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

Caracterização de genes de Resistência a Patógenos em Eucalipto (*Eucalyptus* spp.), Cana-de-açúcar (*Saccharum officinarum)* e Feijão-Caupi (*Vigna unguiculata*).

ANA CAROLINA WANDERLEY NOGUEIRA

RECIFE
2007

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

Caracterização de genes de Resistência a Patógenos em Eucalipto (*Eucalyptus* spp.), Cana-de-açúcar (*Saccharum officinarum)* e Feijão-Caupi (*Vigna unguiculata*).

ANA CAROLINA WANDERLEY NOGUEIRA

Dissertação apresentada ao Programa de Pós-graduação em Genética da Universidade Federal de Pernambuco como requisito para a obtenção do grau de Mestre em Genética.

Orientadora: Profª. Drª. Ana Maria Benko-Iseppon

Recife – PE
- 2007 -

UNIVERSIDADE FEDERAL DE PERNAMBUCO
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

Parecer da comissão examinadora da defesa de dissertação de Mestrado de

ANA CAROLINA WANDERLEY NOGUEIRA

"Caracterização de genes de resistência a patógenos em eucalipto (*Eucalyptus* spp.), cana-de-açúcar (*Saccharum* spp.) e feijão-caupi (*Vigna unguiculata*)"

Área de concentração: Genética Molecular

A comissão examinadora composta pelos professores abaixo considera a candidata

ANA CAROLINA WANDERLEY NOGUEIRA como _____.

Recife, 9 de julho de 2007.

Orientador: Drª. Ana Maria Benko-Iseppon (UFPE)

MembroExterno: Silvia Regina Batistuzzo de Medeiros (UFRN)

Membro Externo: Felipe Rodrigues da Silva (EMBRAPA - DF)

Membro Interno: Tercílio Calsa Jr. (UFPE)

Coordenador do Programa: Marcos Antonio de Moraes Jr. (UFPE)

AGRADECIMENTOS

A Deus, por permitir que tudo isso acontecesse.

A minha mãe, Sheila, pelo apoio incondicional em todos os meus projetos.

A minha irmã, Ana Luiza, porque sem ela as coisas não teriam tanta graça.

A toda a minha família (Meinho, Mamy, Tio Clovinho, Dedé, Paulinho, Clovizinho e Dada), sempre presente, torcendo por mim, onde quer que estejam, a todo momento.

A Renato Vieira, pelo apoio e entusiasmo diante de minhas vitórias e por tornar muito mais doces os meus dias mais difíceis.

À Professora Ana Maria Benko-Iseppon, pela confiança em mim depositada e por conseguir mostrar que podemos ser sempre melhores.

À Nina, amiga e parceira de laboratório, pelas conversas, projetos, milhares de favores e por anos de muitas risadas.

A todos os membros do Laboratório de Genética e Biotecnologia Vegetal, por tornarem mais agradável o meu dia-a-dia.

Aos meus colegas de Mestrado, Ana Carla, Danielle, Ednaldo, Eliane, Gabriel, Helder, Marcela, Michelly, Nina, Patrícia, Rodrigo, Vanessa, Vladimir e Tamara, por dividirem comigo esta etapa tão importante.

A minhas amigas, Ilana, Dani, Claudinha, Carol, Andréa, Suzana, Blenda, Malu, Macau e Polly, pela linda amizade que conseguimos nutrir durante tantos anos.

Aos professores do Programa de Pós-Graduação em Genética.

Ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), pela bolsa concedida durante o desenvolvimento do projeto).

Muito Obrigada!

SUMÁRIO

LISTA DE ABREVIAÇÕES

avr	Avirulência
BLAST	*Basic Local Alignment Search Tool* Ferramenta de Busca por Alinhamento Local
BRACELPA	Associação Brasileira dos Fabricantes de Celulose e Papel
CC	*Coiled Coil* Cauda Espiralada
CENBIO	Centro Nacional de Referência em Biomassa
COOPERSUCAR	Cooperativa de Produtores de Cana, Açúcar e Álcool do Estado de São Paulo
DDBJ	*DNA Database of Japan* Banco de dados de DNA do Japão
DNA	*Desoxiribonucleic Acid* Ácido Desoxirribonucléico
EMBL	*European Molecular Biology Laboratory* Laboratório Europeu de Biologia Molecular
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
EST	*Expressed Sequence Tag* Etiqueta de Seqüência Expressa
FAO	*Food and Agriculture Organization* Organização de Comida e Agricultura
FAPESP	Fundação de Apoio à Pesquisa do Estado de São Paulo
FORESTs	Projeto EST do Eucalipto
HR	*Hipersensitive Response* Reação de Hipersensibilidade
INSD	*International Nucleotide Sequence Database* Banco de Dados Internacional de Seqüências de Nucleotídeos
IRRI	*International Rice Research Institute* Instituto Internacional de Pesquisa Sobre Arroz
LRR	*Leucine Rich Repeats*

	Repetições Ricas em Leucina
MEGA	*Molecular Evolutionary Genetics Analysis* Programa para Análise Molecular e Evolutiva
NBS	*Nucleotide Binding Site* Sítio de Ligação de Nucleotídeo
NCBI	*National Center for Biotechnology Information* Centro Nacional para Informação Biotecnológica
ONSA	*Organization for Nucleotide Sequencing and Analysis* Organização para Sequenciamento e Análise de Nucleotídeos
ORF	*Open Reading Frame* Quadro de Leitura Aberta
pv.	Patovar
R	Resistência
ser	Serina
SOMs	*Self-Organizing Maps* Mapas de Auto-Organização
SUCEST	*Sugarcane EST Project* Projeto EST da Cana-de-açúcar
TIR	*Toll Interleucine Receptor* Receptor Toll-Interleucina
TM	*Transmembrane Domain* Domínio Transmembrana
tre	treonina

LISTA DE FIGURAS

Revisão Bibliográfica

Artigo científico publicado na revista *Genetics and Molecular Research*

Anexo

Artigo publicado na revista *Genetics and Molecular Biology*

Capítulo publicado no livro "Estresses ambientais: danos e/ou benefícios em plantas".

Apêndice

LISTA DE TABELAS

Artigo Científico publicado na revista *Genetics and Molecular Research*

Anexo

Artigo Científico publicado na revista *Genetics and Molecular Biology*

Apêndice

RESUMO

Os genes de resistência (*R*) respondem pela primeira interação entre planta e patógeno, sendo responsáveis pela ativação ou não de mecanismos de resistência em plantas. Este trabalho analisou genes *R* em sequências expressas de eucalipto, cana-de-açúcar e feijão-caupi, geradas através de bibliotecas produzidas a partir de diferentes tecidos em várias fases de desenvolvimento. Depois da análise *in silico* foi possível a identificação de todas as classes de genes de resistência em eucalipto, com destaque para a classe NBS-LRR (*Nucleotide Binding Site;* Sítio de Ligação de Nucleotídeo - *Leucine Rich Repeats;* Repetições Ricas em Leucina) (50% das 208 sequências candidatas que apresentaram domínios completos) e em cana-de-açúcar, com destaque para a classe KINASE (46% das 196 sequências candidatas que apresentaram domínios completos). No feijão-caupi o número de seqüências disponíveis foi escasso, observando-se maior abundância da classe NBS-LRR (80% das 38 sequências candidatas), entretanto estiveram ausentes as classes KINASE e LRR-KINASE. Observaram-se genes *R* em cana e eucalipto em todos os tecidos analisados, em diferentes níveis de expressão sob condições não induzidas. Quando analisados através de alinhamentos múltiplos os genes *R* apresentaram maior semelhança entre espécies pertencentes à mesma família, geralmente agrupando mono e dicotiledôneas em clados distintos, sugerindo que tenham surgido antes da separação entre essas classes. Os resultados do presente estudo têm potencial para colaborar com o desenvolvimento de marcadores moleculares para o melhoramento, para o entendimento da abundância e diversidade e evolução destes genes, com ênfase das espécies estudadas, bem como para identificação dos genes *R* em outras culturas de interesse econômico.

Palavras-chave: 1.Fitopatologia 2.Genes de resistência - cana-de-açúcar 3. Genes de resistência- eucalipto 4. Genes de resistência - Feijão-caupi

ABSTRACT

Resistance genes (*R*) are responsible for the first plant-pathogen interaction, being responsible for activation of resistance pathways is plants. The present work analyzes R genes in sugarcane, eucalyptus and *Vigna unguiculata* expressed sequence tags, generated by libraries produced from different tissues and development stages. After in silico screening was possible to identify all *R*- genes classes in sugarcane, specially KINASE (that presented 46% of 196 candidate sequences with conserved domains), in eucalyptus the most abundant was NBS-LRR (Nucleotide Binding Site-Leucine Rich Repeats) that presented 50% of 208 candidates sequences with conserved domains. Despite the available Vigna unguiculata sequence numbers was scarce, the most abundant class was NBS-LRR (80% of the 38 candidate sequences), however were absent KINASE and LRR-KINASE lasses. Could be observed R- genes in all analyzed sugarcane and eucalyptus tissues, in different expression levels under non induced conditions. When analyzed trough multiple alignments, R-genes presented more similarities among species that pertain to the same family, generally, grouping monocots and dycots in different branches, suggesting their appearing before the separation between these classes. Our result have potential to collaborate to molecular markers developing, with best understanding of these genes diversity and evolution, not only in these studied species, but already, in other economic interests crops.

Key-words: 1. Fitopathology 2. Resistance genes- Eucalyptus 3. Resistance genes- Sugarcane 4. Resistance genes- *Vigna unguiculata*

1.0 - INTRODUÇÃO

Em vegetais, genes de resistência (*R*) estão entre os mais importantes, sendo responsáveis pela defesa contra o ataque de patógenos. De acordo com a natureza de seus produtos, os genes *R* podem ser classificados em cinco diferentes grupos. Sugere-se que essa estrutura conservada seja determinante da especificidade patógeno-hospedeiro, ou seja, se a planta será ou não resistente a um determinado vírus, bactéria, fungo ou nematóide. Em vista do exposto, a identificação destes genes de resistência é de grande relevância para o melhoramento genético de plantas, constantemente expostas ao ataque dos mais diversos patógenos.

O eucalipto cultivado (*Eucalyptus grandis* W. Hill ex Maiden, *E. urophyla* St. Blake, seus híbridos e outras espécies relacionadas) apresenta excelente potencial para a produção de madeira em função de sua diversidade, adaptabilidade e alta produtividade, características que permitem inúmeras utilizações como matéria-prima, como por exemplo, na produção de papel e celulose. Por outro lado, a cana-de-açúcar (*Saccharum officinarum)* está entre as principais culturas vegetais, sendo cultivada em mais de oitenta países tropicais. O Brasil é responsável por aproximadamente 25% de toda a produção mundial, sendo o estado de Pernambuco um dos maiores produtores do país, respondendo a cana-de-açúcar por 40% de sua economia (CENBIO, 2006). Adicionalmente, o feijão-caupi (*Vigna unguiculata* (L.) Walp) é a planta que apresenta mais vantagens nutricionais dentre as tradicionalmente cultivadas e usadas pela população nordestina. 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.

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

2.0 - REVISÃO DA LITERATURA

2.1 – 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, freqüentemente associados às visíveis mudanças que podem ocorrer na morfologia do vegetal (IRRI, 2005).

Conforme levantamentos da FAO (2005) fungos, bactérias, vírus e nematóides 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.

Os produtores freqüentemente gastam altas somas para o controle parcial dos diversos patógenos que atacam suas culturas. Apesar disso, os danos na produção e nas mercadorias resultam anualmente em prejuízos de bilhões de dólares. Apenas nos Estados Unidos, as perdas alcançam nove bilhões de dólares anualmente (Oerke *et al.*, 1994 *apud* Fermin-Muñoz *et al.*, 2000), enquanto em caráter mundial, as doenças em plantas reduzem a produtividade em aproximadamente 12% ao ano (FAO, 2005).

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). A resistência é provavelmente a estratégia de controle mais desejada pelos melhoristas, pois pode ser altamente eficiente, sendo benigna ao meio ambiente e representando pouca ou nenhuma despesa adicional aos produtores (Sagar *et al.*, 2000).

2.2- Resistência contra patógenos

2.2.1- Mecanismos de defesa

As plantas utilizam uma grande variedade de estratégias para se defenderem contra estresses bióticos e abióticos. A primeira delas compreende a formação de uma barreira entre a célula e o patógeno, através do fortalecimento da parede celular com a produção de enzimas ligadas a biossíntese de lignina, bem como a formação de calos após o ataque de um microorganismo (Richter e Ronald, 2000). Segundo Heath (1991), as plantas apresentam resistência à maioria dos microorganismos potencialmente patogênicos, a chamada Resistência Básica (Heath, 1987) e sugere 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". Entretanto, considera-se que a reação de hipersensibilidade (HR; *Hipersensitive Response*) se constitui no principal mecanismo utilizado pelas plantas para se defenderem contra o ataque de patógenos (Bonas e Lahaye, 2002); ela é desencadeada pela "ativação" de um gene de resistência (*R*) e caracterizada por morte celular rápida e localizada impedindo que o patógeno se espalhe e, desta forma, colonize o organismo (Meyers, 2005).

2.2.2- A interação gene-a gene

Flor, em 1942, trabalhando com o patógeno *Melampsora lini* e 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 para cada gene de resistência (*R*) presente no hospedeiro, exista um gene de avirulência (*avr)* no patógeno determinando patogenicidade; 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 (Bonas & Anckerveken, 1999) (Figura 1). Nos primeiros 15 minutos se 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 (Dangl e Jones, 2001). O gene de resistência 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 (Ohtsuki e Sasaki, 2006).

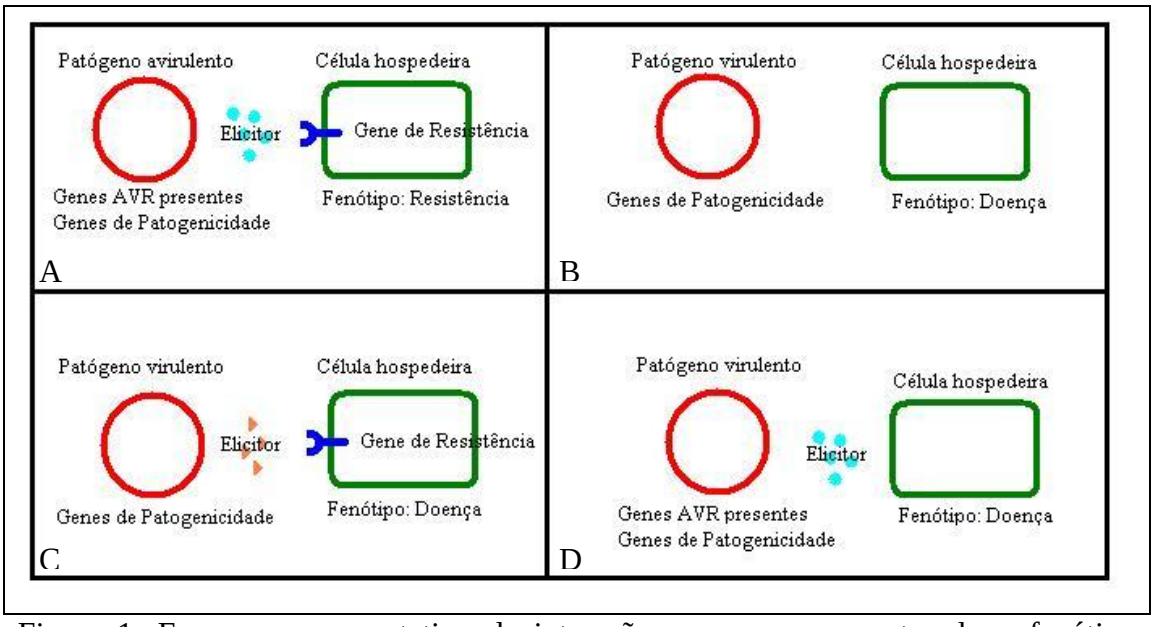


Figura 1: Esquema representativo da interação gene-a-gene, mostrando o fenótipo resistente (A) e os suscetíveis (B-D).

