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

# CARACTERIZAÇÃO BIOINFORMÁTICA DE GENES RELACIONADOS À INTERAÇÃO PATÓGENO-HOSPEDEIRO EM ANGIOSPERMAS

**RECIFE**
**FEVEREIRO/2012**

**ANA CAROLINA WANDERLEY NOGUEIRA**

# CARACTERIZAÇÃO BIOINFORMÁTICA DE GENES RELACIONADOS À INTERAÇÃO PATÓGENO-HOSPEDEIRO EM ANGIOSPERMAS

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Ciências Biológicas, na área de concentração Biotecnologia/ Biologia Celular e Molecular.

**Orientadora: Ana Maria Benko Iseppon**

**RECIFE**
**FEVEREIRO/2012**

**Nogueira, Ana Carolina Wanderley**
**Caracterização bioinformática de genes relacionados à interação patógeno-hospedeiro em angiospermas/ Ana Carolina Wanderley Nogueira. – Recife: O Autor, 2012.**

**217 folhas : il., fig., tab.**
**Orientadora: Ana Maria Benko Iseppon**
**Tese (doutorado) – Universidade Federal de Pernambuco, Centro de Ciências Biológicas. Biotecnologia/ Biologia Celular e Molecular, 2012.**
**Inclui bibliografia e anexos**

**1. Angiosperma 2. Bioinformática 3. Melhoramento genético I. Iseppon, Ana Maria Benko II. Título.**

**582.13 CDD (22.ed.) UFPE/CCB-2012-080**

**COMISSÃO EXAMINADORA**

---

Profª Drª Ana Maria Benko Iseppon
(Orientadora)
UFPE

---

Profª Drª Maria Tereza dos Santos Correia
(Membro Interno)
UFPE

---

Prof Dr Ederson Akio Kido
(Membro Externo)
UFPE

---

Prof Dr Sergio Crovella
(Membro Externo)
UFPE

---

Prof Dr José Miguel Ortega
(Membro Externo)
UFMG

Dedico

*À minha princesa*
*Alice.*

# AGRADECIMENTOS

À minha mãe, Sheila, por todo seu amor; colo sempre disponível e por estar sempre torcendo pelo meu sucesso me fazendo lembrar quando necessário que doutorado é "coisa pra caramba";

À Renato Vieira, meu amor e companheiro, amigo de todas as horas, por quem muitas vezes quis chegar mais longe e de forma mais bonita.

À minha pequena Alice, por ter me ensinado definitivamente o significado da palavra amor;

À minha irmã, Ana Luiza, pela parceiria em todos os momentos;

À Dedé e Tio Clovinho por todo o amor, carinho e admiração a mim dedicados;

À Profª Ana Maria Benko Iseppon, por tantos anos juntas, por tudo que me ensinou mostrando que (por mais impossível que pareça) sempre é possível fazermos mais e melhor;

À Nina, amiga pra todas as horas, com quem dividi (e pretendo continuar dividindo) grande parte do meu sucesso;

À Luis Carlos, pela ajuda sempre disponível, pelo carinho e amizade que conseguimos construir;

À João Pacífico, por nunca se cansar dos meu pedidos de socorro e por atendê-los tão rapidamente;

À todos que fazem do LGBV - estudantes, mestrandos, doutorandos, técnicos e professores – um verdadeiro time;

Às minhas grandes amigas que sempre torceram por mim e admiram minha trajetória;

À Tatianna Ribeiro, pela amizade tão importante pra mim quando estava tão longe de casa;

À Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) pela concessão da bolsa de doutorado;

À Universidade Federal de Pernambuco (UFPE), por meio do Programa de Pós-Graduação em Ciências Biológicas (PPGCB), pela oportunidade e suporte durante todo o período do doutoramento.

*“É do buscar e não do*
*achar que nasce o que*
*eu não conhecia”*

Clarice Lispector

# SUMÁRIO


LISTA DE FIGURAS 9
LISTA DE TABELAS 12
LISTA DE ABREVIATURAS 13
RESUMO 17
ABSTRACT 18
INTRODUÇÃO 19
OBJETIVOS 22
REVISÃO BIBLIOGRÁFICA 23

1 O Ataque de patógenos e as perdas causadas por doenças 24

1.1 Resistência contra patógenos 24
1.1.1 Mecanismos de defesa 24
1.1.2 A interação gene-a-gene 25
1.1.3 Produção e importância econômica 27
1.1.4 A evolução dos genes de resistência 31
1.1.5 Genes Relacionados à Patogenicidade 32

2 A cultura da cana-de-açúcar 35

2.1 Origem, história e citogenética 35
2.2 As doenças da cana-de-açúcar 36
2.3 Produção e importância econômica 38
2.4 O projeto SUCEST 38

3 A cultura da soja 39

3.1 Origem, história e citogenética 39
3.2 As doenças da soja 40
3.3 Produção e importância econômica 41
3.4 O projeto GENOSOJA 42

4 A cultura do feijão-caupi 42

4.1 Origem, história e citogenética 42
4.2 As doenças do feijão-caupi 43
4.3 Produção e importância econômica 44
4.4 O transcriptoma do feijão-caupi e a rede NordEST 45

5 A Bioinformática aplicada às ômicas 45

5.1 História e aplicações 45
5.2 Os bancos de dados e ferramentas 48

REFERÊNCIAS BIBLIOGRÁFICAS 51

**CAPÍTULO I.** Insight on the plant defense mechanisms in Sugarcane transcriptome 72
**CAPÍTULO II.** An overall evaluation of Resistance (*R*) and Pathogenesis- 131

Related (*PR*) superfamilies in soybean, medicago and Arabidopsis
**CAPÍTULO III.** *In Silico* screening for pathogenesis-related-2 gene candidates in *Vigna unguiculata* (L.) Walp. transcriptome 169

**CAPÍTULO IV.** Screenning of biotic defense response genes in *Vigna unguiculata* (L.) Walp. transcriptome 178

CONCLUSÕES GERAIS 210

ANEXOS
**Anexo I.** Instrução para autores: Revista *Bioinformatics and Biology Insights* 211

**Anexo II**. Súmula curricular 213

# LISTA DE FIGURAS

**REVISÃO BIBLIOGRÁFICA**

Figura 1. Esquema representativo da interação gene-a-gene. Em (A) observa-se uma interação compatível que confere resistência onde estão presentes o gene R e o gene avr do patógeno correspondente. Em (B) não há produto nem do gene R nem do avr; em (C) há produto do gene R e de um gene avr não correspondente e em (D) há produto apenas do gene avr e ausência de gene de resistência. Nos três últimos casos o fenótipo é suscetível. 26

Figura 2. Representação gráfica das quatro famílias de Genes de Resistência que apresentam domínios conservados em suas estruturas. 28

Figura 3. Visão geral dos principais mecanismos de sinalização e defesa da planta. 33

**CAPÍTULO I**

Figure 1. Plant-pathogen interaction pathway available at KEGG showing the number of elements found in sugarcane transcriptome for each gene type, followed by the number of exclusively aligned sequences (between parenthesis). Abbreviations: HR, Hypersensitive Response; PAMP: pattern associated to the pathogen; PR, Pathogenesis Related; PaPRR: pattern recognition receptor; R, Resistance; ROS, Reactive Oxygen Species. 126

Figure 2. Distribution of the 314 sugarcane *clusters* that aligned with PR-genes families and of the 107 sugarcane *clusters* that aligned with R-genes with no repetitions. 127

Figure 3. Hierarchical *cluster*ing (*Cluster*3.0) of up-regulated (red) and down-regulated (green) SuperSAGE tags ($p < 0, 05$) related to R (A) and PR classes (B) using FC (Fold Change) of gene expression ratios (experimental/control) under the tested conditions [STS/STC: Salinity-tolerant (100 mM NaCl)/ control; DTS/DTC: Drought- tolerant / control; DSS/DSC: Drought- sensible / control]. The tag number and the contig ID is given together with the gene product. 128

Figure 4. In silico FISH. Schematic representation of *clusters*/groupings that were anchored in the rice genome based on BLAST similarity results. Colored circles next to the *clusters*/grouping names correspond to the synteny events between chromosomes; each one of the 12 chromosomes was identified with a different color, as showed in the legend. Triangles and inverted triangles indicate the similarity of expressed tags (resulted from SuperSAGE) with *clusters*/grouping; the form colors and senses correspond to the library and regulation type, according the legend. 129

Figure 5. Number of sugarcane *clusters* that aligned in each rice chromosome. 130

**CAPÍTULO II**

Figure 1. *R* and *PR* genes encountered in soybean, Arabidopsis and Medicago transcriptomes. *R*-genes are represented in the outer circle and *PR*-genes in the inner circle for each species. 163

Figure 2. Distribution of *R*-gene families in soybean, Arabidopsis and Medicago in the four main *R*-gene categories, considering their conserved domains. Numbers over the columns mean the amount of non-redundant sequences for each class. 164

Figure 3. A. Distribution of *R* and *PR*-genes in soybean, Medicago and Arabidopsis. B. Distribution of *Xa21* and *PR-2* in soybean, Medicago and Arabidopsis. Numbers of matches for each gene category are inside the columns. 165

Figure 4. Dendrograms generated after maximum parsimony analysis showing the relationships among selected plant species considering sequences of (A) *Xa21* and (B) *PR-2*. Keys in (1) represent monocots and in (2) dicots. *Xa21*: The circle on the root of A shows the divergence point between monocots and dicots. *PR-2:* The circle on the root of B shows an ancestor with a symplesiomorphic character. Numbers in the base of the branches regard bootstrap values and bar means evolutionary scale. 166

Figure 5. Number of SuperSAGE tags matching soybean *R* and *PR* gene candidates from three different comparisons among the six libraries: 1- Embrapa-48, drought tolerant stressed *vs.* negative control; 2- BR-16, drought susceptible stressed *vs.* negative control; 3- PI561356 fungus resistant stressed with *Phakopsora pachyrhizi vs.* negative control. 167

Figure 6. Graphic representation of soybean *R* and *PR* sequences positioned on *Medicago truncatula* chromosomes (MtChr) with aid of the CVit-BLAST resource available at the website http://www.medicago.org/. Arrows indicate genes that appear in tandem repetitions. 168

**CAPÍTULO III**

Figure 1. Pipeline to identify PR genes. Black boxes indicate data from automatic annotation. Gray boxes indicate manual annotation steps. Cylinders mean used databases. 172

Figure 2. Twelve conserved motifs characteristic of PR-2 protein in 16 *clusters* from V. unguiculata. The first line shows the conserved motifs generated by the HMMER program using PR-2 proteins from eight different organisms. In light gray, it is possible to observe which motifs appeared in cowpea PR-2 candidates. 173

Figure 3. Dendrogram generated after Maximum Parsimony analysis, showing relationships among the PR-2 seed sequence of *A. thaliana* and orthologs of *V. unguiculata* and other organisms with PR-2 proteins bearing desired domains. Dotted line delimits the main taxonomic units and letters on the right of the dendrogram refer to the grouping. The circle on the root of clade B shows the divergence point between monocots and dicots. Decimal numbers under branches lines means distance values. The numbers between parentheses on the left of the branches nodes corresponding to the Bootstrap values. 173

Figure 4. Graphic representation of *PR-2* isoeletric points. Abscissa (X-axis) represents isoeletric point. Ordinate (Y-axis) represents molecular weight. 174

Figure 5. *PR-2* expression profile. Black indicates higher expression, gray lower expression, and light gray absence of expression in the corresponding tissue and *cluster*. Abbreviations: CT00 (control); BM90 (Leaves of BR14-Mulato genotype); IM90 (Leaves of IT85F genotype collected with 90 minutes after mosaic viruses infection); SS00 (Root of genotype sensitive to salinity without salt stress); SS02 (Root of genotype sensitive to salinity after 2 hours of stress); SS08 (Root of genotype sensitive to salinity after 8 hours of stress); ST00 (Root of genotype tolerant to salinity without salt stress); ST02 (Root of genotype tolerant to salinity after 2 hours of stress); ST08 (Root of genotype tolerant to salinity after 8 hours of stress). 175

**CAPÍTULO IV**

Figure 1. Number of of *R* and *PR* genes in soybean, medicago and cowpea. 205

Figure 2. Distribution of *R*-genes classes in soybean, medicago and cowpea. 206

Figure 3. *R* and *PR* genes transcripts prevalence in 22 different cowpea libraries including: SS00: salinity susceptible accession (cultivar 'Canapu Amarelo') roots without stress; SS02: roots of 'Canapu Amarelo' two hours after salinity stress; SS08: roots of 'Canapu Amarelo' eight hours after salinity stress; ST00: tolerant accession ('Pitiúba') roots without stress (control); ST02: roots of 'Pitiúba' cultivar two hours after salinity stress; ST08: 'Pitiúba' roots two hours after salinity stress; BM01: leaves of the CPSMV (cowpea severe mosaic virus) resistant cultivar 'BR14-Mulato' without stress (control); BM90: 'BR14-Mulato' leave infected with CPSMV (bulk of 30, 60 and 90 min.) after stress; IM01: leave of the CPSMV susceptible accession IT85F-2687 without stress (control); IM90: IT85F-2687 leave infected by CPSMV (bulk of three times 30, 60 and 90 min.) after stress; BUD: growing axillary buds; DSD: developing seeds; LF: leave; LF2: primary leaflets; LFM: leaflets and shoot meristem; MI1: mixed tissues; NDL: nodules; RT2: root hairs; RT3: roots; SD: seeds. 207

Figure 4. Schematic representation of cowpea *R* and *PR* genes alignments in soybean chromosomes. Blue columns regard number of loci per chromosome while red columns represent the number of non-redundant genes. The x-axis represents soybean chromosome number. 208

Figure 5. Distribution and microsyntenic relationships considering *R* (in green) and *PR* (in pink) candidates identified in cowpea against the soybean genome (2n=40). Chromosomes are depicted with the centromere (in red) in their expected position. Scale = 1 Mb. Microsyntenic relationships are shown as links between chromosome regions. 209

# LISTA DE TABELAS

**CAPÍTULO I**

Table 1. Sugarcane *clusters* matching results to each gene of KEGG pathway. 119

Table 2. Sugarcane upregulated SuperSAGE tag (p<0.05) under abiotic stress-related to sugarcane contigs annotated to *R* gene products. Abbreviations: Fam.=Family; Kin.=Kinase; LRR=Leucine-Rich; Prot=Protein; Receptor=recep. 121

Table 3. Sugarcane upregulated SuperSAGE tags (p < 0.05) under abiotic stress related to sugarcane contigs annotated to *PR* gene products. 123

Table 4. Sugarcane *clusters* that presented similarities with distinct segments in a same rice chromosome. 125

**CAPÍTULO II**

Table 1. Soybean *clusters* matching results to each procured *R* and *PR* genes. Showing number of matches to each seed-sequences, e-value, score, size in nucleotide (nt) and amino-acids (aa), presence of conserved domains and number of matches in soybean (S), Medicag*o* (M) and Arabidopsis (A). Abbreviations: (c) = Complete; (i) = Incomplete. 160

**CAPÍTULO III**

Table 1. *V. unguiculata PR-2* candidates, showing the sequence size in nucleotides and amino acids with the conserved start and end sites, the best alignment in NCBI database, including the score and e-value. The accession numbers gb.AAY96764.1, gb.AAY25165.1, emb.CAO61165.1, emb.CAO71593.1 and gbEAU72980.1 corresponding to sequences from *Phaseolus vulgaris*, *Ziziphus jujuba*, *Vitis vinifera*, *Leishmania infantum* and *Synechococcus* sp., respectively. Abbreviations: CD, Glyco-hydro Conserved Domain. 171

**CAPÍTULO IV**

Table 1. *Vigna unguiculata* (Cowpea) identified *PR* and *R* candidates and their features, as compared with the number of matches for *Medicago truncatula (*Medicago) and *Glycine max* (Soybean). 204

# LISTA DE ABREVIATURAS

| | |
|---|---|
| **ABA** | Ácido Abscísico*; Abscisic Acid* |
| **AGROFIT** | Sistema de Agrotóxicos Fitossanitários |
| **ATP** | Adenosina Trifosfato*; Adenosine Triphosphate* |
| **AVR** | Avirulência |
| **BAC** | Cromossomo Artificial de Bactéria; *Bacterial Artificial Chromosome* |
| **BLAST** | Ferramenta Básica de Busca por Alinhamento Local; *Basic Local Alignment Search Tool* |
| **bp** | Pares de bases*; Base pairs* |
| **bZIP** | Ziper de Leucina Básico; *Basic Leucine Zipper* |
| **CAPES** | Coordenação de Aperfeiçoamento de Pessoal de Nível Superior |
| **CC** | Cauda Espiralada; *Coiled coil* |
| **CD** | Domínio Conservado; *Conserved Domain* |
| **cDNA** | DNA Complementar; *Complementary DNA* |
| **CNPq** | Conselho Nacional de Desenvolvimento Científico e Tecnológico |
| **CONAB** | Companhia Nacional de Abastecimento |
| **CPSMV** | Virus do Mosaico Severo do Caupi; *Cowpea Severe Mosaic Virus* |
| **CRKs** | Kinases ricas em cisteína; *Cysteine-rich Kinases* |
| **DDBJ** | Banco de Dados de DNA do Japão; *DNA Databank of Japan* |
| **DNA** | Ácido Desoxirribonucleico; *Desoxyribonucleic Acid* |
| **ET** | Etileno |
| **EMBL** | Laboratório Europeu de Biologia Molecular; *European Molecular Biology Laboratory* |
| **EMBRAPA** | Empresa Brasileira de Pesquisa Agropecuária |
| **EST** | Etiqueta de Sequência Expressa; *Expressed Sequence Tag* |
| **EWAS** | Estudo de Associação Ampla ao Meio Ambiente; *Environment-wide Association Study* |

| | |
|---|---|
| **EUA** | Estados Unidos da América |
| **FACEPE** | Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco |
| **FAO** | Organização das Nações Unidas para Agricultura e Alimentação; *Food and Agriculture Organization* |
| **FAPESP** | Fundação de Amparo à Pesquisa do Estado de São Paulo |
| **FISH** | Hibridização Fluorescente *in situ ; Fluorescent In situ Hibridization* |
| **GenBank** | Banco de Genes; GeneBank |
| **GENOSOJA** | Projeto do Genoma da Soja; *Soybean Genome Project* |
| **GTP** | Trifosfato de guanosina; *Guanosine triphosphate* |
| **GWAS** | Estudo de Associação Genômica Ampla; *Genome-Wide Association Study* |
| **HR** | Resposta Hipersensível; *Hypersensitive Response* |
| **HMM** | *Hidden Markov Model* |
| **INSD** | Banco de Dados Internacional de Sequências de Nucleotídeos; *International Nucleotide Sequence Database*; |
| **JA** | Ácido Jasmônico; *Jasmonic Acid* |
| **KEGG** | Enciclopédia de Genes e Genomas de Kyoto; *Kyoto Enciclopedia of Genes and Genomes* |
| **LRR** | Repetições Ricas em Leucina; *Leucine Rich Repeats* |
| **LTP** | Proteínas Transferidoras de Lipídeos; *Lipid-Transfer Proteins* |
| **MAMP** | Padrões Moleculares Micróbio-Associados; *Microbe-Associated Molecular Patterns* |
| **MAPK** | Proteína Kinase Ativada por Mitose; *Mithogen Activated Protein Kinases* |
| **MEGA** | Análises Genéticas da Evolução Molecular; *Molecular Evolutionary Genetic Analysis* |
| **MYA** | Milhões de Anos Atrás; *Million Years Ago* |
| **MW** | Peso Molecular; *Molecular Weight* |
| **NADPH** | Nicotinamida Adenina Dinucleotídeo-Fosfato*; Nicotinamide Adenine Dinucleotide-phosphate* |

| | |
|---|---|
| **NBS** | Sítio Ligador de Nucleotídeo *; Nucleotide Binding Site* |
| **NCBI** | Centro Nacional para Informação Biotecnológica*; National Center for Biotechnology Information* |
| **NJ** | Agrupamento por Vizinhança*; Neighbor-Joining* |
| **NR** | Não Redundante*; Non-Redundant* |
| **ONSA** | Organização para Sequenciamento e Análise de Nucleotídeos*; Organization for Nucleotide Sequencing and Analysis* |
| **ORF** | Quadro de Leitura Aberto; *Open Reading Frame* |
| **PAMP** | Padrões Moleculares Associados ao Patógeno; *Pathogen-Associated Molecular Patterns* |
| **PCR** | Reação em Cadeia da Polimerase*; Polymerase Chain Reaction* |
| **PheWAS** | Estudo de Associação Fenômica Ampla*; Phenome-Wide Association Study* |
| **pI** | Ponto Isoelétrico; *Isoeletric Point* |
| **PR** | Relacionado à Patogenicidade; *Pathogenesis Related* |
| **PRGdb** | Banco de Dados de Genes de Resistência em Plantas; *Plant Resistance Genes Database* |
| **R** | Resistência; *Resistance* |
| **RLK** | Receptor tipo Kinase; *Receptor Like Kinase* |
| **RENORBIO** | Programa Rede Nordeste de Biotecnologia |
| **RNA** | Ácido Ribonucleico; *Ribonucleic Acid* |
| **ROS** | Espécies Reativas de Oxigênio; *Reactive Oxygen Species* |
| **RRP** | Padrões de Reconhecimento de Receptor; *Recognition Receptor Patterns* |
| **SA** | Ácido Salicílico; *Salycilic acid* |
| **SAR** | Resistência Sistêmica Adquirida; *Systemic Acquired Resistance* |
| **SAGE** | Análises em Série da Expressão Gênica; *Serial Analysis of Gene Expression* |
| **SCMV** | Vírus do Mosaico da Cana-de-açúcar; *Sugarcane Mosaic Virus* |
| **SEAPA** | Secretaria de Estado de Agricultura Pecuária e Abastecimento |

| | |
|---|---|
| **SER/THRE** | Serina/Treonina; *Serine/Threonine* |
| **SNP** | Polimorfismo de Único Nucleotídeo; *Single Nucleotide Polymorphism* |
| **SOM** | Mapas Auto-Organizadores; *Self-Organizing Maps* |
| **SUCEST** | Projeto EST da Cana-de-açúcar *;Sugarcane EST Project* |
| **TAIR** | A Fonte de Informação sobre *Arabidopsis; The Arabidopsis Information Resource* |
| **TF** | Fator de Transcrição*; Transcription Factor* |
| **TIGR** | O Instituto para Pesquisa Genômica; *The Institute for Genomic Research* |
| **TIR** | Receptor Toll- Interleucina; *Toll Interleukine Receptor* |
| **TM** | Transmembrana; *Transmembrane* |
| **UFPE** | Universidade Federal de Pernambuco |
| **UPGMA** | Método não polarizado de Agrupamentos aos Pares com Médias Aritméticas *;Unweighted Pair Group Method with Arithmetic Means* |
| **UTR** | Região Não Traduzida; *Untranslated Region* |
| **WSSE** | Sistema Completo de Troca Instantânea; *Whole System Snapshot Exchange* |

# RESUMO

Os genes de resistência (*R; Resistance*) respondem pela primeira interação entre planta e patógeno, sendo responsáveis pela ativação ou não dos mecanismos de resistência em plantas, como o desencadear da resistência sistêmica adquirida (SAR; *Systemic Acquired Resistance*) e a ativação dos genes relacionados à patogenicidade (*PR; Pathogenesis Related*). Este trabalho analisou genes *R* e *PR* no transcriptoma da cana-de-açúcar, da soja e do feijão-caupi, geradas através de bibliotecas produzidas a partir de diferentes tecidos em várias fases de desenvolvimento. Após análise *in silico* foi possível a identificação de todas as classes de genes *R* em cana, soja e feijão-caupi, com destaque para a classe Sítio de Ligação de Nucleotídeo - Repetições Ricas em Leucina (NBS-LRR; *Nucleotide Binding Site - Leucine Rich Repeats*) nos três organismos. Quanto aos genes *PR,* a família mais representativa foi a *PR-2* em soja e *PR-9* em caupi. Em relação ao padrão de expressão, foram observados os genes *R* e *PR* em diferentes níveis em todos os tecidos analisados nas três espécies estudadas. Quando analisados através de alinhamentos múltiplos tanto os genes *R* quanto os *PR* apresentaram maior similaridade entre espécies pertencentes à mesma família, geralmente agrupando mono e dicotiledôneas em clados distintos, sugerindo que tenham surgido antes da separação entre essas classes; a distribuição e variação no número de cópias em cada espécie parecem ser atribuídas aos processos de duplicação e adaptação que ocorreram durante a evolução desses organismos. Os resultados do presente estudo colaboram com o desenvolvimento de marcadores moleculares para o melhoramento, visando o entendimento da abundância, diversidade e evolução destes genes, com ênfase das espécies estudadas, bem como para identificação dos genes *R* e *PR* em outras culturas de interesse econômico.

**Palavras-chave:** relação patógeno-hospedeiro, estresse biótico, angiospermas, bioinformática.

# ABSTRACT

Resistance (*R*) genes account for the first interaction between plant and pathogen, being responsible for the activation of the resistance mechanisms in plants such as the onset of systemic acquired resistance (SAR) and the activation of pathogenesis-related genes (*PR*). This study analyzed *R* and *PR* genes in the sugarcane, soybean and cowpea transcriptomes from libraries generated of different tissues at various stages of development. The *in silico* analysis allowed the identification of all *R* genes classes in sugarcane, soybean and cowpea, with prevalence of the Nucleotide Binding Site - Leucine Rich Repeats (NBS-LRR) class in all three organisms. As for *PR* genes, the most representative gene family was the *PR-2* in soybean and *PR-9* in cowpea. Additionally, it the expression at different levels from *R* and *PR* genes could be observed in all tissues analyzed, also in the three species studied. Analyzing multiple alignments, both *R* and *PR* families showed greater similarity between species that belong to the same family, usually grouping mono and dicots in distinct clades, suggesting that these genes arose before the separation between these classes. The distribution and variation in the number of copies in each species may be attributed to the replication and adaptation processes that occurred during these organisms evolution. These results collaborate with the development of molecular markers to breeding purposes and to improve the understanding of abundance, evolution and diversity of these genes, with emphasis on the species studied, as well as to identify *R* and *PR* genes in other economic important crops.



**Keywords:** host-pathogen relation, biotic stress, angiosperms, bioinformatics.

# INTRODUÇÃO

As plantas são conhecidas por se defenderem do ataque de patógenos utilizando tanto mecanismos de resistência constitutiva quanto de resistência induzida (Richter e Ronald, 2000). O arsenal de defesa inclui não só barreiras físicas representadas, por exemplo, pela parede celular como também a ativação de metabólitos secundários e proteínas antimicrobianas, que juntos são capazes de impedir a colonização do invasor (Richter e Ronald, 2000). Sabe-se que uma única interação entre planta e patógeno, seja ela compatível ou não, é capaz de recrutar ou silenciar centenas de genes, muitos deles já conhecidos, enquanto outros ainda não foram descritos (Benko-Iseppon et al., 2010).

Imediatamente após o reconhecimento do patógeno pela planta através da interação dos produtos dos genes de avirulência (*avr*; *avirulence*) e resistência (*R*; *Resistance*) respectivamente, a reação de hipersensibilidade (HR; *Hypersensitive response*) é ativada e o desenvolvimento da doença é interrompido; após essa interação, desencadeia-se uma cascata de sinalização que ativa os mecanismos da chamada Resistência Sistêmica Adquirida (SAR; *Systemic Acquired Resistance*), responsável por proteger a planta contra uma ampla gama de patógenos (Wanderley-Nogueira et al., 2007). Esta capacidade de "aprendizado" da planta fez com que há mais de 100 anos pesquisadores sugerissem que – da mesma maneira que os animais – as plantas pudessem ser imunizadas contra determinado micro-organismo quando entrassem em contato com outro micro-organismo ou molécula produzida por ele (Zipfel e Felix, 2005).

Depois do contato com o patógeno e da ativação do sistema específico de defesa, várias reações secundárias são desencadeadas, entre elas a produção de metabólitos, de lignina para reforçar a parede celular e a ativação dos genes relacionados à patogenicidade (*PR; Pathogenesis Related*) que, em parceria com os genes *R*, representam os principais mecanismos de defesa das plantas tanto contra estresses bióticos quanto abióticos (van Loon et al., 2006).

De acordo com a natureza de seus produtos, os genes *R* podem ser classificados em cinco grupos baseados nas estruturas e combinações de seus domínios conservados (Barbosa da Silva et al., 2005; Wanderley-Nogueira et al., 2007; Wanderley-Nogueira et al., 2011),

enquanto os genes *PR* agrupam-se em 17 famílias distintas, algumas delas observadas especificamente em determinadas plantas (Sels et al., 2008).

Visando à identificação e caracterização desses genes, a bioinformática torna-se indispensável. Por unir informática, biologia, estatística e química, o uso de suas ferramentas torna possível 'desvendar' os genomas de plantas de interesse, como por exemplo cana-de-açúcar, soja e feijão-caupi. É possível através da análise de suas sequências comparar genes de organismos relacionados, traçar perfis de expressão de genes e vias de interesse, fazer predições da estrutura e função de proteínas e de suas relações evolutivas entre outros. A genômica comparativa permite que a caracterização de um gene em determinado organismo auxilie na identificação de ortólogos em outros organismos de interesse econômico (Santos e Ortega, 2003). Em vista do exposto, a identificação dos genes *R* e *PR* é de grande relevância para o melhoramento genético de plantas, constantemente expostas ao ataque dos mais diversos agentes patogênicos.

A cana-de-açúcar (*Saccharum officinarum* L.) está entre as principais culturas vegetais, sendo cultivada em mais de oitenta países tropicais. O Brasil é o maior produtor mundial de cana-de-açúcar (cerca de 25% de toda a produção global), sendo o estado de Pernambuco um dos maiores produtores do país, respondendo o cultivo da cana-de-açúcar por 40% de sua economia (EMBRAPA, 2011).

A soja [*Glycine max* L.(Merrill)], por sua vez, é uma importante cultura global, responsável pela produção de óleo e de no mínimo duas vezes mais proteína por acre do que qualquer outro grão ou vegetal (Libault et al., 2010). Economicamente a soja é a fonte de proteína mais valiosa e a maior cultura de óleo comestível no mundo, com destacada importância para a produção de biodiesel, além de ser conhecida como modelo para desenvolvimento de vários processos vegetais (Cannon et al., 2009).

Adicionalmente, o feijão-caupi (*Vigna unguiculata* (L.) Walp.) é a planta que apresenta mais vantagens nutricionais dentre as tradicionalmente cultivadas e usadas pela população de áreas semi-áridas como o Nordeste do Brasil e várias regiões da África. Na década de 80 o Brasil era o segundo maior produtor de feijão-caupi no mundo, contribuindo com 26% da produção mundial. Algumas doenças limitam a produção de feijão-caupi no Nordeste brasileiro, especialmente as infecções provocadas por vírus como o do mosaico

severo, que podem reduzir até 80% da produção em cultivares mais suscetíveis (Cândido e Silva, 2008).

O presente trabalho visou identificar, caracterizar e analisar estruturalmente com o auxílio de ferramentas computacionais, sequências candidatas a genes *R* e *PR* presentes no genoma expresso da cana-de-açúcar, da soja e do feijão-caupi comparando-as às demais sequências depositadas em bancos de dados e descritas na literatura. Os resultados obtidos podem contribuir de forma significativa para o desenvolvimento de marcadores moleculares aplicáveis ao melhoramento, bem como para a identificação de genes *R* e *PR* em outras plantas cultivadas de interesse econômico.

# OBJETIVOS

**- Objetivo Geral**

- Identificar, caracterizar e analisar a estrutura de genes *R, PR* e relacionados à SAR, e outros que venham a fazer parte da relação patógeno-hospedeiro em Angiospermas, avaliando padrões de evolução e expressão.

**-Objetivos Específicos**

**1.** Selecionar na bibliografia relacionada e nos bancos de dados públicos seqüências dos principais genes *R, PR* e demais genes relacionados à SAR, usando-as como seed sequence (sonda) para a identificação de ortólogos no banco de dados do SUCEST, GENOSOJA e NORDEST.

**2.** Identificar e descrever os domínios, motifs e regiões conservadas das seqüências identificadas, comparando-as com os depositados nos bancos de dados.

**3.** Reconhecer e analisar quadros de leitura aberta das seqüências selecionadas.

**4.** Reconhecer padrões evolutivos a partir da análise das seqüências mineradas, através de alinhamentos múltiplos e inferências fenéticas e filogenéticas.

**5.** Estabelecer um perfil da expressão *in silico* dos genes estudados, a partir da análise de sua presença/ausência nos diferentes tecidos e condições de isolamento efetuados na montagem do banco de ESTs de cana-de-açúcar, soja e feijão-caupi.

**6.** Fazer inferências sobre padrões estruturais e a evolução desses genes em plantas superiores.

# REVISÃO BIBLIOGRÁFICA

## 1 – O ataque de patógenos e as perdas causadas por doenças

A existência de doença constitui-se em uma condição anormal que prejudica a planta ou a impossibilita de desempenhar suas funções fisiológicas normais. As doenças são facilmente reconhecidas pelos seus sintomas, frequentemente associados às visíveis mudanças que podem ocorrer na morfologia do vegetal (IRRI, 2011).

Conforme levantamentos da FAO (2011) fungos, bactérias, vírus e nematoides são os principais causadores das doenças, resultando em perdas na agricultura e danos na paisagem, reduzindo assim a produção, a qualidade e a durabilidade, diminuindo não só a estética e o valor nutricional dos alimentos, como também a produtividade e os lucros obtidos com a sua comercialização; além disso, apesar do esforço dos produtores que gastam altas somas para o controle parcial dos diversos patógenos que atacam suas culturas os danos na produção e nas mercadorias (Oerke et al*.,* 1994; ISAAA, 2011) resultam anualmente em prejuízos de bilhões de dólares. As perdas representam de 31 a 42% dos 1,3 trilhões de dólares anuais, podendo chegar a 48% nos países em desenvolvimento, onde a situação é ainda mais crítica (FAO, 2011).

Em contraste com os altos gastos existentes com a aplicação de pesticidas e com os danos gerados ao meio ambiente devido ao uso dos mesmos, deve-se considerar que a engenharia genética pode potencialmente reduzir estes custos, produzindo plantas resistentes a determinados patógenos, uma vez que a resistência natural possui várias vantagens óbvias em relação ao uso de produtos químicos ou a outro método qualquer utilizado no controle de pestes (Fermin-Muñoz et al*.,* 2000). O fato de ser altamente eficiente, benigna ao meio ambiente, não prejudicar a paisagem e representar pouca ou nenhuma despesa adicional aos produtores torna a resistência provavelmente a estratégia de controle mais desejada pelos melhoristas hoje em dia (Shepherd, 2011).

## 1.1- Resistência contra patógenos

### 1.1.1- Mecanismos de defesa

Estudos na última década têm revelado que o sistema de defesa das plantas consiste em diferentes níveis de proteção que evoluíram através da constante batalha de co-evolução entre plantas e seus patógenos (Postel and Kemmerling, 2009). Uma vez que as plantas não possuem um sistema imune como o dos animais, as mesmas precisaram desenvolver uma grande variedade de estratégias na defesa contra estresses bióticos e abióticos (Bolton, 2009).

A primeira estratégia é física, compreende a formação de uma barreira entre a célula e o patógeno, através do fortalecimento da parede celular com a produção de enzimas ligadas à biossíntese de lignina, formação de camadas de cortiça bem como a formação de calos após o ataque de um micro-organismo (Chisholm et al., 2006). Segundo Heath (1991), as plantas apresentam resistência à maioria dos micro-organismos potencialmente patogênicos, a chamada resistência inata, sugerindo que essa resistência possua diversos componentes e que seja complexa e não patógeno-específica. Segundo Salvaudon et al. (2005) esta também seria a segunda estratégia para se defender das doenças: a produção de metabólitos secundários e enzimas hidrolíticas como componentes antimicrobianos, chamada desta vez de resistência não específica ou "resistência de campo"; tais componentes seriam produzidos uma vez que a planta detectasse a presença de padrões moleculares associados a micróbios (MAMPs; *Microbe-Associated Molecular Patterns*) ou padrões moleculares associados a patógenos (PAMPs; *Pathogen-Associated Molecular Patterns)* (Miya et al., 2007). Estes padrões também são os responsáveis pela existência da relação de simbiose entre alguns microrganismos e seus hospedeiros, pois a mesma só se torna possível uma vez que ocorra o reconhecimento das moléculas sinalizadoras da planta por parte do patógeno (Brencic e Winans, 2005). Nos últimos anos, vários aspectos do que agora se é chamado de SAR; vêm sendo elucidados. Entretanto, considera-se que a HR se constitui no principal mecanismo utilizado pelas plantas para se defenderem contra o ataque de patógenos (Bonas e Lahaye, 2002), sendo desencadeada pela "ativação" de um gene *R* e caracterizada por morte celular rápida e localizada impedindo que o patógeno se espalhe e, desta forma, colonize o organismo (Meyers, 2005).

Estudos de expressão diferencial apontam que mudanças substanciais na expressão gênica do hospedeiro são detectadas após o contato com diversos tipos de patógenos e que a indução dessa ampla gama de estratégias de defesa demanda uma redistribuição massiva de energia durante todo o processo (Bolton, 2009; Soto et al*.,* 2009).

### 1.1.2- A interação gene-a-gene

Flor, em 1942, trabalhando com o patógeno *Melampsora lini* em plantas de linho foi o primeiro a estudar a genética da resistência contra doenças baseada na HR, tendo proposto um modelo chamado interação gene-a-gene para demonstrar a relação entre hospedeiro e patógeno.

A interação gene-a-gene determina que o gene de *R* dominante da planta confere resistência a uma variedade de patógenos biotróficos, incluindo vírus, desde que os mesmos possuam o gene *avr* dominante correspondente; desta forma, a planta será resistente e o crescimento do patógeno será interrompido apenas quando ambos os genes, *R* e *avr* forem compatíveis, desencadeando, assim, a reação de hipersensibilidade (Moffet, 2009) (Figura 1).

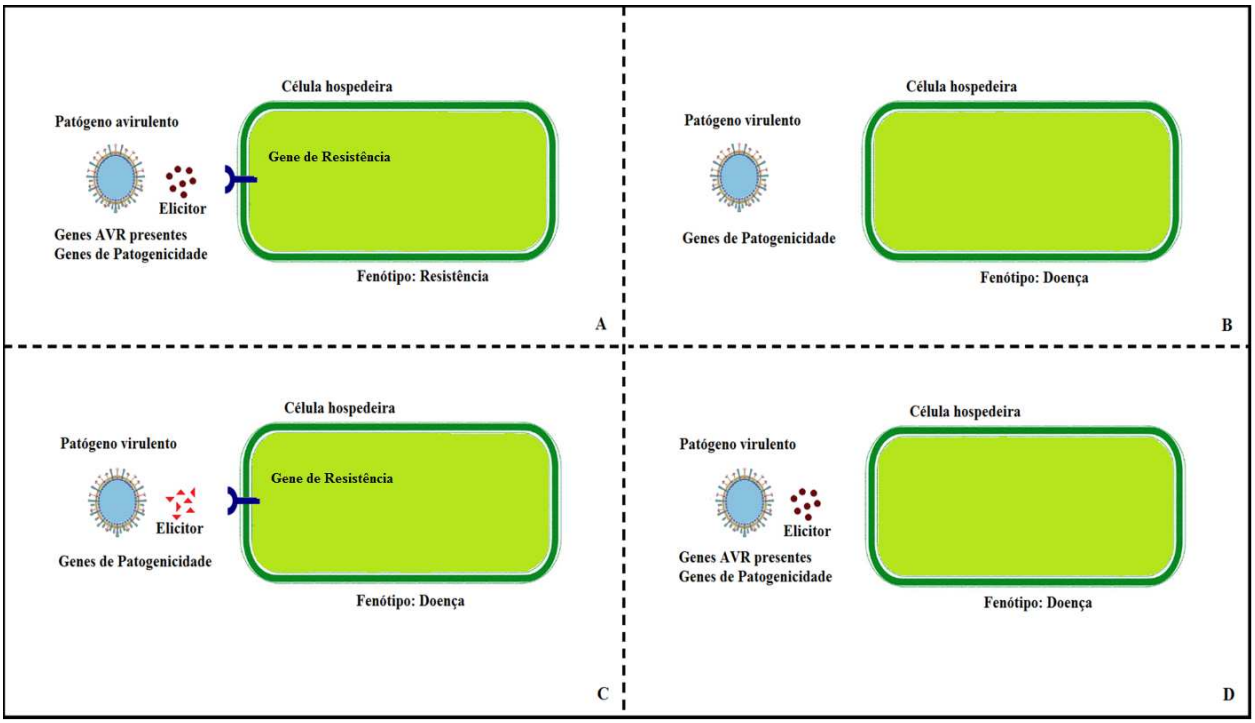



Figura 1: Esquema representativo da interação gene-a-gene. Em (A) observa-se uma interação compatível que confere resistência onde estão presentes o gene *R* e o gene *avr* do patógeno correspondente. Em (B) não há produto nem do gene *R* nem do *avr*; em (C) há produto do gene *R* e de um gene *avr* não correspondente e em (D) há produto apenas do gene *avr* e ausência de gene de resistência. Nos três últimos casos o fenótipo é suscetível.

Nos primeiros 15 minutos o vegetal engatilha uma resposta que induz fluxos iônicos através da membrana plasmática, 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. Após esse tempo, ocorre a síntese de ácido salicílico e etileno e a síntese de compostos antimicrobianos como as fitoalexinas, culminando na morte celular programada (Benko-Iseppon et al., 2010; Sanabria et al., 2010). O gene *R* da planta se refere ao gene que codifica um receptor ou uma enzima responsável pela transdução de sinais para o reconhecimento das moléculas elicitoras (produtos dos genes *avr*) do patógeno (Tang et al., 1999); a planta que não possuir tal gene é chamada de suscetível. Quando ocorre mudança na molécula produzida pelo gene *avr* mutado do patógeno, o mesmo passa a ser considerado virulento, sendo capaz de infectar tanto os hospedeiros antes resistentes, quanto os suscetíveis (de Wit, 2007).

### 1.1.3- Os genes de resistência e suas classes

Em contraste com a grande diversidade dos produtos dos genes de *avr* (Ellis e Jones, 2000) os genes *R* são extremamente conservados e podem ser agrupados em quatro classes distintas de acordo com a natureza de seus produtos além da classe das redutases, que não apresenta estrutura conservada (Hammond-Kosak e Jones, 1997; Morais, 2003; Liu et al., 2004) (Figura 2). Estudos recentes sugerem que os genes *R* 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 (Moffet, 2009).

A primeira classe, chamada classe das Redutases, é representada pelo gene *HM1* do milho, o qual codifica uma redutase NADPH dependente que inativa as toxinas produzidas pelo fungo *Helminthosporium carbonum*. Estudos filogenéticos mostram que este gene se desenvolveu apenas na linhagem das gramíneas (Poaceae), sendo restrito a esta família de plantas (Sindhu et al*.,* 2008). Outros representantes da classe das redutases incluem: (a) o gene *Mlo* de cevada, um provável regulador de defesa contra *Blumenaria graminis* que não teve ainda elucidado seu mecanismo de funcionamento (Reinstädler et al*.,* 2010) e(b) o gene *RPW8* de *Arabidopsis thaliana* L., que codifica uma redutase localizada na membrana, com uma estrutura helicoidal, sem semelhança alguma com qualquer outro gene de resistência (Kobe e Kajava, 2001).

A segunda classe, chamada de classe Quinase, é representada pelo gene *Pto* de *Lycopersicon esculentum* Mill.,que confere resistência à bactéria gram-negativa *Pseudomonas syringae* pv. *tomato*, causadora da mancha foliar do tomate (Salomon et al., 2009). *Pto* foi o primeiro gene de resistência clonado de uma planta que conferia reconhecimento de um patógeno específico (Oh e Martin, 2011). Esta classe caracteriza-se pela presença de uma quinase que fosforila resíduos de serina (*ser*) e treonina (*tre*), capaz de interagir fisicamente com o produto do gene *avrPto* (Tang et al*.,* 1999), tendo sido também caracterizado em outros organismos, como *A. thaliana* e *Phaseolus vulgaris* (Melotto et al., 2004).

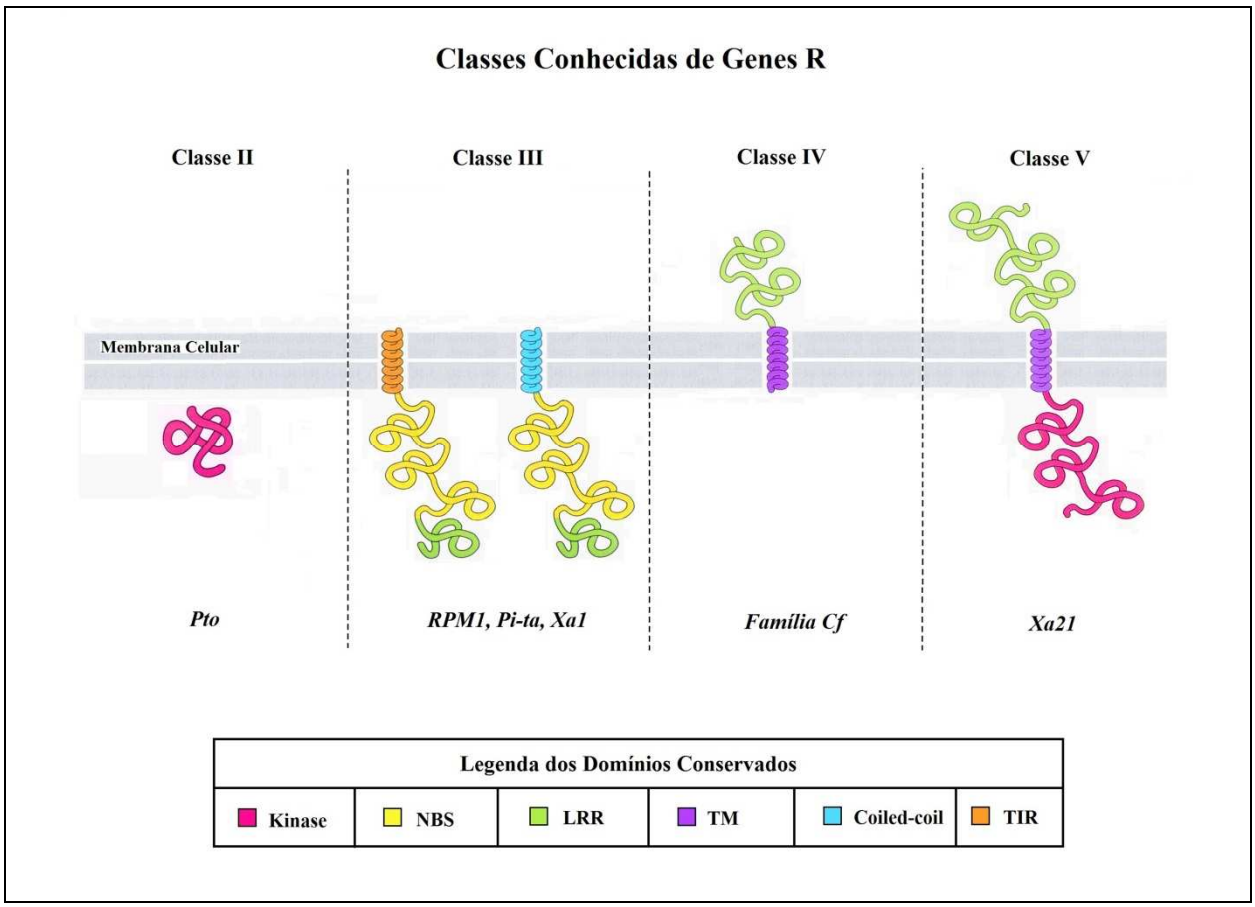



Figura 2: Representação gráfica das quatro famílias de Genes de Resistência que apresentam domínios conservados em suas estruturas. São elas: II=Classe Quinase; III=Classe NBS-LRR; IV=Classe LRR; V=Classe LRR-Quinase. As diferentes cores representam os domínios conservados. Abreviações: NBS: Sítio de Ligação de Nucleotídeo; LRR: Repetições Ricas em Leucina; TM: Domínio Transmembrana; TIR: Receptor Toll/Interleucina (esquema da autora).

A terceira e maior classe de genes *R* é caracterizada por representantes que sintetizam proteínas que contêm domínios chamados Sítio de Ligação de Nucleotídeo (NBS; *Nucleotide Binding Site*) e Repetições Ricas em Leucina (LRR; *Leucine Rich Repeats*), as quais se dividem em duas subclasses com base na estrutura do seu domínio N-terminal (Tarr e Alexander, 2009; Wan et al., 2010).

O domínio LRR está envolvido no reconhecimento do patógeno pelo hospedeiro durante o processo de infecção. Proteínas que contêm LRRs apresentam motivos do aminoácido leucina, repetidas em número de 20 a 30, a intervalos regulares, que agrupados em tandem formam o domínio, podendo conter ainda outros resíduos hidrofóbicos, asparaginas e prolinas (Sun e Wang, 2011). Com base em suas sequências conservadas, os domínios LRR

podem ser classificados em sete subfamílias, das quais apenas uma é específica de plantas (Bella et al., 2008). Os motivos de leucina geram uma estrutura terciária semelhante a uma mola, com cada hélice apresentando um motivo de leucina repetido. A função primária desse domínio parece ser a de 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 envolvidos no processo de resistência (Banerjee et al., 2001). Além da interação proteína-proteína, os domínios LRR também funcionam como sítios de união peptídeo ligante e sítios de interação proteína-carboidrato, relacionados à especificidade da resistência (Kobe e Kajava, 2001). Supõe-se que a porção amino-terminal do domínio LRR, que é altamente variável, possua um papel mais importante no reconhecimento do que a porção carboxi-terminal, que é mais conservada (Lurderer, 2001).

Os NBS, também chamados de P-loops ou NB-ARC, estão presentes em diversas proteínas e são responsáveis por atividade de ligação ao ATP ou GTP, tais como subunidades da ATP sintetase e fatores de elongação do ribossomo (Tameling et al*.,* 2010). A presença desses domínios em genes de resistência sugere que estes sítios estão envolvidos na ativação de proteínas quinases ou proteínas transportadoras (Hammond-Kosack e Jones, 1997), observando-se que mutações em sua estrutura impedem que ocorra a resposta hipersensível, ressaltando a importância da função deste domínio na sinalização (Tsuda e Katagiri, 2010).

A primeira subclasse possui o domínio Cauda Espiralada (CC; *Coiled-coil*) na região amino-terminal, responsável pelo reconhecimento das moléculas elicitoras; são representantes desta classe os genes *Rps2*, *RPP8, RPP13* e *Rpm1* de *A. thaliana; Pib, Pi-ta* e *Xa1* de arroz; *Prf, I2, Mi* e *Sw5* de tomate e *Hero* de batata (Liu et al., 2004). A segunda subclasse possui em sua região amino-terminal o domínio Receptor Toll/Interleucina (TIR; *Toll Interleucine Receptor*) que também está presente em animais, acreditando-se que seja inexistente em monocotiledôneas ou que tenha sido perdido neste grupo. Apesar dos estudos terem sido limitados às espécies economicamente importantes da família das gramíneas, resultados adicionais para *Zingiber* e *Musa* permitem extrapolar as suposições e generalizar a informação para a classe das monocotiledôneas (Tarr e Alexander, 2009) enquanto todas as espécies de dicotiledôneas atualmente estudadas apresentaram este domínio (Goff et al., 2002). Exemplos dessa subclasse tem como representantes os genes *L* (Lawrence et al*.,* 1995) e *P* (Dodds et al*.,*

2001) de linho; *RPP1* (Botella et al*.,* 1998), *RPP4* (van der Biezen et al*.,* 2002), *RPP5* (Parker et al*.,* 1997) e *RPS4* (Gassmann et al*.,* 1999) de *A. thaliana* e *N* (Mestre e Baulcombe, 2006) de tabaco. Apesar de ter sido proposto um papel de sinalizador para o domínio TIR, evidências de que o domínio CC possa desencadear uma resposta sinalizadora ainda são duvidosas (Swiderski et al., 2009); recentemente foi mostrado que o domínio NB-ARC pode desencadear uma resposta de defesa na ausência dos outros dois domínios, sugerindo que o mesmo possa ser responsável pela cascata de sinalização dos genes desta classe por si só (Tameling et al., 2010).

A quarta classe de genes de resistência é representada pela família gênica *Cf* (*Cf-2*, C*f-4 Cf-5 e Cf9)* de tomate, que confere resistência ao fungo *Cladosporium fulvum*, sendo chamada também de classe das proteínas tipo-receptor (RLP; *Receptor-like proteins)* (Stergiopoulos et al., 2010). Esta classe codifica genes que possuem um domínio transmembrana (TM; *Transmembrane Domain*) com o domínio LRR na porção extracelular e uma pequena cauda carboxi-terminal na região intracelular (Kruijt et al*.,* 2005).

A quinta classe é representada pelo gene *Xa21* de arroz (Song et al*.,* 1997) que codifica um receptor tipo quinase caracterizado por um domínio LRR extracelular, como os genes da classe IV, uma região TM e uma ser/tre quinase intracelular, como os genes da classe II; desta forma, a estrutura do gene *Xa21* parece indicar uma ligação evolutiva entre diferentes classes de genes *R* em plantas (Xu et al*.,* 2006). O domínio LRR na maioria dos genes LRR-quinases conhecidos não possui nenhum íntron, ao contrário dos LRR-quinases tipo ERECTA, genes ligados ao desenvolvimento, que são interrompidos por íntrons na primeira leucina do padrão 'xxLxLxx' (Sun e Wang, 2011).

Essa classificação, entretanto, varia de acordo com cada autor. Podemos observar que enquanto Ellis et al. (2000), Morais (2003), van Leeuven et al. (2005), Barbosa da Silva et al.,(2005) e Wanderley-Nogueira et al. (2007) utilizam a classificação acima, Salvaudon et al. (2005) consideram as duas subclasses da terceira classe como classes distintas, enumerando então, seis classes de genes de resistência, enquanto Jones (2001) enumera cinco, mas não mantém a mesma sequência.

### 1.1.4- A evolução dos genes de resistência

Em muitas espécies cultivadas os genes de resistência organizam-se em grupamentos nos cromossomos, denominados *clusters* (Michelmore et al., 1987). Este fenômeno é bem conhecido e estudos moleculares têm demonstrado que esta clusterização geralmente reflete eventos de duplicação cromossômica de regiões que abrigam genes de resistência ancestrais, criando *loci* ricos em genes de resistência (Meyers et al., 2005). As recombinações intragênicas e extragênicas – devido à ocorrência de *crossing over* desigual – são as responsáveis por gerar a diversidade de haplótipos e novas especificidades de resistência (Friedman e Baker, 2007).

Teoricamente *clusters* de genes funcionariam como um reservatório de mutações. No genoma de Arabidopsis, 182 genes formam 20 *clusters* de genes, sendo que quinze destes possuem domínios do tipo LRR (Bergelson et al*.,* 2001). As LRRs provavelmente evoluíram a partir de duplicação, mutação e recombinação de éxons. Por estarem envolvidas no reconhecimento e especificidade de proteínas, são prováveis alvos de pressão e seleção adaptativa (Sun e Wang, 2011). Por exemplo, em tomate, os genes *Pto e Prf* também aparecem em *clusters* (Hulbert, 2001). Para as monocotiledôneas, embora não haja uma colinearidade em relação aos ortólogos nos mapas de arroz, cevada e milho, existe uma organização geral comum entre essas espécies (Hulbert, 2001).

A grande similaridade entre sequências da maioria dos genes *R* clonados, mesmo entre espécies distantes, pode confirmar a hipótese de que os mesmos pertencem a uma grande família multigênica que divergiu após eventos de duplicação e mutação, decorrendo em diferentes especificidades (Richly et al., 2002), surgindo, porém, de um ancestral comum. Tal fato também é refletido quando em determinada análise de similaridade (tanto molecular quanto *in silico)* duas sequências ortólogas, de organismos diferentes como cana-de-açúcar e milho apresentam-se mais parecidas entre si, do que duas sequências de cana-de-açúcar, por exemplo; como ocorre com o gene *RPR1* (Rossi et al., 2003) e também com os genes *Pto, Xa1, Cf-9* (Wanderley-Nogueira et al.,2007) e *Xa21* (Wanderley-Nogueira et al*.,* 2011). Por outro lado, Pryor e Ellis (1993) sugerem que a diversidade da especificidade da resistência

seja reflexo de uma taxa evolutiva muito mais rápida que a observada em mutações espontâneas, principalmente decorrente da pressão de seleção do patógeno sobre o hospedeiro.

