

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

FLÁVIA TADEU DE ARAÚJO

**Genes de resistência a patógenos em feijão-
caupi e em outras leguminosas:
caracterização e diversidade**

RECIFE
2015

FLÁVIA TADEU DE ARAÚJO

**Genes de resistência a patógenos em feijão-caupi e em
outras leguminosas: caracterização e diversidade**

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

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

Coorientadora: Prof^a Dr^a Lidianne L. Barbosa Amorim

**RECIFE
2015**

Catálogo na fonte
Elaine Barroso
CRB 1728

Araújo, Flávia Tadeu de

Genes de resistência a patógenos em feijão-caupi e em outras leguminosas: caracterização e diversidade/ Flávia Tadeu de Araújo– Recife: O Autor, 2015

146 folhas : il., fig., tab.

Orientadora: Ana Maria Benko Iseppon

Coorientadora: Lidiane L. Barbosa Amorim

Dissertação (mestrado) – Universidade Federal de Pernambuco.

Centro de Ciências Biológicas. Genética, 2015

Inclui bibliografia e anexos

- 1. Genética vegetal 2. Feijão-caupi I. Iseppon, Ana Maria Benko (orientadora) II. Amorim, Lidiane L. Barbosa (coorientadora) III. Título**

581.35

CDD (22.ed.)

UFPE/CCB-2015-204

FLÁVIA TADEU DE ARAÚJO

**Genes de resistência a patógenos em feijão-caupi e em
outras leguminosas: caracterização e diversidade**

Aprovada em 02 / 03 / 2015

Banca Examinadora:

Profa. Dr^a. Ana Maria Benko-Iseppon
Universidade Federal de Pernambuco

Prof. Dr. Éderson Akio Kido
Universidade Federal de Pernambuco

Dr. José Ribamar Ferreira Neto
Universidade Federal de Pernambuco

Dr^a. Valesca Pandolfi
Universidade Federal de Pernambuco

RECIFE

2015

*À minha querida Mãe, Maria Czekalski,
fonte diária de inspiração.*

Dedico.

Agradecimentos

A Deus, que está presente constantemente em minha vida, guiando e iluminando os meus passos.

À Pós-Graduação em Genética da Universidade Federal de Pernambuco, pela obtenção do título de Mestre.

À FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco) pelo apoio financeiro, à Embrapa Meio Norte e ao IPA, em especial ao Dr. Antonio Félix da Costa, pela concessão do material genético utilizado nessa pesquisa.

À minha orientadora Prof^a Dr^a Ana Maria Benko-Iseppon, pela oportunidade de fazer parte do seu grupo de pesquisa no Laboratório de Genética e Biotecnologia Vegetal (LGBV-UFPE), gerando conhecimento científico e aperfeiçoando a minha formação profissional.

À minha coorientadora Prof^a Dr^a Lidianne Amorim, pelos ensinamentos, conselhos, broncas, confiança, paciência e amizade.

Ao meu amigo M.Sc. João Pacífico, pelo companheirismo diário, disponibilidade constante e paciência infinita, o que o tornou essencial para o desenvolvimento dessa pesquisa, além dos necessários momentos de risadas.

À minha família, em especial à minha Mãe, Maria Czekalski, por sempre apoiar e incentivar as minhas escolhas e a todos os meus irmãos, Pai e cunhado Daniel que de uma forma ou de outra me fortaleceram para chegar até aqui.

À minha amiga Mitalle Matos, por todo conhecimento e desespero compartilhados, pelos momentos essenciais de descontração e por pensar diferente de mim, o que permitiu que evoluíssemos juntas.

À minha amiga Bruna Piereck por todos os pensamentos científicos construídos em conjunto, por todo carinho, afeto e palavras de conforto nos momentos em que mais precisei.

A todos os membros do LGBV pela convivência, pelo apoio e incentivo constante, em especial à Ana Carolina, Neto, Valesca, Manassés, Wilson, Roberta, Sheyla, Hévila, Marx e Santelmo, pela paciência e ajuda incondicional, assim como à Vanessa, Ivanberto e Jaysa por todo suporte técnico, incentivo, orações e amizade.

“As nuvens mudam sempre de posição, mas são sempre nuvens no céu. Assim devemos ser todo dia, mutantes, porém leais com o que pensamos e sonhamos; lembre-se, tudo se desmancha no ar, menos os pensamentos”.

Paulo Beleki

RESUMO

A cultura do feijão-caupi [*Vigna unguiculata* (L.) Walp.] apresenta importância econômica em nível internacional, entretanto ela é frequentemente acometida por uma diversidade de patógenos. Nesse contexto a família gênica NBS-LRR de genes de Resistência (*R*) se destaca devido ao seu papel fundamental na defesa das plantas contra o ataque de patógenos, sendo a maior e mais diversificada família desse grupo. O objetivo do presente estudo foi caracterizar genes NBS-LRR em feijão-caupi, desenvolver marcadores moleculares RGA (*Resistance Gene Analogs*) e validar genes diferencialmente expressos em uma interação planta-vírus. Inicialmente, sequências candidatas para genes NBS-LRR foram obtidas no banco de dados NordEST, procedendo-se com a anotação dos dados, tradução e identificação dos domínios conservados por meio de ferramentas *in silico*. Um total de 57 sequências NBS-LRR completas foi identificado em feijão-caupi. Como as proteínas codificadas pelos genes *R* apresentam domínios e motivos conservados, foi possível desenvolver marcadores RGAs usando as sequências de feijão-caupi como sonda contra o banco de *Phaseolus vulgaris* L. Foram desenhados 16 pares de iniciadores para *P. vulgaris*, identificando-se um percentual de 87,5% de transferibilidade para o feijão-caupi. Destes, dois foram polimórficos e apresentaram segregação mendeliana em uma população de mapeamento para o vírus do mosaico severo do feijão-caupi (CPSMV). Os 57 candidatos foram ancorados nos 20 pseudocromossomos de *Glycine max* (L.) Merr., verificando-se repetições *in tandem* deste grupo gênico. A análise de expressão gênica diferencial *in silico* foi realizada utilizando dados de RNAseq e SuperSAGE. A validação da expressão gênica via RT-qPCR foi através do desenho de *primers*, com os dados de SuperSAGE, onde três genes alvo apresentaram indução nos níveis de expressão após 16 horas da inoculação com o patógeno. Esses resultados mostram-se valiosos para o melhoramento genético do feijão-caupi.

Palavras-chave: *V. unguiculata*; NBS-LRR; *Resistance Gene Analog*; expressão gênica.

ABSTRACT

The cowpea [*Vigna unguiculata* (L.) Walp.] culture has an international economic importance, however it is often affected by a diversity of pathogens. In this context the NBS-LRR family of resistance (R) genes stands out because of its key role in plant defense against pathogen attack, being the largest and most diverse family of this group. The present work aimed the characterization of NBS-LRR genes from cowpea, the development of RGA (*resistance gene analogs*) markers and the validation of differentially expressed genes in plant-virus interaction. Initially, NBS-LRR gene candidate sequences were obtained from NordEST database, proceeding with data annotation, translation and identification of conserved domains through *in silico* methods. A total of 57 NBS-LRR complete sequences were identified for cowpea. Since *R*-gene encoded proteins exhibit conserved domains and motifs, it was possible to develop RGA markers using cowpea sequences as probes against *Phaseolus vulgaris* L. Sixteen primer pairs of *P. vulgaris* were designed, from which 87.5% were transferable to cowpea. From those, two were polymorphic and showed Mendelian segregation in a mapping population for the cowpea severe mosaic virus (CPSMV). The 57 candidates were anchored in the 20 *Glycine max* (L.) Merr. pseudo-chromosomes, revealing *in tandem* repetitions for this group of gene. Differential gene expression analysis *in silico* was performed using data from RNAseq and SuperSAGE. The validation of gene expression by RT-qPCR was carried out after design of *primers* using SuperSAGE data, from which three target genes presented induction in expression levels at 16 hours after the pathogen inoculation. These results represent valuable data for the genetic improvement of cowpea.

Keywords: *V. unguiculata*; NBS-LRR; Resistance Gene Analog; gene expression.

LISTA DE ILUSTRAÇÕES

REVISÃO BIBLIOGRÁFICA

Figura 1: Representação esquemática dos domínios conservados presentes na família gênica NBS-LRR. N - Amino-terminal, TIR - Receptor Toll- Interleucina, CC - Cauda Espiralada, NBS - Sitio de Ligação ao Nucleotideo, LRR - Repetições Ricas em Leucina, C - Carboxi-terminal, WRKY - Fator de Transcrição.....23

CAPÍTULO I

Figure 1: Main steps of positive (resistant) and negative (susceptible) plant-pathogen interactions using a fungal and a viral pathogens, respectively. PAMPs or MAMPs (Microbe-Associated Molecular Patterns) and DAMPs (Damage-Associated Molecular Patterns) trigger unspecific response due to transcriptional regulation. The positive commensal interaction of plants and soil microorganisms is also illustrated.....42

Figure 2: Graphic representation of five known *R*-genes families including I: Reductases; II: Kinases; III: NBS-LRR; IV: LRR; V: LRR-Kinases. Different colors represent the conserved domains and position is relative to cell membrane. Abbreviations of conserved domains: NBS = Nucleotide Binding Site; LRR = Leucine Rich Repeats; TM = Transmembrane; TIR = Toll-Interleukin Receptor.....49

Figure 3: Anchoring of selected representative gene candidates (*N*, *P2*, *RPP5*, *RPS4*, *L6*, *SSI4*, *RPP4*, *M* and *RPP1*) of the TIR-NBS-LRR gene family (*Toll Interleucine Receptor-Nucleotide Binding Site-Leucine Rich Repeats*) from cowpea [*Vigna unguiculata* L. (Walp.)] from NordEst data bank in *Phaseolus vulgaris* L. genome, available at Phytozome. From the 11 *P. vulgaris* chromosomes, nine presented clusters of this *R*-gene family, mostly in the subterminal position of chromosome arms. The largest number was observed in chromosome 10 (nine

gene categories). The most represented gene was *RPP5* found in eight chromosomes, represented by 41 loci, while *Rpp1* was the less represented, found only in two loci. Tandem repetitions (thicker lines) were observed in chromosomes 2, 7, 10 and 11, indicating possible duplication events.....52

CAPÍTULO II

Figura 1: Alinhamento do domínio NBS presente nas 57 sequências candidatas em feijão-caupi, com os seus motivos conservados *p-loop*, NBS-II, *Kinase-2*, *Kinase-3*, NBS-V e GLPL evidenciados pelos retângulos pretos. Na base da figura encontram-se os gráficos da qualidade do alinhamento e o consenso cujas barras indicam níveis de conservação86 e 87

Figura 2: Alinhamento do domínio LRR presente nos genes candidatos da categoria NBS-LRR em feijão-caupi, caracterizados pela conservação dos resíduos de leucina (em azul).....88

Figura 3: Distribuição e relação da ancoragem de genes, considerando candidatos a genes NBS-LRR identificados em feijão-caupi contra o genoma da soja ($2n = 40$). Os cromossomos são retratados em faixas pretas e as relações de genes ancorados são mostradas como links entre as regiões cromossômicas (linhas coloridas que interligam os cromossomos)90

Figura 4: Árvore fenética construída usando o método de *Neighbor-Joining* para inferência sobre a diversidade de candidatos a genes NBS-LRR em feijão-caupi. Números na base dos ramos referem-se aos valores de *bootstrap*. O grupo I contém os candidatos da subfamília TIR, enquanto o grupo II inclui os candidatos da subfamília não-TIR.....93

Figura 5: Dendrograma gerado pela análise de máxima parcimônia mostrando as relações entre ortólogos ao candidato 'contig98'. Os números na base dos ramos referem-se ao nível de confiabilidade do alinhamento, ou seja, aos valores de *bootstrap*. O círculo vermelho representa a subclasse Asteridae e o círculo azul a

subclasse Rosidae. Chaves ao lado direito do dendrograma indicam da base para o topo: (1) a ordem Fabales (em laranja), (2) ordem Malpighiales (em verde) e (3) a subclasse Asteridae (em roxo)94

Figura 6: *Heatmap* construindo a partir da contagem de *reads* dos genes NBS-LRR em dois genótipos de feijão-caupi contrastantes a seca, para análise de expressão *in silico*. A. Corresponde a expressão do genótipo tolerante a seca (Pingo de Ouro). B. Expressão dos genes no genótipo sensível (Santo Inácio). Na visualização gráfica, a cor branca indica a ausência de expressão e os níveis de azul, a intensidade de expressão na biblioteca em análise, com o intuito de se identificar os principais sítios de expressão dos genes NBS-LRR em feijão-caupi. Quanto maior a intensidade da cor azul, maior o nível de expressão do gene analisado.....96

Figura 7: Representação gráfica dos níveis de expressão relativa dos genes alvo. T30C30 (tratamento com vírus tempo 30 min contra controle tempo 30 min), T60C30 (tratamento com vírus tempo 60 min contra controle tempo 30 min) e T16C16 (tratamento com vírus tempo 16 horas contra controle tempo 16 horas). Genes alvo: VuRGAq3, VuRGAq4, VuRGAq6, VuRGAq10, sendo os três primeiros *up-regulados* em C16T16. Na vertical está demonstrada a taxa de expressão dos genes.....100

LISTA DE TABELAS

CAPÍTULO II

Tabela 1: *Primers* desenhados para identificação de genes NBS-LRR em feijão-caupi, com base no genoma do feijoeiro (*Phaseolus vulgaris*)91

Tabela 2: Dez pares de *primers* desenvolvidos com base nas *tags* selecionadas, para análise da expressão gênica de genes NBS-LRR em feijão-caupi via RT-qPCR.....98

Tabela 3: Dados de expressão gênica dos genes alvo, obtidos a partir do programa REST2009, entre as diferentes combinações: T30C30 (tratamento com vírus tempo 30 min contra controle tempo 30 min), T60C30 (tratamento com vírus tempo 60 min contra controle tempo 30 min) e T16C16 (tratamento com vírus tempo 16 horas contra controle tempo 16 horas), utilizando-se como genes de referência ACT27, Beta-tubulina e SKIP.....99

LISTA DE ABREVIATURAS

AFLP	Polimorfismo de Tamanho de Fragmento Amplificado, <i>Amplified Fragment Length Polymorphism</i>
ATP	Trifosfato de Adenosina, <i>Adenosine triphosphate</i>
Avr	Avirulência, <i>Avirulence</i>
BICMV	Blackeye Cowpea Mosaic Virus
BLAST	Ferramenta Básica de Busca por Alinhamento Local, <i>Basic Local Alignment Search Tool</i>
BTCI	<i>Black eyed-pea Trypsin Chymotrypsin Inhibitor</i>
CABMV	<i>Cowpea Aphid-Borne Mosaic Virus</i>
CC	Cauda Espiralada, <i>Coiled coil</i>
cDNA	DNA Complementar, <i>Complementary DNA</i>
CPGMV	<i>Cowpea Golden Mosaic Virus</i>
CPSMV	Vírus do Mosaico Severo do Feijão-caupi, <i>Cowpea Severe Mosaic Virus</i>
DNA	Ácido Desoxirribonucleico, <i>Deoxyribonucleic Acid</i>
EST	Etiquetas de Sequências Expressas, <i>Expressed Sequence Tag</i>
GeneBank	Banco de Genes, <i>GeneBank</i>
GENOSOJA	Consórcio Nacional para Estudos para Genoma da Soja
GTP	Trifosfato de Guanosina, <i>Guanosine Triphosphate</i>
ha	Hectare
HR	Reação de Hipersensibilidade, <i>Hypersensitive Response</i>
LRR	Repetições Ricas em Leucina, <i>Leucine Rich Repeats</i>
MEGA	Análises Genéticas da Evolução Molecular, <i>Molecular Evolutionary Genetic Analysis</i>
Mb	Mega Pares de Bases, <i>Mega base pairs</i>

mRNA	RNA mensageiro, <i>Messenger RNA</i>
NBS	Sítio de Ligação ao Nucleotídeo, <i>Nucleotide Binding Site</i>
N	Nitrogênio, Nitrogen
NCBI	Centro Nacional para Informação Biotecnológica, <i>National Center for Biotechnology Information</i>
NordEST	Rede Genômica do Nordeste (organizadora do banco de dados do feijão-caupi)
pb	Pares de bases, <i>Base pairs</i>
PCR	Reação em Cadeia da Polimerase, <i>Polymerase Chain Reaction</i>
PRGdb	Banco de Dados de Genes de Resistência em Plantas, <i>Plant Resistance Genes Database</i>
R	Resistência, <i>Resistance</i>
RAPD	DNA Polimórfico Amplificado Randomicamente, <i>Random Amplified Polymorphic DNA</i>
Renorbio	Programa Rede Nordeste de Biotecnologia
RFLP	Polimorfismos em Tamanho de Fragmentos de Restrição, <i>Restriction Fragment Length Polymorphisms</i>
RGA	Análogos de Genes de Resistência, <i>Resistance Genes Analogs</i>
RLK	Receptor Tipo Quinase, <i>Receptor-Like Kinase</i>
RLP	Proteína Tipo Receptora, <i>Receptor-Like Protein</i>
RNA	Ácido Ribonucleico, <i>Ribonucleic Acid</i>
RNAseq	Sequenciamento do RNA; <i>RNA Sequencing</i>
RSTK	Serina/treonina quinase do tipo receptor, <i>Receptor-like serine/threonine kinase</i>
RT-qPCR	PCR quantitativa em Tempo Real, <i>Quantitative Real Time PCR</i>
SAM	Seleção Assistida por Marcadores, <i>Marker-Assisted Selection</i>

SAR	Resistência Sistêmica Adquirida, <i>Systemic Acquired Resistance</i>
SCAR	Regiões Amplificadas a partir de Sequências Caracterizadas, <i>Sequence-Characterized Amplified Region</i>
SSR	Sequências Simples Repetidas, <i>Simple Sequence Repeats</i>
SuperSAGE	Super Análises em Série da Expressão Gênica; <i>Super Serial Analysis of Gene Expression</i>
t	Tonelada, <i>Ton</i>
TIR	Receptor Toll-Interleucina, <i>Toll Interleukine Receptor</i>
VNTR	Repetições em Tandem de Número Variável, <i>Variable Number of Tandem Repeats</i>

SUMÁRIO

1. Introdução	17
2. Revisão da Literatura	20
2.1 Mecanismo de defesa das plantas e os genes de resistência	20
2.2 Família gênica NBS-LRR em plantas.....	22
2.2.1 Características estruturais e funcionais	22
2.2.2 Organização no genoma e evolução dos genes NBS-LRR	24
2.3 Bioinformática na análise de genes de resistência em plantas	26
2.4 Análogos de genes de resistência (RGAs) em plantas	27
2.5 Tecnologias aplicadas a expressão gênica.....	28
2.6 Leguminosas e sua importância econômica	30
2.7 Feijão-caupi (<i>Vigna unguiculata</i>).....	32
2.8 Estudos genômicos em feijão-caupi (Rede NordEST) e outras leguminosas	34
2.9 Ataque de patógenos em feijão-caupi e seu impacto na agricultura	36
3. Objetivos	39
3.1 Geral	39
3.2 Específicos.....	39
4. CAPÍTULO I	40
Plant Elite Squad: first defense line and resistance genes – identification, diversity and functional roles	40

5. CAPÍTULO II	74
Identificação e caracterização de genes NBS-LRR em <i>Vigna unguiculata</i> (L.) Walp.....	74
6. Discussão geral.....	110
7. Conclusões gerais.....	114
8. Referências Bibliográficas	116
9. Anexos	122
Anexo 1: Trinta e cinco genes NBS-LRR previamente caracterizados em plantas, os quais foram alinhados contra o banco de dados do NordEST como sondas (<i>seed sequence</i>) para desenvolvimento desse estudo	122
Anexo 2: Resultados obtidos a partir do BLASTx contra o NCBI, considerando um e-value de e^{-10} . Número de acesso, organismo, descrição, tamanho dos candidatos em nucleótidos (nt) e aminoácidos (aa), bem como os seus valores correspondentes <i>e-value</i> e <i>score</i>	123
Anexo 3: Instruções para autores: Current Protein & Peptide Science	125
Anexo 4: Instruções para autores: Genetics and Molecular Biology	138
10. Curriculum vitae (Lattes)	143

1. Introdução

Ao longo do processo evolutivo, as plantas desenvolveram mecanismos de defesa, pré-existente ou induzidos, os quais são eficientes contra um amplo espectro de patógenos frequentemente encontrados no meio ambiente. O conjunto de mecanismos de defesa inclui barreiras físicas, tais como parede celular, lignina, como também envolve a ativação de metabólitos secundários e proteínas antimicrobianas que, juntos, são capazes de impedir a colonização do agente patogênico. Contudo, alguns microrganismos ainda conseguem ultrapassar estas barreiras, podendo causar danos irreversíveis ao crescimento, desenvolvimento e produtividade dos vegetais.

Após a interação planta-patógeno, a planta reconhece o patógeno através de sinais moleculares que ocorrem entre os produtos de genes de resistência (*R – Resistance*) e as partículas efetoras dos patógenos, chamada de avirulência (*Avr - avirulence*). Em seguida, a reação de hipersensibilidade (*HR - Hypersensitive response*) é ativada e, conseqüentemente, o desenvolvimento da doença é interrompido, por meio de apoptose local. Essa interação desencadeia um mecanismo rápido de defesa via cascata de sinalização, resultando na chamada Resistência Sistêmica Adquirida (*SAR - Systemic Acquired Resistance*), a qual é responsável por proteger a planta contra uma ampla gama de patógenos.

Os genes *R* são classificados em cinco famílias, de acordo com as estruturas e combinações de seus domínios conservados. Dentre essas, a família gênica NBS-LRR é a maior e mais diversificada, sendo todos os seus integrantes associados, de alguma forma, à defesa vegetal, conferindo resistência a bactérias, vírus, fungos, nematóides e afídeos. O domínio NBS (Sitio de Ligação ao

Nucleotídeo; *Nucleotide Binding Site*) é altamente conservado e possivelmente atua na transdução de sinais, enquanto que o domínio LRR (Repetições Ricas em Leucina; *Leucine Rich Repeats*) apresenta uma maior variação, estando envolvido no reconhecimento do patógeno, mediante o ataque do mesmo. Estas proteínas NBS-LRR são divididas em duas subfamílias com base na sua região N-terminal: TIR (Receptor da Toll-Interleucina; *Toll Interleukine Receptor*) e não-TIR. Almejando auxiliar a identificação e caracterização desses genes, a bioinformática torna-se indispensável, uma vez que suas ferramentas englobam conceitos e aplicações de diversas áreas de conhecimento (incluindo biologia, química, estatística e ciências da computação). Tal multidisciplinaridade permite a realização de estudos cada vez mais interligados com as informações acerca de estrutura, localização, regulação e expressão dos diversos genes e de seus produtos. Neste cenário, a genômica comparativa também permite a caracterização de um gene, em determinado organismo e auxilia na identificação de ortólogos em outros organismos de interesse econômico ou não. Diante dessa situação, a identificação de genes *R* representa grande relevância para o melhoramento genético de plantas, as quais estão expostas ao ataque de vários agentes patogênicos.

Devido à importância das leguminosas ao redor do mundo, tanto na agricultura, como no meio ambiente, na nutrição humana e animal, vários projetos buscam integrar estratégias de tecnologia genômica junto a programas de melhoramento, visando ao desenvolvimento de novas cultivares. O desenvolvimento de genótipos de feijão-caupi [*Vigna unguiculata* (L.) Walp.]] resistentes às principais doenças, em especial as infecções causadas por vírus, como o mosaico severo, é uma necessidade para agricultura do Brasil, visto que esse patógeno é capaz de comprometer cerca de 80% da produção dessa cultivar.

Essa cultura representa uma excelente fonte de proteínas, carboidratos, cálcio, ferro e vitaminas na alimentação humana, sendo um componente alimentar básico para as populações rurais e urbanas das regiões Norte e Nordeste do Brasil. Neste contexto, o projeto Brasileiro do Transcriptoma do Feijão-Caupi (rede NordEST) gerou um número significativo de transcritos, os quais começam a ser aplicados no sentido de auxiliar os programas de melhoramento.

Pesquisas voltadas à análise de recursos genéticos e ao melhoramento vegetal estão entre as atividades mais inovadoras no Brasil. Com o auxílio de ferramentas de bioinformática e da biologia molecular é possível identificar os mecanismos genéticos de resistência da planta e assim construir um painel de genes que participam na interação planta-patógeno. Tais pesquisas representam um importante instrumento para redução do tempo de programas de melhoramento convencional que visam à introdução de um ou mais genes de resistência em cultivares com elevado potencial produtivo.

2. Revisão da Literatura

2.1 Mecanismo de defesa das plantas e os genes de resistência

O mecanismo pré-determinado ou defesa pré-formada é considerado o principal mecanismo de defesa no qual as plantas sintetizam peptídeos, proteínas e metabólitos secundários, que irão consequentemente restringir a infecção por patógenos. Muitas barreiras físicas pré-formadas, como paredes celulares, tricomas e cutículas mais espessas, estão incluídas nessa estratégia de defesa (Freeman e Beattie, 2008).

Por outro lado, após o patógeno ter penetrado as estruturas de defesa pré-formadas, as plantas também possuem a capacidade de ativar uma cascata altamente eficiente de respostas de defesa do tipo induzível. Em um estudo realizado com a ferrugem do linho, pelo cientista H. H. Flor, foi possível determinar uma interação ocorrente entre planta e patógeno. Essa hipótese ficou conhecida como interação gene-a-gene, baseada na seguinte afirmação: "*para cada gene que condiciona uma reação de resistência no hospedeiro existe um gene complementar no patógeno que condiciona a avirulência*" (Flor, 1971). Após essa interação, nos primeiros 15 minutos, a planta desencadeia uma resposta que resulta na indução de fluxos iônicos, produção de compostos reativos de oxigênio e óxido nítrico, bem como uma reprogramação da expressão gênica por meio da ação de fatores de transcrição e quinases. Em seguida, ocorre a síntese de ácido salicílico, etileno e compostos antimicrobianos (Benko-Iseppon *et al.*, 2010b; Sanabria *et al.*, 2010), resultando na morte geneticamente programada das células infectadas, evento esse conhecido como reação de Hipersensibilidade (Stakman, 1915).

Essas respostas locais podem, conseqüentemente, desencadear uma resposta duradoura sistêmica - resistência sistêmica adquirida – SAR, a qual vai preparar a planta para responder a uma diversidade de patógenos (Métraux, 2001, Vallad *et al.*, 2004). Logo, a resistência à doença será mediada pelo reconhecimento da molécula elicitora do gene *Avr* que irá ativar uma via de transdução de sinais envolvendo os genes de resistência (*R*). Caso haja alterações ou perdas de genes *R* complementar ao *Avr*, a doença será desencadeada (Ellis e Jones, 1998). Esse tipo de defesa é de caráter genético, altamente específico e só será ativado se a planta detectar alguma partícula estranha.

Mediante a interação entre a planta e o patógeno os genes de resistência são ativados, ou não, resultando em diversas respostas de defesa. Apesar da caracterização e da clonagem de vários genes de resistência, o mecanismo pelo qual eles ativam as respostas de defesa, ainda não está completamente elucidado. No entanto, dados obtidos tanto através da análise genética da interação hospedeiro-patógeno, como através da análise molecular, suportam o modelo no qual os produtos dos genes *R* atuam como receptores para os produtos diretos ou indiretos do gene *Avr* do patógeno (Bent e Mackey, 2007). Sabe-se que uma única interação entre planta e patógeno, seja ela compatível ou não, é capaz de recrutar ou silenciar centenas de genes envolvidos na defesa, onde muitos deles já são conhecidos, enquanto outros ainda não foram descritos (Benko-Iseppon *et al.*, 2010b).

Os genes de resistência, em contraste com a grande diversidade dos produtos dos genes *Avr* (Ellis e Jones, 1998), apresentam regiões extremamente conservadas e podem ser agrupados em famílias de acordo com a sua estrutura e domínios observados. Atualmente, cinco famílias de genes *R* são descritas a

dependem de sua estrutura e da presença de domínios conservados: (1) NBS-LRR - Sítio de ligação ao Nucleotídeo, *Nucleotide Binding Site* - Repetições Ricas em Leucina, *Leucine Rich Repeats*; (2) RLP – Proteína Tipo Receptora, *Receptor like protein*; (3) RLK - Receptor tipo kinase, *Receptor Like Kinase*; (4) RSTK - Receptor de serina/treonina quinase, *Receptor serine/threonine Kinase* e (5) Domínio *coiled coil* ancorado à membrana (Van-Ooijen *et al.*, 2007). Contudo, essa classificação pode apresentar variações, a depender dos autores, com base nos domínios conservados, onde esses genes podem ser agrupados em cinco ou seis classes distintas. Jones (2001) considera as duas subfamílias da primeira família como famílias distintas, enumerando então, seis famílias de genes de resistência. Estudos indicam que as regiões inter-domínios desses genes estão entre os genes de plantas que mais variam, tanto entre populações como também dentro das mesmas, o que não prejudica sua classificação nas famílias em que foram enquadrados (Moffett, 2009), permitindo o reconhecimento de um amplo espectro de patógenos.

2.2 Família gênica NBS-LRR em plantas

2.2.1 Características estruturais e funcionais

Dentre as famílias de genes de resistência em plantas, a NBS-LRR é a maior e mais diversificada, estando todos os seus integrantes associados de alguma forma à defesa, conferindo resistência a bactérias, vírus, fungos, nematoides e afídeos (McHale *et al.*, 2006). As proteínas relacionadas a esses grupos se caracterizam por apresentar um sítio de ligação ao nucleotídeo (NBS), repetições

ricas em leucina (LRR) e domínios amino-terminal e carboxi-terminal variáveis (Figura 1) (McHale *et al.*, 2006).

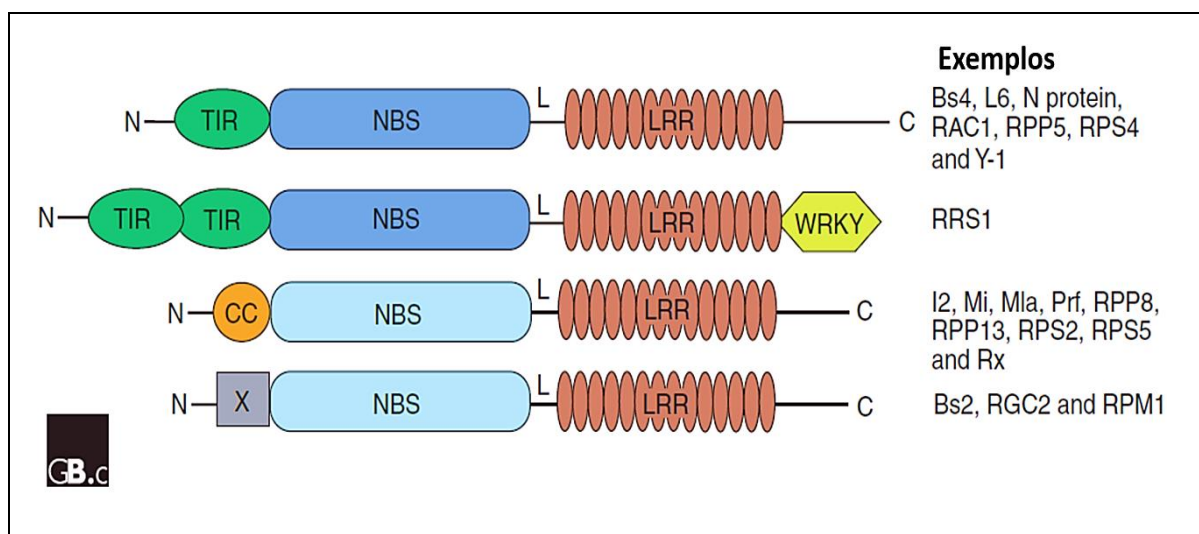


Figura 1: Representação esquemática dos domínios conservados presentes na família gênica NBS-LRR. N - Amino-terminal, TIR - Receptor Toll- Interleucina, CC - Cauda Espiralada, NBS - Sitio de Ligação ao Nucleotideo, LRR - Repetições Ricas em Leucina, C - Carboxi-terminal, WRKY - Fator de Transcrição (Fonte: McHale *et al.*, 2006).

O domínio NBS, também conhecido como *P-loops* ou NB-ARC, é altamente conservado entre diferentes espécies vegetais, sejam elas relacionadas ou não (Tameling *et al.*, 2010). Esse domínio atua, possivelmente, na transdução de sinais, sendo o responsável por atividade de ligação ao ATP (Adenosina Trifosfato, *Adenosine Triphosphate*) ou GTP (Trifosfato de guanosina, *Guanosine triphosphate*), assim como subunidades da ATP sintetase e fatores de alongação do ribossomo (Tameling *et al.*, 2010), entre outros. Sugere-se que os genes de resistência que apresentam o domínio NBS estejam envolvidos na ativação de proteínas quinases ou proteínas transportadoras (Hammond-Kosack e Jones, 1997).

Mutações no domínio NBS que venham a causar uma modificação na sua estrutura, podem resultar em perda de compatibilidade com o gene *Avr*

correspondente, culminando em uma resposta hipersensível frente à ação do patógeno, assim ressaltando-se a importância da função desse domínio nas vias de sinalizações de defesa da planta (Tsuda e Katagiri, 2010).

O domínio LRR é altamente variável e está envolvido no reconhecimento específico do patógeno pelo hospedeiro mediante um processo de infecção. Acredita-se que a função primária do LRR seja promover um ambiente versátil para a mediação da interação entre proteínas, podendo atuar extracelularmente como receptor da molécula produzida pelo patógeno ou ainda intracelularmente, em um passo anterior à via de transdução de sinais no âmbito do processo de resistência (Banerjee *et al.*, 2001). Além disso, esse domínio está envolvido em sítios de união ao peptídeo ligante e a sítios de interação proteína-carboidrato, no que tange à especificidade da resistência (Kobe e Kajava, 2001).

O domínio LRR é formado por motivos do aminoácido leucina, com repetições em número de 20 a 30, em intervalos regulares, que se agrupam em tandem (Sun e Wang, 2011). Esses resíduos de leucina conservados proporcionam uma estrutura terciária semelhante a uma mola, onde as repetições de leucinas são responsáveis pela formação de folhas-beta, resultando em um formato côncavo (Enkhbayar *et al.*, 2003).

2.2.2 Organização no genoma e evolução dos genes NBS-LRR

Diversos estudos de genômica comparativa têm sugerido que os genomas vegetais são capazes de codificar várias centenas de genes NBS-LRR, existindo uma grande diversidade no número e na distribuição das subclasses desses genes (Marone *et al.*, 2013). Os *loci* correspondentes a esses genes podem ser encontrados como genes isolados (*singletons*) ou formando *clusters*, apresentando

uma distribuição irregular nos diferentes cromossomos (Holub, 2001, Marone *et al.*, 2013). Um número grande de sequências que codificam genes NBS têm sido isoladas a partir de várias espécies de plantas por meio de análises genômicas. Em mamão (*Carica papaya* L.) e em pepino (*Cucumis sativus* L.) foram identificados cerca de 50 genes NBS-LRR (Porter *et al.*, 2009, Wan *et al.*, 2013), enquanto em *Arabidopsis thaliana* (L.) foram identificados 159 (Guo *et al.*, 2011). Estes dados mudam constantemente, à medida que rapidamente novas sequências genômicas são produzidas. Porém, tais dados proporcionam uma ideia da dimensão e organização desses genes no genoma de plantas (Marone *et al.*, 2013).

As proteínas NBS-LRR são divididas em duas subfamílias com base na sua região N-terminal: TIR (Receptor da Toll-Interleucina, *Toll Interleukine Receptor*) e não-TIR, conhecida como TNL e CNL, respectivamente. As CNL estão subdivididas em dois grupos, dependendo da presença de um curto ou um prolongado N-terminal, onde estão localizados os genes que apresentam o domínio Cauda Espiralada (CC; *Coiled-coil*) (McHale *et al.*, 2006). Essas duas subfamílias podem ser facilmente identificadas observando-se a estrutura conservada do motivo *Kinase-2* presente no domínio NBS. A classe TIR é caracterizada pela presença do resíduo ácido aspártico (D) na última posição do motivo (Palomino *et al.*, 2006), enquanto que na subfamília não-TIR é reportada a presença do triptofano (W) nessa localização (Meyers *et al.*, 1999). Os membros da subfamília não-TIR estão presentes em monocotiledôneas e em dicotiledôneas, ao contrário dos componentes que apresentam o domínio TIR, os quais estão ausentes nas monocotiledôneas analisadas até o momento (Goff *et al.*, 2002; Tarr e Alexander, 2009).

Acredita-se que a formação e a especificidade de novos genes de resistência sejam geradas a partir de eventos de duplicação, *crossing over* desigual, recombinação ectópica, conversão gênica, seleção natural e divergência de sequências nas regiões intergênicas (Young, 2000; McDowell e Simon, 2006).

Proteínas NBS-LRR que apresentam um alto grau de similaridade são capazes de reconhecer patógenos completamente distintos, ou seja, não há características peculiares entre esses mediadores de resistência que faça distinção entre as diferentes classes de agentes patogênicos (Collier *et al.*, 2009).

2.3 Bioinformática na análise de genes de resistência em plantas

Na segunda metade da década de 1980 surgiu a bioinformática, face ao aumento expressivo na quantidade de sequências geradas a partir dos primeiros projetos de sequenciamento em larga escala, a qual tem por objetivo responder a questionamentos biológicos através de análises com ferramentas computacionais (Baxevanis, 2001).

A diversidade de ferramentas computacionais disponíveis vem facilitando a interpretação e análise de dados genômicos e pós-genômicos, permitindo a associação das informações relacionadas à transcriptômica, proteômica, metabolômica e outras “ômicas”. Com isso, foram criados diversos bancos de dados, a fim de armazenar e disponibilizar os dados para a comunidade científica (Edwards e Batley, 2004). Dentre estes, o GenBank criado pelo Centro Nacional para Informação Biotecnológica (NCBI; *National Center for Biotechnology Information* - <http://www.ncbi.nlm.nih.gov/>), disponibiliza gratuitamente sequências e ferramentas para a sua análise. Adicionalmente, o Banco de Dados de Genes de Resistência em Plantas (PRGdb; *Plant Resistance Genes Database* -

<http://www.prgdb.org>), criado em 2009, foi o primeiro banco de genes *R* identificado por ferramentas de Bioinformática. O PRGBdb engloba mais de 15.000 sequências entre genes *R* caracterizados e preditos pertencentes a 192 espécies distintas de plantas (Sanseverino *et al.*, 2010).

Além dos bancos de dados, diversas ferramentas e programas são rotineiramente utilizados no processo de identificação e caracterização de genes *R*, tais como a Ferramenta de Busca por Alinhamento Local (BLAST; *Basic Local Alignment Search Tool*) (Altschul *et al.*, 1990), o Clustal Omega (Sievers *et al.*, 2014) utilizado para realização de alinhamentos múltiplos, o TreeView (Page, 1996) programa para visualização de imagens, o Programa para Análise Genética e Evolutiva (MEGA; *Molecular Evolutionary Genetics Analysis*) (Tamura *et al.*, 2007) e o CLUSTER (Eisen *et al.*, 1998), o qual permite geração *in silico* de interfaces gráficas para a visualização de diferenças na expressão gênica. Assim, a bioinformática tem ocupado papel significativo na integração de vários dados gerados por estas diferentes abordagens (Edwards e Batley, 2004). A utilização dessas ferramentas favorece a identificação e caracterização de genes de resistência em diversas plantas, tais como em *A. thaliana* (Mondragon-Palomino *et al.*, 2002; Gou *et al.*, 2009), solanáceas como o tomate e a batata (Van-Ooijen *et al.*, 2007), arroz (*Oryza sativa* L.) (Shang *et al.*, 2009) e leguminosas (Shao *et al.*, 2014).

2.4 Análogos de genes de resistência (RGAs) em plantas

Os genes de resistência podem ser identificados por meio de investigação de RGAs (*Resistance Gene Analog*) (Collins *et al.*, 2001), os quais podem ser detectados por similaridade de sequências, através de regiões altamente

conservadas entre diferentes genes *R*, conhecidas como domínios conservados, tais como TIR, NBS e LRR. Essas regiões podem ser observadas tanto em espécies próximas como em espécies remotamente relacionadas. Young (2000) comprovou em estudos, que a família NBS-LRR está presente em quase todas as plantas.

Tendo o conhecimento da estrutura dos genes de resistência é possível desenhar *primers* para os domínios conservados e desenvolver marcadores do tipo RGA através da técnica de PCR (Reação em Cadeia da Polimerase, *Polymerase Chain Reaction*). Frequentemente a identificação de fenótipos de resistência em plantas é de difícil detecção, sendo assim, os marcadores RGAs podem constituir-se como uma ferramenta muito útil para a seleção de genótipos de interesse, construção de mapas genéticos e identificação de genes de resistência em vegetais (Hammond-Kosack e Jones, 1997; Tullu *et al.*, 2006).

Com base na técnica de PCR em busca por domínios conservados, já foram descritos vários genes de resistência em diferentes espécies, tais como uva (*Vitis vinifera* L., Donald *et al.*, 2002), trigo (*Triticum vulgare* Vill., Lacock *et al.*, 2003), banana (*Musa acuminata* Colla, Miller *et al.*, 2008) e arroz (*O. sativa*, Ren *et al.*, 2013).

2.5 Tecnologias aplicadas a expressão gênica

A análise da expressão gênica diferencial compreende uma fonte importante para a biologia molecular, contribuindo com a caracterização de mecanismos de resistência e permitindo a identificação de genes-chave em redes de interação, através da comparação da expressão gênica, mediante

experimentos onde os organismos são submetidos a algum tipo de estresse (Guimarães *et al.*, 2005).

A fim de se compreender a função de um gene, é fundamental a identificação do local (onde) e do tempo (quando) onde o mesmo é expresso. Baseando-se nisso, foram desenvolvidas várias tecnologias baseadas no sequenciamento e contagem de transcritos, tais como EST (Etiquetas de Sequências Expressas, *Expressed Sequence Tag*, Adams *et al.*, 1991), SuperSAGE (Super Análises em Série da Expressão Gênica; *Super Serial Analysis of Gene Expression*, Matsumura *et al.*, 2004;), RNAseq (Sequenciamento do RNA; *RNA Sequencing*, Morin *et al.*, 2008).

As ESTs são sequências geradas a partir da construção de bibliotecas de cDNA (DNA complementar) que corresponde a um mRNA (RNA mensageiro) (Adams *et al.*, 1991), em geral com tamanho entre 200 a 800 pb. As sequências obtidas por EST estão mais sujeitas a erros, tais como deleções, substituições e inserções de bases, quando comparadas a outras técnicas de leitura múltipla, devido ao fato de serem geralmente sequenciadas apenas uma única vez (Wolfsberg e Landsman, 1997).