2.2.3- Os genes de resistência e suas classes

Em contraste com a grande diversidade dos produtos dos genes de avirulência (Ellis e Jones, 2000) os genes *R* são extremamente conservados e podem ser agrupados

em cinco classes distintas de acordo com a natureza de seus produtos (Hammond-Kosak e Jones, 1997; Morais, 2003) (Figura2).

A primeira classe, chamada classe das Redutases, é representada pelo gene *HM1* do milho. Este gene codifica uma redutase que inativa as toxinas produzidas pelo fungo *Helminthosporium carbonum* (Joahal e Briggs, 1992) e pelo gene *Mlo* de cevada, um provável regulador de defesa contra *Blumenaria graminis* (Piffanelli *et al.*, 2002); também representa esta classe o gene *RPW8* de *Arabidopsis*, 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*, que confere resistência à bactéria *Pseudomonas syringae* pv. *tomato*, causadora da mancha foliar do tomate (Song *et al.*, 1997; Melotto *et al.*, 2004; Anderson *et al.*, 2006). Esta classe caracteriza-se pela presença de uma quinase que fosforila resíduos de serina (*ser*) e treonina (*tre*), capaz de interagir com o produto do gene *avrPto* (Tang *et al.*, 1999); este gene também já foi caracterizado em outros organismos, como *Arabidopsis thaliana* e *Phaseolus vulgaris* (Melotto *et al.*, 2004).

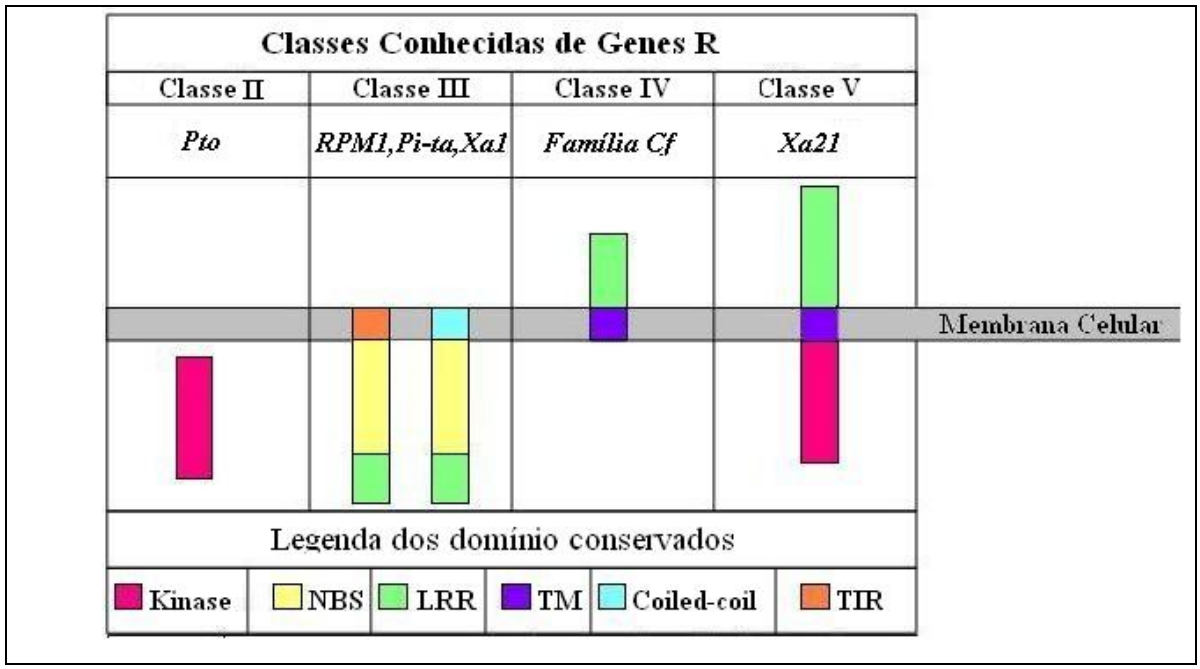


Figura 2: Representação gráfica das quatro famílias de Genes de Resistência que apresentam domínios conservados em suas estruturas. São elas: II: Classe Quinase; III: Classe NBS-LRR; IV: Classe LRR; V: Classe LRR-Quinase. As diferentes cores representam os domínios conservados. NBS: Sítio de Ligação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

A terceira e maior classe é representada por genes que sintetizam proteínas que contêm domínios chamados NBS (*Nucleotide Binding Site*; Sítio de Ligação de Nucleotídeo) e LRR (*Leucine Rich Repeats;* Repetições Ricas em Leucina) e se divide em duas sub-classes (Li *et al.*, 2001; Liang *et al.*, 2005; van Leeuven *et al.*, 2005).

O domínio LRR está envolvido no reconhecimento do patógeno pelo hospedeiro durante o processo de infecção. Proteínas que contêm LRRs apresentam séries do aminoácido leucina, repetidas em número de 24 ou mais a intervalos regulares, podendo conter ainda outros resíduos hidrofóbicos, asparaginas e prolinas (Collins *et al*, 2001). Os motivos de leucina geram uma estrutura terciária semelhante a uma mola, com cada hélice apresentando um motivo de leucina repetido. Afunção das LRRs parece ser a de mediação da interação entre proteínas, podendo atuar estracelularmente como receptor da molécula produzida pelo patógeno ou intracelularmente, em um passo anterior à via de transdução dos 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 interaçãoproteína-carboidrato, relacionados à especificidade da resistência (Kobe e Kajava, 2001). Supões-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 sítios de ligação de nucleotídeos (NBS), também chamados de P-loops, 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. A presença desses domínios os genes de resistência sugere que estes sítios estão envolvidos na ativação de proteínas quinases ou proteínas transportadoras (Hammon-Kosack e Jones, 1997) e mutações em sua estrutura impedem que ocorra a resposta hipersensível, ressaltando a importância da função deste domínio na sinalização (Zhang *et al.*, 2004).

A primeira sub-classe possui o domínio CC (*Coiled-coil;* Cauda Espiralada*)* 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 (Mindrinos *et al.*, 1994; Whitham *et al.*, 1994; Grant *et al.*, 1995; Lawrence *et al.*, 1995; Bent, 1996; Salmeron *et al.*, 1996; Ori *et al.*, 1997; McDowell *et al.*, 1998; Milligan *et al.*, 1998; Yoshimura *et al.*, 1998; Wang *et al.*, 1999; Bittner-Eddy *et al.*, 2000; Bryan *et al.*, 2000; Brommonschenkel *et al.*, 2000; Ernst *et al.*, 2002; Rehmany *et al.*, 2006). A segunda sub-classe possui em sua região amino-terminal ao invés de uma sequência CC, o domínio TIR (*Toll Interleucine Receptor;* Receptor Toll/Interleucina) que também está presente em animais, mas acredita-se que seja inexistente em monocotiledôneas, tendo sido encontrado em todas as espécies de dicotiledôneas atualmente estudadas (Goff *et al.*, 2002); essa sub-classe têm 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* (Whithan *et al.*, 1996 e Mestre e Baulcombe, 2006) de tabaco.

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* (Jones *et al*., 1994; Dixon *et al*., 1996; Kruijt *et al.*, 2005). Esta classe codifica genes que possuem uma proteína ancorada à membrana (TM – *Transmembrane Domain;* Domínio Transmembrana) 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).

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

2.2.4- A evolução dos genes de resistência

Genes de resistência, em muitas espécies cultivadas, 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 resulta de duplicações de sequências em tandem que ocorrem em alguns *loci* gênicos devido à ocorrência de *crossing over* desigual (Meyers *et al.*, 2005).

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 que possuem LRRs (Bergelson *et al,* 2001); 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 ns 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). 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.

Polimorfismos em alelos de genes de resistência podem ser determinantes para a 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* revelaram divergência de seqüência 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)

2.3- A Cultura da cana-de-açúcar

2.3.1- Origem, história e citogenética

A cana-de-açúcar é uma planta perene e alógama (Berding e Roach, 1987), classificada taxonomicamente como membro da divisão *Embriophyta*, incluída na subdivisão *Angiospermae*, classe *Monocotyledoneae*, família *Poaceae* (gramíneas), tribo *Andropogoneae* e gênero *Saccharum* (Quirino, 2003).

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).

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 à 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 alopoliplóide segmental (Roach e Daniels, 1987). Em cana-de-açúcar não se pode assumir herança diplóide, entretanto, segundo Hogarth (1987), evidências sugerem herança semelhante aos diplóides. 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).

2.3.2- As doenças da cana-de-açúcar

No Brasil foram relatadas 40 entre todas as 177 doenças relacionadas em cana-de-açúcar em todo o mundo. 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 (Orsdogoitti *et al*., 1982). 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 em 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 carvão é uma das doenças em cana-de-açúcar de mais fácil identificação. Caracteriza-se pela emergência de um chicote, que consiste em uma modificação do meristema apical do colmo, induzida pelo fungo, com tamanho variável, de alguns centímetros a mais de 1 m de comprimento. O chicote é composto por parte do tecido da

planta e parte do tecido do fungo contendo milhões de esporos. Plantas doentes, antes de emitirem o chicote, têm o ângulo de inserção das folhas mais agudo, limbo foliar estreito e curto, colmos mais finos que o normal e touceiras com superbrotamento. Ocasionalmente, algumas variedades podem produzir sintomas atípicos como galhas, proliferação de gemas e vassoura-de-bruxa (Tokeshi, 1997). Os chicotes surgem em plantas com 2-4 meses de idade, com o pico ocorrendo quando as plantas estão com 6-7 meses de idade. Condições ambientais são determinantes no surgimento de epidemias de carvão. Sob condições de estresse, variedades mesmo com resistência ao fungo, podem apresentar sintomas. Condições de estresse hídrico e calor favorecem a ocorrência da doença.

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 5 a 30% da produtividade, mesmo que o produtor desconheça que seu campo esteja infectado. A importância econômica da doença varia em função de fatores ambientais, principalmente estresse hídrico. As perdas podem ser substanciais, dependendo da variedade empregada e da idade da touceira (Smith, 1999). O raquitismo das soqueiras não é facilmente reconhecido no campo por sintomas externos, uma vez que podem ser confundidos por vários fatores, incluindo práticas culturais, umidade inadequada e deficiência nutricional (Comstock e Lentini, 2002). Para diagnóstico eficaz da doença, várias metodologias têm sido desenvolvidas, como a microscopia de contraste de fase, reações sorológicas de fluorescências, testes sorológicos e técnicas de biologia molecular pelo uso de PCR.

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 á confusão de seus prejuízos com aqueles causados pelo raquitismo das soqueiras (Tokeshi, 1997). Quando a doença se manifesta em variedades extremamente suscetíveis, pode causar perdas de até 100%. Seus prejuízos manifestam-se, principalmente, na má formação dos toletes no plantio, morte das touceiras ou cana, queda na produção e na riqueza de sacarose. A principal forma de controle dessa doença é por meio de variedades resistentes e tolerantes ao patógeno. Entretanto, devido aos sintomas latentes da escaldadura, deve se ter atenção com as variedades, principalmente aquelas tolerantes. No caso de emprego de variedades tolerantes, torna-se necessário

tomar alguns cuidados: evitar plantio de mudas provenientes de campos com a doença; preparo das áreas de viveiros para eliminar bactérias do solo e restos de cultura; desinfecção de equipamentos e ferramentas utilizadas no manejo da cultura.

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). Os sintomas iniciais de infecção consistem em pontos cloróticos com disposição linear no meio ou mais comumente na base das folhas, que evoluem para áreas alongadas formando um mosaico típico, o qual pode aumentar de severidade com a idade da folha. O crescimento das plantas pode ser acentuadamente reduzido conforme a espécie e estirpe do vírus e a variedade de cana, principalmente quando a infecção ocorre nos estágios iniciais de desenvolvimento. Ocasionalmente, em variedades altamente suscetíveis podem ocorrer riscas e estrias nos colmos e encurtamento dos entrenós (Gonçalves *et al*., 2004). Atualmente, a doença está controlada, porém freqüentemente observa-se plantas sintomáticas durante a avaliação de clones para melhoramento e em plantios comerciais. Apesar desse fato, o mosaico não tem ocasionado perdas de grandes proporções, devido à seleção de variedades resistentes e tolerantes associada à prática do “roguing” nos plantios comerciais. A taxa de distribuição do mosaico no campo é influenciada principalmente pelos fatores: nível de tolerância da variedade de cana-de-açúcar; estirpe ou estirpes do vírus presentes; número e distribuição dos focos de infecção; número, tipo e atividade de insetos vetores presentes; condições climáticas afetando a suscetibilidade da cultura e a atividade dos insetos vetores. No entanto, esse vírus incide sobre outras gramíneas, dentre as quais o milho, *Zea mays* (Costa *et al*. 1971) e o sorgo, *Sorghum bicolor* (Pinto, 1984). A introdução de novos cultivares de milho no país nos últimos anos, associada ao grande aumento da área plantada e ao cultivo na safrinha, tem aumentado a incidência do mosaico nessa cultura e em campos de cana-de-açúcar próximos a lavouras de milho (Fernandes e Oliveira, 1997).

2.3.3- 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. O Brasil é responsável por 25% das $1,2x10^9$ toneladas de cana-de-açúcar cultivadas anualmente (COPERSUCAR 2006), produção que ocupa cerca de 18 milhões de hectares do globo terrestre (SUCEST, 2006).

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 um milhão de empregos diretos e indiretos em todo o Brasil (COOPERSUCAR, 2006).

Atualmente Pernambuco é o segundo maior estado produtor nacional de cana-de-açúcar no Brasil. Esta é plantada na zona da mata Pernambucana, 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 (CENBIO, 2006).