A taxa de recombinação nos *R-clusters* não é necessariamente homogênea ou consistente na escala evolutiva. Evidências mais recentes mostram que a recombinação nesses *clusters* é aumentada quando ocorre infecção por um patógeno, sugerindo um mecanismo que induz temporariamente uma instabilidade no genoma em resposta a um estresse extremo. Metilações no DNA e modificações na cromatina podem permitir que tal instabilidade seja regulada e restrita a determinadas regiões do genoma (Sun e Wang, 2011).

Polimorfismos em alelos de genes de resistência podem ser determinantes para uma determinada característica de resistência ou suscetibilidade. Alelos de suscetibilidade geralmente ocorrem por mutação em alelos de resistência (Bergelson et al., 2001). Até o momento pouco se sabe sobre a dinâmica da evolução desses genes. No loco *RPS2* os alelos de resistência são geneticamente mais similares entre si do que os alelos de suscetibilidade. Num estudo de evolução molecular, o gene *RPS2* apresentou um nível de polimorfismo de 1,26%, sendo que aproximadamente metade deste polimorfismo resultou em alteração de aminoácidos e 70% foram alterações não conservadas. Segundo Caicedo et al.(1999) este nível de polimorfismo geralmente não é encontrado em plantas, demonstrando que provavelmente este seja um gene que evolui rapidamente. Entretanto, o loco *RPM1* não foi observado em linhagens de *A. thaliana* suscetíveis a *Pseudomonas syringae* pv. maculicola. Análises de regiões adjacentes ao gene em *A. thaliana* e *A. lyrata* (L.) revelaram divergência de sequência de 10%, um valor alto que indica que provavelmente o polimorfismo foi gerado no momento da separação entre as duas espécies (Bergelson et al., 2001).

### 1.1.5 – Genes relacionados à patogenicidade

Proteínas PR são componentes do arsenal de defesa das plantas e têm sido rotineiramente utilizados como marcadores da SAR após o contato com organismos patogênicos (Figura 3) (Zhang et al., 2010). Em 1970, proteínas PR foram encontradas pela primeira vez em folhas de tabaco infectadas e depois disso passaram a ser descobertas e estudadas em uma grande

variedade de organismos (van Loon e Kammen, 1970). As proteínas PR formam um grupo heterogêneo e são codificadas por genes que são rapidamente induzidos por infecções patogênicas e pelo acúmulo de ácido salicílico (AS; *Salycilic acid*), ácido jasmônico (JA; *Jasmonic acid*) e etileno (ET; *Ethylene*) (Narusaka et al*.,* 2009). Estudos recentes têm mostrado que os genes *PR* também são regulados por fatores ambientais, incluindo estresses abióticos e alterações na luminosidade, bem como por fatores de desenvolvimento, desempenhando um papel importante também nessas duas situações (Seo et al*.,* 2008; Benko-Iseppon et al., 2010).

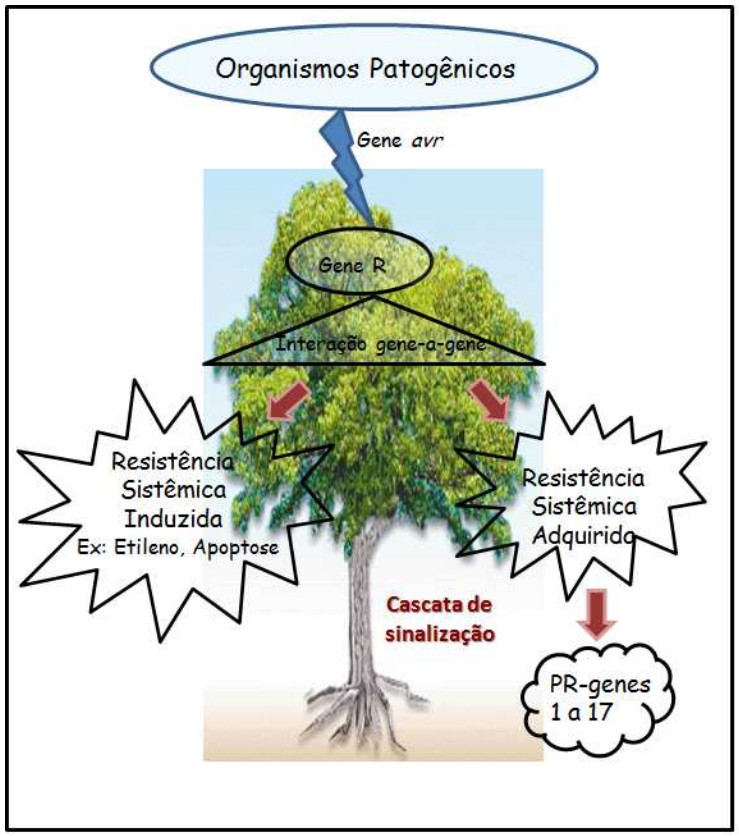



Figura 3. Principais mecanismos de reconhecimento e defesa em plantas. O patógeno secreta o gene *avr* que pode ser compatível com o produto do gene *R* da planta. Interações compatíveis levam à ativação de uma cascata de sinalização induzindo os fatores da resistência sistêmica (como etileno e ácido jasmônico) e da resistência adquirida, representada pelas 17 famílias de genes *PR*. Adaptado de Benko-Iseppon et al. (2010).

As proteínas PR classificam-se em 17 famílias gênicas, formadas por subfamílias ácidas e básicas, localizadas no vacúolo ou extracelulares (Cutt e Klessig, 1992). A similaridade entre sequências, as relações serológicas ou imunológicas e as propriedades enzimáticas são a base

para essa classificação (Van Loon et al*.,* 1999). Funcionalmente, algumas proteínas PR possuem atividade antifúngica, como as quitinases (PR-3) e as β-glucanases (PR-2) (Zhu et al*.*, 1994), algumas já foram bem caracterizadas e estudadas, como as PR-5 (taumatinas) e PR-8 (quitinases tipo III), mas outras ainda têm sua função bioquímica desconhecida como é o caso da PR-1 em tabaco (Gaffney et al*.,* 1993), *Arabidopsis* (Metzler et al*.,* 1991), tomate (Tornero et al*.,* 1997) e maçã; entretanto, a presença de proteínas PR1 é usada rotineiramente como um marcador da SAR. Estudos com plantas frutíferas permitiram o isolamento de cDNAs dos genes *PR-1* e *PR-5* em pêra e *PR1, PR-2, PR-5*e *PR-8* em maçã; tais genes apresentaram-se superexpressos quando as plantas foram submetidas ao ataque da bactéria *Erwinia amylovora* (Bonasera et al*.,* 2006). Em plantas herbáceas a ativação dos genes *PR-1, PR-2, PR-5 e PR-8* e consequentemente da SAR, é bastante influenciada pelo acúmulo de ácido salicílico (SA) em seus tecidos (Mei et al*.*, 2006); em pepino, o gene *PR-8* é altamente induzido pelo SA e seus análogos funcionais (Spoel e Dong, 2008). Os genes do grupo *PR-13* (defensinas ou gama-tioninas) compreendem pequenos peptídeos que também já foram caracterizados e isolados de raízes, sementes, flores e folhas, demonstrando atividade contra fitopatógenos e também contra bactérias patogênicas em humanos (Pelegrini et al*.*, 2011).

Em relação ao estresse abiótico, estudos recentes mostraram que o gene *PR-3* é induzido de maneira significante por altas concentrações de sal, uma vez que plantas que tiveram este gene nocauteado apresentaram uma queda significativa na taxa de germinação de sementes em ambiente salino. Por esta razão foi proposto que o *PR-3* possa mediar os sinais que afetam a germinação de sementes como a resposta ao estresse salino, uma vez que é ABA-dependente em *Arabidopsis*. *PR-4 e PR-5* também participam da regulação da germinação em ambientes salinizados; entretanto, seus efeitos não foram tão evidentes quanto os obtidos para o *PR-3* (Seo et al*.,* 2008). Já foi observado também em monocotiledôneas como cevada, trigo e outras gramíneas, que um grupo de genes *PR* foi superexpresso quando as plantas foram expostas a baixas temperaturas, sugerindo que tais genes também podem estar ligados à resposta ao frio extremo; entretanto, tais respostas ainda não foram extensivamente exploradas (Griffith and Yaish, 2004).

Os genes *PR* apresentam um padrão de expressão distinto tanto entre tecidos como entre diferentes organismos, sugerindo que mais de uma simples via metabólica regule cada família

de genes *PR* ao mesmo tempo (Zhang et al., 2010). Os genes relacionados a essas vias são bastante conservados dentro do Reino Vegetal em relação ao tamanho, à composição de aminoácidos e ao ponto isoelétrico (Bonasera et al., 2006), havendo alguns componentes do sistema que apresentam similaridade com proteínas envolvidas na imunidade inata presente no Reino Animal (Nurnberg e Brunner., 2002).

## 2- A cultura da cana-de-açúcar

### 2.1- Origem, história e citogenética

A cana-de-açúcar é uma planta perene e alógama (Berding e Roach, 1987), classificada taxonomicamente como membro da divisão Embryophyta, incluída na subdivisão Angiospermae, classe Monocotyledoneae, família Poaceae (gramíneas), tribo Andropogoneae e gênero *Saccharum* (Dillon et al., 2007).

As espécies de cana-de-açúcar cultivadas atualmente (*Saccharum* spp.) resultaram de hibridizações interespecíficas envolvendo *S. officinarum*, *S. barberi*, *S. sinense* e as espécies selvagens *S. spontaneum* e *S. robustum,* assim classificadas botanicamente por Brett (1957). Acredita-se que *S. officinarum* foi originalmente domesticada pelo homem, em Papua Nova Guiné por volta de 2500 a.C. (Brandes, 1956), provavelmente a partir do germoplasma de *S. robustum*. Entretanto, Roach e Daniels (1987) e Matsuoka et al. (1999) afirmaram que a cultura teria surgido a partir de *S. spontaneum* ou *Miscanthus* spp. Devido à sua origem multiespecífica, a cana-de-açúcar é conhecida por apresentar um dos genomas mais complexos entre as plantas cultivadas (Ingelbrecht et al., 1999). Tal complexidade é refletida nas características cariológicas das cultivares modernas de cana-de-açúcar, derivadas de cruzamentos de *S. officinarum* (ca. 2n = 80 cromossomos) e *S. spontaneum* (2n = 40-128) cromossomos. Em vista das diferenças cariotípicas entre as duas espécies, os híbridos apresentam proporções variáveis dos dois genomas, com 2n = 100 até 130 cromossomos (Grivet e Arruda, 2002), o que impõe dificuldades significativas na aplicação de melhoramento convencional a esta cultura (Vettore et al., 2001).

A cana-de-açúcar que foi introduzida no Brasil no início do século XVI corresponde a clones híbridos de *S. officinarum* e *S. barberi*, denominada cana-crioula ou cana-da-terra. Entretanto, este híbrido se extinguiu devido à grande suscetibilidade ao vírus do mosaico (Artschwager e Brandes,1958; Berding e Roach, 1987). Posteriormente, as canas-nobres, termo criado por melhoristas holandeses para se referir a genótipos de *S. officinarum* com alto teor de açúcar, dominaram a economia do país e constituíram a principal base para a indústria do açúcar não só no Brasil, como também no mundo. O fim do ciclo da cana-caiana, como era chamada, ocorreu por volta de 1850, após uma epidemia de gomose, doença causada pelo patógeno *Xanthomonas axonopodis* pv. *vasculorum* (Dantas, 1960), o que levou ao surgimento dos primeiros programas de melhoramento da cana-de-açúcar no Brasil.

Praticamente todas as cultivares atuais são férteis e possuem número cromossômico variando entre 2n=70 e 2n=130, representando um alopoliploide segmental (Roach e Daniels, 1987; Wang et al., 2010). Em cana-de-açúcar não se pode assumir herança diploide, entretanto, segundo Hogarth (1987), evidências sugerem herança semelhante aos diploides. É comum encontrar plantas do gênero *Saccharum* apresentando variações quanto ao número cromossômico não só entre células de uma mesma planta, como também entre células de um mesmo tecido (Portieles et al., 2002). Este fenômeno é chamado mosaicismo (Heinz et al.,1969). Supõe-se que o genoma básico da cana-de-açúcar seja composto por 10 cromossomos (x=10), como a maioria das gramíneas, havendo, porém, suposições de que o número básico poderia ser x=8 e x=12 (Ingelbrecht et al., 1999).

## 2.2- As doenças da cana-de-açúcar

No Brasil foram relatadas 58 entre todas as 216 doenças relacionadas em cana-de-açúcar em todo o mundo (EMBRAPA, 2011). As doenças mais importantes são controladas com o uso de variedades tolerantes ou resistentes. Entretanto, como a resistência a doenças nessa cultura apresenta na maioria das vezes caráter quantitativo e não qualitativo; muitas variedades em cultivo podem apresentar certo nível de suscetibilidade a algumas doenças. Como os causadores de doenças são seres vivos, eles podem produzir, também,

novas raças ou variantes que vencem essa resistência e passam a causar novo surto de doença (EMBRAPA, 2011).

Historicamente, no mundo, o carvão, o raquitismo das soqueiras, a escaldadura das folhas e o mosaico da cana-de-açúcar são consideradas as doenças mais importantes que atacam esta cultura (Sanguino, 1998).

O carvão da cana-de-açúcar, causado por *Ustilago scitaminea*, tem ocorrência generalizada no Brasil (Santos, 2008). Os danos causados pelo carvão são variáveis, mas podem causar perdas de até 100% em variedades suscetíveis (Tokeshi, 1997). Comstok e Lentini (2002) afirmaram que certas regiões canavieiras podem permanecer por muitos anos sem relatos de carvão. No entanto, a doença pode reaparecer e devastar rapidamente áreas com variedades suscetíveis. Os danos causados pelo fungo incidem tanto na redução da produção como na perda de qualidade do caldo.

O raquitismo da soqueira é causado pela bactéria *Leifsonia xyli* subsp. *xyli*, antes chamada de *Clavibacter xyli* subsp. *xyli*. Muitos pesquisadores consideram o raquitismo das soqueiras a mais importante doença afetando a cana-de-açúcar em todo o mundo. Pode causar prejuízos de até 100% da produtividade, ainda que o produtor desconheça que seu campo esteja infectado, pois não há ocorrência de sintomas típicos (Almeida, 2008).

A escaldadura das folhas é uma doença provocada pela bactéria *Xanthomonas albilineans* e que tem grande potencial destrutivo, principalmente em variedades suscetíveis. No Brasil, sua importância tem sido subestimada devido aos erros de identificação e à confusão de seus prejuízos com aqueles causados pelo raquitismo das soqueiras (Dinardo-Miranda, 2008).

O principal problema de origem viral da cultura canavieira é o mosaico da cana-de-açúcar. O mosaico consiste em um subgrupo de quatro espécies distintas de *potyvirus*, mas no Brasil, apenas a espécie SCMV é encontrada causando mosaico em cana-de-açúcar (Gonçalves et al., 2004). A transmissão natural do vírus se dá por meio de pulgões e disseminação pela utilização de mudas de canas infectadas, seja para a formação de viveiros ou canaviais comerciais (EMBRAPA, 2011); esse vírus incide também sobre outras gramíneas, dentre as quais o milho, *Zea mays* L. (Costa et al. 1971) e o sorgo, *Sorghum bicolor* L. (Moench) (Pinto, 1984).

## 2.3- Produção e importância econômica

A cana-de-açúcar é uma das principais culturas do mundo, sendo cultivada em mais de 120 países (Matsuoka et al*.,* 1999), incluindo regiões tropicais e subtropicais, representando a mais importante fonte de açúcar e álcool. Em 2006 o Brasil foi responsável por 25% das 1 bilhão e 200 milhões de toneladas de cana-de-açúcar cultivadas anualmente no planeta, o que representa mais de sete milhões de hectares plantados, produzindo mais de 480 milhões de toneladas de cana, colocando o país na liderança mundial em tecnologia de produção de etanol (EMBRAPA, 2011) Estima-se que a produção brasileira deva dobrar na próxima década visando acompanhar essa tendência (Cheavegatti-Gianotto et al., 2011).

Devido à grandeza dos números do setor sucro-alcooleiro no Brasil, não se pode tratar a cana-de-açúcar, apenas como mais um produto, mas sim como o principal tipo de biomassa energética, base para todo o agronegócio sucro-alcooleiro, representado por 350 indústrias de açúcar e álcool e 4,5 milhões de empregos diretos e indiretos em todo o Brasil (PROCANA, 2011). Além de matéria-prima para a produção de açúcar e álcool, seus subprodutos e resíduos são utilizados para co-geração de energia elétrica, fabricação de ração animal e como fertilizantes para as lavouras (EMBRAPA, 2011).

São Paulo é o maior produtor, seguido por Alagoas e Pernambuco, sendo a cana-de-açúcar plantada na zona da mata, numa vasta extensão de terras denominada zona canavieira do estado. Próxima ao oceano Atlântico, essa área atinge 12 mil $km^2$ (12,6% do território estadual). Em Pernambuco, apesar da alta lucratividade das atividades agropecuárias, a cana-de-açúcar ainda desempenha papel importante, representando 40% da economia estadual. Em 1997 a área plantada com cana-de-açúcar no estado estava estimada em 400 mil hectares, existindo 38 usinas e 10 destilarias de álcool instaladas nos municípios da zona açucareira do estado (CONAB, 2009).

## 2.4- O projeto SUCEST

Vários projetos com o objetivo de sequenciar Etiquetas de Sequências Expressas (ESTs; *Expressed Sequence Tags*) em cana-de-açúcar foram conduzidos na África do Sul, Austrália, França e Brasil (Carson e Botha, 2000; Casu et al., 2001; Grivet e Arruda, 2001; Perrin e Wigge, 2002). Juntos, estes projetos produziram mais de 300.000 ESTs de cana. No Brasil um consórcio da rede ONSA (*Organization for Nucleotide Sequencing and Analysis*) lançou em 1999 o projeto SUCEST (*Sugarcane Expressed Sequence Tag Project*) o qual gerou uma base de dados de cerca de 238.000 ESTs, produzidas a partir de 26 bibliotecas de diferentes órgãos e tecidos em diversos estágios de desenvolvimento. A abordagem escolhida por estes projetos se baseou na determinação da sequência do RNA mensageiro através da análise de sequências expressas do genoma. Tal abordagem permitiu a investigação indireta do conteúdo genético dos cromossomos, bem como a determinação da expressão diferencial em cada tecido e em diferentes condições de estresse (SUCEST, 2011).

# 3- A cultura da soja

## 3.1- Origem, citogenética e história

A soja (*Glycine max* L.) é uma leguminosa de grande importância econômica se destacando atualmente não só por produzir alimentos saudáveis como também por gerar energia renovável, conhecida como biodiesel (Prado, 2007; Sousa, 2010). Originária do continente asiático, mais precisamente da China Antiga, há relatos de que a planta representava a base alimentar do povo há mais de 5.000 anos; entretanto, a soja atualmente cultivada é bastante diferente das plantas que se desenvolviam principalmente ao longo do Rio Amarelo; sua evolução começou com o aparecimento de plantas oriundas de cruzamentos naturais, entre duas espécies de soja selvagem, as quais se acredita que foram domesticadas e melhoradas por cientistas da antiga China (Câmara, 1998).

O tamanho estimado do genoma da soja é de 950 Mb distribuídos em 20 pares de cromossomos (2n=40) (Cannon et al., 2009); com a característica marcante de possuir grandes blocos de genes duplicados distribuídos por toda a sua extensão (Schmutz et al., 2010).

No Brasil, a soja começou a ser cultivada comercialmente no Rio Grande do Sul em meados de 1940, expandindo-se para as outras regiões do país a partir da década de 70 e representando hoje 12% do produto interno bruto de todo o agronegócio do país (Yuyama et al., 2007). O Brasil é o quarto maior consumidor mundial de soja, sendo ainda o segundo maior produtor, com mais de 22 milhões de hectares cultivados na safra 2008/09 (AGRIANUAL, 2009). A grande variabilidade existente entre as cultivares de soja, quanto à resposta a fatores ambientais e resistência a pragas e doenças (Dong et al., 2001), proporciona uma ótima adaptabilidade nas mais variadas regiões do Brasil (Carvalho et al., 2010); entretanto, as doenças ainda representam o maior fator limitante da exploração máxima de seu potencial produtivo (da Costa, 2007).

### 3.2- As doenças da soja

Em paralelo à expansão da cultura da soja, é inevitável o surgimento de novas doenças e um aumento na intensidade das doenças pré-existentes. A cultura é infectada por diversos patógenos, ocasionando cerca de 40 diferentes tipos de doenças (EMBRAPA, 2011). A importância econômica de cada doença varia de ano para ano e de região para região, dependendo das condições climáticas de cada safra.

Os fungos são os microrganismos que mais atacam a soja, influenciando na produção final e na qualidade da semente (Nakagawa et al., 2000), podendo acarretar doenças que atacam folhas, como a ferrugem e o Míldio, a vagem, haste ou semente, como a antracnose e também as raízes, como a podridão do carvão, entre outras (EMBRAPA, 2011).

A ferrugem, causada por fungos do gênero *Phakopsora* é a principal doença que acomete a cultura da soja. Devido à forma agressiva pela qual se manifesta, vêm se tornando uma preocupação constante na tentativa de minimizar seus danos, que causam perdas de 10 a 80% da produção (da Costa, 2007).

A antracnose, causada pelo fungo *Colletotrichum dematium var. truncata*, é uma das principais doenças da soja podendo acometer a soja em qualquer área onde for produzida (Ramos et al., 2010). Sob condições de alta umidade, causa apodrecimento e queda das vagens, abertura das vagens imaturas e germinação dos grãos em formação; sua capacidade de

se manter latente por longos períodos, torna o *Colletotrichum dematium var. truncata* o mais importante patógeno pós-colheita da cultura (Bailey et al., 1992)

O causador da podridão do carvão é o fungo *Macrophomina phaseolina (Tass.) Goid.*e costuma ser considerado um patógeno secundário, uma vez que seus principais efeitos sobre a planta aparecem na maioria das vezes nos estágios finais da colheita (Machado, 1987) No Brasil, a doença está frequentemente associada a períodos de seca, nos quais perdas de até 50% puderam ser observadas no nordeste do Paraná (Almeida et al., 2003).

### 3.3- Produção e importância econômica

A soja é considerada um alimento funcional, pois fornece nutrientes ao organismo e traz benefícios para saúde. É rica em proteínas, possui isoflavonas e ácidos graxos insaturados e, segundo pesquisas na área médica, tem ação na prevenção de doenças crônico-degenerativas (Mozaffarian et al., 2006). Também é uma excelente fonte de minerais como ferro, potássio, fósforo, cálcio e vitaminas do complexo B (Messina e Wu, 2009). Além dessas propriedades a soja aparece também como matéria prima para a produção de biodiesel, o biocombustível que apresenta-se como uma relevante alternativa para os combustíveis derivados do petróleo e danosos ao meio ambiente (Fargione et al., 2008).

Nos anos 70 a soja se consolidou como a principal cultura do agronegócio brasileiro, passando de 1,5 milhões de toneladas em 1970 para mais de 15 milhões de toneladas em 1979. A soja foi a única cultura a ter um crescimento expressivo na sua área cultivada ao longo das últimas três décadas (EMBRAPA, 2011)

A revolução socioeconômica e tecnológica protagonizada pela soja no Brasil moderno pode ser comparada ao fenômeno ocorrido com a cana-de-açúcar no Brasil Colônia e com o café no Brasil Império/República, os quais, em épocas diferentes, comandaram o comércio exterior do País. A soja responde por uma receita cambial direta para o Brasil de mais de sete bilhões de dólares anuais e cinco vezes esse valor, se considerados os benefícios que gera ao longo da sua extensa cadeia produtiva (SEAPA, 2010).

### 3.4- O projeto Genosoja

O Consórcio GENOSOJA foi estabelecido no ano de 2009 com o objetivo de integrar várias instituições que tinham interesse em trabalhar com o genoma da soja no Brasil. O objetivo desse projeto foi investigar informações relativas à genômica do organismo, incluindo genômica estrutural e funcional, transcriptômica e proteômica com informações da expressão de genes e as proteínas codificadas por eles, gerando dados para controle de estresses que comprometem a cultura da planta, como secas, doenças e pragas. A maioria dos ensaios incluiu o estudo de sequências geradas sob condições de estresse biótico ou abiótico (Abdelnoor et al., 2009).

## 4- A cultura do feijão-caupi

### 4.1- Origem, citogenética e história

O feijão-caupi, *Vigna unguiculata* (L.) Walp.,é uma espécie autógama (Teófilo et al*.,* 2001), classificada como membro da divisão Embryophyta*,* incluída na subdivisãoAngiospermae*,* na classe Dicotyledoneae na família Fabaceae (NCBI, 2011). Possui um dos menores genomas deste grupo (≅450-500 Mb), apresentando o nível diploide com 2n=22 cromossomos (Benko-Iseppon, 2001).

O feijão-caupi foi introduzido no Brasil proveniente da Europa e oeste da África por colonizadores europeus e escravos africanos durante os séculos 16 e 17. As plantações cultivadas tradicionalmente foram selecionadas para determinados caracteres de interesse ao longo dos últimos três séculos (Passos et al., 2007). Em vista da maior concentração de espécies gênero na África, bem como do número elevado de espécies endêmicas, sugere-se que sua evolução e dispersão provavelmente tenham ocorrido a partir deste continente (Freire-Filho, 1988). O citado autor também destaca que entre as espécies nativas da África, *V. unguiculata*, aparece predominantemente em algumas regiões enquanto suas formas selvagens não têm sido encontradas fora deste continente.

Embora as espécies do gênero *Vigna* estejam distribuídas nas regiões tropicais e subtropicais de todo o mundo, Steele e Mehra (1980) e Ng e Maréchal (1985), citam o oeste da África, mais precisamente a Nigéria, como centro primário de diversidade da espécie. Entretanto, Padulosi et al.(1997) afirmam que provavelmente a região do Transvaal, na República da África do Sul, seja a região de especiação de *V. unguiculata*. Por outro lado, estudos moleculares (Simon et al., 2007) forneceram evidências de que, embora seu centro de diversidade atual se localize na África, o gênero *Vigna* e a maioria das espécies (incluindo *V. unguiculata*) podem ter surgido na Ásia, onde algumas subespécies primitivas ainda ocorrem.

O aumento do tamanho dos grãos, aumento da produtividade média, porte ereto das plantas, floração precoce, bem como a identificação de linhagens resistentes à salinidade e às doenças que mais prejudicam a produção do feijão-caupi são os caracteres mais desejáveis e buscados através de vários cruzamentos (Araújo, 1988; Barreto, 1999; Freire-Filho et al., 1999).

Experimentos têm demonstrado que métodos tradicionais de cruzamento consomem, em geral, inúmeros anos, para que se consiga incorporar genes de resistência em uma determinada cultivar, mas hoje, ferramentas da biotecnologia moderna podem propiciar ao feijão-caupi condições de competitividade e características que atendam às necessidades comerciais internacionais (Timko, 2002).

## 4.2 – As doenças do feijão-caupi

Apesar de sua rusticidade e elevada resistência natural a doenças, a cultura sofre pelo ataque de alguns patógenos capazes de reduzir a sua produtividade e, em alguns casos específicos, inviabilizar sua produção (Freire-Filho, 2008). Esses fatores bióticos são, em parte, responsáveis pela baixa produtividade da cultura observada na região Nordeste. Porém, o ataque por vírus apresenta-se como o fator mais limitante (Rocha et al., 2003). Especialmente o vírus do mosaico severo do caupi (CPSMV, *Cowpea Severe Mosaic Vírus*) da família Comoviridae, os mosaicos de potyvirus da família Potyviridae e o Vírus do Mosaico dourado do caupi da família Geminiviridae, entre outros fatores, têm reduzido drasticamente a produção de feijão-caupi no Brasil com perdas de até 81% (Lima et al., 2005).

O CPSMV encontra-se disseminado e apresenta difícil controle devido à grande variedade de hospedeiros e à numerosa população de vetores (Fernandes et al., 2010). No que diz respeito ao vírus do mosaico dourado do caupi (CPGMV; *Cowpea Gold Mosaic Vírus*) a situação também é preocupante, pois estudos de avaliação do impacto destas doenças em algumas regiões do nordeste do Brasil contabilizaram perdas em torno de 70% da produção (Rodrigues, 2011).

Além das viroses, outras doenças têm sido consideradas importantes, e podem também gerar perdas significativas dependendo da cultivar utilizada (Barreto, 1999).

A mela é causada pelo fungo *Rhizoctonia Solani* e têm como principais sintomas o aparecimento de manchas de aspecto melado nas folhas, mais claras no centro e com a borda escura podendo-se visualizar uma teia branca em cima das folhas à medida em que a doença aumenta (Nechet e Halfeld-Vieira, 2007).

Infelizmente, há poucas fontes de resistência para várias das doenças existentes, estando essas frequentemente em diferentes cultivares, alguns pouco produtivos ou que não atendem às necessidades do mercado (Freire-Filho, 2008).

### 4.3- Produção e importância econômica

O feijão-caupi é uma cultura de importante destaque na economia nordestina e de amplo significado social, constituindo o principal alimento proteico e energético das comunidades rurais em regiões como o Norte e Nordeste. Conhecido também como feijão-de-corda é uma excelente fonte de proteínas (23-25%) e apresenta todos os aminoácidos essenciais, carboidratos (62%), vitaminas e minerais, além de possuir grande quantidade de fibras dietéticas e baixa quantidade de gordura (teor de óleo de 2%, em média) (EMBRAPA, 2011). Apresenta alta rusticidade e adaptabilidade às condições de estiagem prolongadas com capacidade de se desenvolver em solo de baixa fertilidade e por meio da simbiose com bactérias do gênero *Rhizobium*, tem a habilidade para fixar nitrogênio do ar (Oliveira et al., 2003). A capacidade do feijão-caupi de crescer sob condições de solo e ambientes estressantes tem sido reconhecida por muitos cientistas. No Brasil, trata-se do único feijão capaz de sobreviver com sucesso na região norte (alta umidade, muita chuva e solo argiloso) e no

Nordeste (seca, solo arenoso, por vezes salino e muito sol) (Barreto, 1999; Rocha et al.,2007). O mercado do feijão-caupi ainda se restringe a grãos secos, grãos verdes (hidratados) e sementes, havendo já algumas iniciativas para o processamento industrial de feijão-caupi para produção de farinha e produtos pré-cozidos e congelados. O mercado do feijão-caupi ainda tem contornos regionais, concentrando-se, principalmente, nas regiões Nordeste e Norte. Entretanto, há indícios de certa expansão da cultura na região Sudeste, principalmente no norte de Minas Gerais e Rio de Janeiro, predominando nesses estados o grão da Subclasse Fradinho (EMBRAPA, 2011).

### 4.4- O transcriptoma do feijão-caupi e a rede NordEST

A rede NordEST foi estabelecida em 2005, agrupando 12 laboratórios com objetivo de gerar dados genômicos úteis para o melhoramento do feijão-caupi e programas de biotecnologia. 30.000 ESTs foram geradas a partir de 12 bibliotecas distintas sob condições de estresse abiótico (salinidade) e biótico (mosaico severo e potyvirus). Em paralelo, oito bibliotecas de SuperSAGE, submetidas à seca e mosaico severo, foram construídas permitindo a geração e análise de mais de vinte milhões de transcritos (SuperTags) sob diferentes condições. O mapeamento físico dos cromossomos está sendo desenvolvido, 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., 2009; 2010).

## 5- A bioinformática aplicada às ômicas

### 5.1- História e aplicações

O termo “ômicas” se aplica às áreas das ciências biológicas e das engenharias que se dedicam a analisar as interações de dados biológicos derivados de análises de genômica, transcriptômica, proteômica, interactômica, metabolômica, epigenômica comparativamente a dados do fenoma, entre outras. O foco principal visa identificar e associar os dados das

diversas abordagens citadas (tais como genes, proteínas e ligantes) através de ferramentas de bioinformática e biologia de sistemas, encontrando relações de interação entre os mesmos, em geral associando dados depositados em bancos de dados a outros gerados em experimentos laboratoriais e de campo, proporcionando um melhor entendimento das redes e interações entre diferentes fontes de dados (Benko-Iseppon et al., 2012).

A bioinformática vem sendo encarada como uma disciplina especial há pelo menos 15 anos e os primeiros esforços na sua consolidação podem ser localizados no início dos anos 80, a partir do momento em que se iniciou a utilização de ferramentas computacionais para análise de dados bioquímicos, de biologia molecular e biológicos como um todo. Assim surgia a bioinformática, uma nova ciência com raízes nas ciências da computação, na estatística e na biologia, cuja finalidade principal é gerar novos conhecimentos a partir do eficiente acesso e manuseio de grande volume de dados (Carraro e Kitajima, 2002).

A necessidade do desenvolvimento de ferramentas computacionais aplicadas às sequências de DNA e proteínas começou a aparecer com o acúmulo de informações de interesse público ou comum a diversos grupos de pesquisa, que surgiram como resultado do Projeto Genoma Humano, passando a desempenhar um papel essencial em outros projetos genoma. Desde a década de 1990, portanto, os esforços internacionais no sentido de obter sequências genômicas completas levaram à determinação dos genomas de mais de 700 organismos, entre estes, procariotos, leveduras, protozoários, plantas, invertebrados e vertebrados; de acordo com o Banco de dados de genomas *on line* (GOLD; *Genomes on line database*) em 2012, aproximadamente 7.000 outros projetos genoma estão em andamento, representando interesses médicos, comerciais, ambientais e industriais, ou contemplando organismos modelos importantes para o desenvolvimento da pesquisa científica (GOLD, 2012). A consolidação de ciências como a Bioinformática e a Biologia Computacional nas últimas décadas, tem permitido à comunidade científica o uso de abordagens holísticas e ao mesmo tempo inovadoras no estudo da estrutura, organização e evolução de genomas (Binneck, 2004; Abby e Daubin, 2007), no estudo da expressão diferencial de genes e proteínas (Patterson e Aebersold, 2003), na análise da estrutura tridimensional de proteínas (Ginalski, 2006), no processo de reconstrução metabólica e na predição e classificação funcional de genes (Lee et al., 2007; Skrabanek et al., 2008).

Em vista de sua importância, várias universidades, instituições governamentais e empresas farmacêuticas vêm formando grupos de bioinformática. Estes grupos representam importante papel na análise das informações geradas em grande escala pelos sequenciadores de todo o mundo (Prosdocini et al., 2002).

A análise comparativa de genomas possui variadas aplicações em diferentes campos do conhecimento e constitui um campo fértil para pesquisas envolvendo os mecanismos moleculares da patogênese, do espectro de hospedeiros e das particularidades fenotípicas entre agentes patogênicos. Neste sentido, diferentes abordagens têm sido desenvolvidas e empregadas na comparação de sequências genômicas, oferecendo assim múltiplas perspectivas acerca dos organismos estudados (Shendure et al. 2008), Assim, a análise comparativa de genomas de diferentes patógenos trouxe informações importantes para explicar seus mecanismos de penetração e colonização nas plantas e também a identificação de novos agentes envolvidos nestes processos (Abby e Daubin 2007).

Um outro campo em ascensão é a chamada bioinformática traducional que tem como foco a interseção entre bioinformática e doença; tem como objetivo não só a melhor compreensão dos mecanismos de cada doença, mas também a criação de novas ferramentas computacionais para estudá-las (Altman e Miller, 2011). Em 2010 Eriksson et al. utilizaram respostas de questionários recolhidos pela internet e associaram a um banco de dados de Polimorfismo de único nucleotídeo (SNPs; *Single Nucleotide Polymorphism*) para executar múltiplos trabalhos de Estudo de associação ampla ao genoma (GWAS; *Genome-wide* association study), onde a análise de diversas variações genéticas comuns em indivíduos diferentes é relacionada a uma característica. Da mesma forma, Denny et al. (2010) extrapolaram a pesquisa para o que chamamos de Estudo de associação ampla ao fenoma (PheWAS; *Phenome-wide association study*), percebendo que SNPs associadas a uma doença X podem também estar ligadas a uma outra Y; o trabalho fornece uma perspectiva de como essas doenças se relacionam em nível molecular. Numa visão ainda mais ampla, Patel et al. (2010) usaram, para o que foi denominado de Estudo de associação ampla ao meio ambiente (EWAS; *Environment-wide association study*), um conjunto de 226 dados longitudinais de fatores ambientais e estilo de vida que puderam ser relacionados a incidência de diabetes tipo II.

Pensando na dificuldade de reprodução dos experimentos na área da bioinformática, Dudley e Butte descreveram em 2010 o esboço de um sistema de troca instantânea de informações

(WSSE; *Whole system snapshot exchange*) que armazenaria automaticamente todos os dados de determinado trabalho em uma nuvem virtual, que quando acessada recriaria o exato ambiente computacional usado pelos autores permitindo um nível alto de reprodutibilidade, minimizando a redundância na pesquisa e consequentemente acelerando o progresso da ciência.

De acordo com Boulesteix (2010), os bioinformatas são muito otimistas e tendem a buscar sempre dados com significância estatística e publicar apenas resultados positivos. Ela sugere a publicação de resultados negativos, dados mais transparentes e o desenvolvimento de métodos como o WSSE proposto por Dudley e Butte em 2010, além de afirmar que algumas vezes as pessoas são melhores que as máquinas e que na hora de resolver grandes problemas, a *expertise* humana deve ser privilegiada.

## 5.2- Bancos de dados e ferramentas

O Genbank é o mais conhecido banco de dados que disponibiliza gratuitamente sequências e ferramentas para a sua análise. Foi criado inicialmente pelo Centro Nacional para Informação Biotecnológica (NCBI; *National Center for Biotechnology Information*- http://www.ncbi.nlm.nih.gov/);com o objetivo de abrigar sequências submetidas diretamente por autores de publicações científicas. Em 2011 já tinha disponível sequências de mais de 380.000 organismos obtidas inicialmente por submissões dos laboratórios individualmente, englobando projetos de sequenciamento em larga escala, tanto de genomas completos quanto de ESTs (Benson et al., 2011); o dbEST englobava em janeiro de 2011 71.276.166 ESTs, sendo uma das divisões mais ricas em sequências do GenBank (NCBI, 2012). Em paralelo, fazem parte do Banco de Dados Internacional de Sequências de Nucleotídeos (INSD- *International Nucleotide Sequence Database*) o Banco de Dados de DNA do Japão (DDBJ; *DNA Database of Japan* - http://www.ddbj.nig.ac.jp/) e o Laboratório Europeu de Biologia Molecular (EMBL; *European Molecular Biology Laboratory* - http://www.ebi.ac.uk/embl/) que também disponibilizam muitas informações de autores europeus e japoneses, além de manter um importante acordo de cooperação e de intercâmbio de dados (Benson et al., 2011). Embora existam muitos bancos de dados com acesso restrito, atualmente a tendência é que estes dados sejam disponibilizados e estejam acessíveis a todos os pesquisadores, como ocorre no GenBank (Benson et al., 2011).

Em 2009 foi criado o Banco de Dados de Genes de Resistência em Plantas (PRGdb; *Plant Resistance Genes Database* - http://www.prgdb.org); o primeiro banco de genes *R* identificados por ferramentas de Bioinformática. O PRGBdb engloba mais de 15.000 sequências entre genes *R* identificados e preditos pertencentes a 192 espécies diferentes de plantas (Sanseverino et al*.,* 2010*).* Em conjunto com os bancos de dados, várias ferramentas e programas são utilizados no processo de identificação e análise de sequências; entre eles estão a Ferramenta de Busca por Alinhamento Local (BLAST; *Basic Local Alignment Search Tool*) (Altschul et al., 1990), o CLUSTALX e W (Thompson et al., 1997; Larkin et al., 2007), o TreeView (Page, 1996), o Programa para Análise Genética e Evolutiva (MEGA; *Molecular Evolutionary Genetics Analysis)* (Tamura et al*.,* 2011) e o *CLUSTER* (Eisen et al., 1998).

Alinhamentos representam uma poderosa ferramenta não só para comparar sequências desconhecidas com sequências de genes já descritos e depositadas nos bancos de dados, como também para fazer inferências tanto estruturais e funcionais, quanto evolutivas a respeito destas sequências. O BLAST oferece um método de busca rápida nos bancos de dados não só de nucleotídeos, como também de proteínas (Altschul et al., 1990). O *ORF-finder* também é uma ferramenta on-line do NCBI que permite traduzir uma sequência de nucleotídeos (DNA/RNA) para uma sequência proteica em todos os quadros de leitura.

O programa CLUSTAL (X e W) está entre os mais usados para obtenção de múltiplos alinhamentos de sequências, quer sejam de nucleotídeos, ou de aminoácidos. Produz alinhamentos múltiplos biologicamente significativos entre sequências divergentes; calculando qual a melhor sequência para comparar com a selecionada, alinhando-as de acordo com as identidades, similaridades e diferenças (Thompson et al., 1997; Larkin et al., 2007); é capaz também de gerar dendrogramas que podem ser visualizados no programa TreeView (Page, 1996), visto que apesar de possuir o algoritmo para gerá-los, o CLUSTAL não possui uma interface para sua visualização. O programa roda em ambientes operacionais DOS, MAC e UNIX. No entanto, há uma interface do CLUSTALW para Windows, denominada CLUSTALX, bastante fácil de ser manuseada. Apesar de ser muito versátil, o CLUSTAL tem algumas restrições de uso. Ele só deve ser usado para alinhar genes que tenham um mesmo domínio, devendo este aparecer na mesma ordem em todas as sequências. Sequências sem uma ancestralidade comum ou um N ou C terminal muito grande e divergente também devem ser evitados. .

O MEGA foi desenvolvido não só para estimar distâncias genéticas como também para fazer inferências filogenéticas através da construção de dendrogramas, calcular matrizes de distância e fazer análises na composição das sequências, a porcentagem de similaridade entre sequências ou ainda identificação de códons ou marcadores conservados associando-os à características morfológicas e fisiológicas ou ainda variantes de uma característica (Sudhir et al., 1993). A versão 5.0, além de apresentar uma interface mais amigável para o usuário, permite que os arquivos de entrada *.aln* gerados anteriormente apenas pelo CLUSTAL sejam rodados no próprio programa, permitindo também a edição de sequências manualmente (Tamura et al*.,* 2011)

O programa CLUSTER fornece o ambiente computacional e gráfico para análise de sequências genômicas e de grandes quantidades de dados provenientes de complexos experimentos de *microarrays* (avaliação híbrida de microarranjos), Análise Serial de Expressão Gênica (SAGE; Serial Analysis of Gene Expression), EST, entre outros. O programa inclui, por exemplo, uma ferramenta de *clusterização* hierárquica e mapas de auto-organização (SOMs; *Self-Organizing Maps*). Os métodos de *clusterização* hierárquica permitem a análise de expressão simulando um experimento de *microarray*, baseado numa matriz de dados fornecida pelo usuário e estão descritos em Eisen et al. (1998), podendo ser utilizados também na condução de análises *in silico* do perfil de expressão de genes (Eisen et al., 1998).Para cada gene, o perfil de expressão reflete seu comportamento dinâmico sob condições de estresse, estágios de desenvolvimento ou tecidos distintos; genes com perfil de expressão similar são considerados co-expressos.

O pipeline formado por todas essas abordagens contribui para a construção do painel de genes que participam da resposta da planta aos patógenos. O melhoramento genético das plantas é só possível devido à identificação e caracterização *in silico* desses genes que permitem o delineamento dos experimentos *in vitro* e *in vivo* (Santos e Ortega, 2003)

# REFERÊNCIAS BIBLIOGRÁFICAS


Abby S and Daubin V (2007) Comparative genomics and the evolution of prokaryotes. Trends Microbiol 15:135-141.

Abdelnoor RV et al. (2009) GENOSOJA – A Brazilian Soybean Genome Consortium. In: Plant and Animal Genome Conference, XVII, 2009, San Diego - USA. Anais…, abstract P034. San Diego, USA; Scherago International. Disponível em: <http://www.intl-pag.org/17/abstracts/P01_PAGXVII_034.html>.

Almeida AMR, Amorim L, Filho AB, Torres E, Farias JRB, Benato LC, Pinto MC, Valentim M (2003) Progress of soybean charcoal rot under tillage and no-tillage systems in Brazil. Fitopatol Bras 28:131-135.

Almeida IMG (2008) Doenças causadas por bactérias. In: Dinardo-Miranda LL, Vasconcelos ACM, Landell MGA (Eds.) Cana-de-açúcar. Campinas SP. Instituto Agronômico. pp. 437-450.

Altman RB and Miller KS (2011) 2010 Translational bioinformatics year in review. J Am Med Inform Assoc 18:358-366.

Altschul SF,Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool (1990) J Mol Biol 215:403-410.

Araújo JPP (1988) Cowpea breeding in Brazil. In: Watt EE and Araújo JPP (eds) Cowpea research in Brazil. IITA EMBRAPA, Brasília, pp 93-116.

Artschwager E and Brandes EW (1958) Sugarcane (*Saccharum officinarum* L.). Washington: USDA, 307 pp. (Agriculture Handbook, 122).

Bailey JA,O'Connell RJ, Pring RJ, Nash C. (1992) Infection strategies of Colletotrichum

species. In: Bailey JA, Jeger MJ (Eds) Colletotrichum: biology, pathology and control. Wallingford, UK. CAB International. pp.88-120.

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.

Barbosa da Silva A, Wanderley-Nogueira AC, Silva RRM, Berlarmino LC, Soares-Cavalcanti NM and Benko-Iseppon AM (2005) In silico survey of resistance (R) genes in Eucalyptus transcriptome Genet Mol Biol 28(3):562-574.

Barreto PD (1999) Recursos genéticos e programa de melhoramento de feijão-de-corda no Ceará: Avanços e perspectivas. In: Queirós MA de Goedert CO and Ramos SRR (eds) Recursos genéticos e melhoramento de plantas para o Nordeste Brasileiro. EMBRAPA, CPATSA, Petrolina.

Bella J, Hindle KL, McEwan PA, Lovell SC. (2008) The leucine-rich repeat structure. Cell Mol Life Sci 65:2307–2333.

Benko-Iseppon AM (2001) Estudos moleculares e citogenéticos no caupi e em espécies relacionadas: Avanços e perspectivas. Embrapa Documentos 56:327-332.

Benko-Iseppon AM, Kido EA , Monte SJH, Pandolfi V , Calsa-Jr.T , Houllou-Kido LM , Rocha MM , Barbosa PKA , Winter P, Kahl G, Bjorn R , Horres R, Amorim LLB , Onofre AVC , Ferreira-Neto JC, Vasconcelos ATR, Carvalho R , Bellarmino LC , Soares-Cavalcanti NM , Wanderley-Nogueira AC , Barbosa-Silva A , Barros PS , Vieira-Mello GS , Brasileiro-Vidal AC , Bortoleti KCA , Pedrosa-Harand A , Gepts P, Franklin-de-Mello N , Andrade PP , GranjeiroTB , Cunha OL, Meirelles DB , Andrade GP, Pio-Ribeiro G , Sittolin IM , Freire-Filho FR Cowpea (Vigna unguiculata) Transcriptome and Differential Expression: The Project and Applications for Breeding Purposes. In: Plant and Animal Genome Conference, XVII, 2009, San Diego - USA. Anais…, abstract W304. San Diego, USA; Scherago International. Disponível em: <http://www.intl-pag.org/17/abstracts/W40_PAGXVII_304.html>.

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. Outubro/2010, Fortaleza, CE. Pp. 97-98.

Benko-Iseppon AM, Galdino SL, Calsa Júnior T, Kido EA, Tossi A, Belarmino LC, Crovella S (2010) Overview of Plant Antimicrobial Peptides. Curr Prot Pep Sci 11(3): 181-188.

Benko-Iseppon AM, Soares-Cavalcanti NM, Berlarmino LC, Bezerra-Neto JP, Amorim LLB, Ferreira Neto JRC, Pandolfi V, Azevedo HMA, Silva RLO, Santos MG, Alves MVS, Kido EA (2012) 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. Aceito em setembro/2011. *In Press*.

Benson DA et al. (2011) GenBank. Nucl Acids Res 39:D32-D37.

Berding N and Roach BT (1987) Germplasm collection, maintenance and use. In: Sugarcane improvement through breeding. (Ed) DJ Heinz. Elsevier Press, Amsterdam. pp 143-210.

Bergelson J, Kreitman M, Stahl EA and Tian D (2001) Evolutionary dynamics of plant R-genes. Science 292: 2281–2285.

Binneck E (2004) As ômicas: integrando a bioinformação - O papel da bioinformática em expansão. Biotecnol Ciênc 32:28-37.

Bolton MD (2009) Primary metabolism and plant defense-fuel for the fire. Mol Plant Microbe Interact 22:487-497.

Bonas U and Lahaye T (2002) Plant disease resistance triggered by pathogen-derived molecules: Refined models of specific recognition. Curr Opin Microbiol 5:44-50.

Bonasera JM, Kim JF and Beer SV (2006) *PR* genes of apple: identification and expression in response to elicitors and inoculation with Erwinia amylovora. BMC Plant Biol 6:23-34.

Botella MA (1998) Three genes of the Arabidopsis RPP1 complex resistance locus recognize distinct Peronospora parasitica avirulence determinants. Plant Cell 10:1847-1860.

Boulesteix AL (2010) Over-optimism in bioinformatics research. Bioinformatics 26:437-439.

Brandes EW (1956) Origin dispersal and use in breeding of the Melanesian Garden sugarcanes and their derivatives, Saccharum officinarum L. Cong Proc Int Soc Sug Cane Technol 9:709-750.

Brencic A, Winans SC (2005) Detection and response to signals involved in host-microbe interactions by plant-associated bacteria. Microbiol Mol Biol Rev 69:155-194.

Brett PGC (1957) Flowering and pollen fertility in relation to sugarcane breeding in Natal. Cong Proc Int Soc Sug Cane Technol 7:43–56

Caicedo AL, Schaal BA and Kunkel BN (1999) Diversity and molecular evolution of the RPS2 resistance gene in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 96:302–306.

Câmara GMS (1988) Soja: tecnologia da produção. Piracicaba: Publique p.1-14.

Cândido e Silva EK (2008) Cowpea Severe Mosaic Virus: Diagnóstico, Estudo de Herança e Identificação de marcadores moleculares associados à resistência. Dissertação de mestrado, Programa de Pós graduação em Fitopatologia. Universidade Federal Rural de Pernambuco.

Cannon SB, May GD and Jackson SA (2009) Three sequenced legume genomes and many crop species: rich opportunities for translational genomics. Plant Physiol 151:970-977.

Carraro DM and Kitajima JP (2002) Sequenciamento e Bioinformática de Genomas Bacterianos. Biotecnol Ciencia & Desenvolvimento 28:16-20.

Carson DL and Botha FC (2000) Preliminary analysis of expressed sequence tags for sugarcane. Crop Sci 40:1769-1779.

Carvalho ER, Rezende PM, Ogoshi FGA, Botrel EP, Alcantara HP, Santos JP (2010) Desempenho de cultivares de soja [*Glycine max* (L.) Merrill] em cultivo de verão no Sul de Minas Gerais. Ciência e Agrotecnologia, Lavras, 34(4):892-899.

Casu R, Dimmock C, Thomas M, Bower N, Knight D, Grof C, Mcintyre L, Jackson P, Jordan D, Whan V, Drenth J, Tao Y and Manners J (2001) Genetic and expression profiling in sugarcane. Proc Int Soc Sugarcane Technol 24:626-627.

Cheavegatti-Gianotto A, Abreu H, Arruda P et al. (2011) Sugarcane: A Reference Study for the Regulation of Genetically Modified Cultivars in Brazil. Trop Plant Biol 4(1):62-89.

Chisholm ST, Coaker G, Day B, Staskawicz BJ (2006) Host-microbe interactions: Shaping the evolution of the plant immune response. Cell 124:803−814.

Comstock JC and Lentini RS (2002) Sugarcane smut disease. Disponível em:<http://edis.ufl.edu/body_SC008>. Acesso em: Mar 2011

Costa AS, Kitajima EW and Arruda SC (1971) Moléstias de vírus e de micoplasmas do milho em São Paulo. Rev Soc Bras Fitopatol 4:39-41.

Cutt JR and Klessig DF (1992) Pathogenesis-related proteins. In: Genes lnvolved in Plant Defense, T. Bollerand F. Meins, eds (New York Springer-Verlag), pp. 209-243.

Dantas B and Melo JL (1960) A situação das variedades na zona canavieira de Pernambuco (1954/55 a 1957/58) e uma nota histórica sobre as variedades antigas. Boletim Técnico Instituto Agronômico do Nordeste, Recife, IAN/MA, 11:29-82.

de Wit PJ (2007) How plants recognize pathogens and defend themselves. Cell Mol Life Sci 64:2726-2732.

Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D,

Masys DR, Roden DM, Crawford DC (2010). PheWAS: demonstrating the feasibility of a phenome-side scan to discover gene-disease associations. Bioinformatics 26:1205-1210.

Dillon LS, Shapter FM, Henry RJ, Cordeiro G, Izquierdo L, Lee LS. (2007) Domestication to Crop Improvement: Genetic Resources for Sorghum and Saccharum (Andropogoneae). Ann Bot 100:975-989.

Dinardo-Miranda LL and Pragas. In: Dinardo-miranda LL, Vasconcelos ACM, Landell MGA (2008) Cana-de-açúcar. Campinas: Instituto Agronômico, 882 pp.

Dodds PN, Lawrence GJ and Ellis JG (2001) Six amino acid changes confined to the leucine-rich repeat betastrand/beta-turn motif determine the difference between the P and P2 rust resistance specificities in flax. Plant Cell 13:163-178.

Dong YS, Zhao LM, Liu B, Wang ZW, Jin ZQ and Sun H (2001) The genetic diversity of cultivated soybean grown in China. Theor Appl Gen 108:931-936.

Dudley JT and Butte AJ (2010) In silico research in the era of cloud computing. Nature Biotech 28:1181-1185.

Eisen MB et al. (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.

Ellis J and Jones D (2000) Structure and function of proteins controlling strain-specific pathogen resistance in plants. Curr Opin Plant Biol 1:288-293.

Ellis J, Dodds P and Pryor T (2000) Structure, function and evolution of plant disease resistance genes. Curr Opin Plant Biol 3:278-284.

Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, Avey L, Wojcicki A, Pe'er I, Mountain J (2010) Web-based, participant-driven studies yield novel genetic

associations for common traits. PLoS Genet 6:e1000993.

Fargione J, Hill J, Tilman D, Polasky S, Hawthorne P (2008) Land Clearing and the Biofuel Carbon Debt. Science 319:1235-1238.

Fermin-Muñoz GA, Meng B, Ko K, Mazumdar-Leighton S, Gubba A, Carroll JE (2000) Biotechnology: a new era for plant pathology and plant protection. APSnet Feature Story May 2000. Cited march 2011, from <http://www.apsnet.org/online/feature/BioTechnology/Top.html>.

Fernandes CF , Vieira JR, Marcolan AL, Rosa Neto C, Diocleciano JM, Ferro GO, Junior HA, Reis ND, da Silva DSG (2010) Ocorrência do virus do mosaico severo (Cowpea severe mosaic virus - CPSMV) no feijão-caupi (Vigna unguiculata (L.) Walp. Rondônia. Comunicado Técnico, 353. 1ª edição

Flor HH (1942) Current status of the gene-for-gene concept. Annu Rev Plant Pathol 9:275-296.

Freire-Filho FR (1988) Cowpea taxonomy and introduction to Brazil. In: Watt EE and Araújo JPP de (eds) Cowpea research in Brazil. IITA EMBRAPA, Brasília, pp 3-10.

Freire-Filho FR, Ribeiro VQ, Barreto PD and Santos CAF (1999) Melhoramento genético do caupi (*Vigna unguiculata* (L.) Walp.) na região nordeste. In: Queirós MA de, Goedert CO and Ramos SRR (eds) Recursos genéticos e melhoramento de plantas para o Nordeste Brasileiro. EMBRAPA, CPATSA, Petrolina.

Freire-Filho FR, Rocha M, Ribeiro VQ, Sitollin IM (2008) Avanços e perspectivas para a cultura do feijão-caupi. In: Albuquerque, A. C. S.; Silva, A. G. da (Eds.). Agricultura Tropical: quatro décadas de inovações tecnológicas, institucionais e políticas. Brasília: Embrapa Informação Tecnológica 1:285-250.

Friedman AR and Baker BJ (2007) The evolution of resistance genes in multiprotein plant resistance systems. Curr Opin Genet Dev 17:493-499.