Por sua vez, SuperSAGE é uma modificação do procedimento de SAGE convencional, e destaca-se por sua eficiência em gerar perfis de alta definição da transcrição. Essa tecnologia permite análises qualitativas e quantitativas de padrões de transcrição de um ou mais genomas de eucariotos a partir da mudança da frequência de, teoricamente, todo mRNA que possua a cauda Poli Adenosina (PoliA) (Matsumura *et al.*, 2004).

Estudos demonstram que a técnica de RNA-Seq tem inovado a maneira pela qual o transcriptoma vem sendo analisado (Wang *et al.*, 2009). Essa técnica

é baseada na criação de uma população de moléculas de RNA (total ou fracionada) que é convertida em uma biblioteca de fragmentos de cDNA, onde em seguida as *reads* são geradas e alinhadas contra o genoma de referência (Morin *et al.*, 2008). Recentemente a técnica de RNAseq tem sido muito utilizada em trabalhos com vegetais, tais como soja (*G. max*, Severin *et al.*, 2010), *A. thaliana* (Loraine *et al.*, 2013).

2.6 Leguminosas e sua importância econômica

As espécies que compõem o grupo das leguminosas pertencem à família Fabaceae, também conhecida como Leguminosae. Apresentam uma ampla distribuição geográfica, estando presentes em todos os biomas brasileiros e se caracterizam por conter os frutos dentro de vagens (Lewis *et al.*, 2005). Essa família é composta por 430 gêneros e 12.600 espécies e encontra-se dividida em três subfamílias: *Caesalpinoideae*, *Mimosoideae* e *Faboideae*, sendo essa última a de maior interesse econômico. Dentre as espécies dessa subfamília encontram-se o feijão, o grão-de-bico, a fava, a soja e a lentilha, as quais são importantes devido ao valor nutricional dos seus grãos secos (Barroso *et al.*, 2007).

Esse grupo vegetal demonstra relevante importância econômica principalmente pela geração de empregos e por fornecer produtos para alimentação humana e animal. Os diferentes usos dos seus derivados estendem-se às aplicações industriais, tais como, geração de biomassa (carvão, celulose, madeira, arborização), além de produtos farmacêuticos devido aos fitoquímicos com potencial antioxidante e antimicrobiano (Ponnusha *et al.*, 2011; Patel, 2014). Além disso, algumas espécies que apresentam porte arbóreo ou arbustivo são utilizadas na recuperação de solos degradados e na melhoria daqueles que naturalmente têm

baixa fertilidade. Essa prática tem sido bem usual nas regiões tropicais, em áreas destinadas à produção de alimento básico (Locatelli *et al.*, 1991).

O feijoeiro (*P. vulgaris*) apresenta uma ampla adaptação a diferentes ambientes, estando presente em muitos países tropicais, subtropicais e temperados. Esse feijoeiro é a terceira leguminosa produtora de grãos mais comumente cultivada no Brasil (Blair *et al.*, 2009), estando atrás apenas da soja (*G. max*) e do amendoim (*Arachis hypogaea* L.). O feijoeiro é altamente nutritivo com quase o dobro do nível protéico dos cereais, gorduras inferiores à da soja e do amendoim e uma maior quantidade de lisina, fósforo, ferro, zinco, magnésio, cobre e cálcio, comparativamente à quantidade presente nos cereais (Blair *et al.*, 2009).

A soja (*G. max*) apresenta importância mundial, da qual se obtém proveito em diversos segmentos. É amplamente utilizada na alimentação humana, por conter macro e micronutrientes essenciais, além de responder por mais da metade da produção de óleo vegetal e por representar uma fonte potencial para a produção de biodiesel. A soja contribui anualmente para a economia global, com cerca de 46,8 bilhões de dólares, quadro onde o Brasil figura como o segundo maior produtor, perdendo apenas para o EUA, com uma produção superior a 68 milhões de toneladas, na safra de 2010/2011 (MAPA, 2014).

O gênero *Vigna*, por sua vez, representa interesse especial devido aos seus recursos genéticos, o qual se fundamenta principalmente no seu potencial econômico e ecológico por meio da exploração da variabilidade genética de espécies cultivadas e, ainda, pela possibilidade de transferência de características genéticas desejáveis das espécies silvestres às cultivadas, aproveitando eventuais compatibilidades reprodutivas. Dentre as espécies cultivadas do gênero *Vigna*, merecem menção o feijão-caupi (*V. unguiculata*), feijão-mungo [*V. radiata* (L.)

Wilczek], feijão-rajado [*V. mungo* (L.) Hepper], feijão-adzuki [*V. angularis* (Willd.) Ohwi & Ohashi], feijão-arroz [*V. umbellata* (Thunb.) Ohwi & Ohashi] e *V. aconitifolia* (Jacq.) Marechal, os quais se destacam pela sua capacidade de se adaptar a condições climáticas estressantes, crescendo em solos pobres e sem necessidade de suplementação de nitrogênio (Fery, 2002). Essa autonomia no suprimento de nitrogênio se dá devido à capacidade que as leguminosas apresentam em formar simbiose com bactérias conhecidas como rizóbios. Essa interação permite que o nitrogênio presente no ar atmosférico seja convertido e transferido para as plantas em níveis assimiláveis, devido à presença dessas bactérias nos nódulos das raízes (Broughton *et al.*, 2006).

2.7 Feijão-caupi (*Vigna unguiculata*)

O feijão-caupi provavelmente se originou em uma região localizada na parte oeste e central da África (Brito, 2008), pertence à classe Dicotyledoneae, ordem Fabales, família Fabaceae, subfamília Faboideae, tribo Phaseoleae, subtribo Phaseolinae, gênero *Vigna*, seção *Catjang*, espécie *Vigna unguiculata* ssp. *unguiculata* (Padulosi e Ng, 1997). Esse feijão é conhecido por várias denominações em distintas regiões do Brasil. No Nordeste ele recebe prioritariamente os seguintes nomes populares: feijão-de-corda, feijão macassar ou macáçar, feijão-de-vagem, feijão-de-vara, feijão gurutuba, feijão-Vigna. No Norte é chamado de feijão-de-praia, feijão-colônia, enquanto no Sul é denominado feijão miúdo e no Rio de Janeiro de feijão-fradinho (Araújo *et al.*, 1984).

O feijão-caupi apresenta importância, tanto na geração de alimento como na geração de emprego e renda. Trata-se de um alimento rico em proteínas, apresentando todos os aminoácidos essenciais, carboidratos, minerais e grande

quantidade de fibras dietéticas (Freire-Filho *et al.*, 2011). O feijão-caupi constitui um componente alimentar básico nas populações rurais e urbanas das regiões Norte e Nordeste do Brasil. Recentemente o seu consumo está se expandindo de forma mais intensa para as regiões Centro-Oeste e Sudeste do Brasil (Freire-Filho *et al.*, 2011).

O Brasil merece destaque, pois se encontra em terceiro lugar no *ranking* mundial como maior produtor e consumidor de feijão-caupi, com 1,5 milhões de hectares cultivados e uma produtividade de 44,2% (Singh *et al.*, 2002, Freire-Filho *et al.*, 2011). A região Nordeste é responsável por 1.205.839 ha (94,4%) da área cultivada e 412.458 t (87,7%) da produção total de feijão-caupi (Damasceno-Silva, 2007). A cultura do feijão-caupi tem gerado, em média, 1.113.109 empregos por ano, produzindo suprimento alimentar para 28.205.327 milhões de pessoas, apresentando uma produção anual no valor de R\$ 684.825.333 (Freire-Filho *et al.*, 2011).

A forma de consumo dessa cultura pode ser através da vagem verde, grão verde e grão seco, além de outras formas de preparo, como o acarajé (Freire-Filho *et al.*, 2005), produtos de panificação (biscoitos, rocambole), através da farinha de feijão-caupi processada (Frota *et al.*, 2009). Essa leguminosa ainda apresenta grande importância na alimentação de animais, sendo utilizada na forma de forragem verde, feno e ensilagem (Freire-Filho *et al.*, 2005). Estudos indicam a importância desse feijão para a saúde pública, através de análises desenvolvidas com a molécula BTCl (*Black eyed-pea Trypsin Chymotrypsin Inhibitor*), a qual tem a capacidade de matar células cancerígenas humanas sem agredir células saudáveis (Bolgue, 2011).

O feijão-caupi se apresenta como uma excelente cultura para as regiões semiáridas dos trópicos, por apresentar considerável tolerância à seca e ao calor, proporcionar uma excelente cobertura vegetal, desta forma, inibindo o crescimento de ervas daninhas. Também ajuda a evitar a erosão e contribui para a melhoria da fertilidade de solos em áreas marginais, fornecendo matéria e resíduos orgânicos que se incorporam ao substrato após a colheita, melhorando a estrutura edáfica, aumentando a disponibilidade de fósforo e contribuindo para a fixação de nitrogênio atmosférico (N₂) (Valenzuela e Smith, 2002).

A cultura do feijão-caupi explorada tradicionalmente por pequenos produtores em cultivo de sequeiro e com baixo nível tecnológico – antes vista apenas como cultura de subsistência – vem ocupando nas últimas décadas um novo espaço no setor de agronegócios, apresentando um aumento progressivo da participação de empresários de médio e grande porte. Esta nova realidade tem gerado uma demanda por cultivares com características que atendam às necessidades dos sistemas de produção tecnificados, entre elas, o alto potencial de rendimento, resistência a pragas e doenças e qualidade de grãos, além de características de porte e arquitetura adequadas ao maior adensamento e à mecanização da cultura, inclusive da colheita (Bezerra *et al.*, 2008).

2.8 Estudos genômicos em feijão-caupi (Rede NordEST) e outras leguminosas

Considerando a grande importância do feijão-caupi e a redução em sua produtividade devido a diversos estresses ambientais, em 2005 foi estabelecida a rede NordEST (Banco de dados de EST para o feijão-caupi da Rede Nordeste de Biotecnologia), a qual envolveu diversas técnicas de biotecnologia incluindo marcadores moleculares, a genômica, transcriptômica e transformação genética

com financiamento da Rede Nordestina de Biotecnologia (Renorbio), sob coordenação da Universidade Federal de Pernambuco. Na área da transcriptômica, foi realizado o sequenciamento de etiquetas de sequências expressas (ESTs - *Expressed Sequence Tags*), análise serial da expressão gênica (HT-SuperSAGE) e RNA-Seq para a identificação de genes relacionados à resistência ou tolerância a fatores bióticos e abióticos importantes para o melhoramento do feijão-caupi (Benko-Iseppon *et al.*, 2010a), gerando mais de 400 milhões de transcritos sob diversas condições.

Esse projeto integra 10 instituições e 12 Laboratórios, onde foram obtidas 30.000 sequências de EST e mais de 21 milhões de sequências de SuperSAGE relacionadas à expressão de genes de resistência a fatores bióticos e abióticos importantes para o melhoramento do feijão-caupi (Benko-Iseppon *et al.*, 2010a). O banco NordEST representa um dos mais elaborados para o feijão-caupi, com 12 bibliotecas de EST sob condições de estresse biótico (vírus do mosaico severo e potyvirus) e abiótico (salinidade), capacitando o acesso a informações transcricionais de uma diversidade de genes em vários momentos de resposta vegetal frente as situações contrastantes aplicadas experimentalmente (Benko-Iseppon, 2009). O mapeamento físico dos cromossomos está em desenvolvimento, incluindo amostras de microssatélites, rDNA e BACs, o que torna viável a rápida transferência de dados entre culturas visando o melhoramento (Benko-Iseppon *et al.*, 2010a).

Por sua vez, para o feijão comum (*P. vulgaris*) já existe um rascunho do genoma disponibilizado no banco de dados Phytozome (<http://www.phytozome.net/>), o qual auxilia os estudos em outras leguminosas e permite estudos funcionais mais precisos. Alguns estudos de expressão gênica sob

estresse biótico em leguminosas, baseados em RNA-Seq, demonstraram a importância dos genes *R* na resistência a patógeno (Severin *et al.*, 2010).

No Brasil, a soja é considerada o principal produto do agronegócio nacional, representando 28,7 % do montante total exportado por esse setor. Entretanto, apesar da sua relevante importância econômica, essa cultura tem sofrido perdas consideráveis devido a estresses ambientais (Valliyodan e Nguyen, 2008). Com isso, em 2007, o governo brasileiro iniciou o consórcio GENOSOJA (Consórcio Nacional para Estudos para Genoma da Soja) com o objetivo de aperfeiçoar o processo de produção deste grão, com ênfase para fatores que afetam a produção nacional, como a ocorrência de secas, pragas e patógenos (Benko-Iseppon *et al.*, 2012). A soja também já tem o seu genoma sequenciado e disponível no Soybase (<http://soybase.org/>) além de bases integrando dados de ômicas como o banco do projeto GENOSOJA (<http://bioinfo03.ibi.unicamp.br/soja/>).

Os projetos de análise genômica permitem um salto na resolução de informação do genoma das principais leguminosas e ajudam a potencializar as pesquisas nos principais programas de melhoramento genético no Brasil, visando ao desenvolvimento de genótipos resistentes contra as principais doenças.

2.9 Ataque de patógenos em feijão-caupi e seu impacto na agricultura

As doenças são facilmente identificadas devido a seus sintomas os quais são frequentemente associados com mudanças visíveis ocorrentes na morfologia do vegetal (IRRI, 2013).

Os principais patógenos causadores de doenças em leguminosas, são os fungos, bactérias, vírus e nematoides resultando em perdas na agricultura, danos na paisagem e diminuição na produção. Além da redução na estética e no valor

nutricional dos grãos, ocorre uma diminuição significativa nos lucros da comercialização desse produto (FAO, 2014).

Em meio ao amplo espectro de agentes patogênicos que comprometem o desenvolvimento do feijão-caupi, as viroses são indiscutivelmente agentes infecciosos de relevante importância no Brasil. Especialmente, o vírus do mosaico severo do feijão-caupi (CPSMV, *Cowpea Severe Mosaic Vírus*), da família Comoviridae; os mosaicos de potyvirus (CABMV, *Cowpea Aphid-Borne Mosaic Virus* e BICMV, *Blackeye Cowpea Mosaic Virus*), da família Potyviridae, e o vírus do mosaico dourado do feijão-caupi (CPGMV, *Cowpea Golden Mosaic Vírus*) da família Geminiviridae, os quais – conjuntamente com outros fatores – têm reduzido drasticamente a produção desse grão no Brasil e em outros países da América Latina (Gonçalves e Lima, 1982; Pio-Ribeiro e Assis-Filho, 1997; Barreto, 1999). O mosaico severo do feijão-caupi apresenta uma grande variabilidade biológica, sendo eficientemente disseminado na natureza por meio de coleópteros, a maioria deles dos gêneros *Cerotoma* e *Diabrotica* (Freire-Filho *et al.*, 2005) e podendo causar perdas de até 81% (Araújo *et al.*, 1984; Umaharan *et al.*, 1997).

Do ponto de vista das doenças causadas por fungos, existe, atualmente, um grande número de espécies fitopatogênicas identificadas, onde se destacam o *Colletotrichum truncatum* (Schw.) Andrus & Moore, causador da mancha-café; *Mycosphaerella cruenta* Lathan., *Cercospora cruenta* (Sacc.) Deigtom] e *Cercospora canescens* (Ellis e Martin), este último causador da mancha de cercospora ou cercosporiose (Rios, 1988; Athayde-Sobrinho *et al.*, 2000). No que diz respeito às infecções causadas por bactérias, dois grupos estão presentes com alto nível de disseminação: a mancha bacteriana [*Xanthomonas campestris* pv. *vignicola* (Burkholder) Dye (*X. axonopodis* pv. *vignicola*)] e a pústula-bacteriana

(*Xanthomonas* spp.). A pequena importância dada as doenças bacterianas podem ser devida à dificuldade dos produtores na identificação da infecção, quando essa ocorre em sementes e em plântulas (Rios, 1988).

Quanto às doenças provocadas por nematoides, são descritas 30 espécies patogênicas no Brasil, apresentando um elevado grau de patogenicidade. Merecem destaque os nematoides-das-galhas do gênero *Meloidogyne*, principalmente as espécies *M. incognita* e *M. javanica* que possuem extensa disseminação em toda área de cultivo do feijão-caupi (Freire-Filho *et al.*, 2005). De outra forma, o feijão-caupi apresenta sérios problemas quanto ao seu armazenamento, no que diz respeito à conservação dos grãos, pois eles são facilmente comprometidos por insetos pertencentes à ordem Coleoptera, com destaque para as espécies pertencentes à família Bruchidae. Nessa família se encontram o caruncho ou gorgulho [*Callosobruchus maculatus* (Fabr.)], o qual é considerada a principal praga do feijão-caupi armazenado na fase pós-colheita (Sales *et al.*, 2005).

A engenharia genética pode reduzir os altos gastos existentes com a aplicação de pesticidas, minimizando também os danos gerados ao meio ambiente devido ao seu uso, por meio da produção de plantas resistentes a patógenos importantes, uma vez que a resistência natural dispõe de várias vantagens óbvias em relação ao uso de produtos químicos ou a outro método qualquer utilizado no controle de pragas (Fermin-Munoz *et al.*, 2000). Essa tecnologia é altamente eficiente, benigna ao meio ambiente, não prejudica a paisagem e pode representar baixa despesa adicional aos produtores. Sendo assim, a resistência se torna a mais desejada estratégia de controle pelos melhoristas da atualidade (Shepherd *et al.*, 2011). Nesse contexto, a genômica funcional identifica genes candidatos relacionados com vias de resposta a defesa contra o ataque de patógenos, sendo

possível identificar proteínas que são induzidas durante o estresse e aquelas que são cruciais para uma resposta rápida e específica (Abby e Daubin 2007).

3. Objetivos

3.1 Geral

Identificar e caracterizar genes de resistência da classe NBS-LRR, no transcriptoma do feijão-caupi (*V. unguiculata*), validar a expressão dos genes candidatos identificados e desenvolver marcadores moleculares para aplicação em mapeamento genético.

3.2 Específicos

1. Identificar genes NBS-LRR em feijão-caupi e comparar com outras leguminosas, tais como soja e feijão comum, presentes em bancos de dados públicos e de acesso restrito;
2. Inferir sobre a estrutura desses genes, caracterizando e descrevendo suas regiões conservadas (domínios e motivos) nas sequências identificadas;
3. Avaliar padrões de diferenciação e evolução entre genes ortólogos e parálogos a partir da construção de árvores fenéticas e filogenéticas;
4. Desenvolver marcadores moleculares análogos aos genes de resistência (RGAs), com base em genes NBS-LRR selecionados, validando sua utilização em progênies de mapeamento;
5. Inferir sobre o perfil de expressão dos genes de interesse a partir de análises *in silico* e em laboratório via RT-qPCR, utilizando plantas de feijão-caupi submetidas à infecção por vírus.

4. CAPÍTULO I

Plant Elite Squad: first defense line and resistance genes – identification, diversity and functional roles

Ana Carolina Wanderley-Nogueira^{1*}, João Pacífico Bezerra-Neto¹, Ederson Akio Kido¹, Flávia Tadeu de Araújo¹, Lidiane Lindinalva Barbosa Amorim^{1,2}, Sergio Crovella¹, Ana Maria Benko- Iseppon^{1,2}

¹*Universidade Federal de Pernambuco, Genetics Department, Center of Biological Sciences, Av. Prof. Moraes Rego, 1235, 50.670-423, Recife, Pernambuco, Brazil.*

²*Federal Institute of Education, Science and Technology of Piaui, Campus Oeiras, R. Projetada, CEP 64500-000, Oeiras, PI, Brazil.*

³*Genetic Service, IRCCS Burlo Garofolo and Department of Developmental and Reproductive Sciences, University of Trieste -Via dell'Istria, 65/1- 34137 Trieste, Italy.*

**Author to whom correspondence should be addressed; E-Mail: carolwn@gmail.com; Av. Prof. Moraes Rego, 1235, CEP 50.670-420, Recife, PE, Brazil; E- Tel.: +55-81-2126-7816; Fax: +55-82-2126-8522.*

Short Title: **First defense line and resistance genes**

Abstract

Plants exhibit sensitive mechanisms to respond to environmental stresses, presenting some specific and non-specific reactions when attacked by pathogens, including organisms from different classes and complexity, as viroids, viruses, bacteria, fungi and nematodes. A crucial step to define the fate of the plant facing an invading pathogen is the activation of a compatible Resistance (*R*) gene, focus of the present review. Different aspects regarding *R*-genes and their products are discussed, including pathogen recognition mechanisms, signaling and effects on induced and constitutive defense processes, splicing and post transcriptional mechanisms involved. There are still countless challenges to the complete understanding of the mechanisms involving *R*-genes in plants, in particular those related to the interactions with other genes of the pathogen and of the host itself, their regulation, acting mechanisms at transcriptional and post-transcriptional levels, as well as the influence of other types of stress over their regulation. A magnification of knowledge is expected when considering the novel information from the omics and systems biology.

Keywords:

Avirulence, DAMPs, Hypersensitive response, MAMPs, PAMPs, *R*-gene, Signal transduction.

Introduction

Until recently, the term "immunity" was not well accepted because plants have neither a circulatory system nor antibody-like molecules related to an adaptive or acquired system as reported in mammals. Despite of several differences plants like animals are also subjected to pathogen attack leading to diseased condition and in extreme situations death. The causative agents of infectious diseases in plants comprise the same types of organisms that attack other organisms as well. These include fungi, bacteria, viruses and viroids. Out of these viroids are unique to plants alone [1]. As reported for animals, plants defend themselves by a combination of induced and constitutive mechanisms through an intricate network of signals.

The identification of plant genes responsible for defense against pathogen attack has been an important area of research. Among different gene categories identified, Resistance (*R*) genes are essential for plant survival, since they account for the first interaction between plant and pathogen, inducing the hypersensitive response (HR) after the perception of the pathogen through the so called gene-for-gene interaction [2-4]. The identification of a pathogen typically occurs when an avirulence protein, secreted by the pathogen interacts with the product of a plant *R*-gene, activating resistance mechanisms, such as the onset of Systemic Acquired Resistance (SAR) triggering pathogenesis-related (*PR*) genes (Fig. 1). Evidences show that this gene category is ancient in plants, being highly polymorphic, probably due to their coevolution as a response to the pressure of avirulence factors secreted by the pathogens. Even though these genes are so important for breeding purposes, at present automatic bioinformatic annotation tools are unavailable. It is observed that most *R*-genes have a number of conserved functional domains that belong to different classes [2-4] some of them also present in other gene families, making them traceable using bioinformatic tools.

The plant-pathogen recognition leads to the activation of several signal transduction pathways, which interact to form a complex network that culminates in a resistance response, failure of which leads to susceptibility [5]. In the currently proposed model, plants respond to pathogen attack through two types of immune response: (1) baseline resistance and (2) race-specific resistance [6]. Similarly, the genes involved in these types of response are represented in two classes. In the first line of defense, the receptors that trigger an unspecific response are responsible for recognition of Microbial- or Pathogen Associated Molecular Patterns (MAMPs or PAMPs) and DAMPs (Damage-Associated Molecular Patterns). In the second line of defense *R*-genes trigger a specific response to a given pathogen attack after recognition of intracellular elicitors, where *R*-gene product induces changes on gene expression levels, ultimately reshaping the pool of expressed proteins or determining post-translational modifications of available proteins [7].

The present work brings an up-to-date overview on *R*-gene structural diversity, their modes of action, pathogen specificity, classification and evolution, as well as considerations regarding their uses and applications.

Defense Mechanisms

PAMPS, MAMPS and DAMPS

The immune system is a sleeping process in healthy (non-stressed) plants, being switched on after perception of PAMPs from invading pathogens [8]. Animals and also plants are capable of discriminating between self and non-self, a key capacity to defend from pathogens [9]. PAMP alarm signals are perceived by plant Pattern-Recognition Receptors (PRRs), which are receptor-like kinases located in the cell surface, encompassing a ligand-binding extracellular domain and an intracellular kinase, or other receptor-like molecules, which present no recognizable intracellular signaling domain. Plant immune system depends on several messengers to trigger information generated by PAMPs and deliver the information downstream of PRRs to proteins which interpret signals and initiate defense gene expression [10].

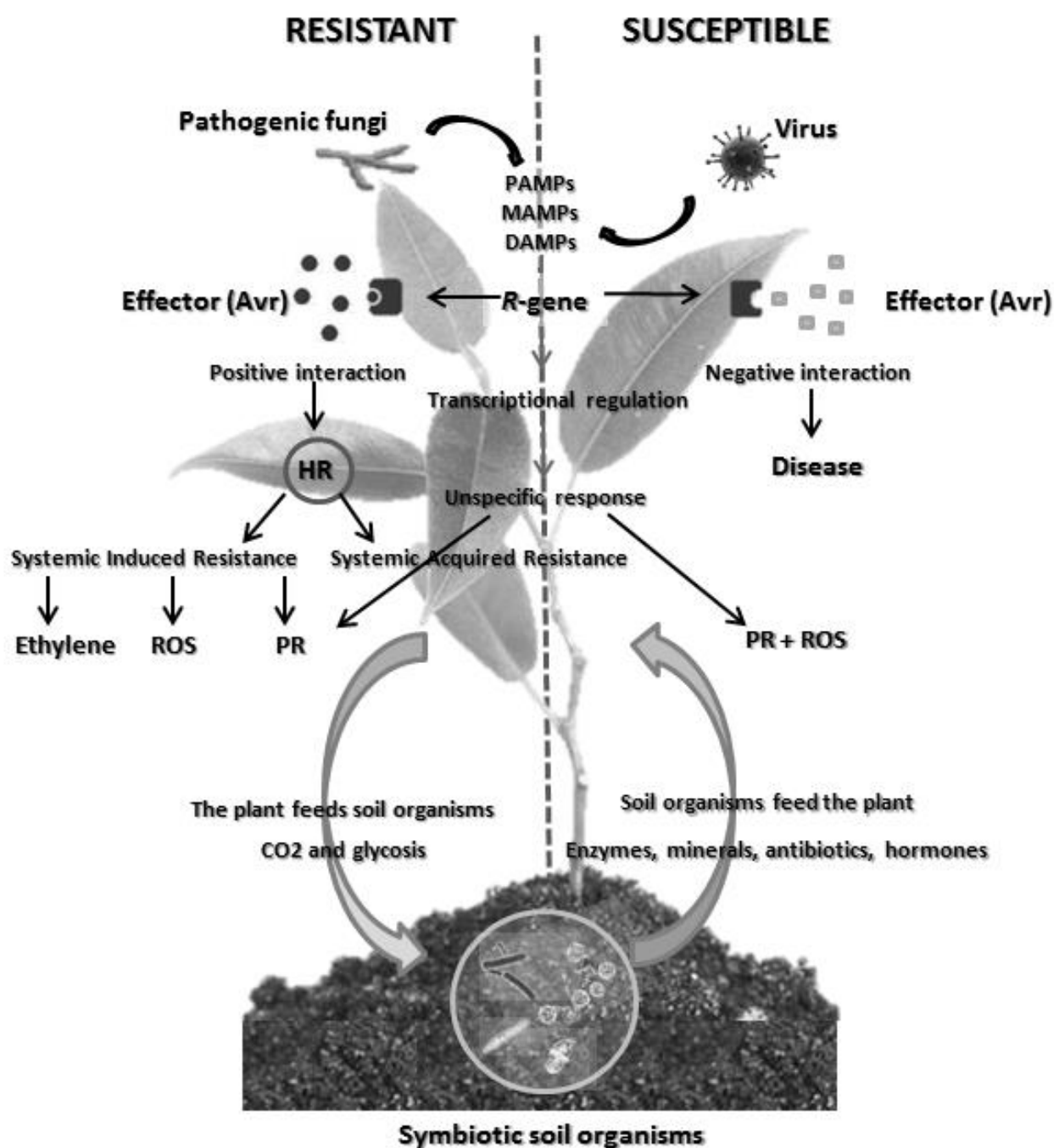


Fig. 1: Main steps of positive (resistant) and negative (susceptible) plant-pathogen interactions using a fungal and a viral pathogen, respectively. PAMPs or MAMPs (Microbial- or Pathogen Associated Molecular Patterns) and DAMPs (Damage-Associated Molecular Patterns) trigger unspecific response due to transcriptional regulation. The positive commensal interaction of plants and soil microorganisms is also illustrated.

The term PAMP has been sometimes criticized considering that not only pathogens, but microbes in general, express the recognized molecules. Therefore, the term MAMP has been proposed. Previously described PAMPs and MAMPs classes can be classified into two types (i) polypeptide and (ii) carbohydrate [11]. Examples of well-characterized MAMP are bacterial lipopolysaccharides, bacterial flagellins, and lipoteichoic acids from gram-positive and also gram-negative bacteria, peptidoglycans and nucleic acids variants normally associated with viruses, such as double-stranded RNA [12]. Several MAMPs and their corresponding PRRs were recognized in the past decades. Plant PRR representatives regard usually receptor kinases, such as transmembrane receptor-like proteins (RLPs) and transmembrane receptor-like kinases (RLKs) [13], whereas the number of RLP and RLK representatives varies among species. For example, *Arabidopsis thaliana* presents half the number of RLK genes when compared to rice, probably due to duplication events occurred in this last species [11,14]. Another example regards *Xa21*, a rice RLK gene that assigns specific resistance to *Xanthomonas oryzae* pv. *oryzae* [15]. The product of this gene also functions as a PRR for the *Ax21* PAMP effector [16]. Diverse researches indicated that conserved PAMPs (like fungal chitin, flg22, bacterial PGN and *Ax21*) can trigger innate immunity when perceived by rice cells [17].

Not only microbial elicitors are able to induce plant defense response, but also plant-derived molecules or chemical signals. This sensing of infectious-self or modified-self (compounds from endogenous origin delivered from structural barriers or other macromolecules, as lytic enzymes synthesized by invader pathogen or by the own host) is generally mediated by DAMPs [18]. Such DAMPs are normally in the apoplast and, like PAMPs, may represent a signal for danger, inducing innate immunity [19]. Oligogalacturonides and cutin released from plant cell walls are examples of DAMPs [20]. Emission of these endogenous signals allows disrupted or injured cells to communicate their damage to the tissue or, systemically, to all organs. For example, systemin is formed in damaged tomato leaves and is further perceived as primary signal, leading to systemic defense [21]. In a similar way in *Arabidopsis*, precursor proteins release a peptide (AtPep1) with 23-residues as a consequence of wounding, triggering an innate immune response that includes recognition through the PEPR1 receptor [22]. A system named damage-associated acts in a comparable manner as the mammalian immune system and detects “danger” signals through a series of molecules derived from other tissues, activating intracellular cascades that lead to an inflammatory response [23].

Plants react in diverse ways to the exposure to various microbial agents, whereas an extensive range of non-pathogenic endophytic microbial agents exist, which, far from harm, bring great benefits to plants. In a broad definition, endophytic organisms are those that live inside plants [24] establishing relationships that may be beneficial (mutualistic symbiosis) or harmful (antagonistic symbiosis) to the host. However, between these extremes there are organisms that can be easily isolated from the interior of the plants, but their presence have no apparent parasitism or mutualistic relation or modifications and are commonly treated as neutralistic cases. Interestingly, within this group some known pathogenic fungi and bacteria are found, indicating that the status of the plant-endophyte relationship may be transient and depends on numerous factors [25-27].

The way in which microbes invade a host cell successfully, suppressing its defenses, has been often attributed to interactions including the bacterial Type III Secretion System (TTSS), which is central to the virulence of many gram-negative bacteria in plants and animals [28]. In this way, bacterial pathogens directly inject effectors into the host cytoplasm. However, TTSS is not limited to pathogenic interactions, since symbionts, such as nodulating rhizobia, also depend of TTSS for communication [29], indicating that different interaction mechanisms may act during the recognition process. There are unsolved questions concerning the immune mechanisms used by plant roots, and one may propose that these are quite different from those used in leaves. In a comparative manner to the digestive tracts of animals, there are symbiotic microorganisms that are usually associated with plant roots, bringing benefits to the plant health, also helping to regulate various immune responses in a spatial and dynamic plan, also interacting with other cellular functions [30].

In order to induce defense mechanisms against a given pathogen, the plant must recognize the threat by perceiving intrinsic factors secreted by the pathogen and also connect these factors in order to counter attack by activating multiple defense signals. This switch is often represented by an *R*-gene, the class discussed in this review. Various aspects including their structures, modes of action, diversity, expression mechanisms, associated metabolic pathways, as well as their biotechnological potential and limitations are discussed.

Studies have revealed that plants defense system consists of different levels of protection that evolved through a constant coevolution battle between plants and their pathogens [31]. Since plants do not possess an immune system, they need to develop a variety of strategies to fight against biotic and abiotic stresses [32]. The first strategy that plants developed to survive and defend themselves against pathogens regards the formation of a physical barrier between the cell and pathogen through cell wall strengthening, a process promoted by the production of enzymes linked to lignin biosynthesis [33]. The second one, the called innate resistance, has many components and is not pathogen-specific. It includes the production of secondary metabolites and hydrolytic enzymes, comprising antimicrobial components. These components are synthesized once the plant detects the presence of PAMPS [34]. Finally the third strategy regards HR being triggered as a result of compatible interaction between the products of a host resistance gene and a pathogen avirulence (*avr*) gene. This response generates programmed localized cell death and, as consequence, SAR, that confers long-lasting resistance against a broad-spectrum of microorganisms [35].

Qualitative and quantitative resistance

Resistance against diseases can be classified into two categories: qualitative resistance conferred by a single resistance (*R*) gene and quantitative resistance orchestrated by several genes (Quantitative Trait Loci, QTL) [36]. Compared with qualitative resistance, the resistance conferred by QTLs is characterized by a partial and durable resistance effect, which is usually not species-specific [37]. This kind of resistance is very important against necrotrophic pathogens as well as against some biotrophic pathogens as *Xanthomonas oryzae* pv. *oryzicola*. Despite being widely used in plant breeding programs, the mechanisms of quantitative resistance are still poorly understood. However, being the most desired, due to increases in the productivity and obtaining of healthier plants with less pesticides, great expectation and good prospects exist for the application of knowledge gained from the study of QTLs for crop improvement [36]. It has been considered the second line of defense, where a dynamic battle between plant and pathogen coevolution also occurs. It leads to a more efficient, fast and strong response, including rapid influx of Calcium ions stored externally to the cytoplasm, the burst of reactive oxygen compounds, activation of MAPK protein kinases, reprogramming of gene expression, cell wall thickening at the site of infection and

finally localized cell death and HR, preventing the pathogen to spread and colonize the plant [5].

However, examining the modulation of gene expression mediated by both lines of defense, there is a considerable overlap between both gene expression profiles, suggesting that the mechanisms used by both types of answer are the same but vary in magnitude [38]. Differential expression studies have indicated that substantial changes in host gene expression may be detected after contact with various types of pathogens and the induction of this wide range of defense strategies demand a massive redistribution of energy throughout the process [32,39]. Within quantitative responses, SAR is a mechanism of induced defense that appears to be conserved across diverse plant families, conferring durable protection to diverse pathogens [40]. This acquired resistance generally spreads within plant tissues and can be activated not only by pathogens that cause necrosis, participating also of HR or, still, as an intrinsic symptom of the disease. The resistance conferred lasts long periods, sometimes throughout the plant life, conferring protection against different viral, bacterial, and fungal pathogens [35]. The reactions responsible for the SAR response include both biochemical and cytological changes, depending upon the transmission of signals from infection site to elicit systemic resistance.

The first systematic study on SAR in *Nicotiana tabacum* (tobacco) showed that SAR prevails for 20 days or more [41] and depends on the coordinated accumulation of both pathogenesis-related proteins and salicylic acid in the whole plant. This phenomenon can be induced by virulent, avirulent, or nonpathogenic microbes to which the plant was exposed, and also artificially using chemicals, including salicylic acid (SA), benzo-(1,2,3)-thiadiazole-7-carbothioic acid S-methyl ester (BTH) or 2,6-dichloro-isonicotinic acid (INA) [40]. Recent studies have revealed that systemic accumulation of SA in the course of SAR is anticipated by diverse metabolic signals, like jasmonate and methyl salicylate; the accumulation of azeleic acid, glycerolipids, whereas the light regime appeared to be also involved in the process [42].

There is growing interest in analyzing the SAR process in association with epigenetic regulatory mechanisms, such as DNA methylation and chromatin condensation which can affect the duration of the immune response, the modulation of gene expression, also influencing plant resistance [43]. Interestingly, it was demonstrated that SAR against a given disease can be passed onto the next generation of plants exposed to an environment which suffers biotic stresses, indicating the existence of an epigenetic mechanism of protection against diseases that functions as a memory related to the stress suffered by the previous generation [44]. In several studies of this nature an increase in somatic homologous recombination in the population of plants under stress was observed, whereas these increased levels of recombination persisted in subsequent generations even in the absence of stress. This genomic flexibility probably collaborated to the adaptive potential due to the stress epigenetic memory [45]. For this reason it was proposed to use the term 'transgenerational SAR' to define the phenomenon of epigenetic inherited memory associated with such stresses [42].

Defense mediated by R-genes

Similarities with animal immune system

An ancient origin of the innate immune response has been already proposed [13]. Common characteristics of innate immunity in animals and plants include receptors for microbe-associated molecules, conserved mitogen-associated protein kinase (MAPKs), signaling cascades and the production of antimicrobial peptides that are able to recognize potential pathogens by detecting PAMPs [46]. It was also proposed that plants and animals perception regarding molecular PAMPs occur in a similar way. This includes the discovery

of molecules in plants that resemble mammalian structures like Toll-like receptors, cytoplasmic nucleotide-binding (NB) oligomerization domains, and leucine-rich repeats (LRR) [47]. Evolutionary studies suggest that these similarities occur due to their pre-existence in an ancient unicellular eukaryote, which lived even before the separation of the plant and animal kingdoms leading to actual similarities in immune systems across kingdom borders [46]. The higher eukaryotes have several differences in regard to the nature of the receptors involved and the pattern of molecular recognition, despite of existing recent discoveries that highlighted common recognition strategies among them [47].

TIR-like genes

TIR (Toll/Interleukin-1 Receptor) domain is present in the immune system of plants and animals [48]. In animal Toll-like receptor occurs as an intracellular signaling domain, which activates an immune response after PAMP perception at the extracellular level [49]. In plants, this domain occurs at the N-terminus of a major subclass of cytoplasmic proteins (NB-LRR family) that – after pathogen perception – induces defense response [33]. The TIR class is found in bryophytes [50] and in gymnosperms [51], but studies in monocots have reported only non-TIR sequences [52,53]. A question remains open if TIR-NBS-LRRs has never developed or has been lost in monocots [54]. No TIR domain was found in recent studies with *Musa* and *Zingiber*, but we can only presuppose they are absent in all monocots, especially due to the fact that most studies have been limited to agriculturally important species of the Poaceae family [54].

NBS domain and its importance in hypersensitive response

The Nucleotide-Binding Site (NBS) conserved domain is responsible for the binding activity to ATP or GTP and is believed to act in signal transduction [55,56], while the LRR domain (variable) is involved in pathogen recognition by the host during infection. The junction of these two well-coordinated functions results in greater efficiency of plant defenses against major diseases [57]. NBS – also called P-loop or NB-ARC – is present in many proteins and is responsible for binding to ATP or GTP, viz., ATP synthetase subunits and ribosomal elongation factors [56]. The presence of these domains in *R*-genes suggests that these sites are involved in activation of kinases or carrier proteins [58]. Thus, mutations in NBS structure prevent the occurrence of HR, emphasizing the important function of this domain in signaling [59].

LRR repeats and their biological significance

The LRR domain is built in a similar way in plants and animals. It includes tandem repeats of a widespread structural motif of 20 to 30 amino acids with conserved leucines. The solenoid structure of this domain is particularly suitable for protein-protein interactions, being the non-conserved residues responsible for the specific interaction with other proteins [60]. Considering their conservation, LRR motifs may be ordered into seven subfamilies, among which a single one is specific to plants [61].

In plants, the major function of this domain seems to be promoting a versatile environment for mediating the interaction between proteins acting both extracellularly as a receptor to pathogen's molecules or intracellularly, in a step before activation of the signal transduction pathway involved in resistance process [62]. In addition to protein-protein interaction, the LRR domains also function as peptide-ligand binding sites and sites of protein-carbohydrate interactions, related to resistance specificity [63]. It is believed that the amino-terminal portion of the LRR domain (which is highly variable) brings a more significant contribution to the recognition than the carboxyl-terminus (which is more conserved) [64].

The development of neurons, differentiation and growth of gonad and glands as well as antifungal response are examples of functions assigned to the LRR-containing receptors in animals. [60]. These features represent a start point of a complex array of signaling pathways converging to enable the transcription of target genes in the nucleus [65]. Some workers propose that general structural similarities among animals and plants PRRs, including the use of a LRR domain for ligand binding, might have arisen through convergent, rather than divergent evolution [46,66,67].

Classification of plant *R*-genes

Resistance genes codify proteins that share a limited number of well conserved motifs and domains [3]. The most commonly accepted *R*-gene families are: (i) Reductases; (ii) Serine/Threonine Kinases; (iii) NBS-LRR (Nucleotide Binding Site-Leucine Rich Repeats), (iv) LRR (Leucine Rich Repeats), (v) LRR-Kinases and (vi) resistance genes that encode distinct proteins. The first five members are grouped according to their structure and the presence of conserved domains [53]. However, this classification varies from author to author [68]. While five distinct *R*-gene classes are considered by some authors [69,70] others enumerate the two subclasses of the third class (NBS-LRR) as distinct ones [71]. Jones [72] enumerates five classes, but in different sequence. In accordance with the majority of the authors, here we will consider the first classification, pointing five different *R*-gene classes that present conserved domains (Fig. 2).

Conserved domains are absent only in the first class i.e. reductases. An example of this class is the *HM1* gene of maize that codifies a NADPH dependent reductase, which inactivates the *Helminthosporium carbonum* fungus toxins [73]. Phylogenetic studies show that this gene developed only in the grass (Poaceae) lineage being possibly restricted to this plant family. *RPW8* of *Arabidopsis* is another representative that codifies a membrane anchored reductase with a helical structure without similarity with any other *R*-genes [63]. Another example regards *Mlo* of barley, which is the putative defense regulator against *Blumenaria graminis*. It has not yet been elucidated how the defense mechanism of *Mlo* works [74]

The second class is characterized by an intracellular serine/threonine-kinase (ser/thre-kinase) domain. This class can be typified by the *Pto* gene that is able to interact with the *avrPto* gene conferring to tomato resistance against the gram-negative bacteria *Pseudomonas syringae* pv. *tomato* [75].