2.3.4- O Projeto SUCEST

Recentemente, vários projetos com o objetivo de sequenciar 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 feitas de diferentes órgãos e tecidos em diferentes estágios de desenvolvimento. A abordagem escolhida por estes projetos se baseia na determinação da seqüência do RNA mensageiro através da análise de seqüências expressas do genoma. Tal abordagem permite 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, 2006).

2.4- A cultura do eucalipto

2.4.1- Origem, citogenética e história

O eucalipto possui como centro de origem a Austrália e regiões próximas como Timor, Indonésia, Molucas, Irian Jaya, Papua Nova Guiné e sul das Filipinas, em uma faixa compreendida entre latitudes 9ºN e 44ºS (Eldridge *et al.*, 1993).

O gênero *Eucalyptus* é considerado um dos mais importantes de toda a família Myrtaceae. Citogeneticamente, esta família apresenta uma variação dos números cromossômicos diplóides (2n =10 a 2n = 77). Porém, o número cromossômico mais comumente representado é 2n = 22 (Goldblatt, 1981; 1984; 1985; 1988; Goldblatt e Johnson, 1990; 1991; 1994; 1996).

Existem cerca de 600 espécies diferentes de eucalipto; compreendendo árvores de grande porte que podem atingir até 80 metros em indivíduos adultos. Trata-se da planta lenhosa mais abundante no globo, ocupando mais de 18 milhões de hectares (Brondani *et al.*, 2006). Enquanto o *E. globulus* é a principal espécie encontrada nas zonas temperadas tais como Portugal, Espanha, Chile e Austrália, os clones híbridos *E. grandis* e *E. urophylla* são bastante utilizados nas indústrias em países de regiões tropicais como Brasil, África do Sul, Índia e Congo devido a sua madeira de qualidade e a seu crescimento rápido, compreendendo por vezes apenas sete anos para a produção de celulose (Eldridge *et al.*, 1993).

O eucalipto foi introduzido no Brasil em 1904 por Edmundo Navarro de Andrade, com o objetivo de suprir as necessidades de lenha, postes e dormentes das estradas de ferro na região Sudeste. Na década de 50 passou a ser produzido, como matéria prima, para o abastecimento das fábricas de papel e celulose sendo sua expansão ampliada nos anos 60. Houve um crescimento da produção até meados dos anos 80, período considerado um marco na silvicultura brasileira, dado os efeitos positivos que gerou no setor (EMBRAPA, 2006), compreendendo hoje cerca de 70% (3,5 milhões de hectares) de toda a área reflorestada no Brasil (Bacha e Barros, 2004).

2.4.2- As doenças do Eucalipto

O eucalipto é atacado por vários patógenos, principalmente fungos, desde a fase de viveiro até os plantios adultos. Geralmente, os problemas são observados nas

plantações, ocorrendo nos mais variados locais, espécies e épocas do ano (Bacha e Barros, 2004). As principais doenças são a ferrugem, o cancro, o oídio e o tombamento, também chamado de “damping off” (Firme et al., 2000).

A ferrugem, causada pelo fungo *Puccinia psidii* é uma doença que atualmente está causando sérios problemas em plantios jovens, viveiros e jardins clonais de *Eucalyptus*, sendo que fotoperíodo, temperatura e umidade são fatores condicionantes para a ocorrência da doença. A maior importância econômica da ferrugem está relacionada a plantios de campo, enquanto em viveiros é mais facilmente controlada com o uso de fungicidas. No campo, o uso de fungicidas para o controle de *Puccinia psidii* não é economicamente viável. A melhor forma de controle é a seleção de materiais genéticos resistentes. Em viveiros e jardins clonais, o controle de ataques intensos utilizando fungicidas é eficiente, sendo recomendado o uso de mancozeb, oxicloreto de cobre, triadimenol, diniconazole ou triforine (Silveira *et al.*, 1998).

O cancro do eucalipto é uma das doenças mais importantes de ocorrência no campo, causado por várias espécies de fungos como *Cryphonectria cubensis*, *Valsa ceratosperma* - fase sexuada, *Cytospora spp.* - fase assexuada e *Botryosphaeria ribis*). O cancro de *Cryphonectria cubensis* foi considerado como a principal doença que afetou a cultura do eucalipto no Brasil na década de 70. Trata-se de uma doença de ampla distribuição geográfica, ocorrendo em regiões tropicais do continente americano (KRUGNER, 1980). Essa doença é caracterizada pela morte dos tecidos da casca, decorrente da ação de vários agentes abióticos e bióticos. Contudo, as condições climáticas parecem ter uma participação maior na manifestação deste tipo de problema, com ligação entre a incidência de certos cancros e as condições adversas ao desenvolvimento da planta. O controle mais recomendado para o cancro seria a utilização de espécies, procedências, progênies ou clones mais resistentes a este patógeno (Silveira, 2000).

O oídio é causado pelo patógeno *Oidium* sp. Esse fungo ataca várias espécies de eucalipto em condições de viveiro, casa de vegetação e campo. A espécie mais suscetível a essa doença é o *Eucalyptus citriodora*. Os sintomas aparecem principalmente em gemas e brotações, causando deformidade ou morte das mesmas. Esses sintomas são caracterizados pelo recobrimento das partes afetadas por estruturas de coloração esbranquiçada, pulvurulenta, constituídas por micélios e estruturas de reprodução do patógeno, típico dos oídios. Em mudas, o ataque sucessivo causa brotamento, resultando em mudas de baixa qualidade. No campo, a ocorrência de oídio

causa perda da dominância apical, afetando a formação de um fuste reto. Para o controle da doença em viveiro pode-se fazer aplicação de benomyl mais enxofre molhável. No campo, a doença tende a desaparecer com o desenvolvimento da planta, através da troca da folhagem juvenil pela adulta (Araújo, 2000).

O tombamento é causado pelos fungos *Cylindrocladium candelabrum, C. clavatum, Rhizoctonia solani, Pythium* spp., *Phytophthora* spp. e *Fusarium* spp. Esses fungos habitam o solo, onde vivem como saprófitas ou na forma de estruturas de repouso, os escleródios, microescleródios, clamidósporos e oósporos, dependendo de cada espécie. Os propágulos desses fungos são disseminados através da água da chuva ou irrigação, vento ou partículas de solo aderidas a suplementos agrícolas, sendo que em ambientes com alta umidade favorecem a ocorrência de tombamento. O ataque compromete as sementes em germinação, afetando os tecidos tenros. Ocorre inicialmente no colo da plântula, podendo se estender ao hipocótilo, com aspecto inicial de encharcamento evoluindo para uma coloração escura, com posterior tombamento e morte da muda. Dependendo da idade da muda pode ocorrer murcha, enrolamento e seca dos cotilédones e das primeiras folhas, porém, esses sintomas são considerados secundários. Por ser uma doença causada por patógenos do solo, o risco de ataque desses fungos aumenta para o caso de semeaduras feitas em sementeiras para posterior repicagem. Esse problema pode ser evitado com o uso da semeadura direta em tubetes suspensos. Porém, cabe ressaltar que a água de irrigação e o substrato devem estar livres de inóculos dos patógenos. O uso de brita como material de cobertura do solo do viveiro evita a contaminação. Ainda em relação ao substrato, este deve apresentar boa drenagem (Firme et al., 2000).

2.4.3- Produção e Importância Econômica

As espécies de eucalipto apresentam excelente potencial para a produção de madeira em função de sua diversidade, adaptabilidade, alta produtividade e características físico-mecânicas, que permitem inúmeras utilizações como matéria-prima para produção de papel e celulose, marcenaria (serraria e madeira processada), caixotaria, lenha e carvão vegetal (Gonzáles, 2002), algumas espécies também produzem óleos essenciais com os quais são fabricados produtos de limpeza, alimentícios, perfumes e remédios. A principal espécie plantada atualmente no Brasil tem sido *E. grandis* (e seus híbridos), com cerca de 1.500.000 hectares de área plantada.

A principal produção ocorre nos estados de Minas Gerais, São Paulo, Bahia e Rio Grande do Sul (Sociedade Brasileira de Silvicultura – SBS, 2004), havendo limitações quanto à sua introdução em áreas mais quentes, úmidas ou semi-áridas, devido ao ataque de patógenos e às condições edafo-climáticas, com ênfase para solos pobres e salinos (Mafia e Alfenas, 2003).

As principais razões do uso extensivo do eucalipto em reflorestamentos comerciais em várias partes do mundo devem-se ao seu rápido crescimento (média de 45 m^3/ha/ano em plantios clonais), quando comparado a outras plantas lenhosas, além de suas excelentes características propícias para a produção de fibras e polpa de madeira de alta qualidade. Além disso, destacam-se também o baixo custo e o curto período de corte (sete anos), fatores que permitem até três rotações sucessivas e econômicas, com ciclo de até 21 anos (Ho *et al.*, 1998).

A produção de papel e celulose compreende componente importante do setor industrial brasileiro, com perspectivas de expansão para os próximos anos. O Brasil apresentou forte crescimento desse setor, tornando-se o principal produtor mundial de celulose a partir do eucalipto como matéria-prima. No ano de 2001 foram produzidas 7,4 milhões de toneladas de papel, com faturamento total de US$ 6,7 bilhões (BRACELPA, 2005). Desse total, cerca de US$ 2,6 bilhões foram obtidos com exportação. No segmento de papel os investimentos vêm sendo realizados no desenvolvimento tecnológico de processos e produtos de maior valor agregado (BRACELPA, 2005).

2.4.4- O Projeto FORESTs

Em novembro de 2001, a FAPESP anunciou o início do projeto FORESTS que propiciou o seqüenciamento de parte do genoma expresso do eucalipto. O citado projeto foi desenvolvido no âmbito do Programa Parceria para Inovação Tecnológica, PITE, contando com a participação de um consórcio de empresas de papel e celulose, tendo como objetivo identificar genes que auxiliem para aumento da produtividade e da qualidade da matéria-prima utilizada na produção de papel e celulose (FAPESP, 2006).

A proposta foi considerada inovadora por envolver as principais indústrias da área num sistema de consórcio pré-competitivo em parceria com as universidades paulistas e a Fapesp. Adicionalmente, o objetivo final não foi apenas a identificação de genes do eucalipto, mas o uso dessa informação na análise da expressão gênica, visando

ao desenvolvimento de uma nova tecnologia para a identificação de genes envolvidos no controle genético da qualidade da madeira, resistência a doenças e pragas, estresses ambientais e tolerância às deficiências nutricionais (FOREST, 2005).

O Projeto Genoma Eucalipto objetivou identificar 15.000 genes através do seqüenciamento de aproximadamente 100.000 ESTs derivadas de bibliotecas de cDNA geradas a partir de diferentes tecidos, estágios de desenvolvimento e condições fisiológicas. Até março de 2005 um total de 17.286 singlets e 15.794 clusters, oriundos de 123.889 reads, haviam sido disponibilizados no banco de dados do FORESTs, ultrapassando o número de sequências inicialmente planejadas pelo projeto (FORESTS, 2005).

2.5- A Cultura do Feijão-Caupi

2.5.1- Origem, Citogenética e História

O feijão-caupi é uma cultura autógama (Teófilo *et al.*, 2001), classificada como membro da divisão *Embriophyta,* incluída na subdivisão *Angiospermae* e na classe *Dicotyledoneae* (NCBI, 2006). Trata-se de uma leguminosa (Fabaceae) do gênero *Vigna* (NCBI, 2006), possuindo um dos menores genomas deste grupo (450-500 Mb), apresentando o nível diplóide 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 (Freire-Filho, 1988).

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 recentes (Simon *et al.*, no prelo) 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.

Vários cruzamentos têm sido efetuados com o intuito de produzir caracteres desejáveis como aumento do tamanho dos grãos, aumento da produtividade média, porte ereto das plantas, floração precoce, bem como para a identificação de linhagens resistentes à salinidade e às doenças que mais prejudicam a produção do feijão-caupi (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).

2.5.2 – As doenças do feijão caupi

Apesar do feijão-de-corda ser uma planta rústica, ela é hospedeira de muitas doenças causadas por vírus, bactérias, fungos e nematóides (Apesar do feijão-de-corda ser uma planta rústica, ela é hospedeira de muitas doenças causadas por vírus, bactérias, fungos e nematóides (Araújo *et al*., 1984). 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% (Araújo *et al*., 1984).

O mosaico severo do caupi (CPSMV) encontra-se disseminado e apresenta difícil controle devido à grande variedade de hospedeiros e à numerosa população de vetores (Freire-Filho *et al*., 1999). No que diz respeito ao vírus do mosaico dourado do

caupi (CPGMV) a situação também é preocupante pois estudos de avaliação do impacto destas doenças em algumas regiões do nordeste do Brasil contabilizam perdas em torno de 70% da produção (Barreto, 1999). Além das viroses, outras doenças são consideradas importantes, e podem também gerar perdas significativas dependendo da cultivar utilizada. Infelizmente, há poucas fontes de resistência para várias das doenças existentes, estando essas freqüentemente em diferentes cultivares, vários pouco produtivos ou que não atendem às necessidades do mercado (Barreto, 1999).

2.5.3- Produção e Importância Econômica

Conhecido também como feijão de corda, o feijão-caupi é uma cultura de importante destaque na economia nordestina e de amplo significado social, constituindo o principal alimento protéico e energético do homem rural; 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, 2006). Apresenta alta rusticidade e adaptabilidade às condições de estiagem prolongadas e 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 e Carvalho, 1988). 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).

2.6- A Bioinformática

2.6.1- História e Aplicações

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 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 seqüências 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. Desta forma, várias universidades, instituições governamentais e firmas farmacêuticas vêm formando grupos de bioinformática. Estes grupos representam importante papel na análise das informações geradas em grande escala pelos seqüenciadores de todo o mundo (Prosdocini *et al*., 2002). 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 (Carraro e Kitajima, 2002).

2.6.2- Bancos de Dados e Ferramentas

O Genbank é o mais conhecido banco de dados que disponibiliza gratuitamente seqüências e ferramentas para a sua análise. Foi criado inicialmente pelo NCBI com o objetivo de abrigar seqüências submetidas diretamente por autores de publicações científicas. Junto com o Genbank, fazem parte do INSD (*International Nucleotide Sequence Database*; Banco de Dados Internacional de Seqüências de Nucleotídeos) o DDBJ (*DNA Database of Japan*; Banco de Dados de DNA do Japão) (http://www.ddbj.nig.ac.jp/), e o EMBL (*European Molecular Biology Laboratory*; Laboratório Europeu de Biologia Molecular; 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 (Tateno *et al.*, 2002). Em conjunto com os bancos de dados, várias ferramentas e programas são utilizados no processo de identificação e análise de seqüências; entre eles estão o BLAST (*Basic Local Alignment Search Tool*; Ferramenta de Busca por Alinhamento Local) (Altschul *et al.*, 1990), o CLUSTALX (Thompson *et al.*, 1997), o TreeView (Page, 1996), o MEGA (*Molecular Evolutionary Genetics Analysis* - Programa para Análise Genética e Evolutiva; Sudhir *et al.*, 2004) e o CLUSTER (Eisen *et al.*, 1998).