Gaffney T, Friedrich L, Vernooij B, Negrotto D, Nye G, Ukness S, Ward E, Kessman H and Ryals J (1993) Requirement of salicylic acid for the induction of systemic acquired resistance. Science 261:754-756.

Gassmann W, Hinsch ME and Staskawicz BJ (1999) The Arabidopsis RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. Plant J 20:265–277.

Ginalski K (2006) Comparative modeling for protein structure prediction. Curr Opin Struct Biol 16(2):172-177.

Goff AS Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J (2002) A draft sequence of rice genome (*Oryza sativa* L. ssp. japônica) Science 296:92-100.

Gonçalves MC, Moreira YJCB, Maia IG, Santos AS, Fantin GM, Chaves ALR, Frangione DSS and Ulian EC (2004) Identificação e caracterização de isolados pertencentes ao subgrupo do Sugarcane mosaic virus no estado de São Paulo. Fitopatologia Brasileira 29:129.

Griffith M and Yaish MWF (2004) Antifreeze proteins in overwintering plants: a tale of two activities. Trends Plant Sci 9:399-405.

Grivet L and Arruda P (2001) Sugarcane Genomics: depicting the complex genome of an important tropical crop. Curr Opin Plant Biol 5:122-127.

Hammond-Kosack KE and Jones JDG (1997) Plant disease resistance genes. Annu Rev Plant physiol Plant Mol Biol 48:575-607.

Heath MC (1991) The role of gene-for-gene interactions in determination of host-specie specificity. Phytopathol 81:127-130.

Heinz DJ, Mee WP and Nickell LG (1969) Chromosome numbers of some Saccharum species hybrids and their cell suspension cultures. Amer J Bot 56:450-456.

Hogarth DM (1987) Genetics of Sugarcane. In: Sugarcane Improvement Through Breeding, edited by D.J. Heinz. New York: Elsevier, pp. 255-271.

Hulbert SH, Webb CA, Smith SM and Sun Q (2001) Resistance gene complexes: Evolution and utilization. Annu Rev Phytopathol 39:285-312.

Ingelbrecht IL, Irvine JE, Mirkov TE (1999) Posttranscriptional gene silencing in transgenic sugarcane. Dissection of homology-dependent virus resistance in a monocot that has a complex polyploid genome. Plant Physiol 119:1187-1198.

Jones DG (2001) Putting the knowledge of plant disease resistance genes to work. Curr Opin Plant Biol 4: 281-287.

Kobe and Kajava (2001) The leucine-rich repeat as a protein recognition motif. Curr Opin in Struct Biol 11:725.

Kruijt M Kip DJ, Joosten MH, Brandwagt BF, de Wit PJ (2005) The Cf-4 and Cf-9 resistance genes against Cladosporium fulvum are conserved in wild tomato species. Mol Plant Microbe Interact 18:1011-1021.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947-2948.

Lawrence GJ, Finnegan EJ, Ayliffe MA and Ellis JG (1995) The L6 gene for flax rust resistance is related to the Arabidopsis bacterial resistance gene RPS2 and the tobacco viral resistance gene N. Plant Cell 7:1195-1206.

Lee D, Redfern O and Orengo C (2007) Predicting protein function from sequence and structure. Nat Rev Mol Cell Biol 8:995-1005.

Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, He J, Xu D, May G and Stacey G (2010) An integrated transcriptome atlas of the crop model *Glycine max* and

its use in comparative analyses in plants. Plant J 63:86-99.

Lima JAA, Nascimento AKQ, Silva GS, Camarço RFEA, Gonçalves MFB (2005) Crotalaria paulinea, Novo Hospedeiro Natural do Vírus do Mosaico Severo do Caupi. Fitopatol bras 30:429-433.

Liu B, Zhang S, Zhu X, Yang Q, Wu S, Mei M, Mauleon R, Leach J, Mew T and Leung H (2004) Candidate defense genes as predictors of quantitative blast resistance in rice. Mol Plant Microbe Int 17:1146-1152.

Lurderer R Rivas S, Nürnberger T, Mattei B, Van den Hooven HW, Van der Hoorn RA, Romeis T, Wehrfritz JM, Blume B, Nennstiel D, Zuidema D, Vervoort J, De Lorenzo G, Jones JD, De Wit PJ, Joosten MH (2001) No evidence for binding between resistance gene product Cf-9 of tomato and avirulence gene product Avr9 of *Cladosporium fulvum*. Mol Plant Microbe Int 14:867-876.

Machado, C.C. (1987) Macrophomina phaseolina: biological behavior of isolates, spatial pattern of microesclerotia in the soil, and incidence on soybeans. (PhD thesis) University of Illinois, Urbana Champaign.

Matsuoka S, Garcia AAF and Arizono H (1999) Melhoramento da cana-de-açúcar In: Aluízio Borém (org) Melhoramento de Espécies Cultivadas. 1Ed. Editora UFV, Viçosa, pp 205-251.

Mei C, Qi M, Sheng G, Yang Y (2006) Inducible Overexpression of a Rice Allene Oxide Synthase Gene Increases the Endogenous Jasmonic Acid Level, PR Gene Expression, and Host Resistance to Fungal Infection. Mol Plant-Microbe Interact 19:1127-1137

Melotto M, Coelho MF, Pedrosa-Harand A, Kelly JD and Camargo LE (2004) The anthracnose resistance locus Co-4 of common bean is located on chromosome 3 and contains putative disease resistance-related genes. Theor Appl Genet 109: 690-699.

Messina M, AH Wu (2009) Perspectives on the soy-breast cancer relation. Am J Clin Nutr 89: 1673-1679.

Mestre P and D Baulcombe (2006) Elicitor-mediated oligomerization of the Tobacco *N* Disease Resistance Protein. Plant Cell 18: 491-501.

Metzler MC, Cutt JR and Klessig DF (1991) Isolation and characterization of a gene encoding a PR-1 like protein from Arabidopsis thaliana. Plant Physiol 96:346-348.

Meyers BC, Kaushik S and Nandety RS (2005). Evolving disease resistance genes. Curr Opin Plant Biol 8:129-134.

Michelmore RW Marsh E, Seely S, Landry B. (1987) Transformation of lettuce (*Lactuca sativa)* mediated by Agrobacterium tumefaciens. Plant Cell Rept 6:439-442.

Miya A, Albert P, Shinya T, Desaki Y, Ichimura K, Shirasu K, Narusaka Y, Kawakami N, Kaku H and Shibuya N (2007) Cerk1, a lysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. Proc Natl Acad Sci USA 104:19613–19618.

Moffet P (2009) Mechanisms of recognition in dominant *R* gene mediated resistance. Adv Virus Res 75:1-33

Morais DAL (2003) Análise bioinformática de genes de resistência à patógenos no genoma expresso da cana-de-açúcar. Dissertação (Mestrado em genética) Universidade Federal de Pernambuco, Pernambuco, Recife.

Mozaffarian D, Katan MB, Ascherio A, Stampfer MJ, WC Willett (2006)Trans fatty acids and cardiovascular disease. New Eng J Med 354: 1601-1613.

Nakagawa T, Zhu H, Morishima N, Li E, Xu J, Yankner BA and Yuan J (2000) Caspase-12 mediates endoplasmic reticulum-specific apoptosis and cytotoxicity by amyloid-beta. Nature 403:98-103.

Narusaka Y, Narusaka M, Abe H, Hosaka N, Kobayashi M, Shiraishi T, Iwabuchi M (2009)

High-throughput screening for plant defense activators using a b -glucuronidase-reporter gene assay in *Arabidopsis thaliana*. Plant Biotechnol 26: 345–349.

Nechet KL Halfeld-Vieira BA (2007) Reação de cultivares de feijão-caupi à mela (*Rhizoctonia solani*) em Roraima. Fitopatol Bras 32:424-428

Ng NQ and Maréchal R (1985) Cowpea taxonomy, origin germ plasm. In: Sinch SR and Rachie KO, eds. Cowpea research, production end utilization. Cheichecter, Johm Wiley pp.11-21.

Nurnberg T and Brunner F (2002) Innate Immunity in plants and animals: emerging parallels between the recognition of general elicitors and pathogen-associated molecular patterns. Cur Opin Plant Biol 5:318-324.

Oerke EC, Dehne HW, Schonbeck F and Weber A (1994) Crop Production and Crop Protection – Estimated Losses in Major Food and CashCrops. Elsevier Science, Amsterdam, 808 pp.

Oh CS and Martin GB (2011) Effector-triggered immunity mediated by the Pto kinase. Trends Plant Sci 16:132-40.

Oliveira AP, Silva VRF, Arruda FP, Nascimento IS, Alves, AU (2003) Rendimento de feijão-caupi em função de doses e formas de aplicação de nitrogênio. Horticultura Brasileira 21(1): 77-80.

Padulosi S, Ng NQ (1997) Origin taxonomy, and morphology of Vigna unguiculata (L.) Walp. In: Singh BB, Mohan R, Dashiell KE, Jackai LEN, eds. Advances in Cowpea Research. Tsukuba; IITA JIRCAS, pp.1-12.

Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. Comp Appl Biosci 12:357-358.

Parker JE, Coleman MJ, Szabò V, Frost LN, Schmidt R, Van der Biezen EA, Moores T, Dean

C, Daniels MJ and Jones JDG (1997) The *Arabidopsis* downy mildew resistance gene RPP5 shares similarity to the Toll and interleukin-1 receptors with N and L6. Plant Cell 9:879–894.

Passos AR, Silva AS, Cruz PJ, Rocha MM, Cruz EMO, Da Rocha MAC, Bahia HF and Saldanha RB (2007) Divergência genética em feijão-caupi. Bragantia 66(4):579-586.

Patel CJ, Bhattacharya J and Butte AJ (2010) An environment-wide association study (EWAS) on type 2 diabetes mellitis. PLoS One 5:e10746.

Patterson SD and Aebersold RH (2003) Proteomics: the first decade and beyond. Nat Genet 33:311-323.

Pelegrini PB, del Sarto RP, Silva ON, Franco OL and Grossi-de-As MF (2011) Antibacterial Peptides from Plants: What They Are and How They Probably Work. Biochemistry Research international, vol. 2011, Article ID 250349, 9 pages. doi:10.1155/2011/250349

Perrin RM and Wigge PA (2002) Genome studies and molecular genetics - plant biotechnology web alert. Curr Opin Plant Biol 5:89-90.

Pinto NFJA Vírus do mosaico comum da cana-de-açúcar em sorgo (Sorghum bicolor (L.) Moench): Caracterização de isolados, reação de cultivares e herança de resistência. 1984. 136p. Tese (Doutorado) - Escola de Agricultura "Luiz de Queiroz". Piracicaba: 1984.

Portieles R, Rodrigues R and Cornide T (2002) Cytogenetics Characterization of new wild clones of the saccharum complex. Cultivos tropicales 25:17-22.

Postel S, Kemmerling B (2009) Plant systems for recognition of pathogen- associated molecular patterns. Semin Cell Dev Biol 20:1025-1031.

Prado RCO (2007) Soja é alimento e energia. In: Yutama, MM, Suzuki S and Camacho SA. Boletim de Pesquisa de Soja. Rondonópolis: Fundação MT, p.11-14.

Prosdocini F, Cerqueira GC, Bineck E, Silva AF, Reis NA, Junqueira ACM, Santos ACF, Junior NA, Wust CI, Filho FC, Kessedjian JL, Petretski JH, Camargo LP, Ferreira RGM, Lima RP, Pereira RM, Jardim S, Sampaio VS, Flatschart VS (2002) Bioinformática: manual do usuário. Biotecnologia, Ciência & Desenvolvimento 29:12-25.

Pryor T and Ellis J (1993) The genetic complexity of fungal resistance genes in plants. Adv Plant Pathol 10:281-305.

Ramos AM, Gally M, García MC, Levin L (2010) Pectinolytic enzyme production by Colletotrichum truncatum, causal agent of soybean anthracnose. Rev Iberoam Micol 27:186-190.

Reinstädler A, Müller J, Czembor J, Piffanelli P and Panstruga R. (2010) Novel induced mlo mutant alleles in combination with site-directed mutagenesis reveal functionally important domains in the heptahelical barley Mlo protein. BMC Plant Biol 10:31.

Richly E, Kurt J and Leister D (2002) Mode of amplification and reorganization of resistance genes during recent Arabidopsis thaliana evolution. Biol Evol 19:76-84.

Richter TE and Ronald PC (2000) The evolution of disease resistance genes. Plant Mol Biol 42:195-204.

Roach BT, Daniels J, (1987) Taxonomy and Evolution. In: Sugarcane Improvement trhough breeding. Heinz DJ ed, Vol 11 . Elsevier, Amsterdam, Netherlands. pp 7-84.

Rocha MM, Campelo JEG, Freire Filho FR, Ribeiro VQ and Lopes AC (2003) Estimativas de parâmetros genéticos em genótipos de feijão-caupi de tegumento branco. Revista Científica Rural 81:135-141.

Rocha MM, Freire Filho FR, Ribeiro VQ, Carvalho HWL, Belarmino Filho J, Raposo JAA, Alcântara JP, Ramos SRR and Machado CF (2007) Adaptabilidade e estabilidade produtiva de genótipos de feijão-caupi de porte semi-ereto na Região Nordeste do Brasil. Pesq agropec Brás 42(9):1283-1289.

Rodrigues MA (2011) Identificação e mapeamento de marcadores associados a locos de resistência a virose do mosaico dourado em feijão caupi (*Vigna unguiculata* (L.) Walp.). Dissertação de Mestrado. Feira de Santana. Universidade Estadual de Feira de Santana. Bahia.

Rossi M, Araujo PG, Paulet F, Garsmeur O, Dias VM, Chen H, Van Sluys MA and D'Hont A (2003) Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. Mol Genet Genomics 269:406-419.

Salomon D, Bonshtien A, Mayrose M, Zhang C, Shokat KM and Sessa G (2009) Bypassing kinase activity of the tomato Pto resistance protein with small molecule ligands. J Biol Chem 284:15289-15298

Salvaudon L, Héraudet V and Shykoff JA. (2005) Parasite-host fitness trade-offs change with parasite identity: genotype-specific interactions in a plant-pathogen system. Evolution 59:2518-2524.

Sanabria NM, Huang JC and Dubery IA (2010) Self/nonself perception in plants in innate immunity and defense. Land Biosc 1(1):40-54

Sanguino A (1998) Diagnóstico e controle do raquitismo da soqueira causado pela bactéria Clavibacter xyli subsp. xyli. STAB, Açúcar, Álcool e Subprodutos 17:26.

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.

Santos AS (2008) Doenças causadas por fungos. In: Leila Luci Dinardo-Miranda, Antônio Carlos Vasconcelos, Marcos Landell. (Org.). Cana-de-açúcar. Campinas, SP: Instituto Agronômico de Campinas.

Santos FR, Ortega JM (2003) Bioinformática. In: Borém A, Del Giúdice M and Sedyiama T (Org.). Melhoramento Genômico. 1. ed. Viçosa, MG: UFV 1:165-185.

Schmutz J et al. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-120.

Sels J, Mathys J, De Coninck BM, Cammue BP and De Bolle MF (2008) Plant pathogenesis-related (PR) proteins: A focus on PR peptides. Plant Physiol Biochem 46:941-50.

Seo PJ, Lee AK, Xiang F and Park CM (2008) Molecular and Functional Profiling of Arabidopsis Pathogenesis-Related Genes: Insights into Their Roles in Salt Response of Seed Germination. Plant Cell Physiol 49:334-344.

Shendure J and Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26:1135-1145.

Shepherd M, Lee D, Turnbull P, Brown T, Bush D, Brawner J, Bartle J, Macdonell P and Henry R (2011). "Eucalypts for biofuel production in northern Australia." Biofuels.

Simon MV, Benko-Iseppon AM, Resende LV, Winter P and Kahl G (2007) Genetic diversity and phylogenetic relationships in Vigna Savi germplasm revealed by DNA amplification fingerprinting (DAF). Genome 50:538-547.

Sindhu A, Chintamanani S, Brandt AS, Zanis M, Scofield SR and Johal GS (2008) A guardian of grasses: specific origin and conservation of a unique disease-resistance gene in the grass lineage. Proc Natl Acad Sci USA 105:1762-1767.

Skrabanek L, Saini HK, Bader GD and Enright AJ (2008) Computational prediction of protein-protein interactions. Mol Biotechnol 38:1-17.

Song WY, Pi LY, Wang GL, Gardner J, Holsten T and Ronald PC (1997) Evolution of the rice

Xa21 disease resistance genes family. Plant Cell 9:1279-1287.

Soto MJ, Dominguez-Ferreras A, Perez-Mendoza D, Sanjuan J and Olivares J (2009) Mutualism versus pathogenesis: the give-and-take in plant–bacteria interactions. Cel Microbiol 11:381–388.

Sousa ES (2010) Resistência de genótipos de soja *Glycine max* (L.) A Nezara viridula (L., 1758) (Hemiptera: Pentatomidae. Dissertação de mestrado Universidade Estadual Paulista"Julio de Mesquita Filho".

Spoel SH and Dong X (2008) Making sense of hormone crosstalk during plant immune responses, Cell Host Micr 3: 348-351.

Steele WM, Mehra KL (1980) Structure, evolution and adaptation to farming system and environment in Vigna. In: Summerfield, DR; Bunting, A.H., eds. Advances in legume science. England, Royol Botanic Gardens, pp.459-468.

Stergiopoulos J, van den Burg HÁ, Ökmen B, Beenen,HG, van Liere S, Kema GHJ and de Wit PJGM. (2010) Tomato *Cf* resistance proteins mediate recognition of cognate homologous effectors from fungi pathogenic on dicots and monocots. PNAS 107:7610-7615.

Sudhir K, Koichiro T and Masatoshi N (1993) MEGA: Molecular Evolutionary Genetics Analysis, version 2 0. The Pennsylvania State University, University Park, PA 16802.

Sun X and Wang G-L (2011) Genome-Wide Identification, Characterization and Phylogenetic Analysis of the Rice LRR-Kinases. PLoS ONE 6(3):e16079. doi:10.1371/journal.pone.0016079

Swiderski MR, Birker D and Jones JD. (2009) The TIR domain of TIR-NB-LRR resistance proteins is a signaling domain involved in cell death induction. Mol Plant Microbe Interact 22:157-165.

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, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28: 2731-2739.

Tang X, Xie M, Kim YJ, Zhou J, Klessing DF and Martin GB (1999) Overexpression of Pto activates defense responses and confers broad resistance. Plant Cell 11:15-29.

Tarr DEK and Alexander HM (2009) TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. BMC Res Notes 2:197.

Teófilo EM, Paiva JB and Filho LM (2001) Polinização artificial em *Vigna unguiculata* (L.) Walp). Ciênc Agrotec 25:220-223.

Thompson JD et al. (1997) The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl Acids Res 25:4876-4882.

Timko MP (2002) Molecular cloning in cowpea: perspectives on the status of genome characterization and gene isolation for crop improvement. In: Fatokun CA, Tarawali SA, Singh BB, Kormawa PM and Tamò M (eds) Challenges and opportunities for enhancing sustainable cowpea production. Proceedings of the World Cowpea Conference III held at the International Institute of Tropical Agriculture, Nigeria, pp 197-212.

Tokeshi H (1997) Doenças da cana-de-açúcar (híbridos de Saccharum spp.) In: Kimati H, Amorim L, Bergamin-Filho A, Camargo LEA, Rezende, JAM (Eds.). Manual de Fitopatologia. São Paulo: Editora Agronômica Ceres 2:207-225.

Tornero P, Gadea J, Conejero V and Vera P (1997) Two PR-1 genes from tomato are differentially regulated and reveal a novel mode of expression for a pathogenesis-

related gene during the hypersensitive response and development. Plant Microb Interact 10:624-634.

Tsuda K and Katagiri F (2010) Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. Curr Opin Plant Biol 13:459-465.

van der Biezen E, Freddie C, Kahn K, JE P and Jones J (2002) Arabidopsis RPP4 is a member of the RPP5 multigene family of TIRNB-LRR genes and confers downy mildew resistance through multiple signalling components. Plant J 29:439–451.

van Leeuwen H, Garcia-Mas J, Coca M, Puigdoménech P and Monfort A. (2005) Analysis of the melon genome in regions encompassing TIR-NBS-LRR resistance genes. Mol Genet Genom 273:240-251

van Loon LC and van Kammen A (1970) Polyacrylamide disc electrophoresis of the soluble leaf proteins from Nicotiana tabacum var. "Samsun" and "Samsun NN". II. Changes in protein constitution after infection with tobacco mosaic virus. Virology 40:190-211.

van Loon LC, Pierpoint WS, Boller Th and Conejero (1999) Recommendations for naming plant pathogenesis-related proteins. Plant Mol Biol Report 12:245-264.

van Loon LC, Rep M, Pieterse CMJ (2006) Significance of inducible defense-related proteins in infected plants. Annu Rev Phytopathol 44:135-162.

Vettore AL, da Silva FR, Kemper EL and Arruda P (2001). The libraries that made SUCEST. Genet Mol Biol 24:1-7.

Wan H, Zhao Z, Malik AA, Qian C and Chen J (2010) Identification and characterization of potential NBS-encoding resistance genes and induction kinetics of a putative candidate gene associated with downy mildew resistance in Cucumis. BMC Plant Biology 10:186-198.

Wanderley-Nogueira AC, Soares-Cavalcanti NM, Morais DAL, Belarmino LC, Barbosa-Silva A and Benko-Iseppon AM (2007) Abundance and diversity of resistance genes in the sugarcane transcriptome revealed by in silico analysis. Genet Mol Res 6(4):866-889.

Wanderley-Nogueira AC, Soares-Cavalcanti NM, Belarmino LC, Kido EA and Benko-Iseppon AM (2011) 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.

Wang X, Liu W, Chen X, Tang C, Dong Y, Ma J, Huang X, Wei G, Han Q, Huang L and Kang Z (2010) Differential gene expression in incompatible interaction between wheat and stripe rust fungus revealed by cDNA-AFLP and comparison to compatible interaction. BMC Plant Biol 10:9.

Xu WH, Wang Y, Liu G, Chen X, Tinjuangjun P, Pi LY, Zhang Y and Song WY (2006) The autophosphorylated Ser686, Thr688 and Ser689 residues within a putative protease cleavage motif in the juxtamembrane domain of *XA2*1 are implicated in the stability control of the rice receptor-like kinase. Plant J 45:740–751

Yuyama MM, Suzuki S, Camacho AS (2007) Boletim de Pesquisa de Soja. n° 11, Rondonópolis: Fundação MT, 274p.

Zhang X, Francis MI, Dawson WO, Graham JH, Orbović V, Triplett EW and Mou Z (2010) Overexpression of the Arabidopsis *NPR1* gene in citrus increases resistance to citrus canker. European J Plant Pathol 128:91-100.

Zhu QE, Maher S, Masoud R, Dixon R and Lamb CJ (1994) Enhanced Protection against fungal attack by constitutive coexpression of chitinase and glucanase genes in transgenic tobacco. Biotechnol 12:807-812.

Zipfel C, Felix G (2005) Plants and animals: a different taste for microbes? Curr Opin Plant Biol 8:353-360.

**Referências da Internet**

Companhia Nacional de Desenvolvimento (CONAB),
http://www.copersucar.com.br/dados.html (Acesso em Novembro 19, 2011).

COPERSUCAR,
http://www.copersucar.com.br/dados.html (Acesso em Novembro 19, 2011).

Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA),
http://www.embrapa.br (Acesso em Abril 13, 2011).

Food and Agriculture Organization (FAO)
http://www.fao.org (Acesso em Abril 13, 2011).

Genomes Online Database (GOLD),
http://www.genomesonline.org/ (Acesso em Janeiro 2, 2012).

International Rice Research Institute (IRRI),
http://irri.org/ (Acesso em Novembro 19, 2011).

International Service for the Acquisition of Agri-Biotech Application (ISAAA),
http://www.isaaa.org/ (Acesso em Janeiro 5, 2012).

National Center for Biotechnology Information (NCBI),
http://www.ncbi.nlm.nih.gov (Acesso em Junho 7, 2011).

PROCANA,
http://www.procana.com.br/ (Acesso em Março 8, 2011).

Sugarcane EST Project (SUCEST),
http://sucest.lad.dcc.unicamp.br/en/ (Acesso em Junho 7, 2011).

# CAPÍTULO I
## (Capítulo de Livro aceito para publicação em Functional Plant Science and Biotechnology on Sugarcane Pathology)

Special Issue (SI) of Functional Plant Science and Biotechnology on Sugarcane Pathology
Global Science Book, UK. Accepted December 2010

### Insight on Pathogen Defense Mechanisms in the Sugarcane Transcriptome

**Ana Carolina Wanderley-Nogueira[1], Ederson Akio Kido[1], Nina da Mota Soares-Cavalcanti[1], Luis Carlos Belarmino[1], João Pacífico Bezerra-Neto[1], William Lee Burnquist[2], Sabrina Moutinho Chabregas[2], José Ivo Baldani[3], Ana Maria Benko-Iseppon[1*]**

[1]Genetics Department,Center of Biological Sciences, Universidade Federal de Pernambuco, Av. Prof. Morais Rego 1235, CEP 50.670-420, Recife, PE, Brazil

[2]Centro de Tecnologia Canavieira (CTC), Bairro Santo Antonio, CEP 13400-970, Piracicaba, SP, Brazil

[3]Embrapa Agrobiologia (CNPAB), Rodovia BR 465 km 7, CEP 23890-000, Seropédica, RJ, Brazil

Corresponding author*: E-mail: ana.benko.iseppon@pq.cnpq.br

**ABSTRACT**

Recognition of pathogen and activation of defense mechanisms is a common feature known from all multicellular organisms. Among higher plants, systemic acquired resistance (SAR) is

known to activate pathogenesis-related (*PR*) genes after recognition of the pathogen mediated by a resistance (*R*) gene. Both gene classes (*R* and *PR*) represent the main mechanism against biotic and sometimes also abiotic stresses. Therefore, the identification of SAR metabolism components is an important concern regarding plant breeding. Sugarcane (*Saccharum officinarum*) is a major tropical and subtropical crop, grown in more than 120 countries, being especially important due to its value for the production of bioethanol, constituting a renewable energy source, besides the sugar production. The present work brings an overview on sugarcane resistance and pathogenesis-related genes, regarding their structure, abundance and role in the plant-pathogen metabolic pathway and also regarding their distribution, as compared with rice. For this purpose a collection of 282,818 expressed sequences tags have been evaluated to identify *R* and *PR* genes as well as important factors identified in the classic plant-pathogen metabolic KEGG pathway using rice full length cDNA as seed-sequences. The identified sugarcane genes have been also used to screen four SuperSAGE libraries with 8,787,313 tags, allowing the identification of the main activated and repressed genes under abiotic stress (drought/salinity) conditions. The 1,460 identified genes have been plotted on a rice virtual karyotype inferring about their distribution, considering a putative synteny as a measure to infer about their relative position within rice and sugarcane chromosomes. The results revealed interesting insights on the variability and complexity of defense genes in sugarcane.

**Key words**:bioinformatics, crosstalk among biotic and abiotic stresses, *PR*-genes, *R*-genes,

**Abbreviations**: **ABA**, abscisic acid; **AS**, salicylic acid; **avr**, avirulence; **BLAST**, basic local alignment search tool; **EST**, expressed sequence tag; **HR**, hypersensitive response; **LRR**, leucine rich repeats; **NBS**, nucleotide binding site; **PR**, pathogenesis related; **R**, resistance; **ROS**, reactive oxygen species; **SAR**, systemic acquired resistance; **Ser/Thre-Kinase**, serine/threonine kinase; **ST**, signal transduction; **SuperSAGE**, super serial analysis of gene expression.

**Running title**: Pathogen defense mechanisms in sugarcane. Wanderley-Nogueira *et al*.

## INTRODUCTION

The prevalence of a disease constitutes an abnormal condition, affecting plant growth and impairing important physiological processes. As highlighted by FAO (2005), fungi, bacteria, virus and nematodes are the main disease agents, resulting in serious losses to agriculture and also native plants, reducing the productivity, nutritional value and overall quality of the produced biomass.

During the plant-pathogen co-evolution, plants develop a complex network of synergic mechanisms to defend against pathogen attack (Pinzón *et al.* 2009). Considering this complex response, one of the most important steps includes detection of the possible invaders by the plant, a step where Resistance (*R*) genes play a crucial role (Mofet 2009). This sensing involves the recognition of a pathogen gene product called avirulence (avr) factor by a matching *R* gene. The plant will be resistant and the pathogen growth and establishment will be impaired when both *avr* and *R* genes are compatible, leading to the so called hypersensitive

response (HR) including local cell death to impair spreading of the pathogen (Jones and Dangl 2006). Besides this localized response, the HR activates a signal cascade which is able to establish resistance against a spectrum of different pathogens (Wang *et al.* 2005), corroborating the observations made at the beginning of the last century that plants, as animals, may be immunized against the attack a of given pathogen after infection by another pathogen (Chester 1933).

In the past decade many aspects of the systemic acquired resistance (SAR) have been elucidated (Ingle *et al*. 2006). The SAR pathway is also common in many incompatible plant-pathogen interactions (Park *et al.* 2010). As soon as the pathogen is detected, the plant induces a set of complex signal molecules that may activate defense proteins (Humphry 2010) or that may have direct antimicrobial effect, as it is the case of the pathogenesis-related (*PR*) genes (Durrant and Dong 2004) or alternatively the production of secondary metabolites that impair pathogen movement or growth within the plant tissues (Sparla *et al.* 2004; Benko-Iseppon *et al.* 2010).

**Categories of *R* genes**

Altogether *R* genes have been recently classified into five different groups or classes, defined by their conserved domains (CD; Bent 1996; Hammond-Kosak and Jones 1997; Ellis and Jones 2000).

The first class is represented by the *HM1* gene of maize that codes for a reductase able to inactivate toxins produced by the fungus *Helminthosporium carbonum* (Joahal and Briggs 1992), being the only *R* gene class where conserved domains are absent. A second class is represented by the *Pto* gene from tomato, that confers resistance against the bacterium

*Pseudomonas syringae* pv. *tomato* – a pathogen that express the corresponding *avrPto* avirulence gene (Martin *et al.*1993). The *Pto* gene is characterized by a serine/threonine-kinase (ser/thre-kinase) domain, able to interact with the *avrPto* gene (Tang *et al.* 1999). This gene was also identified in other plants *viz.Arabidopsis thaliana*, *Phaseolus vulgaris* (Melotto *et al.* 2004) and *Saccharum officinarum* (Wanderley-Nogueira *et al.* 2007).

The third class is represented by genes bearing two domains: leucine-rich repeats (LRR) and nucleotide binding site (NBS) (Liu *et al.* 2004). This is the case of the *Rpm1* and *Rps2* genes from *A. thaliana (*Mindrinos *et al.* 1994), the *N* gene from *Nicotiana tabacum*(Whitham *et al.* 1994), *L6* from *Linum usitatissimum* (Lawrence *et al.* 1995), *Prf* from *Lycopersicon esculentum* (Salmeron *et al.* 1996) and *Rpg1* from *Hordeum vulgare* also found in *P. vulgaris*, *G. max* and *Vicia faba* (Brueggeman *et al.* 2008) The fourth *R* gene class codes for a membrane anchored protein composed by a LRR extracellular domain, a transmembrane region and a short intracellular tail in the C terminal. The *Cf* gene from *L. esculentum* is an example of this class, conferring resistance against *Cladosporium fulvum* (Jones *et al.*1994; Dixon *et al.* 1996).

The *Xa21* gene from rice that confers resistance to the bacterium *Xanthomonas oryzae*pv. *oryzae*is a representative of the fifth class (Song *et al.* 1995; Wang *et al.* 1995). This gene encodes for an extracellular LRR domain (similar to the *Cf* gene), as well as a Ser/Thre-Kinase (Serine/Threonine Kinase) domain (similar to the *Pto* gene), suggesting an evolutionary ligation among different classes/domains in the genesis of plant *R* genes (Song *et al.* 1997).

**Pathogenesis Related (*PR*) gene categories**

The PR proteins comprise of pathogen-induced proteins being routinely classified into 17 families based on their biochemical and molecular biological properties, from PR-1 to PR-17 (van Loon *et al*. 2006).

Similarities among sequences and serologic or immunologic properties are the base of their classification (Van Loon *et al.* 1999). Functionally some PR proteins bear antifungal properties, as the chitinases (PR-3) and β-glucanases (PR-2) (Zhu *et al.* 1994), while others like PR-1 from *N. tabacum* (Atici and Nalbantoğlu 2003), *A. thaliana* (Metzler *et al.* 1991), *L. esculentum* (Tornero *et al.* 1997) and *Malus domestica* (Bonasera *et al.* 2006) present unknown phytochemical functions.Despite of that, this gene class is considered to be a typical SAR marker (Bonasera *et al.* 2006).

Most *PR* genes are expressed to a basal level under normal growth conditions, but are rapidly induced by pathogenic infections. It is notable that several *PR*genes are also regulated during development, for example during leaf senescence and pollen maturation, and also by environmental factors, as osmotic stress, cold and light (Broekaert *et al.* 2000; Zeier *et al*. 2004).

Some PR proteins, including PR-1, chitinases, and thaumatin-like proteins are expressed under cold stress in overwintering monocots, exhibiting antifreeze activities (Hon *et al*. 1995; Atici and Nalbantoğlu 2003; Griffith and Yaish 2004).

Many *PR*genes are constitutively expressed in given plant tissues (Velazhahan and Muthukrishnan 2003; Liu *et al*. 2004), indicating that at least some members of the PR proteins play important roles in plant development other than defense responses, as occur with PR-2 protein in *N.tabacum,* that play an important role in seed germination (Seo *et al.* 2008) .

Consistent with this notion, it has been shown that a tobacco PR-2 protein plays a role in seed germination (Leubner-Metzger and Meins 2000; Leubner-Metzger 2005).

Although many *PR*genes have been identified in different plant species, recent research have focused on their molecular aspects by using them as indicators for pathogenic infections, salicylic acid (SA) signaling, and SAR (van Loon et al. 2006). Transgenic approaches were also employed to obtain information on the role of PR proteins in disease resistance response. As a result, it is unclear how the *PR* genes are regulated by environmental conditions in consonance with plant growth hormones, and to what extent the PR proteins affect plant development.

**Hormones**

Besides *R* and *PR* genes, hormones are also important signaling molecules, playing an important regulatory role in plant development and inducing the expression of many PR proteins. Such hormones are produced at specific sites and in low amounts, also being active in defense pathways, as it is the case of SA, jasmonic acid (JA) and ethylene (ET), considered to be the main molecules activating defense genes(Lu 2009; Divi *et al.* 2010).

Induction of genes encoding *PR-1, PR-2* (β-1,3-glucanase), and *PR-5* (thaumatin-like protein) requires SA signaling. In contrast, genes encoding *PR-3* and *PR-4* (both with endochitinase activities) are independent of SA signaling and depend on JA pathway. For example, in arabidopsis, SA and JA activate distinct sets of *PR* genes in an antagonistic pattern (Thomma *et al*. 1998).

The SA pathway is primarily linked to resistance to biotrophic pathogens i.e. organisms that feed and reproduce on living tissues. This is in contrast to JA and ET, which mediate

resistance mostly against necrotrophic pathogens (organisms which kill their hosts and derive nutrients, live and multiply on dead tissue). This differentiation of defense signaling pathways suggest that plants detect differences between pathogen lifestyle and mode of infection. Genetic evidence for JA antagonism of SA signaling pathways is well documented, but emerging data suggest a more complex signaling network evoking both positive and negative regulatory interactions (Spoel *et al*. 2007; Lópes *et al*. 2008; Vlot *et al*. 2008).

SA is a strong inducer of *PR* genes, and particularly *PR-1* is traditionally used as a marker for SA-mediated defence (Gaffney *et al*. 1993). A different set of genes is activated by JA, such as *VSP2* and *PDF1.2* (Benedetti *et al*. 1995; Penninckx *et al*. 1998).The latter, *PDF1.2*, also responds to ET (Thomma *et al*. 2001). Lately, it has become apparent that plant growth hormones and modulation of developmental processes was not recognized earlier as being partof plant defense. A model proposed by Robert-Seilaniantz *et al*. (2007) showed that auxin and cytokinins promote biotrophic susceptibility by inducing necrotrophic resistance pathways via JA/ET. Plants have evolved mechanisms to suppress auxin signaling as a component of basal defense in order to hinder the invading pathogens from using the hormone as virulence factor (Navarro *et al*. 2006). In parallel, gibberellic acid (GA) induces necrotrophic susceptibility, by inducing the biotrophic SA resistance pathway (Robert-Seilaniantz *et al*. 2007).

However, most of these interactions hitherto remain to be proved. For example, the role of abscisic acid (ABA) in a plant stress context is complex. ABA is a well known component in abiotic stress responses, but has been shown recently to be important in defense to various pathogens, highlighting the known crosstalk among different stress types (Asselbergh *et al*. 2008; Schenke *et al.* 2011). For example, in the interaction among Arabidopsis and

*Leptosphaeriamaculans*, ABA is important for resistance, wherein both callose dependent and independent pathways are present (Kaliff *et al*. 2007). Indeed, some transcriptionfactors and signaling molecules are common molecular players in both biotic and abiotic stresses, common to multiple networks or involved in crosstalk between stress signaling pathways regulated by abscisic acid, salicylic acid, jasmonic acid and ethylene as well as ROS (reactive oxygen species) signaling (Velázquez *et al.* 2011).

**Reactive oxygen species (ROS)**

ROS are produced by all aerobic organisms as a by-product of aerobic metabolism. A substantial increase in the intracellular concentration of ROS is generally associated with deleterious effects, including cell death by apoptosis or necrosis, in pathological conditions such as inflammation. The generation of ROS also appears to be required for many normal cellular functions, including transduction of cell surface receptor signaling (Woo and Rhee 2010).

Depending on the nature of the ROS, some are highly toxic and rapidly detoxified by various cellular enzymatic and non-enzymatic mechanisms. Whereas plants are fortified with mechanisms to combat increased ROS levels during abiotic stress conditions, in other circumstances plants appear to purposefully generate ROS as signaling molecules to control various processes including pathogen defense, programmed cell death, and stomatal behavior (Apel and Hirt 2004).

Altogether, *avr*, *R* and *PR* genes, as well as hormones and ROS are active during plant-pathogen interaction as illustrated in **Fig. 1**.

**The sugarcane expressed sequence tag (EST) database**

Many projects have been carried out aiming to identify expressed sequence tags (ESTs) in sugarcane, including initiatives in Brazil, South Africa, Australia, and France (Grivet and Arruda 2001; Perrin and Wigge 2002). Together these projects produced more than 300,000 sugarcane ESTs. In Brazil, the ONSA (Organization for Nucleotide Sequencing and Analysis;http://watson.fapesp.br/onsa/Genoma3.htm) consortium initiated the SUCEST (Sugarcane Expressed Sequence Tag Project; http://sucest.lad.dcc.unicamp.br/en/) – an initiative that generated about 238,000 ESTs distributed over 26 libraries from different tissues and across developmental stages. The data generated is still not completely evaluated and many gene families remain to be analyzed, as it is the case of the *PR* genes which are here evaluated for the first time. Despite of the considerable genetic variability among sugarcane germplasm revealed by molecular markers (e.g. Parida *et al.* 2010, Costa *et al.* 2011), few evaluations of the genetic diversity and complexity of this crop are available, especially regarding genes associated to the pathogen response.

The present evaluation brings an overview of the mainsequences regarding plant-pathogen interaction in sugarcane, considering the up to date knowledge regarding the model plant rice. It also brings an insight on the expression of such sequences regarding the differential expression in sugarcane by Super-SAGE (Super Serial Analysis of Gene Expression) during abiotic stresses (drought and salinity), showing the dynamic relationship in the response and its modulation during different stress situations, explained by the crosstalk mechanism, i.e., the co-activation of genes among both biotic and abiotic stress types in an interaction.

## MATERIALS AND METHODS

### Rice sequences used as reference database to identify sugarcane candidates

In the present chapter, the first approach for the identification of sugarcane clusters associated with plant pathogen interaction was based on the metabolic pathway map046262 at KEGG (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/). For this purpose, KEGG sequences were downloaded and used in the search of the respective full length cDNA from rice that were subsequently used as reference-sequences. The selected genes included 26 *R* or *PR*-genes, five ROS genes and 18 genes that act with signal transduction and are involved in the hypersensitive response, including hormones.

An additional analysis was performed using 17 *R*-genes previously compiled by Barbosa-da-Silva *et al.* (2005) and Wanderley-Nogueira *et al*. (2007) and 15 *PR*-genes searched by keyword, function and conserved domains at NCBI (PR-1: NP_179068; PR-2: NP_191285.1; PR-3: NP566426.2; PR-4: NP_187123; PR-5: NP_177641.1; PR-6: AAA16881; PR-7: NP_001234257; PR-8: CAM82810; PR-9: AAA34108; PR-10: ACF5101; PR-11: CAA01263; PR-12: P30230; PR-13: AEE35295; PR-14: CAA91436; PR-15: AAB561565.1; PR-16: NP_001061164).

Thus, a total of 81 full length cDNA sequences from *Oryza sativa* from the MSU Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/); composed the reference-sequence database.

### Analysis of the sequences obtained

For the identification of these gene analogs in sugarcane transcriptome, tBLASTx (Basic Local Alignment Tool) alignments were carried out against a local database [Clusterized ESTs

from NCBI (***National Center for Biotechnology Information***) and SUCEST] including 282,878 sequences, using a cut-off value of $e^{-5}$.

Obtained sugarcane clusters were annotated and analyzed regarding score, e-values, sequence size and presence of conserved domains as shown in Table 1.For this purpose sugarcane clusters were translated using the TRANSLATE tool of Expasy (http://us.expasy.org/) and screened for conserved motifs with aid of the RPS-BLAST CD-search tool (Altschul *et al.* 1990). The best match for each gene was submitted to a BLASTx alignment in NCBI GenBank in order to confirm their putative function (results are available in http://150.161.22.10/sugarcane/all.fasta.html). A second manual analysis was also carried out, followed by an elimination of clusters that matched to more than one gene due to common domains. For this purpose, matching clusters to each query sequences were annotated on a local database called 'non-redundant'.

**Evaluation of genes associated to pathogen invasion under abiotic stress by Super-SAGE**

SuperSAGE data was generated by our group as described by Molina *et al*. (2008) in collaboration with GenXPro GmbH (http://www.genxpro.info/) and Brazilian partners (CTC – The Center of Sugarcane Technology, http://www.ctcanavieira.com.br/; CETENE - Centro de Tecnologias Estratégicas do Nordeste, http://www.cetene.gov.br/home/index.php). The RNA samples were obtained from roots of selected sugarcane accessions (*Saccharum* spp.) under salinity (100 mM NaCl) or drought (after 24 h irrigation suppression) stresses in two distinct experiments:

(A) Salinity stress assay: acclimated plants of the clone RB931011 (referred as tolerant to salinity stress by the Brazilian RIDESA program of sugarcane breeding) were grown in a

greenhouse (CETENE, Recife, Brazil) in pots containing washed sand and watered daily with Hoagland solution. For salinity stress application 100 mM NaCl was added to the mentioned solution. After salinity stress induction (30, 60 and 90 minutes) roots were collected from both, stressed and non-stressed (negative control) plants, that were immediately frozen in liquid nitrogen until processing and total RNA extraction. Equimolar amounts of each sample/time were assembled to form the bulks referred to the two SuperSAGE libraries (stressed and control).

(B) Drought stress assay: selected genotypes provided by the Center for Sugarcane Technology breeding program (CTC, Piracicaba, Brazil) were previously identified in a drought tolerance essay. Among them, four genotypes composed the drought sensitive group (CTC9, CTC13, SP90-3414 and SP90-1638) and another four regarded the drought tolerant group (CTC15, CTC6, SP83-2847 and SP83-5073). The plants were grown in 10 liters containers up to three months of development when part remained irrigated and part were submitted to water suppression (24 hours). Roots samples were collected, identified and frozen in liquid nitrogen until processing and total RNA extraction aiming to compose the four SuperSAGE libraries: TD: tolerant under stress; TC: tolerant control; SD: sensitive under stress and SC: sensitive control.

Unique tags (26 bp) differentially expressed (p-value 0.05) were identified using DiscoverySpace4.01 software. All tags were submitted to BLASTn (score = 52; 100% of identity) against the previously identified *PR* and *R* sugarcane contigs dataset previously annotated using Uniprot-SwissProt database (BLASTx; e-value $e^{-10}$). All data were normalized (one million tags by library) and the fold change (FC) of a differentially expressed tag was calculated with the tag frequency in a stressed library in relation to the respective control

(without stress). The normalized data matrix regarding the gene expression was analyzed using a hierarchical clustering approach with aid of the program Cluster3.0.Dendrograms (using the weighted pair-group analysis) were generated using the TreeView program.

**Anchoring sugarcane genes in the rice physical map**

Considering the availability of a genome browser for the rice genome, all sugarcane ESTs and SuperSAGE tags were aligned against *O. sativa* non-redundant clusters anchored on virtual chromosomes from the MSU Rice Genome Annotation Project. For this purpose, the MegaBLAST tool was used with at least 80%identity or direct association with the differentially expressed SuperSAGE tags.

Although the BLAST algorithm was adjusted to make searches as sensitive as possible and considering that this tool only generates local alignments, to ensure that the anchored clusters in the rice genome represented real genes and not just similarity with random segments, the obtained data underwent to a second screening, which considered possible splicing sites, deletions and/or insertions, and the total length of the clusters (considering that the data should represent at least 30% of cluster size).

For better data organization, clusters that represented high similarity with the same region of the rice genome were grouped. The groupings are named in increasing order as GRn, where "GR" means group and "n" the group number. For the virtual karyotype representation a CorelDRAW12 graphic application was used; the rice chromosome information to the schematic representation was obtained from GRAMENE site (http://gramene.org/Oryza_sativa/Location/Genome). For the design of chromosomes, considering the high data amount to be anchored in the genome and the necessity of high

resolution of bands, a random length of 200 mm was adopted for the larger chromosome (number 1); thus, for the band marking, each 1 mm corresponded to 126,344 bp. Moreover, clusters and groupings that schematically appeared very close to each other were indicated as a single band.

## RESULTS AND DISCUSSION

The tBLASTx alignment in sugarcane database using the 49 KEGG pathway reference sequences returned 1,150 clusters, indicating that all members of this pathway are represented in sugarcane (**Fig. 1)**. Among these clusters only 244 were non-redundant, aligning each with one corresponding reference sequence. Other 906 clusters presented repetitions, aligning with two or more reference sequences, confirming the excellent coverage that the existing Sugarcane EST databank comprises, including the most important representatives from different gene families. Regarding the exclusive 244 sugarcane clusters, 74 aligned with *PR* or *R*-genes, 14 aligned with ROS genes and 156 with signal transduction (ST) genes or genes involved in the HR (**Table 1**, **Fig. 1**). From these, 64% aligned with genes of the hypersensitive response, 30% with *R* and *PR* genes while 6% aligned with ROS. As reviewed by Benko-Iseppon et al. (2010) genes of the HR are associated with different stimuli, being activated not only due to pathogen perception, but also under abiotic stress, whilst many of them are constitutively active in low levels, also in the absence of any kind of stress, justifying their prevalence in the present evaluation.Thus, a growing number of evidences supports the notion that plant signaling pathways consist of elaborate networks with frequent crosstalk, thereby allowing plants to regulate both abiotic stress tolerance and disease resistance (Velázquez *et al.* 2011).

Most of the 169 clusters that aligned with two different genes were sequences bearing similar conserved domains like LRR, Kinases or NBS. This occurred among the *Pti4* and *Pti6*, sharing 50 clusters; *WRKY25* and *WRKY29* with 25 identical sequences, *Cf2* and *Cf5* with 7 and *ERF* and *Xa21* that shared three identical sugarcane clusters. Other 737 clusters repeat themselves from three to 20 times among the 49 genes presented in KEGG plant response pathway. This is an expected result also observed in a previous approaches (Barbosa-da-Silva *et al*. 2005; Wanderley-Nogueira *et al.* 2007), due to similar CDs existing in *R* genes among angiosperms.

The 16 classes of *PR* genes search returned 314 non redundant sugarcane clusters. Among them, only four sequences matched with two different reference sequences (*PR-4* and *PR-6*) and the remaining 310 pertained exclusively to one *PR*-gene (**Fig. 2A**).

The most abundant category was *PR-9* (with 93 representatives), including factors classified as peroxidases (**Fig. 2A**). They contribute to plant disease resistance in several ways including (i) the strengthening of plant cell walls through the deposition of lignin, which is thought to be a general defense mechanism against a broad spectrum of pathogens; and (ii) the production of toxic radicals such as hydrogen peroxide (Van Loon and Van Strien 1999).

*PR-2* was the second most frequent gene category with 41 members (**Fig. 2A**), regarding a group of β-1,3-glucanases, whose role in disease resistance is often related to their glucanase activity. PR-2proteins can either directly impair the growth of a fungus by hydrolyzing β-1,3/1,6-glucans within fungal cell walls or by releasing short glucan fragments from pathogen cell walls, which can also be recognized by plants and further induce plant defense responses (Ebel and Cosio 1994; Van Loon and Van Strien 1999).

PR-5 (thaumatin-like), 7 (endoproteinase) and 14 (lipid transfer protein) were also well represented with 37, 35 and 39 matches, respectively (**Fig. 2A**). The PR-5 family belongs to the thaumatin-like proteins with homology to permatins that permeabilize fungal membranes (Vigers *et al.* 1991). Some members of this family have been shown to possess antifungal activity, particularly against oomycetes. Basic PR-5 proteins (osmotin) are induced in tobacco and tomato in response to osmotic stress (Singh *et al.* 1987). PR-7 acts as an endoproteinase. Because lysis of fungal cell walls often requires degradation of cell wall proteins in addition to hydrolysis of chitin and glucan, it seems reasonable to assume that PR-7 serves as an accessory to antifungal action (Goldman and Goldman 1998), while lipid-transfer proteins (LTPs) (PR-14) exhibit both antifungal and antibacterial activity, exerting their effect at the level of the plasma membrane of the target micro-organism. Their relative diversity in sugarcane reinforces previous observations (Garcia-Olmedo *et al.* 1995).

Regarding the additional 17 *R*-genes analysis, other 118 non-redundant sugarcane clusters presented matches with *O. sativa* reference sequences. Eleven of these clusters aligned with two different *R*-genes (*Mi1*and *GPA2*; *HERO* and *Pi-ta*) and 107 aligned exclusively with 10 genes (**Fig. 2B**).

The genes *P* of flax, *I2* and *Mi* oftomato, *HERO*, *Rx1* and *GPA2* of potato*Hrt* of arabidopsis (Cooley *et al.*2000) and *Pib, Pi-ta* and *Xa1* of rice are all members of the NBS-LRR family, that was the most represented class in the sugarcane transcriptome (**Fig. 2B**); this class common feature is the presence of(i) leucine-rich repeats which play a direct role in protein-protein specific recognition event; LRR motif contains 23-25 aa with a consensus sequence (LxxLxxLxLxxNxLt/sgxIpxxLG); (ii) a nucleotide-binding site that is highly conserved among plants, and is similar to mammalian CED-4 and APAF-1 proteins which are

involved in apoptosis (Chinnaiyan *et al.* 1997) and usually signal for programmed cell deathplaying a role in activation of downstream effectors (Bryan *et al.* 2000), and (iii) coiled-coil (CC) or a TIR (Toll Interleukine-Like domain), involved in signal transduction during many cell processes.

Curiously, these 10 genes that confer resistance against bacteria, fungi, virus or nematodes, despite of sharing common domains, presented exclusive alignments with sugarcane sequences while the other 54 *R*-genes (considering kinases, NBS-LRR, LRR and other families) presented alignments with common clusters.

A high number of NBS-LRR class representatives was observed (490 clusters) as compared with other classes previously observed by Barbosa-da-Silva *et al*. (2005) and Wanderley-Nogueira *et al.* (2007) and were expected because most of *R*-genes pertain to this family that play a crucial role in plant defense.

Other 443 sequences aligned with the seven remaining *R*-genes previously identified among the 1,150 sugarcane clusters matching with KEGG pathway. As occurred in most genes selected from theKEGG database, repetitions probably appeared because of the occurrence of different combinations of the same domains (16 genes present LRR, 15 genes present NBS and 20 genes are protein kinases, for example).

Most of the sugarcane clusters that appeared as best matches presented the same conserved domains complete.In the few cases where domains were not identified, their function was confirmed using BLASTx at NCBI.

Considering the taxonomic affiliation of the best matches, as expected, their alignment occurred with monocot species, among them, 74% aligned with *Sorghum bicolor*, 15% with *Zea mays*, 9% with *O. sativa*, and 1% with *Saccharum* hybrid, with a single alignment to a

non-angiosperm (1% to *Laccaria bicolor*, a mushroom). Considering the high abundance of rice sequences in data banks, it would be expected that most matches would belong to sequences of this species. On the other hand, the higher similarity to sorghum and maize is justified by their taxonomic proximity to sugarcane within *Panicoidae* of the *Poaceae*, where sorghum and sugarcane grouped together in the Andropogoneae(Watson and Dallwitz 1992; Bouchenak-Khelladi *et al.* 2008).

**Pathogen defense genes and abiotic stress Super-SAGE assay**

Despite of the focus of the described SuperSAGE being the transcriptome of plants under abiotic stress [drought: water suppression; salinity (100 mM NaCl)], many R (**Table 2**) and PR (**Table 3**) transcripts have been identified, an expected result since biotic and abiotic signaling pathways share multiple nodes and their outputs have significant functional overlap, and are able to be cross-activated. Examples of such events were presented by Chini *et al.* (2004) while establishing enhanced expression of the CC-NBS-LRR gene, *ADR1* (a broad spectrum disease resistance gene, with serine/threonine protein kinases domain) that conferred tolerance to drought.

Tags associated to plant R transcripts representing the major NBS-LRR class were observed to be differentially expressed (**Fig. 3A)**.Usually, R proteins can directly interact with a product of the avirulence gene of the pathogen or indirectly by guarding another protein that is the target of the avirulence protein. Moreover, they can detect pathogen associated molecular pattern (PAMPs) and participate in activation of the innate immune system that will protect the host from infection, or even in degradation by a pathogen toxin. Another important class, the recognition receptors pattern (RRP), which include receptor protein kinases

implicated in PAMP perception were also presented in the SuperSAGE results. Thus some tags annotated as receptor-like kinases (RLK), mainly LRR, were also differentially expressed (**Fig. 3A**). Receptor kinases play important roles in cell signaling being responsible for the information about the cell surroundings. This activation may be justified by the fact that some abiotic stresses open a door to pathogen invasion.

Considering the Figure 3A, the cluster I showed overexpressed tags in salinity-tolerant and in drought-sensitive genotypes. Examples of annotated R genes with high FC are represented by the cluster I were COI1 (an important jasmonate receptor) and TIR1 (auxin receptor) are included. Jasmonate signaling plays a critical role in protecting plants from pathogens or insect attack and in limiting damage from abiotic stresses (Hu *et al.* 2009). The phytohormone auxin has been implicated in developmental plant processes, including apical dominance, tropic responses, vascular development, organ patterning, flower and fruit development. Kepinski and Leyser (2005) provided evidence for a role of auxins in plant defense responses and suggested cross-talk between auxin, abiotic stress and biotic stress signaling pathways. Curiously, the same tags were underexpressed in the drought-tolerant library when compared with the respective control. The clusters II and IV presented upregulated tags in drought-tolerant genotypes, but downregulated in the other comparisons. Examples of tags with this behavior are CRK19 and CRK26. Cysteine-rich Receptor-like Kinases (CRKs) have been suggested to play important roles in the regulation of pathogen defense and programmed cell death and CRK is part of plant Receptor-like kinases (RLK), a group of conserved signaling components that regulate developmental programs and responses to biotic and abiotic stresses (Wrzaczek *et al.* 2010). Some overexpressed RLK members were observed in cluster III (also in the others clusters). Tags of cluster III seem to be related to both stresses in a general

manner. RLK represents nearly 2.5% of arabidopsis protein coding genes (Shiu and Bleecker, 2001). Others RLK members observed were SERK1 (cluster V); SERK2 (cluster IV) and SRF3 (cluster I, II and V), SRF7 (cluster II, III) and SRF8 (cluster II).