The third class is the largest group of predicted intracellular plant immune receptors [76] represented by genes that have two associated domains: LRR and NBS [77]. The NB-LRR class can be subdivided based on deduced N-terminal structural features: many have a domain presenting considerable conservation when compared to *Drosophila*-Toll and mammalian interleukin receptors (TIR-NB-LRR), whereas the other contains putative coiled-coil domains (CC-NB-LRR) [78]. The LRR domain is involved in the recognition of the pathogen by the host during infection process. Proteins with this domain present leucine motifs, including 20-30 repeats grouped in tandem at regular intervals that form the domain and may also present other hydrophobic residues, or asparagines and prolines [60]. The NBS (NB-ARC or p-loop) domain occurs in different classes of proteins and is responsible for binding activities to ATP or GTP [56]. It has been suggested that *R*-genes play a role in activation of protein kinases and that mutations in their structures may impair HR [59]. Genes bearing only the NB-ARC domain can trigger the defense response even when other domains are absent, suggesting that the NB-ARC alone can be responsible for the signaling cascade [56].

The fourth *R*-gene class codifies a membrane anchored protein composed of an extracellular LRR and by a transmembrane region that ends in a short intracellular C terminal

tail. The tomato *Cf* gene family is a representative of this class that confers resistance against the fungus *Cladosporium fulvum* [79].

The rice *Xa21* gene confers resistance to the *Xanthomonas oryzae* pv. *oryzae* bacteria and is a member of the fifth class [80]. This gene codifies an extracellular LRR domain and also a ser/thre-kinase domain (similar to the *Cf* gene from the fourth class and also to the *Pto* gene, from the second class), what suggest an evolutionary ligation among different classes of plant *R*-genes [81]. The LRR domain in most known genes of this class have no introns, unlike the LRR-kinases, as *ERECTA*, a development related gene that is interrupted by introns in the first leucine of the 'xxLxLxx' motif [60].

Copy Number Variations, Genome Distribution and Evolution of *R*-genes

Copy number variations (CNVs) are known as genomic rearrangements that result from losses or gains of DNA pieces, recently reported as a common type of polymorphism in plant genomes [82]. They may result from unbalanced DNA modifications, including structural rearrangements (duplications, deletions, inversions and translocations, as well as events of unequal crossing overs), triggering changes in the number of particular DNA sequence copies.

Among the functionally annotated genes located within CNV regions in soybean, those related to disease resistance and biotic stress response were significantly overrepresented [57]. Similar observations were made for *Arabidopsis* and rice, where *R*-genes represent a significant fraction of genes in CNV regions [83,84]. High incidence of duplication benefits the variability of *R*-genes, being advantageous to face changing environmental conditions. *R*-genes are subject to strong selective pressure promoting coevolution with pathogen effector proteins. Depending on the presence and absence of particular pathogens, the pressure for the selection of corresponding *R*-genes dramatically changes, leading to rapid evolution [57]. Despite NBS-LRR family being the most representative among CNV regions, the large size of this gene category is, per se, not enough to promote CNVs occurrence. Therefore, some gene families appear to be more prone to present CNVs than others. Additionally, members of some gene families seem to be not similarly predisposed to CNV. For instance, clustered genes (especially those organized in tandem arrays) appear to generate CNVs more frequently when compared to isolated family members [57], what is consistent with recombination-based mechanisms of CNV formation, although it does not exclude alternative mechanisms.

In many crop species, the *R*-genes are organized in groups along the chromosomes, called clusters [85]. This phenomenon is well known and molecular studies have demonstrated that clusterization generally reflects events of duplication inside chromosome regions that harbor ancestral resistance genes creating loci rich in *R*-genes [86]. Intragenic and extragenic recombination events – due to unequal crossing-over – are responsible to generate the haplotype diversity and new resistance specificities [87]. Theoretically gene clusters function as a reservoir of mutations. In the *Arabidopsis* genome 182 genes comprise 20 resistance clusters, 15 out of which are LRR-like [87]. LRRs probably evolved from duplication, mutation and exons recombination. By being involved in the recognition and specificity of proteins, are likely targets of selection and adaptive pressure [60]. As an example, in tomato, *Pto* and *Prf* also appear in clusters [61]. For monocots, although there is not collinearity in relation to ortholog maps in rice, barley and corn, there's a general common organization of *R*-genes among these species [61].

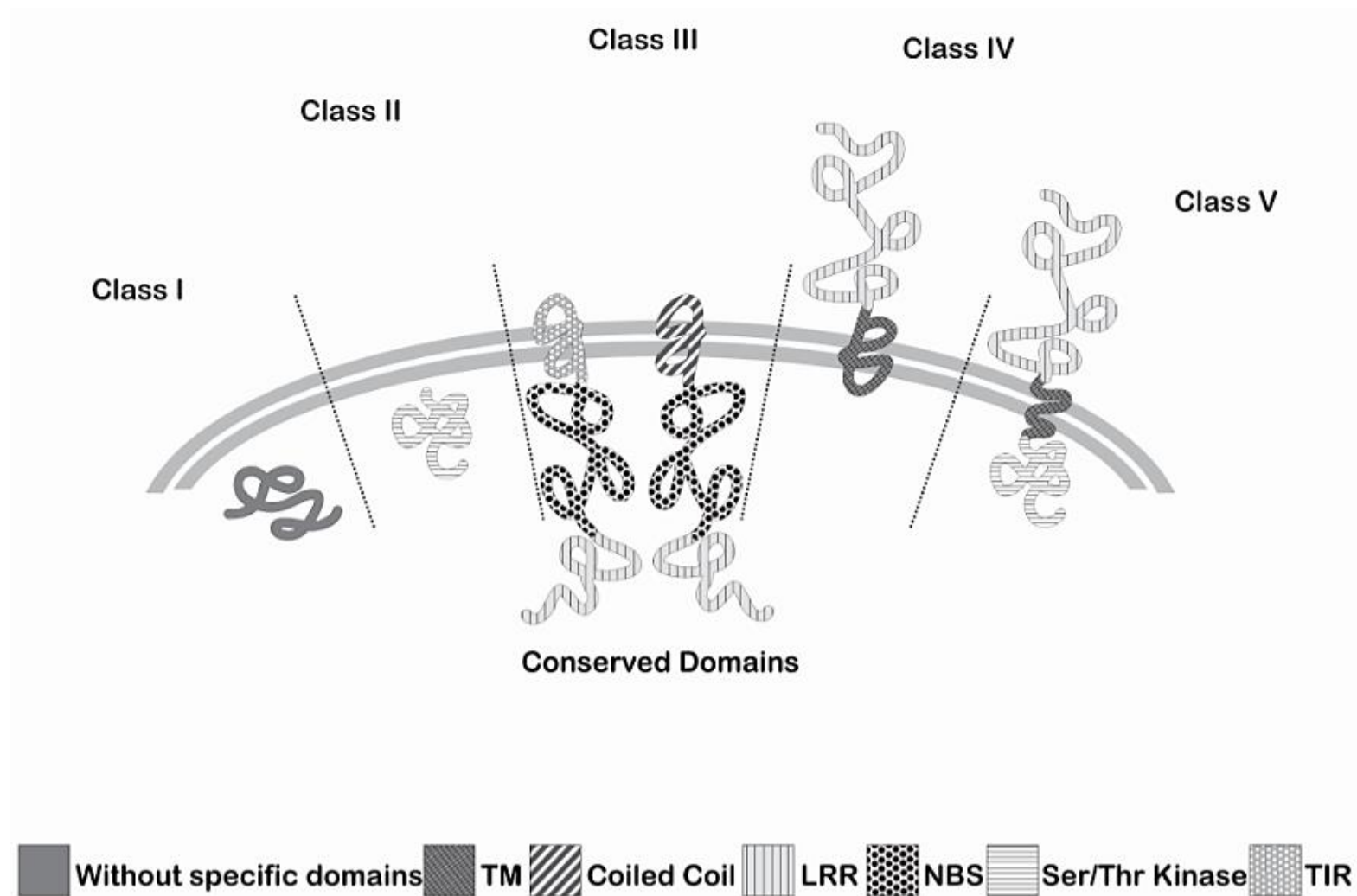


Fig. 2: Graphic representation of five known *R*-genes families including I: Reductases; II: Kinases; III: NBS-LRR; IV: LRR; V: LRR-Kinases. Different colors represent the conserved domains and position is relative to cell membrane. Abbreviations of conserved domains: NBS = Nucleotide Binding Site; LRR = Leucine Rich Repeats; TM = Transmembrane; TIR = Toll-Interleukin Receptor.

The great similarity among sequences of the majority of cloned *R*-genes, even between distant species, may confirm the hypothesis that they belong to a large multigene family that diverged after duplication and mutation events, thus resulting in different specificities [88], emerging, however, from a mutual ancestor. This fact is also reflected when in similarity analyses (both molecular and *in silico*) two orthologous sequences from different organisms such as sugarcane and corn show more similarity to each other, than two sequences of sugarcane, for example; as can be observed in *RPR1 gene* [89] and also in *Pto*, *Xa1*, *Cf-9* [70] and *Xa21* [53]. On the other hand, it was suggested that the diversity of the resistance specificity reflects a much faster evolutionary rate than observed in spontaneous mutations, probably mainly due to the selective pressure of the pathogen on the host [90].

R-clusters recombination rates are not necessarily homogeneous or consistent on the evolutionary scale. Recent evidences show that the recombination in these clusters is enhanced when any pathogen infection occurs, suggesting a mechanism that induces temporarily instability in the genome as an answer to an extreme stress. DNA methylation and changes in chromatin allow the modulation of such instability that is probably restricted to certain genome regions [60]. Polymorphism of *R*-gene alleles can be determinant to achieve resistance or susceptibility response. It is believed that susceptibility alleles generally occur due to mutation in those associated to resistance but little is known about the evolution dynamics of these genes [87].

The family Poaceae is a good example. It has been considered that this family originated about 60 MYA and diversified as a result of polyploidy (also regarded as genome-wide duplication) [91], besides in tandem gene duplication and events of local chromosome changes [92]. Wanderley-Nogueira and coworkers [53] showed the existence of synteny among sugarcane and rice karyotypes, allowing the identification of 1,069 sites in rice chromosomes where sugarcane sequences anchored. Both, repetitions *in tandem* and repeated sequences were detected, with distribution along distinct chromosomes or in the same chromosome, forming *R*-clusters as described in *Zea mays* [93], *Lactuca sativa* [94], *Avena sativa* [95], *Linum usitatissimum* [96] and *Cicer arietinum* [97].

A similar analysis comparing *Glycine max* and *Medicago truncatula* was carried out, where soybean *R*-gene sequences anchored in 1,253 sites in barrel medic virtual chromosomes. These sequences formed clusters especially in four chromosomes, uncovering not only regions bearing tandem repetitions, but also grouping of distinct *R*-genes positioned next to each other, sometimes intercalating [98]. Legumes are considered an ancient family and probably originated at about 54 Mya [99]. In this context, *G. max* and other legumes from the papilionoid subfamily share an ancient duplication (ca. 13 Mya), whereas soybean probably also underwent polyploidy [100]. Such duplications are extensively detectable, being evident in number of duplicated genes and also in large syntenic areas between distinct chromosome regions. For example, a given *M. truncatula* region generally corresponds to two soybean regions, and considering evidences as large similarities in gene densities and distribution, *M. truncatula* could be considered “a simplified draft” of the soybean genome. [101].

To illustrate the clustering of *R*-genes we anchored selected representative gene candidates (*N*, *P2*, *RPP5*, *RPS4*, *L6*, *SSI4*, *RPP4*, *M* and *RPP1*) of the TIR-NBS-LRR gene family from cowpea (*Vigna unguiculata*) in common bean (*Phaseolus vulgaris*) genome. Like other *R*-gene classes they clustered in given chromosome regions, sometimes forming large blocks of tandem repeats (Fig. 3) generally due to duplication events. The results were similar to those observed when anchoring soybean *R*-genes in the legume *M. truncatula* [102]. It is interesting to note that (like other gene families) *R*-genes usually cluster in

terminal or subterminal positions of chromosome arms, possibly facilitating their evolution by reciprocal translocations.

Existing evidences on diversity, abundance and clustering of *R*-genes support assumptions that future efforts regarding increased pathogen resistance may rely on biotechnological inferences that consider whole gene clusters naturally associated in neighboring positions, rather than isolated genes [103], as has been traditionally done.

Data Banks and web-resources for *R*-genes and *in silico* routines

The main focus of "omics" study has been the identification and association of data generated by various approaches of genomics, transcriptomics, proteomics, interactomics, metabolomics and epigenomics, using bioinformatic tools and systems biology. The aim of all these analyses is to reveal relationships between different molecules to uncover their interactions, generally associating data deposited in diverse databases to others generated in laboratory and field experiments [104].

Comparative genome analysis involves diverse applications in various fields of knowledge, especially in approaches regarding molecular mechanisms of pathogenesis, the host range and phenotypical particularities of plant-pathogen relationship. In this sense, different strategies have been used in the comparison of genomic sequences, thus, providing multiple perspectives regarding studied organisms [105]. Additionally, the comparative analysis of the genome of different pathogens brought important information to explain their mechanisms of penetration and colonization in plants and also to identify new agents involved in these processes [106].

The first *R*-gene database generated using bioinformatics tools was PRGdb (Plant Resistance Genes Database, <http://www.prgdb.org>; [107]), developed in 2009. This database includes more than 15000 sequences among identified and predicted *R*-genes pertaining to 192 different plant species.

Bioinformatic analysis allows the generation of databases readily accessible and rich in biological information. Furthermore, a pipeline (DRAGO, Disease Resistance Analysis and Gene Orthology) was proposed, allowing the identification of plant *R*-genes in public datasets such as Unigene and GenBank [107]. Another initiative in the same direction regarded SolRgene, an online database (<http://www.plantbreeding.wur.nl/SolRgenes/>) specialized in exploring genes associated to disease resistance in tuberous *Solanum* taxa, containing information regarding resistance against *P. infestans*, presence of *R*-gene homologs in members of *Solanum* section Petota (1,062 accessions and 12,584 genotypes). Data covering mining of *R*-gene alleles were incorporated and enabled detection of *R*-gene homologs in other germplasm sources [108].

Together with databases, many tools and programs have been used aiming at identifying and analyzing sequences, being chosen depending on the desired approach to the study. This step-by-step study or information workflow is the so called *in silico* routines. Popular tools such as the Basic Local Alignment Search Tool (BLAST) [109] help in gene structural characterization when comparing candidate sequence with any other in any database. CLUSTALX [110] is among the most commonly used tool to obtain biologically meaningful multiple alignments of divergent sequences. Also several solutions have been developed for the identification of genomic structural variants via high-throughput sequencing [111], as a novel transcriptome-to-genome alignment tool named *Dissect* (DIScovery of Structural Alteration Event Containing Transcripts), suitable to highlight and characterize duplication events, as well as inversions, fusions and rearrangements and for discovering structural variations in whole transcriptome involving highly assembled contigs or long reads [112].

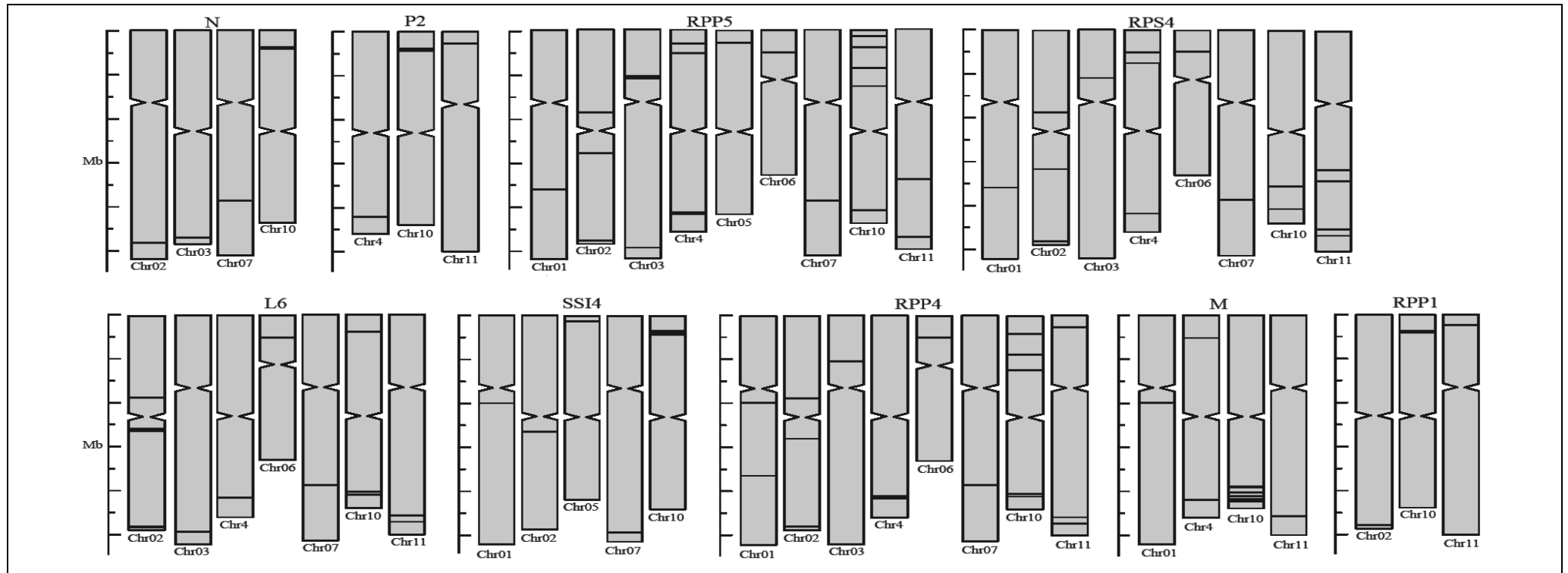


Fig. 3: Anchoring of selected representative gene candidates (*N*, *P2*, *RPP5*, *RPS4*, *L6*, *SSI4*, *RPP4*, *M* and *RPP1*) of the TIR-NBS-LRR gene family (*Toll Interleucine Receptor-Nucleotide Binding Site-Leucine Rich Repeats*) from cowpea [*Vigna unguiculata* L. (Walp.)] from NordEst data bank (<http://bioinfo03.ibi.unicamp.br/vigna/>) in *Phaseolus vulgaris* L. genome available at Phytozome (<http://www.phytozome.net/commonbean.php>). From the 11 *P. vulgaris* chromosomes nine presented clusters of this R-gene family, mostly in the subterminal position of chromosome arms. The largest number was observed in chromosome 10 (nine gene categories). The most represented gene was *RPP5* found in eight chromosomes, represented by 41 loci, while *Rpp1* was the less represented, found only in two loci. Tandem repetitions (thicker lines) were observed in chromosomes 2, 7, 10 and 11, indicating possible duplication events.

In order to accurately compare gene structures, the position and level of similarity within a set of sequences is needed, highlighting conserved regions. Once identified, a special focus to the conservation degree around introns is essential in order to determine whether they can be considered as conserved introns between paralogs/orthologs. GECA is a freely available tool for the analysis of gene exon/intron organization, considered to be user friendly and fast, identifying modifications between members of a given gene family [113].

Another program, the Molecular Evolutionary Genetics Analysis (MEGA) [114] was designed not only to estimate genetic distances as well as to make phylogenetic inferences through dendrogram building. It also enables the calculation of distance matrices, analysis of sequence composition, as well as inferring on the percentage of similarity between sequences, allowing the association of these data to morphological and physiological characteristics, important for inferences on the structure and evolution of sequences of interest.

Differential expression profiling analysis has been also demanded, aiming to identify differentially expressed genes under different tissues or experimental conditions or still in contrasting genotypes (resistant/susceptible) under the same condition. *In silico* analysis of gene expression can also be studied by generation of heat maps using programs as *CLUSTER* [115] where hierarchical clustering methods allow expression analysis simulating microarray experiments based in a data matrix provided by the user. A basic pipeline formed by these approaches contributes to construct a primary panel of *R*-genes that participate in the plants response to pathogens as reported by some authors [53,69,98].

Plant-pathogen interaction and coexpression assays

Plant resistance proteins, in general, recognize avirulence proteins known as pathogen effectors and then trigger a defense cascade. Flor [116] was the first to describe this way of interaction and later his proposition was supplemented by the guard model that explains the so-called indirect interaction, in which the *avr* gene product is recognized by a guard as a target [78].

To understand how the initial recognition event *Avr9/Cf9* is transmitted and triggers the plant defense, *Cf-9* resistance gene from tomato was analyzed against *C. fulvum* [117]. The authors investigated the activation of protein kinases in tobacco expressing the gene *Cf-9*, concluding that two kinases (46 e 48 kD) became rapidly activated (in a gene-for-gene manner) within 2 to 5 min after elicitation with *Avr9* in *Cf9* tobacco plants and cultured cells derived. Both kinases were similar to WIPK (Wound Induced Protein Kinase) and SIPK (Salicylic Acid-Induced Protein Kinase), two MAP kinases. Thus, the data indicated that the *R/Avr* induction was accompanied by several mechanisms and that signaling was connected to MAP kinases, also in response to abiotic stimuli, such as mechanical stresses. In turn, Römer and coworkers (2007) showed that recognition of pathogen *Xanthomonas campestris* pv. *vesicatoria* (Xcv) was mediated by activation of the *Bs3* (a resistance gene) pepper promoter. The acquired data suggested that *Avr* proteins bind and activate the promoter of the cognate *R*-gene in a specific region.

In *Erwinia amylovora* genome an analog to the *Pseudomonas syringae* pv. *tomato* effector protein *AvrRpt2* was identified, being annotated as *AvrRpt2EA* in *Malus robusta* accessions (Mr5) [118]. The *AvrRpt2* *R*-gene is known to activate *RPS2* (from *A. thaliana*) by cleaving *RIN4*, a *RPS2* guard [119]. The study aimed to reinforce the proposition of a resistance-associated mechanism in Mr5 comparable to that known for *A. thaliana*. After inoculation of Mr5 shoots with *E. amylovora* mutant and wild type strains, results showed that while the wild type strains were unable to infect Mr5, the Mr5 resistance was overcome by the mutant strain causing >50% necrosis of the shoots. In turn, after inoculation of pear fruits *avrRpt2EA* with a mutant bearing a deletion, decreased disease symptoms could be

observed, indicating that the mutant carried a potential virulence factor associated to the pear *versus E. amylovora* interaction system [120]. In this regard, a region exhibiting a single nucleotide polymorphism (SNP) could be detected by sequencing the *avrRpt2EA* gene from different strains regarding virulence to Mr5; this SNP altered the amino acid sequence what suggests that MR5 resistance to fire blight may be highly strain-specific [121].

Several studies encompassing plant-pathogen co-expression have been developed in Solanaceae species associated to the pathogen *Pseudomonas infestans*. A set of 54 *avr* genes from *P. infestans* were computationally predicted, being used for identification, characterization and cloning of *R*-genes with broad spectrum in *Solanum* wild relatives [122]. RXLR effector group IpiO (profiled for activation of innate immunity) was able to induce HR in *S. bulbocastanum*, *S. papita* and *S. stoloniferum* containing the *R*-gene *Rpi-blb1*, uncovering response to IpiO and co-segregation of resistance to *P. infestans*. Another example was the study using Agrobacterium infiltration for transient expression of tomato plant resistance gene I-2 and *Fusarium oxysporum* f. sp. *lycopersici* effector (Avr2) protein in *Nicotiana benthamiana* leaves, where HR could be seen in the necrotic area of the leaf. Staining of leaves infiltrated with trypan blue allowed visual identification of HR and fusion between a fluorescent tag to a recombinant protein, facilitating visualization of its sub-cellular position using confocal microscopy [123].

Another study using *N. benthamiana* versus *P. infestans* (*R*-*avr* interaction) was performed [124]. *P. infestans* is known to deliver RXLR effector proteins (PexRD2) to plant cells to modulate host immune signaling and promote colonization. In turn, PexRD2 was proposed to interact with the kinase domain of plant MAPKKK ϵ as a positive regulator of cell death associated with plant immunity. By contrast, homologs of PexRD2 from *P. infestans* had reduced or no interaction with MAPKKK ϵ and did not promote disease susceptibility. Structural mutants of PexRD2 presented no interaction with MAPKKK ϵ and failed to support enhanced pathogen growth or perturb MAPKKK ϵ signaling pathways. This work provided evidence that *P. infestans* RXLR effector PexRD2 evolved to interact with a specific host MAPKKK to perturb plant immunity-related signaling.

Transcriptome data in large scale and gene coexpression networks are the basis for the study of biological networks in plants, where the nodes are represented by genes that are coexpressed and the edges are links between the nodes. Similar patterns of expression in different biological conditions can be observed by two genes connected by an edge [125]. Distinct biological networks comprise information regarding protein-protein as well as protein-DNA interactions and also gene co-expression networks, besides insights on phosphorylation, metabolism, and further structured data sources. A major goal of systems biology today is the identification of new signaling or expression modules from these networks [126]. Functional genomic studies in plants have contributed to a systems-level understanding of how genes function and how biological process is governed by the cooperation of a set of genes. This meta-analysis approach facilitated the comprehensive annotation of functions to unknown genes and the discovery of functional modules in plants [127], including defense mechanisms.

Sequence diversity and marker development

In a host-pathogen relationship genetic variability is important for both sides. If in one hand the *R*-genes of the plant must present polymorphisms to hinder pathogen infection, on the other hand, to pathogens is easier to discard compatibility factors encoded by *Avr* genes in a gene-to-gene resistance interaction, for example, than for the plant to create a completely new recognition capacity.

In the interactions between host and bacteria an effector molecule can act as virulence factor, which modulates the physiology of a susceptible plant, sustaining pathogen growth

or, in the case of a resistant host, as a virulence factor inducing HR [128]. Computational prediction may also allow the identification of *Avr* genes expressed by *P. infestans* responsible for potato late blight [122]. Such an inference permitted the identification of 54 *Avr* candidates in wild *Solanum* species from analyzed sequences including effectors bearing the signal peptide and the RXLR motif for activation of innate immunity (AVR activity). Considering the system potato versus *P. infestans*, the best strategy to improve resistance was to obtain cultivars with broad spectrum *R*-genes, such as those from *S. bulbocastanum* *Rpi-blb1* [129], due to the fact that resistance offered by lower spectra *R*-genes had been broken. *R1* was the first cloned *R*-gene from potato conferring resistance to late blight [130]. This gene belongs to the leucine zipper-NBS-LRR plant *R*-genes class, being mapped on potato chromosome V within a hot spot locus for pathogen resistance. Comparing the susceptibility allele to the resistant one at the *R1* locus, the resistant included a large insertion necessary for a functional *R*-gene. The *Prf* gene, which confers resistance against *P. syringae* is the most similar to *R1* gene, bearing 36% identity [130].

Champouret and coworkers [131] showed that profiling the variation of *Avr-blb1* within a *P. infestans* population allowed the prediction of the effectiveness of *Rpi-blb1*-mediated resistance in potato. The potential use of *Rpi-blb1* comes from molecular analysis of a BAC clone that comprised the target locus, which included a cluster of four resistance gene analog (RGA) candidates pertaining to the CC-NBS-LRR category of plant *R*-genes. The *Rpi-blb1* gene was able to complement the susceptible phenotype in potato and tomato background, demonstrating the potential of interspecific transfer of broad-spectrum late blight resistance to cultivated Solanaceae from sexually incompatible host species [129].

Another important point regarding genetic breeding concerns the response of elite materials that exhibited horizontal resistance to disease. In this sense Tian and coworkers [132] screened for genes induced in leaves of potato accessions with horizontal resistance to *P. infestans*. A cDNA microarray was used to identify *P. infestans*-induced genes from 100 expressed sequence tags (ESTs) from a subtractive cDNA library. From these, 76 transcripts were differentially expressed in infected plants as compared with mock-inoculated controls. Four groups of genes [Group A: strongly induced (>10-fold) at 72 hpi (hours post infection); Group B: up-regulated at all the three time points; Group C: transcripts peaked at 48 hpi; Group D: transcripts up-regulated at 24 hpi and decreased slightly] could be identified according to their expression patterns at three time points [24, 48 and 72 hpi]. The results showed that 35% of the up-regulated genes were associated to primary or secondary metabolism. 18% of the representatives regarded genes involved in cell rescue or protection against diseases, as well as those participating in the cellular organization. The results uncovered significant changes in metabolic pathways associated to defense responses among the various mechanisms involved in the horizontal potato resistance against late blight also revealing 15 genes of unknown function, regarding future targets for breeding purposes.

Within the perspectives of using different types of resistance, Solomon-Blackburn and coworkers [133] sought to distinguish major *R*-gene activated in field trials against potato late blight (*P. infestans*). For this purpose, potato cv. *stirling* (showing a *S. demissum* derived *R*-gene and quantitative field resistance to late blight) was crossed with Maris Piper (susceptible cultivar) to produce an F₁ population from which three genotypes were backcrossed to Maris Piper. F₁ and backcross populations were evaluated considering their resistance to simple race 1,4 (incompatible with *Stirling's R*-gene) and complex race 1,2,3,4,6,7 of *P. infestans* (compatible with *R*-gene) in whole plant glasshouse tests. Comparisons of the results with the simple and complex races apparently enabled F₁ clones to be classified for the presence or absence of the *R*-gene and to be assessed for their level of quantitative field resistance.

Distinction among field resistance and resistance based on *R*-genes seems to be valid for various host-pathogen interactions, excepting potato/*P. infestans* (late blight resistance). In a QTL analysis (regarding resistance to *P. infestans* in *S. microdontum*) Tan and coworkers [134] identified a QTL on chromosome 4. A qualitative genetic analysis resulted in the positioning of the *R* (Pi-mcd1) locus, capable to delay (1 up to 2 weeks) the infection. The position on the short arm of chromosome 4 coincided with a conserved *Phytophthora* *R*-gene cluster previously identified [which included *R2*, *R*(2-like), *R*(Pi-blb3) and *R*(Pi-abpt)], implying that *R*(Pi-mcd1) was the fifth *R*-gene of this NB-LRR cluster.

Resistance Gene Analogs (RGAs) have been isolated from many crops and represent potential tools for breeding disease resistance, especially when associated to marker-assisted selection. Also some *R*-genes have been analyzed based on amplification of their kinase domains or where kinases have been reported as being in the vicinity of *R*-genes, generating interesting polymorphisms. In a similar way, degenerate primers were used to develop RGAs in wheat and soybean, allowing amplification of clone US-01-1158 in sugarcane genome (*Saccharum* spp.) considered resistant against yellow leaf disease provoked by SCYLV (sugarcane yellow leaf virus) [135]. BLASTx results of cloned markers uncovered 20 sequences with homology to NBS-LRR RGAs, and eight to kinase genes with a possible role in disease resistance.

The completion of the *A. thaliana* genome sequence enabled the analysis of a more complete set of *R*-gene homologues in flowering plants. The accumulation and availability of new data opened perspectives for the discovery of new *R*-genes, as happened to RPW8, a new class of protein associated with pathogen recognition conferring resistance to powdery mildew in *A. thaliana* [72]. Similarly, isolation of the barley powdery mildew resistance gene *Mla*, for example, opened doors to understand the extensive allelic diversity of this locus [72]. During infection and resistance response the host generally undergoes changes in transcriptome. To understand these changes, studies have been conducted with a number of methodologies. For example Marathe and coworkers [136] studied the *A. thaliana* resistome in response to cucumber mosaic virus strain Y (CMV-Y) infection using whole genome microarray. Around 80 defense-responsive genes were identified, possibly participating in the defense mediated by *R* gene against both bacterial and viral pathogens.

Transcriptome analysis of the interaction between wheat (*Triticum aestivum* L.) and stripe rust [*Puccinia striiformis* f. sp. *tritici* (Pst)] was studied by Coram and coworkers [137] allowing the identification of 115 transcripts induced during an incompatible interaction, where 61 were HR-specific transcripts. Also, 73 transcripts were induced during a compatible interaction, regarding 19 transcripts associated to biotrophic interaction, including Yr5 gene, a major *R* gene that confers resistance to Pst races in USA. In this analysis, a GeneChip was used to profile the expression changes in wheat isolines (differing for Yr5 gene presence) after inoculation (6, 12, 24 and 48 hpi) with Pst. Transcript annotation revealed that the Yr5 presence resulted in a fast and increased resistance reaction comprising defense-related transcripts and signaling pathways associated to *R*-gene-mediated responses as protein kinase signaling and production of reactive oxygen species (ROS) leading to HR. On the opposite, the lack of *R*-gene signaling resulted in a weaker response.

In silico annotation of transcripts from transcriptomic data banks allowed the identification of *R*-genes in different plant species. For example, Wanderley-Nogueira and coworkers [70] searching for abundance and diversity of *R*-genes in sugarcane transcriptome, comprising 26 libraries (different tissues and development stages; 237,954 ESTs), identified 280 *R*-genes (196 with the expected conserved domains). All *R*-gene classes were observed except the subclass TIR-NBS-LRR, including Toll interleukin-1 receptors), with prevalence of the kinase (*Pto*-like) class. *R*-genes were detected in all libraries (including flowers, root to shoot transition zone, and roots). In leaves, low

expression level was detected for some gene categories, whereas other classes were absent. A high diversity of isoforms was observed in all classes of *R*-genes. Another example regards the evaluation of the transcriptome of *Eucalyptus* species by Barbosa-Silva et al. [69], where 210 *R*-gene candidates could be identified, with prevalence of the LRR-NBS-TIR class, with 93 representatives. Similarly, *in silico* characterization of EST data resulted in the detection of 28 NBS *R*-gene sequences (with NB-ARC domain) in *Curcuma longa* [138]. Most alignments occurred with monocots (67.8%, mainly *Oryza* and *Zingiber*), whereas best alignments with dicots occurred with *A. thaliana*, *Populus trichocarpa* and *Medicago sativa*.

The development of high- throughput sequencing technologies allowed a faster and cheaper generation of large scale transcriptomic data for the analysis of specific genotypes and treatments, resulting in valuable data banks. For instance, a data-mining approach was used to identify tentative *P. vulgaris* *R*-gene-like sequences from about 1.69 million (derived from Roche 454 sequencing) and also 116,716 ESTs available in GenBank [11]. This approach allowed the identification of 365 resistance gene-like sequences (PvRGLs), some of them with expression in leaves confirmed by RT-PCR. From the identified sequences, 105 (29%) PvRGLs could be integrated in the common bean physical map, while 237 (65%) PvRGLs were anchored on the *P. vulgaris* genetic map. Also, 11 sequence-tagged-site (STS) and 19 CAPS (Cleaved Amplified Polymorphic Sequence) markers were developed for 25 PvRGLs. The expressed PvRGLs identified provided a large resource for development of RGL-tagged markers, useful for genetic mapping of disease resistant candidate genes and QTLs.

Splicing approaches regarding *R*-genes

The generation of alternative transcripts has been reported for various *R*-genes, promoting additional diversity putatively important for functional resistance. For example, the biological significance of the alternative transcript *RPS4* from *Arabidopsis* in disease resistance was studied by Zhang and coworkers [139]. The *RPS4* protein pertains to TIR-NBS-LRR class and produces alternative transcripts with truncated open reading frames (ORFs), while the prevalent transcript is generated by intron 3 (or introns 2 and 3) retention. The removal of one or both introns abolished function of *rps4* transgene, whereas the expression was not affected. In turn, a partial resistance in transgenic lines expressing both intron-deficient and truncated transgenes was observed [139]. Authors concluded that both regular and alternative transcripts were necessary for the function. In a continuation of this work Zhang and Gassmann [140] concluded, in summary, that *RPS4* function is regulated at multiple levels, including gene expression, alternative splicing, and protein stability, presumably to fine-tune activity and limit damage inflicted by activated *RPS4* protein.

Biological importance of alternative transcripts was also suggested by Takabatake and coworkers [141] that followed the accumulation of two transcripts of the *N* gene (associated with resistance to tobacco mosaic virus, TMV) probably important for *N* gene-dependent hypersensitive cell death. The *N* gene is a TIR-NBS-LRR type *R*-gene that generates two alternative transcripts [N(S)]: (i) encoding a full-length *N* protein or (ii) a truncated form of the protein [N(L)]. Both types accumulated at the temperature of 20°C but not at 30°C (no permissive temperature, in TMV inoculated leaves). It was proposed that transient expression of p50 (50 kDa helicase domain) of TMV replicase (considered to be *Avr* component of *N*) was responsible for the observed accumulation. Results suggested that the accumulation of both types [N(S) and N(L)] was associated with the function of *N* and, above a certain threshold, triggered *N*-mediated hypersensitive cell death.

However, the results not always have been conclusive, as revealed by the analysis of four transcript classes of the flax rust resistance gene *L6* (containing NBS-LRR domains in the most abundant transcript) [142]. Transgenic flax plants expressing an intronless *L6* gene

were incapable of encoding truncated resistance proteins by alternative splicing, even though expressed L6 resistance products were indistinguishable from that of the wild-type gene. Thus, a definitive role for alternative transcripts and their predicted truncated products could not be assigned in the flax/flax rust system.

On the other hand, post-transcriptional modifications could constitute an efficient source of protein diversity. In *P. vulgaris*, an *R*-gene cluster (*B4*) was identified at the end of linkage group B4. In this cluster, three *R* specificities (Co-9, Co-y and Co-z) and two R-QTLs were considered effective against the fungus *Colletotrichum lindemuthianum* (that causes anthracnose), whereas also four CC-NBS-LRR *R*-like proteins were characterized. Seven truncated *R*-like transcripts were generated as a result of distinct post-transcriptional modifications of *JA1tr* gene (also mapped at the *B4* *R*-gene cluster), corresponding to alternative splicing events involving two introns, exon skipping and multiple splicing events in the ORF [143].

A most comprehensive work was carried out with the RGA5 from rice [144], known for its interaction with RGA4 during effector recognition of two avirulence factors (Avr-pia and Avr1-CO39). In this study two transcripts were generated by alternative splicing from the third of three introns of RGA5. It was observed that both the fully transcribed intron (RGA5RT) as the alternate version (RGA5AT) shared the regions CC, NBS and LRR, being distinguished only in regard to C-terminal region associated with the binding of the protein ATX1 to copper (RATX1). Transformed plants with RGA5AT were susceptible to AVR-pia and *Magnaporthe oryzae* strains expressing Avr1-CO39. Additionally, RGA5RT was necessary in conjunction with RGA4 to impart dual recognition specificity. Tests with the yeast two-hybrid system have shown that Avr-pia physically interacts with the domain RATX1 in the C-terminal of CO39-Avr1, present only in RGA5RT, indicating that RATX1 turns RGA5AT inactive. The study highlighted the importance of the non-LRR region near the C-terminus with respect to the specificity of recognition of the avr factor secreted by the pathogen.

Even though many efforts have been done to understand the role of alternative splicing in pathogen recognition and initiation of defense signaling by *R*-genes, many questions remain open, indicating the need of additional approaches addressed to this question, especially including identification of putative interactions, their validation and functional analysis [145].

Post-transcriptional control (mi- or si-RNAs) of *R*-genes

To test whether the Virus-Induced Gene Silencing (VIGS) system could be utilized as a rapid way to assess gene function of candidate *R*-genes in potato and its wild relatives, Brigneti and coworkers [146] silenced *R1* and *Rx* in *S. tuberosum* and *RB* in *S. bulbocastanum*. The successful post transcriptional silencing of *R1*, *Rx* and *RB* was able to attenuate resistance mediated by *R*-gene and generated susceptible phenotypes confirmed during detached leaf assays. Therefore, VIGS approach was indicated as an effective assay in potato for rapid evaluation of *R*-gene function.

The HR to tobamoviruses can be compromised if genes of the *N* family (that confers resistance to TMV) were silenced in *Nicotiana edwardsonii*. It was demonstrated that a single hairpin transgene is able to silence a block of *R*-genes related by sequence similarity. Additionally the response of *N*-silenced plants to other viruses was altered, whereas HR induced against TBSV (Tomato bushy stunt virus) and CymRSV (Cymbidium ringspot virus) was also affected, indicating that Tombusvirus *R*-gene shares some homology with the *N* gene [147].

The bra-miR1885 miRNA, that is induced specifically by Turnip Mosaic Virus (TuMV) infection, and target TIR-NBS-LRR class of disease-resistant transcripts for

cleavage was recognized as a true functional miRNA [148] in *Brassica*. Similarities between sequences of bra-miR1885 precursor with the target transcripts suggest that bra-miR1885 may be a new miRNA coding sequence originated through inverted duplication events from TIR-NBS-LRR related sequences, which became bra-miR1885 targets.

Li and coworkers [149] identified two miRNAs [nta-miR6019 (22-nt) and nta-miR6020 (21-nt)] able to benefit transcript cleavage regarding the Toll and Interleukin-1 NB-LRR-immune receptor *N* from tobacco (conferring resistance to TMV). The coexpression of *N* with nta-miR6019 and nta-miR6020 resulted in attenuation of *N*-mediated resistance to TMV, indicating a functional role for these miRNAs in NB-LRR regulation. Using bioinformatic tools the authors also identified six 22-nt miRNA and two 21-nt miRNA families from three Solanaceae species (tobacco, tomato, and potato) [149]. Members of these miRNA families seem to cleave transcripts of predicted functional *R*-genes, triggering production of phased secondary 21-nt siRNAs. The results demonstrated a conserved role for miRNAs and secondary siRNAs in NB-LRR/LRR immune receptor gene regulation and pathogen resistance in Solanaceae.

Therefore, it is possible to anticipate that mechanisms associated to *R*-gene regulation are not well understood. The unregulated expression of NB-LRR genes, for example, can trigger autoimmunity in the absence of pathogen infection and inhibit plant growth, and may cause serious consequences to agricultural production. Host cellular function that is regulated by small, and non-coding miRNAs can be altered by plant defense responses. To understand the mechanism of miRNAs in the regulation of cellular functions during stem rust infection in wheat, Gupta and coworkers [150] investigated eight different miRNAs (miR159, miR164, miR167, miR171, miR408, miR444, miR1129 and miR1138) involved in three independent cellular defense responses. The investigation revealed that at the initiation of disease, accumulation of miRNAs putatively plays a key role in HR, which diminishes at the maturation stage, suggesting a putative synergistic association between host and fungal pathogen, resulting in susceptibility. Expression profiling of these miRNAs indicated that the modulation of their expression level (both in presence and absence of *R*-gene) provide a probable explanation for miRNA regulated *R*-gene mediated independent pathways. Other epigenetic mechanisms as histone methylation can alter the expression of *R* or *PR*- genes, as occurs, for example, in *A. thaliana* *PR1* gene when in contact with *Pseudomonas syringae* [151].

Concluding Remarks

Since more than 20 years plant resistance genes provide important evidences on the mechanisms that trigger pathogen recognition. Such genes have been submitted to evolutionary processes and diversifying selection generating one of the most abundant and polymorphic multigene families. Still, such genes in plants have much in common with those reported in studies on the behavior and evolution of the vertebrate immune response. Understanding the role of the high number of *R*-genes in plants and the epigenetic mechanisms involved in their expression also need special efforts, especially considering differences in their activation in time frame analysis, considering different plant accessions and pathogen peculiarities. Also, we have a long way to go towards the understanding of network interactions associated to the defense triggered by these genes as well as post transcriptional modifications that may affect their activity. So, much is expected to occur considering the now available large scale sequencing platforms, protein-protein interaction approaches and transformation approaches in model plants in association with omics and systems biology approaches. Practical applications for breeding purposes still depend on

marker development and selection of promising plant material for crossing, besides the generation of transgenic plants, especially pyramiding defense related genes.