Alinhamentos representam uma poderosa ferramenta não só para comparar seqüências desconhecidas com seqüências de genes já descritos e depositadas nos bancos de dados, como também para fazer inferências tanto estruturais e funcionais quanto evolutivas a respeito destas seqüências. 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 seqüência de nucleotídeo (DNA/RNA) para uma seqüência protéica em todos os quadros de leitura.

O CLUSTALX é um dos programas mais usados para obtenção de múltiplos alinhamentos de seqüências, quer sejam de nucleotídeos, ou de aminoácidos. Produz alinhamentos múltiplos biologicamente significativos entre seqüências divergentes; calculando qual a melhor seqüência para comparar com a seqüência selecionada, alinhando-as de acordo com as identidades, similaridades e diferenças (Thompson *et al.*, 1997); 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 CLUSTALX não possui uma interface para sua visualização.

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 seqüências, a porcentagem de similaridade entre seqüências ou ainda identificação de códons ou marcadores conservados (Sudhir *et al.*, 2004).

O CLUSTER é um programa para análise e visualização de resultados de complexos experimentos de *microarrays.* Realiza vários tipos de análises de seqüências, incluindo clusterização hierárquica e *Self-Organizing Maps* (SOMs). Os métodos de clusterização hierárquica 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).

3. REFERÊNCIAS BIBLIOGRÁFICAS

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

Anderson JC, Pascuzzi PE, Xiao F, Sessa G and Martin GB (2006) Host-mediated Phosphorylation of type III effector AvrPto promotes *Pseudomonas* Virulence and Avirulence in tomato. Plant Cell 18:502-514.

Araújo JPP de, Rios GP, Watt EE, Neves B., Fageria NK, Oliveira IP, Guimarães CM and Silveira-Filho A. (1984). Cultura do Caupi, *Vigna unguiculata* (L.) Walp; descrição e recomendações técnicas de cultivo. EMBRAPA, Circular Técnica pp. 18, 82.

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

Araújo EF (2000) Monitoramento nutricional em plantios de eucalipto na Bahia Sul Celulose – Análise Foliar. Relatório Técnico de Pesquisa da Bahia Sul.. (não publicado).

Artschwager E and Brandes EW (1958) Sugarcane (*Saccharum officinarum* L) Origin, clasification and characteristics of representative clones. In: Agriculture Handbook nº122. Department of Agriculture, Washington, pp 307-308.

Bacha CJC and Barros ALM (2004) Reflorestation in Brazil: recent evolution and the future. Sci Florest 66:191-203.

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. http://www.cpatsa.embrapa.br/livrorg/index.

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

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

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

Bittner-Eddy PD, Crute IR, Holub EB and Beynon JL (2000) *RPP13* is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. Plant J 21:177-88.

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

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.

Botella MA, Parker JE, Frost LN, Bittner-Eddy PD, Beynon JL, Daniels MJ, Holub EB and Jones JD (1998) Three genes of the *Arabidopsis* RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. Plant Cell 10:1847-1860.

Brandes EW (1956) Origin, dispersal and use in breeding of the Malanesian garden sugarcanes and their derivates *Saccharum officinarum* L. Proc Cong Int Soc Sug Technol 9:709-750.

Brett PGC (1957) The identification of more important Sugar Cane Varieties grown in Africa. Sug Assoc Bull 4:18-23.

Brommonschenkel SH, Frary A and Tanksley SD (2000) The broad-spectrum topovirus resistance gene *Sw-5* of tomato is a homolog of the root-knot nematode resistance gene Mi. Mol Plant Micr Interact 13:1130-1138.

Brondani RPV, Williams ER, Brondani C and Grattapaglia D (2006) A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. BMC Plant Biol 6:20-22.

Bryan GT, Wu KS, Farrall L, Jia Y, Hershey HP, McAdams SA, Faulk KN, Donaldson GK, Tarchini R and Valent B (2000) A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant Cell 12:2033-2046.

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

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

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

Comstock JC and Lentini RS (2002) Sugarcane leaf scald disease.Disponível em: http://edis.ufl.edu/body_SC008. Acesso em: 10 jul. 2007.

Comstock JC and Lentini RS (2002) Sugarcane ratoon stunting disease. Disponível em: http://edis.ufl.edu/body_SC008. Acesso em: 10 jul.2007

Costa AS, Kitajima EW and Arruda SC (1971) Moléstias de vírus e de micoplasmas do milho em São Paulo. *Revista da Sociedade Brasileira de Fitopatologia* 4:39-41.

Dantas B (1960) Contribuição para a história da "gomose" da cana-de-açúcar, em Pernambuco e no Brasil. In: Boletim Técnico no 11 Instituto agronômico do Nordeste. Recife, pp 3-17.

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.

Dodds P, Lawrence G and Ellis J (2001) Six amino acid changes confined to the leucine-rich repeat beta-strand/beta-turn motif determine the difference between the P and P2 rust resistance specificities in flax. Plant Cell 13:163-78.

Eisen MB, Spellman PT, Brown PO and Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95:14863-14868.

Eldridge K, Davidson J, Harwood C and van Wyk G (1993) Eucalypt domestication and breeding. Clarendon Press, Oxford, 288 pp.

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.

Ernst K, Kumar A, Kriseleit D, Kloos DU, Phillips MS and Ganal MW (2002) The broad-spectrum potato cyst nematode resistance gene (*Hero*) from tomato is the only member of a large gene family of NBS-LRR genes with an unusual amino acid repeat in the LRR region. Plant J 31:127-136.

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

Fernandes FT and Oliveira E (1997) Principais doenças na cultura do milho. Brasília: EMBRAPA (Circular Técnica n. 26).

Firme DJ, Higashi EM, Silveira RLVA, Leite FP and Gonçalves NA (2000) Efeito da pulverização de cloreto de cálcio no enraizamento das miniestacas de *Eucalyptus* spp. In: FERTIBIO 2000 - Biodinâmica do solo. REUNIÃO BRASILEIRA DE FERTILIDADE DO SOLO E NUTRIÇÃO DE PLANTAS, 25. Santa Maria. Resumos... Santa Maria: Sociedade Brasileira de Ciência do Solo/Sociedade Brasileira de Microbiologia (CD-Rom).

Flor HH (1942) Inheritance of pathogenicity in *Melampsora lini*. Phytopathol 32:653-669.

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, http://www.cpatsa.embrapa.br/livrorg/index.

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.

Goldblatt P (1981) Index to Plant Chromosome Numbers 1975-1978. 5 º Missouri Botanical Garden, Saint Louis, 553 pp.

Goldblatt P (1984) Index to Plant Chromosome Numbers 1979-1981. 8º Missouri Botanical Garden. Saint Louis, 427 pp.

Goldblatt P (1985) Index to Plant Chromosome Numbers 1982-1983. 13º Missouri Botanical Garden. Saint Louis, 224 pp.

Goldblatt P (1988) Index to Plant Chromosome Numbers 1984-1985. 23º Missouri Botanical Garden. Saint Louis, 264 pp.

Goldblatt P and Johnson, D.E. 1990. Index to Plant Chromosome Numbers 1986-1987. 30º Missouri Botanical Garden. Saint Louis, 243 pp.

Goldblatt P and Johnson DE (1991) Index to Plant Chromosome Numbers 1988-1989. 40º Missouri Botanical Garden. Saint Louis, 237 pp.

Goldblatt P and Johnson DE (1994) Index to Plant Chromosome Numbers 1990-1991. 51º Missouri Botanical Garden. Saint Louis, 267 pp.

Goldblatt P and Johnson DE (1996) Index to Plant Chromosome Numbers 1992-1993. 58º Missouri Botanical Garden. Saint Louis, 276 pp.

Gonçalves MC, Moreira YJCB, Maia IG, Santos AS, Fantin GM and Chaves A (2004) Identificação e caracterização de isolados pertencentes ao subgrupo do *Sugarcane mosaic virus* no Estado de São Paulo. Fitopatologia Brasileira 29: 129-129.

Gonzáles E R (2002) Transformação genética de *Eucalyptus grandis* e do híbrido *E. grandis* X *E. urophylla* via *Agrobacterium*. Tese de Doutorado, ESALQ, Piracicaba.

Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW and Dangl JL (1995) Structure of the *Arabidopsis* RPM1 gene enabling dural specificity disease resistance. Science 269:843-846.

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

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D,Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A and Briggs S (2002) A draft sequence of the rice genome (*Oryza sativa* L. spp *japonica*). Science 296:92-100.

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

Heath M (1987) Evolution of resistance and susceptibility to fungal invaders. Can J Plant Pathol 9:389-397.

Heath M (1991) The role of gene-for-gene interaction in the determination of host species specificity. Phytopathology 81:127-130.

Heinz DJ, Mee GWP and Nickel LG (1969) Chromosome number of some *Saccharum* species, hybrids and their cell suspension cultures. Am J Bot 56:450-454.

Ho CK, Chang SH, Tsay JY, Tsay CJ, Chiang VL, Chen ZZ (1998) *Agrobacterium tumefaciens* mediated transformation of *Eucalyptus camaldulensis* and production of transgenic plants. Plant Cell Rep.17:675-680.

Hogarth DM (1987) Genetics of sugarcane In: Heinz DJ (ed) Sugarcane Improvement Through Breeding. Elsevier, New York, pp 255-272.

Ingelbrecht IL, Irvine JE and Mirkov AC (1999) Posttranscriptional and silencing in transgenic sugarcane: Dissection of homology-dependent virus resistance in monocot that has a complex polyploid genome. Plant Physiol 199:1187-1197.

Joahal GS and Briggs SP (1992) Genetic dissection of bacterial speck disease resistance in tomato. Euphytica 79:195-200.

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

Kruijt M, Kip DJ, Joosten MH, Brandwagt BF *et al*. (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.

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.

Li ZK, Sanchez A, Angeles E, Singh S, Domingo J, Huang N and Khush GS (2001) Are the dominant and recesive plant disease resistance genes similar? A case study of rice *R* genes and *Xanthomonas oryzae* pv. *oryzae* races. Genetics 159:757-765.

Liang FS, Kong FN, Zhou CJ, Cao PX, Ye CJ and Wang B (2005) Cloning and characterization of a non-TIR-NBS-LRR type disease resistance gene analogue from peach. DNA Seq 16:103-110.

Mafia RG and Alfenas AC (2003) Diferenciação sintomatológica de manchas foliares em *Eucalyptus* spp. causadas por patógenos fúngicos e bacterianos. Fitopatol Bras 28:688-688.

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

McDowell JM, Dhandaydham M, Long TA, Aarts MG Goff S, Holub EB and Dangl JL (1998) Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of *Arabidopsis*. Plant Cell 10:1861-1874.

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.

Mestre P and Baulcombe DC (2006) Elicitor-mediated oligomerization of the tobacco N disease resistance protein. Plant Cell 18:491-501.

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

Michelmore RW, Hulbert SH, Landry BS and Leung H (1987) Towards a molecular understanding of lettuce downy mildew. In: Day PR, Jellis GJ (eds) Genetics and plant pathogenesis. Cambridge, pp 221-231.

Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I, Zabel P and Williamson VM (1998) The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. Plant Cell 10:1307-1319.

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

Morais DAL (2003) Análise bioinformática de genes de resistência a patógenos no genoma expresso da cana-de-açúcar. Tese de Mestrado, UFPE, Recife.

Ng NQ and Maréchal R (1985) Cowpea taxonomy, origin and germplasm. In: Singh SR and Rachie KO (eds) Cowpea research, production and utilization. Chichester, pp 11-21.

Oerke EC, Dehne HW, Schonbeck F and Weber A (1994) Crop production and crop protection: estimated losses in major food and cash crops. Elsevier, Amsterdam.

Ohtsuki A and Sasaki A (2006) Epidemiology and disease control under gene-for-gene plant-pathogen interactions. J Theor Biol 238:780-794.

Oliveira IP and Carvalho AM (1988) A cultura do caupi nas condições de clima e solo dos trópicos úmidos e semi-áridos do Brasil. In: Araújo JP and Watt EA (eds) O caupi no Brasil. IITA/EMBRAPA, Brasília, pp 65-95.

Ordosgoitti A., González V, Aponte A. (1982) El carbon de la caña de azucar en Venezuela. Agronomía Tropical, 31: 301-307.

Ori N, Eshed Y, Paran I, Presting G, Aviv D, Tanksley S, Zamir D and Fluhr R (1997) The I2C family from the wilt disease resistance locus I2 belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. Plant Cell 9:521-532.

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

Padulosi S, Ng QN and Perrino P (1997) Origin, taxonomy and morphology of *Vigna unguiculata* (L.) Walp. In: Singh BB and Raj M (eds) *Advances in Cowpea Research.*

Parker JE, Coleman MJ, Dean C and Jones JGD (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.

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

Piffanelli P, Zhou F, Casais C, Orme J, Jarosch B, Schaffrath U, Collins NC, Panstruga R and Schulze-Lefert P (2002) The barley MLO modulator of defense and cell death is responsive to biotic and abiotic stress stimuli. Plant Physiol 129:1076-1085.

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

Portieles R, Rodriguez R, Hernández I, Canales E and Cornide MT (2002) Determinación del número cromosómico de um grupo de clones silvestres de origen desconocido y clones de fundacióndel complejo *Saccharum.* Cultivos Tropicales 23:69-72.

Prosdocini F, Cerqueira GC, Binneck E, Silva AF, Reis NA, Junqueira ACM and Santos ACF (2002) Bioinformática: Manual do usuário. Biotecnol Ciên Desenvol 29:12-25.

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

Quirino MS (2003) Polimorfismo de seqüências nucleotídicas em fragmentos genômicos de cana-de-açúcar homólogos a genes de resistência. Tese de Mestrado, ESALQ, Piracicaba.

Rehmany AP, Gordon A, Rose LE, Allen RL, Armstrong MR, Whisson SC, Kamoun S, Tyler BM, Birch PR and Beynon JL (2006) Differential recognition of highly divergent downy mildew avirulence gene alleles by *RPP1* resistance genes from two *Arabidopsis* lines. Plant Cell 17:1839-1850.

Richly E, Kurth J and Leister D. (2002) Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. Mol 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 and Daniels J (1987) A review of the origin and improvement of sugarcane. In: COPERSUCAR International Sugarcane Workshop. COPERSUCAR, Brazil, pp 1-31.

Rocha MM, Lima JA., Freire-Filho FRM, Rosal CJS and Lopes ACA (2003). Resistência de genótipos de caupi (*Vigna unguiculata* L. Walp.) de tegumento branco a isolados de vírus das famílias Bromoviridae, Comoviridae e Potiviridae. Revista Ciência Rural 8: 85-92.

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

Sagar A, Daemmrich A and Ashyia M (2000) The tragedy of the commoners: biotechlology and its publics. Nat Biotechnol 18:2-15.

Salmeron JM, Oldroyd GED, Romens CMT and Scofield SR (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.

Sanguino A. (1998) Situação atual da pesquisa em doenças da cana-de-açúcar. Summa Phytopathol., 24:1 90-91pp.