In relation to the differentially expressed SuperSAGE tags associated with pathogenesis-related proteins presented in the root transcriptome of sugarcane after abiotic stress induction (**Fig. 3B**), some families were observed [PR-2 (Glucan endo-1,3-beta-glucosidase), PR-3 and PR-4 (chitinase), PR-5 (thaumatin-like), PR-6 (proteinase inhibitor), PR-9 (peroxidase), PR-14 (LTP)]. Potential chitinase tags were observed in many clusters (I, II, IV, V, VI). Liu et al. (2004) suggested that the class IV chitinase PmCh4A was involved in the defense response of western white pine to infection and also abiotic stress factors, besides their primary role in plant defense (these enzymes degrade the major component of fungal cell walls, ß-1,4-linked polymers of *N*-acetyl-d-glucosamine). Additionally, some chitinases were also components of plant defense against higher concentrations of specific heavy metals, showing that their biological role is complex and more than expected (Békésiová *et al.* 2007). Another PR family observed in all clusters is the peroxidase group (PER1, PER2, PER3, PER4, PER12, PER15, PER30, PER35, PER36, PER51, PER52, PER72; Fig. 3B). Manandhar *et al.* (1999) presented findings showing that the accumulation level of peroxidase transcripts (and some PR-proteins: PR-1, PR-2, PR-3, PR-4 and PR-5) were higher in rice in response to *Bipolaris sorokiniana* and UV light than in response to avirulent isolate of *Pyricularia oryzae*.

Additional PR components observed in association with SuperSAGE tags (**Fig. 3B**) and abiotic stress were:

a) Cytosolic Ascorbate Peroxidase 1 (cAPX1; tag in cluster IV): Davletova *et al.*(2004) demonstrated that in the absence of the cytosolic $H_2O_2$-scavenging enzyme APX1, the entire

chloroplast $H_2O_2$-scavenging system of *A. thaliana* collapses, $H_2O_2$ levels increase, and protein oxidation with programmed cell death occurs. On the other hand, cAPX2 (cluster IV and VI) products showed to be involved in flooding stress responses in young soybeanseedlings (Shi *et al.* 2008);

b) Thaumatin – PR-5 (tags in clusters I, II, IV, V, VI, VII): transgenic plants with constitutively higher expression of the *GbTLP1* [cotton (*Gossypium barbadense* L.) thaumatin-like protein gene] showed enhanced resistance against different stress agents, particularly, *Verticillium dahliae* and *Fusarium oxysporum*, increasing also tolerance against some abiotic stresses including salinity and drought (Munis *et al.* 2010);

c) Germin-like proteins (tags in clusters IV, V, VII): GLP constitute a ubiquitous family of plant proteins that seem to be involved in many developmental and stress-related processes. Berna and Bernier (1999) showed that expression of the wheat germin gene was also stimulated by some abiotic stresses, especially the heavy metal ions $Cd^{2+}$, $Cu^{2+}$ and $Co^{2+}$;

d) Glucan endo-1,3-beta-glucosidase – PR-2 (synonym: [beta]-1,3-glucanase; tags in clusters I, IV, V, VI): this enzyme acts by reducing the influx of solutes into the membrane vesicles during freezing and thereby reduces osmotic stress and vesicle rupture during thawing (Hincha *et al.*1997);

e) subtilisin-like protein: the salinity stress response in arabidopsis requires a subtilisin-like serine protease (AtS1P), related to membrane-localized b-ZIP transcription factor, AtbZIP17. Liu *et al.* (2007) observed that salinity stress induced a signaling cascade involving the processing of AtbZIP17, its translocation to the nucleus and the upregulation of salinity stress genes;

f) Non-specific lipid-transfer protein (nLTP; tags in clusters III, IV, V, VII): LTP is another protein family involved in plant stress response (Jung *et al.* 2005; Sarowar *et al.* 2009), having the ability to transfer phospholipids between a donor and an acceptor membrane (when this activity is not specific,the peptide sare called non-specific lipid transfer protein).

Besides the valuable identification of important genes associated with abiotic response, as osmoprotectants (Silva *et al.* 2011) by SuperSAGE, the here observed presence of PR genes during abiotic stress induction was also confirmed by Kido *et al.* (2010), that presented a functional review of antimicrobial peptides and an overview of SuperSAGE transcriptional profile of defensin (PR-12), thionin (PR-13) and LTP (PR-14) in libraries of some important crops (cowpea, soybean and sugarcane), again confirming their influence in mechanisms regarding biotic and abiotic genes. SuperSAGE allowed also the generation of a comprehensive panel of the differentially expressed kinases under biotic and abiotic stresses in cowpea (*Vigna unguiculata*), revealing their association with both kinds of stress (Kido *et al.* 2011). Such crosstalk interactions are evident in many differential expression profiling assays and indicate that in the future few genes may be useful to induce a myriad of responses, maybe useful for tolerance/resistance increase in crop plants.

**Anchoring sugarcane sequences in the rice virtual karyotype**

Beyond the BLAST algorithm parameters (score, e-value and percentage of identity) other features were considered (as the probable splicing sites, putative insertions, deletions and cluster full length in bp) for anchoring 73.22% of the 1,460 sugarcane clusters studied in the rice virtual karyotype (**Fig. 4**), in an attempt to infer about their relative position regarding possible synteny and colinearity among sugarcane and rice chromosomes.

Flowering plants originated approximately 200 million years ago (MYA; Wilkstrom et al. 2001) and subsequently diverged into several lineages. The Poaceae family arose about 60 MYA and diverged into different species mainly due to genome-wide amplification (Copley *et al.* 2001), in tandem gene duplication and events of local chromosome changes (Kondrashov *et al.* 2002). It is not surprising that 1,069 sugarcane clusters anchored in some region of the rice chromosomes, and one could assume that this occurred since both are members of Poaceae family; so the evolutionary proximity of rice and sugarcane means that consistent levels of homologymay be expected amongboth species.

Vincentz *et al.* (2004) performed a comparison between sugarcane, rice and arabidopsis transcriptomes and found that 70.5% of sugarcane sequences were similar to arabidopsis (suggesting that their genes probably encoded essential angiosperm functions) and 81.6% had significant match with rice genome, so, these 11.1% sequences represents putative monocot specific material. Additionally, the other 18.4% of sugarcane sequences may correspond to gene losses in rice or fast-evolving sequences that diverged substantially. This can be the reason because 25.88% of our sugarcane clusters did not align to any region of the rice genome.

Sugarcane clusters appeared anchored in all segments of rice chromosomes. The rice chromosomes present heterochromatic regions (Cheng *et al*. 2001), but due to the limitations to sequence those regions, they are not linearly represented along the virtual karyotype, justifying the anchoring of many sugarcane sequences around the centromeres. Additionally, as expected, several sequences clustered along the genome, with some chromosomes rich in resistance genes (e.g. chromosomes 1 and 3) while other regions were relatively poor regarding the evaluated genes (e.g. chromosomes 4, 8, 10 and 12).

Clustering of *R* and *PR* genes confirms the existing theory that a common genetic mechanism has been involved in their evolution. Most resistance genes have been demonstrated to reside in clusters (Kanazin *et al.* 1996) as reported in maize (Dinesh-Kumar et al. 1995), lettuce (Maisonneuve et al. 1994), oat (Rayapati *et al.*1994), flax (Ellis *et al.* 1995) and chickpea (Benko-Iseppon *et al.* 2003). The formation of gene clusters is in general associated with duplication processes followed by diversification through pressure from the pathogen or the environment, in the case of *R* genes and *PR* respectively. It is interesting to note that in short arms of chromosomes 4, 9 and 10 were anchored only one, one and two sequences respectively, which corroborates data from classical cytogenetics indicating that these areas reside in the vicinity of heterochromatic regions (Cheng *et al.* 2001). The distribution of clusters in regions highlighted by classical cytogenetics as heterochromatin, probably indicate the presence ofeuchromatic "islands" throughout the chromosome, closely related to high levels of expression (Yasuhara and Wakimoto 2006).

Some clusters presented association with centromeric regions or nearby existing repeats, as occurred in chromosomes 1, 2, 3, 5, 6, 7, 9 and 11. The presence of these clusters in regions of low gene expression activity is probably due to the alternation of euchromatin and heterochromatin around the centromere (Yan and Jiang 2007), another point to consider is the probability of modifications in histones H3 and H4, which allow genes transcription in this area. Also, previous works have shown that gene transcription may occur near the centromeres of rice chromosomes 8 and 3, so at least some elements could be transcribed in this area (Yan *et al*. 2006).

Regarding the number of sugarcane sequences similar to rice genome per chromosome, it was observed that chromosomes 1 and 3 presented the highest number of anchored sequences,

matching 133 and 124 respectively, while the chromosome 10 presented the lowest number of aligned sugarcane sequences (**Fig. 5**).

Twelve clusters presented similarities with distinct segments in the same chromosome, being considered as duplications, as shown in Table 4 and also indicated by color dots in Figure 5. Those duplicated copies tend to diverge due to mutations and may specialize or optimize to play slightly different roles (Alberts *et al*. 1997).In rice one duplication event is assumed for each 20 million years in regard to its supposed ancestor, including the previous separation of cereals such as sorghum, maize and the Triticeae (Thiel *et al.*2009). Regarding the duplicated segments considering the entire genome, 117 clusters could be identified in at least two distinct chromosomes. Different of what was cited before, repetitions in distinct chromosomes resulted from events of duplication, accidental translocations and sequence divergence, allowing functional diversification (Wendell 2000; Thiel *et al.* 2009). There are also evidences that transpositions outbreaks could be activated by severe environmental biotic or abiotic stress (Levin and Moran 2011).

Regarding synteny evidences, it was observed that chromosome 1 shares more similar regions with chromosome 5; chromosome 2 with chromosomes 4 and 6; chromosome 7 with 1 and 3; chromosome 9 with 8 and 4 and chromosome 10 presented synteny only with chromosome 3. On the telomeric region of chromosome 3 short arm a grouping of sequences could be observed, showing synteny with chromosome 10. The same occurred when the telomeric region of chromosome 11 short arm was analyzed, presenting a cluster of sequences showing synteny with the telomeric region of the chromosome 12 short arm.

Still regarding the analysis of duplication events, a large in tandem repetition was evident in the long arm of chromosome 9, represented by the groups GR005 to GR011. Previous

reports suggest that once duplicated, in tandem repetitions may extend rapidly through events of unequal crossing over, what could confer some advantage (Alberts *et al*. 1997), in the present case, a higher diversity of genes associated to resistance and stress responses.

A remarkable degree of genome conservation has been established in comparative genetic mapping experiments for the Poaceae family, although genome sizes vary as much as 40-fold between some of the species, and despite the fact that they diverged as long as 60 million years ago (Gale and Devos 1998). Genetic mapping experiments in allohexaploid wheat revealed that most gene sequences are triplicated on the A, B and D genomes. Furthermore, the three sets of the seven homeologous chromosomes show overall colinearity. Evidence of a few translocation events was, however, also found (Devos and Gale 1993). Within the Triticeae tribe, extensive colinearity was established, for example, for the homeologous chromosomes of wheat, as compared with*Triticum monococcum*, *Triticum tauschii* and *H. vulgare*, and consensus maps were developed (Van Deynze *et al.*1995; Dubcovsky *et al*. 1996). Microsynteny and colinearity were also observed in other angiosperms when compared to other distant related organisms, as it was the case of regions rich in factors associated with pathogen response chickpea and arabidopsis (Benko-Iseppon *et al.*2003)

Considering the small size of the rice genome as compared with sugarcane, it is clear that higher levels of redundancy are expected in the sugarcane genome. Despite of that, the present evaluation may be valuable for the use of the identified genes for sugarcane breeding, since they may indicate putative linked gene markers for mapping purposes in sugarcane, especially considering most clustered regions here identified.

**SuperSAGE tags BLAST against rice chromosomes**

As result of drought and salinity superSAGE experiments, 76 tags matching resistance genes were upregulated when compared to non-stress conditions and 79 *PR* genes superSAGE tags candidates equally obtained from experiments submitted to abiotic stress were identified as upregulated. A considerable number of tags exposed to abiotic stress as drought and salinity aligned with *R* and *PR* genes including 115 sugarcane sequences that appeared as upregulated if compared to analysis in non-stressed conditions. It is known that both pathogen attack and abiotic stress may trigger a diverse array of plant defense-related genes involved in HR, which is characterized by necrotic lesions resulting from localized host cell death at the site of infection and also activating defense responses in uninfected parts of the plant, expressing so called SAR (Wang *et al*. 2010).

Considering the SuperSAGE output, 63% of *R*-genes candidates (28 tags) presented high similarities with rice genome segments. Matches in all chromosomes could be annotated, although chromosomes 1, 2 and 10 were those that presented higher number of anchored superSAGE tags. In the same way, 78.48% of *PR*-gene candidates (62 tags) matched alignments to rice genome regions, uncovering chromosomes 1, 3 and 6 as the most represented. Although most sequences anchored in rice chromosomes, 37% of the tags regarding *R*-genes and 21.6% of the *PR* ones did not present alignment with any region of the rice genome.They probably represent regions resulted from duplication events followed by divergent evolution (mutations), that are common in large genomes and may be a source of new genetic products that share common domains but may have distinct functions (Alberts *et al.* 1997).It is interesting also to consider that the SuperSAGE tags are expected to anchor at the 3'UTR region that are often outside the most conserved gene regions, maybe also bearing

most frequent establishment of new mutations, when compared with conserved domains or folding regions of the gene.

In contrast, the chromosome 7 presented the lower number of alignments, matching only one *R*-gene and one *PR*-gene tag. Moreover, the search for similarities in the rice genome showed that four *PR* superSAGE tags could be identified anchoring in more than one chromosome. Both tags Sg196058 and Sg181942 presented similarities with chromosomes 1 and 3, while the tag Sg180925 aligned with chromosomes 3, 9 and 11. Finally, the tag Sg218368 performed matches with both chromosomes 2 and 11, also here suggesting the occurrence of duplication and translocation events along the genome.

**CONCLUSIONS**

- The present approach using full length rice reference cDNA sequences permitted the successful identification of 1,460 sugarcane genes associated with the response to pathogen attack.
- The sugarcane transcriptome includes all procured gene categories of KEGG plant-pathogen interaction pathway, unraveling a high abundance of genes associated with HR and SAR, as well as *R* and *PR* genes.
- Most of the identified sequences (74%) presented best matches with *Sorghum bicolor*, followed by *Zea mays* (15%) and *Oryza sativa* (9%) reflecting their taxonomic relationship and also indicating the potential for transferring gene markers from sorghum to sugarcane for mapping purposes.
- Considering the *PR* gene categories identified, most representatives regarded the *PR-9* (peroxidases class), that contributes to plant disease resistance through the deposition of

lignin, conferring resistance against a broad spectrum of pathogens. Other classes (*PR-2, PR5, PR-7*, and *PR-14*) were also well represented, indicating a relative genetic diversity and abundance regarding this category.

- The NBS-LRR *R*-gene class was the most representative, with all families presented matches in sugarcane. Probably the studied sequences represent only part of the diversity and number of *R*-genes that are present in the cultivated sugarcane varieties, especially considering the huge size and redundancy of the sugarcane genome, as compared with most angiosperms, and also the complexity of the epigenetic processes.
- Most of the identified sugarcane sequences matched to rice chromosomes 1 and 3 (133 and 124 respectively), also presenting clustered regions, as expected especially for *R* genes that emerged by duplication events.
- A high number of sequences associated with response to pathogen attack in sugarcane is also active under abiotic stress, especially during drought, confirming observations regarding cross-talk among genes of distinct stress categories. Considering their relative distribution within the rice genome, regions comprising over- and down-expressed SuperSAGE tags are not distributed randomly, presenting consistent co-expression also considering different experimental conditions and stress types.
- The identified sequences represent valuable sources for the sugarcane breeding program, allowing their use in biotechnological approaches, with emphasis on transgene experiments. They are also valuable for mapping purposes, especially considering their putative distribution here uncovered when considering available distribution of genes known from the rice genome. Multifunction genes that are important in pathogen response as well as during abiotic stress in tolerant sugarcane cultivars – here identified by the high throughputSuperSAGE assay – are

most important candidates for approaches aiming to confer multiple (biotic and abiotic) resistance in sugarcane, a very important strategy considering the actual climate changing scenario.

**ACKNOWLEDGEMENTS**

This study has been supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil). We also thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil), FINEP (Financiadora de Estudos e Projetos) and FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco) for the concession of fellowships.

**REFERENCES**

**Altschul SF, Gish W, Miller W, Myers E, Lipman DJ** (1990) Basic local alignment search tool. *Journal Molecular Biology* **215**, 403-410

**Apel K, Hirt H** (2004) Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annual Review of Plant Biology* **55**, 373-399

**Asselbergh B, De Vleesschauwer D, Höfte M** (2008) Global switches and finetuning – ABA modulates plant pathogen defense. *Molecular Plant-Microbe Interact*ions **6**, 709-719

**Atici O, Nalbantoglu B** (2003) Antifreeze proteins in higher plants. *Phytochemistry* **64,** 1187–1196

**Barbosa da Silva A, Wanderley-Nogueira AC, Silva RRM, Belarmino LC, Soares-Cavalcanti NM, Benko-Iseppon AM** (2005) *In silico* survey of resistance (R) genes in *Eucalyptus* transcriptome. *Genetics and Molecular Biology* **28**, 562-574

**Békésiová B, Hraska S, Libantová J, Moravcíková J, Matusíková I** (2007) Heavy-metal stress induced accumulation of chitinase isoforms in plants.*Molecular Biology Reports* **35**, 579-588

**Benedetti CE, Xie D, Turner JG** (1995) COI1-dependent expression of an *Arabidopsis* vegetative storage protein in flowers and siliques and in response to methyl jasmonate. *Plant Physiology* **109**, 567-572

**Benko-Iseppon AM, Winter P, Huettel B, Staginnus C, Muehlbauer FJ, Kahl G** (2003) Molecular markers closely linked to fusarium resistance genes in chickpea show significant alignments to pathogenesis-related genes located on *Arabidopsis* chromosomes 1 and 5. *Theoretical and Applied Genetics* **107**, 379-386

**Benko-Iseppon, AM, Galdino SL, Calsa-Júnior T, Kido EA, Tossi A, Belarmino LC, Crovella S** (2010). Overview of plant antimicrobial peptides.*Current Protein & Peptide Science* **11**, 181-188

**Bent AF** (1996) Plant disease resistance genes: function meets structure.*Plant Cell* **8**, 1751-1771

**Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, Crute IR** (1985) The genetic basis of relationships between microbial parasites and their hosts. In: Fraser RRS (Ed) *Mechanisms of Resistance to Plant Diseases*. Dordrecht, Netherland, pp 80-142

**Berna A, Bernier F** (1999) Regulation by biotic and abiotic stress of a wheat germin gene encoding oxalate oxidase, a H2O2-producing enzyme.*Plant Molecular Biology* **39**, 539-49

**Bonasera JM, Kim JF, Beer SV** (2006) PR genes of apple: identification and expression in response to elicitors and inoculation with *Erwinia amylovora. BMC Plant Biology* **6**, 23-34

**Bouchenak-Khelladi Y, Salamin N, Savolainen V, Forest F, van der Bank M, Chase MW, Hodkinson TR** (2008) Large multi-gene phylogenetic trees of the grasses (Poaceae): Progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution* **47**, 488-505

**Broekaert WF, Terras FRG, Cammue BPA** (2000) Induced and preformed antimicrobial proteins. In: Slusarenko AJ, Fraser RSS and van Loon LC (Eds.) *Mechanisms of Resistance to Plant Diseases*, Kluwer Academic Publishers, London, pp 371-477

**Brueggeman R, Rostoks N, Kudrna D, Kilian A, Han F, Chen J, Druka A, Steffenson B,Kleinhofs A** (2008)The barley stem rust-resistance gene Rpg1 is a noveldisease-resistance gene with homology toreceptor kinases. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 9328-9333

**CETENE - Centro de Tecnologias Estratégicas do Nordeste**; available on line at: http://www.cetene.gov.br

**Chester KS** (1933) The problem of acquired physiological immunity in Plants. *Quarterly Review of Phytopathology* **42**, 185-209

**Cheng Z, Buell CR, Wing RA, Gu M, Jiang J** (2001) Toward a cytological characterization of the rice genome. *Genome Research* **11**, 2133-2141

**Chini A, Grant JJ, Seki M, Shinozaki K, Loake GJ** (2004) Drought tolerance established by enhanced expression of the CC-NBS-LRR gene, ADR1, requires salicylic acid, EDS1 and ABI1. *The Plant Journal* **38**, 810-822

**Chinnaiyan AM, Chaudhary D, O'Rourke K, Koonin E**(1997) Role of CED-4 in the activation of CED-3. *Nature* **388,**728-729

**Cooley MB, Pathirana S, Wu HJ, Kachroo P, Klessig DF (**2000) Members of the *Arabidopsis* HRT/RPP8 family of resistance genes confer resistance to both viral and oomycete pathogens.*Plant Cell* **12**, 663-676

**Copley R, Letunic I, Bork P (**2001) Genome and protein evolution in eukaryotes.*Current Opinion Chemistry Biology* **268**, 78-94

**Costa MLM, Amorim LLB, Onofre AVC, Melo LJOT, Oliveira MBM, Carvalho R, Benko-Iseppon AM** (2011). Assessment of genetic diversity in contrasting sugarcane varieties using inter-simple sequence repeat (ISSR) Markers. *American Journal of Plant Sciences* **2**, 425-432

**CTC - The Center of Sugarcane Technology**; available on line at: http://www.ctcanavieira.com.br

**Davletova S, Rizhsky L, Liang H, Shengqiang Z, Oliver DJ, Coutu J, Shulaev V, Schlauch K, Mittler R** (2004) Cytosolic ascorbate peroxidase 1 is a central component of the reactive oxygen gene network of *Arabidopsis*. *Plant Cell* **17**, 268-281

**Dinesh-Kumar S, Whitham S, Choi D, Hehl R** (1995) Transposon tagging of tobacco mosaic virus resistance gene N: its possible role in the TMV-N-mediated signal transduction pathway. *Proceedings of National Academy ofScience* **92**, 4175-4180

**Dixon MS, Jones DA, Keddie JS, Thomas CT, Harrison K, Jones JDG** (1996) The tomato *Cf2* disease resistance locus comprises two functional genes encoding leucine rich repeats proteins. *Cell* **84**, 451-459

**Devos KM, Gale MD** (1993) The genetic maps of wheat and their potential in plant breeding. *Outlook on Agriculture* **22**, 93-99

**Divi UK, Rahman T, Krishna P** (2010) Brassinosteroid-mediated stress tolerance in Arabidopsis shows interactions with abscisic acid, ethylene and salicylic acid pathways.*BiomedCentral Plant Biology* **10**,151.

**Dubcovsky J, Luo M-C, Zhong G-Y, Bransteitter R, Desai A, Kilian A, Kleinhofs A, Dvoøák J** (1996) Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L. *Genetics* **143**, 983-999

**Durrant WE, Dong X** (2004) Systemic acquired resistance. *Annual Review of Plant Pathology* **42**, 185-209

**Ebel J, Cosio EG** (1994) Elicitors of plant defense responses. *International Review of Cytology* **148**, 1-36

**Ellis J, Jones D** (2000) Structure and function of proteins controlling strain-specific pathogen resistance in plants. *CurrentOpinion in Plant Biology* **1**, 288-293

**Ellis JG, Lawrence GJ, Finnegan EJ, Anderson PA** (1995) Contrasting complexity of two rust resistance loci in flax. *Proceedingsof National Academy ofScience* **92**, 4185-4188

**EXPASY -Expert Protein Analysis System**; available on line at:http://us.expasy.org/

**FAO - Food and Agriculture Organization** (2005) Available online: http://faostat.fao.org/faostat/collections?subset=agriculture.

**Fraire-Velázquez S, Rodríguez-Guerra R, Sánchez-Calderón L** (2011) Abiotic and Biotic Stress Response Crosstalk in Plants. In: Shanker A, Venkateswarlu B (Eds) Abiotic Stress Response in Plants - *Physiological, Biochemical and Genetic Perspectives*. Rijeka, Croatia, pp 3-26

**Gaffney T, Friedrich L, Vernooij B, Negrotto D, Nye G Ukness S, Ward E, Kessman H, Ryals J** (1993) Requirement of salicylic acid for the induction of systemic acquired resistance. *Science* **261**, 754-756

**Gale M, Devos KM** (1998) Plant comparative genetics after 10 years. *Science* **282**, 656-659

**García-Olmedo F, Molina A, Segura A, Moreno M** (1995) The defensive role of nonspecific lipid-transfer proteins in plants. *Trends in Microbiology* **3**, 72-74

**GenXPro - Available on line at:**http://www.genxpro.info/

**Goldman MHS, Goldman GH (1998)***Trichoderma harzianum*transformant has high extracellular alkaline proteinase expression during specific mycoparasitic interactions. *Genetics and Molecular Biology* **21**, 329-333

**GRAMENE;** available on line at: http://www.gramene.org

**Grivet L, Arruda P** (2001) Sugarcane Genomics: depicting the complex genome of an important tropical crop. *Current Opinion in Plant Biology* **5**, 122-127

**Griffith M, Yaish MWF** (2004) Antifreeze proteins in overwintering plants: a tale of two activities. *Trends in Plant Science* **9**, 399-405

**Hammond-Kosack KE, Jones JDG** (1997) Plant disease resistance genes. *Annual Review in Plant Physiology* **48**, 575-607

**Hincha DK, Meins Jr F, Schmitt JM** (1997) [beta]-1,3-Glucanase is cryoprotective *in vitro* and is accumulated in leaves during cold acclimation. *Plant Physiology* **114**, 1077-1083

**Hon WC, Griffith M, Mlynarz A, Kwok YC, Yang DSC** (1995) Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. *Plant Physiology* **109**, 879-889

**Hu X, Li W, Chen Q, Yang Y** (2009) Early signal transduction linking the synthesis of jasmonic acid in plant.*Plant Signal Behavior* **4**, 696-697

**Humphry M, Bednarek P, Kemmerling B, Koh S, Stein M, Göbel U, Stüber K, Pislewska-Bednarek M, Loraine A, Schulze-Lefert P, Somerville S, Panstruga R** (2010) A region conserved in monocot and dicot plants defines a functional module in antifungal plant immunity.*Proceedings of the National Academy of Sciences of United States* **107**, 21896-21901

**Ingle R, Carstens M, Denby K** (2006) PAMP recognition and the plant-pathogen arms race.*BioEssays***28**, 880-889

**Joahal GS, Briggs SP** (1992) Reductase activity encodes by the *Hm1* resistance gene in maize.*Science***198**, 985-987

**Jones DA, Thomas CM, Hammond-Kosack KE, Balint-Kurti PJ, Jones JDG** (1994) Isolation of tomato *Cf9* gene for resistance to *Cladosporium fulvum* by transposons tagging. *Science* **266**, 789-793

**Jones J, Dangl J**(2006) The plant immune system. *Nature* **444**, 323-329

**Jung YH, Lee JH, Agrawal GK, Rakwal R, Kim JA, Shim JK, Lee SK, Jeon JS, Koh HJ, Lee YH, Iwahashi H, Jwa NS** (2005) The rice (*Oryza sativa*) blast lesion mimic mutant, blm, may confer resistance to blast pathogens by triggering multiple defense-associated signaling pathways.*Plant Physiology and Biochemistry* **43**, 397-406

**Kaliff M, Staal J, Myrenas M, Dixelius C** (2007) ABA is required for *Leptosphaeria maculans* resistance via ABI1- and ABI4-dependent signaling. *Molecular and Plant-Microbe Interactions* **20**, 335-345

**Kanazin V, Marek LF, Shoemaker RC** (1996) Resistance gene analogs are conserved and clustered in soybean. *Proceedings of National Academy of Science* **93**, 11746-11750

**KEGG -** Kyoto Encyclopedia of Genes and Genomes; available on line at: http://www.genome.jp/kegg-bin/show_pathway?map04626

**Kepinski S, Leyser O** (2005) Plant development: auxin in loops.*Current Biology* **15**, 208-210

**Kido EA, Pandolfi V, Houllou-Kido LM, Andrade PP, Marcelino FC, Nepomuceno AL, Abdelnoor RV, Burnquist WL, Benko-Iseppon AM** (2010) Plant antimicrobial peptides: an overview of SuperSAGE transcriptional profile and a functional review. *Current Protein and Peptide Science* **11**, 220-230

**Kido EA, Barbosa PK, Ferreira Neto JCR, Pandolfi V, Houllou-Kido LM, Crovella S, Benko-Iseppon AM** (2011) Identification of plant protein kinases in response to abiotic and biotic stresses using SuperSAGE. *Current Protein and Peptide Science* **12**,643-656

**Kondrashov F, Rogozin I, Wolf Y, Koonin E** (2002) Selection in the evolution of gene duplication.*Genome Biology* **3**, 1-9

**Lawrence GJ, Finnegan EJ, Ayliffe MA, Ellis JG** (1995) The *L6* gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene *RPS2* and the tobacco viral resistance gene N.*Plant Cell* **7**, 1195-1206

**Leubner-Metzger G** (2005) β-1,3-glucanase gene expression in low-hydrated seeds as a mechanism for dormancy release during tobacco after-ripening. *Plant Journal* **41**, 133-145

**Leubner-Metzger G, Meins JF** (2000) Sense transformation reveals a novel role for class I β-1-3-glucanase in tobacco seed germination. *Plant Journal* **23**, 215-221

**Levin HL, Moran JV** (2011) Dynamic interactions between transposable elements and their hosts.*Nature Reviews in Genetics* **12**, 615-627

**Liu B, Zhang S, Zhu X, Yang Q, Wu S, Mei M, Mauleon R, Leach J, Mew T, Leung H** (2004) Candidate defense genes as predictors of quantitative blast resistance in rice. *Molecular Plant-Microbe Interactions* **17**, 1146-1152

**Liu J, Liu X, Dai L and Wang G** (2007) Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants .*Journal of Genetics and Genomics* **34**, 765-776

**Lópes MA, Bannenberg G, Castresana C** (2008) Controlling hormone signaling is a plant and pathogen challenge for growth and survival. *Current Opinion in Plant Biology* **11**, 420-427

**Lu H** (2009) Dissection of salicylic acid-mediated defense signaling networks. *Plant Signal Behavior* **8**,713-717

**Maisonneuve B, Bellec Y, Anderson P, Michelmore RW** (1994) Rapid mapping of 2 genes for resistance to downy mildew from *Lactuca serriola* to existing clusters of resistance genes.*Theoretical and Applied Genetics* **89**, 96-10

**Manandhar HK, Mathur SB, Smedegaard-Petersen, Thordal-Christensen H** (1999) Accumulation of transcripts for pathogenesis-related proteins and peroxidase in rice plants triggered by *Pyricularia oryzae*, *Bipolaris sorokiniana* and U.V. light. *Physiological and Molecular Plant Pathology* **55**, 289-295

**Martin GB, Brommonschenkel SH, Chunwongster J, Frary A, Ganal MW, Spivey R, Wu T Earle ED, Tanksley SD** (1993) Map based cloning of a protein kinase gene conferring disease resistance and tomato.*Science* **262**, 1432-1436

**Melotto M, Coelho MF, Pedrosa-Harand A, Kelly JD, Camargo LE** (2004)The anthracnose resistance locus Co-4 of common bean is located on chromosome 3 and contains putative disease resistance-related genes.*Theorethical and Applied Genetics* **109**, 690-699

**Metzler MC, Cutt JR, Klessig DF** (1991) Isolation and characterization of a gene encoding a PR-1 like protein from *Arabidopsisthaliana*. *Plant Physiology* **96**, 346-348

**Mindrinos M, Katagiri F, Yu GL, Ausubel FM** (1994) The *Arabidopsis thaliana* disease resistance gene encodes a protein containing a nucleotide-binding site and leucine rich repeats.*Cell* **78**, 1089-1099

**Mofet P** (2009) Mechanisms of recognition in dominant *R* gene mediated resistance.*Advances in Virus Research* **75**, 1-33

**Molina C, Rotter B, Horres R, Udupa SM, Besse B, Bellarmino L, Baum M, Matsumura H, Terauchi R, Kahl G, Winter P** (2008) The Drought stress responsive transcriptome of chickpea roots. *BioMed Central Genomics* **9**, 553

**MSU Rice Genome Annotation Project**; available on line at: http://rice.plantbiology.msu.edu/

**Munis MF, Tu L, Deng F, Tan J, Xu L, Xu S, Long L, Zhang X** (2010) A thaumatin-like protein gene involved in cotton fiber secondary cell wall development enhances resistance against *Verticillium dahliae* and other stresses in transgenic tobacco. *Biochemistry and Biophysics Research Community* **393**, 38-44

**Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, Voinnet O, Jones JD** (2006) A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science* **312**, 436-439

**ONSA -** Organization for Nucleotide Sequencing and Analysis; available on line at: http://watson.fapesp.br/onsa/Genoma3.htm

**Park CJ, Han SW, Chen X, Ronald PC** (2010)Elucidation of XA21-mediated innate immunity.*Cell Microbiology* **12**, 1017-1025

**Parida SK, Pandit A, Gaikwad K, Sharma TR, Srivastava PS, SinghNK,Mohapatra T (**2010) Functionally relevant microsatellites in sugarcane unigenes. BMC Plant Biology **10**, 251

**Penninckx IA, Thomma BP, Buchala A, Metraux JP, Broekaert WF** (1998) Concomitant activation of jasmonate and ethylene response pathways is required for induction of a plant defensin gene in *Arabidopsis*. *Plant Cell* **8**, 2309-2323

**Perrin RM, Wigge PA** (2002) Genome studies and molecular genetics - plant biotechnology web alert. *Current Opinion in Plant Biology* **5**, 89-90

**Pinzón A, Barreto E, Bernal A, Luke A, Barrios AF, Isea R, Restrepo S** (2009) Computational models in plant-pathogen interactions: the case of *Phytophthora infestans*. *Theoretical Biology & Medical Modelling* **6**, 24

**Rayapati PJ, Lee M, Gregory JW, Wise RP** (1994) A linkage map of diploid Avena based on RFLP loci and a locus conferring resistance to nine isolates of *Puccinia coronata* var. avenae. *Theoretical Applied Genetics* **89**, 831-837

**Robert-Seilaniantz A, Navarro L, Bari R, Jones JDG** (2007) Pathological hormone imbalances. *Current Opinion in Plant Biology* **10**, 372-379

**Sarowar S, Kim YJ, Kim KD, Hwang BK, Ok SH, Shin JS** (2009) Overexpression of lipid transfer protein (LTP) genes enhances resistance to plant pathogens and LTP functions in long-distance systemic signaling in tobacco. *Plant Cell Report* **28**, 419-427

**Salmeron JM, Oldroyd GED, Romens CMT, Scofield SR, Kim HS, Lavelle DT, Dahlbeck D, Staskawicz BJ** (1996) Tomato *Prf* is a member of the leucine rich repeats class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster.*Cell* **86**, 123-133

**Schenke D, Böttcher C, Scheel D** (2011) Crosstalk between abiotic ultraviolet-B stress and biotic (flg22) stress signalling in *Arabidopsis* prevents flavonol accumulation in favor of pathogen defence compound production. *Plant, Cell & Environment* **34**, 1849-1864

**Seo P J, Lee A-K, Xiang F, Park C-M** (2008) Molecular and Functional Profiling of Arabidopsis Pathogenesis-RelatedGenes: Insights into Their Roles in Salt Response of Seed Germination.*Plant Cell Physiology* **49**, 334-344

**Shi F, Yamamoto R, Shimamura S, Hiraga S, Nakayama N, Nakamura T, Yukawa K, Hachinohe M, Matsumoto H, Komatsu S** (2008) Cytosolic ascorbate peroxidase 2 (cAPX 2) is involved in the soybean response to flooding. *Phytochemistry***69,** 1295-1303

**Shiu SH, Bleecker AB** (2001) Plant receptor-like kinase gene family: diversity, function, and signaling.*Science STKE***113**, Review

**Silva RLO, Ferreira Neto JCR, Pandolfi V, Chabregas SM, Burnquist WL, Benko-Iseppon AM, Kido EA** (2011) Transcriptomics of Sugarcane Osmoprotectants Under Drought. In: Vasanthaiah HKN, Kambiranda D (Eds) *Plants and Environment*. Rijeka, Croatia, pp 89-114

**Singh NK, Bracker CA, Hasegawa PM, Handa AK, Buckel S, Hermodson MA, Pfankoch E, Regnier FE, Bressan RA** (1987) Characterization of osmotin. A thaumatin-like protein associated with osmotic adaptations in plant cells. *Plant Physiology* **85**, 529-536

**Song WY,Wang GL, Kim HS,Pi LY, Gardner J, Wang B, Holsten T, Zhai WX, Zhu LH, Fauquet C, Ronald PC** (1995) A receptor kinase-like protein encoded by the rice disease resistance gene *Xa21*. *Science* **270**, 1804-1806

**Song WY, Pi LY, Wang GL, Gardner J, Holsten T, Ronald PC** (1997) Evolution of the rice *Xa21* disease resistance genes family. *Plant Cell* **9**, 1279-1287

**Sparla F, Rotino L, Valgimigli MC, Pupillo P, Trost P** (2004) Systemic resistance induced by benzothisdizole in pear inoculated with the agent of fire blight. *Scientia Horticulturae* **101**, 296-279

**Spoel S, Johnson J, Dong X** (2007) Regulation of tradeoffs between plant defenses against pathogens with different lifestyles.*Proceedings of National Academy of Science* **104**, 18842-18847

**SUCEST -** Sugarcane Expressed Sequence Tag Project; available on line at: http://sucest.lad.dcc.unicamp.br/en/

**Tang X, Xie M, Kim YJ, Zhou J, Klessing DF, Martin GB** (1999) Overexpression of *Pto* activates defense responses and confers broad resistance. *Plant Cell* **11**, 15-29

**Thiel T, Graner A, Waugh R, Grosse I, Close TJ, Stein N** (2009) Evidence and ecolutionary of ancient whole-genome duplication in barley predating the divergence from rice. *BioMed Central Evolutionary Biology* **9**, 209-227

**Thomma BP, Eggermont K, Penninckx IA, Mauch-Mani B, Vogelsang R, Cammue BP, Broekaert WF** (1998) Separate jasmonate-dependent and salicylate-dependent defense-

response pathways in *Arabidopsis* are essential for resistance to distinct microbial pathogens. *Proceedings of National Academy ofScience* **95**, 15107-15111

**Thomma BPHJ, Penninckx IAMA, Broekaert WF, Cammue BPA** (2001) The complexity of disease signaling in *Arabidopsis*. *Current Opinion in Immunology* **13**, 63-68

**Tornero P, Gadea J, Conejero V, Vera P** (1997) Two *PR-1* genes from tomato are differentially regulated and reveal a novel mode of expression for a pathogenesis-related gene during the hypersensitive response and development. *Plant and Microbe Interactions* **10**, 624-634

**Van Deynze AE, Nelson JC, Yglesias ES, Harrington SE, Braga DP, McCouch SR Sorrells ME** (1995) Comparative mapping in grasses. Wheat relationships.*Molecular Genome and Genetics* **248**, 744-754

**Van Loon LC, Van Strien EA** (1999) The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 proteins. *Physiological and Molecular Plant Pathology* **55**, 85–97

**Van Loon LC, Pierpoint WS, Boller Th, Conejero** (1999) Recommendations for naming plant pathogenesis-related proteins.*Plant Molecular Biology Report* **12**, 245-264

**Van Loon LC, Geraats BPJ, Linthorst HJM** (2006) Ethylene as a modulator of disease resistance in plants. *Trends in Plant Science* **11**, 184-191

**Velazhahan R, Muthukrishnan S** (2003) Transgenic tobacco plants constitutively overexpressing a rice thaumatin-like protein (PR-5) show enhanced resistance to *Alternaria alternata*. *Plant Biology* **47**, 347-354

**Vigers AJ, Roberts WK, Selitrennikof CP** (1991) A new family of plant antifungal proteins.*Molecular Plant-Microbe Interactions* **4**, 315-323

**Vincentz M, Cara AAF, Okura VK, da Silva FR, Pedrosa GL, Hemerly AS, Capella NA, Marins M, Fereira PC, França SC, Grivet L, Vettore AL, Kempre EL, Burnquist WL, Targon LPM, Siquiera WJ, Kuramae EE, Marino CL, Camargo LEA, Carrer H, Coutinho LL, Furlan LR, Lemos MVF, Nunes LR, Gomes SL, Santelli RV, Goldman MH, Bacci M, Gigliotii EA, Thiermann OH, Silva FH, Van Sluys M, Nobrega FG, Arruda P, Menck CFM** (2004) Evaluation of monocot and eudicot divergence using the sugarcane transcriptome. *Plant Physiology* **134**, 951-959

**Vlot AC, Klessig DF, Park SW** (2008) Systemic acquired resistance: the elusive signal(s). *Current Opinion in Plant Biology* **11**, 436-442

**Wanderley-Nogueira AC, Mota N, Lima-Morais D, Silva LCB, Silva AB, Benko-Iseppon AM** (2007) Abundance and diversity of resistance (*R*) genes in the sugarcane transcriptome. *Genetics and Molecular Research* **6**, 866-889

**Wang GL, Holsten TE, Song WY, Wang HP, Ronald PC** (1995) Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa21* disease resistance locus.*The PlantJournal* **7**, 525-533

**Wang D, Weaver ND, Kesarwani M, Dong X**(2005) Induction of protein secretory pathway is required for systemic acquired resistance. *Science* **308**,1036-1040

**Wang X , Liu W , ChenX, Tang C, Dong Y, Ma J , Huang X , Wei G, Han Q, Huang L, Kang Z** (2010) Differential gene expression in incompatible interaction between wheat and stripe rust fungus revealed by cDNA-AFLP and comparison to compatible interaction.*BMC Plant Biology* **10**, 9

**Watson L, Dallwitz MJ** (1992) *The Grass Genera of the World*. Wallingford, Oxon, United Kingdom: CAB International

**Wendell J** (2000) Genome evolution in polyploids.*Plant Molecular Biology* **42**, 225-249

**Whitham S, Dinesh-Kumar SP, Choi D, Hehl R, Corr C, Baker B**(1994) The product of tobacco mosaic virus resistance gene *N*: similarity to toll and the interleukin-1 receptor. *Cell* **78**, 1101-1115

**Wilkstrom N, Savolainen V, Chase MW** (2001) Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society Biological Sciences* **268**, 2211-2220

**Woo H, Rhee SG** (2010) Immunoblot detection of proteins that contain cysteine sulfinic or sulfonic acids with antibodies specific for the hyperoxidized cysteine-containing sequence. In: Das D (ed) *Methods in Redox Signaling*, Mary Ann Liebert Editors, Inc, pp. 19-23. Available Online at: http://www.liebertonline.com/doi/pdf/10.1089/9781934854068.19

**Wrzaczek M, Brosché M, Salojärvi J, Kangasjärvi S, Idänheimo N, Mersmann S, Robatzek S, Karpiński S, Karpińska B, Kangasjärvi J** (2010) Transcriptional regulation of the CRK/DUF26 group of receptor-like protein kinases by ozone and plant hormones in *Arabidopsis.BMC Plant Biology* **10**, 95

**Yan H, Ito H, Nobuta K, Ouyang S, Jin W, Tian S, Lu C, Veno RC, Wang GL, Green PJ, Wing RA, Buell CR, Meyers BC, Jiang J** (2006) Genomic and genetic characterization of rice Cen3 reveals extensive transcription and evolutionary implications of complex centromere. *The Plant Cell* **18**, 2123-2133

**Yan H, Jiang J** (2007) Rice as model for centromere and heterochromatin research. *Chromosome Research* **15**, 77-84

**Yasuhara JC, Wakimoto BT** (2006) Oxymoron no more: the expanding world of heterochromatic genes. *Trends in Genetics* **22**, 330-338

**Zeier J, Pink B, Mueller MJ, Berger S** (2004) Light conditions influence specific defense responses in incompatible plant-pathogen interactions: uncoupling systemic resistance from salicylic acid and *PR-1* accumulation. *Planta* **219**, 673-683

**Zhu QE, Maher S, Masoud R, Dixon R, Lamb CJ** (1994) Enhanced Protection against fungal attack by constitutive coexpression of chitinase and glucanase genes in transgenic tobacco. *Biotechnology* **12**, 807-812

Table 1:Sugarcane clusters matching results to each gene of KEGG pathway. Showing number of matches to each seed sequences, number of clusters that aligned exclusively with corresponding gene, e-value, score, size in nucleotide (nt) and aminoacids (aa) and presence of conserved domains. Abbreviations: LRR=Leucine-Rich Repeat; CP=Complete; Fam=Family; IC=Incomplete; Kin=Kinase; Prot=Protein.

| KEGG Pathway | | Sugarcane | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Class** | **Gene** | **Matches up to e-5** | **Exclusive clusters** | **Best match** | **e-value** | **Score** | **Size (nt)** | **Size (aa)** | **Domain complete=CP incomplete=IC** |
| **PR** | *PR1* | 3 | 4 | lcl\|CL2300-Ct1 | 3 $e^{88}$ | 322 | 1178 | 240 | - |
| **R** | *Cf-2* | 110 | 0 | lcl\|CL587-Ct3 | 1 $e^{75}$ | 177 | 2208 | 714 | LRR (CP) |
| | *Cf-5* | 116 | 1 | lcl\|CL10869-Ct1 | 9 $e^{105}$ | 205 | 994 | 311 | LRR (CP) |
| | *Cf-4* | 105 | 1 | lcl\|CL587-Ct3 | 4 $e^{56}$ | 122 | 2208 | 714 | LRR (CP) |
| | *Cf-9* | 96 | 0 | lcl\|CL587-Ct3 | 8 $e^{36}$ | 84.9 | 2208 | 714 | LRR (CP) |
| | *EFR* | 249 | 0 | lcl\|CL587-Ct3 | 4 $e^{85}$ | 150 | 2208 | 714 | LRR (CP) |
| | *FLS2* | 250 | 0 | lcl\|CL587-Ct3 | 1 $e^{102}$ | 207 | 2208 | 714 | LRR (CP) |
| | *Xa21* | 248 | 0 | lcl\|CL587-Ct3 | 1 $e^{76}$ | 89.0 | 2208 | 714 | LRR (CP) |
| | *Pto* | 248 | 7 | lcl\|CL3705-Ct1 | 0.0 | 623 | 1656 | 449 | Ser-Thr Kin. (CP) |
| | *PRF* | 40 | 6 | lcl\|CL11633-Ct1 | 3 $e^{112}$ | 322 | 948 | 294 | - |
| | *WRKY25* | 41 | 8 | lcl\|CL3534-Ct1 | 7 $e^{68}$ | 165 | 1424 | 279 | WRKY (CP) |
| | *WRKY29* | 31 | 0 | lcl\|CL4442-Ct1 | 4 $e^{33}$ | 140 | 784 | 108 | - |
| | *Pti1* | 249 | 7 | lcl\|CL1219-Ct3 | 0.0 | 419 | 1667 | 207 | Prot.Kin. (c-IC) |
| | *Pti 4* | 52 | 1 | lcl\|CL13275-Ct1 | 2 $e^{47}$ | 104 | 853 | 256 | AP2 (n-IC) |
| | *Pti5* | 1 | 0 | lcl\|CL13275-Ct1 | 5 $e^{06}$ | 49 | 853 | 256 | AP2 (n-IC) |
| | *Pti6* | 61 | 12 | lcl\|CL13048-Ct1 | 1 $e^{30}$ | 117 | 453 | 150 | AP2 (n-IC) |
| | *RIN4* | 17 | 0 | lcl\|CL11031-Ct1 | 5 $e^{29}$ | 81.3 | 905 | 233 | - |
| | *RPM1* | 26 | 2 | lcl\|CL4449-Ct1 | 1 $e^{105}$ | 85.4 | 3165 | 470 | NBS (n-IC)/LRR (CP) |
| | *RPS2* | 16 | 0 | lcl\|CL9468-Ct1 | 3 $e^{19}$ | 44 | 675 | 126 | NBS (n-IC) |
| | *RPS4* | 50 | 7 | lcl\|CL12379-Ct1 | 5 $e^{73}$ | 243 | 746 | 66 | - |
| | *RPS5* | 5 | 0 | lcl\|CL2439-Ct2 | 9 $e^{07}$ | 53.8 | 1261 | 374 | NBS/LRR (CP) |
| | *PBS1* | 249 | 10 | lcl\|CL4898-Ct1 | 0.0 | 602 | 1277 | 356 | Prot.Kin. (CP) |
| | *RAR1* | 3 | 0 | lcl\|CL20348-Ct1 | 4 $e^{16}$ | 63.4 | 685 | 223 | CHORD (CP) |
| | *MLA10* | 36 | 4 | lcl\|CL12379-Ct1 | 2 $e^{80}$ | 166 | 746 | 66 | - |
| | *L6* | 27 | 3 | lcl\|CL17937-Ct1 | 5 $e^{45}$ | 128 | 623 | 164 | NBS (n-IC) |
| | *RRS1* | 32 | 0 | lcl\|CL9540-Ct1 | 4 $e^{85}$ | 299 | 685 | 217 | NBS (n-IC) |
| **ST +HR** | *CERK1* | 249 | 0 | lcl\|CL2995-Ct1 | 8 $e^{180}$ | 398 | 1017 | 316 | Prot.Kin. (CP) |
| | *JAZ* | 15 | 0 | lcl\|CL1899-Ct2 | 8 $e^{80}$ | 266 | 1401 | 270 | Tify super Fam. (CP) |
| | *MPK4* | 181 | 2 | lcl\|CL681-Ct2 | 1 $e^{174}$ | 515 | 1554 | 375 | Prot.Kin.-like (CP) |
| | *FRK1* | 249 | 0 | lcl\|CL3705-Ct1 | 0.0 | 623 | 1656 | 449 | Prot.Kin. (CP) |
| | *EIX* | 113 | 6 | lcl\|CL587-Ct3 | 6 $e^{46}$ | 82.6 | 2208 | 714 | LRR (CP) |
| | *MYC2* | 24 | 23 | lcl\|CL3629-Ct1 | 4 $e^{73}$ | 189 | 1023 | 150 | - |
| | *MIN7* | 8 | 8 | lcl\|CL836-Ct4 | 0.0 | 397 | 1523 | 496 | - |
| | *HSP90* | 43 | 43 | lcl\|CL143-Ct3 | 0.0 | 593 | 2125 | 610 | HTPase_c (CP) |
| | *MPK3/6* | 207 | 3 | lcl\|CL4057-Ct1 | 0.0 | 797 | 1467 | 278 | Prot. Kin. (CP) |
| | *MKK1/2* | 234 | 0 | lcl\|CL3935-Ct1 | 1 $e^{136}$ | 319 | 1021 | 312 | PKc_MAPKK (CP) |
| | *COI1* | 24 | 24 | lcl\|CL464-Ct1 | 0.0 | 907 | 1789 | 390 | - |
| | *NHO1* | 1 | 1 | lcl\|CL15012-Ct1 | 4 $e^{110}$ | 354 | 1038 | 302 | FGGY_N super Fam. (IC) |
| | *BAK1/SER 4* | 249 | 2 | lcl\|CL1733-Ct2 | 0.0 | 763 | 1916 | 276 | Prot. Kin. (CP) |
| | *MEKK1* | 220 | 13 | lcl\|CL7215-Ct1 | 3 $e^{158}$ | 534 | 1179 | 288 | Prot. Kin. (CP) |
| | *CEBiP* | 3 | 0 | lcl\|CL18653-Ct1 | 2 $e^{69}$ | 179 | 735 | 225 | LysM (CP) |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *MKK4/5* | 196 | 14 | lcl\|CL11106-Ct1 | 3 $e^{47}$ | 187 | 573 | 149 | - |
| **ROS** | *STG1* | 17 | 16 | lcl\|CL470-Ct2 | 0.0 | 270 | 1530 | 362 | TPR (CP) |
| | *CNGCs* | 16 | 0 | lcl\|CL2014-Ct2 | 0.0 | 648 | 1933 | 628 | CAP_ED |
| | *CDPKs* | 183 | 11 | lcl\|CL319-Ct6 | 0.0 | 1070 | 2194 | 535 | Prot.Kin./EF-hand/EF-hand |
| | *Rboh* | 16 | 0 | lcl\|CL1560-Ct2 | 0.0 | 653 | 1781 | 287 | NOX_Duox |
| | *CaMCML* | 56 | 1 | lcl\|CL59-Ct10 | 8 $e^{106}$ | 360 | 756 | 149 | EF-hand/EF-hand |
| | *NOS* | 2 | 2 | lcl\|CL12950-Ct1 | 2 $e^{93}$ | 245 | 775 | 195 | GTPase_YqeH(n-IC) |

Table 2. Sugarcane upregulated SuperSAGE tag ($p < 0.05$) under abiotic stress-related to sugarcane contigs annotated to R gene products. Key for abbreviations: Fam.=Family; Kin.=Kinase; LRR=Leucine-Rich; Prot=Protein; Receptor=recep.