Abbreviations

AFLP, Amplified Fragment Length Polymorphism
 AGP, arabinogalactan protein
 Avr, avirulence
 BAK, BRI1-associated kinase
 BHT, benzo-(1,2,3)-thiadiazole-7-carbothioic acid S-methyl ester
 BLAST, Basic Local Alignment Search Tool
 BRI, brassinosteroid insensitive (receptor)
 BSA, bulked segregant analysis
 CAPS, Cleaved Amplified Polymorphic Sequence
 CC, coiled coil
 CNV, Copy Number Variation
 CRR, cysteine-rich repeat
 DAMP, damage-associated molecular pattern
Dissect, DIScovery of Structural Alteration Event Containing Transcripts
 DNA, Deoxyribonucleic Acid
 DRAGO, Disease Resistance Analysis and Gene Orthology
 EFR, elongation factor Tu receptor
 EGF, epidermal growth factor
 ETI, effector-triggered immunity
 ETS, effector-triggered susceptibility
 FLS, flagellin(sensing) binding receptor
 GAP, GPI-associated protein
 GlcNac, N-acetyl glucosamine
 GPI, glycosylphosphatidylinositol
 HR, hypersensitive response
 INA, 2,6-dichloro-isonicotinic
 IRAK, interleukin-1 receptor-associated kinase
 JA, jasmonic acid
 KEGG, Kyoto Encyclopedia of Genes and Genomes
 LAR, local induced/acquired resistance
 LPS, lipopolysaccharides
 LRR, leucine-rich repeat
 LRX, leucine-rich extensin
 LysM, lysine motif
 MAPK, mitogen activated protein kinase
 MYA, Million Years Ago
 NBS, nucleotide-binding site
 ORF, Open Reading Frame
 P/MAMP, pathogen microbe-associated molecular pattern
 PCR, Polymerase Chain Reaction
 PERK, proline-rich extensin-like receptor kinase
 PR, pathogenesis related
 PRR, pattern recognition receptor
 PTI, pathogen-triggered immunity

QTL, Quantitative Trait Loci
 R, resistance
 RD, arginine-aspartate
 RFLP, Restriction Fragment Length Polymorphism
 RGA, Resistance Gene Analog
 RLCK, receptor-like cytoplasmic kinase
 RLK, receptor-like kinase
 RLP, receptor-like protein
 RNA, Ribonucleic Acid
 RPK, receptor protein kinase
 RSTK, receptor serine/threonine kinase
 RTK, receptor tyrosine kinase
 SA, salicylic acid
 SAR, Systemic Acquired Resistance
 SI, self-incompatibility
 SIPK, Salicylic Acid-Induced Protein Kinase
 SLG, self-incompatibility locus glycoprotein
 SRK, S receptor kinase
 STS, Sequence Tagged Site
 SSH, Suppression Subtractive Hybridization
 TDF, Transcript-derived Fragments
 TIR, toll/interleukin-1 receptor
 TMRK, Transmembrane receptor kinase
 TMV, Tobacco Mosaic Virus
 TNFR, tumor necrosis factor receptor
 TuMV, Turnip Mosaic Virus
 VIGS, Virus-Induced Gene Silencing
 WAK, wall-associated kinase
 WIPK, Wound-Induced Protein Kinase
 WRKY, trp-arg-lys-tyr

Conflict of Interest

The authors declare that this article content has no conflict of interest.

Acknowledgements:

The authors thank CAPES (Brazilian Coordination of Improvement of Higher Education Personnel), FACEPE (Fundação de Amparo à Pesquisa do Estado de Pernambuco), CAPES (*Coordination for the Improvement of Higher Level Education Personnel*) and CNPq (Brazilian Council of Scientific and Technological Development) for important financial support and fellowships.

References

- [1] Hassen, F.; Massart, S.; Motard, J.; Roussel, S.; Parisi, O.; Kummert, J.; Fakhfakh, H.; Marrakchi, M.; Perreault, J.P.; Jijakli, M.H. Molecular Features of new peach latent mosaic viroid variants suggest that recombination may have contributed to the evolution of this infectious RNA. *Virology*, **2007**, 360, 50-57.

- [2] Ellis, J.; Dodds, P.; Pryor, T. The generation of plant disease resistance gene specificities. *Trends Plant Sci*, **2000**, 5, 373-379.
- [3] Dangl, J.L.; Jones, J.D.G. Plant pathogens and integrated defence responses to infection. *Nature* **2001**, 411, 826-833.
- [4] Benko-Iseppon, A.M.; Galdino, S.L.; Calsa Júnior, T.; Kido, E.A.; Tossi, A.; Belarmino, L.C. Crovella, S. Overview of plant antimicrobial peptides. *Curr Prot Pept Sci*, **2010**, 1, 181-188.
- [5] Panstruga, R.; Parker, J.E.; Schulze-Lefert, P. SnapShot: plant immune response pathways. *Cell*, **2009**, 136, 978.
- [6] Jones, J.D.; Dangl, J.L. The plant immune system. *Nature*, **2006**, 44, 323-329.
- [7] Benschop, J.J.; Mohammed, S.; O' Flaherty. Quantitative phosphoproteomics in early elicitor signaling in *Arabidopsis*. *Molecul. Cel. Prot.*, **2007**, 6, 1198-1214.
- [8] Dangl, J.L.; Horvath, D.M.; Staskawicz, B.J. Pivoting the Plant Immune System from Dissection to Deployment. *Science*, 2013, 6147, 746- 751.
- [9] Fraiture, M.; Brunner, F. Killing two birds with one stone: trans kingdom suppression of PAMP/MAMP-induced immunity by T3E from enteropathogenic bacteria. *Front. Microbiol.*, **2014**, 320, 1-7.
- [10] Macho, A.P.; Zipfel, C. Plant PRRs and the Activation of Innate Immune Signaling. *Mol. Cell*, 2014, 2, 263–272.
- [11] Liu, B.; Li, J.F. Ao, Y.; Qu, J.; Li, Z.; Su, J.; Zhang, Y.; Liu, J.; Feng, D.; Qi, K. Lysin motif-containing proteins LYP4 and LYP6 play dual roles in peptidoglycan and chitin perception in rice innate immunity. *Plant Cell*, **2012**, 24, 3406–3419.
- [12] Newman, M.A.; Sundelin, T.; Nielsen, J.T.; Erbs, G. MAMP (microbe-associated molecular pattern) triggered immunity in plants. *Front. Plant Sci.*, **2013**, 4, 1-14.
- [13] Zipfel, C.; Felix, G. Plants and animals: a different taste for microbes? *Curr. Opin. Plant Biol.*, **2005**, 8, 353–360.
- [14] Shiu, S.H.; Karlowski, W.M.; Pan, R.; Tzeng, Y.H.; Mayer, K.F.; Li, W.H.. Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell*, **2004**. 16, 1220–1234.
- [15] Song, W.Y.; Wang, G.L.; Chen, L.L.; Kim, H.S.; Pi, L.Y.; Holsten, T.; Gardner, J.; Wang, B.; Zhai, W.X.; Zhu, L.H. A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science*, **1995**, 270, 1804–1806.
- [16] Lee, S.W.; Han, S.W.; Sriyanum, M.; Park, C.J.; Seo, Y.S.; Ronald, P.C. A type I-secreted, sulfated peptide triggers XA21-mediated innate immunity. *Science*, **2009**, 326, 850–853.

- [17] Liu, Z.; Crampton, M.; Todd, A.; Kalavacharla, V. Identification of expressed resistance gene-like sequences by data mining in 454-derived transcriptomic sequences of common bean (*Phaseolus vulgaris* L.). *BMC Plant Biol.*, **2012**, *12*, 1:42.
- [18] Boller, T.; Felix, G. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.*, **2009**, *60*, 379–406.
- [19] Henry, G.; Thonart, P.; Ongena, M. PAMPs, MAMPs, DAMPs and others: an update on the diversity of plant immunity elicitors. *Biotechnol. Agron. Soc. Environ.*, **2012**, *2*, 257–268.
- [20] Brutus, A.; Sicilia, F.; Macone, A.; Cervone, F.; DeLorenzo, G. A domains wap approach reveals a role of the plant wall- associated kinase1 (WAK1) as a receptor of oligogalacturonides. *Proc. Natl. Acad. Sci. U.S.A.*, **2010**, *107*:9452–9457.
- [21] Thomma, B.P.H.J.; Nürnberger, T.; Joostena, M.H.A.J. Of PAMPs and Effectors: The Blurred PTI-ETI Dichotomy. *Plant Cell*, **2011**, *231*, 4–15.
- [22] Yamaguchi, Y.; Pearce, G.; Ryan, C.A. The cell surface leucine-rich repeat receptor for AtPep1, an endogenous peptide elicitor in Arabidopsis, is functional in transgenic tobacco cells. *Proc. Natl. Acad. Sci. USA.*, **2006**, *103*, 10104-10109.
- [23] Lotze, M.T.; Zeh, H.J.; Rubartelli, A.; Sparvero, L.J.; Amoscato, A.A.; Washburn, N.R. The grateful dead: damage associated molecular pattern molecules and reduction/oxidation regulates immunity. *Immunol Rev.*, **2007**, *220*: 60–81.
- [24] Hardoim, P.R.; Van Overbeek, L.S.; Elsas, J.D.V. Properties of bacterial endophytes and their proposed role in plant growth. *Trends microbiol.*, **2008**, *10*, 463-471.
- [25] Schulz, B.; Boyle, C. The endophytic continuum. *Myc. Res.*, **2005**, *6*, 661-686.
- [26] Rosenblueth, M.; Martinez-Romero, E. Bacterial endophytes and their interactions with hosts. *Mol. Plant Micr. Interact.*, **2006**, *3*, 827-837.
- [27] Pirofski, L.A.; Casadevall, A.Q.A. What is a pathogen? A question that begs the point. *BMC Biol.*, **2012**, *1*, 6.
- [28] Alfano, J.R.; Collmer, A. Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annu. Rev. Phytopathol*, **2004**, *42*, 385.
- [29] Marie, C.; Broughton, W.J.; Deakin, W.J. *Rhizobium* type III secretion systems: legume charmers or alarmers? *Curr. Opin. Plant Biol.*, **2001**, *4*, 336t
- [30] Spoel, S.H.; Dong, X. How do plants achieve immunity? Defense without specialized immune cells. *Nature Rev. Immun.*, **2012**, *12*,89-100.
- [31] Postel, S.; Kemmerling, B. Plant systems for recognition of pathogen-associated molecular patterns. *Semin. Cell Dev. Biol.*, **2009**, *20*, 1025-1031.

- [32] Bolton, M.D. Primary metabolism and plant defense-fuel for the fire. *Mol Plant Microbe Interact*, **2009**, 22, 487-497.
- [33] Chisholm, S.T.; Coaker, G.; Day, B.; Staskawicz, B.J. Host-Microbe Interactions: Shaping the Evolution of the Plant Immune Response. *Cell*, **2006**, 124, 803–814.
- [34] Miya, A.; Albert, P.; Shinya, T.; Desaki, Y.; Ichimura, K.; Shirasu, K.; Narusaka, Y.; Kawakami, N.; Kaku, H.; Shibuya, N. CerK1, a lysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. *Proc. Natl. Acad. Sci. USA*, **2007**, 104, 19613–19618.
- [35] Durrant, W.E.; Dong, X. Systemic acquired resistance. *Annu. Rev. Plant Pathol.*, **2004**, 42, 185-209.
- [36] Poland, J.A.; Balint-Kurti, P.J.; Wisser, R.J.; Pratt, R.C.; Nelson, R.J. Shades of gray: the world of quantitative disease resistance. *Trends Plant Sci.*, **2009**, 14, 21-29.
- [37] Wisser, R.J.; Sun, Q.; Hulbert, S.H.; Kresovich, S.; Nelson, R.J. Identification and characterization of regions of the rice genome associated with broad-spectrum, quantitative disease resistance. *Genetics*, **2005**, 169, 2277-229.
- [38] Dodds, P.; Rathjen, J. Plant immunity: towards an integrated view of plant–pathogen interactions. *Genetics*, **2010**, 11, 539-548.
- [39] Soto, M.J.; Dominguez-Ferreras, A.; Perez-Mendoza, D.; Sanjuan, J.; Olivares, J. Mutualism versus pathogenesis: the give-and-take in plant–bacteria interactions. *Cel Microbiol.*, **2009**, 11, 381–388.
- [40] Vallad, G.E.; Goodman, R.M. Systemic Acquired Resistance and Induced Systemic Resistance in Conventional Agriculture. *Crop Sci.*, **2004**, 44, 1920–1934.
- [41] Ross, A.F. Systemic Acquired Resistance induced by localized virus infection in plants. *Virology*, **1961**, 14, 340-358.
- [42] Luna, E.; Bruce, T.J.A.; Roberts, M.R.; Flors, V.; Ton, J. Next-Generation Systemic Acquired Resistance. *Plant Physiol.*, **2012**, 158, 844– 853.
- [43] Alvarez, M.E.; Nota, F.; Cambiagno, D.A. Epigenetic control of plant immunity. *Mol. Plant Pathol.*, **2010**, 11, 563–576.
- [44] Kathiria, P.; Sidler, C.; Golubov, A.; Kalischuk, M.; Kawchuk, L.M.; Kovalchuk, I. Tobacco mosaic virus infection results in an increase in recombination frequency and resistance to viral, bacterial, and fungal pathogens in the progeny of infected tobacco plants. *Plant Physiol.*, **2010**, 153, 1859–70.
- [45] Geoghegan, J.L.; Spencer, H.G. Population-epigenetic models of selection. *Theor. Popul. Biol.*, **2012**, 81, 232–42.

- [46] Ausubel, F.M.. Are Innate immune signaling pathways in plants and animals conserved? *Nature Immunol.*, **2005**, 6, 973-979.
- [47] Nürnberger, T.; Brunner, F.; Kemmerling, B.; Piater, L. Innate immunity in plants and animals: striking similarities and obvious differences. *Immunol Rev.*, **2004**, 198,249-266.
- [48] Bernoux, M.; Ve, T.; Williams, S.; Warren, C.; Hatters, D.; Valkov, E.; Zhang, X.; Ellis, J.G.; Kobe, B.; Dodds, P.N. Structural and Functional Analysis of a Plant Resistance Protein TIR Domain Reveals Interfaces for Self-Association, Signaling and Autoregulation. *Cell Host Microbe.*, **2011**, 9, 200–211.
- [49] Tapping, R.I. Innate immune sensing and activation of cell surface Toll-like receptors. *Semin Immunol.*, **2009**, 21,175–184.
- [50] Akita, M.; Valkonen, J.P. A novel gene family in moss (*Physcomitrella patens*) shows sequence homology and a phylogenetic relationship with the TIR-NBS class of plant disease resistance genes. *J. Mol. Evol.*, **2002**, 55, 595-605.
- [51] Liu, J.J.; Ekramoddoullah, A.K. Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (*Pinus monticola* Dougl. ex. D. Don.). *Mol. Genet. Genomics*, **2003**, 5, 432-441.
- [52] Bai, J.; Pennill, L.A.; Ning, J.; Lee, S.W.; Ramalingam, J.; Webb, C.A.; Zhao, B.; Sun, Q.; Nelson, J.C.; Leach, J.E.; Hulbert, S.H. Diversity in nucleotide binding site leucine-rich repeat genes in cereals. *Genome Res.*, **2002**, 12, 1871-1884.
- [53] Wanderley-Nogueira, A.C.; Soares-Cavalcanti, N.M.; Belarmino, L.C.; Kido, E.A. Benko-Iseppon, A.M. Insight on Pathogen Defense Mechanisms in the Sugarcane Transcriptome In: *Special Issue (SI) of Functional Plant Science and Biotechnology on Sugarcane Pathology* ed.UK : Global Science Book., **2011**.
- [54] Tarr, D.E.K.; Alexander, H.M. TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. *BMC Research Notes*, **2009**, 2, 197.
- [55] Ellis, J.; Jones, D. Structure and function of proteins controlling strain-specific pathogen resistance in plants. *Curr. Opin. Plant. Biol.*, **2000**, 1, 288-293.
- [56] Tameling, W.I.; Nooijen, C.; Ludwig, N.; Boter, M.; Sloodweg, E.; Goverse, A.; Shirasu, K.; Joosten, M.H. RanGAP2 mediates nucleocytoplasmic partitioning of the NB-LRR immune receptor Rx in the Solanaceae, thereby dictating Rx function. *Plant Cell*, **2010**, 22, 4176-4194.
- [57] McHale, L.K.; Haun, W.J.; Xu, W.W.; Bhaskar, P.B.; Anderson, J.E.; Hyten, D.L.; Gerhardt, D.J.; Jeddeloh, J.A.; Stupar, R.M. Structural variants in the soybean genome localize to clusters of biotic stressresponse genes. *Plant Physiol.*, **2012**, 159, 1295–1308.

- [58] Hammond-Kosack, K.E.; Jones, J.D.G. Plant disease resistance genes. *Ann. Rev. Plant. Physiol.*, **1997**, *48*, 575-607.
- [59] Tsuda, K.; Katagiri, F. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Curr Opin Plant Biol*, **2010**, *13*, 459-465.
- [60] Sun, X.; Wang, G-L. Genome-Wide Identification, Characterization and Phylogenetic Analysis of the Rice LRR-Kinases. *PLoS ONE*, **2011**, *6*:16079.
- [61] Bella, J.; Hindle, K.L.; McEwan, P.A.; Lovell, S.C. The leucine-rich repeat structure. *Cell Mol Life Sci.*, **2008**, *65*, 2307–2333.
- [62] Banerjee, D.; Zhang, X.; Bent, A.F. The leucine-rich repeat domain can determine effective interaction between RPS2 proteins and other host factors in Arabidopsis RPS2-mediated disease resistance. *Genetics*, **2001**, *158*, 439–450.
- [63] Kobe, K.; Kajava, J. The leucine-rich repeat as a protein recognition motif. *Curr Opin in Struct Biol.*, **2001**, *11*:725.
- [64] Lurderer, R.; Rivas, S.; Nürnberger, T.; Mattei, B.; Van den Hooven, H.W.; Van der Hoorn, R.A.; Romeis, T.; Wehrfritz, J.M.; Blume, B.; Nennstiel, D.; Zuidema, D.; Vervoort, J.; De Lorenzo, G.; Jones, J.D.; De Wit, P.J.; Joosten, M.H. No evidence for binding between resistance gene product Cf-9 of tomato and avirulence gene product Avr9 of *Cladosporium fulvum*. *Mol Plant Microbe Int.*, **2001**, *14*, 867-876.
- [65] Diévar, A.; E. Clark LRR-containing receptors regulating plant development and defense. *Development*, **2004**, *131*, 251-261.
- [66] Ronald, P.C.; Beutler, B. Plant and animal sensors of conserved microbial signatures. *Science*, **2010**, *330*, 1061–1064.
- [67] Norio, M.N.; Miyashita, H. Leucine-Rich Repeat (LRR) Domains Containing Intervening Motifs in Plants *Biomolecules*, **2012**, *2*, 288–311.
- [68] Wanderley-Nogueira, A.C.; Soares-Cavalcanti, N.M.; Bellarmino, L.C.; Bezerra-Neto, J.P.; Benko-Iseppon, A.M. An overall evaluation of the Resistance (R) and Pathogenesis Related (PR) superfamilies in soybean, medicago and arabidopsis. *Gen Mol Biol.*, **2012**, *35*, 260-271.
- [69] Barbosa da Silva, A.; Wanderley-Nogueira, A.C.; Silva, R.R.M.; Belarmino, L.C.; Soares-Cavalcanti, N.M.; Benko-Iseppon, A.M. In silico survey of resistance (R) genes in *Eucalyptus* transcriptome. *Gen Mol Biol.*, **2005**, *28*, 562-574.
- [70] Wanderley-Nogueira, A.C.; Soares-Cavalcanti, N.M.; Morais, D.A.L.; Belarmino, L.C.; Barbosa-Silva, A.; Benko-Iseppon, A.M. Abundance and diversity of resistance genes in the sugarcane transcriptome revealed by in silico analysis. *Genet Mol Res.*, **2007**, *4*, 866-889.

- [71] Salvaudon, L.; Héraudet, V.; Shykoff, J.A. Parasite-host fitness trade-offs change with parasite identity: genotype-specific interactions in a plant-pathogen system. *Evolution*, **2005**, *59*, 2518-2524.
- [72] Jones, D.G. Putting the knowledge of plant disease resistance genes to work. *Curr Opin Plant Biol.*, **2001**, *4*, 281-287.
- [73] Sindhu, A.; Chintamanani, S.; Brandt, A.S.; Zanis, M.; Scofield, S.R.; Johal, G.S. A guardian of grasses: specific origin and conservation of a unique disease-resistance gene in the grass lineage. *Proc Natl Acad Sci USA.*, **2008**, *105*, 1762-1767.
- [74] Reinstädler, A.; Müller, J.; Czembor, J.; Piffanelli, P.; Panstruga, R. Novel induced *mlo* mutant alleles in combination with site- directed mutagenesis reveal functionally important domains in the heptahelical barley Mlo protein. *BMC Plant Biol.*, **2010**, *10*, 31.
- [75] Tang, X.; Xie, M.; Kim, Y.J.; Zhou, J.; Klessing, D.F.; Martin, G.B. Overexpression of *Pto* activates defense responses and confers broad resistance. *Plant Cell*, **1999**, *11*, 15-29.
- [76] Swiderski, M.R.; Birker, D.; Jones, J.D. The TIR domain of TIR-NB- LRR resistance proteins is a signaling domain involved in cell death induction. *Mol Plant Microbe Interact.*, **2009**, *22*, 157-165.
- [77] Liu, B.; Zhang, S.; Zhu, X.; Yang, Q.; Wu, S.; Mei, M.; Mauleon, R.; Leach, J.; Mew, T.; Leung, H. Candidate defense genes as predictors of quantitative blast resistance in rice. *Mol Plant Microbe Int.*, **2004**, *17*, 1146-1152.
- [78] Dangl, J.L.; Jones, J.D.G. Plant pathogens and integrated defence responses to infection. *Nature*, **2001**, *411*, 826-833.
- [79] Dixon, M.S.; Jones, D.A.; Keddie, J.S.; Thomas, C.T.; Harrison, K.; Jones, J.D.G. The tomato *Cf2* disease resistance locus comprises two functional genes encoding leucine rich repeats proteins. *Cell*, **1996**, *84*, 451-459.
- [80] Wang, G.L.; Holsten, T.E.; Song, W.Y.; Wang, H.P.; Ronald, P.C. Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa21* disease resistance locus. *The Plant J.*, **1995**, *7*, 525-533.
- [81] Song, W.Y.; Pi, L.Y.; Wang, G.L.; Gardner, J.; Holsten, T.; Ronald, P.C. Evolution of the rice *Xa21* disease resistance genes family. *Plant Cell*, **1997**, *9*, 1279-1287.
- [82] Zmiénko, A.; Samelack, A.; Koslowiski, T.; Figlerowicz, P. Copy number polymorphism in plant genomes. *Theor Appl Genet.*, **2014**, *127*, 1-18.
- [83] Xu, X.; Liu, X.; Ge, S.; Jensen, J.D.; Hu, F.; Li, X.; Dong, Y.; Gutenkunst, R.N.; Fang, L.; Huang, L.; Li, J.; He, W.; Zhang, G.; Zheng, X.; Zhang, F.; Li, Y.; Yu, C.; Kristiansen, K.; Zhang, X.; Wang, J.; Wright, M.; McCouch, S.; Nielsen, R.; Wang, J.;

- Wang, W. Resequencing 50 accessions of cultivated and wild rice fields markers for identifying agronomically important genes. *Nat Biotechnol.*, **2011**, *30*, 105–111.
- [84] Lu, P.; Han, X.; Qi, J.; Yang, J.; Wijeratne, A.J.; Li, T.; Ma, H. Analysis of Arabidopsis genome-wide variations before and after meiosis and meiotic recombination by resequencing *Landsberg erecta* and all four products of a single meiosis. *Genome Res.*, **2012**, *22*, 508–518.
- [85] Micheltore, R.W.; Marsh, E.; Seely, S.; Landry, B. Transformation of lettuce (*Lactuca sativa*) mediated by *Agrobacterium tumefaciens*. *Plant Cell Rept.*, **1987**, *6*, 439-442.
- [86] Meyers, B.C.; Kaushik, S.; Nandety, R.S. Evolving disease resistance genes. *Curr Opin Plant Biol.*, **2005**, *8*, 129-134.
- [87] Friedman, A.R.; Baker, B.J. The evolution of resistance genes in multiprotein plant resistance systems. *Curr Opin Genet Dev.*, **2007**, *17*:493-499.
- [88] Richly, E.; Kurt, J.; Leister, D. Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. *Biol Evol.*, **2002**, *19*, 76-84.
- [89] Rossi, M.; Araujo, P.G.; Paulet, F.; Garsmeur, O.; Dias, V.M.; Chen, H.; Van Sluys, M.A.; D'Hont, A. Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. *Mol Genet Genomics*, **2003**, *269*, 406-419.
- [90] Pryor, T.; Ellis, J. The genetic complexity of fungal resistance genes in plants. *Adv Plant Pathol.*, **1993**, *10*, 281-305.
- [91] Copley, R.; Letunic, I.; Bork, P. Genome and protein evolution in eukaryotes. *Curr Opin Chem Biol.*, **2001**, *268*, 78-94.
- [92] Kondrashov, F.; Rogozin, I.; Wolf, Y.; Koonin, E. Selection in the evolution of gene duplication. *Genome Biol.*, **2002**, *3*, 1-9.
- [93] Dinesh-Kumar, S.; Whitham, S.; Choim D.; Hehl, R. Transposon tagging of tobacco mosaic virus resistance gene N: its possible role in the TMV-N-mediated signal transduction pathway. *Proc Nat Acad Sci.*, **1995**, *92*, 4175-4180.
- [94] Maisonneuve, B.; Bellec, Y.; Anderson, P.; Micheltore, R.W. Rapid mapping of two genes for resistance to downy mildew from *Lactuca serriola* to existing clusters of resistance genes. *Theor Appl Genet.*, **1994**, *89*, 96-104
- [95] Rayapati, P.J.; Lee, M.; Gregory, J.W.; Wise, R.P. A linkage map of diploid *Avena* based on RFLP loci and a locus conferring resistance to nine isolates of *Puccinia coronata* var. 'avenae'. *Theor Appl Genet.*, **1994**, *89*, 831-837.
- [96] Ellis, J.G.; Lawrence, G.J.; Finnegan, E.J.; Anderson, P.A. Contrasting complexity of two rust resistance loci in flax. *Proc Nat Acad Sci.*, **1995**, *92*, 4185-4188.

- [97] Benko-Iseppon, A.M.; Winter, P.; Huettel, B.; Staginnus, C.; Muehlbauer, F.J.; Kahl, G. Molecular markers closely linked to fusarium resistance genes in chickpea show significant alignments to pathogenesis-related genes located on *Arabidopsis* chromosomes 1 and 5. *Theor Appl Genet.*, **2003**, *107*, 379-386.
- [98] Wanderley-Nogueira, A.C.; Soares-Cavalcanti, N.M.; Belarmino, L.C.; Barbosa-Silva, A.; Kido, E.A.; Pandolfi, V.; Calsa Jr, T.; Benko-Iseppon, A.M. *In Silico* Screening for Pathogenesis Related-2 Gene Candidates in *Vigna unguiculata* Transcriptome. *Lec Notes Comp Sci.*, **2010**, *6160*, 70 – 81.
- [99] Petrov, D.A. Evolution of genome size: new approaches to an old problem. *Trends in genet.*, **2001**, *17*, 23-28.
- [100] Shoemaker, R.C.; Schlueter, J.; Doyle, J.J. Paleopolyploidy and gene duplication in soybean and other legumes. *Curr Opin Plant Biol.*, **2006**, *9*, 104-109.
- [101] Mudge, J.; Cannon, S.B.; Kalo, P.; Oldroyd, G.E.; Roe, B.A.; Town, C.D.; Young, N.D. Highly syntenic regions in the genomes of soybean, *Medicago truncatula* and *Arabidopsis thaliana*. *BMC Plant Biol.*, **2005**, *5*, e15.
- [102] Wanderley-Nogueira, A.C.; Soares-Cavalcanti, N.M.; Bellarmino, L.C.; Bezerra-Neto, J.P.; Benko-Iseppon, A.M. An overall evaluation of the Resistance (R) and Pathogenesis Related (PR) superfamilies in soybean, medicago and arabidopsis. *Gen Mol Biol.*, **2012**, *35*, 260-271.
- [103] Dafny-Yelin, M.; Tzfira, T. Delivery of multiple transgenes to plant cells. *Plant Physiol.*, **2007**, *145*, 1118-1128.
- [104] Benko-Iseppon, A.M.; Soares-Cavalcanti, N.M.; Belarmino, L.C.; Bezerra-Neto, J.P.; Amorim, L.L.B.; Ferreira-Neto, J.R.C.; Pandolfi, V.; Azevedo, H.M.A.; Silva, R.L.O.; Santos, M.G.; Alves, M.V.S.; Kido, E.A. Prospecção de genes de resistência à seca e à salinidade em plantas nativas e cultivadas. Volume Especial do III Simpósio de Mudanças Climáticas e Desertificação no Semiárido Brasileiro. *Revista Brasileira de Geografia Física*, **2012**, *4*, 6.
- [105] Shendure, J.; Porreca, G.J.; Reppasm N.B.; Lin, X.X.; McCutcheon, J.P.; Rosenbaum, A.M.; Wang, M.D.; Zhang, K.; Mitra, R.D.; Church, G.M. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, **2005**, *309*, 1728–1732
- [106] Abby, S.; Daubin, V. Comparative genomics and the evolution of prokaryotes. *Trends Microbiol.*, **2007**, *15*, 135-141.
- [107] Sanseverino, W.; Roma, G.; De Simone, M.; Faino, L.; Melito, S.; Stupka, E.; Frusciante, L.; Ercolano, M.R. PRGdb: a bioinformatics platform for plant resistance gene analysis. *Nucl Acids Res.*, **2010**, *38*, D814-D821.
- [108] Vleeshouwers, V.G.; Finkers, R.; Budding, D.; Visser, M.; Jacobs, M.M.; van Berloo, R.; Pel, M.; Champouret, N.; Bakker, E.; Krennek, P.; Rietman, H.; Huigen, D.; Hoekstra,

- R.; Goverse, A.; Vosman, B.; Jacobsen, E.; Visser, R.G. SolRgene: an online database to explore disease resistance genes in tuber-bearing *Solanum* species. *BMC Plant Biol.*, **2011**, *11*, 116.
- [109] Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J Mol Biol.*, **1990**, *215*, 403-410.
- [110] Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; Thompson, J.D.; Gibson, T.J.; Higgins, D.G. Clustal W and Clustal X version 2.0. *Bioinformatics.*, **2007**, *21*, 2947-2948.
- [111] Asmann, Y.W. A novel bioinformatics pipeline for identification and characterization of fusion transcripts in breast cancer and normal cell lines. *Nucleic Acids Res.*, **2011**, *39*, e100.
- [112] Yorukoglu, D.; Hach, F.; Swanson, L.; Collins, C.C.; Birol, I.; Sahinalp, S.C. Dissect: detection and characterization of novel structural alterations in transcribed sequences. *Bioinformatics*, **2012**, *12*, i179-i187.
- [113] Fawal, N.; Savelli, B.; Dunand, C.; Mathé, C. GECA: a fast tool for gene evolution and conservation analysis in eukaryotic protein families. *Bioinformatics*, **2012**, *10*, 1398-1399.
- [114] Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol.*, **2011**, *28*, 2731-2739.
- [115] Eisen, M.B.; Spellman, P.T.; Brown, P.O.; Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Nat Acad Sci USA.*, **1998**, *95*, 14863–14868.
- [116] Flor, H.H. Current status of gene-for-gene concept. *Ann Rev of Phytopathol.*, **1971**, *9*, 275–296.
- [117] Romeis, T.; Piedras, P.; Zhang, S.; Klessig, D.F.; Hirt, H.; Jones, J.D. Rapid Avr9- and Cf-9 -dependent activation of MAP kinases in tobacco cell cultures and leaves: convergence of resistance gene, elicitor, wound, and salicylate responses. *Plant Cell.*, **1999**, *2*, 273-87
- [118] Zhao, Y.; He, S-H.; Sundin, G.W. The *Erwinia amylovora* avrRpt2F. gene contributes to virulence on pear and AvrRpt2 EA is recognized by *Arabidopsis* RPS2 when expressed in *Pseudomonas syringae*. *MPMI*, **2006**, *19*, 644-654.
- [119] Day, B.; Dahlbeck, D.; Huang, J.; Chisholm, S.T.; Li, D.; Staskawicz, B.J.. Molecular basis for the *RIN4* negative regulation of *RPS2* disease resistance. *Plant Cell*, **2005**, *4*, 1292-305.

- [120] Peil, A.; Flachowsky, H.; Hanke, M-V.; Richter, K.; Rode, J.. Inoculation of *Malus 9 robusta* 5 progeny with a strain breaking resistance to fire blight reveals a minor QTL on LG5. *Acta Horticult.*, **2011**, 896, 357–362.
- [121] Voght, I.; €ohner, T.W.; Richter, K.; Flachowsky, H.; Sundin, G.W.; Wensing, A.; Savory, E.A.; Geider, K.; Day, B.; Hanke, M-V.; Peil, A.. Gene-for-gene relationship in the host–pathogen system *Malus 9 robusta 5–Erwinia amylovora*. *New Phytologist*, **2013**, 197, 1262–1275.
- [122] Vleeshouwers, V.G.; Rietman, H.; Krennek, P.; Champouret, N.; Young, C.; Oh, S.K.; Wang, M.; Bouwmeester, K.; Vosman, B.; Visser, R.G.; Jacobsen, E.; Govers, F.; Kamoun, S.; Van der Vossen, E.A. Effector genomics accelerates discovery and functional profiling of potato disease resistance and *Phytophthora infestans* avirulence genes. *PLoS One.*, **2008**, 8, e2875.
- [123] Ma, L.; Lukasik, E.; Gawehns, F.; Takken, F.L. The use of agroinfiltration for transient expression of plant resistance and fungal effector proteins in *Nicotiana benthamiana* leaves. *Methods Mol Biol.*, **2012**, 835, 61-74. doi: 10.1007/978-1-61779-501-5_4.
- [124] King, S.R.; McLellan, H.; Boevink, P.C.; Armstrong, M.R.; Bukharova, T.; Sukarta, O.; Win, J.; Kamoun, S.; Birch, P.R.; Banfield, M.J. *Phytophthora infestans* RXLR effector PexRD2 interacts with host MAPKKK ϵ to suppress plant immune signaling. *Plant Cell*, **2014**, 3, 1345-1359.
- [125] Sweetman, C.; Wong, D.; Ford, C.; Drew, D. Transcriptome analysis at four developmental stages of grape berry (*Vitis vinifera* cv. Shiraz) provides insights into regulated and coordinated gene expression. *BMC Genomics*, **2012**, 13, 691.
- [126] Wong, C.J.D.; Sweetman, C.; Drew, D.P.; Ford, C.M. VTCdb: a gene co-expression database for the crop species *Vitis vinifera* (grapevine). *BMC Genomics*, **2013**, 14, 882.
- [127] Barrett, T.; Troup, D.B.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M. NCBI GEO:archive for functional genomics data sets—10 years on. *Nucleic Acids Res.*, **2011**, 39 (suppl 1), D1005–D1010.
- [128] Wang, L.; Tang, X.; He, C. The bifunctional effector AvrXccC of *Xanthomonas campestris* pv. *campestris* requires plasma membrane-anchoring for host recognition. *Mol Plant Pathol.*, **2007**, 8, 491-501.
- [129] van der Vossen, E.; Sikkema, A.; Hekkert, B.L.; Gros, J.; Stevens, P.; Muskens, M.; Wouters, D.; Pereira, A.; Stiekema, W.; Allefs, S. An ancient R gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. *Plant J.*, **2003**, 36, 867-82.
- [130] Ballvora, A.; Ercolano, M.R.; Weiss, J.; Meksem, K.; Bormann, C.A.; Oberhagemann, P.; Salamini, F.; Gebhardt, C. The R1 gene for potato resistance to late blight

(*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. *Plant J.*, **2002**, 30, 361-371.

- [131] Champouret, N.; Bouwmeester, K.; Rietman, H.; van der Lee, T.; Maliepaard, C.; Heupink, A.; van de Vondervoort, P.J.; Jacobsen, E.; Visser, R.G.; van der Vossen, E.A.; Govers, F.; Vleeshouwers, V.G. *Phytophthora infestans* isolates lacking class I ipiO variants are virulent on Rpi-blb1 potato. *Mol Plant Microbe Interact.*, **2009**, 22, 1535-1545.
- [132] Tian, Z.D.; Liu, J.; Wang, B.L.; Xie, C.H. Screening and expression analysis of *P. infestans* induced genes in potato leaves with horizontal resistance. *Plant Cell Rep.*, **2006**, 25, 1094-1103.
- [133] Solomon-Blackburn, R.M.; Stewart, H.E.; Bradshaw, J.E. Distinguishing major-gene from field resistance to late blight (*Phytophthora infestans*) of potato (*Solanum tuberosum*) and selecting for high levels of field resistance. *Theor Appl Genet.*, **2007**, 115, 141-9.
- [134] Tan, M.Y.; Hutten, R.C.; Celis, C.; Park, T.H.; Niks, R.E.; Visser, R.G.; van Eck, H.J. The R(Pi-mcd1) locus from *Solanum microdontum* involved in resistance to *Phytophthora infestans*, causing a delay in infection, maps on potato chromosome 4 in a cluster of NBS-LRR genes. *Mol Plant Microbe Interact.*, **2008**, 21, 909-918.
- [135] Glynn, N.C.; Comstock, J.C.; Sood, S.G.; Dang, P.M.; Chaparro, J.X. Isolation of nucleotide binding site-leucine rich repeat and kinase resistance gene analogues from sugarcane (*Saccharum* spp.). *Pest Manag Sci.*, **2008**, 64, 48-56.
- [136] Marathe, R.; Guan, Z.; Anandalakshmi, R.; Zhao, H.; Dinesh-Kumar, S.P. Study of *Arabidopsis thaliana* resistome in response to cucumber mosaic virus infection using whole genome microarray. *Plant Mol Biol.*, **2004**, 55, 501-520.
- [137] Coram, T.E.; Wang, M.; Chen, X. Transcriptome analysis of the wheat-Puccinia striiformis f. sp. tritici interaction. *Mol Plant Pathol.*, **2008**, 9, 157-169.
- [138] Joshi, R.K.; Kar, B.; Nayak, S. Survey and characterization of NBS-LRR (R) genes in *Curcuma longa* transcriptome. *Bioinformation*, **2011**, 6, 360-380.
- [139] Zhang, X.C.; Gassmann, W. RPS4-mediated disease resistance requires the combined presence of RPS4 transcripts with full-length and truncated open reading frames. *Plant Cell.*, **2003**, 15, 2333-2342.
- [140] Zhang, X.C.; Gassmann, W. Alternative splicing and mRNA levels of the disease resistance gene RPS4 are induced during defense responses. *Plant Physiol.*, **2007**, 145, 1577-87.
- [141] Takabatake, R.; Seo, S.; Mitsuhashi, I.; Tsuda, S.; Ohashi, Y. Accumulation of the two transcripts of the N gene, conferring resistance to tobacco mosaic virus, is probably

- important for *N* gene-dependent hypersensitive cell death. *Plant Cell Physiol.*, **2006**, *47*, 254-61.
- [142] Ayliffe, M.A.; Frost, D.V.; Finnegan, E.J.; Lawrence, G.J.; Anderson, P.A.; Ellis, J.G. Analysis of alternative transcripts of the flax L6 rust resistance gene. *Plant J.*, **1999**, *17*, 287-92.
- [143] Ferrier-Cana, E.; Macadré, C.; Sévignac, M.; David, P.; Langin, T.; Geffroy, V. Distinct post-transcriptional modifications result into seven alternative transcripts of the CC-NBS-LRR gene JA1tr of *Phaseolus vulgaris*. *Theor Appl Genet.* , **2005**, *110*, 895-905
- [144] Cesari, S.; Thilliez, G.; Ribot, C.; Chalvon, V.; Michel, C.; Jauneau, A.; Rivas, S.; Alaux, L.; Kanzaki, H.; Okuyama, Y. The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. *Plant Cell* , **2013**, *25*, 1463–1481.
- [145] Kwon, Y.J.; Park, M.J.; Kim, S.G.; Baldwin, I.T.; Park, C.M. Alternative splicing and nonsense- mediated decay of circadian clock genes under environmental stress conditions in Arabidopsis. *BMC Plant Biol.*, **2014**, *14*,136.
- [146] Brigneti, G.; Martín-Hernández, A.M.; Jin, H.; Chen, J.; Baulcombe, D.C.; Baker, B.; Jones, J.D. Virus-induced gene silencing in *Solanum* species. *Plant J.*, **2004**, *39*, 264-72.
- [147] Balaji, B.; Cawly, J.; Angel, C.; Zhang, Z.; Palanichelvam, K.; Cole, A.; Schoelz, J. Silencing of the N family of resistance genes in *Nicotiana edwardsonii* compromises the hypersensitive response to tombusviruses. *Mol Plant Microbe Interact.*, **2007**, *20*, 1262-1270.
- [148] He, X.F.; Fang, Y.Y.; Feng, L.; Guo, H.S. Characterization of conserved and novel microRNAs and their targets, including a TuMV- induced TIR-NBS-LRR class *R* gene-derived novel miRNA in *Brassica*. *FEBS Lett.*, **2008**, *582*, 2445-2452.
- [149] Li, F.; Pignatta, D.; Bendix, C.; Brunkard, J.O.; Cohn, M.M.; Tung, J.; Sun, H.; Kumar, P.; Baker, B. MicroRNA regulation of plant innate immune receptors. *Proc Natl Acad Sci U S A* ., **2012**, *109*, 1790- 1795.
- [150] Gupta, O.P.; Permar, V.; Koundal, V.; Singh, U.D.; Praveen, S. MicroRNA regulated defense responses in *Triticum aestivum* L. during *Puccinia graminis* f.sp. *tritici* infection. *Mol Biol Rep*, **2012**, *39*, 817-24.
- [151] De-La-Peña, C.; Rangel-Cano, A.; Alvarez-Venegas, R. Regulation of disease-responsive genes mediated by epigenetic factors: interaction of Arabidopsis-Pseudomonas. *Mol Plant Pathol.*, **2012** ,*4*, 388-98.

5. CAPÍTULO II

(Artigo a ser submetido à Revista *Genetics and Molecular Biology*)

Identificação e caracterização de genes NBS-LRR em *Vigna unguiculata* (L.) Walp.

