assessment using AutoDock Vina. *Brief Bioinform*. 2015;16:1045-1056. doi:10.1093/bib/bbv008.
 308. Audie J, Swanson J. Advances in the prediction of protein-peptide binding affinities: implications for peptide-based drug discovery: protein-peptide binding affinities. *Chem Biol Drug Des*. 2013;81:50-60. doi:10.1111/cbdd.12076.
 309. Audie J, Swanson J. Recent work in the development and application of protein-peptide docking. *Future Med Chem*. 2012;4:1619-1644. doi:10.4155/fmc.12.99.
 310. London N, Raveh B, Schueler-Furman O. Peptide docking and structure-based characterization of peptide binding: from knowledge to know-how. *Curr Opin Struct Biol*. 2013;23:894-902.
 311. Sousa SF, Ribeiro AJM, Coimbra JTS, et al. Protein-ligand docking in the new millennium—a retrospective of 10 years in the field. *Curr Med Chem*. 2013;20:2296-2314. doi:10.2174/0929867311320180002.
 312. Wilson A. Inhibition of protein-protein interactions using designed molecules. *Chem Soc Rev*. 2009;38:3289-3300. doi:10.1039/b807197g.
 313. Unal EB, Gursoy A, Erman B, VitALE: Viterbi algorithm for de novo peptide design. *PLoS ONE*. 2010;5:e10926. doi:10.1371/journal.pone.0010926.
 314. Lensink MF, Velankar S, Wodak SJ. Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins*. 2017;85:359-377. doi:10.1002/prot.25215.
 315. Lensink MF, Wodak SJ. Docking, scoring, and affinity prediction in CAPRI. *Proteins*. 2013;81:2082-2095. doi:10.1002/prot.24428.
 316. Barh D, Tiwari S, Silva Andrade B, et al. Potential chimeric peptides to block the SARS-CoV-2 spike receptor-binding domain. *F1000Research*. 2020;9:576.
 317. Pant S, Singh M, Ravichandiran V, Murty US, Srivastava HK. Peptide-like and small-molecule inhibitors against Covid-19 [published online ahead of print May 6, 2020]. *J Biomol Struct Dyn*. doi:10.1080/07391102.2020.1757510.