| Tag | Contig | R protein | Description | Stress* | Chromosome |
|---|---|---|---|---|---|
| Sg232385 | lcl\|CL4049Ctg-1 | FBL3 | F-box/LRR-repeat prot. 3 | Salt | - |
| Sg309748 | lcl\|CL3705Ctg-2 | HERK1 | Recep. prot. kin. HERK 1 | Salt | 6 |
| Sg309667 | lcl\|CL444Ctg-9 | RKL1 | Recep. kin. | Salt | - |
| Sg313217 | lcl\|CL1Ctg-731 | RLK | Recep. kin. | Salt | 12 |
| Sg27138 | lcl\|CL4449Ctg-1 | RPM1 | Disease resistance prot. RPM1 | Salt | 11 |
| Sg350189 | lcl\|CL1621Ctg-1 | SRF3 | Prot. strubbelig recep. fam. 3 | Salt | 2 |
| Sg58453 | lcl\|CL1621Ctg-1 | SRF3 | Prot. strubbelig recep. fam. 3 | Salt | 2 |
| Sg145047 | lcl\|CL138Ctg-2 | TRXH | Thioredoxin H-type | Salt | - |
| Sg14055 | lcl\|CL3962Ctg-2 | APK1B | Prot. kin. APK1A, chloroplastic | Drought | 9 |
| Sg41941 | lcl\|CL3193Ctg-1 | LRR | LRR-repeat (LRR) prot. | Drought | 3 |
| Sg284448 | lcl\|CL6862Ctg-1 | LRR-RLK | LRR-repeat recep. kin. | Drought | - |
| Sg171649 | lcl\|CL357Ctg-3 | MYB4 | Myb-related prot. Myb4 | Drought | - |
| Sg307209 | lcl\|CL1086Ctg-2 | RLK | Recep. kin. | Drought | 5 |
| Sg108593 | lcl\|CL2628Ctg-2 | RLK | Recep. kin. | Drought | 2 |
| Sg164846 | lcl\|CL3899Ctg-1 | SERK1 | Somatic embryogenesis recep.-kin.1 | Drought | - |
| Sg32183 | lcl\|CL1415Ctg-1 | SRF3 | Prot. strubbelig recep. fam. 3 | Drought | - |
| Sg168565 | lcl\|CL6296Ctg-1 | STPK | Ser/Thr prot. kin. | Drought | 3 |
| Sg86044 | lcl\|CL8140Ctg-1 | STPK | Ser/Thr prot. kin. | Drought | 2 |
| Sg108292 | lcl\|CL13024Ctg-1 | TIR1 | Transport inhibitor resp. 1 prot. | Drought | - |
| Sg159973 | lcl\|CL955Ctg-1 | TRX2 | Thioredoxin H-type 2 | Drought | - |
| Sg51485 | lcl\|CL138Ctg-2 | TRXH | Thioredoxin H-type | Drought | - |
| Sg85924 | lcl\|CL1Ctg-1314 | BRL1 | Ser/Thr prot. kin. BRI1 1 | Salt/ Drought (S) | 12 |
| Sg67140 | lcl\|CL2957Ctg-1 | CLV1 | Recep. prot. kin. CLAVATA1 | Salt/ Drought (S) | 3 |
| Sg112945 | lcl\|CL262Ctg-8 | COI1 | Coronatine-insensitive prot. 1 | Salt/ Drought (S) | - |
| Sg359087 | lcl\|CL2155Ctg-1 | DSK1 | LRR-recep. Ser/Thr prot. kin. | Salt/ Drought (S) | 1 |
| Sg177002 | lcl\|CL7232Ctg-1 | EXS | LRR-repeat / prot. kin. EXS | Salt/ Drought (S) | 2 |
| Sg55869 | lcl\|CL2503Ctg-4 | FER | Recep. prot. kin. FERONIA | Salt/ Drought (S) | 1 |
| Sg261220 | lcl\|CL3311Ctg-1 | HSL1 | Recep. prot. kin. HSL1 | Salt/ Drought (S) | 8 |
| Sg269246 | lcl\|CL592Ctg-6 | HT1 | Ser/Thr prot. kin. HT1 | Salt/ Drought (S) | 6 |
| Sg29268 | lcl\|CL9800Ctg-1 | LRR-RLK | LRR recep. prot. kin. | Salt/ Drought (S) | 4 |
| Sg136128 | lcl\|CL3699Ctg-2 | MYB1 | Myb-related prot. Hv1 | Salt/ Drought (S) | - |
| Sg2721 | lcl\|CL1056Ctg-3 | RLK | Recep. kin. | Salt/ Drought (S) | 5 |
| Sg165539 | lcl\|CL1Ctg-731 | RLK | Recep. kin. | Salt/ Drought (S) | 12 |
| Sg242481 | lcl\|CL1963Ctg-3 | TIR1 | Transport inhibitor resp. 1 prot. | Salt/ Drought (S) | - |
| Sg327823 | lcl\|CL12575Ctg-1 | EXS | LRR-repeat recep. kin. EXS | Salt/ Drought (T) | 1 |
| Sg202457 | lcl\|CL7304Ctg-1 | MEE39 | LRR recep. Sr/Th-prot. kin. MEE39 | Salt/ Drought (T) | 9 |
| Sg321931 | lcl\|CL1750Ctg-1 | STPK | Ser/Thr prot. kin. | Salt/ Drought (T) | 10 |
| Sg276480 | lcl\|CL3066Ctg-1 | STPK | Ser/Thr prot. kin. | Salt/ Drought (T) | 4 |
| Sg285898 | lcl\|CL16830Ctg-1 | WAK4 | Wall-associated recep. kin. 4 | Salt/ Drought (T) | - |
| Sg90440 | lcl\|CL1963Ctg-2 | AFB2 | Transport inhibitor resp. 1 prot. | Drought (T) | 4 |
| Sg112939 | lcl\|CL6590Ctg-2 | AFB2 | Transport inhibitor resp. 1 prot. | Drought (T) | - |
| Sg69207 | lcl\|CL900Ctg-4 | APK1A | Prot. kin. APK1A, chloroplastic | Drought (T) | 6 |
| Sg161113 | lcl\|CL5649Ctg-1 | APK2B | Prot. kin. 2B, chloroplastic | Drought (T) | 10 |
| Sg104430 | lcl\|CL1Ctg-1314 | BRL1 | Ser/Thr prot. kin. BRI1 1 | Drought (T) | 12 |

| | | | | | |
|---|---|---|---|---|---|
| Sg185147 | lcl\|CL262Ctg-8 | COI1 | Coronatine-insensitive prot. 1 | Drought (T) | - |
| Sg84672 | lcl\|CL9670Ctg-1 | CRK19 | Cysteine-rich recep. prot. kin. 19 | Drought (T) | 11 |
| Sg106653 | lcl\|CL5650Ctg-1 | CRK26 | Cysteine-rich recep. prot. kin. 26 | Drought (T) | - |
| Sg328269 | lcl\|CL14146Ctg-1 | CTR1 | Ser/Thr prot. kin. CTR1 | Drought (T) | 10 |
| Sg249703 | lcl\|CL3947Ctg-1 | CTR1 | Ser/Thr prot. kin. CTR1 | Drought (T) | - |
| Sg78151 | lcl\|CL3521Ctg-1 | HT1 | Ser/Thr prot. kin. HT1 | Drought (T) | 9 |
| Sg32977 | lcl\|CL592Ctg-6 | HT1 | Ser/Thr prot. kin. HT1 | Drought (T) | 16 |
| Sg139561 | lcl\|CL4853Ctg-1 | LECRKA4.3 | Lectin recep. kin. prot. | Drought (T) | 10 |
| Sg150666 | lcl\|CL11059Ctg-1 | LRR RLK | LRR-repeat recep. kin. | Drought (T) | 2 |
| Sg197538 | lcl\|CL1856Ctg-2 | LRR RLK | LRR-repeat recep. kin. | Drought (T) | 8 |
| Sg42130 | lcl\|CL2271Ctg-1 | LRR RLK | LRR-repeat recep. kin. | Drought (T) | 3 |
| Sg258732 | lcl\|CL8133Ctg-1 | LRR RLK | LRR-repeat recep. kin. | Drought (T) | - |
| Sg126264 | lcl\|CL2866Ctg-2 | LRR | LRR-repeat (LRR) prot. | Drought (T) | - |
| Sg84166 | lcl\|CL2090Ctg-3 | LRR-RLK | LRR-recep. prot. kin. | Drought (T) | 11 |
| Sg134899 | lcl\|CL2439Ctg-2 | LRR UBL | LRR-ubiquitin Fam. prot. | Drought (T) | 10 |
| Sg81195 | lcl\|CL2439Ctg-2 | LRR UBL | LRR-ubiquitin Fam. prot. | Drought (T) | 10 |
| Sg205102 | lcl\|CL3699Ctg-1 | MYB1 | Myb-related prot. Hv1 | Drought (T) | - |
| Sg95975 | lcl\|CL357Ctg-3 | MYB4 | Myb-related prot. Myb4 | Drought (T) | - |
| Sg164597 | lcl\|CL9007Ctg-1 | NAK | Ser/Thr prot. kin. NAK | Drought (T) | 6 |
| Sg44998 | lcl\|CL5525Ctg-1 | PUB34 | U-box domain-containing prot. 34 | Drought (T) | 6 |
| Sg157473 | lcl\|CL1296Ctg-5 | RLK | Recep. like kin. | Drought (T) | 5 |
| Sg13069 | lcl\|CL4024Ctg-1 | RLK | Recep. like kin. | Drought (T) | 1 |
| Sg16260 | lcl\|CL7850Ctg-1 | RLK | Recep. like kin. | Drought (T) | 12 |
| Sg190665 | lcl\|CL289Ctg-2 | SERK2 | Somatic embryogen. recep. kin. 2 | Drought (T) | 1 |
| Sg255086 | lcl\|CL1356Ctg-1 | SRF3 | Prot. strubbelig recep. Fam. 3 | Drought (T) | 7 |
| Sg47088 | lcl\|CL586Ctg-1 | SRF7 | Prot. strubbelig recep. Fam. 7 | Drought (T) | - |
| Sg273272 | lcl\|CL2666Ctg-1 | SRF8 | Prot. strubbelig recep. Fam. 8 | Drought (T) | 1 |
| Sg103811 | lcl\|CL19234Ctg-1 | STPK | Ser/Thr prot. kin. | Drought (T) | - |
| Sg254990 | lcl\|CL3066Ctg-1 | STPK | Ser/Thr prot. kin. | Drought (T) | 4 |
| Sg186844 | lcl\|CL1963Ctg-3 | TIR1 | Transport inhibitor resp. 1 prot. | Drought (T) | - |
| Sg64443 | lcl\|CL955Ctg-1 | TRX2 | Thioredoxin H-type 2 | Drought (T) | - |
| Sg51480 | lcl\|CL138Ctg-7 | TRXH | Thioredoxin H-type | Drought (T) | - |
| Sg169613 | lcl\|CL218Ctg-3 | TRXH | Thioredoxin H-type | Drought (T) | - |

Table 3. Sugarcane upregulated SuperSAGE tags ($p < 0.05$) under abiotic stress* related to sugarcane contigs annotated to PR gene products.

| Tag | Contig | PR protein | Description | Stress* | Chromosome |
|---|---|---|---|---|---|
| Sg89596 | lcl\|CL1656Ctg-4 | BGL14 | Glucan endo-1,3-beta-glucosidase 14 | Salt | 6 |
| Sg156201 | lcl\|CL4004Ctg-1 | BGL6 | Glucan endo-1,3-beta-glucosidase 6 | Salt | 2 |
| Sg234615 | lcl\|CL5380Ctg-1 | CHT2 | Chitinase 2 | Salt | 5 |
| Sg9165 | lcl\|CL2359Ctg-1 | PER1 | Peroxidase 1 | Salt | - |
| Sg102410 | lcl\|CL857Ctg-3 | PER12 | Peroxidase 12 | Salt | 4 |
| Sg280254 | lcl\|CL3370Ctg-2 | PER2 | Peroxidase 2 | Salt | 5 |
| Sg319382 | lcl\|CL3370Ctg-2 | PER2 | Peroxidase 2 | Salt | 5 |
| Sg103631 | lcl\|CL81Ctg-11 | PER30 | Peroxidase 30 | Salt | 1 |
| Sg8601 | lcl\|CL6421Ctg-1 | PER51 | Peroxidase 51 | Salt | 8 |
| Sg44533 | lcl\|CL2278Ctg-1 | PR6 | Pathogenesis-related Prot. 6 | Salt | 1 |
| Sg40564 | lcl\|CL2300Ctg-1 | PRMS | Pathogenesis-related Prot. PRMS | Salt | 1 |
| Sg134165 | lcl\|CL3953Ctg-1 | PRX74 | Peroxidase 1 | Salt | 5 |
| Sg114624 | lcl\|CL2475Ctg-1 | SLP3 | Serine-type peptidase 3 | Salt | 6 |
| Sg184829 | lcl\|CL170Ctg-7 | TL1 | Thaumatin-like Prot. 1 | Salt | 10 |
| Sg196058 | lcl\|CL9293Ctg-1 | GER8 | Germin-like Prot. 8 | Salt/ Drought (T) | 1,3 |
| Sg260023 | lcl\|CL1686Ctg-2 | LIC2 | Lichenase-2 (Fragment) | Salt/ Drought (T) | 5 |
| Sg192751 | lcl\|CL148Ctg-1 | LTP3 | Non-specific lipid-transfer Prot. 3 | Salt/ Drought (T) | 11 |
| Sg78722 | lcl\|CL81Ctg-11 | PER30 | Peroxidase 30 | Salt/ Drought (T) | 1 |
| Sg237293 | lcl\|CL12204Ctg-1 | PER35 | Peroxidase 35 | Salt/ Drought (T) | 4 |
| Sg48247 | lcl\|CL5Ctg-2 | PR4 | Pathogenesis-related Prot. 4 | Salt/ Drought (T) | - |
| Sg240549 | lcl\|CL2278Ctg-1 | PR6 | Pathogenesis-related Prot. 6 | Salt/ Drought (T) | 1 |
| Sg127746 | lcl\|CL7441Ctg-1 | TL1 | Thaumatin-like Prot. 1 | Salt/ Drought (T) | 10 |
| Sg105253 | lcl\|CL172Ctg-1 | TLP | Thaumatin-like Prot. | Salt/ Drought (T) | - |
| Sg209573 | lcl\|CL172Ctg-3 | TLP | Thaumatin-like Prot. | Salt/ Drought (T) | 12 |
| Sg32358 | lcl\|CL172Ctg-3 | TLP | Thaumatin-like Prot. | Salt/ Drought (T) | 12 |
| Sg301849 | lcl\|CL172Ctg-4 | TLP | Thaumatin-like Prot. | Salt/ Drought (T) | 12 |
| Sg18972 | lcl\|CL949Ctg-2 | TLP | Thaumatin-like Prot. | Salt/ Drought (T) | - |
| Sg48178 | lcl\|CL5335Ctg-1 | ZLP | Zeamatin-like Prot. | Salt/ Drought (T) | 3 |
| Sg188704 | lcl\|CL54Ctg-1 | APX1 | L-ascorbate peroxidase 1 cytosolic | Drought | 3 |
| Sg13934 | lcl\|CL1Ctg-628 | APX2 | L-ascorbate peroxidase 2 cytosolic | Drought | 7 |
| Sg32000 | lcl\|CL1Ctg-769 | APX2 | L-ascorbate peroxidase 2 cytosolic | Drought | - |
| Sg20615 | lcl\|CL4012Ctg-2 | ARA12 | Subtilisin-like protease | Drought | 4 |
| Sg326868 | lcl\|CL4788Ctg-2 | BGL(GIV) | Glucan endo-1,3-beta-glucosidase GIV | Drought | 1 |
| Sg326865 | lcl\|CL11799Ctg-1 | BGL | Glucan endo-1,3-beta-glucosidase GVI | Drought | 1 |
| Sg3555 | lcl\|CL1656Ctg-3 | BGL14 | Glucan endo-1,3-beta-glucosidase 14 | Drought | 6 |
| Sg171075 | lcl\|CL1656Ctg-4 | BGL14 | Glucan endo-1,3-beta-glucosidase 14 | Drought | 6 |
| Sg103796 | lcl\|CL1061Ctg-5 | CHT1 | Chitinase 1 | Drought | - |
| Sg139239 | lcl\|CL2946Ctg-2 | CHT12 | Chitinase 12 | Drought | - |
| Sg181942 | lcl\|CL2832Ctg-1 | GLP | Germin-like Prot. | Drought | 1,3 |
| Sg275910 | lcl\|CL148Ctg-2 | LTP3 | Non-specific lipid-transfer Prot. 3 | Drought | - |
| Sg53947 | lcl\|CL2173Ctg-2 | PER | Peroxidase | Drought | - |
| Sg105048 | lcl\|CL5161Ctg-2 | PER1 | Peroxidase 1 | Drought | 1 |
| Sg20895 | lcl\|CL3370Ctg-2 | PER2 | Peroxidase 2 | Drought | 5 |
| Sg219297 | lcl\|CL4132Ctg-2 | PER2 | Peroxidase 2 | Drought | 3 |
| Sg57316 | lcl\|CL18012Ctg-1 | PER4 | Peroxidase 4 | Drought | 6 |

| | | | | | |
|---|---|---|---|---|---|
| Sg96280 | lcl\|CL6421Ctg-1 | PER51 | Peroxidase 51 | Drought | 8 |
| Sg168319 | lcl\|CL81Ctg-8 | POD | Peroxidase 15 | Drought | - |
| Sg180951 | lcl\|CL5594Ctg-1 | PRB1-2 | Pathogenesis-related Prot. PRB1-2 | Drought | 10 |
| Sg180925 | lcl\|CL10368Ctg-1 | TLP | Thaumatin-like Prot. | Drought | 3, 9, 11 |
| Sg177465 | lcl\|CL4842Ctg-1 | ARA12 | Subtilisin-like protease | Drought (S) | 3 |
| Sg310886 | lcl\|CL867Ctg-1 | CTL1 | Chitinase-like Prot. 1 | Drought (S) | 9 |
| Sg345193 | lcl\|CL867Ctg-1 | CTL1 | Chitinase-like Prot. 1 | Drought (S) | 9 |
| Sg175504 | lcl\|CL58Ctg-12 | LTP110-A | Non-specific lipid-transfer Prot. 3 | Drought (S) | - |
| Sg29841 | lcl\|CL2173Ctg-2 | PER | Peroxidase | Drought (S) | - |
| Sg62895 | lcl\|CL7859Ctg-1 | PER12 | Peroxidase 12 | Drought (S) | - |
| Sg224114 | lcl\|CL13415Ctg-1 | PER2 | Peroxidase 2 | Drought (S) | - |
| Sg218991 | lcl\|CL7657Ctg-1 | PER2 | Peroxidase 2 | Drought (S) | 8 |
| Sg167970 | lcl\|CL1172Ctg-3 | PER3 | Peroxidase 3 | Drought (S) | 6 |
| Sg291746 | lcl\|CL12204Ctg-1 | PER35 | Peroxidase 35 | Drought (S) | 4 |
| Sg76238 | lcl\|CL3102Ctg-2 | PER36 | Peroxidase 36 | Drought (S) | - |
| Sg96280 | lcl\|CL6421Ctg-1 | PER51 | Peroxidase 51 | Drought (S) | 8 |
| Sg102099 | lcl\|CL17223Ctg-1 | PER52 | Peroxidase 52 | Drought (S) | 2 |
| Sg162753 | lcl\|CL81Ctg-10 | PER72 | Peroxidase 72 | Drought (S) | 1 |
| Sg168319 | lcl\|CL81Ctg-8 | POD | Peroxidase 15 | Drought (S) | - |
| Sg218368 | lcl\|CL17093Ctg-1 | RIXI | Xylanase inhibitor Prot. 1 | Drought (S) | 2, 11 |
| Sg3555 | lcl\|CL1656Ctg-3 | BGL14 | Glucan endo-1,3-beta-glucosidase 14 | Drought (T) | 6 |
| Sg171075 | lcl\|CL1656Ctg-4 | BGL14 | Glucan endo-1,3-beta-glucosidase 14 | Drought (T) | 6 |
| Sg103796 | lcl\|CL1061Ctg-5 | CHT1 | Chitinase 1 | Drought (T) | - |
| Sg95918 | lcl\|CL12126Ctg-1 | ECH | Chitinase 12 | Drought (T) | 4 |
| Sg338428 | lcl\|CL402Ctg-4 | ECH | Chitinase 12 | Drought (T) | 4 |
| Sg192751 | lcl\|CL148Ctg-1 | LTP3 | Non-specific lipid-transfer Prot. 3 | Drought (T) | 11 |
| Sg219297 | lcl\|CL4132Ctg-2 | PER2 | Peroxidase 2 | Drought (T) | 3 |
| Sg180925 | lcl\|CL10368Ctg-1 | TLP | Thaumatin-like Prot. | Drought (T) | 3, 9, 11 |

* Stress [Salt: 100 mM NaCl; drought: 24h after water suppression]

Table 4: Sugarcane clusters that presented similarities with distinct segments in a same rice chromosome.

| Chromosome | Cluster | Position | |
|---|---|---|---|
| | | Start | End |
| 1 | CL13048Contig1 | 6813508 | 6813684 |
| | | 26733776 | 26733608 |
| 2 | CL19373Contig1* | 969864 | 970285 |
| | | 981312 | 981732 |
| 3 | CL13191Contig1* | 17681689 | 17681192 |
| | | 17702859 | 17702476 |
| | CL2713 Contig1 e Contig 2 | 510781 | 512759 |
| | | 997006 | 999145 |
| | CL2740Contig1 | 22702577 | 22703551 |
| | | 31482289 | 31481744 |
| | CL8128Contig1 | 28750774 | 28748635 |
| | | 35477873 | 35477018 |
| 7 | CL15614Contig1* | 21122766 | 21122970 |
| | | 21130188 | 21130392 |
| 9 | CL13870Contig1 | 10155655 | 10156233 |
| | | 10366331 | 10366736 |
| | CL22Contig10 | 18514955 | 18514293 |
| | | 18525801 | 18525139 |
| | | 18537260 | 18536598 |
| | | 21015489 | 21015930 |
| | CL22Contig8 | 18515301 | 18514527 |
| | | 18526115 | 18525376 |
| | | 18537574 | 18536835 |
| | | 21015124 | 21015815 |
| 11 | CL12189Contig1 | 26771155 | 26770944 |
| | | 26790178 | 26789967 |
| | | 26871113 | 26870902 |

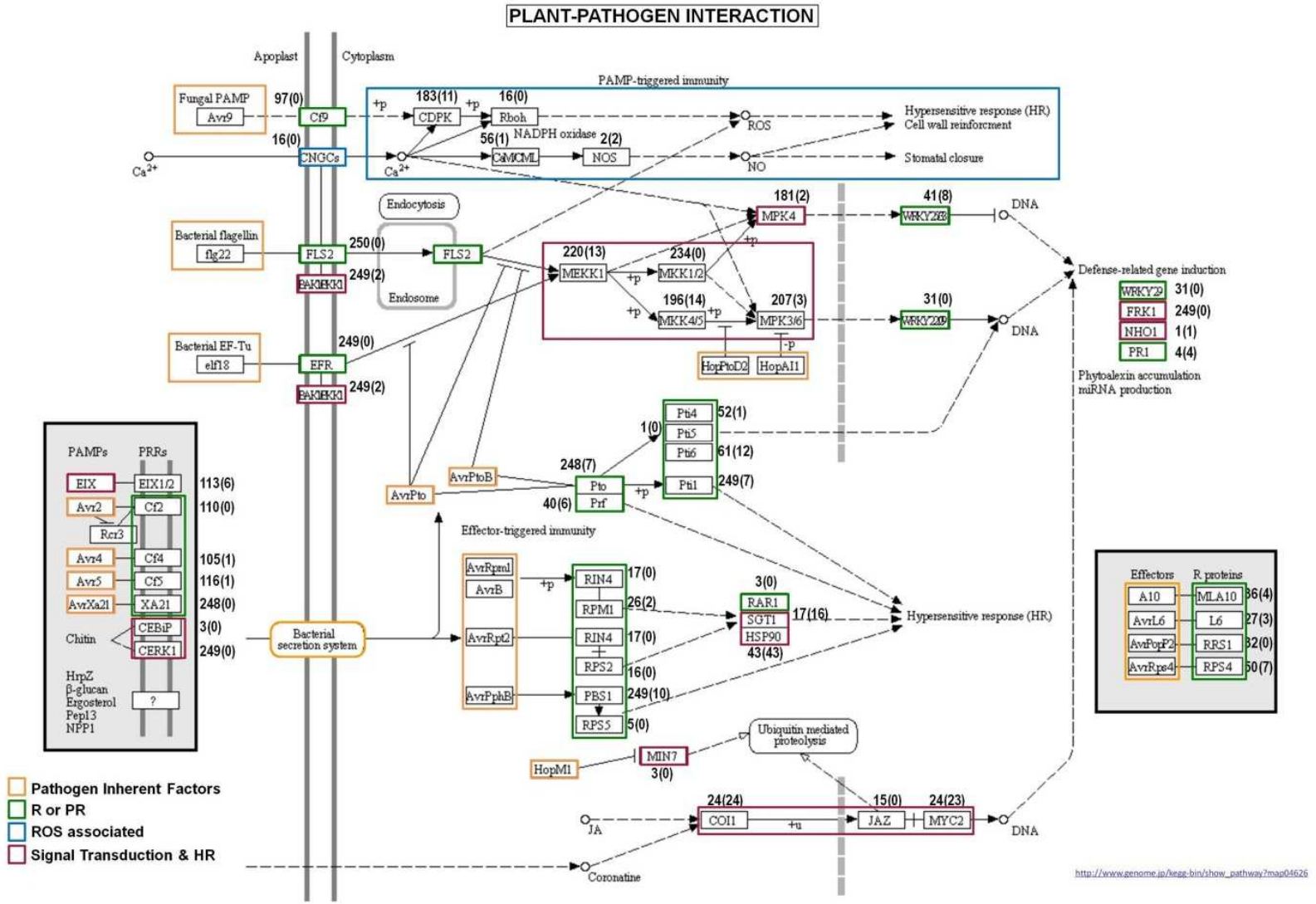



**Fig. 1 Plant-pathogen interaction pathway available at KEGG (Kyoto Encyclopedia of** Genes and Genomes, http://www.genome.jp/kegg-bin/show_pathway?map04626) showing the number of elements found in sugarcane transcriptome for each gene type, followed by the number of exclusively aligned sequences (between parenthesis). Abbreviations: HR, Hypersensitive Response; PAMP: pattern associated to the pathogen; PR, Pathogenesis Related; PaPRR: pattern recognition receptor; R, Resistance; ROS, Reactive Oxygen Species.

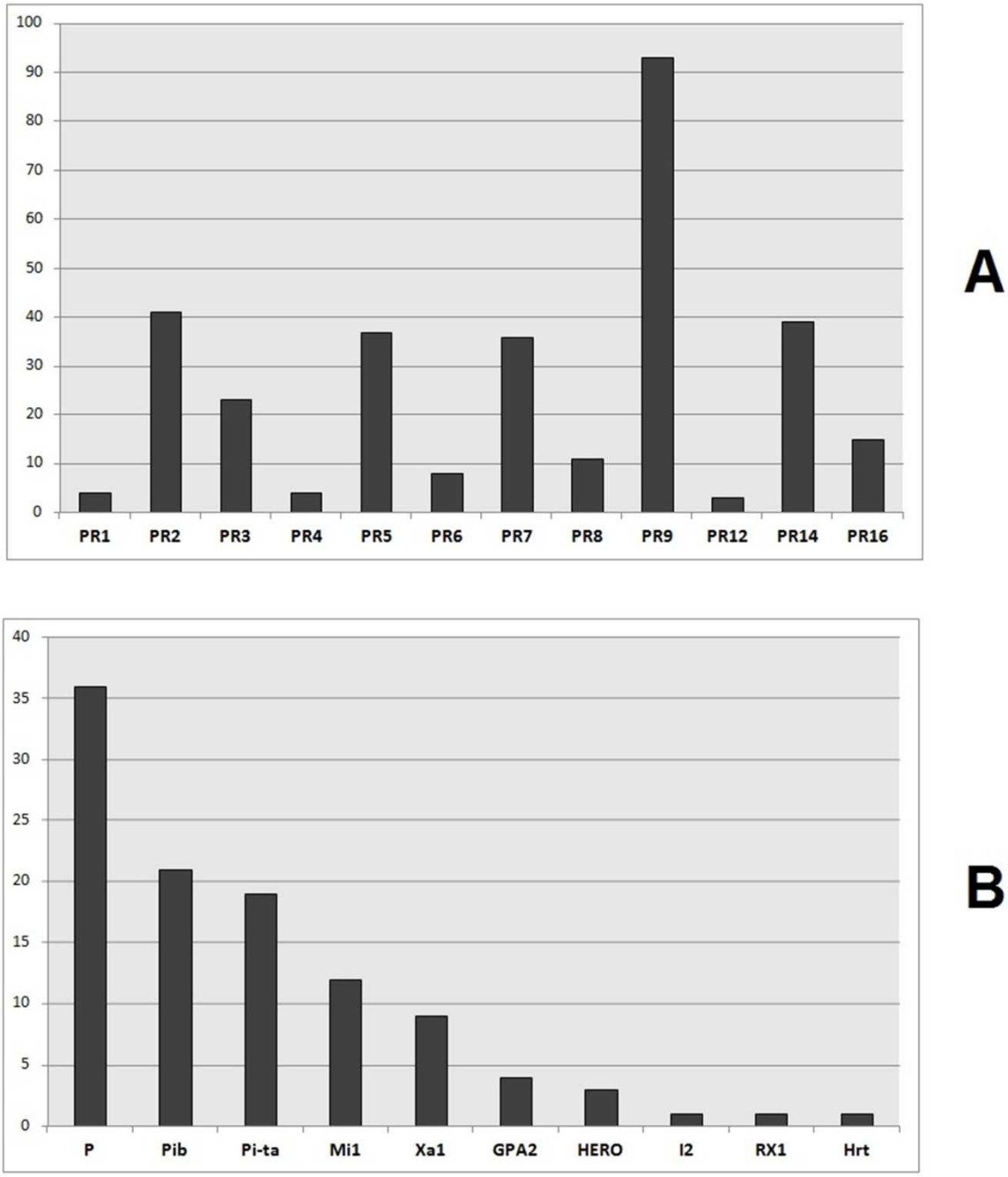



**Fig. 2 Distribution of the 314 sugarcane clusters that aligned with *PR*-genes families (A)** and of the 107sugarcane clusters that aligned with *R*-genes with no repetitions **(B).**

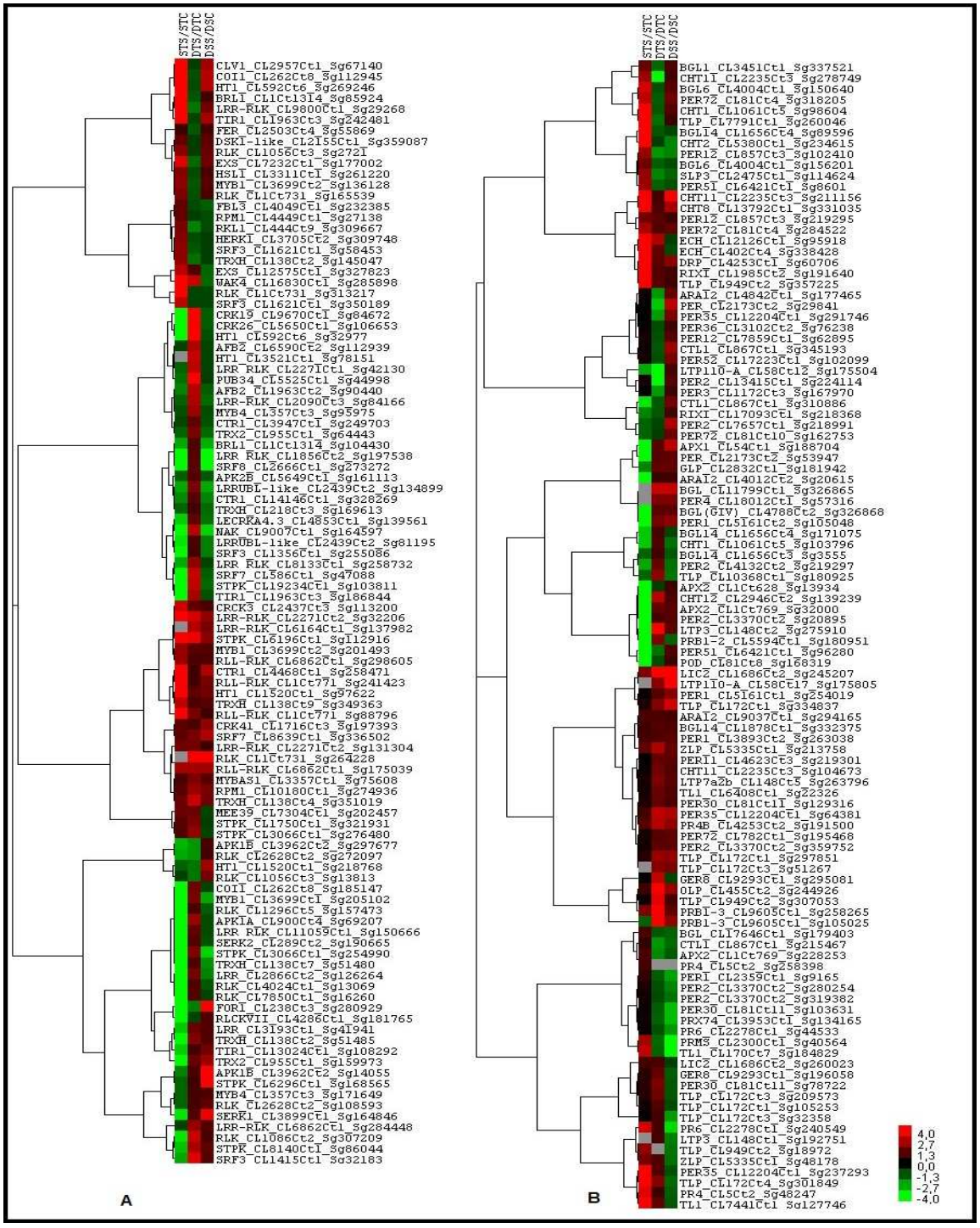



**Fig. 3 Hierarchical clustering (Cluster3.0) of up-regulated (red) and down-regulated** (green) SuperSAGE tags ($p < 0,05$) related to R (**A**) and PR classes (**B**) using FC (Fold Change) of gene expression ratios (experimental/control) under the tested conditions [STS/STC: Salinity-tolerant (100 mMNaCl)/ control; DTS/DTC: Drought- tolerant / control; DSS/DSC: Drought- sensible / control]. The tag number and the contig ID is given together with the gene product.

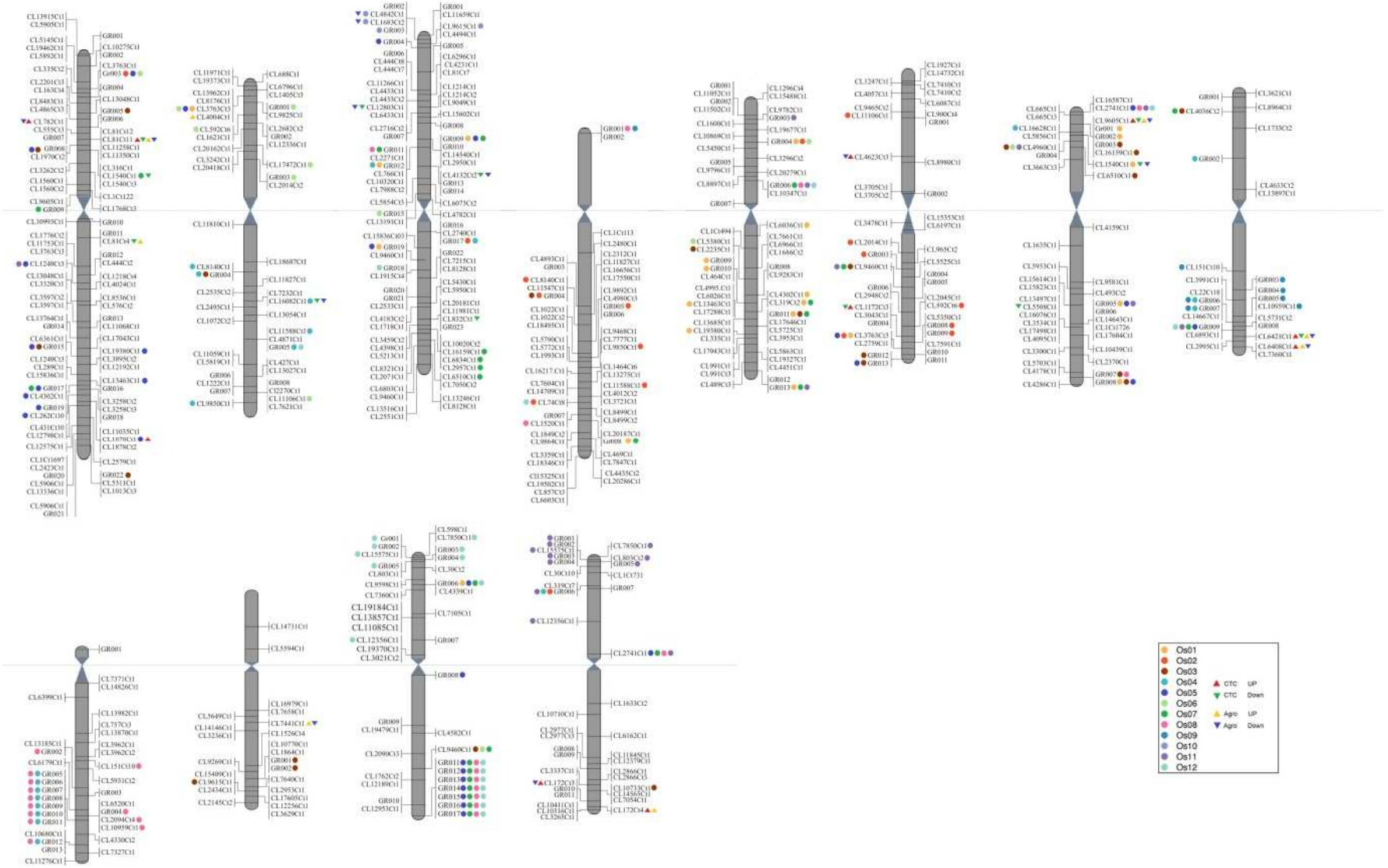



**Fig. 4** ***In silico*** **FISH.** Schematic representation of clusters/groupings that were anchored in the rice genome based on BLAST similarity results. Colored circles next to the clusters/grouping names correspond to the synteny events between chromosomes; each one of the 12 chromosomes was identified with a different color, as showed in the legend. Triangles and inverted triangles indicate the similarity of expressed tags (resulted from SuperSAGE) with clusters/grouping; the form colors and senses correspond to the library and regulation type, according the legend.

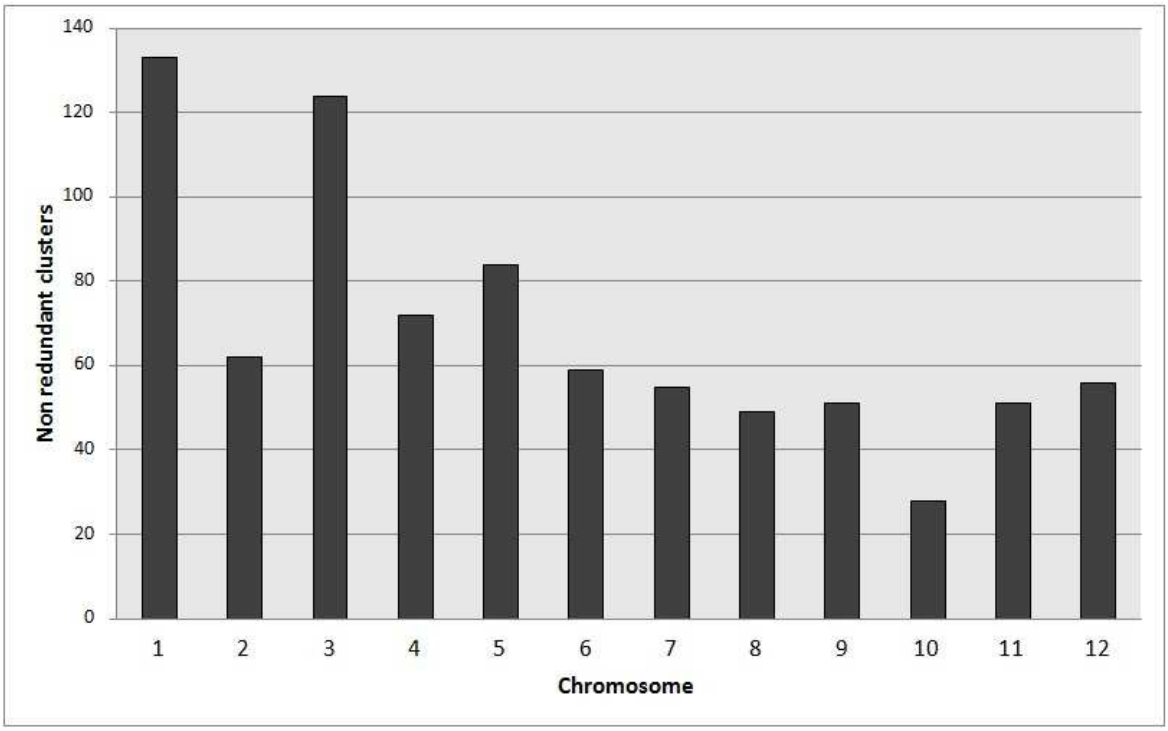



**Fig. 5** Number of sugarcane clusters that aligned in each rice chromosome.

## CAPÍTULO II
## (Artigo Aceito para Publicação na Revista Genetics and Molecular Biology)

# An overall evaluation of the Resistance (*R*) and Pathogenesis-Related (*PR*) superfamilies in soybean, as compared with *Medicago* and *Arabidopsis*

Ana C. Wanderley-Nogueira[1], Luis C. Belarmino[1], Nina da M. Soares-Cavalcanti[1], João P. Bezerra-Neto[1], Ederson A. Kido[1], Valesca Pandolfi[1], Ricardo V. Abdelnoor[2], Eliseu Bineck[2], Marcelo F. Carazzole[3,4] and Ana M. Benko-Iseppon[1]

[1]Department of Genetics, Center of Biological Sciences, Universidade Federal de Pernambuco, Recife, PE, Brazil.

[2]Embrapa Soja, Londrina, PR, Brazil.

[3]Laboratório de Genômica e Expressão, Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas-SP, Brazil.

[4]Centro Nacional de Processamento de Alto Desempenho em São Paulo, Universidade Estadual de CampinasCampinas, SP, Brazil.

Short running title: *R* and *PR* genes in soybean

Keywords: pathogen response, biotic stress, bioinformatics, *Glycine max*, *Medicago truncatula*.

Send correspondence to Ana Maria Benko-Iseppon. Department of Genetics, Center of Biological Sciences, Universidade Federal de Pernambuco, Av. Prof. Morais Rego 1235, 50.670-420, Recife, PE, Brazil. E-mail: ana.iseppon@gmail.com

## Abstract

Plants have the ability to recognize and respond to a multitude of pathogens, resulting in a massive reprogramming of the plant to activate defense responses including Resistance (*R*) and Pathogenesis-Related (*PR*) genes. Abiotic stresses can also activate *PR* genes and enhance pathogen resistance, representing valuable genes for breeding purposes. The present work offers an overview of soybean *R* and *PR* genes present in the GENOSOJA (Brazilian Soybean Genome Consortium) platform, regarding their structure, abundance, evolution and role in the plant-pathogen metabolic pathway, as compared with *Medicago* and *Arabidopsis*. Searches revealed 3,065 *R* candidates (756 in Soybean, 1,142 in *Medicago* and 1,167 in *Arabidopsis*), and *PR* candidates matching to 1,261 sequences (310, 585 and 366 for the three species, respectively). The identified transcripts were also evaluated regarding their expression pattern in 65 libraries, showing prevalence in seeds and developing tissues. Upon consulting the SuperSAGE libraries, 1,072 *R* and 481 *PR* tags were identified in association with the different libraries. Multiple alignments were generated for *Xa21* and *PR-2* genes, allowing inferences about their evolution. The results revealed interesting insights regarding the variability and complexity of defense genes in soybean, as compared with *Medicago* and *Arabidopsis*.

## Introduction

In order to prevent the effects of pathogen attack, plants evolved the ability to recognize the threat and struggle against the invader as well as trigger an effective response (Bolton, 2009). One of the most important steps of this complex response lies in the detection of pathogen invaders by the plant, a step where *R* (Resistance) genes play a crucial role. This sensing involves the recognition of a pathogen gene product called avirulence (avr) factor by a correspondent *R* gene. The plant will be resistant and the pathogen growth and establishment will be impaired when both *avr* and *R* genes are compatible, leading to the so-called Hypersensitive Response (HR) that triggers diverse responses, including local cell death to impair spreading of the pathogen (Bonas and Anckerveken, 1999). Besides this local reaction, the HR activates a signal cascade — including hormones and *PR* (Pathogen Related) genes, among others — that are able to establish resistance against a spectrum of different pathogen classes, this corroborating observations made at the beginning of the last century that plants, as well as animals (Benko-Iseppon *et al.*, 2010), may be immunized against the attack a of given pathogen after infection by another pathogen (Chester, 1933).

Besides a local reaction, plants may also display the Systemic Acquired Resistance (SAR). The SAR pathway is also common in many non-compatible plant-pathogen interactions (Nurnberg and Brunner, 2002). As soon as the pathogenic agent is detected, the plant induces a complex set of signal molecules able to activate defense proteins that may have a direct antimicrobial effect, as in the case of Pathogenesis-Related (*PR*) genes (Durrant and Dong, 2004). Alternatively, they may induce the production of secondary metabolites that impair pathogen movement or growth within the plant tissues (Sparla *et al.*, 2004).

Resistance genes are generally classified into five different groups or classes, defined according to their conserved domains (CD) (Bent, 1996; Hammond-Kosack and Jones, 1997; Ellis and Jones, 2000). The first class is represented by the *HM1* gene of maize that encodes a reductase able to inactivate toxins produced by the fungus *Helminthosporium carbonum* (Joahal and Briggs, 1992). It is the only *R* gene class where conserved domains are absent. A second class is represented by the *Pto* gene from tomato that confers resistance against the bacterium *Pseudomonas syringae* pv. *tomato.* It is characterized by a serine/threonine-kinase (ser/thre-kinase) domain, able to interact with the *avrPto* gene (Tang *et al.*, 1999). This gene was also identified in other plants, such as *Arabidopsis thaliana*, *Phaseolus vulgaris* (Melotto *et al.*, 2004), eucalyptus (Barbosa-da-Silva *et al.*, 2005) and sugarcane (Wanderley-Nogueira *et al.*, 2007).

The third class is represented by genes bearing two domains, viz. LRR (*Leucine Rich Repeats*) and NBS (*Nucleotide Binding Site*) (Liu *et al.*, 2004). This is the case of the *Rpm1* and *Rps2* genes from *A. thaliana*, the *N* gene from tobacco, *L6* from flax, *Prf* from tomato and *Rpg1* from soybean also found in common bean, and faba bean (Mindrinos *et al.*, 1994; Lawrence *et al.*, 1995; Salmeron *et al.*, 1996; Ashfield *et al.*, 2003). The fourth *R* gene class encodes a membrane-anchored protein composed of an extracellular LRR domain, a transmembrane region and a short intracellular tail in the C terminal. The *Cf* gene from tomato is an example of this class, conferring resistance against *Cladosporium fulvum* (Dixon *et al.*, 1996).

The *Xa21* gene from rice confers resistance to the bacteria *Xanthomonas oryzae* pv. *oryzae* and is a representative of the fifth class (Song *et al.*, 1995; Wang *et al.*, 1995). This gene encodes an extracellular LRR domain (similar to the *Cf* gene), as well as a ser/thre-kinase

domain (similar to the *Pto* gene), suggesting an evolutionary connection among different classes in the genesis of plant *R* genes (Song *et al.*, 1997).

PR proteins comprise pathogen-induced proteins that are routinely classified into 17 families based on their biochemical and molecular biological properties, from PR-1 to PR-17 (van-Loon *et al.*, 2006). Similarities among sequences and serological or immunological properties form the basis of their classification (van-Loon *et al.*, 1999). Although most PR proteins are known to have antifungal activities, their active molecular mechanisms are not well understood except for PR-2 (β-glucanases) and PR3 (chitinases) (Kitajima and Sato, 1999). PR1 is the most abundantly accumulated protein after pathogen infection and its genes have been cloned in many plants, such as tobacco (Gaffney *et al.*, 1993), *A. thaliana* (Metzler *et al.*, 1991), tomato (Tornero *et al.*, 1997) and apple. Although its phytochemical functions are unknown in all these species, this gene class is nonetheless considered to be a typical SAR marker (Bonasera *et al.*, 2006). PR-5 is a thaumatin-like protein with high antifungal activity, being also expressed under cold stress in overwintering monocots where it exhibits antifreeze activities (Hon *et al.*, 1995, Atici and Nalbantoğlu, 2003, Griffith and Yaish, 2004). Other families like PR-8 (Glycosyl hydrolase), PR-9 (secretory peroxydase), PR-14 (lipid transfer proteins), PR-15 (oxalate oxydase) and PR-17 (basic secretory proteins) (Nanda *et al.*, 2010) have been well studied and are believed to be involved in plant defense responses, although their molecular mechanisms have yet to be determined (Bolton, 2009). Most *PR* genes are expressed at a basal level under normal growth conditions, but are rapidly induced after pathogen infection. It is worthy of note that several *PR* genes are also regulated during development, leaf senescence and pollen maturation, as well as by environmental factors, such as osmotic, cold and light stress (Zeier *et al.*, 2004).

Soybean (*Glycine max*) is a globally important crop, providing oil and at least twice as much protein per acre as any other major grain (Libault *et al.*, 2010). Economically, soybean is the most valuable protein and edible oil crop in the world and serves as a model for seed and other developmental processes (Cannon *et al.*, 2009).

The present evaluation offers an overview of the main available sequences regarding plant-pathogen interaction of the *R* and *PR* classes in the soybean transcriptome, here compared with data available from *Arabidopsis* and *Medicago*, providing insights on the expression of such sequences in different tissues and inferring as to how these genes may have behaved over the course of evolution.

## Material and Methods

### Search and screening for R and PR genes in soybean, *Medicago* and *Arabidopsis* databases

For this purpose 59 proteins that play important roles in plant defense response were selected as seed sequences. The selected protein sequences were related to the 42 *R* and 17 *PR* gene classes described above. The *R* genes were previously compiled by Barbosa-da-Silva *et al.* (2005) and Wanderley-Nogueira *et al.* (2007), and *PR* seed sequences are available in Table S1 (Supplementary Material). All 59 seed sequences regarded full cDNAs that were obtained from the NCBI database and conceptually translated to improve search strategies.

For the identification of these gene analogs in soybean, *Medicago* and *Arabidopsis* transcriptomes, tBLASTx alignments were carried out against three platforms: GENOSOJA (The Brazilian Soybean Genome Consortium), TIGR (The Institute for Genomic Research) and TAIR (The Arabidopsis Information Resource), using $1e^{-05}$ as the cut-off value.

Obtained clusters were annotated and analyzed for score, e-values, sequence size and presence of conserved domains, as shown in Table 1. For this purpose all clusters were translated using the TRANSLATE tool of Expasy and screened for conserved motifs with the aid of the rps-BLAST CD-search tool (Altschul *et al.*, 1990). The best match for each gene in each studied species was submitted to a BLASTx alignment in NCBI GenBank in an effort to confirm their putative function.

In a second manual analysis redundancies, i.e. clusters that matched more than one gene due to common domains, were eliminated. For this purpose, clusters matching each query sequence were annotated on a local database (called 'non-redundant').

The third step of the analysis aimed at comparing the number of *R* and *PR* candidate sequences obtained after the tBLASTn searches against the soybean, *Arabidopsis* and *Medicago* databases by direct counting of non-redundant clusters for each one of the 59 genes studied.

**Phylogenetic analysis**

Aiming to analyze the relationships among these genes, some *R* and *PR* gene candidates were selected from all three studied species for an evolutionary analysis using the maximum parsimony method and bootstrap function with 5,000 replicates. For this purpose CLUSTALx alignments were submitted to the program MEGA (Molecular Evolutionary Genetic Analysis), Version 4 for Windows (Tamura *et al.*, 2007).

**Studying syntenic regions among the soybean and *Medicago* genomes**

Best matches for all selected soybean genes were aligned against the *M. truncatula* pseudogenome aiming to anchor the 59 soybean sequences in virtual chromosomes through the CVit-BLAST procedure implemented in the *Medicago* sequencing resource website. BLAST algorithm parameters (score, e-value and percentage of identity) were adjusted to infer about the position of soybean sequences along the *Medicago* virtual chromosomes.

***In silico* expression assay based on GENOSOJA EST sequences**

A preliminary analysis of the prevalence regarding the 59 genes in the soybean libraries was verified by direct correlation of the read frequencies of each cluster in various GENOSOJA cDNA libraries. Information regarding the 65 libraries that constitute the GENOSOJA database is available on The Soybean Genome Project Website. For practical purposes we combined some libraries that comprised different stages of the same tissue/organ (for example, B01 and B02 are here referred to as "B"), resulting in a total of 16 libraries (**B**: vegetable buds of field grown plants; **C**: cotyledons; **EN**: endosperm; **EP**: epicotyls; **F**: flowers; **H**: hypocotyls; **LV**: leaves; **R**: roots; **SH**: germination shoots; **ST**: stems; **SO**: somatic embryos; **SC**: soybean submitted to drought; **LI**: leaves infected with Asian rust; **MJ**: soybean submitted to *Meloidogyne javanica*; **SD**: seeds and **UK**: unknown). To generate an overall picture of selected *R* and *PR* gene expression patterns in soybean, a hierarchical clustering approach (Eisen *et al.*, 1998) was applied using normalized data and a graphic representation constructed with the aid of the CLUSTER program. Dendrograms including both axes (using the weighted pair-group for each cluster and library) were generated with aid of the TreeView program (Page, 1996). In these graphics, light yellow means no expression and red indicates all degrees of expression.

### *In silico* expression assay based on the GENOSOJA SuperSAGE libraries

*R* and *PR* candidates were also used to screen the six SuperSAGE libraries generated by the GENOSOJA consortium. For the drought experiment, four libraries were generated using roots of two contrasting soybean genotypes, viz. Embrapa-48 (tolerant) and BR-16 cultivar (susceptible), both submitted to dehydration in the dark for 25 up to 150 minutes (all times bulked together), as compared with non-stressed controls. The other stressed library was generated using leaves of the resistant accession PI561356 inoculated with rust fungus and collected 12, 24 and 48 hours post inoculation. For the composition of the pathogen-stressed library, equimolar amounts of the three inoculation times were used, as compared with the negative, non-inoculated control of the same genotype. The libraries were constructed at GenXPro GmbH (Frankfurt, Germany), essentially as described by Matsumura *et al.* (2008), and were subsequently sequenced via a SOLEXA platform.

Aiming to perform an overview of the GENOSOJA SuperSAGE data associated with *R* and *PR* genes, SuperSAGE tags were submitted to a BLASTn (maximum e-value $1e^{-05}$) against the database generated from three comparisons of the six available libraries (1- Embrapa-48, drought tolerant stressed vs. negative control; 2- BR-16, drought susceptible stressed vs. negative control; 3- PI561356 fungus resistant stressed vs. negative control). Each SuperSAGE tag was annotated considering the respective library comparison and also the respective aligned ESTs.

## Results

### Description and distribution of *R* and *PR* genes in soybean, *Medicago* and *Arabidopsis*

The tBLASTn alignment against the soybean transcriptome using the 59 known *R* and *PR* gene probes returned 1,066 non-redundant sequences from the contigs and singlets deposited in the GENOSOJA database. Among them, 700 represented contigs and 366 singlets, which together encompassed 26,653 reads. Regarding the tBLASTn searches in the *Medicago* transcriptome, a total of 1,727 sequences were positive matches. In *Arabidopsis,* 1,533 sequences returned matches after the same procedure.

A screening of *R* and *PR* genes in these three species resulted in the identification of 4,326 candidates, of which 3,065 were *R* and 1,261 *PR* gene candidates. A graphical representation regarding the prevalence of these sequences and how they are distributed among the soybean, *Medicago* and *Arabidopsis* transcriptomes is shown in Figure 1.

After analyzing all results it was observed that only one *PR* (*PR-13*) and two *R* genes (*L6* and *M*) were absent from the soybean transcriptome, while all the other 56 genes presented positive results in the tBLASTn searches. The same was denoted in the *Medicago* tBLASTn results for these three genes. Also in *Arabidopsis* no matches could be found for the two R genes *L6* and *M*, but four candidate sequences could be identified for the *PR-13* class, as shown in Table 1. A comparison of the distribution of non-redundant sequences in the three species revealed that the NBS-LRR family was the most frequent one in all cases, while the LRR-kinase class was the least represented in all studied organisms (Figure 2). Moreover, it was observed that while *Arabidopsis* presented a higher number of *R* gene candidates, *Medicago* matched the high number of *PR* genes. In both cases, soybean presented the lowest number of matches (Figure 3A).

The three most represented *R* and *PR* genes in all species were the same, with *Xa21*, *EFR* and *Pti6* representing *R* genes and *PR-2*, *PR-7* and *PR-9* representing *PR* genes. Due to

this abundance, both *Xa21* and the *PR-2* genes were selected for the construction of a dendrogram and expression analysis. Matching of *Xa21* and *PR-2* candidates in soybean, *Medicago* and *Arabidopsis* did not follow a regular distribution pattern, since soybean presented fewer matches for both genes, and most of the *Xa21* candidate sequences were found in *Medicago*, whereas most *PR-2* candidates were found in *Arabidopsis* (Figure 3B).

Among the 310 *PR* genes of soybean only 40 matched with more than one seed sequence, all the others being exclusive to a given *PR* gene family. On the other hand, almost all *R* genes matched sequences that aligned with more than one probe, requiring manual sorting. Exceptions occurred only with respect to *RAR*, *RIN*, *P*, *WRKY29*, and *Xa21,* which aligned in most cases with exclusive sequences.

**Phylogenetic analysis of *Xa21* and *PR-2* genes**

Dendrograms generated for *Xa21* and *PR-2* genes using the soybean sequences and orthologs clearly divided dicots and monocots into distinct clades (Figure 4). In the *Xa21* analysis, the fern *Selaginella moellendorffii* was placed in a basal position from which the two branches representing monocots and dicots emerged (Figure 4A). The monocots group included members of the Poaceae family in one branch, with a bootstrap CI of 95%, associated in the same branch with the palm *Elaeis guineensis*. Regarding the dicot group, it was observed that both Fabaceae members (*G. max* and *M. truncatula*) were positioned together, while the other branch included members of the suborder Eurosidae I (*Vitis vinifera* and *Ricinus communis*), together with *A. thaliana,* a member of the Eurosidae II suborder.

Considering the *PR-2* dendrogram (Figure 4B), the grasses (Poaceae represented by rice and maize) occupied a basal position, from which a clade containing two monocots, ginger

(*Zingiber officinale*) and banana (*Musa paradisiaca*), emerged. Moreover, a large clade containing all dicots was split into two subclades that behaved as merophyletic groups. For example, tobacco (*Nicotiana tabacum*) and coffee (*Coffea arabica*), members of the Asterid order, remained together, but potato (*Solanum tuberosum*) of the same order was positioned on another branch. Soybean and *Medicago* were also positioned in separate subclades.

**Expression pattern of *R* (*Xa21*) and *PR* (*PR-2*) genes in the soybean transcriptome**

From the 26,653 reads identified, an *in silico* expression assay was carried out considering transcripts from both genes *Xa21* (2,980 reads) and *PR-2* (1,099 reads). This allowed identifying their prevalence and normalizing their distribution among the tissues and conditions represented in the 65 different libraries. Graphic illustrations of these comparisons are available as Figures S1 and S2 (Supplementary Material).