Flávia T. de Araújo¹, Mitalle K. da Silva Matos¹, João P. Bezerra-Neto¹, Lidiane L. Barbosa Amorim^{1,2}, Ana M. Benko-Iseppon¹

¹Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Av. Prof. Moraes Rego, 1235, CEP 50.670-420, Recife, PE, Brasil.

²Instituto Federal de Ciência, Tecnologia e Ensino do Piauí-IFPI, Campus Oeiras, Rua projetada s/n, Bairro Uberaba II, CEO 63.500-000, Oeiras, PI, Brasil.

Autor para correspondência:

Ana Maria Benko Iseppon, Universidade Federal de Pernambuco, ¹Departamento de Genética, Centro de Ciências Biológicas, Av. Prof. Moraes Rego 1235, CEP 50.670-420, Recife, PE, Brasil. Telefone: 55-81-2126-7816; Fax: 55-82-2126-8522; Email: ana.benko.iseppon@pq.cnpq.br

Resumo

O feijão-caupi apresenta importância econômica em nível global, sendo, porém, frequentemente acometida por estresse biótico. As plantas apresentam mecanismos que são capazes de reconhecer e responder a uma diversidade de patógenos, através da interação entre os produtos dos genes *R* (*Resistance*) e *Avr* (*Avirulence*), sendo que a maioria desses genes apresentam os domínios NBS-LRR (Sitio de Ligação ao Nucleotídeo, *Nucleotide Binding Site* - Repetições Ricas em Leucina, *Leucine Rich Repeats*). O objetivo desse estudo foi caracterizar genes NBS-LRR em feijão-caupi, desenvolver marcadores RGA (Análogos a genes de Resistência, *Resistance Gene Analogs*) e analisar expressão diferencial de genes *R* em uma interação planta-vírus em diferentes tempos (30 min, 60 min e 16 h). Inicialmente, sequências candidatas para genes NBS-LRR foram obtidas no banco NordEST, utilizando sondas (*seed sequence*), procedendo-se com a sua anotação, tradução e identificação dos domínios conservados por meio de ferramentas *in silico*, identificando-se 57 sequências NBS-LRR completas. Destas, duas foram usadas para desenvolver marcadores do tipo RGA que se mostraram polimórficos em parentais contrastantes de feijão-caupi para o vírus do mosaico severo, segregando na população de mapeamento. Nas análises de RT-qPCR, observou-se uma maior expressão de genes NBS-LRR após 16 horas da inoculação do patógeno. Os resultados obtidos podem ser úteis para o melhoramento do feijão-caupi.

Palavras-chave: *Nucleotide Binding Region*, *Leucine Rich Repeat*, *Receptor Toll-Interleucina*, Defesa Vegetal.

Introdução

O feijão-caupi (*Vigna unguiculata* (L) Walp.) é uma leguminosa de alta importância econômica, rica em proteínas, carboidratos, vitaminas e sais minerais, além de possuir baixa quantidade de gordura e grande quantidade de fibras dietéticas. Assim, representa um alimento básico para as populações de regiões de baixa renda, como no caso da região Nordeste do Brasil (Assunção *et al.*, 2005). Apesar da indiscutível importância agrônômica e nutricional do feijão-caupi, sua produtividade ainda se encontra em níveis médios, sendo afetada por estresses bióticos e abióticos (Oliveira, 2006, Gao *et al.*, 2007).

Os genes de resistência são responsáveis pela interação planta-patógeno, também conhecida como interação gene-a-gene, onde um gene *R* (*Resistance*) da planta reconhece um gene *Avr* (*Avirulence*) do patógeno, desencadeando uma reação de hipersensibilidade (HR - Resposta Hipersensível; *Hypersensitive Response*), impedindo o desenvolvimento da doença (Flor, 1971). Sabe-se que uma única interação, seja ela compatível ou não, é capaz de recrutar ou silenciar centenas de genes envolvidos na defesa, onde muitos deles já são conhecidos, enquanto outros ainda não foram descritos (Benko-Iseppon *et al.*, 2010).

Os genes *R* conferem resistência a bactérias, vírus, fungos, nematóides e são classificados em cinco famílias, com base nos domínios conservados (Van-Ooijen *et al.*, 2007). A maioria dos genes de resistência se caracteriza por apresentar os domínios conservados NBS (Sitio de Ligação ao Nucleotídeo; *Nucleotide Binding Site*) e LRR (Repetições Ricas em Leucina; *Leucine Rich Repeats*), os quais estão associados com atividade de ATP e reconhecimento do patógeno, respectivamente. A família gênica NBS-LRR é dividida em duas

subfamílias, TIR (Receptor Toll- Interleucina; *Toll Interleukine Receptor*) e não-TIR (McHale *et al.*, 2006). A identificação e a clonagem de genes de resistência podem permitir a identificação de materiais promissores, contribuindo significativamente para a melhoria futura do germoplasma de feijão-caupi (Gowda *et al.*, 2002).

O elevado nível de conservação dos domínios entre os genes de resistência pertencentes à família NBS-LRR torna possível o desenho de *primers* para sua amplificação pela reação em cadeia da polimerase (PCR - *Polymerase Chain Reaction*), gerando marcadores conhecidos como análogos a genes de resistência (RGA - *Resistance Genes Analogs*) em plantas, tais como a uva (*Vitis vinifera* L., Donald *et al.*, 2002), trigo (*Triticum vulgare* Vill., Lacock *et al.*, 2003), banana (*Musa acuminata* Colla, Miller *et al.*, 2008) e arroz (*Oryza sativa* L., Ren *et al.*, 2013).

Várias tecnologias foram desenvolvidas com o intuito de acessar a expressão gênica diferencial, tais como EST (Etiquetas de Sequências Expressas, *Expressed Sequence Tag*, Adams *et al.*, 1991), RNAseq (Sequenciamento do RNA; *RNA Sequencing*, Morin *et al.*, 2008) e SuperSAGE (Super Análises em Série da Expressão Gênica; *Super Serial Analysis of Gene Expression*, Matsumura *et al.*, 2004). Tais tecnologias também têm contribuído com a caracterização de genes associados à resistência a doenças desempenhando um papel chave em redes de interação, especialmente em experimentos onde as plantas são submetidas a algum tipo de estresse biótico (Guimarães *et al.*, 2005).

O objetivo deste estudo foi identificar e caracterizar genes *R* da classe NBS-LRR em feijão-caupi, desenvolver marcadores moleculares do tipo RGA e validar a expressão de genes NBS-LRR candidatos para resistência a patógenos.

Material e Métodos

Identificação e caracterização de transcritos em *Vigna unguiculata*

Para a obtenção de genes candidatos da categoria NBS-LRR em feijão-caupi foram utilizadas 35 sondas (*seed sequence*), correspondentes a genes NBS-LRR já descritos em várias espécies (Sanseverino *et al.*, 2010) (Anexo 1). As sondas foram alinhadas através da ferramenta tBLASTn contra o banco de dados de transcritos do feijão-caupi - NordEST (<http://bioinfo03.ibi.unicamp.br/Vigna/>) adotando-se um *e-value* igual ou inferior a e^{-10} . Candidatos redundantes foram removidos, procedendo-se em seguida a uma reclusterização através do programa CodonCODE (<http://www.codoncode.com/aligner>), utilizando parâmetros de acordo com o padrão do programa, a fim de gerar sequências mais completas.

Os candidatos ESTs de feijão-caupi foram submetidos à ferramenta BLASTx (considerando um *e-value* igual ou inferior a e^{-10}) contra o banco de dados do NCBI (<http://www.ncbi.nlm.nih.gov/>) a fim de confirmar sua identidade, considerando os domínios conservados. As sequências obtidas foram traduzidas em proteínas, utilizando-se a ferramenta ORF-Finder (NCBI), sendo a seguir alinhadas através do programa Clustal ômega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). Os domínios conservados foram identificados com o auxílio da ferramenta CD-Search (*Batch*) (NCBI) e a predição dos motivos foi feita usando o software MEME (<http://meme.nbcr.net/meme/cgi-bin/meme.cgi>).

Obtenção do material vegetal e extração do DNA genômico

Foram utilizados os acessos de feijão-caupi BR 14-Mulato e IT85F-2687, os quais foram obtidos no banco de germoplasma da Embrapa Meio Norte CPAMN

(Empresa Brasileira de Pesquisas Agropecuárias, Centro de Pesquisas Agropecuária do Meio-Norte), Teresina/Piauí, juntamente com os 94 genótipos referentes à população de mapeamento gerada a partir do cruzamento dos acessos BR 14-Mulato e IT85F-2687, contrastantes quanto à resistência ao vírus do mosaico severo do caupi (CPSMV, *Cowpea Severe Mosaic Virus*). O DNA genômico foi extraído segundo o método de Weising *et al.*, (2005), sendo suas concentrações determinadas por eletroforese em gel de agarose a 1,2%.

Desenho de *primers* e desenvolvimento de RGA

Pares de *primers* foram desenhados com base no genoma de *P. vulgaris* através da ferramenta Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>), visando à amplificação dos domínios conservados TIR e NBS em feijão-caupi. Esses *primers* foram testados nos dois genótipos (BR14-Mulato, resistente, e IT85-F2687, susceptível ao vírus do mosaico severo - *Cowpea severe mosaic virus*, CPSMV).

Para o desenvolvimento de RGA, foram realizadas reações de PCR, contendo um volume final de 20 µL, seguindo as seguintes condições: 50 ng de DNA, 3,0 µL de MgCl₂ (25 mM), 1,5 µL de dNTPs (10 mM), 1 µL de cada *primer* (10 mM), 1.0 U Taq polimerase, 2,0 µL de solução tampão 10x (500 mM KCl, 100 mM Tris-HCl, pH8.3 e 1,4 mM MgCl₂). As condições de temperatura foram de acordo com o programa: 94°C por 1 minuto, 57°C por 1 minuto, 72° por um minuto, durante 40 ciclos. Os produtos de PCR foram submetidos à eletroforese em gel de agarose a uma concentração de 1,8% e em gel de poliacrilamida com concentração de 6%, usando como padrão de peso molecular DNA *Ladder* com 100 pares de base - pb (fabricado pela Promega). A visualização das bandas no gel de agarose foi feita por

meio de luz ultravioleta, enquanto que a coloração do gel de poliacrilamida foi realizada utilizando nitrato de prata.

Ancoragem dos genes NBS-LRR em cromossomos virtuais de soja

Os candidatos a genes NBS-LRR de *V. unguiculata* foram alinhados contra o pseudogenoma de *Glycine max* no servidor web SoyBase database (<http://bioinfo03.ibi.unicamp.br/soja/>) utilizando a ferramenta BLASTn (*cut-off* de e^{-20}) para identificação da localização das sequências no genoma. Esta abordagem permitiu a identificação de sua distribuição, posição relativa e abundância, inferida a partir de uma representação do cariótipo virtual gerado com auxílio do programa Circos (*Circular Genome Data Visualization*, Visualização Circular de Dados do Genoma).

Análise fenética e filogenética dos genes NBS-LRR

O alinhamento múltiplo de sequências de proteínas foi realizado utilizando o programa ClustalW, inserido no pacote de programas MEGA 6.06 (Tamura *et al.*, 2007), por meio do método de *Neighbor-Joining*, considerando *p-distance* e um *bootstrap* de 2.000 repetições, a fim de construir uma árvore UPGMA (Método não Polarizado de Agrupamentos aos Pares com Médias Aritméticas; *Unweighted Pair Group Method with Arithmetic Means*), para observação da distinção dos genes de resistência ao longo da evolução em feijão-caupi.

A análise filogenética foi realizada utilizando-se sequências ortólogas ao candidato 'contig98' (o qual apresentou os domínios TIR, NBS e LRR completos) comparativamente a sequências de espécies distintas que apresentaram alinhamentos significativos (e^{-10}), as quais foram selecionadas no banco de dados do Phytozome (<http://phytozome.jgi.doe.gov/pz/portal.html>), alinhadas usando o

ClustalX, sendo em seguida submetidas ao programa MEGA 5.05 para geração de uma árvore filogenética pelo método de máxima parcimônia e teste *bootstrap* de 5.000 replicações.

Análise *in silico* da expressão dos genes

Por meio da contagem direta das *reads*, correspondentes aos genes NBS-LRR em feijão-caupi, foi construído um perfil de expressão das sequências obtidas no banco de dados do NordEST, as quais foram geradas a partir da técnica de RNAseq, correspondendo a dados de dois genótipos contrastantes à tolerância à seca (Pingo de ouro - tolerante, Santo inácio - sensível). Essas *reads* foram agrupadas segundo a biblioteca/tecido de expressão, após a normalização. A análise do perfil de expressão desses genes em *V. unguiculata* foi realizada com o auxílio do programa CLUSTER 3.0 (Eisen *et al.*, 1998), gerando um *heatmap* visualizado através do software TreeView (Page, 1996).

Inoculação, extração de RNA, síntese de cDNA e reações de RT-qPCR

Para inoculação do vírus do mosaico severo, foi utilizada a cultivar BR-14 Mulato (resistente), onde foram preparados 60 vasos, contendo como substrato terra preta. Foram semeadas cinco sementes por vaso, resultando em 300 plantas. O experimento foi conduzido em telado anti-afídeo na casa de vegetação do Instituto Agrônomo de Pernambuco (IPA), com fotoperíodo natural. O vírus foi inoculado depois de 21 dias da semeadura. Após 30, 60 minutos e 16 horas da inoculação do patógeno, foi realizada a coleta do material vegetal (tecido foliar), o qual foi congelado em N₂ (Nitrogênio) líquido e armazenado a - 80°C. O RNA total foi extraído utilizando o kit “SV Total RNA Isolation System” (Promega), e a síntese

de cDNA foi feita por meio do kit Improm-II™ Reverse Transcription Systems (Promega), seguindo as recomendações do fabricante.

Para a análise da expressão gênica, foram desenhados pares de *primers* utilizando as sequências de ESTs candidatas a genes NBS-LRR em feijão-caupi, as quais foram alinhadas com as *tags* obtidas a partir da técnica de SuperSAGE em feijão-caupi (sob estresse biótico), essas *tags* possuíam no máximo um *mismatch* (divergência entre bases no alinhamento *tag*-EST), selecionando-se aquelas que apresentaram expressão diferencial (aumento ou diminuição da expressão gênica) mediante realização da comparação de uma biblioteca em relação ao seu controle. Essas bibliotecas foram construídas com base no tempo após a aplicação do estresse biótico (BRC1: controle negativo, BRM3090: mistura de amostra de tecido foliar nos tempos de 30 e 90 minutos após a inoculação do vírus, BRM16: idem, 16 h após inoculação).

As reações de RT-qPCR foram realizadas utilizando o Kit "GoTaq® qPCR Master Mix" (Promega), seguindo as recomendações do fabricante com otimização, utilizando três replicas biológicas e três replicas técnicas. A amplificação foi feita no equipamento *Fluorescent Quantitative Quantification System* (BIOER®) considerando os parâmetros utilizados por Da Silva *et al.* (2013). A quantificação relativa, utilizando como genes de referência o ACT (actina), Beta-Tubulina e o SKIP (dados não publicados) foi analisada pelo software REST (*Relative Expression Software Tool*) versão 2.0.13 (Pfaffl, Horgan e Dempfle, 2002).

Resultados

Organização das sequências, montagem de *contigs* e análise dos motivos

Foram obtidos 3.613 alinhamentos após tBLASTn contra o banco de dados do NordEST, relacionadas à família NBS-LRR de genes *R*. Após a retirada de redundâncias, as sequências tratadas geraram 331 *contigs* e 201 *singletons*, onde apenas 54 *contigs* e três *singletons* apresentaram o domínio conservado NBS completo (Figura 1).

Os resultados obtidos através do BLASTx dos 57 transcritos em feijão-caupi mostraram que essas sequências estão associadas com genes de resistência já descritos na literatura. O tamanho das sequências variou de 1264 a 8542 pb, no que diz respeito aos nucleotídeos, e de 350 a 1344, nas sequências de aminoácidos (Anexo 2).

A fim de se observar as estruturas presentes nos genes NBS-LRR, as sequências foram traduzidas em aminoácidos para identificação dos motivos que caracterizam a região NBS em plantas. A partir do alinhamento múltiplo gerado com o domínio NBS dos 57 candidatos, foi possível observar que a maioria dessas proteínas preditas é altamente conservada em *V. unguiculata*. Analisando o alinhamento presente na figura 1, foi possível identificar seis motivos conservados: *p-loop*, NBS-II, *Kinase-2*, *Kinase-3*, NBS-V e GLPL, sendo o *p-loop*, *kinase-2*, *kinase-3* e o GLPL os mais conservados, quanto ao tipo e à ordem dos aminoácidos presentes, ao contrário dos motivos NBS-II e NBS-V, que apresentaram um maior nível de variação nesse estudo.

Com base no motivo *Kinase-2* as sequências foram divididas em duas subfamílias, onde 25 candidatos foram agrupados na subfamília não-TIR-NBS-

LRR, por apresentarem o resíduo triptofano (W) conservado na última posição. Os outros 32 membros pertenciam à subfamília TIR-NBS-LRR, caracterizada pela presença do ácido aspártico (D) na última posição, podendo apresentar variações, tais como a presença de uma asparagina (N), serina (S), alanina (A), treonina (T) ou leucina (L) (Figura 1). Foi observado nessas regiões conservadas, dentro do motivo *Kinase-2*, que em algumas posições o aminoácido leucina (L) foi alterado por outro aminoácido, como isoleucina (I) ou valina (V).

O motivo *p-loop* foi caracterizado pela sequência de aminoácidos GGLGKTTLA (glicina - glicina - leucina - glicina - lisina - tronina - tronina - leucina - alanina) (Figura 1), estando estes aminoácidos presentes em 33 sequências analisadas, enquanto as demais apresentaram algumas mutações pontuais. No motivo NBS-II o aminoácido fenilalanina (F) foi observado na posição 321 na maioria das sequências, com exceção de dez candidatos.

Na figura 1 foi notada nos membros não-TIR uma maior conservação dos aminoácidos alanina e triptofano nas posições 325 e 326, enquanto que nas sequências TIR os aminoácidos cisteína e fenilalanina respectivamente foram mais conservados nessas localizações. Todos os candidatos não-TIR apresentaram uma serina conservada na posição 331. Ainda, quanto a esse motivo, observou-se que na subfamília TIR existe um padrão CFL (cisteína - fenilalanina - leucina) de aminoácidos conservados em 12 sequências. Por sua vez, o motivo *Kinase-3* se caracterizou pela sequência de aminoácidos GSRILVTTR (glicina - serina - arginina - isoleucina - leucina - valina - treonina - treonina - arginina), estando presente em 56 candidatos, dentre os quais 22 sequências apresentaram mutações pontuais.

O motivo NBS-V foi caracterizado pela presença de resíduos pontuais conservados (Figura 1). Na posição 457 foi possível observar o aminoácido tirosina

(Y) conservado em 34 sequências analisadas, sendo que as demais sequências apresentaram variações nessa posição, onde podem estar presentes os aminoácidos histidina (H), fenilalanina (F), leucina (L) ou cisteína (C). Nas posições 470, 471, 475 e 476 foi identificada uma maior conservação dos resíduos L, F, A e F, respectivamente, existindo substituições por uma valina ou isoleucina.

O motivo GLPL foi localizado em 28 sequências, caracterizado pela sequência de aminoácidos: GLPL (glicina - leucina - prolina - leucina), observando-se que as demais sequências apresentaram alterações em alguns aminoácidos (Figura 1).

O domínio LRR foi identificado através do alinhamento realizado no Clustal ômega, embora o mesmo não tenha sido detectado em todos os genes candidatos (Figura 2). Analisando as sequências conservadas de aminoácidos, foram observados dois padrões mais frequentes de repetição de leucinas: LXXLXXL (LRR_1) e LXXLXL (LRR_2), onde o L corresponde à presença do aminoácido leucina (observando-se nessa posição uma isoleucina, valina ou metionina) enquanto o X representa qualquer outro aminoácido. Por sua vez, o domínio CC (*coiled-coil*) não foi identificado através dos programas de predição e também não foi possível fazer a identificação da sequência conservada, característica desse domínio.

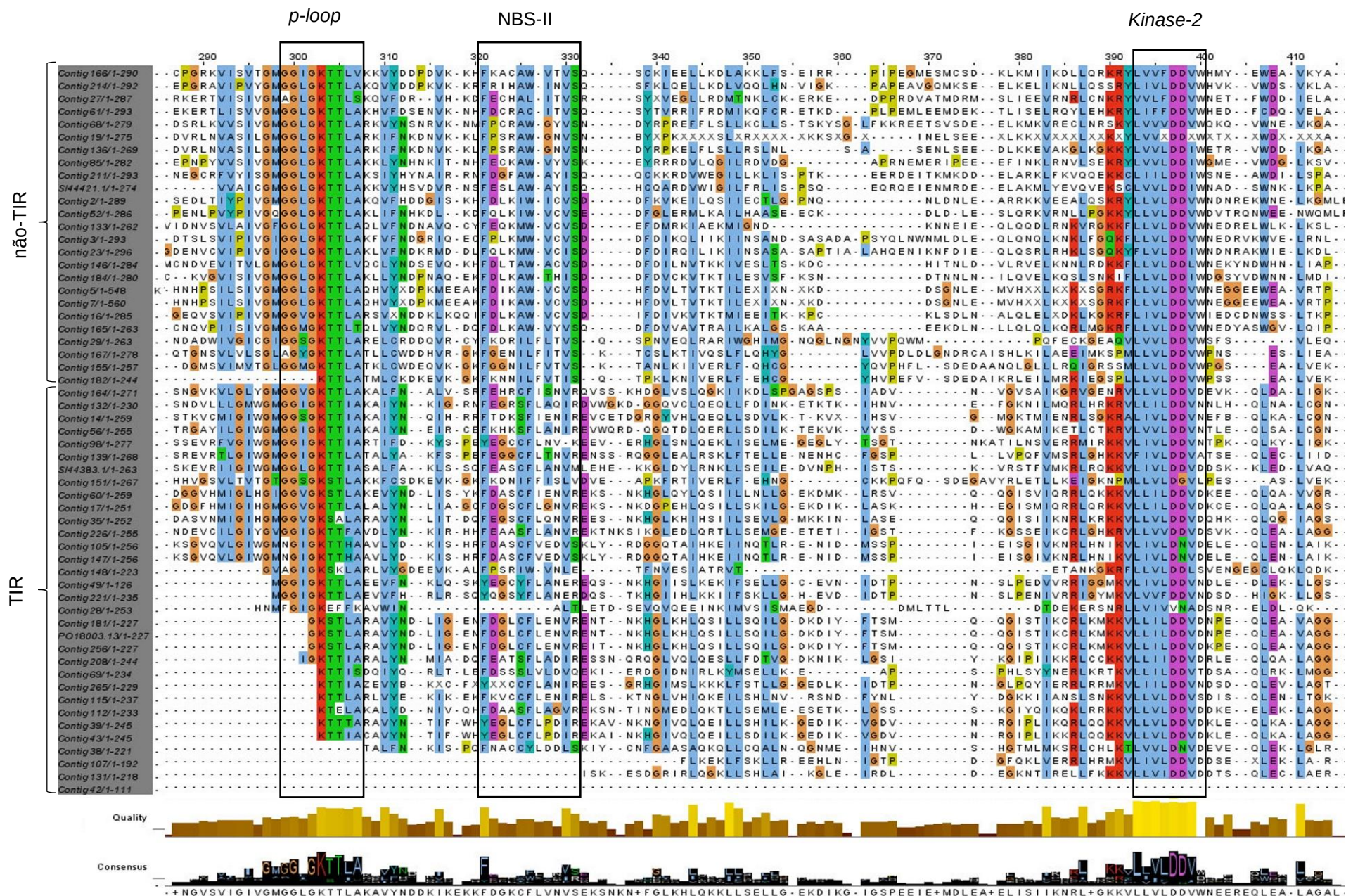


Figura 1: Alinhamento do domínio NBS presente nas 57 seqüências candidatas em feijão-caupi, com destaque para seus motivos conservados *p-loop*, NBS-II, *Kinase-2*, *Kinase-3*, NBS-V e GLPL evidenciados pelos retângulos pretos. Na base da figura encontram-se os gráficos da qualidade do alinhamento e o consenso cujas barras indicam níveis de conservação.

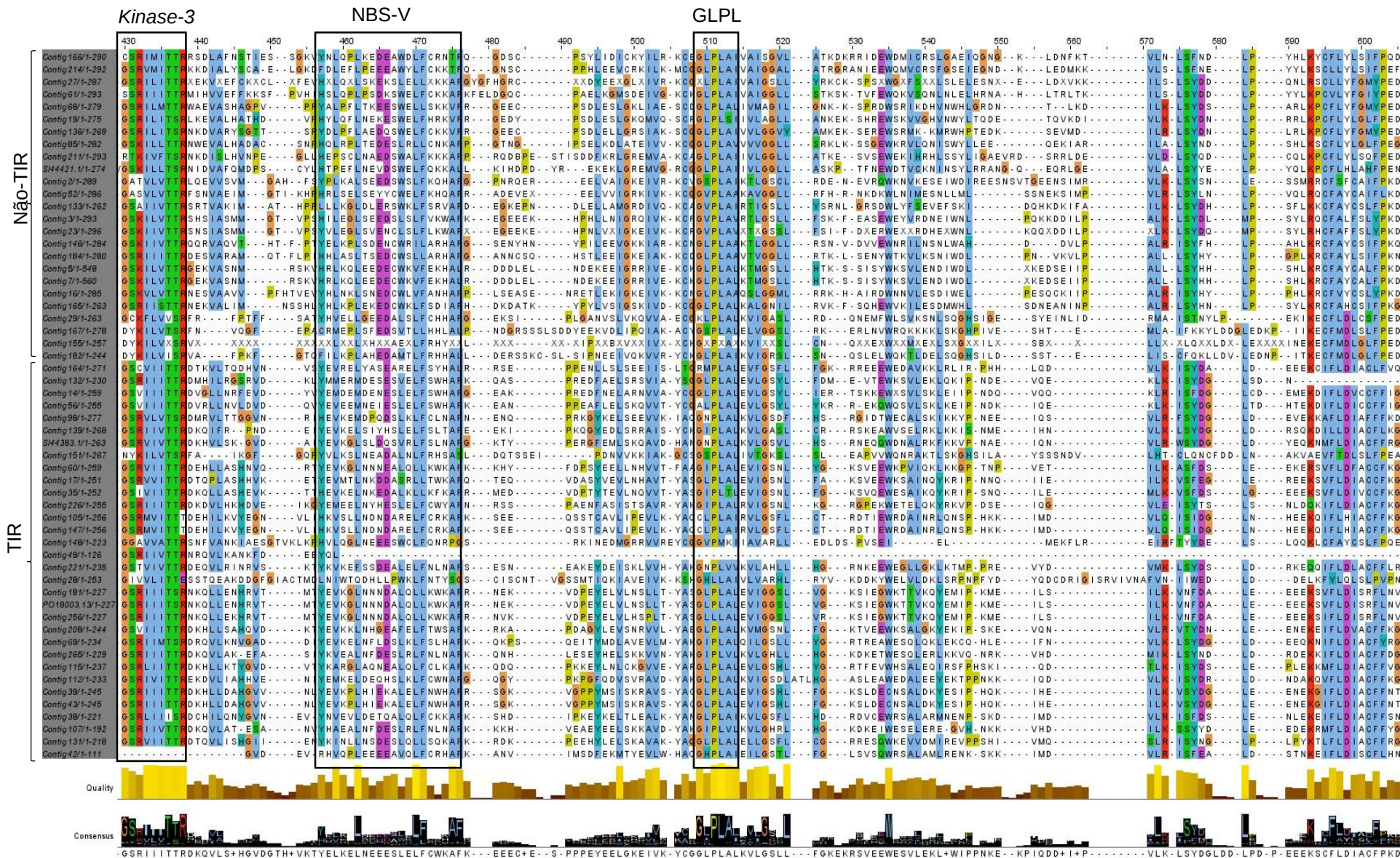


Figura 1. Continuação.

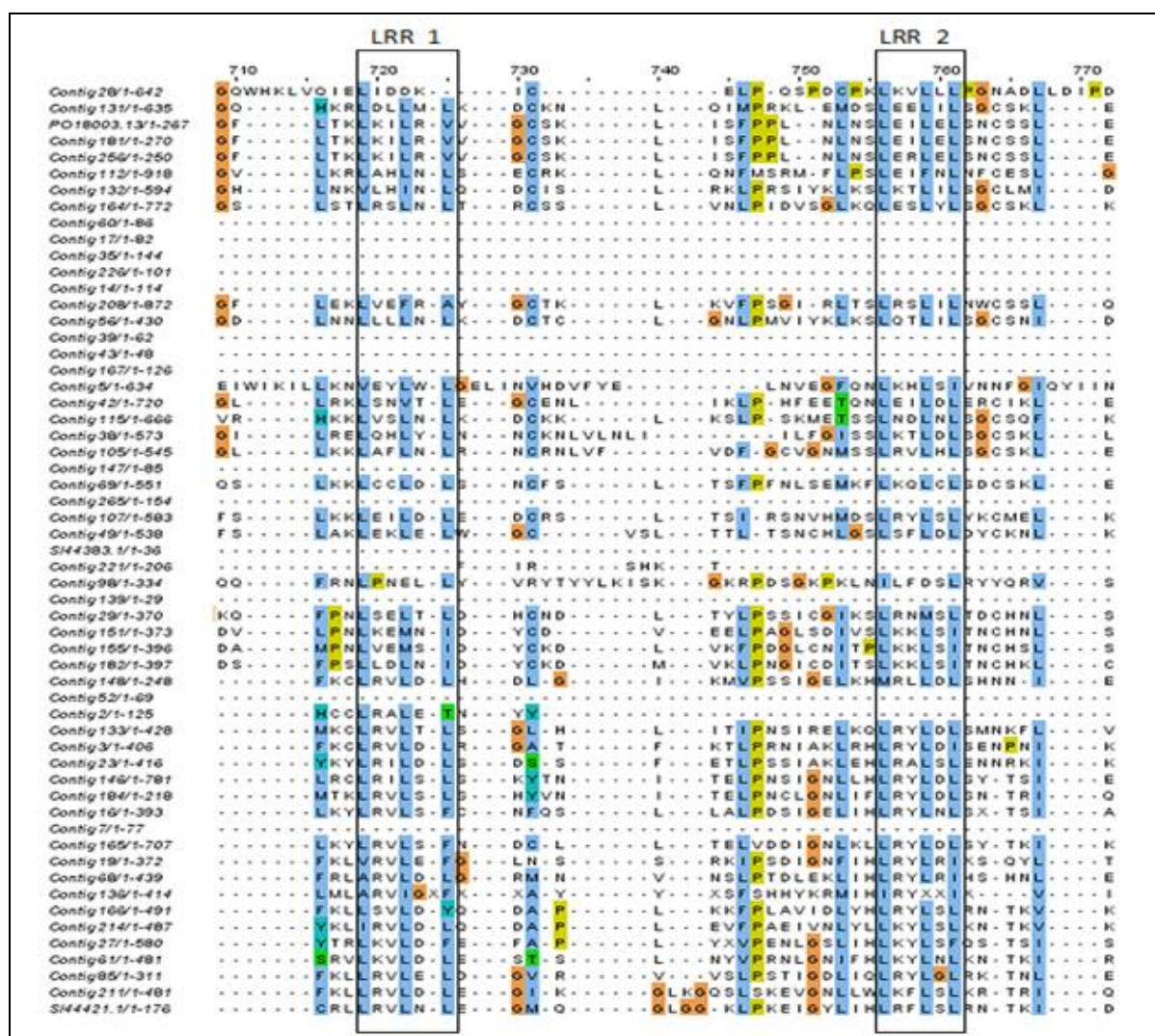


Figura 2: Alinhamento do domínio LRR presente nos genes candidatos da categoria NBS-LRR em feijão-caupi, caracterizados pela conservação dos resíduos de leucina (em azul).

Desenvolvimento de análogos a genes de resistência (RGAs)

Foram desenhados 16 pares de *primers* para os candidatos a genes NBS-LRR em feijão-caupi, com base no genoma de *P. vulgaris*. Destes, sete foram desenhados para o domínio TIR e nove para o NBS (Tabela 1). O tamanho dos *amplicons* foi dentro do esperado, variando de 403 a 500 pb (dados não mostrados).

Esses pares de *primers* foram inicialmente submetidos a PCR de gradiente, a fim de otimizar as condições das reações, especificamente quanto à temperatura de anelamento. Em seguida os mesmos foram testados nos parentais de feijão-caupi contrastantes a virose (BR14-Mulato X IT85F), onde 14 amplificaram, demonstrando um percentual de 87,5% de transferibilidade, passando a ser chamados de RGAs. Apenas dois iniciadores não amplificaram, os quais foram desenhados para o domínio NBS (PvNBS42; PvNBS107.1). Dentre os 14 RGAs analisados em feijão-caupi, dez apresentaram especificidade (fragmento único), enquanto que nos demais foi possível observar a presença de mais de um fragmento. Dentre os RGAs específicos, dois (PvTIR107; PvNBS51.1) apresentaram polimorfismo entre os genótipos analisados, os quais foram aplicados na população de mapeamento do feijão-caupi, composta por 96 indivíduos, onde foi observado a ocorrência de segregação mendeliana (dados não mostrados).

Ancoragem de sequências de *Vigna unguiculata* no pseudogenoma de *Glycine max*

O alinhamento dos 57 genes de feijão-caupi contra os cromossomos virtuais de soja revelou 16.271 alinhamentos distribuídos em todos os vinte cromossomos (Figura 3). Devido à abundância de dados, foi considerado um *e-value* igual a zero ($e^{-0.0}$), para realizar uma melhor análise da distribuição dessas sequências ao longo do pseudogenoma da soja. Pode-se observar que 46 genes candidatos apresentaram semelhanças com segmentos distintos no mesmo cromossomo ou apareceram duas vezes em diferentes cromossomos. Onze sequências candidatas ancoraram apenas em um *locus* exclusivo. Foi identificado um total de 258

alinhamentos nos 20 pseudocromossomos de soja, onde apenas os cromossomos 3, 16 e 18 apresentaram o maior número de sequências ancoradas, correspondentes a 130 *loci* do quantitativo total. Por outro lado, o cromossomo 10 apresentou o menor número de genes ancorados, compreendendo apenas dois *loci*. O cromossomo 3 apresentou o maior número de duplicações, combinando 64 *loci*, sendo a maioria deles organizados em tandem. Quanto à posição relativa das sequências nos cromossomos, as regiões subteloméricas dos cromossomos de *G. max* foram as que apresentaram uma maior frequência de *loci* correspondentes a genes NBS-LRR.

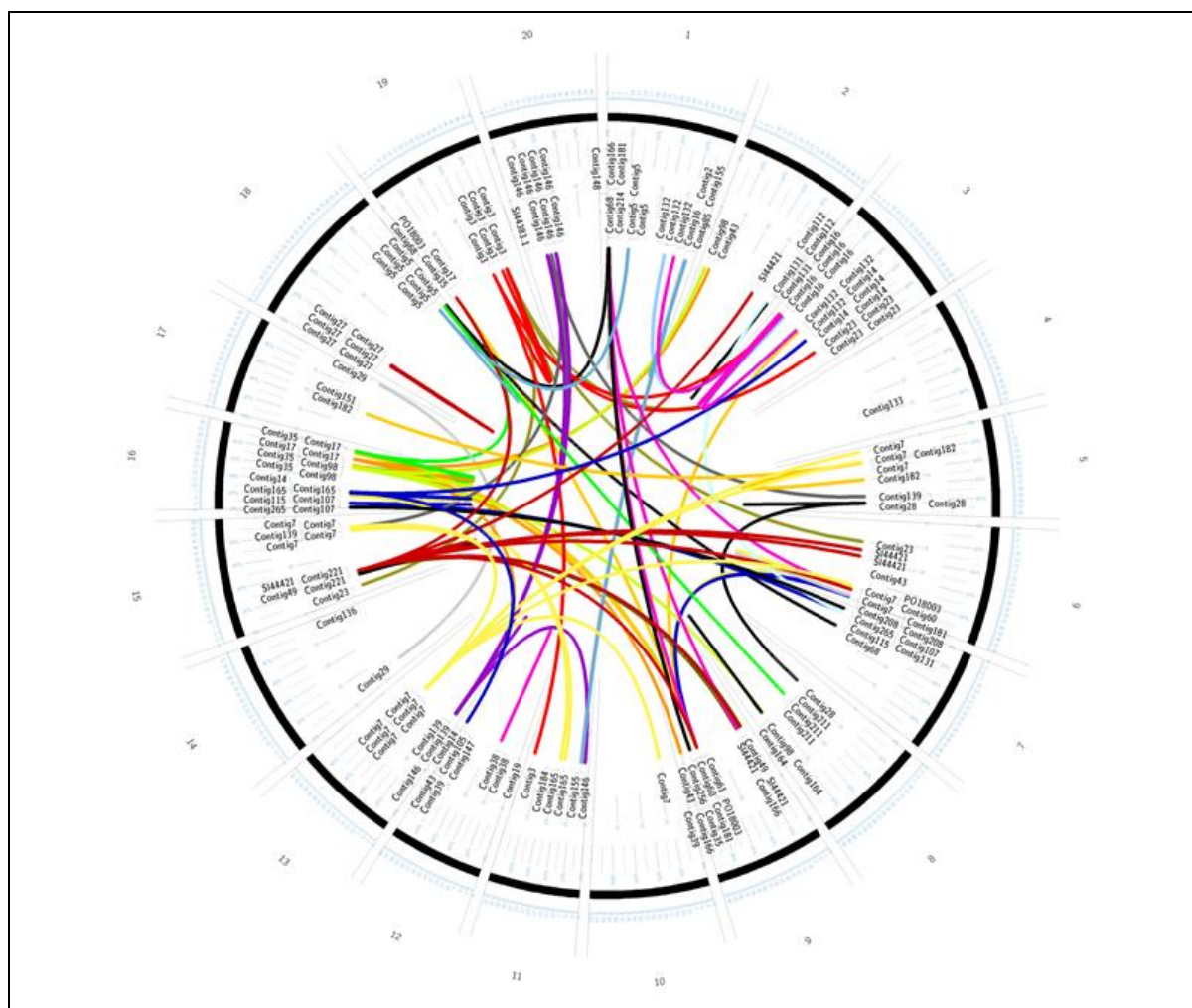


Figura 3: Distribuição e relação da ancoragem de genes, considerando candidatos a genes NBS-LRR identificados em feijão-caupi contra o genoma da soja ($2n = 40$). Os cromossomos são representados por faixas pretas e as relações de genes ancorados são mostradas como links entre as regiões cromossômicas (linhas coloridas que interligam os cromossomos).

Tabela 1: Primers desenhados para identificação de genes NBS-LRR em feijão-caupi, com base no genoma do feijoeiro (*Phaseolus vulgaris*).

Primer	Sequência (5' - 3')	Sequência (3' - 5')
PvTIR39	TCTCTTCCTCCAGGAATCCAT	TGTGCCTTCTCATTGTCAGC
PvTIR42	ATGGCTTCTAACGCCATCAT	GATTGGCGACTTGTGTGAGA
PvTIR51.1	AGGCTGCATCCACCTATGAT	CCGTCTTCGAAATGATAGCC
PvTIR51.2	ATGCTTCTTCCACCTGTTGC	CTTTCTGGAGACCCAATCCA
	TGCTTCTTCAAGATGGTGTTC	AAGTTCAGCATCATCTCTGTG
PvTIR107	G	C
PvTIR115	AGGATGTTTCATGGTCCAACC	GGCATCTCTCCATCTTTCCA
PvTIRNBS3		
9	GCTGACAATGAGAAGGCACA	TGACATCAGACCCATCTCCA
PvNBS39	CGATTGCTCGTGCTGTGTAT	GAAACTGCACGGTTGGAAAT
PvNBS42	ACTTCCGAAGGATGAACTGG	GACGACTCTGCTCCCTTCAC
PvNBS51.1	CTCTTCGGCTGCTTAAATGG	TCATGTCCCGTAGCAAGTCA
PvNBS51.2	AGCTGATTGGTTTGGTCCTG	CTTCAACCTCTGCCAATTCA
PvNBS85	ACAAAGCATTTCTGGAACG	ATGGTGGATGCGAATTGTTT
PvNBS107.1	TGCAAGCAGAGTACCAGGAA	CATAAACGCCATCAGGATCA
PvNBS107.2	GGCATGGGTGGTATTGGTAA	GGAATGCCTTTGGCATAATC
PvNBS112	TTTCTTGGACATAGCTTGCTTC	CTTCTCGCTGAGGAGGATCA
PvNBS115	TTTGTGCAAAGGAGTGTTG	CCTGGGTCAATTTGGAGATTC

Análise Fenética e Filogenética

A árvore fenética foi construída a partir de 50 sequências candidatas (sendo os sete candidatos retirados para melhorar o alinhamento), sendo 25 TIR-NBS-LRR e 25 não-TIR-NBS-LRR (Figura 4). Pôde-se observar que essas foram divididas primeiramente em dois grupos fenéticos (numerados em algarismo romano **I** e **II**, sendo que no primeiro (**I**) agruparam-se todas as sequências pertencentes à subfamília TIR. Esse grupo incluiu dois subgrupos, estando a maioria das sequências presente no primeiro subgrupo (cuja raiz encontra-se circulada em vermelho, com *bootstrap*=95), enquanto o segundo (circulado em azul, *bootstrap*=99) abrigou apenas oito membros TIR. As sequências do grupo fenético **II** correspondem à subfamília não-TIR, onde houve a divisão em dois subgrupos, cada um com 100% de nível de confiabilidade (circulado em verde, *bootstrap*=100). No primeiro subgrupo foram agrupadas quatro sequências, apresentando alto nível

de similaridade. Enquanto que no segundo subgrupo ocorreu o alinhamento das outras sequências com similaridade menor entre si. Este alinhamento justifica-se pelas semelhanças estruturais presentes nos motivos NBS-II e Kinase-2 do domínio NBS dessas sequências.

Para análise evolutiva de genes NBS-LRR foi utilizado um representante de feijão-caupi (contig98), comparativamente a 13 sequências (Figura 5) já caracterizadas e disponíveis na literatura em diferentes espécies. A relação entre esses ortólogos foi refletida em grupamentos de acordo com a taxonomia clássica. Analisando a distribuição evolutiva das sequências foi possível encontrar um ortólogo ao gene candidato 'contig98' na briófita *Physcomitrella patens* L., a qual funcionou como grupo externo. Podem-se observar membros da subclasse Rosidae separados em dois clados distintos, um englobando a ordem Fabales, onde se agruparam quatro leguminosas de interesse econômico: feijão-caupi, feijão comum, alfafa e soja. O outro clado apresentou comportamento merofilético polifilético, unindo mamona, linho, mandioca, *Populus* e mamão, todos membros da ordem Malpighiales, com exceção do mamão, que é uma Brassicaceae, membro de outra subclasse. O clado que mostra a subclasse Asteridae está representado por três membros da família Brassicaceae (*Arabidopsis thaliana*, *Capsella rubella* e *Brassica rapa*), apresentando-se aqui como um grupo merofilético parafilético.