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

Silveira RLVA (2000) Efeito do potássio no crescimento, nas concentrações dos nutrientes e nas características da madeira juvenil de progênies de *Eucalyptus grandis* Hill ex Maiden cultivadas em solução nutritiva. Piracicaba, 2000. 169p. Tese (Doutorado) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo.

Silveira RLVA, Gonçalves NA and Krugner (1998) Estado nutricional de *Eucalyptus citriodora* Hook cultivado sob diferentes doses de boro e sua relação com a agressividade de *Botryosphaeria ribis*. *Scientia Forestalis*, 53: 57-70.

Simon MV, Benko-Iseppon AM, Resende LV, Winter P and Kahl G (no prelo). Genetic diversity and phylogenetic relationships in *Vigna* Savi germplasm revealed by DNA amplification fingerprinting (DAF). Genome (Canada), submetido Agosto/2006, aceito Jan/2007, Vol. 50.

Smith DT (1999) Crop profile for sugarcane in Texas – Diseases. Disponível em: http://www.pestdata.ncsu.edu/cropprofiles/docs/txsugarcane.html. Acesso em: 10 jul 2007.

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

Steele WM and Mehra KL (1980) Structure, evolution and adaptation to farming systems and environment in *Vigna*. In:Summerfiel RJ and Bunting AH (eds) Advances in Legume Science.Royal Botanic Gardens, England, pp 459-468.

Sudhir K, Koichiro T and Masatoshi N (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5:150-163.

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.

Tateno Y, Imanishi T, Miyazaki S, Fukami-Kobayashi K, Saitou N, Sugawara H and Gojobori T (2002) DNA databank of Japan (DDBJ) for genome scale research in life science. Nucleic Acids Res 25:4876-4882.

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

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic 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 (IITA) Nigeria, pp 197-212.

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

Van der Biezen EA, Freddie CT, Kahn K, Parker JE and Jones JD (2002) *Arabidopsis* RPP4 is a member of the RPP5 multigene family of TIR-NB-LRR genes and confers downy mildew resistance through multiple signaling components. Plant J 29:439-51.

Van Leeuwen H, Garcia-Mas J, Coca M, Puigdomenech P and Monfort A (2005) Analysis of the melon genome in regions encompassing TIR-NBS-LRR resistance genes. Mol Genet Genomics 273:240-51.

Wang ZX, Yano M, Yamanouchi U and Iwamoto M (1999) The Pib gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant J 19:55-64.

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

Whithan S, McCormick S and Baker B (1996) The N gene of tobacco confers resistance to tobacco mosaic virus in transgenic tomato. Proc Natl Acad Sci USA 93:8776-8781.

Xu WH, Wang YS, Liu GZ, Chen X, Tinjuangjun P, Pi LY and Song WY (2006) The autophosphorylated Ser686, Thr688, and Ser689 residues in the intracellular juxtamembrane domain of *Xa21* are implicated in stability control of rice receptor-like kinase. Plant J 45:740-751.

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang ZX, Kono I, Kurata N, Iwata N and Sasaki T (1998) Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc Natl Acad Sci USA 95:1663-1668.

INTERNET RESOURCES

Associação Nacional dos Fabricantes de Celulose e Papel (BRACELPA), http://www.bracelpa.org.br (October 23, 2006)

Centro Nacional de Referência em Biomassa (CENBIO), http://www.cenbio.org.br/pt/index.html (September 16, 2006)

Cooperativa de Produtores de Cana, Açúcar e Álcool do Estado de São Paulo (COPERSUCAR), http://www.copersucar.com.br (December 12, 2006)

Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), http://www.embrapa.br (October 23, 2006)

Eucalyptus genome sequencing consortium (FORESTs), https://forests.esalq.usp.br (June 10, 2005)

Food and Agriculture Organization (FAO), http://www.fao/unesco/org.sw (February 10, 2005)

Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), http://www.fapesp.br (December 12, 2006)

International Rice Research Institute (IRRI), http://www.irri.org/ (February 10, 2005)

Nucleotide Center for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov (October 19, 2006)

Sociedade Brasileira de Silvicultura (SBS), http://www.sbs.org.br (June 10, 2005)

Sugarcane Expressed Sequence Tag Project (SUCEST), http://www sucest lad dcc unicamp br/en/ (December 12, 2006)

4.0 – ARTIGO CIENTÍFICO

Abundance and Diversity of Resistance (*R*) Genes in the Sugarcane Transcriptome Revealed by *In Silico* Analysis

Ana Carolina Wanderley-Nogueira; Nina da Mota Soares-Cavalcanti; David Anderson de Lima Morais; Luiz Carlos Belarmino; Adriano Barbosa-Silva and Ana Maria Benko-Iseppon

Universidade Federal de Pernambuco, Center for Biological Sciences (CCB), Departamento de Genética, Laboratório de Genética e Biotecnologia vegetal, Recife, PE, Brazil.

Short running title: Resistance (*R*) genes in Sugarcane.

Key words: data mining, sugarcane, serine-threonine kinase, nucleotide binding site, leucine-rich repeats, gene-for-gene interaction.

Corresponding Author:
Ana Maria Benko-Iseppon, UFPE/CCB/Genética, Laboratório de Genética e Biotecnologia Vegetal, Av. Prof. Moraes Rego, s/ nº; 50732-970, Recife, PE, Brazil. E-mail: ana.benko.iseppon@pesquisador.cnpq.br

ABSTRACT:

Resistance genes (*R*-genes) are responsible for the first interaction of the plant with pathogens being responsible for the activation (or not) of the defense response. Despite their importance and abundance, no tools for their automatic annotation are still available. The present work analyzed *R*-genes in the SUCEST (sugarcane expressed sequence tags) database that includes 26 libraries of different tissues and development stages comprising 237,954 ESTs. A new annotation routine was used in order to avoid redundancies and overestimation of *R*-genes number, common mistakes in previous evaluations. After *in silico* screening 280 resistance genes have been identified, 196 bearing the complete domains expected. Regarding the alignments most of the sugarcane's clusters got best matches with proteins from *Oryza sativa*, probably due to the prevalence of sequences of this monocot in data banks. All *R*-gene classes were found except subclass LRR-NBS-TIR including Toll Interleukin-1 receptors, with prevalence of the kinase (*Pto*-like) class. *R*-genes were expressed in all libraries but flowers, transition root to shoot and roots were the most representative suggesting that in sugarcane the expression of *R*-genes in non induced conditions prevail in these tissues. On leaves only low level of expression was found for some gene classes while others were completely absent. A high allelic diversity was found in all classes of *R*-genes, sometimes presenting best alignments with dicotyledons, despite the great number of genes from rice, maize and other grasses deposited in data banks. The results and future possibilities regarding R-genes in sugarcane research and breeding are further discussed.

Key words: *Saccharum*, defense, pathogens, NBS-LRR, kinase, expression profile.

INTRODUCTION:

A major concern regarding plant genome research is to recognize genes responsible for important traits, including defence genes against infection by pathogens. Because plants are sessile, they cannot move to avoid biotic attack or abiotic stress, or to find mating partners. Thus, they depend heavily on chemical signals. Large scale sequencing revealed that plants present many more genes than animals mainly due to polyploidy or large-scale duplication (Borevitz and Ecker, 2004).

Different defense mechanisms are responsible for the protection of plants and animals against their biotic environment. No major histoincompatibility complex genes or antibody-like genes could be identified in plants; however, plant resistance (*R*) genes are abundant and can be grouped into subfamilies (Jones, 2001; Meyers et al., 2003).

Sugarcane is one of the most important sources of sugar and alcohol in the world and is cultivated in tropical and subtropical areas in more than 80 countries around the globe. In 2004/2005 2.7x 10^7 tons of sugarcane were produced only in Brazil in an area estimated on one million ha and was used mainly for sugar consumption or as energy source (ethanol), contributing for 25% of the world's production (UDOP, 2007). The cultivated sugarcane varieties are the result of interespecific hybridization involving *Saccharum officiarum*, *S. barberi*, *S. sinense* and both wild species *S. spontaneum* and *S. robustum*. It is thought that *S. officinarum* was originally selected by humans in Papua New Guinea, perhaps from *S. robustum* germplasm. Because of its multispecific origin, sugarcane is thought to have one of the most complex plant genomes, carrying also variable

chromosome numbers (generally 2n=70-120) with a commensurately large DNA content (Lu et al., 1994).

A large-scale sequencing of sugarcane expressed sequence tags (ESTs) was carried out as a first step in depicting the genome of this important tropical crop. Twenty-six unidirectional cDNA libraries were constructed from a variety of tissues sampled from thirteen different sugarcane cultivars. A total of 291,689 cDNA clones were sequenced in their 5' and 3'end regions. After trimming low-quality sequences and removing vector and ribosomal RNA sequences, 237,954 ESTs potentially derived from protein-encoding messenger RNA (mRNA) remained. The average insert size in all libraries was estimated to be 1,250 bp with the insert length varying from 500 to 5,000 bp. Clustering the 237,954 sugarcane ESTs resulted in 43,141 clusters (Vettore et al., 2001). No general evaluation of resistance (*R*) genes is still available for the sugarcane transcriptome.

Despite the importance of such genes for breeding purposes, no automatic annotation tools are still available. This may be justified by the nature of *R* genes that combine a limited number of related functional domains also regarding different gene classes (Ellis *et al.*,1998, 1999). Previous grouping using domains as primary seed sequence resulted in overestimation of gene number and misclassification. This can be explained by evidences that known *R*-genes combine a limited number of related functional domains (Ellis *et al.*, 1998,1999). Thus, a better understanding of the nature of these genes is necessary in order to understand the difficulties and potentialities regarding automatic annotation, especially in complex genomes like sugarcane.

Plant resistance genes (*R*) are responsible for the specific defense response and are the most important group of genes used by breeders for disease control (Rommens and Kishore, 2000). These genes evolve rapidly, since they suffer constant selection pressure by pathogen evolution. For each *R* gene there is a corresponding gene in the pathogen, called

avirulence (*avr*) gene, which determines pathogenicity. Plant will be resistant and the growth of the pathogen will be arrested only when both genes, *R* and *avr*, are present and compatible (Ellis and Jones, 2000). So, for each *R*-gene a corresponding *Avr* gene co-exists: this is the basis of the gene-for-gene concept, suggested by Flor (1956, 1971).

This gene-for-gene interaction is very specific (Meyers et al., 2005; Salvaudon et al., 2005; Ohtsuki and Sasaki, 2006). The *avr* genes determine the inability of a given pathogenic strain to infect a plant that carries the corresponding *R*-gene breaking out the hypersensitive reaction (HR; Bonas and Anckerveken, 1999). This relationship that is hypersensitive, race specific, and governed by interactions between avirulence genes in pathogens and resistance genes in hosts, is called qualitative resistance (Nelson, 1972).

On the opposite of *R*-genes, *avr* genes products actually described do not comprise a defined family of related proteins, since no sharing similar motifs or domains could be identified (Richter and Ronald, 2000). Resistance genes are members of a very large multigene family, are highly polymorphic and have diverse recognition specificities (Pryor and Ellis, 1993). The cloned resistance genes were grouped into five classes, based on the predicted protein structure (Song et al., 1997).

The first class includes the tomato gene *Pto*, which confers resistance to *Pseudomonas syringae* pv. *tomato*, this gene encodes an active serine/threonine kinase that plays a direct role in both signaling processes and pathogen effectors (Tang et al., 1999; Anderson et al., 2006).

Regarding the second class, the common feature is the presence of leucine-rich repeats (LRRs) which play a direct role in protein-protein specific recognition event; a nucleotide-binding site (NBS) that usually signalizes for programmed cell death and a leucine zipper (LZ) or a coiled-coil (CC) sequence, involved in signal transduction during many cell processes. *R*-genes of this class could be found in many plants as *Arabidopsis*

thaliana (*Rps2, RPP8, RPP13* and *Rpm1*), rice (*Pib, Pi-ta* and *Xa1),* tomato (*Prf, I2, Mi and Sw5*) and potato (*Hero*) *R*-genes (Mindrinos et al., 1994; Whitham et al., 1994; Grant et al., 1995; Lawrence et al., 1995; Bent, 1996; Salmeron et al., 1996; Ori et al., 1997; McDowell et al., 1998; Milligan et al., 1998; Yoshimura et al., 1998; Wang et al., 1999; Bittner-Eddy et al., 2000; Bryan et al., 2000; Brommonschenkel et al., 2000; Ernst et al., 2002 and Rehmany et al., 2005).

The third class includes similar proteins as described for the second class (often both types are classified together in a single gene class), but instead of a CC sequence at the amino terminal region (Meyers et al., 1999) it presents a TIR (Toll Interleucine-Like domain), including the genes *L* (Lawrence et al., 1995), and *P* (Dodds et al., 2001) of flax; *RPP1* (Botela et al., 1998), *RPP4* (van der Biezen et al., 2002), *RPP5* (Parker et al., 1997) and *RPS4* (Gassmann et al., 1999) of *A. thaliana* and *N* (Whithan et al., 1996 and Mestre and Baulcombe, 2006) of tobacco. The TIR domain is also present in animals and is supposed to be absent in monocotyledonous plants (Ellis and Jones, 2000), being present in all dicotyledonous taxa actually studied.

A fourth class of resistance genes is represented by tomato *Cf g*ene family (*Cf-2, Cf-4 Cf-5 and Cf9),* that mediates resistance to the fungal pathogen *Cladosporium fulvum* (Jones et al., 1994; Dixon et al., 1996; Kruijt et al., 2005). This gene encodes a putative membrane-anchored protein (TM, Transmembrane domain) with the LRR motif in the presumed extracellular domain and a short C-terminal tail in the intracellular domain.

The fifth class is represented by the rice gene *Xa21* (Song et al., 1995; Wang et al., 1996) which encodes an extracellular receptor-like kinase, including also a TM, LRRs and an intracellular serine/threonine kinase domain. Thus, the structure of *Xa21* indicates an evolutionary link between different classes (I and IV) of plant disease resistance genes (Song et al., 1997; Xu et al., 2006).

There is still a sixth class, which encodes genes of the reductase group with no conserved domains like cited above. This class is represented by the maize *Hm1* gene, that confers resistance against the toxin produced by the fungus *Cochliobolus carbonum* (Joahal and Briggs, 1992) and *Mlo* from barley, a putative regulator of defense against *Blumenaria graminis* (Piffanelli et al., 2002).

Previous evaluations have shown that the automatic annotation of R genes may bring to redundancy and wrong classification of *R* genes (Meyers *et al.*, 1999). In the present paper we use of 30 complete sequences of previously described *R*-genes as template and propose a new approach for unambiguous identification and classification of *R*-genes candidates in plants.