The analysis of their expression pattern in soybean, obtained from normalized data, revealed that all libraries presented almost the same number of reads. The most representative library was from seed tissues (SD), presenting 10% of the identified reads. Expression in tissues from leaves (LV), roots (R) and flowers (F) presented similar expression, representing 9% of all reads in each tissue. The remaining tissues also presented significant expression (ranging from 5% to 8%), except in the case of libraries made from tissues submitted to the nematode *Meloidogyne javanica* (MJ), where no reads were identified.

**Expression considering the SuperSAGE libraries**

BLASTn results revealed that 944 soybean EST candidates aligned with 1,553 SuperSAGE tags when considering a cut-off value of $\leq e^{-5}$. Among all tags, 1,072 aligned with

the *R* gene candidates from different classes, with emphasis on the NBS-LRR class. Additionally, 481 tags aligned with *PR* gene candidates, most of them with the *PR-9* secretory peroxidase family (Figure 5). Data concerning sequence-tag association are available as supplementary material (Tables S2, S3 and S4). The best results were obtained for comparison 1 (BR-16, drought susceptible stressed vs. negative control), which matched 613 non-redundant tags, while 465 were found for comparison 2 (Embrapa-48, drought tolerant stressed vs. negative control), and for comparison 3 (PI561356 fungus resistant stressed vs. negative control) 475 SuperTags were represented (Figure 5). It is noteworthy that many tags matched in more than one comparison.

### Anchoring soybean *R* and *PR* genes in *Medicago* virtual chromosomes

The alignment of 59 soybean genes against the *Medicago* virtual chromosomes revealed 1,253 sites in all nine chromosomes, also including sub-telomeric regions (Figure 6). 58 genes presented similarities with distinct segments in the same chromosome or appeared twice in distinct chromosomes. Only the *PR-1* sequence anchored in an exclusive chromosome (2).

The highest number of anchored genes was found in chromosome 8, matching 32 of the 59 genes in 85 sites. On the other hand, chromosome 6 presented the lowest number of anchored genes (12). Nonetheless, this chromosome presented the highest number of duplications, matching 228 sites, most of them in tandem positions. Such tandem repetitions could be also observed in three sites of chromosome 3. The lowest gene density was observed in the long arm of chromosome 3. Syntenic regions were evident in chromosomes 2 and 4 (Figure 6).

Several sequences clustered along the genome, with some chromosomes rich in resistance genes, especially chromosomes 2, 7, 8 and 9, with at least four distinct genes in very close positions. These blocks of genes always matched *R* genes, while *PR* genes generally appeared in the same chromosomes in distinct sites.

## Discussion

The 1,066 soybean sequences resulting from tBLASTx alignments confirmed the excellent coverage that the existing GENOSOJA databank comprises, including the most important representatives from different gene families.

Legumes are plants known to be able to withstand many kinds of stresses, including rapid climate changes, drought tolerance, exposure to diseases and pests, water logging and flooding (Cannon *et al.*, 2009), which could explain the higher number of *PR* genes encountered in *Medicago* in comparison to *Arabidopsis*, since these families of genes can be activated by different kinds of biotic or abiotic stress (Glombitza *et al.*, 2004). The low number of *R* and *PR* gene candidates found in soybean is curious when compared to *Arabidopsis* and *Medicago*, since these have smaller genomes (157 Mb and 583 Mb respectively) than that of *G. max* (1,115 Mb). This may be due to the analyzed sample, which was restricted to expressed sequence tags, whereas the databases of both *Arabidopsis* and *Medicago* are larger. Previous studies on legumes showed that despite the relatively large difference in genome sizes of soybean and *Medicago*, gene densities are similar, indicating that a given *Medicago* region is likely to correspond well with two soybean regions (Mudge *et al.*, 2005). This leads us to believe that additional expression assays in soybean may reveal important genes that are expressed under very specific conditions.

The number of soybean clusters that aligned with more than one *R* gene seed sequence is not surprising. Similar results were observed in previous studies regarding *R* genes of eucalyptus (Barbosa-da-Silva *et al.*, 2005) and sugarcane (Wanderley-Nogueira *et al.*, 2007). This occurs due to the common domains shared by *R* genes, as for example the LRR domain that is present in the LRR, NBS-LRR and LRR-kinase gene families, facilitating alignments with more than one gene. This is rarer when considering *PR* gene categories that are more distinct in structure and function (Kitajima and Sato, 1999), as also observed herein. A higher number of sequences matching NBS-LRR families, when compared to other classes, was also reported by Barbosa-da-Silva *et al.* (2005) and Wanderley-Nogueira *et al.* (2007), confirming the general observation that most *R* genes are members of this class.

Dendrograms generated from these data revealed a similar picture in both gene classes selected (*Xa21* and *PR-2*). In the case of *Xa21*, the positioning of *Selaginella moellendorffii* as an outgroup was expected, since this species figures as a member of an ancient vascular plant lineage that first appeared 400 million years ago, and thus represents a basal node on the plant evolutionary tree (Weng *et al.*, 2008). The analysis of the *Xa21* orthologs from different species reflected their relationship according to classic taxonomy. Lilliopsida class (monocots) appeared as a monophyletic group uniting on the same branch *Oryza sativa*, *Zea mays* and *Sorghum bicolor,* which are all annual cereal grains of the Poaceae family, while the palm *Elaeis guineensis* (Arecaceae) was positioned on another branch. Considering the Magnoliopsida (dicots), the same occurred, since *Medicago* and soybean, both legumes and members of Fabaceae, appeared in a subclade, separated from the remaining species. *R* genes are considered fast evolving, due to their co-evolution with specific pathogens (Michelmore and Meyers, 1998). In the case of *Xa21* the most polymorphic region is its extracellular LRR

domain, which is responsible for pathogen specificity (Ellis *et al.*, 2000), defining the relationships of the dendrogram presented here.

The *PR-2* dendrogram topology showed two main clades, as expected, monocots and dicots. The grouping of monocots followed the taxonomic relationship, segregating *Musa* and *Zingiber* (Zingiberales) from *Oryza* and *Zea* (Poaceae). It was possible to identify that a symplesiomorphic character united all dicots, reflecting their common origin. Moreover, considering the Magnoliopsida group, the evolutionary model of the *PR-2* class seemed to follow a synapomorphic pattern, leading to their diversification in different groups comprising families and orders, this probably reflecting divergent processes regarding this *PR* gene.

The studied organisms presented different centers of origin, habitats and cycles of life, as well as tolerance, resistance and sensitivity to diverse kinds of biotic and abiotic stresses. Nonetheless, from an overall perspective and considering the position of different species in the dendrograms, it is evident that both *Xa21* and *PR-2* pathways genes were present in a common ancestor of the angiosperms, since they appear relatively conserved in different plant groups.

Many *PR* genes are constitutively expressed in given plant tissues (Velazhahan and Muthukrishnan, 2003; Liu *et al.*, 2004), suggesting a link between biotic and abiotic stresses and indicating that at least some members of the PR proteins play important roles in plant development, besides their role in defense responses. This fact may explain why the expression of *PR-2* gene can be observed at a basal level in almost all tissues, as seen when considering their frequencies in the soybean libraries.

Studies carried out by Li *et al.* (2008) and Libault *et al.* (2010) revealed consistent differences in gene expression patterns among diverse tissues, especially between roots and

aerial tissues, but also revealed similarities between expression levels in tissues such as flowers and leaves, corroborating our results. The most represented library was for seeds, including different development stages, which is not surprising, since previous evaluations also revealed that the soybean grain contained the vast majorities of expressed genes and regulatory sequences in the plant (Cannon et al., 2009). In the case of the *PR-2* protein, it is interesting to note that previous evaluations carried out by Leubner-Metzger (2005) in tobacco suggest that this gene could play a role in seed germination. Furthermore, the expression of both genes was also increased in leaves, roots and flowers, confirming their prevalence in developing tissues.

As mentioned above, abiotic stress is able to trigger diverse plant responses. After an initial massive distribution of energy triggered by stress, a wide array of defense mechanisms is activated by *R* genes, inducing a signal cascade and increased *PR* gene transcription (Vergne *et al.*, 2010). This may justify the considerable amount of soybean SuperSAGE tags related to these genes among the three comparisons considered, with considerable representation in both biotic and abiotic (drought) conditions, as well as in the negative controls, with many tags represented in more than one treatment. The high number of tags that matched with BR-16 drought susceptible library vs. control could be explained by the ability of the plant to continue expressing genes related to systemic acquired resistance as a consequence of contact with any kind of previous stress, a crosstalk previously reported for other plants (Durrant and Dong, 2004; Kido *et al.*, 2011). Comparing the distribution between *R* and *PR* genes, both were representative with 1,072 tags matching *R* genes and 481 tags matching *PR* candidates, indicating that additional analytical efforts regarding the SuperSAGE

candidates will reveal not only associations with specific situations, but also allelic differences important in the definition of biotic and abiotic stress responses.

Flowering plants originated approximately 200 million years ago (Wilkstrom *et al.*, 2001) and subsequently diverged into several lineages. Legumes are an old family believed to have originated approximately 54 Mya (Lavin *et al.*, 2005). Soybean and other papilionoid legumes show evidence of an older shared duplication and probably soybean underwent polyploidy 13 Mya (Shoemaker *et al.*, 2006). These duplications are widely evident, both in number of similar duplicated genes and in large areas of synteny between chromosomal regions. Previous evidence indicates extensive similarities in gene densities and distribution among soybean and *Medicago*, inferring that a given *Medicago* region is likely to correspond well with two soybean regions (Mudge *et al.*, 2005). This evidence suggests that *Medicago* could represent "a simplified draft" of the soybean gene distribution, making an evaluation regarding *R* and *PR* soybean ortholog distribution in this crop most desirable. Hence, it is not surprising that all identified soybean *R* and *PR* transcripts appeared anchored in 1,253 sites in all segments of *Medicago* virtual chromosomes.

The rich *R* gene regions found in chromosomes 2, 7, 8 and 9 confirm previous observations that most resistance genes reside in clusters (Kanazin *et al.*, 1996), as reported in maize (Dinesh-Kumar *et al.*, 1995), lettuce (Maisonneuve *et al.*, 1994), oat (Rayapati *et al.*, 1994) and flax (Ellis *et al.*, 1995). The formation of gene clusters is in general associated with a common ancestor, and the diversification of these genes is the result of duplication processes followed by diversification due to pathogen or environmental pressure.

Clustering of *R* genes corroborates the existing theory that a common genetic mechanism involving duplication has been responsible for the evolution and diversification of

this gene superfamily (Hulbert *et al.*, 2001). The four clusters presented similarities with distinct segments in the same chromosome, probably reflecting tandem gene duplication mechanisms. Such duplicated copies tend to diverge by acquiring additional mutations and may specialize or optimize to play slightly different roles (Alberts *et al.*, 2002).

Regarding the duplicated segments considering the entire genome, 58 genes could be identified in at least two distinct chromosomes. Unlike tandem duplications, repetitions in distinct chromosomes resulted from events of duplication followed by translocations and sequence divergence, also allowing functional diversification (Wendell, 2000; Thiel *et al.*, 2009). There is also evidence that transposition outbreaks could be activated by severe environmental biotic or abiotic stress.

Still regarding the duplication event analysis, a large in tandem repetition was evident in both chromosomes 3 and 6, represented by the genes *Xa1*/*I2* and *RRS1,* respectively. Previous reports suggested that once duplicated, genes in tandem repetitions may expand rapidly through events of unequal crossing over, since the character could confer advantage to the organism (Alberts *et al.*, 2002), in this case a higher diversity of genes associated with resistance and stress response. This evidence supports 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 (Dafny-Yelin and Tzfira, 2007), as has been traditionally done.

In conclusion, the here identified sequences represent valuable resources for the soybean breeding program, allowing their use in biotechnological approaches, with emphasis on transgenes. They are also valuable for mapping purposes, considering the putative distribution

here uncovered when considering available distribution of genes known from the *Medicago* genome.

Considering gene diversity revealed especially by the SuperSAGE approach, their association with specific responses to biotic or abiotic stress conditions may reveal important gene variants for germplasm screening in the search for new accessions useful for breeding purposes, especially in association with marker assisted selection (MAS), saving decades of laborious research.

## Acknowledgments

The authors would like to thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco), and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for their financial support.

## References

Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P (2002) Molecular Biology of the Cell. 4th edition, Garland Publishing Company, New York & London, 1616 pp.

Altschul SF, Gish W, Miller W and Myers E (1990) Basic local alignment search tool. J Mol Biol 215:403-410.

Ashfield T, Bocian A, Held D, Henk AD, Marek LF, Danesh D, Penuela S, Meksem K, Lightfoot DA, Young ND, *et al* (2003) Genetic and physical localization of the

soybean *Rpg1-b* disease resistance gene reveals a complex locus containing several tightly linked families of NBS-LRR genes. Mol Plant Microbe Interact 16:817-826.

Atici O and Nalbantoğlu B (2003) Antifreeze proteins in higher plants. Phytochemistry 64**:**1187-1196.

Barbosa-da-Silva A, Wanderley-Nogueira AC, Silva RRM, Belarmino LC, Soares-Cavalcanti, NM and Benko-Iseppon AM (2005) *In silico* survey of resistance (*R*) genes in *Eucalyptus* transcriptome. Genet Mol Biol 28:562-574.

Benko-Iseppon AM, Galdino SL, Calsa Júnior T, Kido EA, Tossi A, Belarmino LC and Crovella S (2010) Overview of plant antimicrobial peptides. Curr Prot Pept Sci 11:181-188.

Bent AF (1996) Plant disease resistance genes: Function meets structure. Plant Cell 8:1751-1771.

Bolton M (2009) Primary metabolism and plant defense – fuel for the fire. Mol Plant Microbe Interact 22:487-497.

Bonas U and Anckerveken GV (1999) Gene-for-gene interactions: Bacterial avirulence proteins specify plant disease resistance. Curr Opin Plant Biol 2:94-98.

Bonasera JM, Kim JF and Beer SV (2006) *PR* genes of apple: Identification and expression in response to elicitors and inoculation with *Erwinia amylovora.* BMC Plant Biol 6:23-34.

Cannon SB, May GD and Jackson SA (2009) Three sequenced legume genomes and many crop species: Rich opportunities for translational genomics. Plant Physiol 151:970-977.

Chester KS (1933) The problem of acquired physiological immunity in plants. Quart Rev Phytopathol 42:185-209.

Dafny-Yelin M and Tzfira T (2007) Delivery of multiple transgenes to plant cells. Plant Physiol 145:1118-1128.

Dinesh-Kumar SP, Whitham S, Choi D, Hehl R, Corr C and Baker B (1995) Transposon tagging of tobacco mosaic virus resistance gene *N*:I its possible role in the TMV-N-mediated signal transduction pathway. Proc Natl Acad Sci USA 92:4175-4180.

Dixon MS, Jones DA, Keddie JS, Thomas CT, Harrison K and Jones JDG (1996) The tomato *Cf2* disease resistance locus comprises two functional genes encoding leucine rich repeats proteins. Cell 84:451-459.

Durrant WE and Dong X (2004) Systemic acquired resistance. Annu Rev Plant Pathol 42:185-209.

Eisen MB, Spellman PT, Brown PO and Botstein B (1998) Cluster analysis and display of genome-wide expression patterns. Genetics 25:14863-14868.

Ellis J and Jones D (2000) Structure and function of proteins controlling strain-specific pathogen resistance in plants. Curr Opin Plant Biol 1:288-293.

Ellis J, Lawrence GJ, Finnegan EJ and Anderson PA (1995) Contrasting complexity of two rust resistance loci in flax. Proc Natl Acad Sci USA 92:4185-4188.

Ellis J, Dodds P and Pryor T (2000) Structure, function and evolution of plant disease resistance genes. Curr Opin Plant Biol 3:278-284.

Gaffney T, Friedrich L, Vernooij B, Negrotto D, Nye G, Ukness S, Ward E, Kessman H and Ryals J (1993) Requirement of salicylic acid for the induction of systemic acquired resistance. Science 261:754-756.

Glombitza S, Dubuis P-H, Thulke O, Welzl G, Bovet L, Götz M, Affenzeller M, Geist B, Hehn A, Asnaghi C, *et al* (2004) Crosstalk and differential response to abiotic and biotic stressors reflected at the transcriptional level of effector genes from secondary metabolism. Plant Mol Biol 54:817-835.

Griffith M and Yaish MWF (2004) Antifreeze proteins in overwintering plants: A tale of two activities. Trends Plant Sci 9:399-405.

Hammond-Kosack KE and Jones JDG (1997) Plant disease resistance genes. Annu Rev Plant Physiol 48:575-607.

Hon WC, Griffith M, Mlynarz A, Kwok YC and Yang DSC (1995) Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. Plant Physiol 109:879-889.

Hulbert SH, Webb CA, Smith SM and Sun Q (2001) Resistance gene complexes: Evolution and utilization. Annu Rev Phytopathol 39:285-312.

Joahal GS and Briggs SP (1992) Reductase activity encodes by the *Hm1* resistance gene in maize. Science 198:985-987.

Kanazin V, Marek LF and Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. Proc Natl Acad Sci USA 93:11746-11750.

Kido EA, Barbosa PK, Ferreira Neto JCR, Pandolfi V, Houllou-Kido LM, Crovella S and Benko-Iseppon AM (2011) Identification of plant protein kinases in response to abiotic and biotic stresses using SuperSAGE. Curr Prot Pept Sci 12:643-656.

Kitajima S and Sato F (1999) Plant pathogenesis-related proteins: Molecular mechanisms of gene expression and protein function. J Biochem 125:1-8.

Lavin M, Herendeen PS and Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Syst Biol 54:575-594.

Lawrence GJ, Finnegan EJ, Ayliffe MA and Ellis JG (1995) The *L6* gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene *RPS2* and the tobacco viral resistance gene *N*. Plant Cell 7:1195-1206.

Leubner-Metzger G (2005) β-1,3-glucanase gene expression in low-hydrated seeds as a mechanism for dormancy release during tobacco after-ripening. Plant J 41:133-145.

Li L, He H, Zhang J, Wang X, Bai S, Stolc V, Tongprasit W, Young ND, Yu O and Deng XW (2008) Transcriptional analysis of highly syntenic regions between *Medicago truncatula* and *Glycine max* using tiling microarrays. Genome Biol 9:R57.

Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, He J, Xu D, May G and Stacey G (2010) An integrated transcriptome atlas of the crop model *Glycine max* and its use in comparative analyses in plants. Plant J 63:86-99.

Liu B, Zhang S, Zhu X, Yang Q, Wu S, Mei M, Mauleon R, Leach J, Mew T and Leung H (2004) Candidate defense genes as predictors of quantitative blast resistance in rice. Mol Plant Microbe Int 17:1146-1152.

Maisonneuve B, Bellec Y, Anderson P and Michelmore RW (1994) Rapid mapping of two genes for resistance to downy mildew from *Lactuca serriola* to existing clusters of resistance genes. Theor Appl Genet 89:96-104.

Matsumura H, Kruger DH, Kahl G and Terauchi R (2008) SuperSAGE: A modern platform for genome-wide quantitative transcript profiling. Curr Pharm Biotechnol 9:368-374.

Melotto M, Coelho MF, Pedrosa-Harand A, Kelly JD and Camargo LE (2004) The anthracnose resistance locus *Co-4* of common bean is located on chromosome 3 and contains putative disease resistance-related genes. Theor Appl Genet 109:690-699.

Metzler MC, Cutt JR and Klessig DF (1991) Isolation and characterization of a gene encoding a PR-1 like protein from *Arabidopsis thaliana*. Plant Physiol 96:346-348.

Michelmore RW and Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113- 1130

Mindrinos M, Katagiri F, Yu GL and Ausubel FM (1994) The *Arabidopsis thaliana* disease resistance gene encodes a protein containing a nucleotide-binding site and leucine rich repeats. Cell 78:1089-1099.

Mudge J, Cannon SB, Kalo P, Oldroyd GE, Roe BA, Town CD and Young ND (2005) Highly syntenic regions in the genomes of soybean, *Medicago truncatula* and *Arabidopsis thaliana*. BMC Plant Biol 5:e15.

Nanda AK, Andrio E, Marino D, Pauly N and Dunand C (2010) Reactive Oxygen Species during plant-microorganism early interactions. J Integr Plant Biol 52:195-204.

Nurnberg T and Brunner F (2002) Innate immunity in plants and animals: Emerging parallels between the recognition of general elicitors and pathogen-associated molecular patterns. Curr Opin Plant Biol 5:318-324.

Page RD (1996) Treeview program, version 1 6 1. Comp Appl Biosci 12:357-358.

Rayapati PJ, Lee M, Gregory JW and Wise RP (1994) 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 89:831-837.

Salmeron JM, Oldroyd GED, Romens CMT, Scofield SR, Kim HS, Lavelle DT, Dahlbeck D and Staskawicz BJ (1996) Tomato *Prf* is a member of the leucine rich repeats class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. Cell 86:123-133.

Shoemaker RC, Schlueter J and Doyle JJ (2006) Paleopolyploidy and gene duplication in soybean and other legumes. Curr Opin Plant Biol 9:104-109.

Song WY, Pi LY, Wang GL, Gardner J, Holsten T and Ronald PC (1997) Evolution of the rice *Xa21* disease resistance genes family. Plant Cell 9:1279-1287.

Song WY, Wang GL, Kim HS, Pi LY, Gardner J, Wang B, Holsten T, Zhai WX, Zhu LH, Fauquet C, *et al* (1995) A receptor kinase-like protein encoded by the rice disease resistance gene *Xa21*. Science 270:1804-1806.

Sparla F, Rotino L, Valgimigli MC, Pupillo P and Trost P (2004) Systemic resistance induced by benzothisdizole in pear inoculated with the agent of fire blight. Sci Hortic 101:269-279.

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.

Tang X, Xie M, Kim YJ, Zhou J, Klessing DF and Martin GB (1999) Overexpression of *Pto* activates defense responses and confers broad resistance. Plant Cell 11:15-29.

Thiel T, Graner A, Waugh R, Grosse I, Close TJ and Stein N (2009) Evidence and evolutionary of ancient whole-genome duplication in barley predating the divergence from rice. BMC Evol Biol 9:209-227.

Tornero P, Gadea J, Conejero V and Vera P (1997) Two *PR-1* genes from tomato are differentially regulated and reveal a novel mode of expression for a pathogenesis-related gene during the hypersensitive response and development. Plant Microbe Interact 10:624-634.

van-Loon LC, Geraats BPJ and Linthorst HJM (2006) Ethylene as a modulator of disease resistance in plants. Trends Plant Sci 11:184-191.

van-Loon LC, Pierpoint WS, Boller T and Conejero V (1999) Recommendations for naming plant pathogenesis-related proteins. Plant Mol Biol Rep 12:245-264.

Velazhahan R and Muthukrishnan S (2003) Transgenic tobacco plants constitutively overexpressing a rice thaumatin-like protein (PR-5) show enhanced resistance to *Alternaria alternata*. Plant Biol 47:347-354.

Vergne E, Grand X, Ballini R, Chalvon V, Saindrenan P, Tharreau D, Nottéghem J-L and Morel J-B (2010) Preformed expression of defense is a hallmark of partial resistance to rice blast fungal pathogen *Magnaporthe oryzae*. BMC Plant Biol 10:e206.

Wanderley-Nogueira AC, Mota N, Lima-Morais D, Silva LCB, Silva AB and Benko-Iseppon AM (2007) Abundance and diversity of resistance (*R*) genes in the sugarcane transcriptome. Genet Mol Res 6:866-889.

Wang GL, Holsten TE, Song WY, Wang HP and Ronald PC (1995) Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa21* disease resistance locus. Plant J 7:525-533.

Wendell J (2000) Genome evolution in polyploids. Plant Mol Biol 42:225-249.

Weng JK, Banks JA and Chapple C (2008) Parallels in lignin biosynthesis: A study in *Selaginella moellendorffii* reveals convergence across 400 million years of evolution Comm Int Biol 1:20-22.

Wilkstrom N, Savolainen V and Chase MW (2001) Evolution of the angiosperms: Calibrating the family tree. Proc Soc Biol Sci 268:2211-2220.

Zeier J, Pink B, Mueller MJ and Berger S (2004) Light conditions influence specific defense responses in incompatible plant-pathogen interactions: Uncoupling systemic resistance from salicylic acid and *PR-1* accumulation. Planta 219:673-683.

**Internet Resources**

Expert Protein Analysis System (Expasy), http://expasy.org.uk (August 18, 2010).

Medicago sequencing resource website, http://www.medicago.org (October 19, 2010).

The Arabidopsis Information Resource (TAIR), http://www.arabidopsis.org (September 1, 2010).

The Brazilian Soybean Genome Consortium (GENOSOJA), http://bioinfo03.ibi.unicamp.br/soja (August 8, 2010).

The Institute for Genomic Research (TIGR), http://plantta.jcvi.org (August 8, 2010).

**Supplementary Material** – the following online material is available for this article:

Figure S1 - *Xa21* expression profile in 16 different libraries from GENOSOJA.

Figure S2 - *PR-2* expression profile in 16 different libraries from GENOSOJA.

Table S1 - Accession number of reference *PR* genes used as seed sequences.

Table S2 - Number of SuperSAGE tags per comparison.

Table S3: SuperSAGE tags that matched genes.

Table S4: Number of tag repetitions in comparisons matching *R* and *PR* genes.

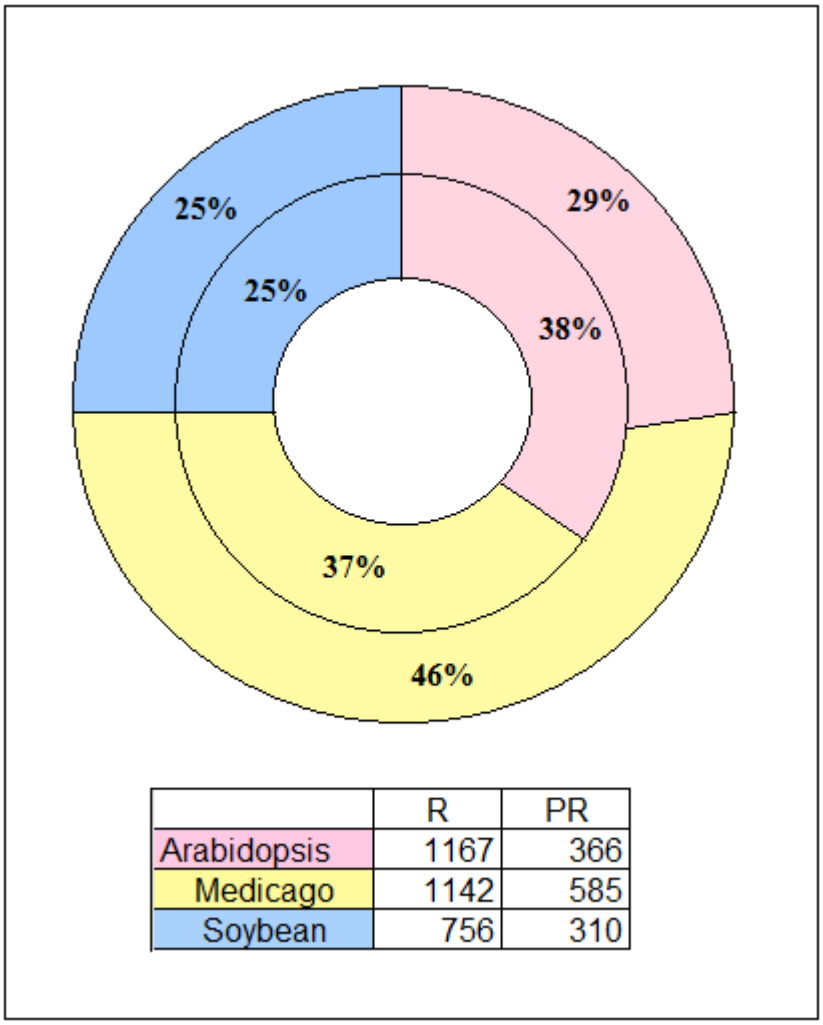
This material is available as part of the online article from http://www.scielo.br/gmb.

Table 1: Soybean clusters matching results to each procured *R* and *PR* genes. Showing number of matches to each seed-sequences, e-value, score, size in nucleotide (nt) and amino-acids (aa), presence of conserved domains and number of matches in soybean, medicago and arabidopsis. Abbreviations: (c)=Complete; (i)=Incomplete; M= *Medicago truncatula*; A= *Arabidopsis thaliana*.

| | Soybean | | | | | | | M | A |
|---|---|---|---|---|---|---|---|---|---|
| Gene | Best match | e-value | score | Size (nt) | Size (aa) | Conserved Domain (c/i) | # matches | # matches | # matches |
| *PR1* | Contig 5043 | 7,00E-47 | 181 | 498 | 165 | SCP (c) | 8 | 19 | 22 |
| *PR2* | Contig 9520 | 1,00E-102 | 369 | 1047 | 348 | Glyco-Hydro (c ) | 86 | 214 | 95 |
| *PR3* | Contig 5557 | 4,00E-48 | 187 | 957 | 318 | Chitinase (c ) | 7 | 21 | 15 |
| *PR4* | Contig 10145 | 2,00E-67 | 250 | 636 | 211 | Chitin binding/Barwin(c ) | 2 | 14 | 2 |
| *PR5* | Contig 29866 | 5,00E-60 | 226 | 1041 | 345 | Thaumatin (c ) | 21 | 36 | 29 |
| *PR6* | Contig 5043 | 1,00E-46 | 181 | 495 | 164 | SCP (c ) | 11 | 17 | 23 |
| *PR7* | Contig 66 | 5,00E-141 | 481 | 2283 | 760 | Peptidase/Subtilisin (c ) | 82 | 97 | 50 |
| *PR8* | Contig 14006 | 4,00E-89 | 232 | 894 | 297 | hevamine (c) | 11 | 22 | 1 |
| *PR9* | Contig 1796 | 1,00E-120 | 428 | 978 | 325 | Secretory peroxidase(c ) | 31 | 46 | 66 |
| *PR10* | Contig 4865 | 6,00E- 26 | 112 | 410 | 160 | Bet v 1(c ) | 18 | 18 | 34 |
| *PR11* | Contig 5806 | 9,00E-79 | 289 | 1098 | 365 | plant chitinase class V (c ) | 1 | 11 | 9 |
| *PR12* | Contig 13869 | 1,00E-09 | 58 | 291 | 96 | Gamma-thionin (i) | 1 | 15 | 8 |
| *PR13* | No match | - | - | - | - | - | - | - | 4 |
| *PR14* | Contig 13114 | 6,00E-18 | 86 | 357 | 118 | lipid-transfer protein (c ) | 18 | 36 | 16 |
| *PR15* | SJ01-E1-UK1-089-G01-UC.F | 1,00E-48 | 188 | 660 | 219 | Cupin2 (c ) | 27 | 47 | 37 |
| *PR16* | Contig 13716 | 2,00E-59 | 223 | 666 | 221 | Cupin2 (c ) | 27 | 51 | 37 |
| *PR17* | Contig 25189 | 2,00E-73 | 271 | 678 | 225 | basic secretory proteins (c ) | 2 | 1 | 5 |
| *Pto* | Contig 5707 | 2,00E-143 | 505 | 2502 | 833 | Ser-Thre Kinase (i ) | 238 | 239 | 248 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Prf* | Contig 5666 | 4,00E-34 | 142 | 2736 | 920 | P-loop NTPase domain (c ) | 5 | 25 | 49 |
| *Pti4* | SJ05-E1-S06-021-E06-UC.F | 6,00E-33 | 136 | 825 | 274 | DNA-binding domain (c ) | 89 | 90 | 119 |
| *Pti5* | Contig 25338 | 6,00E-45 | 176 | 645 | 214 | DNA-binding domain (c ) | 70 | 70 | 89 |
| *Pti1* | SJ05-E1-UK1-024-H07-UC.F | 2,00E-33 | 138 | 759 | 252 | DNA-binding domain (c ) | 104 | 112 | 138 |
| *Pti6* | Contig 10050 | 2,00E-146 | 514 | 1086 | 361 | Tyr Kinase (i ) | 248 | 249 | 249 |
| *RAR1* | Contig 27196 | 1,00E-76 | 281 | 672 | 223 | CHORD superfamily (c ) | 1 | 2 | 1 |
| *RIN4* | Contig 20845 | 7,00E-25 | 109 | 741 | 246 | AvrRpt-cleavage (c ) | 2 | 8 | 1 |
| *RPM1* | Contig 25089 | 5,00E-29 | 125 | 2781 | 926 | P-loop NTPase - LRR (c ) | 14 | 73 | 90 |
| *RPS2* | SJ01-E1-L06-046-G05-UC.F | 7,00E-10 | 62 | 2538 | 845 | P-loop NTPase - LRR (c ) | 4 | 36 | 90 |
| *PBS1* | Contig 26006 | 3,00E-132 | 467 | 1152 | 383 | Protein Kinase (c ) | 239 | 247 | 251 |
| *RPS5* | Contig 10273 | 1,00E-17 | 87 | 1941 | 646 | P-loop NTPase - LRR (c ) | 5 | 36 | 65 |
| *MLA10* | SJ18-P1-S12-046-B20-UC.F | 4,00E-07 | 51 | 913 | 305 | P-loop NTPase - LRR (c ) | 0 | 21 | 30 |
| *L6* | Contig 16939 | 5,00E-55 | 210 | 3198 | 1065 | TIR- P-loop-LRR (c ) | 24 | 123 | 171 |
| *RRS1* | Contig 14438 | 1,00E-30 | 107 | 2211 | 736 | P-loop NTPase - LRR (c ) | 102 | 142 | 239 |
| *RPS4* | Contig 16939 | 1,00E-35 | 148 | 3198 | 1065 | TIR- P-loop-LRR (c ) | 50 | 198 | 226 |
| *Xa1* | Contig 5507 | 5,00E-63 | 238 | 3609 | 1202 | P-loop NTPase - LRR (c ) | 17 | 108 | 91 |
| *Hrt* | Contig 16939 | 3,00E-54 | 207 | 3198 | 1065 | TIR- P-loop-LRR (c ) | 61 | 208 | 181 |
| *Mi1* | Contig 12827 | 2,00E-08 | 58.2 | 2733 | 910 | TIR- P-loop-LRR (c ) | 1 | 29 | 50 |
| *BS2* | Contig 10273 | 3,00E-14 | 76 | 1941 | 646 | P-loop NTPase - LRR (c ) | 9 | 68 | 135 |
| *GPA2* | SJ14-E1-S07-021-C03-UC.F | 1,00E-22 | 104 | 2733 | 910 | P-loop NTPase - LRR (c ) | 10 | 50 | 123 |
| *RX1* | Contig 5666 | 4,00E-39 | 159 | 2736 | 920 | P-loop NTPase - LRR (c ) | 14 | 61 | 112 |
| *Pi-ta* | SJ14-E1-S07-021-C03-UC.F | 1,00E-23 | 107 | 2733 | 910 | P-loop NTPase - LRR (c ) | 2 | 17 | 62 |
| *I2* | Contig 5507 | 8,00E-64 | 241 | 3609 | 1202 | P-loop NTPase - LRR (c ) | 22 | 109 | 108 |
| *RPP8* | SJ14-E1-S07-021-C03-UC.F | 3,00E-19 | 94 | 2733 | 910 | P-loop NTPase - LRR (c ) | 11 | 71 | 129 |
| *HERO* | SJ14-E1-S07-021-C03-UC.F | 1,00E-08 | 58 | 2733 | 910 | P-loop NTPase - LRR (c ) | 5 | 39 | 78 |
| *L6* | no match | - | - | - | - | - | - | - | - |
| *RPP13* | SJ14-E1-S07-021-C03-UC.F | 1,00E-23 | 107 | 2733 | 910 | P-loop NTPase - LRR (c ) | 2 | 51 | 77 |
| *RP1* | Contig 10273 | 2,00E-26 | 86 | 1941 | 646 | P-loop NTP-ase (c ) | 14 | 71 | 69 |
| *N* | Contig 16939 | 2,00E-51 | 198 | 3198 | 1065 | TIR- P-loop-LRR (c ) | 64 | 196 | 171 |
| *P* | Contig 20164 | 3,00E-11 | 64 | 585 | 194 | Dirigent super family (c ) | 17 | 37 | 18 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *M* | no match | - | - | - | - | - | - | - | - |
| *WRKY25* | Contig 3637 | 4,00E-65 | 244 | 1761 | 586 | WRKY superfamily 2 (c ) | 68 | 52 | 77 |
| *WRKY33* | Contig 3637 | 7,00E-78 | 287 | 1761 | 586 | WRKY superfamily 1 (c ) | 71 | 58 | 85 |
| *WRKY29* | SJ01-E1-L08-116-F02-UC.F | 5,00E-21 | 97.4 | 768 | 255 | WRKY superfamily 2 (c ) | 28 | 22 | 37 |
| *Cf2* | Contig 17295 | 1,00E-71 | 267 | 3132 | 1043 | Multiple LRRs (c ) | 249 | 250 | 266 |
| *Cf4* | Contig 14446 | 4,00E-40 | 162 | 2256 | 751 | Multiple LRRs (c ) | 116 | 208 | 249 |
| *Cf5* | Contig 6299 | 1,00E-39 | 160 | 2955 | 984 | Multiple LRRs (c ) | 123 | 207 | 219 |
| *Cf9* | Contig 14446 | 5,00E-53 | 204 | 2256 | 751 | Multiple LRRs (c ) | 107 | 188 | 267 |
| *Xa21* | Contig 439 | 3,00E-69 | 259 | 2913 | 970 | LRRs - Kinase (c ) | 251 | 249 | 247 |
| *FLS2* | Contig 6299 | 6,00E-66 | 233 | 2955 | 984 | LRRs - Kinase (c/i ) | 174 | 251 | 249 |
| *EFR* | Contig 439 | 2,00E-59 | 227 | 2913 | 970 | LRRs - Kinase (c ) | 250 | 239 | 253 |




| | R | PR |
|---|---|---|
| Arabidopsis | 1167 | 366 |
| Medicago | 1142 | 585 |
| Soybean | 756 | 310 |

Figure 1: Percentage of *R* and *PR* genes encountered in soybean, arabidopsis and medicagotranscriptomes. *R*-genes are represented in the outer circle and *PR*-genes in the inner circle.

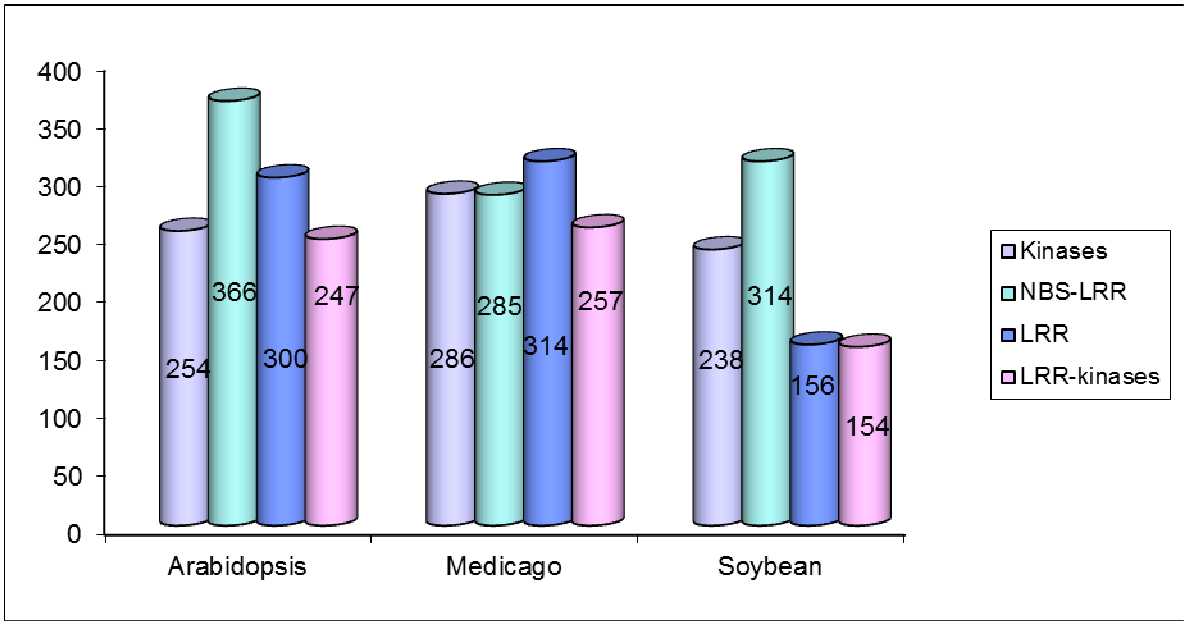



Figure 2: Distribution of *R*-genes families in soybean, arabidopsis and medicago. Numbers inside columns mean the amount of non-redundant sequences for each class.

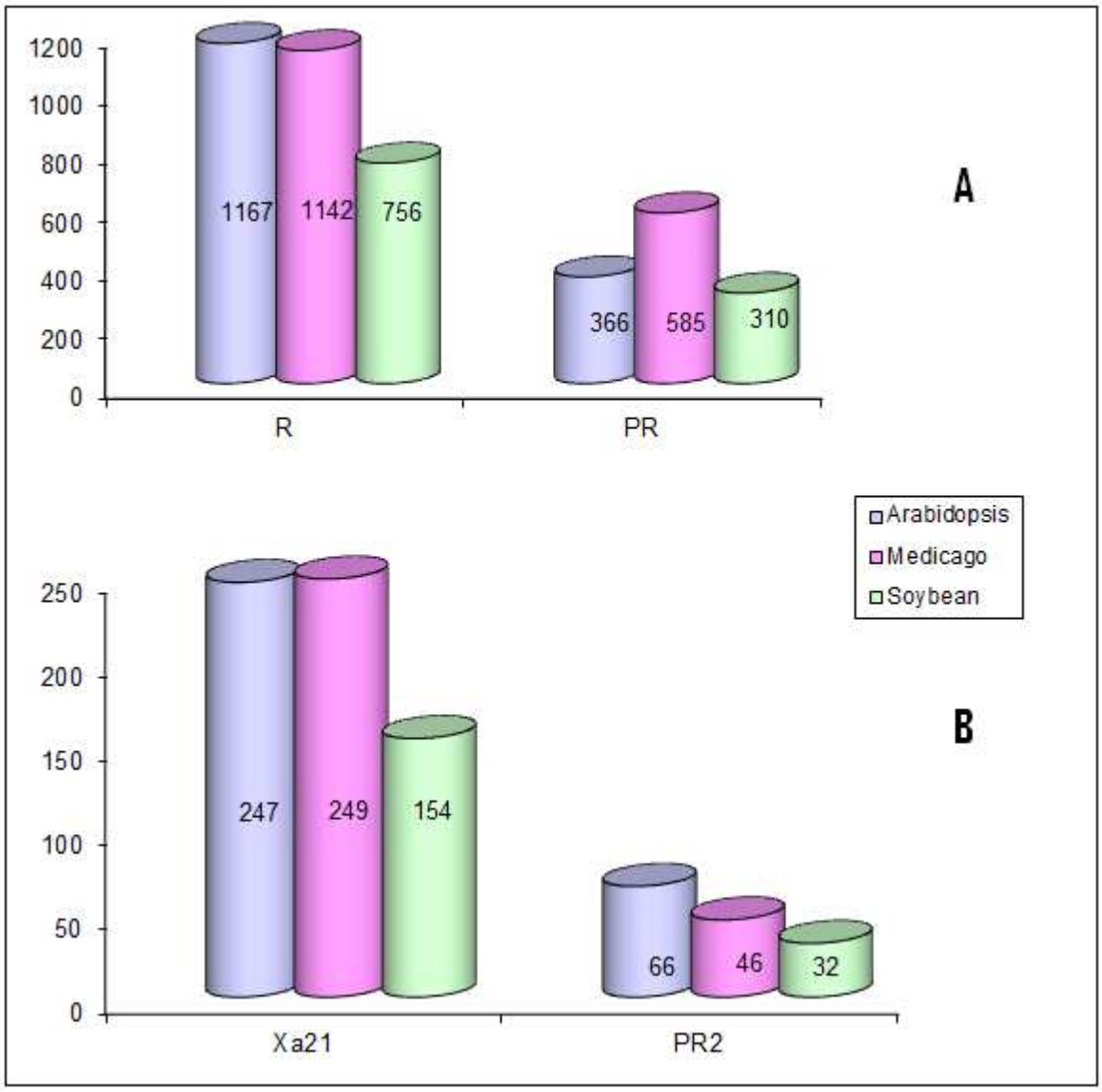



Figure 3: A. Distribution of *R* and *PR*-genes in soybean, medicago and arabidopsis. B. Distribution of *Xa21* and *PR-2* in soybean, medicago and arabidopsis. Numbers of matches for each gene category are inside the columns.

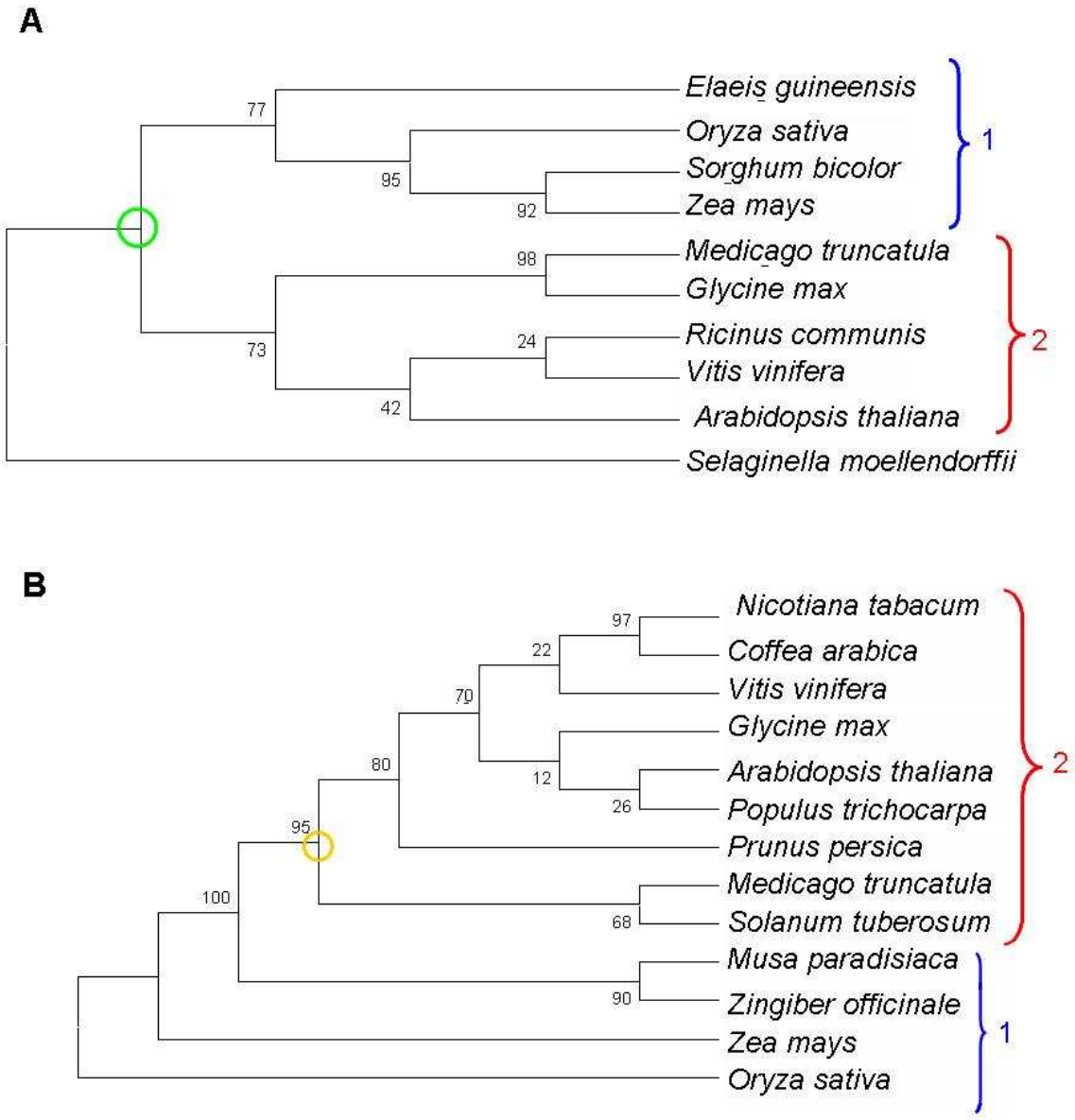



Figure 4: Dendrograms generated after maximum parsimony analysis showing the relationship among (A) *Xa21* and (B) *PR-2*orthologs.Keys in blue represent monocots (1) and in red dicots (2).*Xa21*: The green circle on the root shows the divergence point between monocots and dicots. *PR-2:* The yellow circle on the root shows a symplesiomorphyc character. Numbers are bootstrap values.

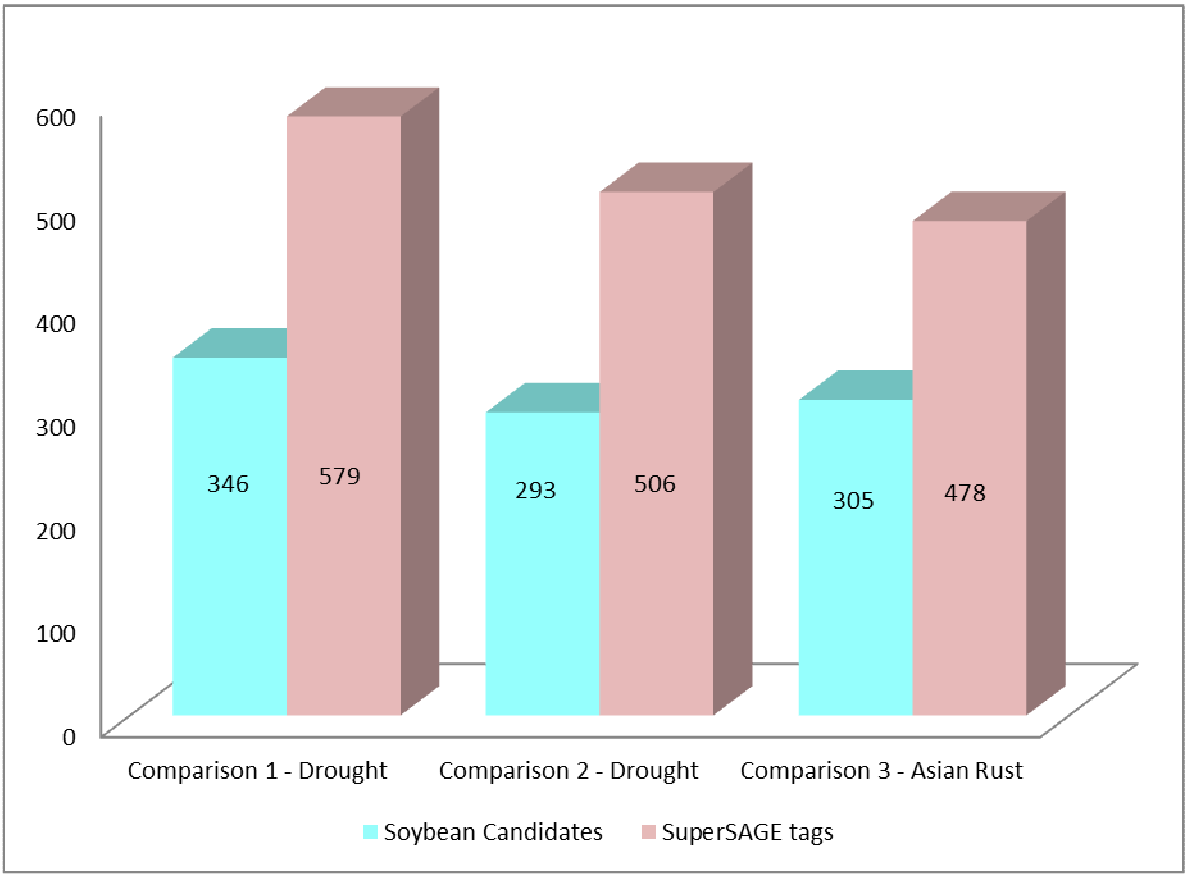
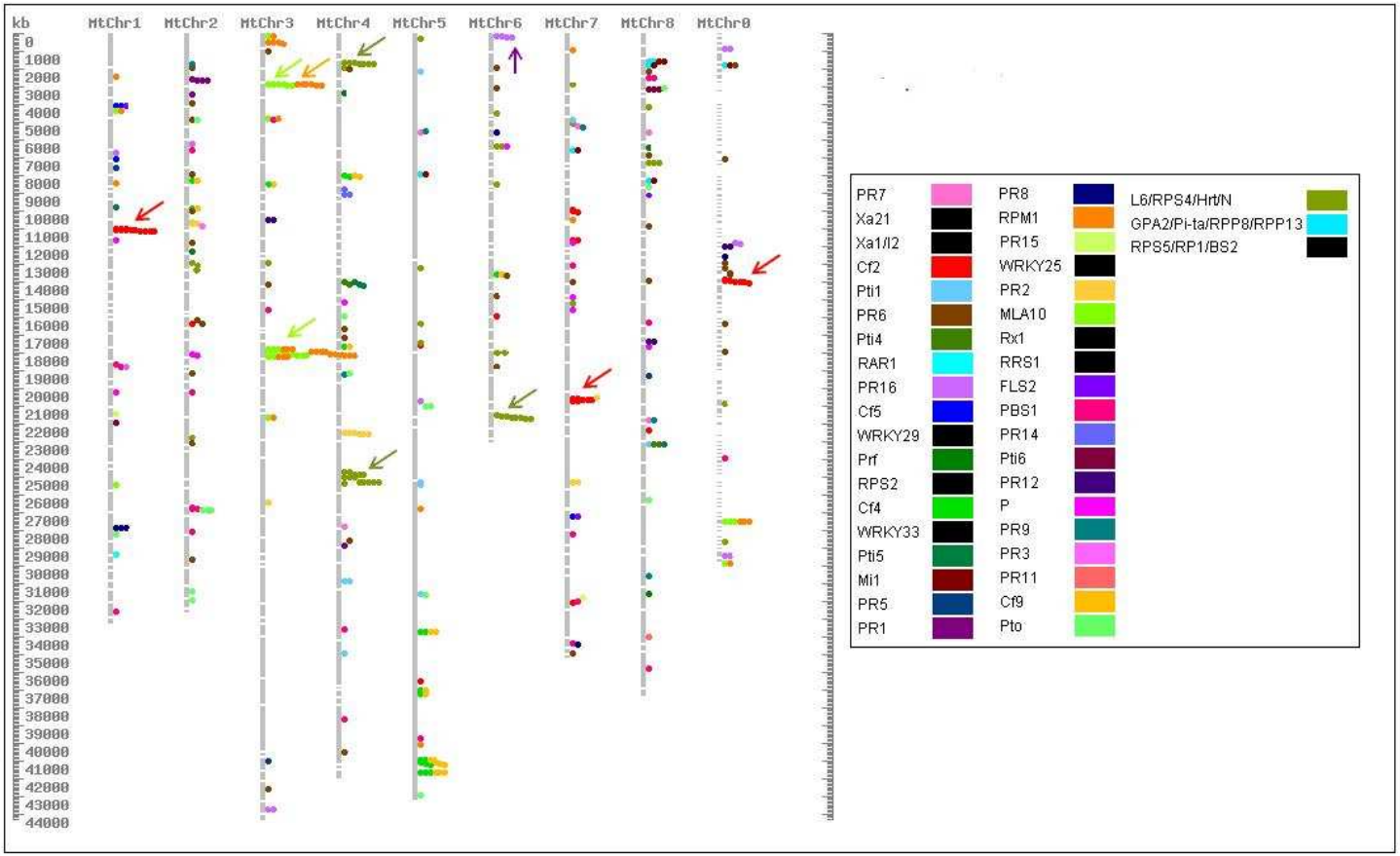



Figure 5: Number of SuperSAGE tags matching soybean *R* and *PR* gene candidates from three different comparisons among the six libraries: 1-Embrapa-48, drought tolerant stressed *vs.* negative control; 2- BR-16, drought susceptible stressed *vs.* negative control; 3- PI561356 fungus resistant stressed with *Phakopsora pachyrhizi vs.* negative control

Figure 6: Graphic representation of soybean R and PR sequences positioned on *Medicago truncatula* chromosomes (MtChr) with aid of the CVit-Blast resource available at the website http://www.medicago.org/. Arrows shows genes that appear in tandem repetitions.