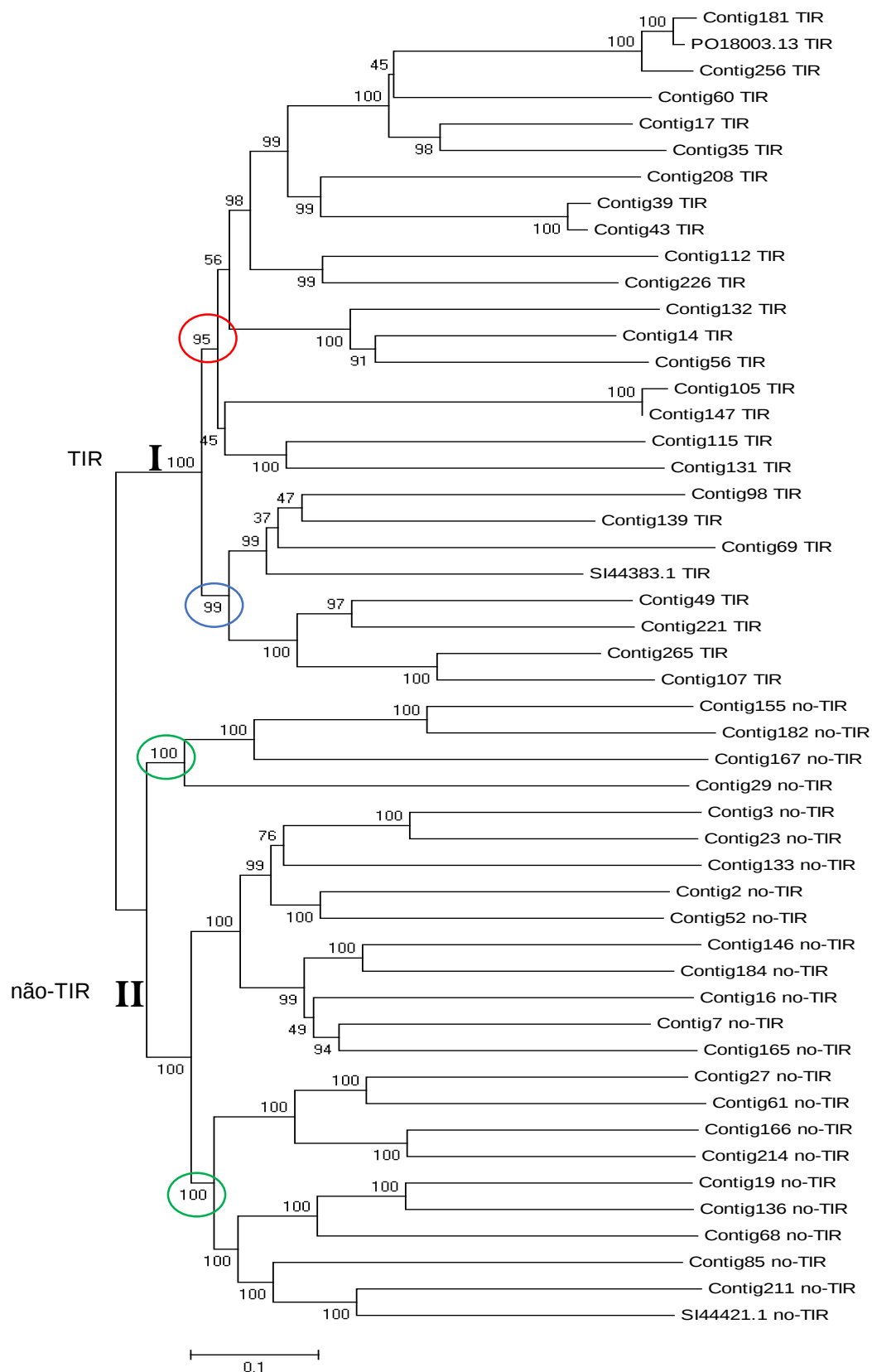


Figura 4: Árvore fenética construída usando o método de *Neighbor-Joining* para inferência sobre a diversidade de candidatos a genes NBS-LRR em feijão-caupi. Números na base dos ramos referem-se aos valores de *bootstrap* (2.000 repetições). O grupo I contém os candidatos da subfamília TIR, enquanto o grupo II inclui os candidatos da subfamília não-TIR.

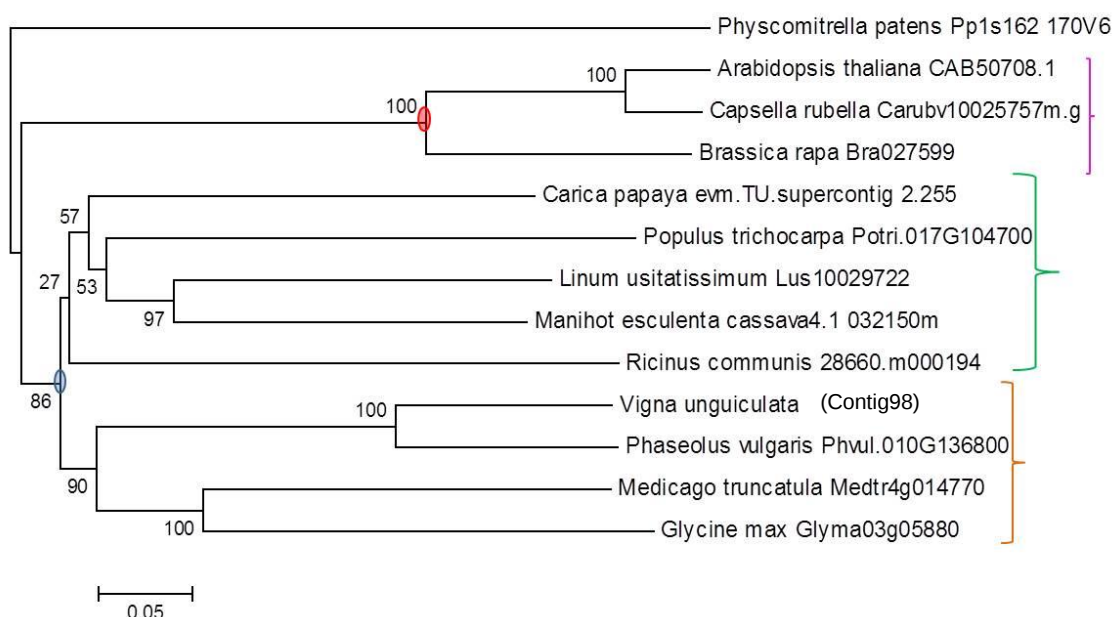


Figura 5: Dendrograma gerado pela análise de máxima parcimônia mostrando as relações entre ortólogos ao candidato 'contig98'. Os números na base dos ramos referem-se ao nível de confiabilidade do alinhamento, ou seja, aos valores de *bootstrap*. O círculo vermelho representa a subclasse Asteridae e o círculo azul a subclasse Rosidae. Chaves ao lado direito do dendrograma indicam da base para o topo: (1) a ordem Fabales (em laranja), (2) ordem Malpighiales (em verde) e (3) a subclasse Asteridae (em roxo).

Análise de expressão gênica *in silico*

Utilizando os dados de expressão obtidos para os dois órgãos (tecidos foliares e de raízes), a partir de dados de RNAseq, um perfil de expressão por clusterização hierárquica foi construído baseado na abundância de transcritos de feijão-caupi (Figura 6). Em geral os candidatos a genes NBS-LRR apresentaram expressão em ambos os órgãos vegetais, demonstrando uma expressão diferencial (aumentando ou reprimindo a sua expressão) de acordo com o tecido analisado, mediante condições normais e sob estresses (dados não mostrados). Para essa análise foram considerados dois genótipos de feijão-caupi: Pingo de Ouro (tolerante à seca) e Santo Inácio (sensível à seca). Quanto à expressão gênica observada no

genótipo tolerante, notou-se a existência de um nível mais elevado da expressão dos genes NBS-LRR no tecido radicular, onde vários genes se comportaram de forma similar. Porém alguns candidatos, tais como contig146_PO32777.1 e contig148_38337.1 (Figura 6.A - caixa vermelha), apresentavam uma menor expressão e à medida em que o tempo de déficit hídrico aumentava (RP75 e RP150 - estresse hídrico, com 75 e 150 minutos, respectivamente), esses genes eram induzidos a aumentar a sua expressão para responder ao estresse. Por sua vez, em outros candidatos, foi possível observar um processo inversamente proporcional, onde o aumento do período de seca, resultava na diminuição de expressão (p.ex. contig133_PO23292.1; contig23_PO43183.1) (Figura 6.A - caixas verdes). Observou-se também que os genes presentes no tecido foliar tendem a apresentar uma expressão gênica mais elevada, ao longo do período de estresse hídrico, quando comparados ao tecido radicular (Figura 6.A).

Ao se analisar o genótipo sensível à seca (Santo Inácio) pode-se observar uma menor expressão na biblioteca FI0 (controle do tecido foliar no tempo zero; única biblioteca de tecido foliar disponível para este acesso). Quanto ao tecido radicular, observou-se um comportamento semelhante ao genótipo tolerante, porém ocorreu uma redução significativa na expressão de genes NBS-LRR.

Comparando as bibliotecas RI0 e FI0 (controle tecido radicular tempo zero; controle tecido foliar tempo zero, respectivamente), nota-se que existe uma maior expressão desses genes no tecido radicular. Também é possível observar que alguns candidatos, como os transcritos Contig23_SI34678.6 e Contig139_SI33999.1 (Figura 6.B - caixa roxa) apresentaram expressão diferencial em tecidos/órgãos específicos, sugerindo uma expressão tecido-específica.

Pingo de Ouro

Santo Inácio

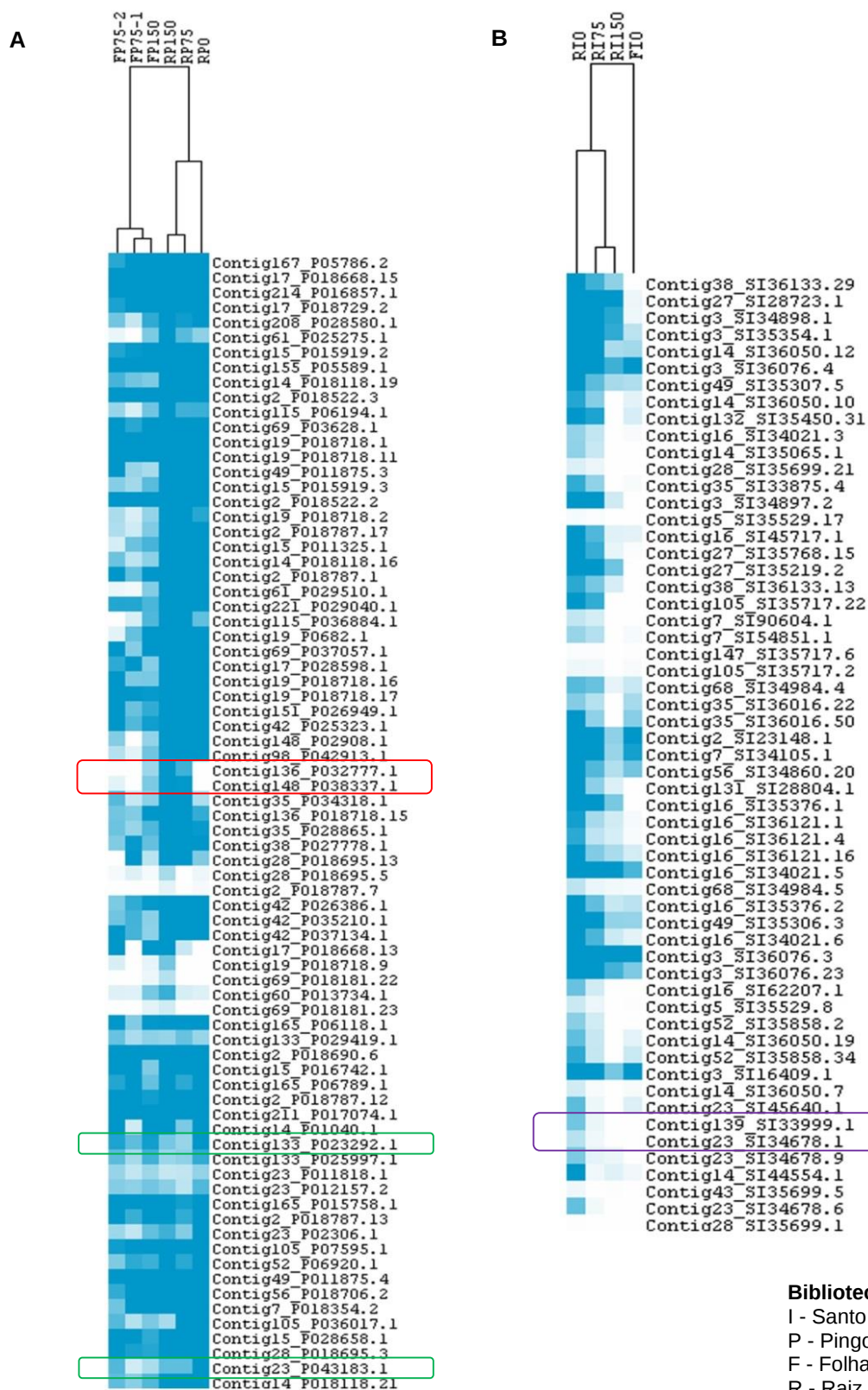


Figura 6: Heatmap construindo a partir da contagem de reads dos genes NBS-LRR em dois genótipos de feijão-caupi contrastantes a seca, para análise de expressão *in silico*. A. Corresponde a expressão do genótipo tolerante a seca (Pingo de Ouro). B. Expressão dos genes no genótipo sensível (Santo Inácio). Na visualização gráfica, a cor branca indica a ausência de expressão e os níveis de azul, a intensidade de expressão na biblioteca em análise, com o intuito de se identificar os principais sítios de expressão dos genes NBS-LRR em feijão-caupi. Quanto maior a intensidade da cor azul, maior o nível de expressão do gene analisado.

Validação de alvos revelados pela expressão gênica de SuperSAGE via RT-qPCR

Com base nas *tags* selecionadas a partir de bibliotecas de SuperSAGE com o genótipo BR14-Mulato infectado com o CPSMV, foram desenhados dez pares de *primers* para genes de resistência pertencentes a família NBS-LRR em feijão-caupi (Tabela 2). Inicialmente foi realizada uma triagem para definir a concentração ideal dos *primers* (0,2 a 0,5 μ M), considerando-se o ciclo inicial de detecção de amplificação. Seis genes de referência foram analisados quanto à sua eficiência de amplificação em triplicatas técnicas e bulks de cDNA das amostras inoculadas com o CPSMV e seus respectivos controles, considerando o cDNA concentrado e três diluições seriadas (1:10, 1:100, 1:1.000). Em seguida foi realizado o teste de estabilidade, para obtenção da melhor combinação, sendo identificados os genes ACT + Beta-tubulina + SKIP, como mais estáveis. Esses genes de referência foram utilizados por já existir relatados demonstrando sua eficiência em tecido de feijão-caupi, submetido ao estresse abiótico (dados não mostrados).

Dos dez pares de *primers* desenhados, quatro foram inespecíficos (ou seja, quando se utilizou o cDNA de feijão-caupi, sob estresse do CPSMV, observou-se mais de uma banda no gel,) sendo os demais seis *primers* testados em reações de RT-qPCR para observação da sua eficiência. Dos sete, quatro foram eficientes [VuRGAq3 (*tag* CMV_77013), VuRGAq4 (*tag* CMV_95506), VuRGAq6 (*tag* CMV_99622) e VuRGAq10 (*tag* CMV_34474), os quais foram avaliados na quantificação relativa a partir de cDNAs do genótipo tolerante ao CPSMV (BR14-Mulato), sob condições normais e sob estresse, usando-se como genes de referência ACT, Beta-tubulina e SKIP.

Tabela 2: Dez pares de *primers* desenvolvidos com base nas *tags* selecionadas, para análise da expressão gênica de genes NBS-LRR em feijão-caupi via RT-qPCR.

Primer	Sequência (5' - 3')	Sequência (3' - 5')	Tamanho (pb)
VuRGAq1	TGGAGCGATTCATGTTATGG	CCCTTCTTGACCAATCTTGC	20/20
VuRGAq2	AGCCCTGATTGGTTTGGTC	CATCAACCTGTTCCGTTTGG	19/20
VuRGAq3	GCTCTCCTCTCTCTCGATCTTTC	GGGTAACAAACCCTCGAACC	23/20
VuRGAq4	GTTGAAGGAGAGCCACCAAG	GCCACACCATAAACCTGTC	20/20
VuRGAq5	CCTGAATGGTGGGTGAGACT	AAACCAGAACCTCGGCTTC	20/19
VuRGAq6	TTGCCTTCCCATCTCATCTC	AAGGGTCTTCGTAAGGTCTCG	20/21
VuRGAq7	TTGGTGAGAAGGAGCATCAG	TGGAGGCTTAGGAAAGACTACG	20/22
VuRGAq8	ACGAATGGCTCCACATTACC	TTGAATGCTACTCGGCAGTG	20/20
VuRGAq9	TGGTGTCTTGATGAGCTTGC	AGGCTTTCTTCCACCTCTGC	20/20
VuRGAq10	ACATGCCACACATCATCCAG	GCTGCAAGAGCTTCTCAAGG	20/20

A tabela 3 apresenta os níveis de expressão relativa, referentes aos genes alvo, mediante diferentes combinações entre controle e condições de estresse causadas pelo vírus do mosaico severo. Fazendo a comparação entre à condição de estresse com o vírus no tempo 30 min em relação ao controle no tempo 30 min (T30C30), pôde-se observar que três alvos não apresentaram expressão diferencial significativa, onde apenas o alvo VuRGAq4 teve sua expressão reprimida. Enquanto na comparação entre à condição de estresse com o vírus no tempo 60 min em relação ao controle no tempo 30 min (T60C30), o alvo VuRGAq3, apresentou uma redução na sua expressão, ao contrário dos demais, que se mantiveram não diferencialmente expressos. Ao se analisar a condição comparativa entre à condição de estresse com o vírus no tempo de 16 horas em relação ao controle no tempo 16 horas (T16T16), observou-se que os alvos VuRGAq3, VuRGAq4 e VuRGAq6, foram *up*-regulados, apresentando os seguintes níveis de expressão relativa: *fold-change* 1,291; 1,341; 1,279, respectivamente, indicando uma resposta tardia dos genes NBS-LRR frente ao estresse causado

pelo CPSMV. Por sua vez, o alvo VuRGAq10 manteve a sua expressão diferencial não significativa.

Tabela 3: Dados de expressão gênica dos genes alvo, obtidos a partir do programa REST2009, entre as diferentes combinações: T30C30 (tratamento com vírus tempo 30 min contra controle tempo 30 min), T60C30 (tratamento com vírus tempo 60 min contra controle tempo 30 min) e T16C16 (tratamento com vírus tempo 16 horas contra controle tempo 16 horas), utilizando-se como genes de referência ACT27, Beta-tubulina e SKIP.

		T30C30		T60C30		T16C16	
Alvos	Contig EST	Expressão relativa		Expressão relativa		Expressão relativa	
VuRGAq3	SI3507.1	0,898	ns	0,825	<i>Down</i>	1,291	<i>Up</i>
VuRGAq4	SI4993.4	0,698	<i>Down</i>	0,973	ns	1,341	<i>Up</i>
VuRGAq6	PO11988.1	0,824	ns	0,961	ns	1,279	<i>Up</i>
VuRGAq10	SI23875.2	1,043	ns	1,208	ns	0,992	Ns

* *Down* - expressão reprimida; *Up* - expressão induzida; ns- não significativo.

Um esquema gráfico entre a relação da expressão relativa dos genes alvo com o desvio padrão atribuído pelo programa REST2009 (Std. Error) encontra-se apresentado na Figura 7.

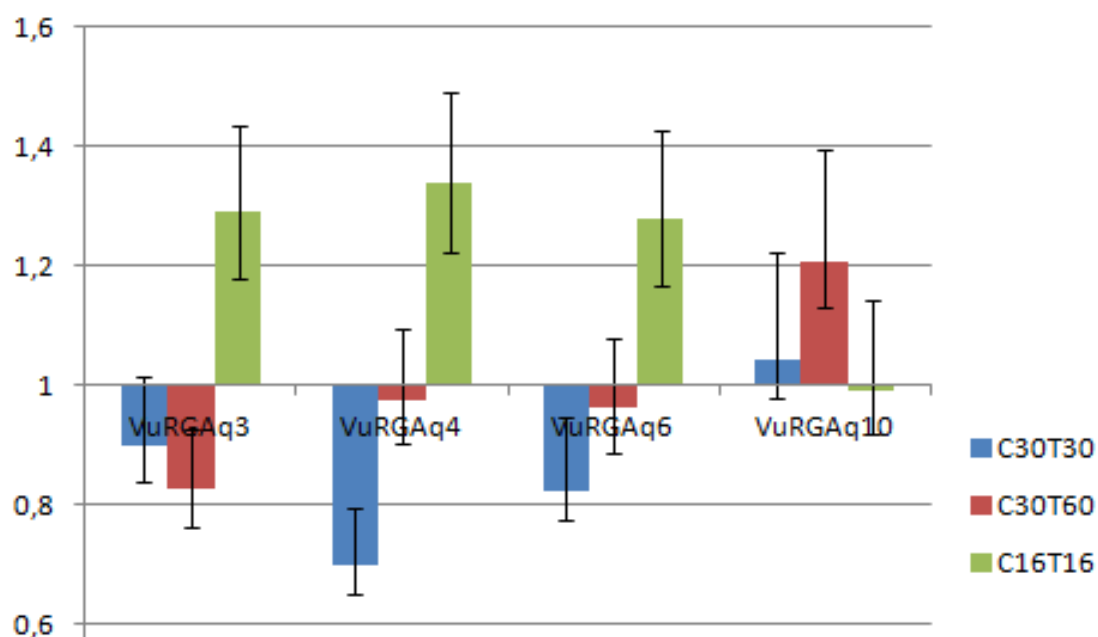


Figura 7: Representação gráfica dos níveis de expressão relativa dos genes alvo. T30C30 (tratamento com vírus tempo 30 min contra controle tempo 30 min), T60C30 (tratamento com vírus tempo 60 min contra controle tempo 30 min) e T16C16 (tratamento com vírus tempo 16 horas contra controle tempo 16 horas). Genes alvo: VuRGAq3, VuRGAq4, VuRGAq6, VuRGAq10, sendo os três primeiros *up*-regulados em C16T16. Na vertical está demonstrada a taxa de expressão dos genes.

Discussão

Análises comparativas de genomas de vegetais indicam que estes são capazes de codificar centenas de genes NBS-LRR, existindo uma grande diversidade no número e na distribuição das subfamílias características desses genes (Marone *et al.*, 2013). Diversos estudos têm revelado a presença desses genes no genoma de várias plantas, tais como *Medicago truncatula* (Ameline-Terrogrosa *et al.*, 2008), soja (Wanderley-Nogueira, *et al.*, 2012a), arroz (Liu, *et al.*, 2013), entre outros.

A redução significativa, observada no presente estudo, entre as sequências obtidas no banco de dados do NordEST e os candidatos a genes NBS-LRR em feijão-caupi, pode ser justificada pelo fato do domínio LRR ser compartilhado com

mais duas famílias pertencentes ao grupo de genes de resistência em plantas, RLP (Proteína Tipo Receptora, *Receptor-Like Protein*) e RLK (Receptor Tipo Quinase, *Receptor-Like Kinase*) (Van-Ooijen *et al.*, 2007).

Os motivos do domínio NBS identificados no presente estudo foram semelhantes aos observados por Palomino *et al.* (2006) nos genomas de *Vicia faba* e *Cicer arietinum*, assim como no trabalho realizado por Gowda *et al.* (2002) em feijão-caupi.

Vários trabalhos desenvolvidos com genes NBS-LRR se baseiam no motivo *kinase-2* para distinguir as duas subfamílias TIR e não-TIR (Palomino *et al.*, 2006; Ameline-Torregrosa *et al.*, 2008; Maiti *et al.*, 2012). No presente estudo os membros correspondentes à subfamília não-TIR foram caracterizados pela presença do aminoácido triptofano (W) na última posição do motivo *kinase-2*, resultado semelhante ao encontrado pelos autores acima mencionados. No que se refere às sequências TIR, estas foram caracterizadas pela presença do ácido aspártico (D), porém foram observadas algumas variações, tais como a presença dos aminoácidos N, S, A, T ou L nessa posição. Esses resultados não corroboram com os relatados por Gowda *et al.* (2002), onde foi identificado apenas o ácido aspártico na última posição em feijão-caupi. Porém, alterações como essas já foram descritas em *V. faba* e *C. arietinum* (Palomino *et al.*, 2006). No motivo NBS-II foi possível observar que existe um padrão conservado (CFL – cisteína - fenilalanina - leucina) em feijão-caupi na subfamília TIR (Gowda *et al.*, 2002), o qual está ausente em *V. faba* e *C. arietinum* (Palomino *et al.*, 2006).

O nível de homologia dos genes *R* é determinado em termos de semelhança ou identidade dos aminoácidos. A similaridade permite a variação conservativa, isto é, a substituição de um resíduo hidrofóbico como leucina por outro como valina,

isoleucina ou metionina, assim como a substituição de um resíduo polar, como arginina por outro como lisina. Sabe-se que a alteração da estrutura primária de um polipeptídeo, através de uma substituição conservadora, pode não alterar significativamente a atividade do referido peptídeo (Gebhardt *et al.*, 2012).

O domínio LRR desempenha um importante papel no reconhecimento de produtos específicos de genes *Avr*, para promover a resistência das plantas a patógenos (Mondragon-Palomino *et al.*, 2002). Padrões de repetições semelhantes aos descritos nesse estudo para esse domínio têm sido relatados em outras análises, onde se acredita que uma forte seleção positiva possa estar agindo nos resíduos não-específicos (X), proposto como mecanismo de diversificação de sequências (Ratnaparkhe *et al.*, 2011). O motivo “LXXLXL” do domínio LRR, também foi identificado por Butterbach (2007) em estudo com o gene *Rx* em batata (*Solanum tuberosum*). Bent e Mackey (2007) e Maiti *et al.* (2012) acreditam que tais motivos formem folhas-beta onde, por sua vez, os resíduos X são solventes expostos e disponíveis para detectar potenciais partículas de patógenos. No presente estudo o domínio CC não foi identificado, de forma semelhante ao observado por Ameline-Torregrosa *et al.*, (2008) em *M. truncatula*, devido a dificuldade para detecção desse domínio nos programas de predição disponíveis.

Vários estudos realizados com os genes NBS-LRR com o intuito de desenvolver RGAs utilizaram os domínios conservados para o desenho de *primers*. Nesse trabalho foram encontrados sete RGAs para sequências completas contendo os domínios TIR-NBS-LRR e sete para sequências completas apresentando o NBS-LRR. Ratnaparkhe *et al.*, (2011) identificaram seis RGAs no amendoim (*Arachis hypogaea* L.), onde dois tinham os domínios TIR-NBS-LRR completos, enquanto que nos outros quatro faltavam um ou dois domínios. Em

estudos com *V. mungo* foram desenvolvidos três marcadores moleculares do tipo RGA (Maiti *et al.*, 2010; Maiti *et al.*, 2012), os quais foram utilizados em populações de mapeamento genético. Esses marcadores estão frequentemente associados a *loci* de genes *R* (Wang *et al.*, 2006), ou com parte de genes *R* (Maiti *et al.*, 2010). Maiti *et al.* (2010) mostraram claramente que o marcador RGA-CYR1 está totalmente ligado à resistência de *V. mungo* ao vírus MYMIV, segregando fielmente com essa característica em uma população com 335 indivíduos.

Tem-se observado que a distribuição dos genes NBS-LRR nos cromossomos é muito irregular no genoma da maioria das espécies previamente estudadas. Assim, como nesse estudo, foi possível observar que poucos cromossomos retêm a maioria desses genes, como no caso da batata, onde os cromossomos 4 e 11 suportam cerca de 15% dos genes mapeados até o momento (Lozano *et al.*, 2012). Nas leguminosas, também é possível identificar essa distribuição desigual, Ameline-Torregrosa *et al.* (2008) observaram que apenas os cromossomos 3, 4 e 6 correspondem a 54% dos genes *R* descritos em *M. truncatula*. Os genes pertencentes a essa família tendem a se organizar em agrupamentos (*clusters*), muitas vezes formando superaglomerados. Tais aglomerados são esperados devido à existência de domínios comuns compartilhados por genes *R* (Wanderley-Nogueira *et al.*, 2012a). Acredita-se que os eventos de duplicação sejam responsáveis pelo surgimento das repetições ao longo de cada cromossomo. Por sua vez, fenômenos de duplicação seguidos de transposição ou translocação são considerados para justificar a presença de sequências de um mesmo gene em locais distintos de um mesmo cromossomo ou em cromossomos diferentes, respectivamente (Hulbert *et al.*, 2001).

Regiões conservadas observadas neste estudo, também foram relatadas por Wanderley-Nogueira *et al.* (2012c) onde compararam o feijão-caupi e a soja, e descreveram a existência de uma maior prevalência de genes *R* nas regiões terminais dos cromossomos, fato que está relacionado com a variabilidade genética presente na família gênica NBS-LRR, especialmente com a alta variação do domínio LRR, visto que essas regiões estão mais propícias à ocorrência de mutações.

Wan *et al.* (2012) em estudos com genes NBS-LRR em *Capsicum annuum* L., observaram que os genes pertencentes a essa família tendem a agrupar-se em *clusters* diferentes, de acordo com sua subfamília. Assim como nesse estudo, eles identificaram seis motivos dentro do domínio NBS, e acreditam que a presença dos motivos de consenso não-TIR e TIR-NBS-LRR encontram-se agrupados (clusterizados), fornecendo mais uma prova de que os membros dessa família gênica apresentam uma origem em comum, padrão observado também em *M. truncatula* (Ameline-Torregrosa *et al.*, 2008).

Com base no ortólogo ao gene candidato 'contig98' encontrado na briófitas *Physcomitrella patens* L., acredita-se que o gene tenha surgido antes do aparecimento das plantas vasculares (angiospermas) e que tenha sido perdido pelas monocotiledôneas (Meyers *et al.*, 1999) no processo evolutivo, após a separação entre as duas classes. O agrupamento que ocorreu com os membros pertencentes à ordem Fabales no presente estudo, corroborou com o trabalho de Wanderley-Nogueira *et al.* (2012a), onde alfafa e soja também posicionaram-se no mesmo grupo, no que diz respeito à organização de genes *R*. Analisando genes de resistência, Wanderley-Nogueira *et al.* (2012a) observaram que as espécies *A. thaliana* (Brassicaceae) e *Ricinus communis* (Euphorbiaceae) agruparam-se no

mesmo clado, apresentando-se como um grupo merofilético parafiléticos, enquanto que no presente estudo, essas duas espécies foram separadas em clados distintos, de acordo com a sua ordem, sendo que *A. thaliana* pertence à ordem Brassicales e *R. communis* Malpighiales.

A análise de expressão gênica apresenta-se como uma ferramenta importante que pode contribuir para a caracterização de resistência. Como revisado por Benko-Iseppon *et al.* (2010) genes que resultam em HR estão associados a diferentes estímulos, sendo ativados não só devido à percepção do patógeno, mas também sob estresse abiótico, enquanto muitos deles são constitutivamente ativos em níveis baixos, também na ausência de qualquer tipo de estresse, justificando sua prevalência nos dados de RNAseq, quanto a expressão gênica *in silico*, observado no presente estudo. Assim, um número crescente de evidências apoia a noção de que as vias de sinalização de plantas consistem de redes com presença frequente de resposta cruzada (*crosstalk*), permitindo assim que as plantas possam regular genes envolvidos com a tolerância ao estresse abiótico, como também aqueles associados à resistência a doenças (Velazquez *et al.* 2011).

Estudos demonstram a aplicabilidade da RT-qPCR para observação das variações na expressão gênica de transcritos de genes NBS-LRR em vegetais, como em *P. vulgaris* (David *et al.*, 2010) e *Poncirus trifoliata* (Boava *et al.*, 2011).

Maiti *et al.*, (2012) analisando o nível de expressão diferencial dos genes NBS-LRR em *Vigna mungo*, mediante uma infecção por vírus, utilizaram como padrão de normalização um gene de referência desenvolvido para actina. No presente estudo foi usado o gene de referência ACT27, o que pode indicar a importância desse gene para validar estudos de resposta a viroses no gênero *Vigna*.

Foi observado que os genes NBS-LRR em feijão-caupi aumentaram a sua expressão após 16 horas da inoculação do vírus, culminando em uma resposta tardia. Esses dados corroboraram com os relatados por Maiti *et al.*, (2012), onde a expressão aumentou significativamente após 12 horas da inoculação do vírus MYMIV (*Mungbean yellow mosaic India virus*) em *V. mungo*, atingindo o nível máximo de expressão às 18 horas, seguido de repressão gradativa até as 72 horas. Em trabalho realizado com café (*Coffea* spp.) infectado por ferrugem, observou-se uma resposta ainda mais tardia, ocorrendo a maior expressão de genes NBS-LRR entre 24 e 72 horas após a inoculação (Diola *et al.*, 2013).

Agradecimentos

Os autores agradecem ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco) e CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo apoio financeiro e a EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) e IPA (Instituto Agrônômico de Pernambuco) pela concessão do material genético.

Referências

- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF *et al.* (1991) Complementary DNA sequencing: expressed sequence *tags* and human genome project. *Science* 252:1651–1656.
- Ameline-Torregrosa C, Wang BB and O’Bleness MS (2008) Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol* 146(1):5–21.
- Assunção IP, M-Filho RL, Resende LV, Barros MCS, Lima GSA, Coelho RSB e Lima JAJ (2005) Genes diferentes podem conferir resistência ao *Cowpea severe mosaic virus* em caupi. *Fitopatol Bras* 30:274-278.

- Benko-Iseppon AM (2010) Brazilian cowpea transcriptome project: over 20 million expressed sequence *tags* to analyze and breed salinity and virus resistance. Resumos do III Congresso Brasileiro de Biotecnologia 97-98 pp.
- Bent AF and Mackey D (2007) Elicitors, effectors, and *R* genes: the new paradigm and a life time supply of questions. *Annu Rev Phytopathol* 45: 399–436.
- Boava LP[†], Cristofani-Yaly M, Mafra VS, Kubo K, Kishi LT, Takita MA, Ribeiro-Alves R and Machado MA (2011) Global gene expression of *Poncirus trifoliata*. *Genomics* 12(39):1471-2164.
- Butterbach P (2007) Molecular evolution of the disease resistance gene *Rx* in *Solanum*. PhD thesis, University Wageningen, The Netherlands, ISBN: 90-8504-724-7.
- Da Silva M, Oliveira-Silva RL, Ferreira-Neto JRC, Guimarães ACR, Veiga DT, Chabregas SM, Burnquist WL, Kahl G, Benko-Iseppon AM and Kido EA (2013) Expression Analysis of Sugarcane Aquaporin Genes under Water Deficit. *J Nucl Acids* 1-14 pp.
- David P, Francs-Small CC, SéVignac M, Thareau V, Macadré C, Langin T and Geffroy v (2010) Three highly similar formate dehydrogenase genes located in the vicinity of the B4 resistance gene cluster are differentially expressed under biotic and abiotic stresses in *Phaseolus vulgaris*. *Theor Appl Genet* 121:87–103
- Diola V, Brito GG, Caixeta ET, Pereira LF, Loureiro ME (2013) A new set of differentially expressed signaling genes is early expressed in coffee leaf rust race II incompatible interaction. *Funct Integr Genomics* 133:379-89.
- Donald TM, Pellerone F, Adam-Blondon AF, Bouquet A, Thomas MR and Dry IB (2002) Identification of resistance gene analogs linked to a powdery mildew resistance locus in grapevines. *Theor Appl Genet* 104:610-618.
- Eisen MB, Spellman PT, Brown PO[†] and Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95:14863–14868.
- Flor HH (1971) Current status of the gene-for-gene concept. *Annu Rev Phytopathol* 9:275–296.
- Gao JP, Chao DY and Lin HX (2007) Understanding Abiotic Stress Tolerance Mechanisms: Recent Studies on Stress Response in Rice. *J Integr Plant Biol* 49:742–750.
- Gebhardt C, Ballvora A, Ercolano MR, Weiss J and Salamini F (2012) Gen de resistência derivado de planta. Patente 2 375 933.
- Gowda BS, Miller JL, Rubin SS, Sharma DR and Timko MP (2002) Isolation, Sequence analysis, and linkage mapping of resistance-gene analogs in cowpea (*Vigna unguiculata* L. Walp.). *Euphytica* 126:365–377.
- Guimarães PM, Proite k, Leal-Bertioli SC e Bertioli D (2005) Análise *in silico* da expressão gênica diferencial de *Arachis stenosperma* inoculado com *Meloidogyne arenaria*. *Boletim de Pesquisa e Desenvolvimento, EMBRAPA Recursos Genéticos e Biotecnologia* 85.
- Hulbert SH, Webb CA, Smith SM and Sun Q (2001) Resistance gene complexes: Evolution and utilization. *Annu Rev Phytopathol* 39:285–312.
- Jones DG (2001) Putting the knowledge of plant disease resistance genes to work. *Curr Opin Plant Biol* 4: 281-287.
- Lacock L, Van-Niekerk C, Loots S, Du Preez F and Botha AM (2003) Functional and comparative analysis of expressed sequences from *Diuraphis noxia* infested

- wheat obtained utilizing the conserved Nucleotide Binding Site. *Afr J Biotechnol* 2:75-81.
- Liu Y, Liu B, Zhu X, Yang J, Bordeos A, Wang G, Leach JE and Leung H (2013) Fine-mapping and molecular marker development for Pi56(t), a NBS-LRR gene conferring broad-spectrum resistance to *Magnaporthe oryzae* in rice. *Theor Appl Genet* 126:985–998.
- Lozano R, Ponce O, Ramirez M, Mostajo N and Orjeda G (2012) Genome-Wide Identification and Mapping of NBS-Encoding Resistance Genes in *Solanum tuberosum* Group Phureja. *PLoS ONE* 7(4):e34775.
- Maiti S, Basak J, Kundagrami S, Kundu A and Pal A (2010) Molecular marker-assisted genotyping of Mungbean yellow mosaic India virus resistant germplasms of mungbean and urdbean. *Mol Biotechnol* 47:95–104.
- Maiti S, Paul S and Pal A (2012) Isolation, Characterization, and Structure Analysis of a Non-TIR-NBS-LRR Encoding Candidate Gene from MYMIV-Resistant *Vigna mungo*. *Mol Biotechnol* 52:217-233.
- Marone D, Russo MA[†], Laidò G, Leonardis AM and Mastrangelo AM (2013) Plant Nucleotide Binding Site-Leucine-Rich Repeat (NBS-LRR) Genes: Active Guardians in Host Defense Responses. *Int J Mol* 14:7302-7326.
- McHale L, Tan X, Koehl P and Michelmore CW (2006) Plant NBS-LRR proteins: adaptable guards REVIEW. *Genome Biol* 7(212).
- Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW and Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J* 20:317-332.
- Miller RNG[†], Bertoli DJ, Baurens FC, Santos CMR, Alves PC, Martins NF, Togawa RC, Souza-Júnior MT and Pappas-Júnior GJ (2008) Analysis of non-TIR NBS-LRR resistance gene analogs in *Musa acuminata* Colla: Isolation, RFLP marker development, and physical mapping. *BMC Plant Biol* 8(15).
- Mondragon-Palomino M, Meyers BC, Michelmore RW and Gaut BS (2002) Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Res* 12:1305–1315.
- Morin R, Bainbridge M, Fejes A, Hirst M, Krzywinski M, Pugh TJ, McDonald H, Varhol R, Jones SJM and Marra MA (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* 45:81-94.
- Oliveira JTA (2006) Mecanismos de defesa do feijão-caupi [*Vigna unguiculata* (L.) walp.] contra patógenos. Congresso Nacional de Feijão caupi, Anais, VI Reunião Nacional de Feijão caupi, 4 pp.
- Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Comp Appl Biosci* 12:357-358.
- Palomino C, Satovic Z, Cubero JI and Torres AM (2006) Identification and characterization of NBS–LRR class resistance gene analogs in faba bean (*Vicia faba* L.) and chickpea (*Cicer arietinum* L.). *Genome* 49(10): 1227-37.
- 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.
- Ratnaparkhe MB, Wang X, Li J, Compton RO, Rainville LK, Lemke C, Kim C, Tang H and Paterson AH (2011) Comparative analysis of peanut NBS-LRR gene clusters suggests evolutionary innovation among duplicated domains and erosion of gene microsynteny. *New Phytol* 192:164–178.

- Ren J, Yu Y, Gao F, Zeng L, Lu X, Wu X, Yan W and Ren G (2013) Application of resistance gene analog markers to analyses of genetic structure and diversity in rice. *Genome* 56:377–387.
- Sanseverino W, Roma G, De Simone M, Faino L, Melito S, Stupka E, Frusciante L and Ercolano MR (2010) PRGdb: a bioinformatics platform for plant resistance gene analysis. *Nucl Acids Res* 38:D814-D821.
- Tamura K, Dudley J, Nei M and Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599.
- Van-Ooijen, G, Van-Den Burg HA, Cornelissen BJ and Takken FL (2007) Structure and function of resistance proteins in solanaceous plants. *Annu Rev Phytopathol* 45:43–72.
- Velazquez FS, Guerra RR and Calderon SL (2011) Abiotic and Biotic Stress Response Crosstalk in Plants. *In: Shanker A and Venkateswarlu B (Eds) Abiotic Stress Response in Plants Physiol Biochem Genetic Perspectives* 3-26 pp.
- Wan H, Yuan W, Ye Q, Wang R, Ruan M, Li Z, Zhou G, Yao Z, Zhao J, Liu S and Yang Y (2012) Analysis of TIR- and non-TIR-NBS-LRR disease resistance gene analogous in pepper: characterization, genetic variation, functional divergence and expression patterns. *BMC Genomics* 13(502).
- Wanderley-Nogueira AC, Kido EA, Soares-Cavalcanti NM, Belarmino L, Bezerra-Neto JP, Burnquist WL, Chabregas SM, Baldani JI and Benko-Iseppon AM (2012a) Insight on pathogen defense mechanisms in the sugarcane transcriptome. *Functional Plant Sci Biotech* 6:134-148.
- Wanderley-Nogueira AC, Belarmino LC, Soares-Cavalcanti NM, Bezerra-Neto JP, Kido EK, Pandolfi V, Abdelnoor RV, Binneck E, Carazzole MF and Benko-Iseppon AM (2012b) An overall evaluation of the resistance (*R*) and pathogenesis-related (*PR*) superfamilies in soybean, as compared with *Medicago* and *Arabidopsis*. *Genet Mol Biol* 35:260-271.
- Wanderley-Nogueira (2012c) Caracterização bioinformática de genes relacionados a interação patógeno-hospedeiro em angiospermas. Tese de Doutorado, Universidade Federal de Pernambuco, 217 p.
- Wang HY, Yang WX, Zhang LR, Meng QF, Liu DQ and Zhang T (2006) Isolation and characterization of resistance gene homology sequence from wheat. *J Phytopathol* 154:670–675.
- Weising K, Nybom H, Pfenninger M, Wolff K, Kahl G (2005) DNA fingerprinting in plants: Principles, Methods, and Applications. (Ed), CRC Press, Boca Raton, 444 pp.