Further important questions regarding the present work include: How many *R*-genes can be identified in the SUCEST database? Do they correspond to the known *R*-gene classes with the same combinations of conserved domains? Are they preferably similar to other Poaceae (e.g. rice, wheat, maize) available sequences in databases? In which tissues are they expressed in non induced conditions? Considering the allopolyploid and hybrid origin of the sugarcane genome, can one expect a larger diversity of alleles regarding the expressed resistance genes as compared with diploids as rice, maize and *Arabidopsis*? The present work attempts to bring light to some of these open questions using a data mining-based analysis of plant disease *R*-genes in SUCEST database, as compared with available information from other plants deposited in public databases.

MATERIAL AND METHODS

The sugarcane ESTs used in the present work are available in Genbank (NCBI, National Center for Biotechnology Information, www.ncbi.nih.gov/). The clusterized ESTs are available at www.biotec.icb.ufmg.br/sucest**.** Information regarding the 26 libraries that

constitute the SUCEST libraries, including experimental conditions and pipeline routines have been described before (Grivet and Arruda, 2001; Vettore et al., 2001). For practical purposes we united some libraries that comprised different stages of the same tissue/organ (AM1 and AM2 are here referred simply as "AM", for example), resulting in a total of 13 libraries (**AD:** tissues infected by *Gluconacetobacter diazotroficans,* **AM:** Apical meristem; **CL:** Callus; **FL:** Flower; **HR**: tissues infected with *Herbaspirillum rubrisubalbicans*; **LB:** Lateral Bud; **LR:** Leaf Roll ; **LV:** Leaves; **RT**: Root; **RZ:** Stem-Root transition; **SB:** Stalk Bark; **SD:** Seeds; **ST:** Stem) considered for a better evaluation of the results of the present work.

For the identification of sugarcane *R*-genes a search was carried out using sequences of known *R*-genes selected from the literature against the SUCEST database (see Attachment I at the appendix). Members of the sixth class (reductases) have not been included in the present evaluation. The genes selected include 27 *R*-genes previously compiled by Barbosa-da-Silva et al. (2005) including all five gene classes previously described. To this work we included three sequences, namely the genes *Pi-ta* and *Pib* from rice and *RPM1* from *A. thaliana* (accession numbers AAK00132, BAA76282 and AC016827_19, respectively) all belonging to the second class (LRR-NBS) described before.

For the identification of *R*-genes, tBLASTn alignments were carried out against SUCEST database using the 30 seed sequences related above. After this search, sugarcane sequences that were found to match *R*-genes with a cut-off of e^{-20} have been used for a homology screening of *R*-genes in Genbank (NCBI) using the BLASTx (Altschul et al., 1990). The clusters frame of the tBLASTn alignment was used to predict the Open Reading Frames (ORFs) for each selected cluster.

A second general analysis using a cut-off of e^{-10} was also carried out followed by an elimination of some redundancies (genes that matched more than one gene class due to common domains). For this purpose, matching clusters to each query sequences were annotated on a local database (called 'non-redundant'). Cluster name was adopted as primary key in order to identify and prevent inclusion of the same cluster in different gene classes due to the presence of common domains.

Exclusively in the case of the third class of *R*-genes (LRR+NBS+TIR) an additional tBLASTn search was carried out using only the TIR domain to confirm its presence/absence in sugarcane.

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). Multiple alignments with CLUSTALx program allowed the structural analysis of the sequences including conserved and diverging sites as well as elimination of non aligned terminal segments.

For each R-gene class, one resistance gene (*Pto, Xa1, Cf* and *Xa21*, respectively) was selected to perform a phenetic UPGMA (Unwheited Pair Groups Method using Arithmetic Averages) analysis using bootstrap function with 1,000 replicates. For this purpose CLUSTALx alignments have been submitted to the program MEGA (Molecular Evolutionary Genetic Analysis), Version 3 for Windows, kindly provided by the authors (Kumar et al., 2004). The resulting dendrogram was created with the program TreeView for Windows (Page, 1996) kindly provided by Dr. Robert Page (University Glasgow, Scotland).

A preliminary analysis of *R*-gene distribution patterns in sugarcane libraries was verified by direct correlation of the reads' frequency of each cluster in various SUCEST cDNA libraries (Figure 3).

To generate an overall picture of *R*-genes expression patterns in sugarcane, a hierarchical clustering approach (Eisen et al., 1998) 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 gene class and lybrary) were generated by the TreeView program (Eisen et al., 1998). On the graphics (Figure 4), yellow means no expression and red all degrees of expression. This approach was previously employed by other plant EST projects as in rice (Ewing et al., 1999) and also in sugarcane (Lambais, 2001).

RESULTS AND DISCUSSION

Using 30 well known *R*-genes as template we could identify 196 clusters in SUCEST database bearing the complete expected domains. Considering the identity of these genes with the queries, a total of 151 clusters could be identified as non-redundant while 45 other sequences aligned with more than one *R*-gene used as query (Table 1).

The use of several previously described and sequenced *R*-genes as seed sequences proved to be a useful and low-time consuming strategy in the search for *R*-genes candidates in plants. This approach allowed the identification of a large set of candidate sequences by using various representative genes per class, while former works (e.g. Koczyk and Chelkowski, 2003) employed few genes. Previous works have shown that using only domains or few genes as template per class resulted in double grouping of some genes in different classes and caused some level of redundancy (Meyers et al., 1999). In other cases a higher stringency had to be used (e^{-50} or less) resulting on exclusion of important gene candidates Rossi et al., 2003). For example, the kinase domain (present in classes I and V) or the NBS domain (present in classes (II and III) often bring to the generation of mixed grouping. Furthermore the imperfect nature of the LRR domain alone may bring some

problems regarding automatic annotation and classification showing that this domain is not adequate for this purpose (Barbosa-da-Silva et al., 2005). Thus, the strategy of generating a local database (here called non-redundant) by adopting the cluster number as a primary key register was very effective in the solution of this problem, helping their recognition, classification, elimination of duplicates and inferences about candidates of orthologs and paralogs. We recommend this procedure for the future development tools specific for *R* gene automatic annotation quantification and classification.

Lowering the cut-off value (from e^{-20} to e^{-10} during tBLASTn) additional 84 *R*-gene-clusters could be identified in the sugarcane transcriptome but many presented only partial sequences or incomplete domains. Altogether, this means that sugarcane encodes a significant number of transcriptionally active *R*-genes (at least 280) with considerable allelic diversity. This number is much higher than the 88 sugarcane sequences identified for development of RGA (resistance gene analog) markers by Rossi et al. (2003). The authors used keyword search and 17 RGA-related seed sequences with a stringent BLASTn cutoff (e^{-50}). Many important sequences bearing complete domains have been excluded using this approach, what is justified by some needs for use in comparative *in silico* mapping, the main focus of this work. Despite the undoubted identification of *R* genes using this approach, no overall picture of *R* genes abundance and diversity within an EST database is possible by using this procedure.

It is interesting to note that only two of the 26 sugarcane libraries were obtained under influence of microorganisms (tissues infected by *Gluconacetobacter diazotroficans* and tissues infected with *Herbaspirillum rubrisubalbicans*) and that none of them are pathogenic and, on the opposite, are symbiotic organisms. Under exposition to pathogen the number of genes will probably increase and one may suppose that additional sequences may be identified.

Clusters representing exclusive *R*-gene classes were: (I) kinase: 92; (II) LRR-NBS-CC: 62; (IV) TM-LRR: 27 and (V) kinase-TM-LRR: 15. Clusters that aligned with *R*-genes class II and III (TIR-NBS-LRR) showed only CC, NBS and LRR domains (no TIR domain) and were therefore included in the group II, since the presence of TIR is the distinctive factor between both classes.

The prediction of clusters coding regions revealed that ORFs were oriented in both forward and reverse reading frames, with an average of 394 amino acids (aa) in length. ORF sizes varied from 992 (cluster SCCCLR1001A03.g of the LRR class) to 102aa. Regarding the average ORF length in each *R*-gene class, we observed 380 aa for class I (kinase), 262 aa for class II (LRR-NBS-CC), 492 aa for class IV (TM-LRR) and 442 aa for class V (kinase-TM-LRR) (Figure 1).

Most of the 196 clusters that aligned with known *R*-genes were from monocotyledonous class (169 clusters), represented by eight different species of the Poaceae family, with emphasis on rice. From dicots only three families appeared as best matches (27 clusters), including four different species. A comprehensive inventory of all species that aligned with sugarcane with their taxonomic affiliation is presented in Table 2.

Lowering the cut-off value from e^{-20} to e^{-10} during tBLASTn alignments additional *R*-gene clusters could be identified in sugarcane (Figure 1), with exception of the kinase group (class I), where the same number of clusters (92) was identified with both approaches. Considering the remaining classes a higher number of clusters could be identified, even though in some of them domains were incomplete or missing. This is the case of the class II (LRR-NBC-CC) where 120 clusters have been identified (instead of 62 previously identified at e^{-20}), similar to class IV (TM-LRR) with 41 instead of 27 and class V (TM-LRR) with 27 instead of 15 (Figure 1). With this lower cut-off the total number of putative transcribed *R*-genes in sugarcane increased from 196 to 280.

General considerations about conserved domains of sugarcane *R*-genes

Some *R*-genes pertaining to different classes were able to align significantly to the same cluster on SUCEST database, an occurrence also observed during mining of the *Eucalyptus* transcriptome (Barbosa-da-Silva et al., 2005), probably because known *R*-genes combine a limited number of related functional domains (Ellis et al., 1999).

The conserved domains (CDs) identified during this investigation showed that most of the sugarcane predicted sequences presented the same motifs shared by previously known disease *R*-genes. The CD with the higher level of sampling was kinase, which was present in class I and class V with a total of 107 occurrences.

The CD-search revealed conserved regions (Figure 1) in all of the 196 analyzed clusters. From the 107 clusters that presented the kinase domain, and 92 of them matched to *Pto* gene (class I) while 15 matched the *Xa21* gene (class V).

The NBS domain was present in 62 clusters in sugarcane (120 with cut-off of e^{-10}). After search with genes encoding class III (LRR-NBS-TIR) only non-TIR sequences matching with class II (due to the common NBS domain) could be identified. No significant matches have been found also after tBLASTn search using exclusively the TIR domain.

NBS domain 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), with the additional proposition that NB-ARC plays a role in activation of downstream effectors (Bryan et al., 2000). Transmembrane motifs were found only in 19 of all analyzed sequences, where 14 were related to *Cf* gene and five to *Xa21*.

The other frequent domain shared was LRR, matching 104 occurrences in 77 different clusters in all classes except kinase (class I) represented by the *Pto* gene. LRR can

appear like a receptor to recognize the avr proteins, as in *Cf* (27 clusters) and *Xa21* (15 clusters) or intracellular, like in class II of R-proteins (62 clusters). The LRR motif contains 23-25aa with a consensus sequence (LxxLxxLxLxxNxLt/sgxIpxxLG) but this pattern is often imperfect (Jones, 2001) and may be difficult to recognize with available *in silico* tools, so it is possible that a larger number would be recognized with lower cut-off and additional manual search.

Class I: *Pto*-like *R*-genes (solely kinase domain)

The first class includes the tomato gene *Pto*, which confers resistance to *Pseudomonas syringae* pv. *tomato*. In tomato *Pto* has been described as a small gene. The open reading frame (ORF) consists of 963 nucleotides, it has no introns, and encodes a functional serine-threonine kinase (Loh and Martin, 1995). 52 alleles of this gene were found in a search including seven *Lycopersicon* species, bearing 41 variant amino acid positions among these alleles (Rose et al., 2005). 92 clusters of sugarcane presented high significant alignments to the used *Pto* seed sequence (accession 2112354A). Size of the clusters varied between 1926 and 759nt with ORFs from 642 to 253aa (Table 1, Figure 1) indicating that gene size within this class may vary significantly.

In the sugarcane transcriptome no redundancy was observed between this class and the *Xa21* class, which also presents a kinase domain. This is in accordance with the observations of Vallad et al. (2001) that used bootstrap analysis to determine that five *Pto*-like kinase families from bean were distinctly different from other kinases. They also found that *Pto*-kinase subdomains VIa, VIb, VIII and IX of the Pto-like class are unique in plant species. This conservation is confirmed by the fact that sugarcane genes of this group aligned to 92 sequences deposited in the Genbank (both cut-offs, e^{-20} and e^{-10}), 22 of them being dicots (Table 2).

Class II and III: Sugarcane clusters bearing NBS-LRR and NBS-LRR-TIR domains

Evidences have shown that *R*-genes are quite abundant in higher plants, but the most functionally defined *R*-genes belong to the NBS-LRR class (here including classes II and III for a better understanding of their common and distinctive attributes), considered also the largest of plant disease resistance genes.

Unlike in *A. thaliana* and other dicots, the NBS-LRR gene class coding for a Toll interleukin-1 receptor (TIR) domain has been absent in all studied monocots even though mainly members of the Poaceae family have been analyzed. Most cereal genes are similar in structure to the members of the non-TIR class of dicots, although many do not code for a coiled-coil (CC) domain in their amino termini (Bai et al., 2002).

A total of 85 TIR-NBS-LRR have been identified in *A. thaliana* genome (The Arabidopsis Genome Initiative, 2000) and 93 in *Eucalyptus* transcriptome (Barbosa-da-Silva et al., 2005). The availability of the rice whole genome sequence enabled the global characterization of NBS-LRR genes revealing that this crop carries about 500 NBS-LRR genes (at least three to four times the complement found in *A. thaliana*). Over 100 of these genes were predicted to be pseudogenes in the rice cultivar Nipponbare, but some of these were functional in other rice lines. In rice over 80 other NBS-encoding genes were identified that belonged to four different subclasses, but only two of which are present in dicotyledonous plant sequences present in databases (Monosi et al., 2004). Zhou et al. (2004) considered that 76% of all gene families with a five fold size are bigger in rice as compared with *A. thaliana*. In the sugarcane transcriptome this class was represented by 62 clusters bearing complete domains selected with a cut-off of e^{-20} while 120 clusters could be identified with a lower cut-off (e^{-10}) (Figure 1). Considering the allopolyploid and hybrid nature of sugarcane and also that no libraries under pathogen induction are available in the SUCEST database, an increase in the number of NBS-LRR sequences is expected under

different experimental conditions, possibly more than rice that presents a smaller genome and is diploid.

In sugarcane as in rice and other cereals only non-TIR sequences matching NBS-LRR *R*-genes could be identified. No significant matches have been found also after tBLASTn search using exclusively the TIR domain. This result confirms previous assumptions that the TIR domain (also present in animals) may be absent in monocotyledonous plants (Ellis and Jones, 2000) like Poaceae, being present in all dicotyledonous taxa actually studied. Therefore it has been suggested that the TIR domain may have been lost in the course of differential evolution between mono and dicotyledoneous plants (Pan et al., 2000).