# *IN SILICO* SCREENING FOR PATHOGENESIS RELATED-2 GENE CANDIDATES IN *VIGNA UNGUICULATA* TRANSCRIPTOME

Ana Carolina Wanderley-Nogueira[1], Nina da Mota Soares-Cavalcanti[1], Luis Carlos Belarmino[1], Adriano Barbosa-Silva[1,2], Ederson Akio Kido[1], Semiramis Jamil Hadad do Monte[3], Valesca Pandolfi[1,3], Tercilio Calsa-Junior[1], Ana Maria Benko-Iseppon[1]

(1)Universidade Federal de Pernambuco
Center of Biological Sciences, Departament of Genetics, Laboratory of Plant Genetics and Biotechnology, R. Prof. Moraes Rêgo s/no., Recife, PE, Brazil
ana.benko.iseppon@pesquisador.cnpq.br

(2) Max-Delbrueck Center for Molecular Medicine
Computational Biology and Data Mining Group, Robert-Roessle-Str. 10, 13125, Berlin, Germany

(3) Universidade Federal do Piaui
Laboratory of Immunogenetics and Molecular Biology, Campus Petrônio Portela bloco 16, Teresina, PI, Brazil

*Keywords*: bioinformatics, pathogenesis-related, Beta-glucanases, biotic stress, cowpea.

**abstract.** Plants evolved diverse mechanisms to struggle against pathogen attack, for example the activity of Pathogenesis-Related (*PR*) genes. Within this category *PR*-2 encodes a Beta-glucanase able to degrade the polysaccharides present in the pathogen cell wall. The aim of this work was to screen the NordEST database to identify *PR*-2 members in cowpea transcriptome and analyze the structure of the identified sequences as compared with data from public databases. After search for *PR*-2 sequences in NordEST; CLUSTALx and MEGA4 were used to align *PR*-2 orthologs and generate a dendrogram. CLUSTER program revealed the expression pattern trough differential display. A new tool was developed aiming to identify plant PR-2 proteins based in the HMMER analysis. Among results, a complete candidate from cowpea could be identified. Higher expression included all libraries submitted to biotic (cowpea severe mosaic virus, CPSMV) stress, as well as wounded and salinity stressed tissues, confirming PR expression under different kind of stresses. Dendrogram analysis showed two main clades, the outgroup and Magnoliopsida where monocots and dicots organisms were positioned as sister groups. The developed HMM model could identify PR-2 also in other important plant species, allowing the development of a bioinformatic routine that may help the identification not only of pathogenesis related genes but any other genes, classes that present similar conserved domains and motifs.

## 1 Introduction

In response to persistent challenge by a broad spectrum of microorganisms, plants have evolved diverse mechanisms to prevent infection [1]. The first one is the hypersensitive response (HR), that is immediate and starting with signal recognition from a pathogen elicitor by a host Resistance (*R*) gene leading to rapid cell death [2]; the second strategy is the systemic activation of genes encoding mithogen-activated protein kinases (MAPKs) and pathogenesis-related (*PR*) proteins which are directly or indirectly inhibitory towards pathogens and have been associated with the phenomenon of systemic acquired resistance (SAR) [3].

The PR proteins include several gene families, formed by acidic and basic subfamilies located in the vacuole or extracellular [4]. Functionally, some PR proteins have antifungal activity, such as Beta-glucanases. This enzyme, product of *PR*-2 gene activity, is able to degrade the polysaccharides present in the pathogen cell wall, preventing the colonization of the host by these organisms [5]. Studies suggested that PR-2 proteins play a protective role through two distinct mechanisms. First, the enzyme can impair microbial growth and proliferation directly by hydrolyzing the Beta-1,3/1,6-glucan of the cell walls rendering the cells susceptible to lysis and possibly to other plant defense responses. Second, an indirect defensive role is suggested by the observation that specific Beta-1,3/1,6-glucan oligosaccharides, released from the pathogen walls by the action of glucanases, can induce a wide range of plant defense responses and hence the SAR [6]. In herbaceous plants the *PR*-2 activation are strongly influenced by the accumulation of salicylic acid (SA) in their tissues [7]. Genes related to this pathway are highly conserved within the plant kingdom in relation to size, amino acid composition and isoelectric point [8], whilst some components of the system show similarity to proteins involved in innate immunity in the animal kingdom [9].

There is no previous evaluation regarding these metabolic pathways in *V. unguiculata*, which presents great economic importance in semi-arid regions throughout the world. The present work aimed to perform a data mining-based identification of *PR*-2 gene in the NORDEST database, comparing them with sequences deposited in public databases and literature data.

## 2 Methods

For the identification of *PR*-2 gene candidates, a tBLASTn alignment was carried out against NordEST database constructed using 18,984 transcripts isolated from *V. unguiculata*. An *Arabidopsis thaliana* sequence (AT3G57260.1) was used as seed sequence. After this search, *PR*-2 matching sequences (cutoff e-5) have been used to screen for homology in Genbank (NCBI) using the BLASTx tool [10].

Cowpea clusters were translated at Expasy [11] and screened for conserved motifs with aid of the RPS-BLAST CD-search tool [10]. Multiple alignments with CLUSTALx (available at link `http://www.nordest.ufpe.br/~lgbv/PR-2_Alignment`) allowed the structural analysis of conserved and diverging sites as well as elimination of non aligned terminal segments. CLUSTALx alignments were submitted to the program MEGA (Molecular Evolutionary Genetic Analysis V.4) [12] aiming to create a dendrogram using the maximum parsimony method and bootstrap function (1,000 replicates). Conserved motif evaluation was carried out using *.aln files (from CLUSTALx) out of eight *PR*-2 candidates from eight different species as input to the HMMER (Hidden Markov Models) program, that allowed the search of *PR*-2 typical patterns in 16 cowpea selected sequences.

To establish an overall picture of *PR*-2 gene distribution pattern in cowpea, we carried out a direct correlation of the reads frequency of each protein sequence in various NordEST cDNA libraries (available at link `http://www.nordest.ufpe.br/~lgbv/PR-2_Candidates_Reads`). Afterwards a hierarchical clustering approach was applied using normalized data and a graphic representation constructed with aid of the CLUSTER program. Dendrograms including both axes (using the weighted pair-group for each cluster and library) were generated by the TreeView program [13]. On the graphics, light gray means no expression and black all degrees of expression (see Fig. 4).

The analysis of protein migratory behavior, using only the best matches of each seed sequence, was generated with the JvirGel program [14] using an analytical mode for serial calculation of MW (molecular weight), pI (Isoelectric Point), pH-dependent charge curves and hydrophobicity probes, generating a virtual 2D gel as JavaTM applet.

Table 1: *V. unguiculata PR-2* candidates, showing the sequence size in nucleotides and amino acids with the conserved start and end sites, the best alignment in NCBI database, including the score and e-value. The acess numbers gb.AAY96764.1, gb.AAY25165.1, emb.CAO61165.1, emb.CAO71593.1 and gbEAU72980.1 corresponding to sequences from *Phaseolus vulgaris*, *Ziziphus jujuba*, *Vitis vinifera*, *Leishmania infantum* and *Synechococcus* sp., respectivelly. Abbreviations: CD, Glyco-hydro Conserved Domain.

| *Vigna* sequence | Size(nt/aa) | CD | Start/End | Best match | Score/E-value |
|---|---|---|---|---|---|
| Contig2370 | 804/140 | yes | 2/121 | gb.AAY96764.1 | 255/8.00E-67 |
| Contig2265 | 833/198 | yes | 1/164 | gb.AAY25165.1 | 286/6.00E-76 |
| Contig277 | 1244/242 | yes | 2/107 | emb.CAO61165.1 | 345/1.00E-93 |
| VUPIST02011D02.b00 | 867/103 | no | - | emb.CAO71593.1 | 43/0.005 |
| VUABTT00001D02.b00 | 552/78 | no | - | gbEAU72980.1 | 33/4.6 |

The analysis of SuperSage data included evaluation of 298,060 tags (26 bp) distributed in four libraries submitted to injury or mosaic virus infection against a negative control (available at link `http://www.nordest.ufpe.br/~lgbv/Sage_Tags`). The transcripts were screened for homology using as seed sequence a cDNA full length sequence corresponding to the *A. thaliana PR-2* gene. A local BLASTn against the SuperSage tags database (cutoff e-10) was performed using the Bioedit program. The obtained tags were classified into up or down-regulated comparing the control with infected and injured libraries; for this purpose the Discovery Space program was used.

Moreover, it has been developed a web-tool in order to identify amino acid sequences similar to the protein PR-2. For this purpose we used a set of sequences highly similar to PR-2 from *A. thaliana* extracted from the Entrez Protein.

The tool has been developed as a HTML platform which accepts a unique (or list) of FASTA formatted sequence(s) and some alignment parameters like e-value, score, coverage and dust filters. Behind the platform the software BLAST [10] is used to perform the batch pairwise alignments and PHP (available at link `http://www.php.net`) to parse the alignment results in order to generate the tool output.

In parallel to the alignment, it is executed for each query sequence, which matched PR-2 sequences over the desired thresholds, a HMMPFAM using the software HAMMER [15] is conducted in order to screen the relevant biological motifs present in PR-2 proteins, like glyco-hydro domain. Finally, a distance tree based on protein's similarities is generated for the local PR-2 proteins and the user's selected sequences. For this purpose we have used the PHYLIP package (`http://evolution.gs.washington.edu/phylip.html`) and the Neighbor Joining method. The developed web-tool can be accessed at the supplementary material website in the corresponding link (available at link `http://www.nordest.ufpe.br/~lgbv/HAMMER_consensus`). This bioinformatic routine may help the identification not only of pathogenesis related genes but any other genes classes that present similar conserved domains and motifs as showed in Figure 1.

## 3 Results

After the tBLASTn five *PR-2* candidates were identified in the NordEST database as including three clusters presenting the desired glyco-hydro motif as conserved domain (Tab. 1) and two singlets (sequences available at link `http://www.nordest.ufpe.br/~lgbv/vigna_nucleotide_sequences`). Best alignments using identified clusters occurred with chinese jujube (*Ziziphus jujuba*), grape (*Vitis vinifera*) and common bean (*Phaseolus vulgaris*).

The three clusters presented the desired domain incomplete at the amino-terminal, although only the Contig277 (Vigna_1) presented the complete desired domain. The

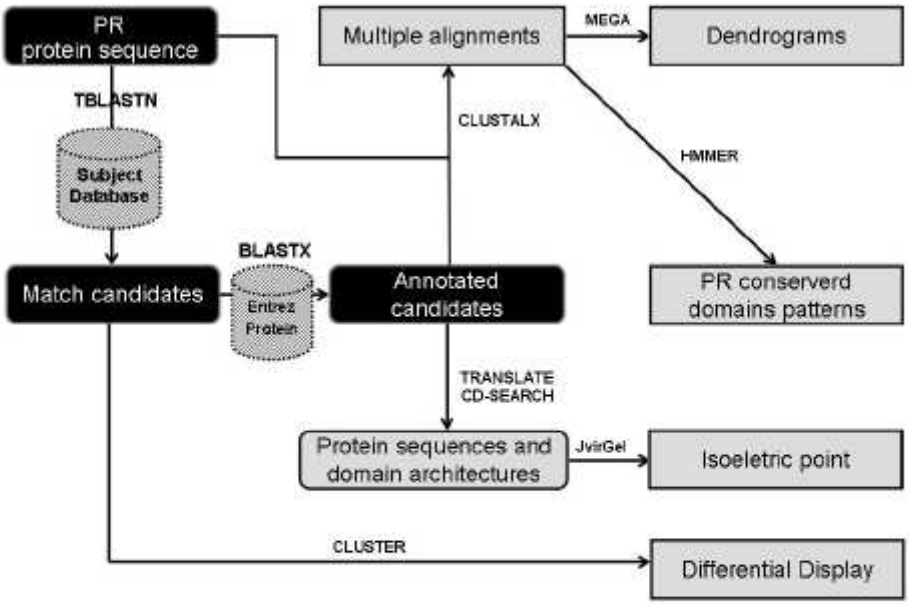



Figure 1: Pipeline to identify PR genes. Black boxes indicates data from automatic annotation. Gray boxes indicates manual annotation steps. Cylinders means used databases.

HMMER search of glyco-hydro domains generated a pattern of conserved motifs characteristic of these proteins when applied to *V. unguiculata* PR-2 sequences. Among the 16 NCBI cowpea sequences, three presented all 12 motifs that determine the Beta-glucanase activity according the HMMER consensus. Among the other sequences 23,07% presented nine from 12 conserved sites, 53,84% presented eight, 23,07% presented six and only 0,07% presented five sites (Fig. 2).

As the aim was to reconstruct the evolutionary history of *PR*-2 family genes considering the recent sequencing *V. unguiculata*, the most ancestral organisms (*Pinus pinaster* and *Physcomitrella patens*) were selected as out group (Fig. 3).

The topology showed two main clades, as expected: (A) outgroup and (B) Magnoliophyta group, that appeared as a monophyletic clade. Moreover, in this last one, the monocots and dicots organisms were identified as sister groups (Fig. 3B). Considering the Magnoliopsida subclade, the organisms were grouped according to their family, but the Rosid subclass behaved as a merophyletic paraphyletic group, sharing characteristics with the Asterid.

The virtual electrophoresis evaluation of PR-2 proteins from the 11 analyzed species presented isoelectric point from 3.0 to 9.45. Considering the molecular mass, values varied from 22.16MW to 41.15MW (Fig. 4). Closely related species did not present similar pIs except in case of corn and sugarcane.

Transcripts obtained in NordEST database were used to perform a hierarchical clustering analysis from ESTs permitting an evaluation of expression intensity considering co-expression in different libraries (black upper dendrogram) (Fig. 5). The three *V. unguiculata* contigs were formed by 95 reads with expression in seven of the nine NordEST libraries. Higher expression (21%) occurred in the library SS02 (roots of genotype sensitive to salinity after 2 hours of stress). Furthermore, IM90 (leaves collected after 90 min. of virus inoculation) and CT00 (negative control) presented no expression.

Figure 2: Twelve conserved motifs characteristic of PR-2 protein in 16 clusters from *V. unguiculata*. The first line shows the conserved motifs generated by the HMMER program using PR-2 proteins from eight different organisms. In light gray, it is possible to observe which motifs appeared in cowpea *PR-2* candidates.

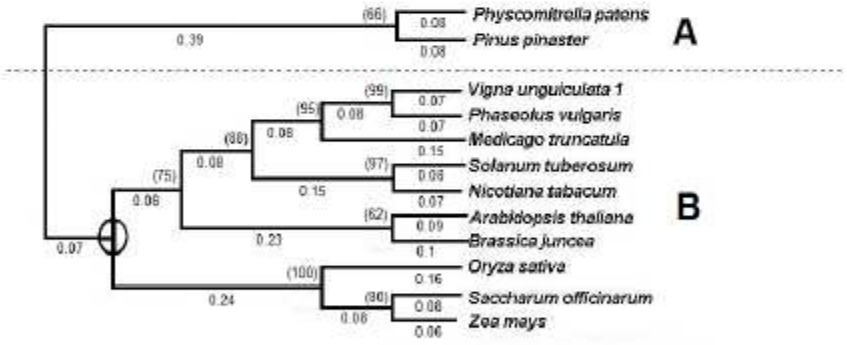



Figure 3: Dendrogram generated after Maximum Parsimony analysis, showing relationships among the PR2 seed sequence of *A. thaliana* and orthologs of *V. unguiculata* and other organisms with PR2 proteins bearing desired domains. Dotted line delimits the main taxonomic units and letters on the right of the dendrogram refer to the grouping. The circle on the root of clade B shows the divergence point between monocots and dicots organisms. Decimal numbers under branches lines means distance values. The numbers between parenthesis on the left of the branches nodes corresponding to the Bootstrap values.

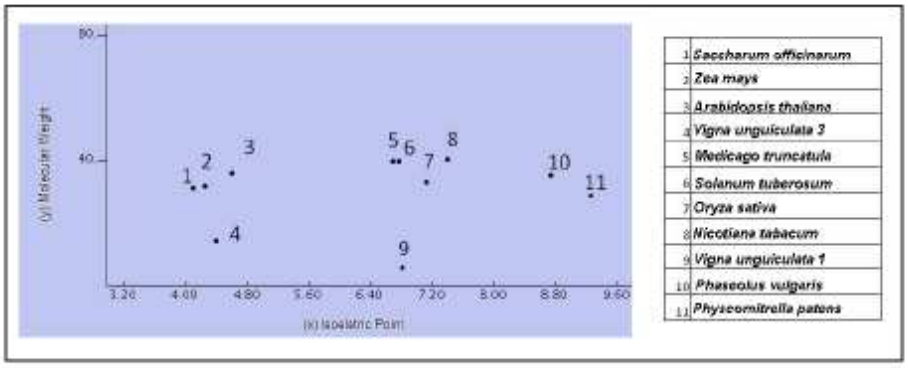



Figure 4: Graphic representation of PR-2 isoeletric points. Abscissa (X-axis) represents isoeletric point. Ordinate (Y-axis) represents molecular weight.

Regarding the SuperSAGE analysis, 31 tags were obtained with the local BLASTn in our database, among them, eight (25.8%) were up-regulated in both injured (BRC2) and infected (BRM) libraries when compared with the control. No *PR*-2 tag was down-regulated and 23 presented no differential regulation.

## 4 Discussion

Considering the number of sequences available hitherto the NordEST project, the number of *PR*-2 candidate sequences was higher as expected, revealing five clusters that aligned with this gene, formed by 95 reads.

Regarding the search of conserved motifs, essential to protein activity, it is interesting to note that some sequences presented sites that did not match with the consensus generated by the HMMER program, although most of the differences occurred with synonymous amino acids like methionine, leucine, valine or isoleucine, probably not affecting the protein activity; as methionine is a hydrophobic amino acid, and can nearly be classed with other aliphatic amino acids, it prefers substitution with other hydrophobic amino acids (like leucine, valine and isoleucine) [16]. Using the profile hidden Markov models (profile HMMs) may be a valuable tool to identify not only PR proteins but any other family built from the seed alignment and an automatically generated full alignment which contains all detectable protein sequences [17].

In the generated dendrogram it was possible to indentify that symplesiomorphic characteristics united all Magnoliopsida organisms, as expected, since these plants evolved from a common ancestor. In the Magnoliopsida group, the evolutionary model of the PR-2 seemed to follow a sinapomorphic pattern, leading to their presence in different families and subclasses. Moreover, it was possible to perceive that the different Magnoliopsida families were grouped based on unique autapomorphies. In the monocot grouping all members were annual cereal grains of the Poaceae family. However, while maize and sugarcane were grouped in a terminal clade, rice was grouped separately. This may be justified by the fact that *Zea* and *Saccharum* belong to the Panicoideae subfamily, while *Oryza* belongs to the Ehrhartoideae subfamily.

The studied organisms presented different centers of origin, habitats and cycles of life, as well as tolerance, resistance and sensitivity to diverse kind of biotic and abiotic stresses. Despite of that, in a general view, it is evident that the PR-2 pathway, that is activated under any type of stressor component, was a characteristic present in a common ancestor, being relatively conserved in the different groups during the evolution.

Considering our evaluation using virtual bidimensional electrophoresis migration for PR-2 sequences, one group could be clearly identified, with only two sequences (corn and sugarcane) deviating especially with respect to their molecular weight, despite the

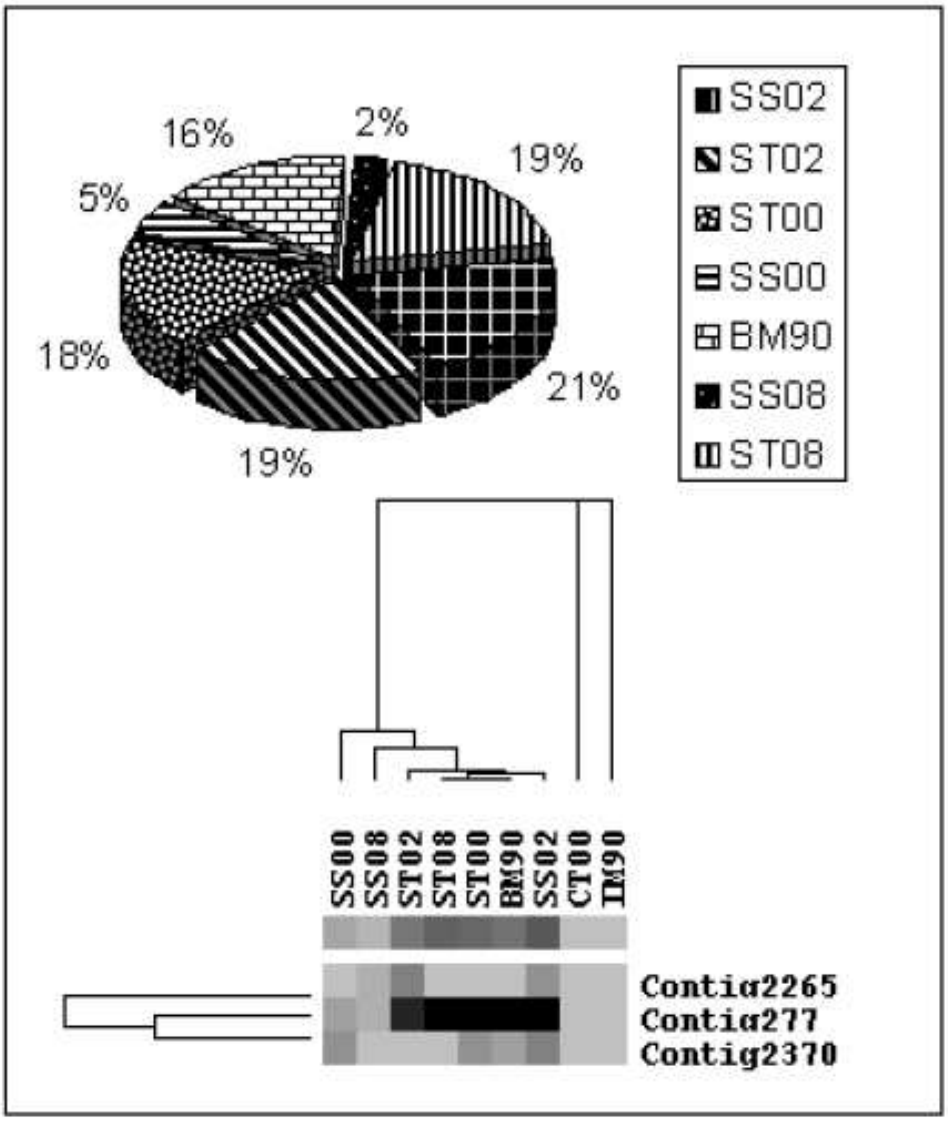



Figure 5: *PR*-2 expression profile. Black indicates higher expression, gray lower expression, and light gray absence of expression in the corresponding tissue and cluster. Abbreviations: CT00 (control); BM90 (Leaves of BR14-Mulato genotype); IM90 (Leaves of IT85F genotype collected with 90 minutes after mosaic viruses infection); SS00 (Root of genotype sensitive to salinity without salt stress); SS02 (Root of genotype sensitive to salinity after 2 hours of stress); SS08 (Root of genotype sensitive to salinity after 8 hours of stress); ST00 (Root of genotype tolerant to salinity without salt stress); ST02 (Root of genotype tolerant to salinity after 2 hours of stress); ST08 (Root of genotype tolerant to salinity after 8 hours of stress).

similar pIs and also confirming their similarity revealed in the dendrogram. The second group included most species of the 10 remaining species. Since the functional domains presented high degree of conservation, probably the divergent pattern of migration reflected divergences of the extra-domain regions that are responsible for the acid and basic character of the proteins. Such extra-domain variations are also responsible for the diversity of the sequences and probably for differences in their overall structure and transcription control.

The pattern of expression showed that PR-2 transcripts appeared in seven from nine available libraries from NordEST project. Being excluded from negative control (CT00, no stress) and IM90 (leaves collected from IT85F genotype after 90 minutes after virus inoculation). The absence in the non stressed plants was expected and its absence in the genotype IT85F (that is susceptible to this virus infection) reveals that in the resistant genotype (BR14-Mulato) this category of gene is probably activated earlier after injury or virus infection. A similar pattern was confirmed by our SuperSAGE data with high expression after salinity stress, as compared with the negative control. The activation of *PR* genes after different categories of abiotic stress (wounding and salinity) confirmed the theory that these genes can be activated and expressed systematically as response to any kind of stress, biotic or abiotic [7, 8].

The identified sequences represent valuable resources for the development of markers for molecular breeding and development of pathogenesis related genes specific for cowpea and other related crops. Additionally, the bioinformatic routine may help biologists to identify not only *PR*-2 genes but any other class of genes that bear similar conserved domains and motifs in their structure in higher plants. This platform represents the first step towards the development of on line tools to identify genes and factors associated with response to pathogen infection in plants.

**Acknowledgments**

This work was developed as part of the Northeastern Biotechnology Network (RENORBIO). The authors are grateful to Brazilian Mininstry of Science and Technology (MCT/CNPq) and to Foundation for Support of Science and Technology of Pernambuco State - Brazil (FACEPE) for supporting our research.

References

[1] K.F. Pedley and G.B. Martin "Role of mithogen-activated protein kinases in plant immunity". *Current Opinion in Plant Biology*, vol.8, pp.541-547, 2005

[2] U. Bonas U and G.V. Anckerveken "Gene-for-gene interactions: bacterial avirulence proteins specify plant disease resistance". *Current Opinion in Plant Biology*, vol.2, pp. 94-98, 1999.

[3] A.I. Bokshi, S.C. Morris, K. McDonald and R.M. McConchie "Application of INA and BABA control pre and post harvest diseases of melons through induction of systemic acquired resistance". *Acta Horticulturae 694*, 2003

[4] J.R. Cutt and D.F. Klessig "Pathogenesis Related Proteins. In: Boller T and Meins F (eds) Genes involved in plant defense. Springer, Vienna, pp. 209-243, 1992.

[5] Q.E. Zhu QE, S. Maher, R. Masoud, R. Dixon and C.J. Lamb "Enhanced Protection against fungal attack by constitutive coexpression of chitinase and glucanase genes in transgenic tobacco". *Biotechnology*, vol.12, pp. 807-812, 1994.

[6] F. Cte and M.G. Hahn "Oligosaccharins: Structures and signal transduction". *Plant Molecular Biology*, vol.26, pp. 1379-1411, 1994.

[7] T.P. Delaney, S. Ukness, B. Vernooij, L. Friedrich, K. Weymann, D. Negrotto, T. Gaffney, M. Gut-Rella, H. Kessman, E. Ward and J. Ryals "A central role of salicylic acid in plant disease resistance". *Science*, vol.266, pp. 754-756, 1994.

[8] J.M. Bonasera, J.F. Kim and S.V. Beer "PR genes of apple: identification and expression in response to elicitors and inoculation with *Erwinia amylovora*". *BMC Plant Biology*, vol.6, pp. 23-34, 2006.

[11] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A. Bairoch "ExPASy: The proteomics server for in-depth protein knowledge and analysis". *Nucleic Acids Research*, vol.31, no.13, pp. 3784-3788, 2003.

[12] S. Kumar, K. Tamura and M. Nei "MEGA 3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment". *Briefings in bioinformatics*, vol.5, pp. 150-163, 2004.

[13] M.B. Eisen, P.T. Spellman, P.O. Brown and B. Botstein "Cluster analysis and display of genome-wide expression patterns". *Genetics* vol.25, pp. 14863-14868, 1998.

[14] R.M. Ewing, A.B. Kahla, O. Poirot and F. Lopez "Large-scale statistical nalyses of rice ESTs reveal correlated patterns of gene expression". *Genome Research* vol.9, pp. 950-959, 1999.

[15] S.R. Eddy "Profile hidden Markov models". *Bioinformatics*, vol.14, pp. 755-763, 1998.

[16] M.J. Betts and R.B. Russell "Amino acid properties and consequences of subsitutions". M.R. Barnes, I.C. Gray (eds) "In Bioinformatics for Geneticists". Wiley, 2003

[17] R.D. Finn, J. Tate, J. Mistry, P.C. Coggill, S.J. Sammut, H.R. Hotz, G. Ceric, K. Forslund, S.R. Eddy, L.E. Sonnhammer and A. Bateman "The Pfam protein families database". *Nucleic Acids Research*, vol.36, pp. 281-288, 2008.

## CAPÍTULO IV
## (Artigo a ser submetido à Revista *Bioinformatic and Biology Insights*)

**Screening of biotic defence response genes in *Vigna unguiculata*** (L.) Walp. transcriptome

Ana Carolina Wanderley-Nogueira[1], Luis Carlos Belarmino[1], Ederson Akio Kido[1], Valesca Pandolfi[1], Ana Maria Benko-Iseppon[1]

[1]Genetics Department, Center of Biological Sciences, Universidade Federal de Pernambuco, Av. Prof. Morais Rêgo, 1235, CEP 50.670-420, Recife, PE, Brazil.

Short running header: *R* and *PR* genes in cowpea, medicago and soybean

Corresponding author:

Ana Maria Benko Iseppon, Universidade Federal de Pernambuco, Genetics Department, Center of Biological Sciences, Av. Prof. Morais Rêgo 1235, CEP 50.670-420, Recife, PE, Brazil. Phone: 55-81-2126-7816; Fax: 55-82-2126-8522; Email: ana.benko.iseppon@pq.cnpq.br

**Abstract**

Plant Resistance *(R*) genes encode immune receptors that recognize pathogens and activate defence mechanisms triggering the hypersensitive response (HR) and also pathogenesis related (*PR*) genes expression. Aiming to perform an overview on cowpea *R* and *PR* genes, regarding their structure, abundance and roles, the available cowpea transcriptome was evaluated, revealing 1,133 matching sequences (854 *R* and 279 *PR* candidates). The most representative candidates regarded the NBS-LRR class (44.49%) for *R* genes and Secretory peroxidase *PR-9* family (27.24%), among *PR* candidates. Analysing 22 different libraries it could be observed that leave tissues presented higher expression (17%) while remaining tissues enclosed 3% to 5% of all reads. After anchoring 57 cowpea *R* and *PR* candidates in the soybean genome, a clusterization of *R* and *PR* genes was observed, especially in terminal chromosome positions. Tandem duplications were prevalent regarding *R* genes, in consonance with their proposed evolutionary mechanism regarding gene duplication and neofunctionalization.

Keywords: Pathogen response, environmental stress, bioinformatics, legumes.

## Introduction

In order to combat the effects of pathogen attack, plants evolved the ability to recognize the pathogens' threat and struggle against the invader as well as trigger an effective response (Bolton, 2009). Considering this complex response, one of the most important steps regards the detection of possible invaders by the plant, a stage where *R* (Resistance) genes play a crucial role. This sensing involves the recognition of a pathogen gene product called avirulence (*avr*) factor by a correspondent *R* gene. The plant will be resistant when both *avr* and *R* genes are compatible, leading to the so called Hypersensitive Response (HR) characterized by rapid cell death in and around the initial infection site avoiding the pathogen growth and establishment (Park, 2005).

Besides the mentioned local reaction, the HR activates a signal cascade able to establish resistance against a spectrum of different pathogen classes, corroborating observations made at the beginning of the last century that plants, as animals, may be immunized against the attack of a given pathogen after infection by another pathogen (Chester, 1933). In the past decade many aspects of the Systemic Acquired Resistance (SAR) have been elucidated. As soon as the pathogenic agent is detected the plant induces a complex set of signal molecules able to activate defence proteins that may have direct antimicrobial effect as it is the case of the Pathogenesis-Related (*PR*) genes (Li et al, 2010). Alternatively they may induce the production of secondary metabolites that impair pathogen movement or growth within the plant tissues (Sparla et al., 2004).

Resistance genes are generally classified into five different groups or classes, defined by their conserved domains (CD) (Bent, 1996; Hammond-Kosak and Jones, 1997; Ellis and Jones, 2000).

The only *R* gene class where conserved domains are absent is the class represented by the *HM1* gene of maize codifying a reductase able to inactivate toxins produced by the fungus *Helminthosporium carbonum* (Joahal and Briggs, 1992). Due to the lacking of conserved domains, this gene class will be not considered in the present evaluation.

Another *R* gene, *Pto* from tomato confers resistance against the bacterium *Pseudomonas syringae* pv. *tomato* and is characterized by a serine/threonine-kinase (ser/thre-kinase) domain, able to interact with the *avrPto* gene (Tang et al., 1999). *Pto* was the first *R* gene cloned from a plant that recognized a specific pathogen (Oh and Martin, 2011). This gene was also identified in other plants as *Arabidopsis thaliana* L., *Phaseolus vulgaris* L. (Melotto et al., 2004) and sugarcane (Wanderley-Nogueira et al., 2007) and represents the second class of *R* genes.

The third class is the largest class of predicted intracellular plant immune receptors (Swiderski et al., 2009) represented by genes bearing two domains: LRR (*Leucine Rich Repeats*) and NBS (*Nucleotide Binding Site*) (Liu et al., 2004). This is the case of the *Rpm1* and *Rps2* genes from *A. thaliana*, the *N* gene from tobacco, *L6* from flax, *Prf* from tomato and *Rpg1* from soybean also found in common bean, soybean and faba bean (Liu et al., 2004).

The fourth *R* gene class codifies a membrane anchored protein composed by a LRR extracellular domain, a transmembrane region and a short intracellular tail in the C terminal. The *Cf* gene from tomato is an example of this class, conferring resistance against *Cladosporium fulvum* (Dixon et al., 1996).

The *Xa21* gene from Rice that confers resistance to the bacteria *Xanthomonas oryzae* pv. *oryzae* is a representative of the fifth class (Song et al., 1995; Wang et al.,

1995). This gene codifies an extracellular LRR domain (similar to the *Cf* gene), as well as a ser/thre-kinase domain (similar to the *Pto* gene), suggesting an evolutionary ligation among different classes in the genesis of plant *R* genes (Song et al., 1997).

The PR proteins comprise pathogen-induced proteins being routinely classified into 17 families (from PR-1 to PR-17) based on their biochemical and molecular biological properties suggesting a general role for these proteins in adaptation to biotic and abiotic stress conditions (van Loon et al., 2006). Similarities among sequences and serologic or immunologic properties are the base of their classification (Van Loon et al., 1999).

Although most PR proteins are known to have antifungal activities, their active molecular mechanisms are not well understood except in PR-2 (β-glucanases), PR-3 (chitinases) (Kitajima and Sato, 1999) and PR-4 (Hevein or Barwin) (Wang et al., 2011). PR1 is the most abundantly family accumulated after pathogen infection and has been cloned in tobacco (Gaffney et al*.,* 2003), *Arabidopsis* (Metzler et al., 1991), tomato (Tornero et al., 1997) and apple, always presenting unknown phytochemical functions; despite of that, this gene class is considered to be a typical SAR marker (Bonasera et al*.,* 2006). PR-5 is a thaumatin-like protein with high antifungal activity, being also expressed under cold stress in overwintering monocots, exhibiting antifreeze activities (Hon et al., 1995, Atici and Nalbantoğlu, 2003, Griffith and Yaish, 2004). Genes that pertain to *PR-13* group (defensin or gamma-thionin) comprise small peptides that were also isolated and characterized from roots, seeds, flowers and leaves, showing phytopathogenic activity and activity also against pathogenic bacteria in humans (Pelegrini et al*.*, 2011). Other families like PR-8 (Glycosyl hydrolase), PR-9 (secretory peroxidase), PR-14 (lipid transfer proteins), PR-15 (oxalate oxydase) and PR-17 (basic secretory proteins) (Nanda et al., 2010) have been well studied and are believed to be involved in plant defence responses, although their molecular mechanisms should yet be

determined (Bolton, 2009). Most *PR* genes are expressed to a basal level under normal growth conditions but are rapidly induced after pathogen infection. Previous approaches have shown that *PR* genes are also regulated by environmental factors as osmotic, cold and light stress (Seo et al*.,* 2008; Benko-Iseppon et al., 2010) and during development, for example during leaf senescence and pollen maturation (Zeier et al., 2004).

Cowpea (*Vigna unguiculata* (L.) Walp.) is a food legume with high protein and carbohydrate content, being important not only for the human diet, but also suitable as high protein feed and fodder to livestock (Muchero et al, 2009). With its greater tolerance to heat, drought, diseases and low soil fertility, *V. unguiculata* is a particularly valuable crop well adapted to poor semi-arid regions of the tropics where other food legumes do not perform well. It is also able to enhance soil fertility through nitrogen fixation (Lu and Yang, 2010) under these adverse conditions. The estimated worldwide area under cowpea cultivation was estimated to achieve 14 million ha in warm regions of the world with substantial quantity produced in South America, especially in semiarid north-eastern of Brazil (Singh et al., 2003). Despite of its importance, no previous overall identification of *R* and *PR* gene candidates of cowpea was carried out up to date.

The present evaluation brings an overview of the abundance and diversity of the available sequences regarding plant-pathogen interaction in cowpea transcriptome, bringing insights on their prevalence in different tissues.

**Material and Methods**

For the identification of *R* and *PR* genes in cowpea 57 proteins that play important role in plant defence response were selected as seed sequences from other angiosperms.

The selected protein sequences were related to 40 *R*- and 17 *PR* genes classes described and validated before. The used *R* genes were previously compiled by Barbosa-da-Silva et al. (2005) and Wanderley-Nogueira et al. (2007), while *PR* sequences were also used in a previous evaluation on soybean by Wanderley-Nogueira et al. (in press). All 57 seed sequences regarded full cDNAs obtained from NCBI and translated to perform best search results.

For the identification of these gene analogs in cowpea transcriptome, tBLASTn alignments were carried out against the cowpea genome sequences that were downloaded from NCBI and assembled by CAP3 with default parameters, using $e^{-5}$ as cut-off value.

Obtained clusters were annotated and analyzed regarding score, e-values, sequence size and presence of conserved domains as shown in Table 1. For this purpose all clusters were translated using the TRANSLATE tool of Expasy and screened for conserved motifs with aid of the RPS-BLAST CD-search tool (Altschul et al., 1990). The best match for each gene in each studied species was submitted to a BLASTx alignment in NCBI GenBank with the aim to confirm their putative function. Following, a self BLASTn search was made with the now classified cowpea candidates aiming to locate more distantly related sequences to be identified.

A second manual analysis was also carried out, followed by an elimination of redundancies (clusters that matched more than one gene due to common domains). For this purpose, matching clusters to each query sequences were annotated on a local database (called 'non-redundant').

The third step of the analysis was to compare the number of R and PR candidate sequences obtained after the tBLASTn against the soybean and medicago databases by direct counting of non-redundant clusters for each one of the 57 genes studied.

An analysis of the distribution of the 57 genes in the cowpea libraries was carried out by direct correlation of the reads' frequency regarding each cluster in all 22 different libraries including all sequences deposited in the NordEST database, as follows: **(1)** NordEST Salinity essay including: **SS00**: susceptible accession (cultivar 'Canapu Amarelo') roots without stress (control); **SS02**: roots of 'Canapu Amarelo' two hours after submission to 100 mM NaCl solution; **SS08**: roots of 'Canapu Amarelo' eight hours after submission to 100 mM NaCl; **ST00**: tolerant accession ('Pitiúba') roots without stress (control); **ST02**: roots of 'Pitiúba' cultivar two hours after submission to 100 mM NaCl; **ST08**: 'Pitiúba' roots two hours after submission to 100 mM NaCl; **(2)** NordEST virus (CPSMV, Cowpea Severe Mosaic Virus) inoculation essay: **BM01**: leaves of the resistant cultivar 'BR14-Mulato' without stress (control); **BM90**: 'BR14-Mulato' leaves infected by the CPSMV virus (bulk of three times 30, 60 and 90 minutes) after stress; **IM01**: leaves of the virus susceptible accession IT85F-2687 without stress (control); **IM90**: IT85F-2687 leaves infected by CPSMV Mosaic virus (bulk of three times 30, 60 and 90 minutes) after stress; **(3)** Libraries available at NCBI from different tissues (no treatment given): **BUD**: growing axillary buds; **DSD**: developing seeds; **LF**: leaves; **LF2**: primary leaflets; **LFM:** leaflets and shoot meristem; **MI1**: mixed tissues; **NDL**: nodules; **RT2**: root hairs; **RT3**: roots; **SD**: seeds.

*V. unguiculata* best match for all 57 selected genes (Table 1) were aligned against *Glycine max* L. Merr. pseudogenome using the soybean genome browser tool available at the SoyBase web server (http://soybase.org). For this purpose the MegaBLAST tool was used to identify the exact location of the candidate sequences in the genome,

considering at least 80% of identity as parameter. This approach allowed the identification of their distribution, relative position, and abundance as inferred from a virtual karyotype representation generated with aid of the Circos (Circular Genome Data Visualization) program.

## Results

### 1-Collection and distribution of genes in cowpea, soybean and medicago

The first tBLASTn alignment against the cowpea transcriptome using the 57 known *R* and *PR* gene probes returned 1,133 non-redundant sequences among clusters and singlets available in the NordEST database and NCBI. Among them, 670 regarded clusters and 279 represented singlets, which together enclosed 7,051 reads. Regarding the tBLASTn in medicago transcriptome, a total of 1,727 sequences matched positive results. In soybean 1,066 sequences returned matches after the same procedure.

A screening of *R* and *PR* genes in these three species resulted in the identification of 3,899 candidates, being 2,752 *R* and 1,174 *PR* gene candidates. A graphical representation of these sequences and how they are distributed among soybean, medicago and cowpea libraries is shown in Figure 1.

After analyzing all results it could be observed that only two *PR* gene (*PR-10* and *PR-13*) were absent in the cowpea transcriptome, while all other 55 genes presented positive results after the first tBlastn. In medicago and soybean tBlastn the same results were observed for the *PR-13* gene as shown in Table 1. A comparison of the distribution of non-redundant sequences in the three species revealed that the NBS-

LRR family was the most frequent in soybean, medicago and cowpea, performing 44,49% of all matches in the last one (Figure 2).

Moreover, it could be observed that medicago presented a higher number of *R* genes and *PR* genes candidates. Cowpea was the species that presented the lowest number of matches for *PR* genes.

While in soybean and medicago the three most represented *R* genes were the same, *Xa21*, *EFR* and *Pti6,* cowpea presented more candidates of *PBS1, Cf-5 and FLS2* genes. Regarding *PR* genes the classes *PR-2, PR-7* and *PR-9* were the most abundant in all three species. Among the 279 *PR* genes of cowpea only 23 matched with more than one seed sequence; all the remaining candidates were exclusive of one *PR* gene family. By the other hand almost all *R* genes matched with sequences that aligned with more than one probe. Exceptions occurred only in *RAR1*, *RIN4*, *P*, *N, Xa21*, *Pti1*, *Pti6* and *WRKY29*, that aligned in most cases with exclusive sequences.

**2- Prevalence of reads among cowpea libraries**

Distribution of transcripts among the different tissues and conditions available in the 22 cowpea libraries after normalized data revealed that all libraries presented almost the same number of reads. The most representative library was from leave (LF) tissues performing 17% of all matching reads. Expression in all other tissues varied from 3% to 5%, except in the case of the libraries made from nodule tissues (NDL), where no reads were identified (Figure 3).

**3- Anchoring *Vigna unguiculata R* and *PR* genes in soybean chromosomes**

All procured *R* and *PR* cowpea best matches (57 sequences) were anchored on the soybean chromosomes matching 360 corresponding loci. From these, 56 presented similarities with distinct segments in the same chromosome or appeared twice in distinct chromosomes. Only the sequence corresponding to *RPM1* gene anchored exclusively in a single site (chromosome 3). The higher number of anchored genes appeared in chromosomes 1, 2 and 3 matching 17, 16 and 16 genes and 20, 16 and 30 loci, respectively. On the other hand, chromosomes 4, 6 and 7 presented the lowest number of correspondences, matching only five genes and five loci each (Figures 4 and 5).

In tandem repetitions could be observed along almost all chromosomes, with emphasis on the *R* genes group, especially for *RPM1* (nine repetitions) and *PR-14* genes (four repetitions) in chromosome 3 and *RPS2* gene (seven repetitions) in chromosome 18. Exceptions are represented by chromosomes 4, 6 and 7, that enclose non duplicated sequences. Blocks of *R* genes appeared clustered along the genome, with at least four distinct genes in very close positions while *PR* genes generally appeared in the same chromosomes in distinct sites.

## Discussion

The relative abundance of available nitrogen for legumes increases their chances of adapting in more challenging environments (Cannon et al., 2009). So, it is not surprising that cowpea, the same way as other important legumes – soybean and medicago – presented in genome almost all procured gene families involved in defence and adaptive responses. The 1,133 sequences that resulted from tBLASTx alignments

confirm some coverage that the existing cowpea databank comprises, including most important representatives from different gene families.

The lowest number of *R* and *PR* gene candidates found in *V. unguiculata* when compared to medicago reflects by one side the still limited number of transcripts sequenced for this species. By the other side, it may reflect the compact character of the cowpea genome, since cowpea presents the smallest genome among the three taxa (450 Mb against 1,115 Mb of soybean and 583 Mb of medicago respectively). The ancient polyploidization events during the evolution of the Glycinea (Shoemaker et al., 2006) and the high level of repetitive sequences in the soybean genome (Innes et al., 2008) indicate that despite the relative large difference in the genome sizes of soybean and medicago, gene densities are similar. This fact probably explains the similar amount of sequences found comparing soybean and medicago.

Legumes are plants known to be able to support many kinds of stress as rapid climate changes, drought tolerance, exposition to diseases and pests, water logging and flooding (Cannon et al., 2009). Considering that diverse *PR* gene families can be activated not only by biotic but also by such abiotic stresses, the *PR* genes encountered in the transcriptome of the three species also under biotic stress is not surprising.

The similar amount of reads in libraries made from tissues submitted both to abiotic (salt stress essay) and biotic (infection by CPSMV) stress could be explained by the ability of the plant to continue expressing genes related to systemic acquired resistance as a consequence of the contact with any kind of previous biotic or abiotic stress (Durrant and Dong, 2004); this process named crosstalk was reported before in cowpea (Kido et al., 2011) and also in soybean by Wanderley-Nogueira et al. (in press).

*R* genes are known by sharing common domains as, for example, the LRR domain that is present in the LRR family, NBS-LRR family and LRR-kinase family, facilitating alignments with more than one gene, so the number of cowpea clusters that aligned with more than one seed sequence was expected. Similar results were observed in previous studies regarding eucalyptus (Barbosa-da-Silva et al., 2005), sugarcane (Wanderley-Nogueira et al., 2007) and soybean (Wanderley-Nogueira et al., in press). This fact is rarer in *PR* genes classes that are really distinct in structure and functions (Kitajima and Sato, 1999) as observed in our results.

NBS-LRR class best alignments, as reported in Barbosa-da-Silva et al. (2005), Wanderley-Nogueira et al. (2007) and Wanderley-Nogueira et al. (in press) confirm that the general observation that most of *R* genes are members of this class.

Considering the prevalence of the procured *R* and *PR* genes in leave and other aerial tissues, this is in consonance with Libault et al. (2010), that recognized consistent differences in gene expression among tissues of aerial organs and other plant tissues, a fact that can explain why our results showed higher expression in such tissues when compared with other organs as roots, after biotic and abiotic stresses.

The present approach included the anchoring of cowpea *R* and *PR* genes in the soybean genome. The angiosperms originated approximately 200 million years ago (mya; Wilkstrom et al. 2001) and subsequently diverged into several lineages. According to Gepts et al. (2005) the Leguminosae evolved ca. 65-65 mya, from which the Papilionoideae (including the genera *Medicago*, *Glycine* and *Vigna*) originated about 59 mya. Within this last group, *Glycine* and *Vigna* are most related, being included in the Phaseoloid tribe, estimated to have a more recent origin ~45 mya).

Synteny and colinearity have been observed not only within plant families, but also among far related species, including genes responsible for the defence against pathogen attack, as observed by Benko-Iseppon et al. (2003) that reported microsyntenic region rich in defence genes among the linkage group 5 of the legume chickpea (*Cicer arietinum* L.) and the chromosomes 1 and 5 of *A. thaliana*. Similarly to that, Yan et al. (2004) observed microsyntenic relations between the legumes *G. max* and *M. truncatula* and among both and the far related *A. thaliana*. Thus, in the absence of the cowpea complete genome sequence, the most adequate annotation approach to infer about the possible distribution of the cowpea defence sequences regards the *G. max* genome, available at the SoyBase genome browser.

Such a comparison is highly valuable, even considering that cowpea is diploid and presents 2n=20, while soybean bears 2n=40 chromosomes (Benko-Iseppon, 2001). The cultivated *G. max* (2n=40) is supposed to have originated from a cross between *G. soya* (2n=20) and another parental *Glycine* species originating the allopolyploid soybean actually cultivated worldwide (Chung and Singh, 2008; Gill et al., 2009). Therefore the anchoring of most of the procured sequences in at least two chromosomes was expected, confirming previous analysis that observed that, in average, 61.4% of the soybean genes are present in two chromosomes, 5.63% in three and 21.53% in four (Schmutz et al., 2010).

Many of the redundancies observed in the soybean genome by previous evaluations are evident also in regard to the actual *R* and *PR* gene distribution. For example, Schmutz et al. (2010) observed a high colinearity among the long arms of chromosomes 10 and 20, both showing many common genes in our evaluation, especially regarding the terminal portions of the long arms.

Besides inter-chromosomal redundancies, it is interesting to note that many sequences presented similarities with distinct segments in the same chromosome. Especially in the case of *R*

genes, '*in tandem*' repetitions could be observed along almost all chromosomes, with emphasis on the *R* genes group, as for example the *RPM1* sequence (nine repetitions) in chromosome 3 and *RPS2* gene (seven repetitions) in chromosome 18. In a lower grade, repetitions of *PR* genes have been also observed, as for example regarding the *PR-14* candidate with four repetitions in chromosome 3.

Previous works also highlighted that most resistance genes reside in clusters (Kanazin et al. 1996) as reported in maize (Dinesh-Kumar et al. 1995), lettuce (Maisonneuve et al. 1994), oat (Rayapati et al. 1994), flax (Ellis et al. 1995), arabidopsis (The Arabidopsis Genome Initiative, 2000), chickpea (Benko-Iseppon et al. 2003) and sugarcane (Wanderley-Nogueira et al., 2012).

The clustering of *R* genes was recognized also in genetic mapping approaches by Pryor (1987) that suggested a faster evolving manner for this superfamily. Later works (Walsh, 1995) indicated a possible role of the incorrect meiotic pairing (mainly unequal crossing-over) in the genesis of such clusters, an event also observed in regard to the locus *Rp1* in maize by Sudupak et al. (1993). Such clustering of *R* genes and their higher evolutionary rate (as compared with most other coding sequences), lead to the existing theory that a common genetic mechanism has been involved in their evolution in association with duplication processes followed by neofunctionalization through pressure from the pathogen or the environment (Pryor e Ellis, 1993). Anyway, even considering their faster evolutionary behaviour, such a redundancy within and among chromosomes is not exclusive of genes associated to pathogen response, being observed also in regard to other gene families as aquaporins in soybean (Belarmino et al., in press) or transcription factors in other legumes (Soares-Cavalcanti et al., in press).

**Concluding remarks and perspectives**

The results revealed interesting insights on the variability and complexity of defense genes in cowpea as compared with soybean and medicago. The application of molecular biology in parallel with bioinformatics is now thought to be one of the most important topics in plant science and has come as main source in understanding biotic and abiotic stress responses. Using this strategy many stress-inducible genes could be isolated and their functions precisely characterized in transgenic plants. Thus, the here identified sequences represent valuable sources for application on cowpea breeding or also for related legumes. The next steps regard differential expression profiling and in vivo cell biology, using the identified candidates among resistant and susceptible cowpea accessions to uncover their most specific roles, including pathogen specificity and spatial-temporal regulation.

## Author Contributions

Conceived and designed the experiments: ACWN, LCB AMBI. Analysed the data: ACWN. Wrote the first draft of the manuscript: ACWN. Contributed to the writing of the manuscript: LCB, EAK, VP. Agree with manuscript results and conclusions: ACWN, LCB, EAK, VP, AMBI. Jointly developed the structure and arguments for the paper: ACWN, AMBI. Made critical revisions and approved final version: EAK, VP, AMBI. All authors reviewed and approved of the final manuscript.

## Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

## Acknowledgements

The authors thank to CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco), and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the financial support and fellowships.

## References


Alberts B, Johnson A, Walter P. *Biologia Molecular da Célula*. 3rd ed. Artes médicas; 1997.

Altschul SF, Gish W, Miller W, Myers E. *Basic local alignment search tool.* J Mol Biol. 1990; 2:124-143.

Ashfield T, Bocian A, Held D et al. *Genetic and physical localization of the soybean Rpg1-b disease resistance gene reveals a complex locus containing several tightly linked families of NBS-LRR genes* Mol Plant Mic Int.2003; 16:817-826.

Atici O, Nalbantoglu B. *Antifreeze proteins in higher plants*. Phytochem; 2003; 64:1187-1196.

Barbosa da Silva A, Wanderley-Nogueira AC, Silva RRM , Berlarmino LC, Soares-Cavalcanti NM, Benko-Iseppon AM. *In silico survey of resistance (R) genes in Eucalyptus transcriptome*. Gen Mol Biol. 2005;28:562-574.

Belarmino LC, Oliveira ARS, Brasileiro-Vidal AC et al. *Mining plant genome browsers as a mean for efficient connection of physical, genetic and cytogenetic mapping: an example using soybean.* Gen Mol Biol 2012;in press.

Benko-Iseppon AM, Lins Galdino S, Calsa Jr T et al. *Overview of Plant Antimicrobial Peptides*. Curr Prot Pept Sci. 2010; 11:181-188.

Benko-Iseppon AM. *Estudos moleculares e citogenéticos no caupi e em espécies relacionadas: Avanços e perspectivas.* Embrapa Documentos. 2001;56:327-332.

Benko-Iseppon AM, Winter P, Huettel B, Staginnus C, Muehlbauer FJ, 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 Gen. 2003;107:379-386.

Bent AF. *Plant disease resistance genes: function meets structure*. Plant Cell. 1996; 8:1751-1771.

Bolton M. *Primary metabolism and plant defense – fuel for the fire*. Mol Plant Microbe Interact. 2009; 22:487-497.

Bonas U, Anckerveken GV. *Gene-for-gene interactions: bacterial avirulence proteins specify plant disease resistance*. Curr Opin Plant Biol. 1999; 2:94-98.

Bonasera JM, Kim JF, Beer SV. *PR genes of apple: identification and expression in response to elicitors and inoculation with Erwinia amylovora.* BMC Plant Biol. 2006; 6:23-34.

Broekaert WF, Terras FRG, Cammue BPA. In A. J. Slusarenko, R. S. S. Fraser, L. C. van Loon, eds. *Induced and preformed antimicrobial proteins. Mechanisms of Resistance to Plant Diseases*. Dordrecht: Kluwer; 2000:371-477.

Cannon SB, May GD, Jackson SA. *Three sequenced Legume Genomes and many crop species: Rich opportunities for Translational Genomics*. Plant Physiol. 2009; 151:970-977.

Chester KS. *The Problem of Acquired Physiological Immunity in Plants*. Quart Rev Phytopathol. 1933; 42:185-209.

Chung G., Singh RJ. *Broadening the genetic base of soybean: A multidisciplinary approach.* Crit Rev Plant Sci. 2008;27:295-341.

Dinesh-Kumar S, Whitham S, Choi D and 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.

Dixon MS, Jones DA, Keddie JS, Thomas CT, Harrison K, Jones JDG. *The tomato Cf2 disease resistance locus comprises two functional genes encoding leucine rich repeats proteins*. Cell. 1996; 84:451-459.

Durrant WE, Dong X. *Systemic Acquired Resistance*. Ann Rev Plant Pathol. 2004; 42:185-209.

Eisen MB, Spellman PT, Brown PO, Botstein B. *Cluster analysis and display of genome-wide expression patterns*. Genetics. 1998; 25:14863-14868

Ellis J, Jones D. *Structure and function of proteins controlling strain-specific pathogen resistance in plants*. Curr Opin Plant Biol. 2000; 1:288-293.

Ellis JG, Lawrence GJ, Finnegan EJ, Anderson PA. *Contrasting complexity of two rust resistance loci in flax.* Proc Nat Acad Sci, 1995; 92:4185-4188.

Gaffney T, Friedrich L, Vernooij B et al. *Requirement of Salycilic Acid for the Induction of Systemic Acquired Resistance*. Science. 1993; 261:754-756.

Gill N, Findley S, Walling JG, et al. *Molecular and chromosomal evidence for allopolyploidy in soybean, Glycine max (L.) Merr.* Plant Physiol. 2009;151:1167-1174.