6. Discussão geral

Assim como esse estudo, vários trabalhos têm sido realizados com o intuito de realizar uma investigação a respeito da quantidade e distribuição dos genes NBS-LRR no genoma dos vegetais, tais como em *Arabidopsis thaliana*, tomate, arroz, banana, alfafa, soja, cana-de-açúcar, amendoim, feijão comum, entre outras. Esses estudos mostraram que os genes *R* tendem a se organizar formando agrupamentos em regiões específicas dentro dos cromossomos. Esses agrupamentos podem variar muito, sendo o menor aglomerado relatado em tomate, contendo apenas dois genes (CF2/CF5), enquanto que o maior, apresentando 24 genes, foi descrito na alface (*loci* Dm3).

O quadro responsável pela organização e evolução das sequências correspondentes a genes NBS-LRR em plantas é semelhante ao descrito para o complexo de histocompatibilidade (MHC), receptor de células T e os genes da imunoglobulina em vertebrados. Portanto, as forças evolutivas que compõem os sistemas de defesa das plantas e vertebrados podem ser semelhantes, embora os tipos específicos de genes envolvidos sejam diferentes. Provavelmente o evento de duplicação seja o passo inicial mais óbvio para o surgimento dos complexos aglomerados de genes *R*, pois podem criar novos *loci*, alterar o número de membros da família de genes e/ou gerar sequências repetidas dentro de um gene. Os agrupamentos desses genes fornecem um número maior de oportunidades para que as plantas sejam capazes de enfrentar o rápido desafio da evolução dos patógenos.

Todos os 57 candidatos tiveram resultados positivos para o domínio NBS e codificavam proteínas com os motivos esperados. Esse domínio é composto por

vários motivos característicos, que correspondem a regiões altamente conservadas dentro do domínio NBS, tais como *P-loop*, *Kinase-2* e GLPL. O motivo *P-loop* corresponde à região de ancoragem, enquanto os demais estão relacionados à orientação da molécula de ATP. Estudos mostraram que o domínio NBS pode desencadear uma resposta de defesa, sugerindo que o mesmo possa ser capaz de desencadear uma cascata de sinalização por si só, dos genes pertencentes a essa família. Os motivos conservados identificados dentro do domínio NBS em feijão-caupi são evolutivamente conservados tanto em plantas próximas como também em plantas que são remotamente relacionadas, sendo encontrados em diferentes famílias, ordens e grupos vegetais.

A metodologia baseada em PCR foi eficaz para o desenvolvimento de RGA no feijão-caupi, aproveitando a relativa conservação existente nos domínios TIR e NBS. Essa técnica é bastante utilizada em estudos que visam o desenvolvimento desses marcadores em vegetais, como em *Medicago truncatula*, em *Phaseolus vulgaris* e em *Oryza Sativa*, devido à sua eficiência, facilidade e baixo custo. O domínio LRR presente nos genes NBS-LRR, caracterizado por repetições de leucinas conservadas (LXXLXXL e LXXLXL), é responsável pelo evento de reconhecimento e, portanto, de maior interesse para o desenvolvimento de novas especificidades de reconhecimento aos patógenos.

Apesar de existirem vários estudos realizados sobre o desenvolvimento de marcadores moleculares, os programas de melhoramento no Brasil ainda não foram beneficiados com o uso de marcadores RGAs. Com isso, mapas genéticos foram desenvolvidos e devem ser saturados para identificação desses marcadores ligados a características agrônômicas de interesse. Para tal finalidade, novos marcadores análogos a genes de resistência polimórficos precisam ser

identificados, para serem utilizados em progênie de mapeamento genético, assim contribuindo para o aperfeiçoamento do melhoramento genético de culturas de interesse econômico no Brasil.

Em geral, existem poucas informações sobre os mecanismos específicos de regulação e correção da resistência de leguminosas cultivadas frente aos estresses bióticos. Esses genes estão no topo da cascata de reconhecimento do patógeno e apresentam alta complexidade, sendo variáveis devido aos mecanismos genéticos de coevolução na interação planta-patógeno. É importante salientar que esses genes não atuam de forma independente, ou seja, dependem de interações com outros genes/proteínas. Nesse estudo foi observado que vários genes de resistência NBS-LRR são expressos de forma similar em feijão-caupi, corroborando com diversos trabalhos caracterizados na literatura. Essa análise deve ser confirmada através de técnicas como *Northern blot*, RT-PCR e PCR quantitativa em tempo real (RT-qPCT). Por sua vez, outros genes apresentam baixos níveis de expressão quando não estão infectados por patógenos, indicando uma expressão induzida por partículas efetoras, como no caso dos genes *RPM1* e *RPP5* em *A. thaliana*.

Em vista da importância econômica da cultura do feijão-caupi para as regiões Norte e Nordeste do Brasil, além de sua importância ecológica, ajudando a evitar a erosão, devido à formação de cobertura vegetal, pode-se considerar que a rede genômica NordEST e o presente trabalho colaboram significativamente para o entendimento dos processos de defesa na cultura do feijão-caupi e indiretamente no gênero *Vigna*, embora o quantitativo identificado seja claramente inferior ao potencial existente, uma vez que os genes NBS-LRR compõem boa parte do genoma de plantas. No genoma de *A. thaliana*, por exemplo, com 172 Mb (mega

pares de bases), foram descritos 149 genes NBS-LRR, o que representa aproximadamente 1% do genoma. Assim como em arroz, com tamanho de 402 Mb, foram localizados 600 genes NBS-LRR, também representando cerca de 1% do genoma.

Apesar da grande conservação identificada no domínio NBS, o alinhamento múltiplo realizado com as sequências candidatas de feijão-caupi, mostrou grande variação entre as sequências, especialmente nas regiões entre os motivos conservados, como esperado, também consideradas importantes para funções bioquímicas, apesar de acumularem um maior número de mutações.

7. Conclusões gerais

1. Os 57 candidatos a genes NBS-LRR em feijão-caupi apresentaram os domínios conservados típicos da família gênica estudada, sendo relatada a presença de membros pertencentes as duas subfamílias.
2. Foram desenvolvidos dois marcadores moleculares do tipo RGA em feijão-caupi, com base nas regiões interdomínios significativamente variáveis, os quais apresentam possível aplicação no melhoramento assistido para o desenvolvimento de cultivares de feijão-caupi resistentes a viroses;
3. Considerando a distribuição desses genes no genoma de *Glycine max*, acredita-se que estes estejam organizados de forma clusterizada nos cromossomos de feijão-caupi;
4. Os genes analisados foram agrupados em dois grupos fenéticos, sendo clusterizados de acordo com as subfamílias existentes, devido à conservação do motivo *kinase-2* presente no domínio NBS;
5. As leguminosas foram agrupadas em um único clado, indicando que existe uma menor diversidade de genes NBS-LRR em espécies próximas, do que em espécies remotamente relacionadas;

6. Os maiores níveis de expressão de genes NBS-LRR em feijão-caupi ocorreram após 16 horas da inoculação do patógeno, resultando em uma resposta de defesa tardia;
7. Os resultados obtidos nesse trabalho podem contribuir para um aperfeiçoamento do conhecimento referente aos mecanismos genéticos relacionados à resistência aos estresses bióticos, disponibilizando informações novas para os programas de melhoramento genético de leguminosas.

8. Referências Bibliográficas

- Abby S and Daubin V (2007) Comparative genomics and the evolution of prokaryotes. *Trends Microbiol* 15:135-141.
- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merrill CR, Wu A, Olde B, Moreno RF *et al.* (1991) Complementary DNA sequencing: expressed sequence *tags* and human genome project. *Science* 252:1651–1656.
- Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403-410.
- Araújo JPP, Rios GP, Watt EE, Neves BP, Fageria NK, Oliveira IP, Guimarães CM, Filho S (1984) A Cultura do Caupi, *Vigna unguiculata* (L.) Walp; descrição e recomendações técnicas de cultivo. EMBRAPA, Circular Técnica, 18:82.
- Athayde-Sobrinho C, Viana FMP and Santos AA (2000) Doenças do feijão caupi. In: Cardoso MJ (Org) A Cultura do Feijão Caupi no Meio-Norte do Brasil, Teresina, Embrapa Meio-Norte, (Embrapa Meio-Norte, Circular Técnica, 28:157-183.
- Banerjee D, Zhang X and Bent AF (2001) The leucine-rich repeat domain can determine effective interaction between RPS2 proteins and other host factors in *Arabidopsis* RPS2-mediated disease resistance. *Genetics* 158:439–450.
- Barreto PD (1999) Recursos genéticos e programa de melhoramento de feijão-de-corda no Ceará: Avanços e perspectivas. In: Queirós MA, Goedert CO, Ramos SRR (eds) Recursos genéticos e melhoramento de plantas para o Nordeste Brasileiro. EMBRAPA, CPATSA, 27.
- Barroso MR, Magalhães MJ, Carnide S, Vegas CA e Cachón MR (2007) Caracterização e avaliação de diferentes espécies de leguminosas grão na região de trás-os-montes. Coleção uma agricultura com Norte. ISBN: 978-972-85-6-67-4.
- Baxevanis AD (2001) Bioinformatics and the internet. *Methods Biochem Anal* 43:1-17.
- Benko-Iseppon AM (2009) Identificação e Validação de Genes Importantes para a Biotecnologia e Produtividade a partir do Transcriptoma do Feijão-Caupi (*Vigna unguiculata*). Projeto Submetido ao Edital MCT/CNPq/CT-AGRO/CT-BIOTEC Nº 42/2009. Programa Genoprot - Rede Integrada de Estudos Genômicos e Proteômicos, 1-34 pp.
- Benko-Iseppon AM (2010a) Brazilian cowpea transcriptome project: over 20 million expressed sequence *tags* to analyze and breed salinity and virus resistance. Resumos do III Congresso Brasileiro de Biotecnologia, 97-98 pp.
- Benko-Iseppon AM, Galdino SL, Calsa Junior T, Kido EA, Tossi A, Belarmino LC and Crovella S (2010b) Overview of Plant Antimicrobial Peptides. *Curr Prot Pep Sci* 113:181-188.
- Benko-Iseppon AM, Nepomuceno AL and Abdelnoor RV (2012) GENOSOJA – The Brazilian Soybean Genome Consortium: High throughput omics and beyond. *Genet Mol Biol* 35:i-iv.
- Bent AF and Mackey D (2007) Elicitors, effectors and *R* genes: The new paradigm and a lifetime supply of questions. *Annu Rev Phytopathol* 45:399-436.
- Bezerra CAA, Tavora FAJF, Freire-Filho RF e Ribeiro QR (2008) Morfologia e produção de grãos em linhagens modernas de feijão caupi em diferentes densidades populacionais. *Rev Biol Ciênc Terra* 8:85-93.

- Blair MW, Astudillo C, Grusak MA, Graham R and Beebe SE (2009) Inheritance of seed iron and zinc concentrations in common bean (*Phaseolus vulgaris* L.). *Mol Breed* 23:197-207.
- Bolgue H (2011) Pesquisadora descobre alternativa para o tratamento do câncer de mama Molécula extraída do feijão-de-corda teve boa resposta contra células cancerígenas. UnB Agência (Universidade de Brasília).
- Brito ES (Ed) (2008) Feijão caupi. Fortaleza: Embrapa Agroindústria Tropical, 97.
- Broughton WJ, Hanin M, Relic B, Kopcińska J, Golinowski W, Simsek S, Ojanen-Reuhs T, Reuhs B, Marie C, Kobayashi H, *et al.* (2006) Flavonoid-inducible modifications to rhamnan O antigens are necessary for *Rhizobium* sp. strain NGR234- Legume Symbioses. *J Bacteriol* 188:3654-3663.
- Collins N, Park R, Spielmeyer W, Ellis J and Pryor AJ (2001) Resistance gene analogs in barley and their relationship to rust resistance. *Gene* 44:375-381.
- Damasceno-Silva KJ (2007) O feijão-caupi para processamento industrial e exportação. Portal do Agronegócio. Disponível em: <<http://www.infoteca.cnptia.embrapa.br/handle/doc/69639>>.
- Donald TM, Pellerone F, Adam-Blondon AF, Bouquet A, Thomas MR and Dry IB (2002) Identification of resistance gene analogs linked to a powdery mildew resistance locus in grapevines. *Theor Appl Genet* 104:610-618.
- Edwards D and Batley J (2004) Plant bioinformatics: from genome to Phenome. *Trends in Biotechnol* 22:232-237.
- Eisen MB, Spellman PT, Brown PO⁺ and Botstein D⁺ (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95(25):14863–14868.
- Ellis J and Jones D (1998) Structure and function of proteins controlling strain-specific pathogen resistance in plants. *Curr Opin Plant Biol* 1:288-293.
- Enkhbayar P, Kamiya M, Osaki M, Matsumoto T and Matsumisha M (2003) Structural principles of Leucine Rich Repeat proteins. *Proteins* 54:393-403.
- Fermin-Muñoz GA, Meng B, Ko K, Mazumdar-Leighton S, Gubba A and Carroll JE (2000) Biotechnology: a new era for plant pathology and plant protection. APSnet Feature. Disponível em <<http://www.apsnet.org/publications/apsnetfeatures/Pages/Biotechnology.aspx>>.
- Fery FL (2002) New opportunities in *Vigna*. In: Janick J and Whipkey A (eds), *Trends in new crops and new uses*. ASHS Press, 424–428 pp.
- Flor HH (1971) Current status of the gene-for-gene concept. *Annu Rev Phytopathol* 9:275–296.
- Food and Agriculture Organization (FAO) <http://www.fao.org> (Acesso em setembro 13, 2014).
- Freeman BC and Beattie GA (2008) An overview of plant defenses against pathogens and herbivores. APSnet Feature, Disponível em <<http://www.apsnet.org/publications/apsnetfeatures/Pages/Biotechnology.aspx>>.
- Freire-Filho FR, Ribeiro VQ, Barreto PD e Santos AA (2005) Melhoramento Genético. In: Freire-Filho FR e Lima JAA (eds) *Feijão-caupi: Avanços Tecnológicos*. Embrapa Informação Tecnológica, Brasília, DF, 29-92 pp.
- Freire-Filho FR, Ribeiro VQ, Rocha MM, Damasceno-Silva KJ, Nogueira MSRN e Rodrigues EV (2011) Produção, melhoramento genético e potencialidades do feijão-caupi no Brasil. IV Reunião de Biofortificação.

- Frota KMG, Morgano MA, Silva MG, Araújo MAM e Moreira-Araújo RSR (2009) Utilização da farinha de feijão-caupi (*Vigna unguiculata* L. Walp) na elaboração de produtos de panificação. *Ciênc Tecnol Aliment* 30(1):44-50.
- Goff AS, Ricke D, Lan TH, Presting G, Wang R, Dunn M and Glazebrook J (2002) A draft sequence of rice genome (*Oryza sativa* L. ssp. *japonica*), *Science* 296:92-100.
- Gonçalves MFB e Lima JAA (1982) Efeitos do "cowpea severe mosaic virus" sobre a produtividade do feijão-de-corda. *Fitopatol Bras* 7(547).
- Gou M, Su N, Zheng J, Huai J, Wu G, Zhao J, He J, Tang D, Yang S and Wang G (2009) An F-box gene, CPR30, functions as a negative regulator of the defense response in *Arabidopsis*. *Plant J* 60:757–770.
- Guimarães PM, Proite k, Leal-Bertioli SC e Bertioli D (2005) Análise *in silico* da expressão gênica diferencial de *Arachis stenosperma* inoculado com *Meloidogyne arenaria*. *Boletim de Pesquisa e Desenvolvimento/ EMBRAPA Recursos Genéticos e Biotecnologia*, 85.
- Guo YL, Fitz J, Schneeberger K, Ossowski S, Cao J and Weigel D (2011) Genome-wide comparison of nucleotide-binding site-leucine-rich-repeat-encoding genes in *Arabidopsis*. *Plant Physiol* 157:757–769.
- Hammond-Kosack KE and Jones JDG (1997) Plant disease resistance genes. *Annu Rev Plant Physiol*, 48:575-607.
- Holub EB (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. *Nature Rev Genet* 2:516-527.
- International Rice Research Institute (IRRI), <http://irri.org/> (Acesso em Novembro, 2013).
- Jones DG (2001) Putting the knowledge of plant disease resistance genes to work. *Curr Opin Plant Biol* 4: 281-287.
- Kobe B and Kajava AV (2001) The leucine-rich repeat as a protein recognition motif. *Curr Opin in Struct Biol* 11:725.
- Lacock L, Van-Niekerk C, Loots S, Du Preez F and Botha AM (2003) Functional and comparative analysis of expressed sequences from *Diuraphis noxia* infested wheat obtained utilizing the conserved Nucleotide Binding Site. *Afr J Biotechnol* 2:75-81.
- Lewis GP, Schrire B, Mackinder B and Lock M (2005) *Legumes of the world*. Kew: Royal Botanic Gardens, 577 pp.
- Locatelli MT, Palm CA, Smyth TJ e Ricci MSF (1991) Seleção de leguminosas para cultivo *alley-cropping* sob condições de Latossolo Amarelo em Porto Velho. EMBRAPA. CPAF Rondônia, 125.
- Loraine AE, McCormick S, Estrada A, Patel K and Qin P (2013) RNA-seq of *Arabidopsis* pollen uncovers novel transcription and alternative splicing. *Plant Physiol* 162:1092–1109.
- MAPA (Ministério da Agricultura e Pecuária), <http://goo.gl/QPrkHL>. Acessado em 15 de setembro, 2014.
- Marone D, Russo MA[†], Laidò G, Leonardis AM and Mastrangelo AM (2013) Plant Nucleotide Binding Site-Leucine-Rich Repeat (NBS-LRR) Genes: Active Guardians in Host Defense Responses. *Int J Mol* 14:7302-7326.
- Matsumura H, Reich S, Reuter M, Krüger DH, Winter P, Kahl G and Terauchi R (2004) SuperSAGE: A Potent Transcriptome Tool for Eukaryotic Organisms Tool for Eukaryotic Organisms. In: San Ming Wang (eds) *Current Techn App* 77-90 pp.

- McDowell JM and Simon SA (2006) Recent insights into *R* gene evolution. *Mol Plant Pathol* 7(5):437-48.
- McHale L, Tan X, Koehl P and Michelmore CW (2006) Plant NBS-LRR proteins: adaptable guards Review. *Genome Biol* 7(212).
- Métraux JP (2001) Systemic acquired resistance and salicylic acid: current state of knowledge. *Eur J Plant Pathol* 107:13-18.
- Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW and Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J* 20:317–332.
- Miller RNG[†], Bertoli DJ, Baurens FC, Santos CMR, Alves PC, Martins NF, Togawa RC, Souza-Júnior MT and Pappas-Júnior GJ (2008) Analysis of non-TIR NBS-LRR resistance gene analogs in *Musa acuminata* Colla: Isolation, RFLP marker development, and physical mapping. *BMC Plant Biol* 8(15).
- Moffett P (2009) Mechanisms of recognition in dominant *R* gene mediate resistance. *Adv Virus Res* 75:1-33.
- Mondragon-Palomino M, Meyers BC, Michelmore RW and Gaut BS (2002) Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Res* 12:1305–1315.
- Morin R, Bainbridge M, Fejes A, Hirst M, Krzywinski M, Pugh TJ, McDonald H, Varhol R, Jones SJM and Marra MA (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* 45:81-94.
- Padulosi S and Ng NQ (1997) Origin taxonomy, and morphology of *Vigna unguiculata* (L.) Walp. In: Singh BB, Mohan R, Dashiell KE and Jackai LEN (ds.eds). *Advances in Cowpea Research*, IITA JIRCAS, Tsukuba, 1-12 pp.
- Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Comp Appl Biosci* 12:357-358.
- Palomino C, Satovic Z, Cubero JI and Torres AM (2006) Identification and characterization of NBS–LRR class resistance gene analogs in faba bean (*Vicia faba* L.) and chickpea (*Cicer arietinum* L.). *Genome* 49(10):1227-37.
- Patel A (2014) Isolation, characterization and production of a new recombinant lectin protein from leguminous plants. *Biochemical Compounds*, ISSN 2052-9341.
- Pio-Ribeiro G and Assis FM (1997) Doenças do caupi (*Vigna unguiculata* (L.) Walp.). In: Kimari H, Amorim L, Bergamin-Filho A, Camargo LEA and Rezende JAM (eds) *Manual de Fitopatologia*. Agronômica Ceres, São Paulo, 233-244 pp.
- Ponnusha BS, Subramaniyam S, Pasupathi P, Subramaniyam B and Virumandy R (2011) Antioxidant and Antimicrobial properties of *Glycine Max*-A review. *Int J Cur Bio Med Sci* 1:49–62.
- Porter BW, Paidi M, Ming R, Alam M, Nishijima WT and Zhu YJ (2009) Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. *Mol Genet Genomics* 281:609–626.
- Ren J, Yu Y, Gao F, Zeng L, Lu X, Wu X, Yan W, and Ren G (2013) Application of resistance gene analog markers to analyses of genetic structure and diversity in rice. *Genome* 56:377–387.
- Rios GP (1988) Doenças fúngicas e bacterianas do caupi. In: Araújo JPP and Watt EE (eds) *O Caupi no Brasil*. IITA/EMBRAPA, Brasília, 547-589 pp.
- Sales MP, Andrade LBS, Ary MB, Miranda MRA, Teixeira FM, Oliveira AS, Fernandes KVS and Xavier-Filho J (2005) Performance of bean bruchids

- Callosobruchus maculatus* and *Zabrotes subfasciatus* (Coleoptera: Bruchidae) reared on resistant (IT81D-1045) and susceptible (Epace 10) *Vigna unguiculata* seeds: Relationship with trypsin inhibitor and vicilin excretion. *Comp Biochem Physiol A Mol Integr Physiol* 142:422-426.
- Sanabria NM, Huang JC and Dubery IA (2010) Self/nonself perception in plants in innate immunity and defense. *Land Biosc* 11:40-54.
- Sanseverino W, Roma G, De Simone M, Faino L, Melito S, Stupka E, Frusciante L and Ercolano MR (2010) PRGdb: a bioinformatics platform for plant resistance gene analysis. *Nucl Acids Res* 38:D814-D821.
- Severin AJ, Woody JL, Bolon Y-T, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE *et al.* (2010) RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biol* 10:160–176.
- Shang J, Tao Y, Chen X, Zou Y, Lei C, Wang J, Li X, Zhao X, Zhang M, Lu Z *et al.* (2009) Identification of a new rice blast resistance gene, *Pid3*, by genome wide comparison of paired nucleotide-binding site leucine-rich repeat genes and their pseudogene alleles between the two sequenced rice genomes. *Genetics* 182:1303–1311.
- Shao ZQ, Zhang YM, Hang YY, Xue JY, Zhou GC, Wu P, Wu XY, Wu XZ, Wang Q, Wang B *et al.* (2014) Long-Term Evolution of Nucleotide-Binding Site-Leucine-Rich Repeat Genes: Understanding Gained from and beyond the Legume Family. *Plant Physiol* 166:217–234.
- Shepherd M, Lee D, Turnbull P, Brown T, Bush D, Brawner J, Bartle J, Macdonell P and Henry RJ (2011) Eucalypts for biofuel production in Northern Australia: identifying species from current and future testing programs, Rural Industries Research & Development Corporation, Canberra, ACT. ISBN: 9781742542485
- Sievers F and Higgins DG (2014) UNIT 3.13 Clustal Omega. *Curr Protoc Bioinformatics* 12(3).
- Singh BB, Ehlers JD, Sharma B and Freire-Filho FR (2002) Recent progress in cowpea breeding. In: Fatokun CA, Tarawali SA, Singh BB, Kormawa PM, Tamo M (eds) *Challenges and opportunities for enhancing sustainable cowpea production*. IITA, Ibadan, 22-40 pp.
- Stakman EC (1915) Relation between puccinia graminis and plants highly resistant to its attack. *Journal of Agricultural Research*, IV:3.
- Sun X and Wang GL (2011) Genome-Wide Identification, Characterization and Phylogenetic Analysis of the Rice LRR-Kinases. *PLoS ONE* 6:e16079.
- Tameling WI, Nooijen C, Ludwig N, Boter M, Slootweg E, Goverse A, Shirasu K and Joosten MH (2010) RanGAP2 mediates nucleocytoplasmic partitioning of the NB-LRR immune receptor *Rx* in the Solanaceae, thereby dictating *Rx* function. *Plant Cell* 22:4176-4194.
- Tamura K, Dudley J, Nei M and Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599.
- Tarr DEK and Alexander HM (2009) TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. *BMC Res Notes* 2(197).
- Tsuda K and Katagiri F (2010) Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Curr Opin Plant Biol* 13:459-465.
- Tullu A, Tar'an B, Breitschultz C, Buchwaldt L, Banniza S, Warkentin TD and Vandenberg AA (2006) Quantitative-trait locus for resistance to ascochyta

- blight (*Ascochyta lentis*) map close to a gene for resistance to anthracnose (*Colletotrichum truncatum*) in lentil. Canadian J of Plant Pathol 28:588-595.
- Umaharan P, Ariyanayagan RP and Haque SQ (1997) Resistance to cowpea severe mosaic virus, determined by three dosage dependent genes in *Vigna unguiculata* (L.) Walp. Euphytica 95:49-55.
- Valenzuela H and Smith J (2002) Cowpea. In: Sustainable Agriculture Green Manure Crops. Cooperative Extension Service, College of Tropical Agriculture & Human resources, 1-3 pp.
- Vallad GE, Goodman RM (2004) Systemic Acquired Resistance and Induced Systemic Resistance in Conventional Agriculture. Crop Science, 44:1920–1934.
- Valliyodan B and Nguyen H (2008) Genomics of abiotic stress in soybean. In: Stacey, Gary (eds). Genetics and Genomics of Soybean, 1^a ed. Heidelberg: Springer, 18:343-372.
- Van-Ooijen G, Burg VD HA, Cornelissen BJ and Takken FL (2007) Structure and function of resistance proteins in solanaceous plants. Annu Rev Phytopathol 45:43–72.
- Wan H, Yuan W, Bo K, Shen J, Pang X and Chen J (2013) Genome-wide analysis of NBS-encoding disease resistance in *Cucumis sativus* and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. BMC Genomics 14(109).
- Wanderley-Nogueira AC, Kido EA, Soares-Cavalcanti NM, Bellarmino LC, Bezerra-Neto JP, Chabragas SM and Benko-Iseppon AM (2012) Insight on pathogen defense mechanisms in the sugarcane transcriptome. Func Plant Sci Biotech 6:134-148.
- Wang Z, Gerstein M and Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature Rev Genet 16.
- Wolfsberg TG and Landsman D (1997) A comparison of expressed sequence tags (ESTs) to human genomic sequences. Nucleic Acids Res 25:1626–1632.
- Young ND (2000) The genetic architecture of resistance. Curre Opin Plant Biol 3:285-290.

9. Anexos

Anexo 1: Trinta e cinco genes NBS-LRR previamente caracterizados em plantas, os quais foram alinhados contra o banco de dados do NordEST como sondas (*seed sequence*) para desenvolvimento desse estudo, com base em Sanseverino *et al.* (2010).

Planta	Gene R	ID Proteína	Estrutura	Planta	Gene R	ID Proteína	Estrutura
<i>Arabidopsis thaliana</i> L.	<i>Rpm1</i>	CAA61131	NBS-LRR	<i>Hordeum vulgare</i> L.	<i>Mla6</i>	CAC29242	NBS-LRR
<i>Arabidopsis thaliana</i> L.	<i>Rpp1</i>	AAC72977	TIR-NBS-LRR	<i>Hordeum vulgare</i> L.	<i>Mla1</i>	AAG37354	NBS-LRR
<i>Arabidopsis thaliana</i> L.	<i>Rpp4</i>	AAM18462	TIR-NBS-LRR	<i>Linum usitatissimum</i> L.	<i>L6</i>	AAA91022	TIR-NBS-LRR
<i>Arabidopsis thaliana</i> L.	<i>Rpp5</i>	AAF08790	TIR-NBS-LRR	<i>Linum usitatissimum</i> L.	<i>M</i>	AAB47618	TIR-NBS-LRR
<i>Arabidopsis thaliana</i> L.	<i>Rpp8</i>	AAC78631	NBS-LRR	<i>Linum usitatissimum</i> L.	<i>P2</i>	AAK28806	TIR-NBS-LRR
<i>Arabidopsis thaliana</i> L.	<i>Rpp13</i>	AAF42832	NBS-LRR	<i>Lactuca sativa</i> L.	<i>Dm3</i>	AAD03156	NBS-LRR
<i>Arabidopsis thaliana</i> L.	<i>Rps2</i>	AAM90883	NBS-LRR	<i>Nicotiana tabacum</i> L.	<i>N</i>	AAA50763	TIR-NBS-LRR
<i>Arabidopsis thaliana</i> L.	<i>Rps4</i>	CAB50708	TIR-LRR	<i>Nicotiana tabacum</i> L.	<i>Hero</i>	CAD29728	NBS-LRR
<i>Arabidopsis thaliana</i> L.	<i>Rps5</i>	AAC26126	NBS-LRR	<i>Zea mays</i> L.	<i>Rp1-d</i>	AAD47197	NBS-LRR
<i>Arabidopsis thaliana</i> L.	<i>Ssi4</i>	AAN86124	TIR-NBS-LRR	<i>Solanum lycopersicum</i> L.	<i>I2</i>	AAD27815	NBS-LRR
<i>Arabidopsis thaliana</i> L.	<i>Rcy1</i>	AAM13905	NBS-LRR	<i>Solanum lycopersicum</i> L.	<i>Mi-1</i>	AAC67238	NBS-LRR
<i>Solanum tuberosum</i> L.	<i>R1</i>	AAU95638	NBS-LRR	<i>Solanum lycopersicum</i> L.	<i>Prf</i>	AAC49408	NBS-LRR
<i>Solanum tuberosum</i> L.	<i>Rgc1</i>	AAF76163	NBS-LRR	<i>Solanum lycopersicum</i> L.	<i>Sw-5</i>	ACQ42910	NBS-LRR
<i>Solanum tuberosum</i> L.	<i>Rx</i>	CAB50786	NBS-LRR	<i>Triticum aestivum</i> L.	<i>Cre3</i>	CAD12795	NBS-LRR
<i>Oryza sativa</i> L.	<i>Pib</i>	BAA76282	NBS-LRR	<i>Triticum aestivum</i> L.	<i>VRGL1</i>	AAF19148	NBS-LRR
<i>Oryza sativa</i> L.	<i>Pi-ta</i>	BAF91352	NBS-LRR	<i>Triticum aestivum</i> L.	<i>Yr10</i>	AAG42168	NBS-LRR
<i>Oryza sativa</i> L.	<i>Rpr1</i>	BAA75812	NBS-LRR	<i>Capsicum frutescens</i> L.	<i>Bs2</i>	AAF09256	NBS-LRR
<i>Oryza sativa</i> L.	<i>Xa1</i>	BAA25068	NBS-LRR				

Anexo 2: Resultados obtidos a partir do BLASTx contra o NCBI, considerando um *e-value* de e^{-10} . Número de acesso, organismo, descrição, tamanho dos candidatos em nucleótidos (nt) e aminoácidos (aa), bem como os seus valores correspondentes *e-value* e *score*.

Candidatos	BLASTx			Tamanho		Padrões	
	Acesso	Organismo	Descrição	Nt	aa	e-value	score
Contig2	ESW30163.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	5033	577	0.0	897
Contig3	ESW34218.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	6532	868	0.0	1374
Contig5	ESW11941.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	5408	1084	0.0	1696
Contig7	ESW23024.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	4660	539	0.0	811
Contig14	ESW06173.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	2241	572	0.0	869
Contig16	ESW06623.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	2834	868	0.0	1417
Contig17	ESW24557.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3588	388	0.0	608
Contig19	ESW03443.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3368	814	0.0	1048
Contig23	ESW34190.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3212	763	0.0	1280
Contig27	ESW18560.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3856	880	0.0	942
Contig28	ESW23120.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	6863	1237	0.0	899
Contig29	ESW32809.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	5891	824	0.0	1608
Contig35	ESW24554.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	4185	459	0.0	887
Contig38	ESW04962.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	5062	831	0.0	1122
Contig39	ESW23433.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	4159	536	0.0	911
Contig42	ESW06510.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	4506	856	0.0	443
Contig43	ESW23433.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3678	518	0.0	799
Contig49	XP_003547433.1	<i>Glycine max</i>	Protein Supressor OF npr1-1, Constitutive 1-like	4535	538	0.0	1162
Contig52	ESW29499.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3312	527	0.0	789
Contig56	ESW06168.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	5041	879	0.0	1536
Contig60	ESW24555.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	4387	748	0.0	908
Contig61	XP_003526348.1	<i>Glycine max</i>	Disease resistance protein RPM1-like	8542	1014	0.0	1494
Contig68	ESW11258.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	7068	885	0.0	1592
Contig69	ESW32444.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	5119	873	0.0	1022
Contig85	ESW30597.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	4157	753	0.0	1440
Contig98	NP_001235633.1	<i>Glycine max</i>	TIR-NBS-LRR type disease resistance protein	2820	810	0.0	1080
Contig105	ESW24484.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3763	1039	0.0	1567
Contig107	ESW07460.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	2639	777	0.0	1081
Contig112	ESW13426.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	4791	1324	0.0	1431
Contig115	ESW07099.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3811	1120	0.0	1965
Contig131	XP_003518448.2	<i>Glycine max</i>	TMV resistance protein N-like	4334	1106	0.0	1648
Contig132	XP_003522058.1	<i>Glycine max</i>	TMV resistance protein N-like	3769	1077	0.0	1548

Contig133	ESW20934.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3594	875	0.0	1375
Contig136	XP_006603606.1	<i>Glycine max</i>	Disease resistance protein RPP13-like	3313	850	0.0	868
Contig139	ESW05466.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	2095	350	0.0	601
Contig146	ESW18167.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	4091	1236	0.0	2191
Contig147	ESW24484.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3991	568	0.0	872
Contig148	ESW30226.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3867	656	0.0	1048
Contig151	XP_003551061.1	<i>Glycine max</i>	Probable disease resistance protein	3208	805	0.0	1165
Contig155	XP_007156839.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	2461	658	0.0	982
Contig164	XP_007153879.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	4353	1237	0.0	2018
Contig165	XP_007148989.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	4313	1168	0.0	1983
Contig166	XP_007158212.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3408	945	0.0	1764
Contig167	XP_003629563.1	<i>Medicago truncatula</i>	LRR protein soc-2-like protein	3136	564	0.0	665
Contig181	XP_007152561.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	4645	870	0.0	790
Contig182	XP_007155962.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3595	837	0.0	1483
Contig184	XP_003598470.1	<i>Medicago truncatula</i>	Resistance protein	3024	715	0.0	867
Contig208	XP_007148814.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	4158	1344	0.0	1943
Contig211	XP_007139456.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3708	932	0.0	1687
Contig214	XP_007158209.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	3236	942	0.0	1698
Contig221	XP_003534888.1	<i>Glycine max</i>	Putative disease resistance protein	2386	441	0.0	563
Contig226	XP_007154567.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	1857	560	0.0	1002
Contig256	AAO23074.1	<i>Glycine max</i>	R 10 protein	2732	645	8,00e ⁻¹⁵⁹	501
Contig265	XP_007135470.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	1760	570	0.0	984
SI44383.1	XP_007157781.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	1264	414	0.0	748
PO18003.13	XP_007152561.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	4801	724	0.0	1058
SI44421.1	XP_007138746.1	<i>Phaseolus vulgaris</i>	HypotHetical protein	2205	635	0.0	1144

Anexo 3: Instruções para autores: *Current Protein & Peptide Science*

ONLINE MANUSCRIPT SUBMISSION:

An online submission and tracking service *via* Internet facilitates a speedy and cost-effective submission of manuscripts. The full manuscript has to be submitted online via Bentham's Content Management System (CMS) at bsp-cms.eurekaselect.com / **View Submission Instructions**.

Manuscripts must be submitted by one of the authors of the manuscript, and should not be submitted by anyone on their behalf. The principal/corresponding author will be required to submit a Copyright Letter along with the manuscript, on behalf of all the co-authors (if any). The author(s) will confirm that the manuscript (or any part of it) has not been published previously or is not under consideration for publication elsewhere. Furthermore, any illustration, structure or table that has been published elsewhere must be reported, and copyright permission for reproduction must be obtained.

For all online submissions, please provide soft copies of all the materials (main text in MS Word or Tex/LaTeX), figures / illustrations in TIFF, PDF or JPEG, and chemical structures drawn in ChemDraw (CDX) / ISISDraw (TGF) as separate files, while a PDF version of the entire manuscript must also be included, embedded with all the figures / illustrations / tables / chemical structures etc. It is advisable that the document files related to a manuscript submission should always have the name of the corresponding author as part of the file name, i.e., "Cilli MS text.doc", "Cilli MS Figure 1", etc.

It is imperative that before submission, authors should carefully proofread the files for special characters, mathematical symbols, Greek letters, equations, tables, references and images, to ensure that they appear in proper format.

References, figures, tables, chemical structures *etc.* should be referred to in the text at the appropriate place where they have been first discussed. Figure legends/captions should also be provided.

A successful electronic submission of a manuscript will be followed by a system-generated acknowledgement to the principal/corresponding author. Any queries therein should be addressed to ilyas@benthamscience.org

Editorial Policies:

The editorial policies of Bentham Science Publishers on publication ethics, peer-review, plagiarism, copyrights/ licenses, errata/corrections, and article retraction/ withdrawal can be viewed at [Editorial Policy](#)

MANUSCRIPTS PUBLISHED:

The Journal accepts mini- and full-length review articles written in English. Single topic/thematic issues may also be considered for publication.

Single Topic Issues:

These special issues are peer-reviewed and may contain invited or uninvited review/mini-review articles. Single Topic Special Editor will offer a short Perspective and coordinate the solicitation of manuscripts between 5-8 (for a mini-thematic issue) to 8-12 (for full-length thematic issue) from leading scientists. Authors interested in editing a single topic issue in an emerging topic of protein and peptide science may submit their proposal to the Editor-in-Chief at atcps@benthamscience.org for consideration. Visit www.benthamscience.com/cpps/Special-Issues.htm for viewing forthcoming issues.

Conference Proceedings:

For proposals to publish conference proceedings in this journal, please contact us at email: proceedings@benthamscience.org.

MANUSCRIPT LENGTH:

Review Articles:

The total number of words for a published comprehensive review article is from 8000 to 40000 words, and for mini-review articles from 3000 to 6000 words.

There is no restriction on the number of figures, tables or additional files e.g. video clips, animation and datasets, that can be included with each article online. Authors should include all relevant supporting data with each article (Refer to Supplementary Material section).

MANUSCRIPT PREPARATION:

The manuscript should be written in English in a clear, direct and active style. All pages must be numbered sequentially, facilitating in the reviewing and editing of the manuscript.

MICROSOFT WORD TEMPLATE:

It is advisable that authors prepare their manuscript using the template available on the Web, which will assist in preparation of the manuscript according to Journal's Format. [Download the Template](#).

Our contracted service provider [Eureka Science](#) can, if needed, provide professional assistance to authors for the improvement of English language and figures in manuscripts.

MANUSCRIPT SECTIONS FOR PAPERS:

Manuscripts may be divided into the following sections:

- Copyright Letter
- Title
- Title page
- Abstract
- Graphical Abstract
- Keywords
- Text Organization
- List of Abbreviations (if any)
- Conflict of Interest
- Acknowledgements
- References
- Appendices
- Figures/Illustrations (if any)
- Chemical Structures (if any)
- Tables (if any)
- Supportive/Supplementary Material (if any)

Copyright Letter:

It is mandatory that a signed copyright letter also be submitted along with the manuscript by the author to whom correspondence is to be addressed, delineating the scope of the submitted article declaring the potential competing interests, acknowledging contributions from authors and funding agencies, and certifying that the paper is prepared according to the 'Instructions for Authors'. All inconsistencies in the text

and in the reference section, and any typographical errors must be carefully checked and corrected before the submission of the manuscript. The article should not contain any such material or information that may be unlawful, defamatory, fabricated, plagiarized, or which would, if published, in any way whatsoever, violate the terms and conditions as laid down in the copyright agreement. The authors acknowledge that the publishers have the legal right to take appropriate action against the authors for any such violation of the terms and conditions as laid down in the copyright agreement. Copyright letter can be downloaded from the journal's Web site. [Download the Copyright letter](#)

Title:

The title of the article should be precise and brief and must not be more than 120 characters. Authors should avoid the use of non-standard abbreviations. The title must be written in title case except for articles, conjunctions and prepositions.

Authors should also provide a short 'running title'. Title, running title, byline, correspondent footnote and keywords should be written as presented in original manuscripts.

Title Page:

Title page should include paper title, author(s) full name and affiliation, corresponding author(s) names complete affiliation/address, along with phone, fax and email.

Abstract:

The abstract should not exceed 250 words for review papers summarizing the essential features of the article.

Graphical Abstract:

A graphic must be included with each manuscript for use in the Table of Contents (TOC). This must be submitted separately as an electronic file (preferred file types are EPS, PDF, TIFF, Microsoft Word, PowerPoint and CDX etc.). A graphical abstract, not exceeding 30 words along with the illustration, helps to summarize the contents of the manuscript in a concise pictorial form. It is meant as an aid for the rapid viewing of the journals' contents and to help capture the readers' attention. The graphical abstract may feature a key structure, reaction, equation, etc. that the manuscript elucidates upon. It will be listed along with the manuscript title, authors' names and affiliations in the contents page, typeset within an area of 5 cm by 17 cm, but it will not appear in the article PDF file or in print.

Graphical Abstracts should be submitted as a separate file (must clearly mention graphical abstract within the file) online via Bentham's Content Management System by selecting the option "supplementary material".

Keywords:

Provide 6 to 8 keywords in alphabetical order.