As in rice, also in sugarcane most alignments using class II genes occurred with monocots including only two sequences significantly homolog to dicot sequences deposited in databases (Table 2), both from *A. thaliana*. Exceptionally this is not the most abundant *R*-gene class found in SUCEST database when we consider the most stringent procedure (cut-off e^{-20}), with only 62 sequences included in this group against 92 *Pto*-like sequences that bear exclusively the kinase domain (Figure 1). A more permissive e-value (cut-off e^{-10}) allowed the identification of 120 clusters of the NBS-LRR class, almost the double of the number identified previously, while in the case of the *Pto*-like sequences (kinase domain) the same number of clusters (92) was revealed in both approaches. Previous workers identified that overall sequence homology among *R*-genes of the NBS-LRR class is lower than in the kinase class. By the other hand, the NBS contains some sequence motifs as P-loop, kinase-2, kinase-3 and GLPAL that are highly conserved even among distantly related plants (Hammond-Kosack and Jones, 1996). The wide distribution of NBS-LRR genes in the plant kingdom and their prevalence in both mono and dicots indicate that they are ancient. This was confirmed by Liu and Ekramoddoullah (2003) that amplified TIR-

NBS-LRR in gymnosperms (*Pinus monticola*, white pine) confirming that they share a common origin with R genes from angiosperms. Thus, to identify the whole diversity of class II genes, low stringency alignments are advisable.

In the present work, the low number of dicots with best alignments regarding this class of gene may be explained by the imperfect nature of the LRR domain not ever recognizable with available *in silico* tools.

Class IV: *Cf*-like *R*-genes (TM-LRR domains)

Genes of the *Cf*-family mediate resistance to the fungal pathogen *Cladosporium fulvum* in tomato. They encode a putative membrane-anchored protein (also named transmembrane, TM) with the LRR motif in the presumed extracellular domain and a short C-terminal tail in the intracellular domain (Jones et al., 1994; Dixon et al., 1996).

In our study 25 clusters presenting best matches with this protein belonged to monocots while only two matched to dicot (*A. thaliana*) sequences (at e^{-20}). All selected clusters presented the putative TM-LRR domains. Central for this model for *Cf* protein function is the concept that the highly variable regions within the LRRs are responsible for the recognition of pathogen-encoded avirulence determinants either directly or indirectly through some co-receptor (Dixon et al., 1996). The LRR domains are known to play a role in protein-protein interactions. In tomato the size of these sequences varied between 968 and 855aa (Dixon et al., 1996), while ORFs identified in sugarcane varied between 992 and 614aa.

As observed in the NBS-LRR group, also in the *Cf*-related genes the number of putative related clusters is increased to 41 when the cut-off value is e^{-10} (as compared with 27 clusters with stringent e^{-20} conditions). It seems that also in this case the imperfect nature of the LRR domains make difficult to identify sequences of this group and additional mining is needed since under stringent conditions new variants may be not recognized. By

the other hand, permissive conditions often result in redundancies that may lead to the classification of the same cluster in different gene classes.

Class V: LRR-TM-kinase (*Xa21*-like *R*-genes)

The structure of this class indicates an evolutionary link between the here presented classes I and IV of *R* genes (Song *et al.*, 1997). Overall annotation revealed that *Arabidopsis* also carries homologues to the LRR-kinase-*Xa21* group (Jones, 2001), while eight clusters with significant homology to *Xa21* were found also in far related woody dicots as in case of *Eucalyptus* transcriptome (Barbosa da Silva et al., 2005).

In the present evaluation 15 of the analyzed clusters corresponded to this class with high e-values, but this number will probably increase if only the receptor-like kinase sequence is used as template, since the LRR may be quite variable between sugarcane and rice. Regarding the best alignments, one of the best matches revealed LRR-kinase of carrot (a dicot from the family Apiaceae) and 14 of monocots (all belonging to Poaceae with 10 rice sequences; Table 2).

Using a PCR-based approach Song et al. (1997) cloned seven *Xa21* members in rice and found the presence of 15 transposable element sequences, two of them in coding sequences, confirming the influence of such sequences in the evolution of these genes. Whole sequencing in the rice genome revealed this class of genes in two chromosomes: in the short arm of chromosome 12 (including 12 tandem arrays of *Xa21*-like sequences), while the first described sequence for this gene was found on the long arm of chromosome 11 (The Rice Chromosomes 11 and 12 Sequencing Consortia, 2005). According to these authors both regions are full of defense related genes, confirming the clusterized organization of these sequences in chromosomes. These results reveal that the number of *Xa21* representatives in rice (13) is lower than that found in the sugarcane transcriptome in

non induced libraries (15), suggesting that they are probably still more abundant in sugarcane, including some of the largest contigs included in our evaluation (up to 2184 bp).

Exceptional *R*-genes have proven to provide durable disease control, due to the fast evolving pathogen genome that breaks resistance. The *Xa21* gene is an important exception to this rule that reveals the full potential of *R*-genes for breeding purposes (Rommens and Kishore, 2000). This may be very valuable for sugarcane breeding, especially considering the possibility of pyramidization of such genes in important crops, increasing the potentiality of an effective specific *R-Avr* interaction.

Bootstrap analysis of selected *R*-gene groups

The phenetic UPGMA bootstrap analysis revealed grouping between monocots and dicots in most dendrograms considering the four genes used as template (*Pto*, *Xa1*, *Cf9* and *Xa21*). In all cases sugarcane R clusters appeared in two or more clades within the each dendrogram. Also grouping of species belonging to different taxonomic families was observed in all cases (Figure 2).

Sometimes best alignments match to *O. sativa*, as in case of *Pto* and *Cf9* groups (Figure 2A) but in all cases where *Sorghum bicolor* sequences were available they presented the best matches, as it occurred in the *Xa1* and *Xa21*-groups (Figure 2B, D). By the other hand in all four evaluations some sugarcane clusters presented lower levels of similarity and in three cases (*Pto*, *Xa1* and *Cf9)* a sugarcane cluster was the most divergent, remaining in a basal position in the cladogram that included sequences from mono and dicotyledoneous plants. This makes clear that sugarcane bears a high allelic diversity of *R*-genes, also considering that the present study do not include libraries obtained under conditions of pathogen stress.

It is interesting to note that all trees grouped species pertaining different plant families as expected, since phenetic analysis considers only similarity aspects, not evolutionary. These grouping confirm that *R*-genes, including these four chosen for the phenetic analysis, appeared before divergent evolution of monocots and dicots.

Distribution of ESTs in the SUCEST Lybraries

Considering the distribution of the 2108 reads (comprised in the 196 clusters selected) in the 13 analyzed libraries, a higher prevalence could be observed in flower (FL, 28%), stem-root transition tissues (RZ=12.47%) and apical meristem (AM=12%). It would be expected to find higher levels of expression in root, since this is the main entrance for many pathogenic bacteria, fungi and nematodes. Considering both libraries root (RT) and transition root-stem (RZ) tissues together included 18% of the *R*-genes expressed. Taking in mind that the tissues have been cultivated under controlled non stressed conditions, one may suppose that these results represent *R*-genes that are regularly expressed in root and root to shoot tissues.

Surprisingly a higher prevalence of *R*-genes could be detected in growing tissues as apical meristem (AM) leaf roll (LR) and lateral bud (LB) that altogether comprised 24% of the expressed *R*-genes. If we consider also flower libraries (that included five different early stages of development) as young/growing tissues, 46% of the here annotated expressed *R*-genes were expressed within this group (Figure 3A). Otherwise flower libraries together comprised 63,774 reads (26.8 % of all SUCEST reads) what may explain the presence of some clusters comprising reads mainly from flower libraries.

Lowest prevalence of reads representing *R*-genes was observed in leaves (1%), tissues infected with *H. rubrisubalbicans* (2%), callus (4%) and seed (4%) libraries, respectively (Figure 3A-B).

Strikingly differences could be observed between the prevalence of *R*-genes in the two libraries obtained under influence of microorganisms (tissues infected by *G.* and with *H. rubrisubalbicans*). While the first presented significant presence of *R*-genes (7%) identified among the 18,144 reads sequenced (7.6% of all SUCEST reads) the second (that comprised 12,000 reads/5% of all SUCEST reads) presented one of the lowest (2%) levels of expressed *R*-genes. Both libraries (AD1 and HR1) were constructed with plantlets inoculated with *G. diazotroficans* or *H. diazotroficans,* which are endophytic nitrogen-fixing bacteria that naturally colonize sugarcane tissues (Lee et al., 2000).

The association of endophytic diazotrophic bacteria with plants is quite different than other nitrogen-fixing associations. Diazotrophic bacteria colonizes intercellular spaces and vascular tissues of most organs of the host plant, without causing visible plant anatomical changes or disease symptoms (Reinhold-Hurek and Hurek, 1998).

It has been described that the endophytic diazotrophs produce plant growth-regulating hormones, such as auxin (Fuentes-Ramirez et al., 1993) and more recently, gibberellin (Bastian et al., 1998). The mechanisms involved in the establishment of this particular type of interaction and what kind of molecules mediate signaling between plant and bacteria remain unclear. In addition, very little is known about the role of the plant in the association. Differences in the contribution of BNF to the plant nitrogen balance in distinct sugarcane cultivars suggest that the plant is controlling, at least in part, the efficiency of the process (Urquiaga et al., 1992). The plant could control bacterial colonization by sending the proper signals and/or providing the best physiological conditions for bacterial survival. Another question to be addressed is how the association benefits the plant. The endophytic diazotrophs promote plant growth when inoculated in sugarcane plantlets, possibly by supplying nitrogen and/or plant hormones (Sevilla et al., 2001).

Nogueira et al. (2001) evaluated the SUCEST data regarding both libraries but for this purpose they pooled both libraries together avoiding the identification of distinctive sequence classes between both. Most functional categories identified in the study included transporters, transcription factors and protein kinases. Our results suggest that the interaction of sugarcane with each bacteria was clearly distinctive indicating that the reaction to the infection with *H. diazotroficans* is clearly permissive with a very low level of *R*-gene expression, while *R*-genes seemed to be activated during the interaction with *G. diazotroficans.* These results suggest that a comparative analysis of both libraries using non infected seedlings as control would reveal different physiological conditions.

Considering the distribution of reads correlated with the classification of *R*-genes (Figure 3B) it is clear that reads of class I (with kinase domain) were most abundant in all libraries, followed by class II (LRR-NBS-CC) and class V (kinase-TM-LRR), with exception of flower library (FL) and stalk bark (SB) where the third more representative group was class IV (TM-LRR).

On the opposite, leaves – traditionally the main entrance for viral infections – comprised only 1% of the sequenced R genes. Otherwise one may consider that this was one of the smallest libraries, comprising only 6,342 reads (2.7 % of all SUCEST reads). However, all tissues had *R*-genes expression, suggesting that *R*-genes are overall expressed in constitutively low levels. This is in accordance with the observation of Tang et al. (1999) that detected a basal *Pto*-kinase activity maintained within a low level even when *avrPto* was not present. In the presence of the pathogen *Pto*-kinase is immediately available and its abundance immediately increased. In contrast, *Xa1* (a gene that confers resistance to rice against *Xantomonas oryzae* pv. *oryzae*) mRNA was detected from rice leaves at 5 days after cutting and inoculation of both compatible and incompatible strain of *Xantomonas oryzae* pv. *oryzae*, but was not detected in intact leaves (Yoshimura et al., 1998). These

findings suggest that *R*-genes expression may be induced either by stimulus of wounding involved in the pathogen infection or in tissues subjected to attacks.

One of the lowest *R*-genes expression levels (4%) was observed on *in vitro* cultivated calli treated with contrasting temperatures (cold and hot) and alternating dark and light exposure. This low expression suggests that such abiotic stimuli may recruit some responses that indirectly suppress R gene expression.

Expression Pattern

In silico evaluation of gene expression can be inferred only with normalized differential display data, an approach that considers reordered data matrices. This method also allows the identification of clusters bearing similar expression patterns in cDNA libraries, suggesting that they might be co-regulate *in vivo*. Lambais (2001) studying defense-related (PR-class) proteins in sugarcane, a signal cascade also induced by *R* genes, argued that genes with similar functions or cDNA libraries are expected to have similar patterns of gene expression and also to cluster together in chromosomes. Classical genetic mapping has demonstrated that *R*-genes tend to be clustered in few chromosomes in the genome (Winter et al., 2000; Benko-Iseppon et al., 2003). In *A. thaliana* they are clustered in two chromosome arms (The Arabidopsis Genome Initiative, 2000) as well as in rice (The Rice Chromosomes 11 and 12 Sequencing Consortia, 2005). The same *R* genes have been observed clustered and almost in the same order in tomato (Ku et al., 2000) and chickpea (Benko-Iseppon et al., 2003), confirming that genes order and proximity are important for their proper functionality.

In the case of the present analysis four different approaches (Figure 4) have been used to evaluate sites of expression and patterns of co-regulation of *R*-genes in sugarcane

considering the clusters significantly aligned with the genes *Xa1* (Figure 4A), *Cf* (Figure 4B), *Xa21* (Figure 4C) and *Pto* (Figure 4D). The prevalence of expression in libraries constructed from flowers tissues (FL) is clear for all four groups, with must clusters co-expressing within this library. In the case of the genes *Xa1* and *Cf* the same clusters were also expressed in apical meristems and stalk bark, while in *Xa21* almost only flower and apical meristem presented significant expression (Figure 4D).

For *Cf9* and *Xa21* an almost complete lack of expression was observed in most libraries (Figure 4B-C), while in the case of *Xa1* higher expression (two to 10 fold) could be detected in most libraries.

Pto analogs represent the most interesting case study in sugarcane, since many allelic variants of this group could be found. They appeared expressed in all libraries studied but a clear prevalence of expression in FL (flower), followed by RZ (transition zone root to shoot) and also RT (root) could be observed with many clusters co-expressing in these libraries. Also in the case of *Pto*-like clusters high co-expression levels regarding five sugarcane clusters could be observed in the AD library (seedlings infected with *G. diazotroficans*).

The co-expressed clusters presented in Figure 4 represent important candidate sequences for fine mapping of *R*-gene rich linkage groups in sugarcane, especially using primers designed for conserved flanking sequences, that may reside close in the same linkage group.

It is interesting to note that also after normalized expression analysis of all four selected *R*-genes their representation in leaves was very low, suggesting that the level expression in this tissue may be low (less than 1% in the case of *Xa1* and *Pto*) or absent (*Cf9* and *Xa21*).

Concluding Remarks

Using bioinformatic tools it was possible to identify and classify *R*-genes in sugarcane transcriptome, and also to make some inferences regarding their expression pattern under non induced conditions. All five classes of *R*-genes with their respective conserved domains could be found in sugarcane except the TIR domain that has been absent as in all studied monocots previously studied.

The 196 identified sequences represent valuable resources for the development of markers for molecular breeding and development of RGA (resistance gene analogs) or gene-specific markers specific for sugarcane and other related cereal crops. The identified clusters constitute also excellent probes for physical mapping of genes in sugarcane, giving support to genetic mapping programs and synteny studies. This may be especially useful for a comparative mapping between sugarcane and *Sorghum*, getting around the difficulties of mapping a large and complex genome as it is the case of sugarcane.