Griffith M, Yaish MWF. *Antifreeze proteins in overwintering plants: a tale of two activities*. Trends Plant Sci. 2004;9:399-405.

Hammond-Kosack KE, Jones JDG. *Plant disease resistance genes.* Ann Rev Plant Physiol. 1997; 48:575-607.

Hon WC, Griffith M, Mlynarz A, Kwok YC, Yang DSC. *Antifreeze proteins in winter rye are similar to pathogenesis-related proteins*. Plant Physiol. 1995; 109:879-889.

Innes RW, Torregrosa CA, Tom Ashfield, et al. *Differential accumulations of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidyin the ancestor of soybean*. Plant Physiol. 2008; 148:1740-1759.

Joahal GS, Briggs SP. *Reductase activity encodes by the Hm1 resistance gene in maize.* Science. 1992; 198:985-987.

Kanazin V, Marek LF, Shoemaker RC. *Resistance gene analogs are conserved and clustered in soybean*. Proc Nat Acad Sci. 1996; 93:11746-11750.

Kido EA, Barbosa PK, Ferreira Neto JCR et al. *Identification of plant protein kinases in response to abiotic and biotic stresses using SuperSAGE.* Curr Prot Pep Sci. 2011; 12:643-656.

Kitajima S, Sato F. *Plant pathogenesis-related proteins: Molecular mechanisms of gene expression and protein function*. J Biochem. 1999; 125:1-8.

Kumar S, Tamura K, Nei M. *MEGA 3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment*. Brief Bioinform. 2004; 5:150-163.

Lavin M, Herendeen PS, Wojciechowski MF. *Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary*. Syst Biol. 2005; 54:575-594.

Lawrence GJ, Finnegan EJ, Ayliffe MA, Ellis JG. *The L6 gene for flax rust resistance is related to the Arabidopsis bacterial resistance gene RPS2 and the tobacco viral resistance gene N*. Plant Cell, 1995; 7:1195-1206.

Leubner-Metzger G. *β-1,3-glucanase gene expression in low-hydrated seeds as a mechanism for dormancy release during tobacco after-ripening*. Plant J. 2005; 41:133-145.

Li L, He H, Zhang J et al. *Transcriptional analysis of highly syntenic regions between Medicago truncatula and Glycine max using tiling microarrays*. Genome Biol. 2008; 9:R57.

Li X, Xia B, Jiang Y et al. *A new pathogenesis-related protein, LrPR4, from Lycoris radiata, and its antifungal activity against Magnaporthe grisea*. Mol Biol Rep. 2010; 37:995–1001.

Libault M, Farmer A, Joshi T et al. *An integrated transcriptome atlas of the crop model Glycine max and its use in comparative analyses in plants*. Plant J. 2010;63:86-99.

Liu B, Zhang S, Zhu X et al. *Candidate defense genes as predictors of quantitative blast resistance in rice.* Mol Plant Microbe Int. 2004; 17:1146-1152.

Maisonneuve B, Bellec Y, Anderson P, Michelmore RW. *Rapid mapping of two genes for resistance to downey mildew from Lactuca serriola to existing clusters of resistance genes.* Theor Appl Genet.1994; 89:96-104.

Melotto M, Coelho MF, Pedrosa-Harand A, Kelly JD, Camargo LE. *The anthracnose resistance locus Co-4 of common bean is located on chromosome 3 and contains putative disease resistance-related genes*. Theor Appl Genet. 2004; 109:690-699.

Metzler MC, Cutt JR, Klessig DF. *Isolation and characterization of a gene encoding a PR-1 like protein from Arabidopsis thaliana*. Plant Physiol. 1991; 96:346-348.

Mindrinos M, Katagiri F, Yu GL, Ausubel FM. *The Arabidopsis thaliana disease resistance gene encodes a protein containing a nucleotide-binding site and leucine rich repeats*. Cell. 1994; 78:1089-1099.

Nanda AK, Andrio E, Marino D, Pauly N, Dunand C. *Reactive Oxygen Species during plant-microorganism early interactions*. J Integr Plant Biol. 2010; 52:195-204.

Nurnberg T, Brunner F. *Innate Immunity in plants and animals: emerging parallels between the recognition of general elicitors and pathogen-associated molecular patterns*. Curr Opin Plant Biol. 2002; 5:318-324.

Oh CS, Martin GB. *Effector-triggered immunity mediated by the Pto kinase.* Trends Plant Sci. 2011; 16:132-40.

Page RD. *Treeview program, version 1 6 1*. Comp Appl Biosci. 1996; 12:357-358.

Park JM. *The Hypersensitive Response. A Cell Death during Disease Resistance*. Plant Pathol J. 2005; 21:99-101.

Gepts P, Beavis WD, Brummer EC et al. *Legumes as a Model Plant Family. Genomics for Food and Feed Report of the Cross-Legume Advances through Genomics Conference. Plant Physiol*. 2005;137:1228-1235.

Pelegrini PB, del Sarto RP, Silva ON, Franco OL, Grossi-de-As MF *Antibacterial Peptides from Plants: What They Are and How They Probably Work*. Biochem Res int. 2011;doi:10.1155/2011/250349.

Pryor, T. *Origin and structure of disease resistance genes in plants.* Trends Genet. 1987;3:157-161.

Pryor T, Ellis J. *The genetic complexity of fungal resistance genes in plants*. Adv. Plant Pathol. 1993;10:281-305.

Rayapati PJ, Lee M, Gregory JW, Wise RP. *A linkage map of diploid* Avena *based on RFLP loci and a locus conferring resistance to nine isolates of* Puccinia coronata *var avenae.* Theoret Appl Genet. 1994;89:831-837.

Salmeron JM, Oldroyd GE, Rommens CM et al. *Tomato Prf is a member of the leucine rich repeats class of plant disease resistance genes and lies embedded within the Pto kinase gene cluster.* Cell. 1996;86:123-133.

Schmutz J, Cannon SB, Schlueter JA et al. Genome sequence of the paleopolyploid soybean. Nature. 2010;463:178-183.

Seo PJ, Lee AK, Xiang F, Park CM. *Molecular and Functional Profiling of Arabidopsis Pathogenesis-Related Genes: Insights into Their Roles in Salt Response of Seed Germination*. Plant Cell Physiol. 2008;49:334-344.

Shoemaker RC, Schlueter J, Doyle JJ. *Paleopolyploidy and gene duplication in soybean and other legumes*. Curr Opin Plant Biol. 2006;9:104-109.

Soares-Cavalcanti NM, Belarmino LC, Kido EA et al. *Overall picture of expressed Heat Shock Factors in Glycine max, Lotus japonicus and Medicago truncatula.* Gen Mol Biol. 2012;(in press).

Song WY, Wang GL, Chen LL et al. *A receptor kinase-like protein encoded by the rice disease resistance gene Xa21*. Science. 1995; 270:1804-1806.

Song WY, Pi LY, Wang GL, Gardner J, Holsten T, Ronald PC. *Evolution of the rice Xa21 disease resistance genes family*. Plant Cell. 1997; 9:1279-1287.

Sparla F, Rotino L, Valgimigli MC, Pupillo P, Trost P. *Systemic resistance induced by benzothisdizole in pear inoculated with the agent of fire blight*. Sci Hortic. 2004; 101:296-279.

Sudupak MA, Bennetzer JL, Hulbert SH. *Unequal exchange and meiotic instability of disease-resistance genes in the Rp1 region of maize*. Genetics. 1993;133:119-125.

Swiderski MR, Birker D, Jones JD. *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.

Tang X, Xie M, Kim YJ, Zhou J, Klessing DF, Martin GB. *Overexpression of Pto activates defense responses and confers broad resistance*. Plant Cell. 1999; 11:15-29.

The *Arabidopsis* Genome Initiative (2000). *Analysis of the genome sequence of the flowering plant Arabidopsis thaliana*. Nature. 2000;408:796-815.

Thiel T, Graner A, Waugh R, Grosse I, Close TJ, Stein N. *Evidence and ecolutionary of ancient whole-genome duplication in barley predating the divergence from rice.* BMC Evol Biol. 2009;9:209-227.

Tornero P, Gadea J, Conejero V, Vera P. *Two PR-1 genes from tomato are differentially regulated and reveal a novel mode of expression for a pathogenesis-related gene during the hypersensitive response and development*. Plant Micr Interact. 1997; 10:624-634.

Van Loon LC, Geraats BPJ, Linthorst HJM. *Ethylene as a modulator of disease resistance in plants*. Trends Plant Sci. 2006;11:184-191.

Van Loon LC, Pierpoint WS, Boller Th, Conejero T. *Recommendations for naming plant pathogenesis-related proteins*. Plant Mol Biol Rep. 1999;12:245-264.

Velazhahan R, Muthukrishnan S. *Transgenic tobacco plants constitutively overexpressing a rice thaumatin-like protein (PR-5) show enhanced resistance to Alternaria alternata*. Plant Biol. 2003;47:347-354.

Walsh, J.B. *How often do duplicated genes evolve new fuction?* Genetics. 1995;139:421-428.

Wanderley-Nogueira AC, Mota N, Lima-Morais D, Silva LCB, Silva AB, Benko-Iseppon AM. *Abundance and Diversity of Resistance (R) Genes in the Sugarcane Transcriptome*. Gen Mol Res. 2007; 6:866-889.

Wanderley-Nogueira, AC, Kido EA, Soares-Cavalcanti NM, Belarmino LC, Benko-Iseppon AC. *Insight on Pathogen Defense Mechanisms in the Sugarcane Transcriptome.* Special Issue (SI) of Functional Plant Science and Biotechnology on Sugarcane Pathology. Global Science Book, UK 2012.

Wanderley-Nogueira AC, Soares-Cavalcanti NM, Bellarmino LC, Bezerra-Neto JP, Benko-Iseppon AM. *An overall evaluation of the Resistance (R) and Pathogenesis Related (PR) superfamilies in soybean, medicago and arabidopsis*. Gen Mol Biol. 2012; in press.

Wang GL, Holsten TE, Song WY, Wang HP, Ronald PC. *Construction of a rice bacterial artificial chromosome library and identification of clones linked to the Xa21 disease resistance locus*. Plant J. 1995; 7:525-533.

Wang N, Xiao B, Xiong L. *Identification of a cluster of PR4-like genes involved in stress responses in rice.* J Plant Physiol. 2011; doi:10.1016/j.jplph.2011.07.013.

Wendell J. *Genome evolution in polyploids*. Plant Mol Biol. 200; 42:225-249.

Weng JK, Banks JA, Chapple C. *Parallels in lignin biosynthesis: A study in Selaginella moellendorffii reveals convergence across 400 million years of evolution*. Comm Int Biol. 2008; 1:20-22.

Wilkstrom N, Savolainen V, Chase MW. *Evolution of the angiosperms: calibrating the family tree.* Proc Soc Biol Sci. 2001; 268:2211-2220.

Yan HH, Mudge J, Kim DJ, Shoemaker RC, Cook DR, Young ND. *Comparative physical mapping reveals features of microsynteny between Glycine max, Medicago truncatula, and Arabidopsis thaliana.* Genome. 2004;47:141–155.

Zeier J, Pink B, Mueller MJ, Berger S. *Light conditions influence specific defense responses in incompatible plant-pathogen interactions: uncoupling systemic resistance from salicylic acid and PR-1 accumulation*. Planta. 2004; 219:673-683.

**Internet Resources**

Expert protein Analysis System (Expasy). Available at: http://expasy.org.uk. Acessed January 18 2012.

Medicago sequencing resource website. Available at: http://www.medicago.org. Acessed October 19 2010.

SoyBase. Available at: http://soybase.org. Acessed January 25 2012.

Table 1. *Vigna unguiculata* (Cowpea) identified *PR* and *R* candidates and their features, as compared with the number of matches for *Medicago truncatula* (Medicago) and *Glycine max* (Soybean).

| | Cowpea | | | | | | Medicago | Soybean | |
|---|---|---|---|---|---|---|---|---|---|
| Gene | Best match | e-value | score | Size (nt) | Size (aa) | Domain | match es | match es | match es |
| *PR1* | CL2909Contig1 | 2e-53 | 203 | 814 | 182 | - | 6 | 19 | 8 |
| *PR2* | CL2718Contig1 | 3e-101 | 363 | 1255 | 341 | Glyco_hydro_17 superfamily | 27 | 46 | 31 |
| *PR3* | CL2010Contig1 | 2e-124 | 441 | 1567 | 322 | chitinase_glyco_hydro_19 | 12 | 21 | 7 |
| *PR4* | CL919Contig1 | 2e-66 | 247 | 815 | 202 | - | 3 | 14 | 2 |
| *PR5* | CL4214Contig1 | 5e-65 | 243 | 1396 | 316 | Thaumatin-like protein | 19 | 36 | 21 |
| *PR6* | CL2909Contig1 | 2e-50 | 193 | 814 | 66 | - | 4 | 17 | 11 |
| *PR7* | CL812Contig1 | 0.0 | 760 | 2617 | 718 | Peptidase S8 family domain | 60 | 97 | 82 |
| *PR8* | CL6270Contig1 | 2e-94 | 341 | 1115 | 300 | Hevamine-glycosyl hydrolase 18 | 7 | 22 | 11 |
| *PR9* | CL1415Contig1 | 9e-129 | 455 | 1360 | 329 | Secretory peroxidase | 77 | 214 | 86 |
| *PR10* | - | - | - | - | - | - | - | 18 | 18 |
| *PR11* | VUPIST08021A05.B00 | 7e-25 | 89 | 534 | 97 | - | 1 | 11 | 1 |
| *PR12* | CL83Contig2 | 3e-14 | 72.4 | 564 | 95 | Knot1 - plant defensin | 6 | 15 | 1 |
| *PR13* | - | - | - | - | - | - | - | - | - |
| *PR14* | CL220Contig1 | 3e-26 | 112 | 744 | 117 | Lipid-transfer protein | 28 | 36 | 18 |
| *PR15* | CL2531Contig1 | 7e-50 | 192 | 965 | 217 | Cupin | 23 | 47 | 27 |
| *PR16* | CL1672Contig1 | 4e-64 | 239 | 819 | 226 | Cupin2 | 25 | 51 | 27 |
| *PR17* | CL2038Contig1 | 2e-72 | 267 | 1155 | 225 | Plant basic secretory protein | 4 | 1 | 2 |
| *Pto* | CL4211Contig2 | 4e-169 | 590 | 972 | 291 | Ser/Thre Kinase | 140 | 239 | 238 |
| *Prf* | VU524SDGMI1239A05GB | 9e-25 | 110 | 758 | 223 | NB-ARC | 3 | 25 | 5 |
| *Pti4* | CL2861Contig1 | 2e-56 | 214 | 1142 | 248 | APETALA | 60 | 90 | 89 |
| *Pti5* | CL11265Contig1 | 2e-36 | 147 | 1039 | 281 | APETALA | 62 | 70 | 70 |
| *Pti1* | CL1209Contig2 | 2e-149 | 524 | 1341 | 361 | Kinase | 98 | 112 | 104 |
| *Pti6* | CL3091Contig1 | 4e-22 | 100 | 2058 | 264 | APETALA | 99 | 249 | 248 |
| *RAR1* | CL3813Contig1 | 5e-68 | 252 | 1056 | 209 | Zn binding domain | 2 | 2 | 1 |
| *RIN4* | CL1984Contig1 | 1e-35 | 144 | 1238 | 245 | AvrRpt2 auto-cleavage site | 4 | 8 | 2 |
| *RPM1* | VUMIXFLGLFM149C06GB | 8e-42 | 166 | 777 | 258 | NB-ARC | 14 | 73 | 14 |
| *RPS2* | VUITKSDGMI1175E08GB | 6e-23 | 103 | 814 | 261 | NB-ARC | 9 | 36 | 4 |
| *PBS1* | CL4962Contig1 | 3e-148 | 314 | 1854 | 193 | Kinase | 69 | 247 | 239 |
| *RPS5* | VUITKSDGMI1175E08GB | 2e-23 | 105 | 814 | 261 | NB-ARC | 9 | 36 | 5 |
| *MLA10* | VUMIXFLGLFM149C06GB | 2e-10 | 60.8 | 777 | 258 | NB-ARC | 2 | 21 | 0 |
| *L6* | VUUCRSDGMI1249D03GB | 1e-34 | 141 | 769 | 254 | TIR / NB-ARC | 23 | 123 | 24 |
| *RRS1* | CL12357Contig1 | 1e-40 | 104 | 1191 | 258 | LRR | 43 | 142 | 102 |
| *RPS4* | CL12357Contig1 | 1e-40 | 103 | 1191 | 258 | LRR | 43 | 198 | 50 |
| *Xa1* | VUMIXFLGLFM149C06GB | 3e-59 | 224 | 777 | 258 | NB-ARC | 31 | 108 | 17 |
| *Hrt* | CL12357Contig1 | 5e-53 | 130 | 1191 | 258 | LRR | 43 | 208 | 61 |
| *Mi1* | VUITKSDGMI1175E08GB | 8e-25 | 110 | 814 | 261 | NB-ARC | 9 | 29 | 1 |
| *BS2* | VUITKSDGMI1175E08GB | 2e-38 | 155 | 814 | 261 | NB-ARC | 9 | 68 | 9 |
| *GPA2* | VUITKSDGMI1175E08GB | 6e-42 | 167 | 814 | 261 | NB-ARC | 9 | 50 | 10 |
| *RX1* | CL15438Contig1 | 3e-15 | 80 | 709 | 227 | NB-ARC | 21 | 61 | 14 |
| *Pi-ta* | VUITKSDGMI1175E08GB | 1e-25 | 112 | 814 | 261 | NB-ARC | 6 | 17 | 2 |
| *I2* | VUMIXFLGLFM149C06GB | 6e-56 | 214 | 777 | 258 | NB-ARC | 35 | 109 | 22 |
| *RPP8* | CL15438Contig1 | 6e-23 | 106 | 709 | 227 | NB-ARC | 13 | 71 | 11 |
| *HERO* | VUITKSDGMI1175E08GB | 3e-31 | 132 | 814 | 261 | NB-ARC | 8 | 39 | 5 |
| *RPP13* | VUITKSDGMI1175E08GB | 1e-37 | 152 | 814 | 261 | NB-ARC | 15 | 51 | 2 |
| *RP1* | VUMIXFLGLFM149C06GB | 1e-39 | 159 | 777 | 258 | NB-ARC | 23 | 71 | 14 |
| *N* | CL12357Contig1 | 1e-49 | 134 | 1191 | 258 | LRR | 22 | 196 | 64 |
| *P* | CL4681Contig1 | 1e-49 | 191 | 786 | 182 | Dirigent superfamily | 18 | 37 | 17 |
| *WRKY25* | CL3337Contig1 | 2e-75 | 278 | 2043 | 548 | WRKY | 63 | 52 | 68 |
| *WRKY33* | CL3337Contig1 | 9e-94 | 339 | 2043 | 548 | WRKY | 66 | 58 | 71 |
| *WRKY29* | CL4331Contig1 | 1e-26 | 116 | 1069 | 139 | - | 23 | 22 | 28 |
| *Cf2* | CL12438Contig1 | 9e-85 | 237 | 1149 | 233 | LRR | 91 | 250 | 249 |
| *Cf4* | CL14600Contig1 | 3e-26 | 116 | 858 | 173 | LRR | 189 | 208 | 116 |
| *Cf5* | CL1797Contig1 | 2e-35 | 146 | 1721 | 367 | LRR | 285 | 207 | 123 |
| *Cf9* | CL14600Contig1 | 2e-29 | 125 | 858 | 173 | LRR | - | 188 | 107 |
| *Xa21* | CL11330Contig1 | 6e-78 | 288 | 1261 | 335 | LRR / Kinase | 162 | 249 | 154 |
| *FLS2* | CL9027Contig1 | 4e-57 | 219 | 783 | 118 | Kinase | 137 | 251 | 251 |
| *EFR* | CL11330Contig1 | 1e-65 | 247 | 1261 | 335 | LRR / Kinase | 101 | 239 | 250 |

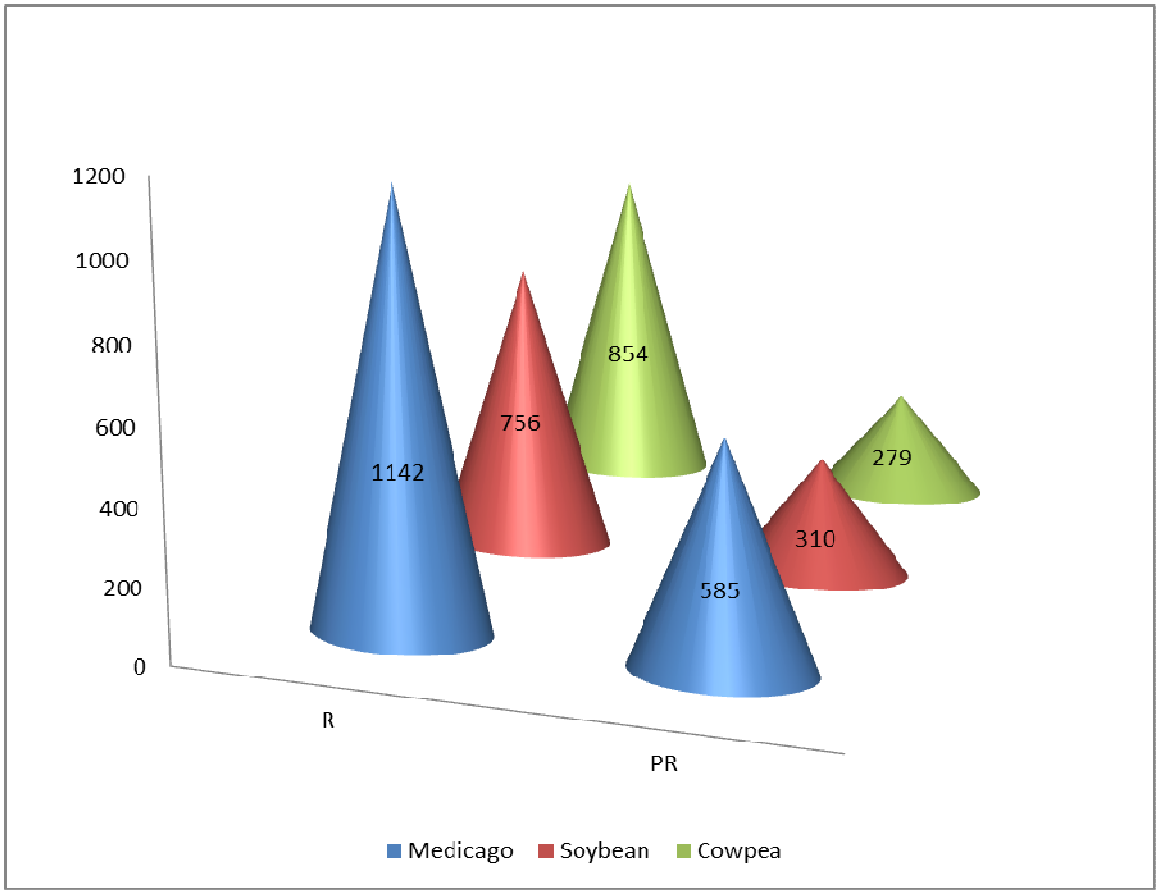



Figure 1: Number of *R* and *PR* genes in *Medicago truncatula* (blue), *Glycine max* (red) and *Vigna unguiculata* (green).

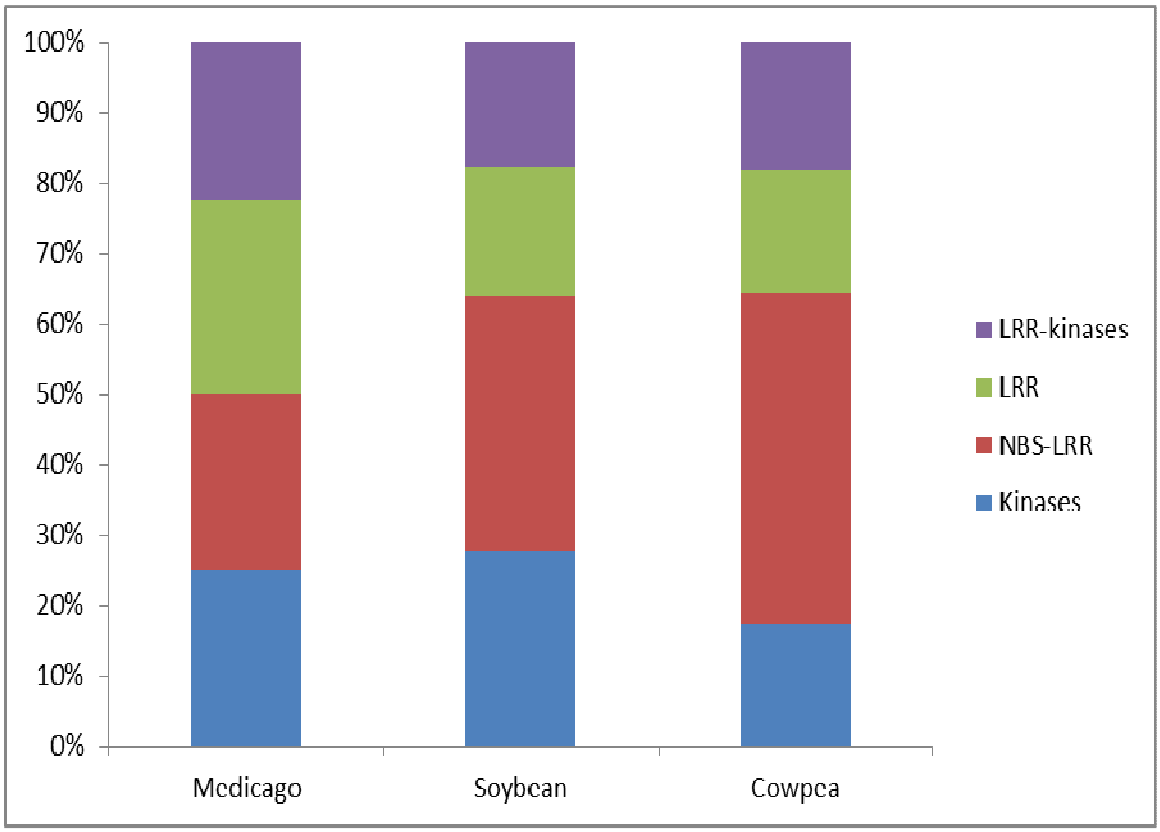



Figure 2: Distribution of the main *R* genes classes (indicated by different colours) in *Medicago truncatula*, *Glycine max* and *Vigna unguiculata*.

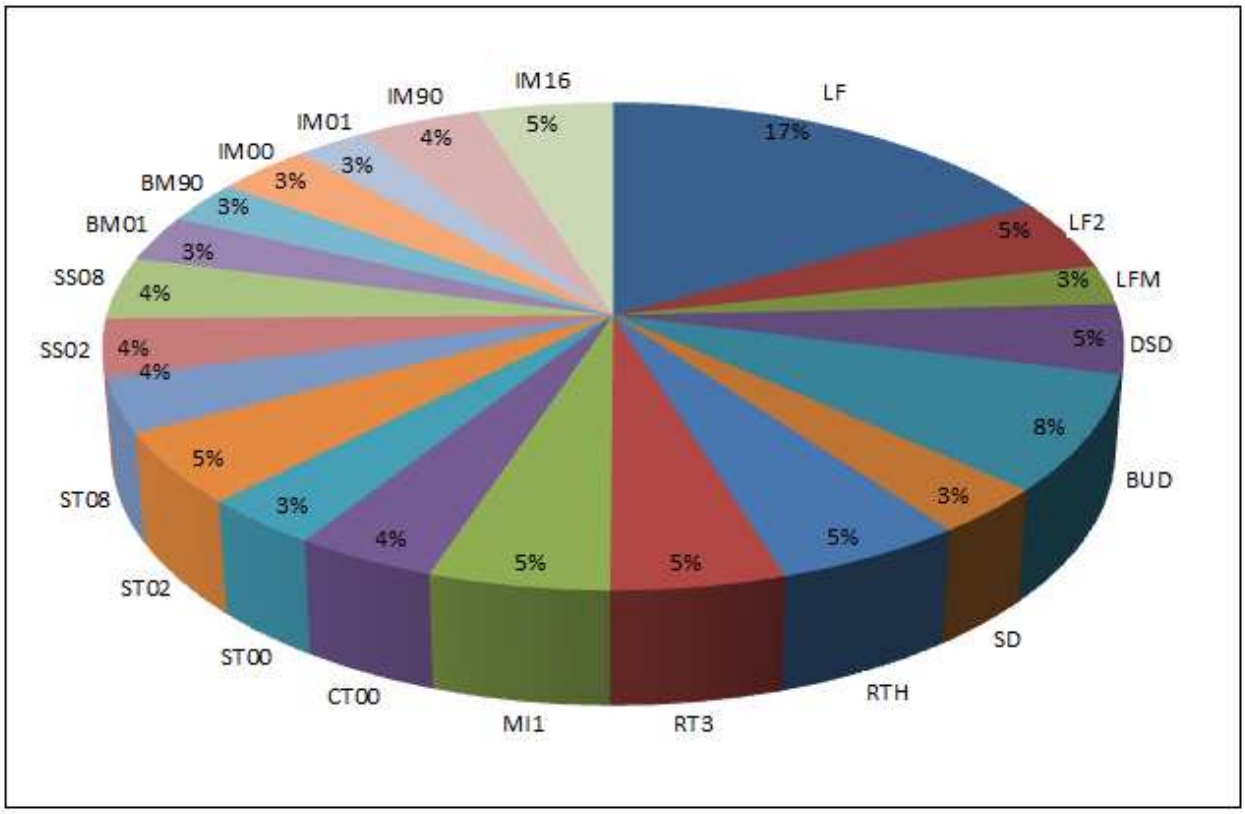



Figure 3: *R* and *PR* genes transcripts prevalence in 22 different cowpea libraries including: **SS00**: salinity susceptible accession (cultivar 'Canapu Amarelo') roots without stress; **SS02**: roots of 'Canapu Amarelo' two hours after salinity stress; **SS08**: roots of 'Canapu Amarelo' eight hours after salinity stress; **ST00**: tolerant accession ('Pitiúba') roots without stress (control); **ST02**: roots of 'Pitiúba' cultivar two hours after salinity stress; **ST08**: 'Pitiúba' roots two hours after salinity stress; **BM01**: leaves of the CPSMV (cowpea severe mosaic virus) resistant cultivar 'BR14-Mulato' without stress (control); **BM90**: 'BR14-Mulato' leave infected with CPSMV (bulk of 30, 60 and 90 min.) after stress; **IM01**: leave of the CPSMV susceptible accession IT85F-2687 without stress (control); **IM90**: IT85F-2687 leave infected by CPSMV (bulk of three times 30, 60 and 90 min.) after stress; **BUD**: growing axillary buds; **DSD**: developing seeds; **LF**: leave; **LF2**: primary leaflets; **LFM:** leaflets and shoot meristem; **MI1**: mixed tissues; **NDL**: nodules; **RT2**: root hairs; **RT3**: roots; **SD**: seeds.

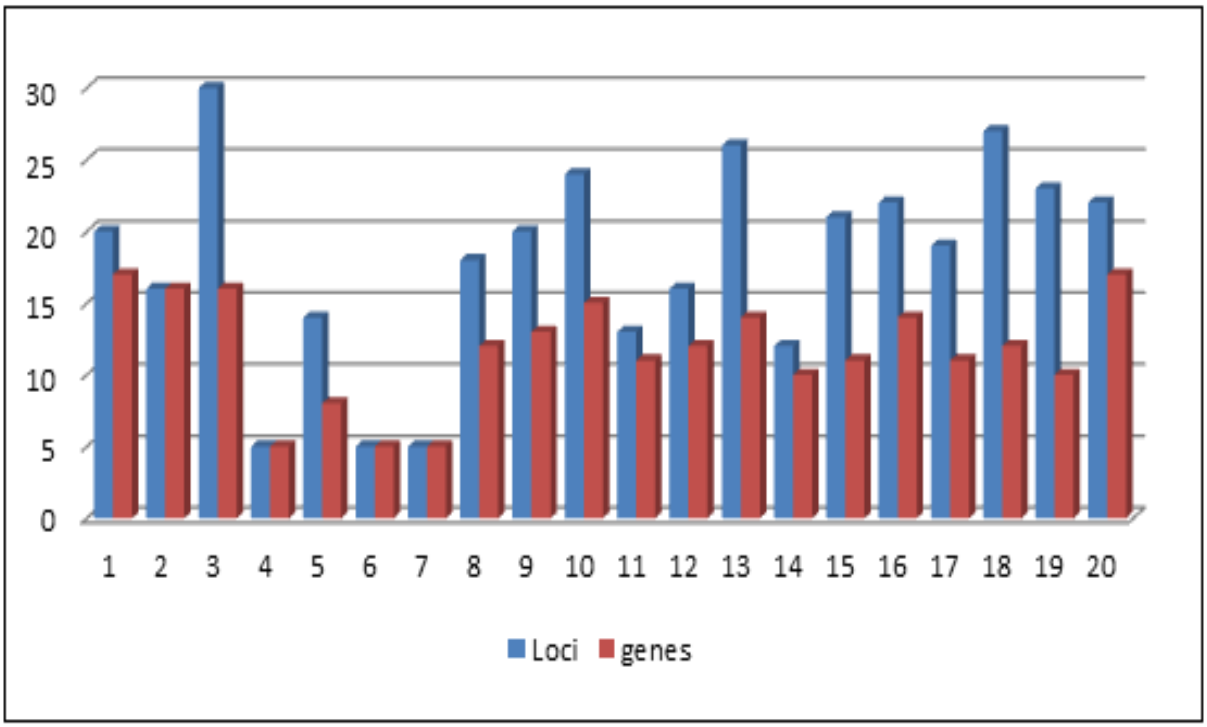



Figure 4: Schematic representation of cowpea *R* and *PR* genes alignments in soybean chromosomes. Blue columns regard number of loci per chromosome while red columns represent the number of non-redundant genes. The x-axis represents soybean chromosome number.

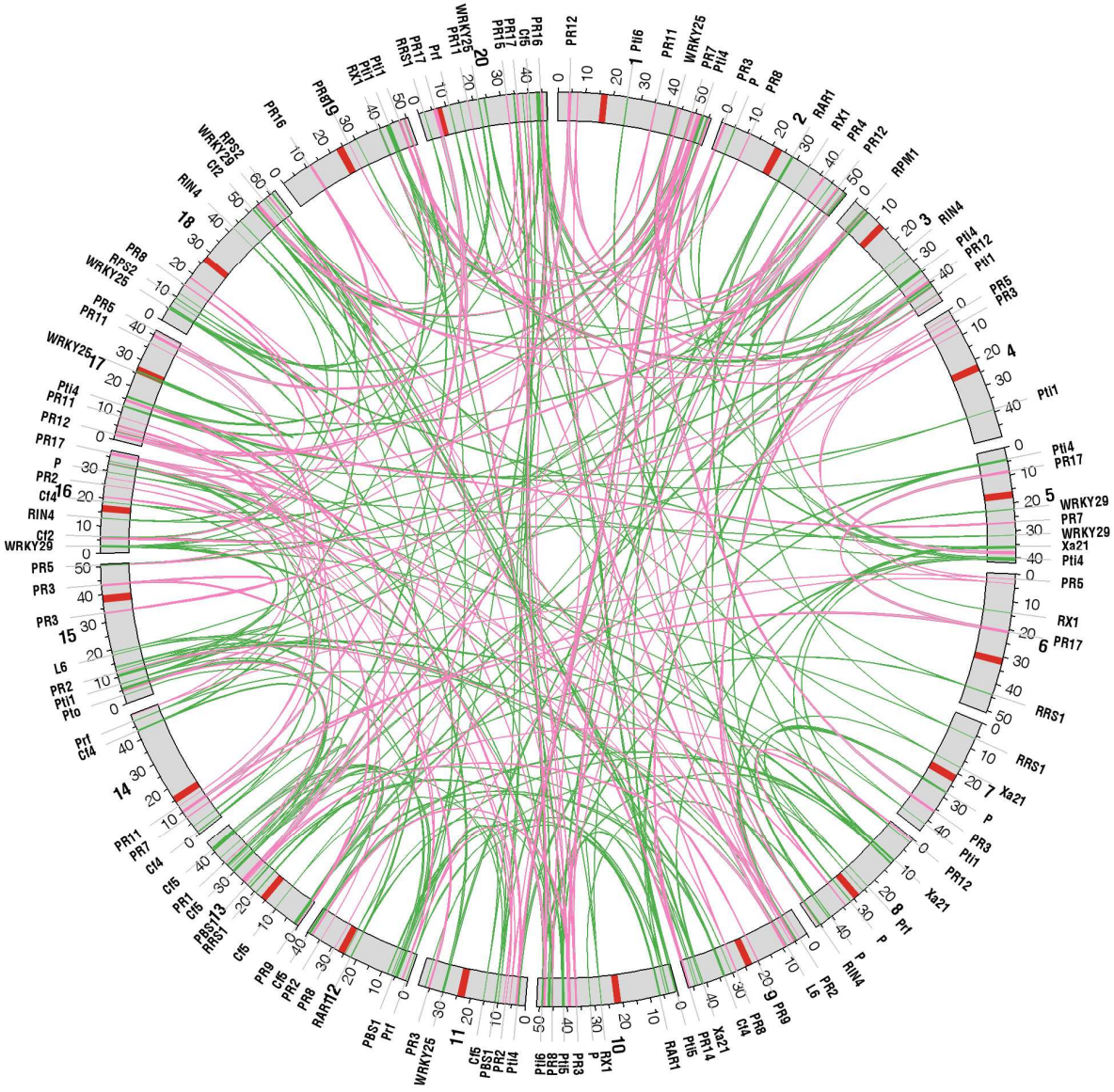


Figure 5: Distribution and microsyntenic relationships considering *R* (in green) and *PR* (in pink) candidates identified in cowpea against the soybean genome (2n=40). Chromosomes are depicted with the centromere (in red) in their expected position. Scale = 1 Mb. Microsyntenic relationships are shown as links between chromosome regions.

# CONCLUSÕES

- As plantas superiores analisadas apresentam os principais mecanismos de resposta ao ataque de patógenos indicando a existência destes mecanismos no ancestral comum das mesmas;

- Os genomas da soja, caupi e cana-de-açúcar apresentam ortólogos dos principais genes envolvidos nas vias responsivas ao estresse biótico, sejam os que atuam diretamente nas respostas, ou aqueles que funcional em etapas intermediárias;

- Com raras exceções, em todos os genomas, pelo menos um representante dos ortólogos dos genes *R* e *PR* procurados foi identificado;

-A classe de genes *R* mais representativa foi a NBS-LRR em todos os organismos estudados;

- De uma forma geral, o fato de os genes estudados serem mais parecidos em organismos pertencentes à mesma família indica que os mesmos surgiram antes da separação entre mono e dicotiledôneas;

- O número de cópias dos genes *R* e *PR* estudados bem como seu posicionamento dentro dos genomas parecemser resultado do acúmulo de duplicações e *crossing over* ocorridos durante o curso da evolução e especiação;

- Os tecidos que apresentam maior expressão, na maioria das vezes, são aqueles que primeiro entram em contato com o patógeno, como raiz e folha;

-O padrão de expressão dos genes *PR* mostra que os mesmos são ativados tanto por estresse biótico quanto abiótico.

- A presença dos genes descritos tanto em mono quanto em dicotiledôneas abre caminhos promissores para futuros estudos, especialmente considerando o melhoramento genético de plantas de interesse agrícola.

Anexo I. Instruções para autores *Bioinformatics and Biology Insights*

## Description of the format

References to a work or a part of a work inside the manuscript appear as numbered citations. At the end of the manuscript each reference is listed in full prefixed by the number used in the manuscript.

Within the manuscript numbered citations are generally formatted as superscripted numbers starting from 1 and numbered in the order in which they appear in the manuscript. Numbers appear after the closest following comma or full stop, but before the closest following colon or semicolon.

Where multiple sources need to be cited at the same point separate the numbers with commas.

Where page numbers need to be cited after a reference number, they should be placed in parentheses following the citation number.

Parenthetical citations are used within the text for sources not suitable for inclusion in the references list at the end of the paper. Typically personal communications and unpublished data would be treated in this way. Within the brackets, the author's name, including their initials, and an indication of the type of source, i.e. "unpublished data", and the month and year should be given.

Authors' surnames are given in full, and for their first and middle names only the initials are given following the surname. Where there are six or less authors each name is separated by a comma without any other words. Where there are more than six authors only the first three authors are given followed by "et al."

## Monographs

Capitalize the significant words of the title. Words such as "and," "the," "to" and "or" are not capitalized unless they appear at the beginning of the title. Italicize the entire title.

Smith AA. *Lorem Ipsum Dolor Sit Amet, Consectetuer Adipiscing Elit*. New York: Oxford University Press; 1988.

## Monographs with specific pages or multiple authors

Note that more than six authors are treated differently. A page or page range is given after the copyright year and separated from the copyright year using a colon.

Smith AA, Brown BB, Jones CC. *Lorem Ipsum Dolor Sit Amet, Consectetuer Adipiscing Elit*. New York: Oxford University Press; 1988:23-36.

## Monographs in second or later editions

Smith AA, Brown BB, Jones CC. Lorem Ipsum Dolor Sit Amet, *Consectetuer Adipiscing Elit*. 9th ed. New York: Oxford University Press; 1988.

### Where a monograph has no listed author or editor

Where this is the case, substitute the translator or editor's name.

### Edited volumes

Indicate this using "ed." and by appending information indicating the series to which the book belongs where appropriate.

Smith AA, ed. *Lorem Ipsum Dolor Sit Amet, Consectetuer Adipiscing Elit*. New York: Oxford University Press; 1988. New American Latin Series; No. 4.

### Articles or separately authored chapters in edited volumes

Smith AA. Lorem Ipsum Dolor. In: Brown BB, Jones CC, eds. *Lorem Ipsum Dolor Sit Amet, Consectetuer Adipiscing Elit*. New York: Oxford University Press; 1988:32-46

### Multi-volume works

Smith AA, Brown BB, Jones CC. *Lorem Ipsum Dolor Sit Amet, Consectetuer Adipiscing Elit*. Vol 2. New York: Oxford University Press; 1988.

### Electronic and online resources

Where the resource is accessed through a database indicate which database. If it is accessed through a website provide the URL. In both cases indicate the type of resource and when it was accessed.

Smith AA, Brown BB, Jones CC. *Lorem Ipsum Dolor Sit Amet, Consectetuer Adipiscing Elit* [e-book]. New York: Oxford University Press; 1988. Available from: Netlibrary. Accessed June 4 2004.

Smith AA, Brown BB, Jones CC. *Lorem Ipsum Dolor Sit Amet, Consectetuer Adipiscing Elit* [e-book]. New York: Oxford University Press; 1988. Available at: http://www.loremipsumdolor.net Accessed June 4 2004.

### Scholarly journals

Capitalize the first letter in the first word of article titles and also any proper names or abbreviations which are normally capitalized, i.e. IBM. Journal titles are fully capitalized and italicized.

Brown BB. *Lorem ipsum dolor sit amet, consectetuer adipiscing elit*. Lorem Ipsum Dolor. 1988;2:124-143.

Anexo II. Súmula curricular no período de 2008 a 2012

## Ana Carolina Wanderley Nogueira

### Formação Acadêmica/Titulação

**2008** Doutorado em Ciências Biológicas.
Universidade Federal de Pernambuco, UFPE, Recife, Brasil
Título: CARACTERIZAÇÃO BIOINFORMÁTICA DE GENES RELACIONADOS À INTERAÇÃO PATÓGENO-HOSPEDEIRO EM ANGIOSPERMAS
Orientador: Ana Maria Benko-Iseppon
Bolsista da: Fundação de Amparo à pesquisa de Pernambuco

**2005 - 2007** Mestrado em Genética.
Universidade Federal de Pernambuco, UFPE, Recife, Brasil
Título: Caracterização e Análise In silico de genes de resistência a patógenos no genoma expresso da Cana-de-Açúcar, Eucalipto e feijão Caupi., Ano de obtenção: 2007
Orientador: Ana Maria Benko-Iseppon
Bolsista do: Conselho Nacional de Desenvolvimento Científico e Tecnológico

**2000 - 2004** Graduação em Bacharelado em Ciências Biológicas.
Universidade Federal de Pernambuco, UFPE, Recife, Brasil
Título: Identificaçãoe Análise do gene Xa21 no genoma da cana-de-açúcar
Orientador: Ana Maria Benko-Iseppon
Bolsista da: Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco

### Formação complementar

**2010 - 2010** Curso de curta duração em Curso ISO 14001 - Gestão Ambiental nas Empresas.
Centro Nacional de Educação a Distância, CENED, Brasil

### Atuação profissional

### 1. Universidade Federal de Pernambuco - UFPE

**Atividades**

**03/2001 - Atual** Pesquisa e Desenvolvimento, Centro de Ciências Biológicas
*Linhas de Pesquisa:*
*Bioinformática - Genômica e Proteômica*

**03/2001 - Atual** Estágio, Centro de Ciências Biológicas, Departamento de Genética
*Estágio:*
*Genética-data mining*

**08/2000 - 02/2001** Estágio, Centro de Ciências Biológicas, Departamento de Antibióticos
*Estágio: Microbiologia*

## Linhas de pesquisa

**1.** Bioinformática - Genômica e Proteômica

Objetivos: Data-mining e análise estrutural de genes de Resistência a patógenos, seca e salinidade no genoma expresso de cana-de-açúcar, eucalipto e feijão Caupi.

## Áreas de atuação

**1.** Genética
**2.** Bioinformática
**3.** Biologia Geral

## Idiomas

**Inglês** Compreende Bem, Fala Bem, Escreve Bem, Lê Bem

**Espanhol** Compreende Bem, Fala Bem, Escreve Bem, Lê Bem

## Produção em C, T&A

## Produção bibliográfica

**Artigos completos publicados em periódicos**

1. WANDERLEY-NOGUEIRA, A.C, Mota, N.S.C, BELLARMINO, 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 VignaUnguiculataTranscriptome.Lecture Notes in Computer Science. , v.6160, p.70 - 81, 2010.
*Referências adicionais : Inglês. Meio de divulgação: Vários*

2. Soares-Cavalcanti N.M, WANDERLEY-NOGUEIRA, A.C, BELLARMINO, L. C., Barros dos Santos, P., Benko-Iseppon, A. M.
Comparative In Silico Evaluation of MYB Transcription Factors in Eucalyptus, Sugarcane and Rice Transcriptomes. Lecture Notes in Computer Science. , v.5488, p.44 - 55, 2009.
*Referências adicionais : Português. Meio de divulgação: Vários*

3. Barros dos Santos, P., Soares-Cavalcanti N.M, Vieira-Melo G.S., WANDERLEY-NOGUEIRA, A.C, Calsa Jr T., Benko-Iseppon, A. M.
In Silico Evaluation of Osmoprotectants in Eucalyptus Transcriptome.Lecture Notes in Computer Science. , v.5488, p.66 - 77, 2009.
*Referências adicionais : Português. Meio de divulgação: Vários*

**Artigos aceitos para publicação**

1. BELLARMINO, L. C., WANDERLEY-NOGUEIRA, A.C, Soares-Cavalcanti N.M, Pandolfi, V., Amorim L. L. B., Kahj, Winter, Benko-Iseppon, A. M.
Analysis and Annotation of Chickpea Bacterial Artificial Chromosome Sequences. Lecture Notes in Computer Science. , 2012.
*Referências adicionais : Inglês.*

2. Amorim L. L. B., BELLARMINO, L. C., Piereck B., WANDERLEY-NOGUEIRA, A.C, Soares-Cavalcanti N.M, Brasileiro-Vidal A.C., Benko-Iseppon, A. M.
BIOINFORMATICS ANALYSIS OF CLASS II TRANSPOSABLE ELEMENTS IN THE COWPEA GENOME AS COMPARED WITH THE ACTUAL SOYBEAN KNOWLEDGE. Lecture Notes in Computer Science. , 2012.
*Referências adicionais : Inglês.*

3. 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. Genetics and Molecular Biology. 2011.
*Referências adicionais : Inglês.*

4. Soares-Cavalcanti N.M, WANDERLEY-NOGUEIRA, A.C, BELLARMINO, L. C., Barros dos Santos, P., Benko-Iseppon, A. M.
Comparative in silico evaluation of MYB transcription factors in eucalyptus, sugarcane and rice transcriptomes. Lecture Notes in Computer Science. , 2011.
*Referências adicionais : Inglês.*

5. Soares-Cavalcanti N.M, BELLARMINO, L. C., Bezerra-Neto, J. P., WANDERLEY-NOGUEIRA, A.C, LIRA, R., ARRUDA, H., KIDO, E. A., Pandolfi, V., Benko-Iseppon, A. M.
DIFFERENTIAL GENE EXPRESSION IN THE PROSPECTION OF TOLERANCE GENES AGAINST DROUGHT AND SALINITY IN NATIVE AND CULTIVATED PLANTS.Revista Caatinga (UFERSA. Impresso). , 2011.
*Referências adicionais : Português.*

6. Soares-Cavalcanti N.M, BELLARMINO, L. C., WANDERLEY-NOGUEIRA, A.C, Bezerra-Neto, J. P., LIRA, R., Benko-Iseppon, A. M.
Evaluation of the main gene families for drought and salinity in the GENOSOJA database, as compared with Medicago and Arabidopsis.Genetics and Molecular Biology (Impresso). , 2011.
*Referências adicionais : Inglês.*

7. Barros dos Santos, P., BELLARMINO, L. C., Soares-Cavalcanti N.M, WANDERLEY-NOGUEIRA, A.C, Vieira-Melo G.S., Calsa Jr T., Benko-Iseppon, A. M.
In silico evaluation of osmoprotectants in eucalyptus transcriptome.Lecture Notes in Computer Science. , 2011.
*Referências adicionais : Inglês.*

8. Soldatos T.G., O' Donoghue S.I., Satagopan V., Barbosa-Silva, A>, Pavlopoulos G. A., WANDERLEY-NOGUEIRA, A.C, Soares-Cavalcanti N.M, Schneider Reinhard
Caipirini: using gene sets to search literature. BMC Bioinformatics. , 2010.
*Referências adicionais : Inglês.*

**Capítulos de livros publicados**

1. WANDERLEY-NOGUEIRA, AC, Soares-Cavalcanti, N M, Belarmino, LC, Kido, EA, 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

**Trabalhos publicados em anais de eventos (completo)**

1. Soares-Cavalcanti N.M, WANDERLEY-NOGUEIRA, A.C, BELLARMINO, L. C., Barros dos Santos, P., Benko-Iseppon, A. M.
Comparative In Silico Evaluation of MYB Transcription Factors in Eucalyptus,Sugarcane and Rice. In: Internationa meeting on computational intelligence methods for bioinformatics and biostatistics, 2008, Salerno.
**Lecture Notes on Bioinformatics LNBI/LNCS series of Springer Verlag**. , 2008.
*Referências adicionais : Itália/Inglês. Meio de divulgação: Meio digital*

2. Barros dos Santos, P., Soares-Cavalcanti N.M, Vieira-Melo G.S., WANDERLEY-NOGUEIRA, A.C, Benko-Iseppon, A. M., Calsa Jr T.
In Silico Evaluation of Osmoprotectants in Eucalyptus Transcriptome. In: Internationa meeting on computational intelligence methods for bioinformatics and biostatistics, 2008, Salerno.
**Lecture Notes on Bioinformatics LNBI/LNCS series of Springer Verlag**. , 2008.
*Referências adicionais : Itália/Inglês. Meio de divulgação: Meio digital*

**Trabalhos publicados em anais de eventos (resumo)**

1. Soares-Cavalcanti N.M, WANDERLEY-NOGUEIRA, A.C, BELLARMINO, L. C., Pandolfi, V., Benko-Iseppon, A. M.
In silico evaluation of NAC transcription factor associated to abiotic stress in the sugarcane transcriptome In: CongressoBrtasileiro de Fisiologia Vegetal, 2009, Fortaleza.
**Anais doXII Congresso Brtasileiro de Fisiologia Vegetal**. , 2009.
*Referências adicionais : Brasil/Inglês. Meio de divulgação: Meio digital*

2. Soares-Cavalcanti N.M, WANDERLEY-NOGUEIRA, A.C, Pandolfi, V., Calsa Jr T., Barbosa-Silva, A>, Benko-Iseppon, A. M.
Quantitativeanalysisofcowpeasagetagsrelatedtoabioticandbiotic responses In: II Simposio Brasileiro de Genética Molecular de Plantas, 2009
**Anais doII Simposio Brasileiro de Genética Molecular de Plantas**. , 2009.
*Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital*

3. Vieira-Melo G.S., Soares-Cavalcanti N.M, Barros dos Santos, P., WANDERLEY-NOGUEIRA, A.C, Casteleti, C. H. M., Calsa Jr T., Benko-Iseppon, A. M.
Analise computacional do gene DMI3 no genoma expresso do eucalipto In: Congresso Brasileiro de genética, 2008, Salvador.
**Anais do Congresso Brasileiro de genética**. , 2008.
*Referências adicionais : Brasil/Português.*

4. Vieira-Melo G.S., Barros dos Santos, P., Soares-Cavalcanti N.M, WANDERLEY-NOGUEIRA, A.C, Benko-Iseppon, A. M.
Estudo in silico do gene GMN 70 de soja no transcriptoma do feijão-caupi In: 59 congresso nacional de botânica, 2008, Natal.
**Anais do 59 congresso nacional de botânica**. , 2008.
*Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital*

5. Vieira-Melo G.S., Soares-Cavalcanti N.M, WANDERLEY-NOGUEIRA, A.C, Barros dos Santos, P., Benko-Iseppon, A. M.
Estudo in silico do gene GM70 de soja no genoma do feijão-caupi In: Congresso Latinoamericano y del Caribe de cactáceas y Otras Suculentas, 2008
**Anais do Congresso Latinoamericano y del Caribe de cactáceas y Otras Suculentas**. , 2008.
*Referências adicionais : Brasil/Inglês.*

6. Barros dos Santos, P., Soares-Cavalcanti N.M, WANDERLEY-NOGUEIRA, A.C, Vieira-Melo G.S., Benko-Iseppon, A. M.
Mineração revela ortólogos aos genes CMO e BADH no genoma do feijão-caupi In: Congresso

Latinoamericano y del Caribe de cactáceas y Otras Suculentas,, 2008, Natal - RN.
**Anais do . Congresso Latinoamericano y del Caribe de cactáceas y Otras Suculentas**. , 2008.
*Referências adicionais : Brasil/Inglês. Meio de divulgação: Meio digital*

**Apresentação de Trabalho**

1. Benko-Iseppon, A. M., KIDO, E. A., Calsa Jr T., Pandolfi, V., Amorim L. L. B., Brasileiro-Vidal A.C., Soares-Cavalcanti N.M, WANDERLEY-NOGUEIRA, A.C, BELLARMINO, L. C., Barbosa-Silva, A>, BRUNELLI, K., Vieira-Melo G.S., Barros dos Santos, P., ANDRADE, P. P., Freire-Filho FR
**Cowpea (Vignaunguiculata) TranscriptomeAnd Differential Expression: The Project And Applications For Breeding Purposes**, 2011. (Conferência ou palestra,Apresentação de Trabalho)
*Referências adicionais : Brasil/Português.*

## Produção Técnica
**Demais produções técnicas**

1. Soares-Cavalcanti N.M, BELLARMINO, L. C., WANDERLEY-NOGUEIRA, A.C
**Apostila do Curso de Introdução à Bioinformática**, 2010. (Desenvolvimento de material didático ou instrucional)
*Referências adicionais : Brasil/Português.*

## Eventos
**Participação em eventos**

1. **A sustentabilidade e o Setor Industrial**, 2011. (Seminário)

2. Apresentação de Poster / Painel no**I Simposio de Inovação Cientifica**, 2008. (Simpósio)
Análise in silico dos genes que participam das respostas aos estresses bióticos nos genomas do eucalipto, cana-de-açúcar e feijão-caupi.

**Organização de evento**

1. Benko-Iseppon, A. M., BELLARMINO, L. C., Amorim L. L. B., Soares-Cavalcanti N.M, ARRUDA, H., WANDERLEY-NOGUEIRA, A.C, Sotero D, Alves G
**1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy**, 2010
*Referências adicionais : Brasil/Inglês. Meio de divulgação: Outro*

2. WANDERLEY-NOGUEIRA, A.C
**I Simpósio de Inovação em Ciências Biológicas**, 2008.
*Referências adicionais : Brasil/Português. Meio de divulgação: Outro*