Text Organization:

The main text should begin on a separate page and should be divided into separate sections. The manuscript should be divided into title page, abstract and the main text. The text may be subdivided further according to the areas to be discussed, which should be followed by the Acknowledgements and Reference sections. The review article should mention any previous important recent and old reviews in the field and contain a comprehensive discussion starting with the general background of the field. It should then go on to discuss the salient features of recent developments. The authors should avoid presenting material which has already been published in a previous review. The authors are advised to present and discuss their observations in brief. The manuscript style must be uniform throughout the text and 10 pt Times New Roman fonts should

be used. The full term for an abbreviation should precede its first appearance in the text unless it is a standard unit of measurement. The reference numbers should be given in square brackets in the text. Italics should be used for Binomial names of organisms (Genus and Species), for emphasis and for unfamiliar words or phrases. Non-assimilated words from Latin or other languages should also be italicized e.g. *per se*, *et al.* etc.

Standard Protocol on Approvals, Registrations, Patient Consents & Animal Protection:

All clinical investigations must be conducted according to the Declaration of Helsinki principles. For all manuscripts reporting data from studies involving human participants, formal review and approval by an appropriate institutional review board or ethics committee is required. For research involving animals, the authors should indicate whether the procedures followed were in accordance with the standards set forth in the eighth edition of *Guide for the Care and Use of Laboratory Animals* (grants.nih.gov/grants/olaw/guide-for-the-care-and-use-of-laboratory-animals_prepub.pdf); published by the National Academy of Sciences, The National Academies Press, Washington, D.C.).

A specific declaration of such approval must be made in the copyright letter and in a stand-alone paragraph at the end of the Methods section especially in the case of human studies where inclusion of a statement regarding obtaining the written informed consent from each subject or subject's guardian is a must. The original should be retained by the guarantor or corresponding author. Editors may request to provide the original forms by fax or email.

Greek Symbols and Special Characters:

Greek symbols and special characters often undergo formatting changes and get corrupted or lost during preparation of manuscript for publication. To ensure that all special characters used are embedded in the text, these special characters should be inserted as a symbol but should not be a result of any format styling (*Symbol* font face) otherwise they will be lost during conversion to PDF/XML.

Authors are encouraged to consult reporting guidelines. These guidelines provide a set of recommendations comprising a list of items relevant to their specific research design. Chemical equations, chemical names, mathematical usage, unit of measurements, chemical and physical quantity & units must conform to SI and Chemical Abstracts or IUPAC.

All kinds of measurements should be reported only in International System of Units (SI).

List of Abbreviations:

If abbreviations are used in the text either they should be defined in the text where first used, or a list of abbreviations can be provided.

Conflict of Interest:

Financial contributions to the work being reported must be clearly acknowledged, as should any potential conflict of interest under the heading 'Conflict of Interest'. Authors must list the source(s) of funding for the study, for each author, and for the manuscript preparation.

Acknowledgements:

All individuals listed as authors must have contributed substantially to the design, performance, analysis, or reporting of the work and are required to indicate their specific contribution. Anyone (individual/company/institution) who has substantially contributed to the study for important intellectual content, or who was involved in the article's drafting the manuscript or revising must also be acknowledged.

Guest or honorary authorship based solely on position (e.g. research supervisor, departmental head) is discouraged.

The specific requirements for authorship have been defined by the International Committee of Medical Journal Editors (ICMJE; www.icmje.org). Examples of authors' contributions are: 'designed research/study', 'performed research/study', 'contributed important reagents', 'collected data', 'analyzed data', 'wrote paper' etc. This information must be included in the submitted manuscript as a separate paragraph under the heading 'Acknowledgements'. The corresponding author is responsible for obtaining permission from all co-authors for the submission of any version of the manuscript and for any changes in the authorship.

References:

References must be listed in the ACS Style only. All references should be numbered sequentially [in square brackets] in the text and listed in the same numerical order in the reference section. The reference numbers must be finalized and the bibliography must be fully formatted before submission.

See below few examples of references listed in the ACS Style:

Journal Reference:

- [1] Bard, M.; Woods, R.A.; Bartón, D.H.; Corrie, J.E.; Widdowson, D.A. Sterol mutants of *Saccharomyces cerevisiae*: chromatographic analyses. *Lipids*, **1977**, 12(8), 645-654.
- [2] Zhang, W.; Brombosz, S.M.; Mendoza, J.L.; Moore, J.S. A high-yield, one-step synthesis of o-phenylene ethynylene cyclic trimer via precipitation-driven alkyne metathesis. *J. Org. Chem.*, **2005**, 70, 10198-10201.

Book Reference:

- [3] Crabtree, R.H. *The Organometallic Chemistry of the Transition Metals*, 3rd ed.; Wiley & Sons: New York, **2001**.

Book Chapter Reference:

- [4] Wheeler, D.M.S.; Wheeler, M.M. In: *Studies in Natural Products Chemistry*; Atta-ur-Rahman, Ed.; Elsevier Science B. V: Amsterdam, **1994**; Vol. 14, pp. 3-46.

Conference Proceedings:

- [5] Jakeman, D.L.; Withers, S.G.E. In: *Carbohydrate Bioengineering: Interdisciplinary Approaches*, Proceedings of the 4th Carbohydrate Bioengineering Meeting, Stockholm, Sweden, June 10-13, 2001; Teeri, T.T.; Svensson, B.; Gilbert, H.J.; Feizi, T., Eds.; Royal Society of Chemistry: Cambridge, UK, **2002**; pp. 3-8.

URL(WebPage):

- [6] National Library of Medicine. Specialized Information Services: Toxicology and Environmental Health. sis.nlm.nih.gov/Tox/ToxMain.html (Accessed May 23, **2004**).

Patent:

- [7] Hoch, J.A.; Huang, S. Screening methods for the identification of novel antibiotics. U.S. Patent 6,043,045, March 28, **2000**.

Thesis:

- [8] Mackel, H. *Capturing the Spectra of Silicon Solar Cells*. PhD Thesis, The Australian National University: Canberra, December **2004**.

E-citations:

- [9] Citations for articles/material published exclusively online or in open access (free-to-view) , must contain the exact Web addresses (URLs) at the end of the reference(s), except those posted on an author's Web site unless editorially essential, e.g. 'Reference: Available from: URL'.

Some important points to remember:

- All references must be complete and accurate.
- All authors must be cited and there should be no use of the phrase *et al*.
- Online citations should include the date of access.
- Journal abbreviations should follow the *Index Medicus/MEDLINE*.
- Special care should be taken of the punctuation convention as described in the above-mentioned examples.
- Superscript in the in-text citations and reference section should be avoided.
- Abstracts, unpublished data and personal communications (which can only be included if prior permission has been obtained) should not be given in the reference section but they may be mentioned in the text and details provided as footnotes.
- The authors are encouraged to use a recent version of EndNote (version 5 and above) or Reference Manager (version 10) when formatting their reference list, as this allows references to be automatically extracted.

Appendices:

In case there is a need to present lengthy, but essential methodological details, appendices must be used, which can be a part of the article. An appendix must not exceed three pages (Times New Roman, 12 point fonts, 900 max. words per page).The information should be provided in a condensed form, ruling out the need of full sentences. A single appendix should be titled APPENDIX, while more than one can be titled APPENDIX A, APPENDIX B, and so on.

Figures/Illustrations:

All authors must strictly follow the guidelines below for preparing illustrations for publication in ***Current Protein & Peptide Science***. If the figures are found to be sub-standard, then the manuscripts will be rejected and the authors offered the option of figure improvement professionally by Eureka Science. The costs for such improvement will be charged to the authors.

The authors should provide the illustrations as separate files, as well as embedded in the text file, numbered consecutively in the order of their appearance. Each figure should include a single illustration. Each figure should be closely cropped to minimize the amount of white space surrounding the illustration.

If a figure consists of separate parts, it is important that a single composite illustration file be submitted, containing all parts of the figure.

Photographs should be provided with a scale bar if appropriate, as well as high-resolution component files.

Scaling/Resolution:

For Line Art image type, which is generally an image based on lines and text and does not contain tonal or shaded areas, the preferred file format is TIFF or EPS, with colour mode being Monochrome 1-bit or RGB, in a resolution of 900-1200 dpi.

For Halftone image type, which is generally a continuous tone photograph and contains no text, the preferred file format is TIFF, with colour mode being of RGB or Grayscale, in a resolution of 300 dpi.

For Combination image type, which is generally an image containing halftone in addition to text or line art elements, the preferred file format is TIFF, with colour mode being of RGB or Grayscale, in a resolution of 500-900 dpi.

Formats:

For illustrations, the following file formats are acceptable:

- **Illustrator**
- **EPS** (preferred format for diagrams)
- **PDF** (also especially suitable for diagrams)
- **PNG** (preferred format for photos or images)
- **Microsoft Word** (version 5 and above; figures must be a single page)
- **PowerPoint** (figures must be a single page)
- **TIFF**
- **JPEG** (conversion should be done using the original file)
- **BMP**
- **CDX** (ChemDraw)
- **TGF** (ISISDraw)

Bentham Science does not process figures submitted in GIF format.

If the large size of TIFF or EPS figures acts as an obstacle to online submission, authors may find that conversion to JPEG format before submission results in significantly reduced file size and upload time, while retaining acceptable quality. JPEG is a 'lossy' format, however in order to maintain acceptable image quality, it is recommended that JPEG files are saved at High or Maximum quality.

Files should not be compressed with tools such as Zipit or Stuffit prior to submission as these tools will in any case produce negligible file-size savings for JPEGs and TIFFs, which are already compressed.

Please do not:

1. Supply embedded graphics in your word processor (spreadsheet, presentation) document;
2. Supply files that are optimized for screen use (like GIF, BMP, PICT, WPG); the resolution is too low;
3. Supply files that are too low in resolution;
4. Submit graphics that are disproportionately large for the content.

Image Conversion Tools:

There are many software packages, many of them freeware or shareware, capable of converting to and from different graphics formats, including PNG.

Good general tools for image conversion include GraphicConverter on the Macintosh, PaintShop Pro, for Windows, and ImageMagick, which is available on Macintosh, Windows and UNIX platforms.

Note that bitmap images (e.g. screenshots) should not be converted to EPS, since this will result in a much larger file size than the equivalent JPEG, TIFF, PNG or BMP, with no increase in quality. EPS should only be used for images produced by vector-drawing applications such as Adobe Illustrator or CorelDraw. Most vector-drawing applications can be saved in, or exported as, EPS format. In case the images have been originally prepared in an Office application, such as Word or PowerPoint, then the original Office files should

be directly uploaded to the site, instead of being converted to JPEG or another format that may be of low quality.

Color Figures/Illustrations:

- The cost for each individual page of color figures/plates/illustrations is **US\$ 950**.
- Color figures should be supplied in CMYK not RGB colors.

Chemical Structures:

Chemical structures must be prepared in ChemDraw/CDX and provided as separate file.

Structure Drawing Preferences:

[As according to the ACS style sheet]

Drawing Settings:

Chain angle	120°
Bond spacing	18% of width
Fixed length	14.4 pt (0.500cm, 0.2in)
Bond width	2.0 pt (0.071cm, 0.0278in)
Line width	0.6 pt (0.021cm, 0.0084in)
Margin width	1.6 pt (0.096cm)
Hash spacing	2.5 pt (0.088cm, 0.0347in)

Text settings:

Font	Times New Roman / Helvetica
Size	8 pt

Under the Preference Choose:

Units	Points
Tolerances	3 pixels

Under Page Setup Use:

Paper	US letter
Scale	100%

Tables:

- Data Tables should be submitted in Microsoft Word table format.
- Each table should include a title/caption being explanatory in itself with respect to the details discussed in the table. Detailed legends may then follow.
- Table number in bold font *i.e.* Table 1, should follow a title. The title should be in small case with the first letter in caps. A full stop should be placed at the end of the title.
- Tables should be embedded in the text exactly according to their appropriate placement in the submitted manuscript.
- Columns and rows of data should be made visibly distinct by ensuring that the borders of each cell are displayed as black lines.
- Tables should be numbered in Arabic numerals sequentially in order of their citation in the body of the text.
- If a reference is cited in both the table and text, please insert a lettered footnote in the table to refer to the numbered reference in the text.
- Tabular data provided as additional files can be submitted as an Excel spreadsheet.

Supportive/Supplementary Material:

We do encourage to append supportive material, for example a PowerPoint file containing a talk about the study, a PowerPoint file containing additional screenshots, a Word, RTF, or PDF document showing the original instrument(s) used, a video, or the original data (SAS/SPSS files, Excel files, Access Db files etc.) provided it is inevitable or endorsed by the journal's Editor.

Supportive/Supplementary material intended for publication must be numbered and referred to in the manuscript but should not be a part of the submitted paper. In-text citations as well as a section with the heading "Supportive/Supplementary Material" before the "References" section should be provided. Here, list all Supportive/Supplementary Material and include a brief caption line for each file describing its contents.

Any additional files will be linked to the final published article in the form supplied by the author, but will not be displayed within the paper. They will be made available in exactly the same form as originally provided only on our Web site. Please also make sure that each additional file is a single table, figure or movie (please do not upload linked worksheets or PDF files larger than one sheet). Supportive/ Supplementary material must be provided in a single zipped file not larger than 4 MB.

Authors must clearly indicate if these files are not for publication but meant for the reviewers'/editors' perusal only.

PERMISSION FOR REPRODUCTION:

Bentham Science has collaborated with the Copyright Clearance Center to meet our customer's licensing, besides rights & permission needs.

The Copyright Clearance Center's RightsLink® service makes it faster and easier to secure permission from Bentham Science's journal titles. Simply visit [Journals by Title](#) and locate the desired content. Then go to the article's abstract and click on "Rights and Permissions" to open the RightsLink's page. If you are unable to locate the content you wish to use or you are unable to secure the rights you are seeking, please e-mail us at permissions@benthamscience.org

Published/reproduced material should not be included unless written permission has been obtained from the copyright holder, which should be forwarded to the Editorial Office in case of acceptance of the article for publication.

AUTHORS AND INSTITUTIONAL AFFILIATIONS:

The author will be required to provide their full names, the institutional affiliations and the location, with an asterisk in front of the name of the principal/corresponding author. The corresponding author(s) should be designated and their complete address, business telephone and fax numbers and e-mail address must be stated to receive correspondence and galley proofs.

PAGE CHARGES:

No page charges will be levied to authors for the publication of their review articles.

LANGUAGE AND EDITING:

Manuscripts submitted containing many English typographical errors will not be published. Manuscripts which are accepted for publication on condition that the written English submitted is corrected, will be sent a quote by Eureka Science, a professional language editing company. Authors from non-English language countries who have poor English language written skills, are advised to contact the language editing company prior to submitting their manuscript to the journal. Please contact Eureka Science for a language editing quote at e-mail: info@eureka-science.com stating the total number of words of the article to be edited.

언어 및 편집:

영문 오타가 많은 원고는 출판되지 않을 것입니다. 영문 오타를 없애겠다는 조건으로 받은 원고는 영어 편집 전문회사인 유럽 공동 기술개발 기구로부터 가격 견적서가 보내 질 것입니다. 영어 작문에 어려움이 있는 비영어권 국가의 저자들은 원고를 학술지에 제출하기 전에 영어 편집회사와 접촉할 것을 권합니다. 영어 편집 견적서를 받기 위해서 교정될 원고의 단어수를 적은 메일을 유럽 공동 기술개발 기구 메일인 info@eureka-science.com 로 보내시기 바랍니다.

语言和编辑:

含有很多英文印刷错误的提交稿将不予发表。接受发表的稿件其英文写作应是正确的；专业的语言编辑公司（尤里卡科学），可对稿件的英文润色提供报价。建议非英语国家、且英文写作欠佳的作者在投稿前先与语言编辑公司联系。请与尤里卡科学联系 info@eureka-science.com。

EDITION ET LANGUE:

Les manuscrits soumis avec plusieurs erreurs typographiques en Anglais ne seront pas publiés en l'état. Les manuscrits sont acceptés pour publication à la condition que l'anglais utilisé soit corrigé après la soumission et seront envoyés pour examen à Eureka Science, une société d'édition de langue professionnelle. Les auteurs en provenance de pays où la langue est différente de l'anglais et qui ont de médiocres compétences en anglais écrit, sont priés de contacter la société d'édition de langue avant de soumettre leur manuscrit à la revue. Merci de contacter Eureka Science à info@eureka-science.com pour un devis en indiquant le nombre total de mot de l'article à éditer.

PROOF CORRECTIONS:

Authors will receive page proofs of their accepted paper before publications. To avoid delays in publication, proofs should be checked immediately for typographical errors and returned **within 48 hours**. Major changes are not acceptable at the proof stage. If unable to send corrections **within 48 hours** due to some reason, the author(s) must at least send an acknowledgement on receiving the galley proofs or the article will be

published exactly as received and the publishers will not be responsible for any error occurring in the published manuscript in this regard.

The corresponding author will be solely responsible for ensuring that the revised version of the manuscript incorporating all the submitted corrections receives the approval of all the co-authors of the manuscript.

REPRINTS:

Printed reprints and e-prints may be ordered from the Publisher prior to publication of the article. First named authors may also order a personal print and online subscription of the journal at 50% off the normal subscription rate by contacting the subscription department at e-mail: subscriptions@benthamscience.org

OPEN ACCESS PLUS:

Accepted articles can be published online for free open access for all to view. Open access publishing provides the maximum dissemination of the article to the largest audience. Authors must pay for this service. All authors will be asked to indicate whether or not they wish to pay to have their paper made freely available on publication. If authors do not select the "Open Access Plus" option, then their article will be published with standard subscription-based access at no charge, subject always to the applicable Standard Terms & Conditions and Copyright Letter signed by the author.

Bentham Science offers authors the choice of "Open Access Plus" publication of their articles at a fee of **US\$ 2,900** per published article.

All editors, board members and those authors who have contributed more than two articles in Bentham Science publications are entitled to a **40% discount** on "Open Access Plus" fees.

For more information please contact us at e-mail: openaccess@benthamscience.org.

REVIEWING AND PROMPTNESS OF PUBLICATION:

All manuscripts submitted for publication will be immediately subjected to peer-reviewing, usually in consultation with the members of the Editorial Advisory Board and a number of external referees. Authors may, however, provide in their Copyright Letter the contact details (including e-mail addresses) of four potential peer reviewers for their paper. Any peer reviewers suggested should not have recently published with any of the authors of the submitted manuscript and should not be members of the same research institution.

All peer-reviewing will be conducted *via* the Internet to facilitate rapid reviewing of the submitted manuscripts. Every possible effort will be made to assess the manuscripts quickly with the decision being conveyed to the authors in due course.

COPYRIGHT:

Authors who publish in Bentham Science print & online journals will transfer copyright to their work to **Bentham Science Publishers**. Submission of a manuscript to the respective journals implies that all authors have read and agreed to the content of the Copyright Letter or the Terms and Conditions. It is a condition of publication that manuscripts submitted to this journal have not been published and will not be simultaneously submitted or published elsewhere. Plagiarism is strictly forbidden, and by submitting the article for publication the authors agree that the publishers have the legal right to take appropriate action against the authors, if plagiarism or fabricated information is discovered. By submitting a manuscript the authors agree that the copyright of their article is transferred to the publishers if and when the article is accepted for publication. Once submitted to the journal, the author will not withdraw their manuscript at any stage prior to publication.

SELF-ARCHIVING

By signing the Copyright Letter the authors retain the rights of self-archiving. Following are the important features of self-archiving policy of Bentham Science journals:

1. Authors can deposit the first draft of a submitted article on their personal websites, their institution's repositories or any non-commercial repository for personal use, internal institutional use or for permitted scholarly posting.
2. Authors may deposit the **ACCEPTED VERSION** of the peer-reviewed article on their personal websites, their institution's repository or any non-commercial repository such as PMC, arXiv after **12 MONTHS of publication** on the journal website. In addition, an acknowledgement must be given to the original source of publication and a link should be inserted to the published article on the journal's/publisher's website.
3. If the research is funded by NIH, Wellcome Trust or any other Open Access Mandate, authors are allowed the archiving of published version of manuscripts in an institutional repository after the mandatory embargo period. Authors should first contact the Editorial Office of the journal for information about depositing a copy of the manuscript to a repository. Consistent with the copyright agreement, Bentham Science does not allow archiving of **FINAL PUBLISHED VERSION** of manuscripts.
4. The link to the original source of publication should be provided by inserting the DOI number of the article in the following sentence: "The published manuscript is available at EurekaSelect via <http://www.eurekaselect.com/>[insert DOI]."
5. There is no embargo on the archiving of articles published under the **OPEN ACCESS PLUS** category. Authors are allowed deposition of such articles on institutional, non-commercial repositories and personal websites immediately after publication on the journal website.

PLAGIARISM PREVENTION:

Bentham Science Publishers uses the iThenticate software to detect instances of overlapping and similar text in submitted manuscripts. iThenticate software checks content against a database of periodicals, the Internet, and a comprehensive article database. It generates a similarity report, highlighting the percentage overlap between the uploaded article and the published material. Any instance of content overlap is further scrutinized for suspected plagiarism according to the publisher's Editorial Policies. Bentham Science allows an overall similarity of 20% for a manuscript to be considered for publication. The similarity percentage is further checked keeping the following important points in view:

Low Text Similarity:

The text of every submitted manuscript is checked using the Content Tracking mode in iThenticate. The Content Tracking mode ensures that manuscripts with an overall low percentage similarity (but which may have a higher similarity from a single source) are not overlooked. The acceptable limit for similarity of text from a single source is 5%. If the similarity level is above 5%, the manuscript is returned to the author for paraphrasing the text and citing the original source of the copied material.

It is important to mention that the text taken from different sources with an overall low similarity percentage will be considered as a plagiarized content if the majority of the article is a combination of copied material.

High Text Similarity:

There may be some manuscripts with an overall low similarity percentage, but a higher percentage from a single source. A manuscript may have less than 20% overall similarity but there may be 15 % similar text taken from a single article. The similarity index in such cases is higher than the approved limit for a single source. Authors are advised to thoroughly rephrase the similar text and properly cite the original source to avoid plagiarism and copyright violation.

Types of Plagiarism:

We all know that scholarly manuscripts are written after thorough review of previously published articles. It is therefore not easy to draw a clear boundary between legitimate representation and plagiarism. However, the following important features can assist in identifying different kinds of plagiarized content. These are:

- Reproduction of others words, sentences, ideas or findings as one's own without proper acknowledgement.
- Text recycling, also known as self-plagiarism. It is an author's use of a previous publication in another paper without proper citation and acknowledgement of the original source.
- Paraphrasing poorly: Copying complete paragraphs and modifying a few words without changing the structure of original sentences or changing the sentence structure but not the words.
- Verbatim copying of text without putting quotation marks and not acknowledging the work of the original author.
- Properly citing a work but poorly paraphrasing the original text is considered as unintentional plagiarism. Similarly, manuscripts with language somewhere between paraphrasing and quoting are not acceptable. Authors should either paraphrase properly or quote and in both cases, cite the original source.
- Higher similarity in the abstract, introduction, materials and methods, and discussion and conclusion sections indicates that the manuscript may contain plagiarized text. Authors can easily explain these parts of the manuscript in many ways. However, technical terms and sometimes standard procedures cannot be rephrased; therefore Editors must review these sections carefully before making a decision.

Plagiarism in Published Manuscripts:

Published manuscripts which are found to contain plagiarized text are retracted from the journal website after careful investigation and approval by the Editor-in-Chief of the journal. A 'Retraction Note' as well as a link to the original article is published on the electronic version of the plagiarized manuscript and an addendum with retraction notification in the journal concerned.

E-PUB AHEAD OF SCHEDULE:

Bentham Science Publishers are pleased to offer electronic publication of accepted papers prior to scheduled publication. These peer-reviewed papers can be cited using the date of access and the unique DOI number. Any final changes in manuscripts will be made at the time of print publication and will be reflected in the final electronic version of the issue. Articles ahead of schedule may be ordered by pay-per-view at the relevant links by each article stated *via* the **E-Pub Ahead of Schedule**.

Disclaimer:

Articles appearing in E-Pub Ahead-of-Schedule sections have been peer-reviewed and accepted for publication in this journal and posted online before scheduled publication. Articles appearing here may contain statements, opinions, and information that have errors in facts, figures, or interpretation. Accordingly, Bentham Science Publishers, the editors and authors and their respective employees are not responsible or liable for the use of any such inaccurate or misleading data, opinion or information contained in articles in the E-Pub Ahead-of- Schedule.

Member of Cope



Anexo 4: Instruções para autores: *Genetics and Molecular Biology*

SUBMISSION OF PAPERS

There is a publication charge for manuscripts once they are accepted. For price information, exemptions and waiver policies, please consult the journal homepage <http://www.gmb.org.br>.

1. Manuscripts must be submitted through our online submission platform hosted at: <http://mc.manuscriptcentral.com/gmb>

The cover letter should be addressed to:

Carlos C. F. Menck, Editor-in-Chief, Genetics and Molecular Biology

2. For submission the following instructions must be observed:

- The manuscript must be submitted by the Corresponding Author. This is the person who will also check the page proofs, and arranges for any payment that may incur during the editorial process.
- Entering the following metadata is required: (i) the manuscript title, (ii) a short running title (max. 35 characters), (iii) the Abstract, and (iv) up to five keywords. All these items must be exactly the same as those figuring in the first two pages of the manuscript file. Furthermore, a cover letter addressed to the Editor is required. It must be edited by the corresponding author inserted within the reserved data field.
- Statements are required informing that the data have not been published and are not under consideration elsewhere, and that all authors have approved the submission of the manuscript. Furthermore, possible conflicts of interest (e.g. due to funding, consultancies) must also be disclosed.
- The names of all co-authors, including institutional affiliations and e-mail addresses must be entered, as contact information for the Editorial Office.
- In the referee suggestions field, up to five reviewer names can be entered by the author(s); valid e-mail contact addresses for these are required, in case they are selected by the editor. These suggestions can be made separately as preferred and not-preferred reviewer(s).
- Files must be uploaded separately and identified according to file types. The main text file must include references and, if applicable, figure legends, which must be typed on a separate page following the References and Internet Resources sections. Each table, figure and element containing supplementary material must be saved and uploaded in a separate file. Formats for text and tables are Word or RTF in Windows platform. Figures should be in TIFF or JPEG formats (see detailed instructions in 3.1.h).
- Manuscripts including photos or any other identifiable data of human subjects must be accompanied by a copy of the signed consent by the individual or his/her guardian.

Failure to adhere to these guidelines can delay the handling of your contribution and manuscripts may be returned before being reviewed.

Special attention should be given to the structuring of the manuscript and correct language usage. These are important factors in the smooth running of the editorial and peer-review process, and can result in faster publication.

3. Categories of Contribution

3.1. Research Articles

Manuscripts must be written in English in double-spaced, 12-point type throughout; marked with consecutive line and page numbers, beginning with the cover page.

The following elements must start on a new page and be ordered as they are listed below:

a) **The title page** must contain: a concise and informative title; the authors' names (first name at full length); the authors' institutional affiliation, including department, institution, city, state or province, and country; different affiliations indicated with superscript Arabic numbers; a short running title of up to 35 characters (including spaces); up to five key words; the corresponding author's name, full postal, and email address.

b) **The Abstract** must be a single paragraph that does not exceed 200 words and summarizes the main results and conclusions of the study. It should not contain references.

c) **The text** must be as succinct as possible. *Text citations*: articles should be referred to by authors' surnames and date of publication; citations with two authors must include both names; in citations with three or more authors, name the first author and use *et al.* List two or more references in the same citation in chronological order, separated by semi-colons. When two or more works in a citation were published in the same year, list them alphabetically by the first author surname. For two or more works by the same author(s) in a citation, list them chronologically, with the years separated by commas. (Example: Freire-Maia *et al.*, 1966a, 1966b, 2000). Only articles that are published or in press should be cited. In the case of personal communications or unpublished results, all contributors must be listed by initials and last name (*et al.* should not be used). *Numbers*: In the text, numbers nine or less must be written out except as part of a date, a fraction or decimal, a percentage, or a unit of measurement. Use Arabic numerals for numbers larger than nine. *Binomial Names*: Latin names of genera, species and infraspecific taxa must be printed in italics; names of orders and families should appear in the Title and also when first mentioned in the text. URLs for programs, data or other sources should be listed in the Internet Resources Section, immediately following the References Section, not in the text.

The text includes the following elements:

Introduction - Description of the background that led to the study.

Material (or Subjects) and Methods - Details relevant to the conduct of the study. Statistical methods should be explained at the end of this section.

Results - Undue repetition in text and tables should be avoided. Statistical analyses should be presented as complete as possible, i.e. not only *P*-values should be shown, but also all other test variables required for full appreciation of the results by the reviewers and readers. Comments on relevance of results are appropriate but broader discussion should be part of the Discussion section.

Discussion - The findings of the study should be placed in context of relevant published data. Ideas presented in other publications should not be discussed solely to make an exhaustive presentation. Some manuscripts may require different formats appropriate to their content.

d) **The Acknowledgments** must be a single paragraph that immediately follows the discussion and includes references to grant support.

e) **The References Section**: References must be ordered alphabetically by the first author surname; references with the same first author should be ordered as follows: first, as single author in chronological order; next, with only one more co-author in alphabetical order by the second author; and finally followed by references with more than two co-authors, in chronological order, independent of the second author surnames. In references with more than 10 authors only the first ten should be listed, followed by *et al.* Use standard abbreviations for journal titles as suggested by NCBI (<http://www.ncbi.nlm.nih.gov/journals/>).

Only articles that are published or in press should be included in this section. Works submitted for publication but not yet accepted, personal communications and unpublished data must be cited within the text. "Personal communication" refers to information obtained from individuals other than the authors of the manuscript being submitted; "unpublished data" refers to data produced by at least one of the authors of the manuscript under consideration. Works of restricted circulation (e.g., theses not available in public

databases, congress abstracts not published in regular journals or public databases) should not be listed in this section.

Sample journal article citation:

Breuer ME and Pavan C (1955) Behaviour of polytene chromosomes of *Rhynchosciara angelae* at different stages of larval development. *Chromosoma* 7:371-386.

Yonenaga-Yassuda Y, Rodrigues MT and Pellegrino KCM (2005) Chromosomal banding patterns in the eyelid-less microteiid lizard radiation: The X1X1X2X2:X1X2Y sex chromosome system in *Calyptommatius* and the karyotypes of *Psilophthalmus* and *Tretioscincus* (Squamata, Gymnophthalmidae). *Genet Mol Biol* 28:700-709.

Sample book citation:

Dobzhansky T (1951) *Genetics and Origin of Species*. 3rd edition. Columbia University Press, New York, 364 pp.

Sample chapter-in-book citation:

Crawford DC and Howard-Peebles PN (2005) Fragile X: From cytogenetics to molecular genetics. In: Gersen SL and Keagle MB (eds) *The Principles of Clinical Cytogenetics*. 2nd edition. Humana Press, New Jersey, pp 495-513.

Sample electronic article citation:

Gotzek D, Ross KG (2009) *Current status of a model System: The gene Gp-9 and its association with social organization in fire ants*. *PLoS One* 4:e7713.

f) **Internet Resources Section:** this section should contain a list of URLs referring to data presented in the text, as well as software programs and other Internet resources used during data processing. Date of consultation must be stated.

Sample Internet resource citation:

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/OMIM> (September 4, 2009)
LEM Software, http://dir.niehs.nih.gov/dirbb/weinbergfiles/hybrid_design.htm (September 4, 2009)

g) **Tables:** must be in Word format prepared with the table tool (do not use space bar or tabulator). A concise title should be provided above the table. Tables must be numbered consecutively in Arabic numerals. Each column must have a title in the box head. Footnotes typed directly below the table should be indicated in lowercase superscript letters. Tables that are to appear in the printed version must be saved in Word format and not as figures, so that they can later be fitted during typesetting. Each table must be saved and uploaded as a separate file.

h) **Figures** must be numbered consecutively using Arabic numerals. Images should be in TIFF or JPEG format. Figures in Word, PowerPoint or Excel format cannot be published. Only sequence data can be presented in Word format. Journal quality reproduction will require grayscale resolution yielding 300 dpi, color figures should be at 600 dpi. These resolutions refer to the output size of the file, that is the size in which it will appear printed in the journal; if it is anticipated that images will be enlarged or reduced, the resolutions should be adjusted accordingly. Figures composed of several elements should be sent as a single panel, obeying the print size definitions of the journal (single or two columns width). Scanned figures should not be submitted. Color illustrations are accepted. Each figure/panel must be saved and uploaded as a separate file. When uploading, identify each illustration by the first author name and the number of the respective figure.

Figure legends must be included at the end of the main text file and should be typed on a new page.

i) **Nomenclature:** Taxonomic names should be in accordance with current international standards. For rules concerning gene names and gene symbols, please see separate Instruction form.

j) **Sequences** may appear in text or in figure. DNA, RNA and protein sequences equal to or greater than 50 units must be entered into public databases and accession numbers must be provided upon acceptance of the article. Failure to do so will inadvertently delay publication.

k) **Data access:** reference should be made to availability of detailed data and materials used for reported studies.

l) **Ethical issues:** Reports of experiments on live vertebrates must include a statement in the text that the institutional review board approved the work and the protocol number must be provided. For experiments involving human subjects, authors must also include a statement that informed consent was obtained from all subjects. If photos or any other identifiable data are included, a copy of the signed consent must be uploaded during manuscript submission.

m) **Supplementary Material:** Data that the authors consider of importance for completeness of a study, but which are too extensive to be included in the published version, can be submitted as Supplementary Material. At publication, this material will be made available together with the electronic version. In case a manuscript contains such material, it should be appropriately identified within the text file. Supplementary material in tables should be identified as Table S1, Table S2, etc., in case of figures they should be named accordingly, Figure S1, Figure S2. In addition, a list of this material should be presented at the end of the manuscript text file, containing the following statement:

Supplementary material - the following online material is available for this article:

- *Table S1 – < short title >*
- *Figure S1 – < short title >*

This material is available as part of the online article from <http://www.scielo.br/gmb>

3.2 Short Communications

Short Communications present brief observations that do not warrant full-length articles. They should not be considered preliminary communications;

- should be 15 or fewer typed pages in double spaced 12-point type, including literature cited;
- should include an Abstract;
- but no further subdivision, with introduction, material and methods, results and discussion; all in a single section and without headers.
- up to four items (tables and/or figures) may be submitted;

Note: The title page, abstract and reference section format is that of a full-length Research Article. For Supplementary Material see instructions in item 3.1.m.

3.3 Letters to the Editor

Relate or respond to recent published items in the journal. Discussions of political, social and ethical issues of interest to geneticists are also welcome in this form.

3.4 Review Articles

Review Articles are welcome. The Editor must be contacted prior to submission. Please, provide an Abstract and a list of your recent publications in the area.

3.5 Book Reviews

Publishers are invited to submit books on Genetics, Evolution and related disciplines, for review in the journal. Aspiring reviewers may propose writing a review.

3.6 History, Story and Memories

These are accounts on historical aspects of Genetics relating to Brazil.

4. Articles accepted for publication

Once an article is accepted, the Editorial Office will send it to copy editor for language and technical corrections. If major corrections were proposed, the manuscript with the highlighted corrections will be returned to the corresponding author for approval. The final version approved by the authors must be free of any text/correction markings when returned to the Editorial Office.

After typesetting, page proofs will be sent to the corresponding author. Changes made to page proofs, apart from typesetting errors, will be charged to the authors. Notes added in proof require Editorial approval.

Together with the proofs, a form of consent to publish and transfer of copyright is sent to the corresponding author. The latter will have to sign this form, also on behalf of any co-authors, and send it by fax to the Editorial Office.

5. Reprints

Reprints are free of charge and will be provided as a pdf-file.

10. Curriculum vitae (Lattes)

Flávia Tadeu de Araújo

Curriculum Vitae

Janeiro/2015

Flávia Tadeu de Araújo

Curriculum Vitae

Dados pessoais

Nome Flávia Tadeu de Araújo
Filiação Edson Tadeu de Araújo e Maria Czekalski
Nascimento 01/12/1987 - Recife/PE - Brasil
Carteira de Identidade 7339228 SDS-PE - PE - 05/08/2003
CPF 064.053.234-98

Formação acadêmica/titulação

2013 Mestrado em Genética.
 Universidade Federal de Pernambuco, UFPE, Recife, Brasil
 Título: Genes de resistência a patógenos em feijão-caupi e em outras leguminosas: caracterização e diversidade.
 Orientadora: Ana Maria Benko-Iseppon
 Bolsista do(a): Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE)

2009 - 2013 Graduação em Bacharelado em Ciências Biológicas.
 Universidade Federal Rural de Pernambuco, UFRPE, Recife, Brasil
 Título: Caracterização molecular de acessos de feijão-caupi com potencial de uso no melhoramento genético assistido por marcadores microsatélites.
 Orientadora: Ana Maria Benko-Iseppon
 Bolsista do(a): Empresa Brasileira de Pesquisa Agropecuária

2006 - 2008 Ensino Profissional de nível técnico .
 Colégio Agrícola Dom Agostinho Ikas, CODAI, Brasil

Formação complementar

2013 - 2013 Curso de curta duração em PCR em Tempo Real: Princípios Básicos e Aplicações. Universidade Federal de Pernambuco, UFPE, Recife, Brasil

Produção Bibliográfica

Artigos aceitos para publicação

1. WANDERLEY-NOGUEIRA, A. C., KIDO, E. A., BEZERRA NETO, J. P., **ARAÚJO, F. T.**, AMORIM, L. L. B., CROVELLA, S., BENKO-ISEPPON, A. M. Plant Elite Squad: first defense line and resistance genes – identification, diversity and functional roles. Current Protein and Peptide Science, 2015.

Trabalhos publicados em anais de eventos (completo)

1. **ARAÚJO, F. T.**, WANDERLEY-NOGUEIRA, A. C., AMORIM, L. L. B., BELARMINO, L. C. S., BENKO-ISEPPON, A. M. Prospecção de genes R da subfamília TIR-NBS-LRR em feijão-caupi In: I Congresso Internacional de Ciências Biológicas I Congresso Nacional de Ciências Biológicas VI Simpósio de Ciências Biológicas, Recife-PE, 2013.

Trabalhos publicados em anais de eventos (resumo)

1. **ARAÚJO, F. T.**, MATOS, M. K. S., OLIVEIRA, A. R. S., SILVA-LIMA, S. C. B., PIERECK, B., BEZERRA NETO, J. P., AMORIM, L. L. B., BENKO-ISEPPON, A. M. Caracterização e ancoragem de gene de resistência do tipo NBS-LRR em soja [*Glycine max* (L.) Merr., Fabaceae] In: XX Encontro de Genética do Nordeste (ENGINE), Campina Grande-PB, 2014.
2. **ARAÚJO, F. T.**, MATOS, M. K. S., BEZERRA NETO, J. P., AMORIM, L. L. B., BENKO-ISEPPON, A. M. Desenvolvimento de marcadores do tipo RGA (*Resistance Gene Analog*) em feijão-caupi In: XX Encontro de Genética do Nordeste (ENGINE), 2014, Campina Grande-PB, 2014.
3. BENKO-ISEPPON, A. M., **ARAÚJO, F. T.**, BEZERRA NETO, J. P., AMORIM, L. L. B., WANDERLEY-NOGUEIRA, A. C. Identificação e ancoragem de genes de resistência TIR-NBS-LRR no genoma do feijão-caupi In: XI Congresso Latinoamericano de Botânica, 65º Congresso Nacional de Botânica, Salvador-BA, 2014.
4. CESAR, R., OLIVEIRA, W. D., **ARAÚJO, F. T.**, MATOS, M. K. S., AMORIM, L. L. B., BENKO-ISEPPON, A. M. Marcadores SRR polimórficos entre genótipos de feijão-caupi contrastantes para tolerância à seca In: XI Congresso Latinoamericano de Botânica, 65º Congresso Nacional de Botânica, Salvador-BA, 2014.
5. ARAUJO, E. C., ARAUJO, W. J., **ARAÚJO, F. T.**, SILVA-ALMEIDA, S. L. Programa saúde na escola (PSE) e sua efetividade nas escolas do município de Recife-PE In: XIV Jornada de ensino, pesquisa e extensão, Recife-PE, 2014.
6. ARAUJO, E. C., ARAUJO, W. J., SILVA-ALMEIDA, S. L., **ARAÚJO, F. T.** Projeto Olhar Recife e sua importância no aprendizado do aluno In: XIV Jornada de ensino, pesquisa e extensão, Recife-PE, 2014.
7. MATOS, M. K. S., BEZERRA NETO, J. P., AMORIM, L. L. B., **ARAÚJO, F. T.**, WANDERLEY-NOGUEIRA, A. C., BENKO-ISEPPON, A. M. Structural domain analysis and 'in silico' expression of R2R3-MYB subfamily in cowpea (*Vigna unguiculata* (L.) Walp.) transcriptome In: 60º Congresso Brasileiro de Genética, Guarujá-SP, 2014.
8. **ARAÚJO, F. T.**, WANDERLEY-NOGUEIRA, A. C., AMORIM, L. L. B., PANDOLFI, V., BENKO-ISEPPON, A. M. Ancoragem de genes candidatos à subfamília TIR-NBS-LRR de feijão-caupi em pseudocromossomos de *Phaseolus vulgaris* L. In: III Jornada da Pós-Graduação em Genética da UFPE, Recife, 2013.

Trabalhos publicados em anais de eventos (resumo expandido)

1. D'ASSUNCAO, C. G., **ARAÚJO, F. T.**, OLIVEIRA, M. A. B. Análise comportamental do porco-do-

mato (*Tayassu tajacu*) em cativeiro no parque estadual dois irmãos, recife - PE In: XIII Jornada de ensino, pesquisa e extensão – JEPEX 2013, Recife, 2013.

Apresentação de trabalho e palestra

1. **ARAÚJO, F. T.** Análise de genes de resistência a patógenos em feijão-caupi e em outras leguminosas: caracterização, mapeamento e diversidade, 2014. (Seminário, Apresentação de Trabalho)
2. **ARAÚJO, F. T.** Isolation and characterization of NBS–LRR resistance gene analogues from mango, 2014. (Seminário, Apresentação de Trabalho)
3. **ARAÚJO, F. T.** Comparative analysis of peanut NBS-LRR gene clusters suggests evolutionary innovation among duplicated domains and erosion of gene microsynteny, 2013. (Seminário, Apresentação de Trabalho)
4. **ARAÚJO, F. T.** Identification and Characterization of Nucleotide-Binding Site-Leucine-Rich Repeat Genes in the Model Plant *Medicago truncatula*, 2013. (Seminário, Apresentação de Trabalho)
5. **ARAÚJO, F. T.** Identification of expressed resistance gene-like sequences by data mining in 454-derived transcriptomic sequences of common bean (*Phaseolus vulgaris* L.), 2013. (Seminário, Apresentação de Trabalho)
6. **ARAÚJO, F. T.** NBS-LRR work a 'bait and switch' on pathogens, 2013. (Seminário, Apresentação de Trabalho)

Participação em eventos

1. GE Day | Microarray, 2013. (Encontro)

Outra

1. Avaliadora dos pôsteres na área de Ciências Biológicas no I Congresso Internacional de Ciências Biológicas, I Congresso Nacional de Ciências Biológicas, VI Simpósio de Ciências Biológicas, 2013, Universidade Católica de Pernambuco.