The studied sequences probably represent only part of the diversity and number of *R*-genes that are present in cultivated sugarcane. It is to expect that an evaluation of tissues under conditions of stress induced by pathogen infection would reveal additional information about *R*-genes and their expression, especially considering the huge size and complexity of the sugarcane genome, as compared with most angiosperms.

Furthermore, it is necessary to manipulate the expression of these genes in economically important plant species in order to improve disease resistance. Although the field is still in its very much infancy some reports indicate that this strategy is feasible.

ACKNOWLEDGEMENTS

The present authors thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FACEPE (Fundação de Amparo à Pesquisa do Estado de Pernambuco) for the concession of fellowships. To FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) and SUCEST coordination we thank for the access to the Sugarcane EST data bank.

REFERENCES

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

Anderson JC, Pascuzzi PE, Xiao F, Sessa G et al. (2006). Host-mediated Phosphorylation of type III effector AvrPto promotes *Pseudomonas* Virulence and Avirulence in tomato. *Plant Cell* 18:502-514.

Bai J, Pennill LA, Ning J, Lee SW et al. (2002). Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Res.* 12:1871-84.

Barbosa da Silva A, Wanderley-Nogueira AC, Silva RRM, Belarmino LC et al. (2005). In silico survey of resistance (*R*) genes in *Eucalyptus* transcriptome. *Genet. Mol. Biol.* 28:562-574.

Bastian F, Cohen A, Piccoli P, Luna V et al. (1998). Production of indole-3-acetic acid and giberellins A(1) and A(3) by *Gluconacetobacter diazotrophicus* and *Herbaspirillum seropedicae* in chemically-defined culture media. *Plant Growth Regul.* 24:7-11.

Benko-Iseppon AM, Winter P, Huettel B, Staginnus C et al. (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. *Theor. Appl. Genet.* 107:379-386.

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

Bittner-Eddy PD, Crute IR, Holub EB and Beynon JL (2000). RPP13 is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. *Plant J.* 21:177-88.

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

Borevitz JO and Ecker JR (2004). Plant Genomics: The Third Wave. *Annu. Rev. Genomics Hum. Genet.* 5:443–77.

Botella MA, Parker JE, Frost LN, Bittner-Eddy PD et al. (1998). Three genes of the *Arabidopsis* RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. *Plant Cell* 10:1847-1860.

Brommonschenkel SH, Frary A and Tanksley SD (2000). The broad-spectrum tospovirus resistance gene *Sw-5* of tomato is a homolog of the root-knot nematode resistance gene *Mi*. *Mol. Plant-Microbe Interact.* 13:1130-38.

Bryan GT, Wu KS, Farrall L, Jia Y et al. (2000). A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. *Plant Cell* 12:2033-2046.

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

Dixon MS, Jones JGD, Keddie JS, Thomas CM et al. (1996). The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell* 84:451-459.

Dodds P, Lawrence G and Ellis J (2001). Six amino acid changes confined to the leucine-rich repeat beta-strand/beta-turn motif determine the difference between the P and P2 rust resistance specificities in flax. *Plant Cell* 13:163-78.

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 (1998). Structure and function of proteins controlling strain-specific pathogen resistance in plants. *Curr. Opin. Plant Biol.* 1:288-293.

Ellis JG, Lawrence GJ, Luck JE and Dodds N (1999). Identification of regions in alleles of the flax rust resistance gene *L* that determines differences in gene-for-gene specificity. *Plant Cell* 11:495-506.

Ernst K, Kumar A, Kriseleit D, Kloos DU et al. (2002). The broad-spectrum potato cyst nematode resistance gene (*Hero*) from tomato is the only member of a large gene family of NBS-LRR genes with an unusual amino acid repeat in the LRR region. *Plant J*. 31:127-136.

Ewing RM, Kahla AB, Poirot O, Lopez F et al. (1999). Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res. 9*: 950-959.

Flor HH (1956). The complementary genetic systems in flax and flax rust. *Adv. Genet*. 8:29-54.

Flor HH (1971). Current status of the gene-for-gene concept. *Annu. Rev. Plant Pathol.* 9:275-296.

Fuentes-Ramirez LE, Jimenez-Salgado T, Abarca-Ocampo IR and Caballero-Mellado J (1993). *Gluconacetobacter diazotrophicus*, an indolacetic acid-producing bacterium isolated from sugarcane cultivars in Mexico. *Plant Soil* 154:145-150.

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.

Grant MR, Godiard L, Straube E, Ashfield T et al. (1995). Structure of the *Arabidopsis RPM1* gene enabling dural specificity disease resistance. *Science* 269:843-846.

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 (1996). Resistance Gene-Dependent Plant Defense Response. *Plant Cell* 8:1773-1791.

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

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

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

Ku H-M, Vision T, Liu J and Tanksley SD (2000). Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication following selective gene loss creates a network of synteny. *Proc. Natl. Acad. Sci. USA* 97:9121–9126.

Kruijt M, Kip DJ, Joosten MH, Brandwagt BF et al. (2005). The *Cf-4* and *Cf-9* resistance genes against *Cladosporium fulvum* are conserved in wild tomato species. *Mol. Plant Microbe Interact.* 18:1011-21.

Kumar S, Tamura K, and Nei M (2004). MEGA 3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Brief. Bioinf*. 5:150-163.

Lambais MR (2001). *In silico* differential display of defense-related expressed sequence tags from sugarcane tissues infected with diazotrophic endophytes. *Genet. Mol. Biol.* 24:1-4.

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 S, Reth A, Meletzus D, Sevilla M et al. (2000). Characterization of a major cluster of *nif, fix*, and associated genes in a sugarcane endophyte, *Acetobacter diazotrophicus. J. Bacteriol.* 182:7088-7091.

Liu JJ and Ekramoddoullah AK (2003). Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (*Pinus monticola* Dougl. ex. D. Don.). *Mol. Genet. Genomics* 270:432-441.

LohYT and Martin GB (1995). The *Pto* bacterial resistance gene and the *Fen* insecticide sensitivity gene encode functional protein kinases with serine/threonine specificity. *Plant Physiol.* 108:1735-1739.

Lu YH, D'Hont A, Paulet F, Grivet L et al. (1994). Molecular diversity and genome structure in modern sugarcane varieties. *Euphytica* 78:217-226.

McDowell JM, Dhandaydham M, Long TA, Aarts MG et al. (1998). Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of *Arabidopsis*. *Plant Cell* 10:1861-1874.

Mestre P and Baulcombe DC (2006). Elicitor-mediated oligomerization of the tobacco N disease resistance protein. *Plant Cell* 18: 491-501.

Meyers BC, Diekcman AW, Michelmore RW, Sivaramakrishnan S et al. (1999). Plant disease resistance genes encodes members of an ancient and diverse protein family within the nucleotide-biding superfamily. *Plant J.* 20:317-332.

Meyers BC, Kozik A, Griego A, Kuang H et al. (2003). Genome wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15:809–34.

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

Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I et al. (1998). The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. *Plant Cell* 10:1307-19.

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

Monosi B, Wisser RJ, Pennil L and Hulbert SH (2004). Full-genome analysis of resistance gene homologues in rice. *Theor. Appl. Gen.* 109:1434-47.

Nelson RR (1972). Stabilizing racial populations of plant pathogens backby use of resistance genes. *J. Environ. Qual.* 3:220-227.

Nogueira EM, Vinagre F, Masuda HP, Vargas C et al. (2001). Expression of sugarcane genes induced by inoculation with *Gluconacetobacter diazotrophicus* and *Herbaspirillum rubrisubalbicans*. *Genet. Mol. Biol.* 24:1-4.

Ohtsuki A and Sasaki A (2006). Epidemiology and disease control under gene-for-gene plant-pathogen interactions. *J. Theor. Biol*. 238:780-794.

Ori N, Eshed Y, Paran I, Presting G et al. (1997). The I2C family from the wilt disease resistance locus I2 belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. *Plant Cell* 9:521-532.

Page RDM (1996). TreeView: An application to display phylogenetic trees on personal computers. *Comp. Applic. Biosci.* 12: 357–358.

Pan Q, Wendel J and Fluhr R (2000). Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J. Mol. Evol.* 3:203-213.

Parker JE, Coleman MJ, Dean C and Jones JGD (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.

Piffanelli P, Zhou F, Casais C, Orme J et al. (2002). The barley *Mlo* modulator of defense and cell death is responsive to biotic and abiotic stress stimuli. *Plant Physiol.* 129:1076-1085.

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

Rehmany AP, Gordon A, Rose LE, Allen RL et al. (2005). Differential recognition of highly divergent downy mildew avirulence gene alleles by *RPP1* resistance genes from two *Arabidopsis* lines. *Plant Cell* 17:1839-50.

Reinhold-Hurek B and Hurek T (1998). Life in grasses: diazotrophic endophytes. *Trends Microbiol.* 6:139-144.

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

Romeis T (2001). Protein kinases in the plant defense response. *Curr. Opin. Plant Biol.* 4:407-414.

Rommens CM and Kishore GM (2000). Exploiting the full potential of disease resistance genes for agricultural use. *Curr. Opin. Biotechnol.* 11:120-125.

Rose LE, Langley CH, Bernal AJ and Michelmore RW (2005). Natural Variation in the *Pto* Pathogen Resistance Gene Within Species of Wild Tomato (*Lycopersicon*). I. Functional Analysis of *Pto* Alleles. *Genetics* 171: 345-357.

Rossi M, Araujo PG, Paulet F and Garsmeur O (2003). Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. *Mol. Gen. Genomics* 269:406-419.

Salmeron JM, Oldroyd GED, Romens CMT, Scofield SR et al. (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.

Salvaudon L, Heraldet 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.

Sevilla M, Burris RH, Gunapala N and Kennedy C (2001). Comparison of benefit to sugarcane plant growth a $^{15}N_2$ incorporation following inoculation of sterile plants with *Gluconacetobacter diazotrophicus* wild-type and Nif-mutant strains. *Mol. Plant-Microbe Interact.* 14:358-366.

Somssich IE and Hahlbrock K (1998). Pathogen defence in plants – a paradigm of biological complexity. *Trends Plant Sci.* 3:86-90.

Song WY, Wang GL, Kim HS, Pi LY et al. (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 et al. (1997). Evolution of the rice *Xa21* disease resistance genes family. *Plant Cell* 9:1279-1287.

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

The Arabidopsis Genome Iniciative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815.

The Rice Chromosomes 11 and 12 Sequencing Consortia (2005). The sequence of rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. *BMC Biol.* 3:20.

Urquiaga S, Cruz HS and Boddey RM (1992). Contribution of nitrogen fixation to sugarcane: nitrogen-15 and nitrogen balance estimates. *Soil Sc. Soc. Am. J.* 56:105-114.

Vallad G, Rivkin M, Vallejos C and McClean P (2001). Cloning and homology modeling of a Pto-like protein kinase family of common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 103:1046-1058.

Van der Biezen EA, Freddie CT, Kahn K, Parker JE et al. (2002). *Arabidopsis RPP4* is a member of the *RPP5* multigene family of TIR-NB-LRR genes and confers downy mildew resistance through multiple signaling components. *Plant J.* 29:439-51.

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

Wang GL, Holsten TE, Song WY, Wang HP et al. (1996). Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa21* disease resistance locus. *Plant J.* 7:525-533.

Wang ZX, Yano M, Yamanouchi U, Iwamoto M et al. (1999). The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. *Plant J.* 19:55-64.

Whitham S, Dinesh-Kumar SP, Choi D, Heh R et al. (1994). The product of Tobacco mosaic virus resistance gene *N*: Similarity to Toll and the interleukin-1 receptor. *Cell* 78:1101-1115.

Whithan S, McCormick S and Baker B (1996). The *N* gene of tobacco confers resistance to tobacco mosaic virus in transgenic tomato. *Proc. Natl. Acad. Sci. U.S.A.* 93:8776-81.

Winter P, Benko-Iseppon AM, Hüttel B, Pfaff T et al. (2000). A linkage map of the chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *C. arietinum* x *C. reticulatum* cross: localization of resistance genes for fusarium wilt races 4 and 5. *Theor. Appl. Gen.* 101:1155-1163.

Xu WH, Wang YS, Liu GZ, Chen X et al. (2006). The autophosphorylated Ser686, Thr688, and Ser689 residues in the intracellular juxtamembrane domain of *Xa21* are implicated in stability control of rice receptor-like kinase. *Plant J.* 45:740-751.

Yoshimura S, Yamanouchi U, Katayose Y, Toki S et al. (1998). Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *Proc. Natl. Acad. Sci. U.S.A.* 95:1663-1668.

Zhou T, Wang Y, Chen JQ, Araki H et al. (2004). Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol. Genet. Genomics* 271: 402–415.

INTERNET RESOURCES

Expert Protein Analysis System (ExPASy), http://expasy.org (May 25, 2006).

National Center For Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov (June, 7, 2006)

Sugarcane EST Genome Project (SUCEST), http://sucest.lad.dcc.unicamp.br (May 2, 2006).

União dos Produtores de Bioenergia (UDOP), http://www.udop.com.br (February 12, 2007).

Table 1. Main sugarcane clusters similar to known R-genes. tBLASTn results and sequence evaluation of sugarcane *R*-genes including the five best matches of each *R*-gene class: (I) Features and evaluation results with sugarcane cluster number, cluster size in nucleotides (n), ORF (Open Reading Frame) size in amino-acids (aa), e-value; score and frame. (II) Data about BLASTx best alignment: NCBI gi|-number and plant species. Some clusters matched significant alignments with genes belonging to class III but presented no TIR domain.

(I) Cluster Features and Evaluation								(II) BLASTx Information	
Gene class and expected domain	Gene name	Sugarcane Cluster Nr.	Size (n)	ORF (aa)	E-value	Score and Frame		NCBI gi|-Nr.	Plant Species
Class I kinase	*Pto*	SCEZLB1010F11.g	1350	450	e^{-112}	402	+2	46981335	*Oryza sativa*
	Pto	SCCCLR1072H06.g	1788	596	e^{-100}	360	-1	34909896	*Oryza sativa*
	Pto	SCCCLR1024A02.g	984	328	6.0 e^{-99}	357	-3	34909896	*Oryza sativa*
	Pto	SCSGHR1066F10.g	759	253	6.0 e^{-62}	234	+1	37536416	*Oryza sativa*
	Pto	SCCCRZ1001B02.g	1926	642	7.0 e^{-61}	230	-3	15222211	*Arabidopsis thaliana*
Class II LRR-NBS-CC and Class III LRR-NBS-TIR	*Rps5, RPP8, RPM1, Hrt, R1, BS2,Gpa2, RX2*	SCSFSB1103A12.g	1830	610	2.0 e^{-74}	277	-1	4519936	*Oryza sativa*
	Xa1, Rpp8, RPM1, Hero, I2,Sw5, R1,RP1, Bs2, Rx2	SCEPAM2014D05.g	1725	575	3.0 e^{-58}	224	+1	27542759	*Sorghum bicolor*
	RPP8, RPM1, Hero, Pib, Sw5 ,RPP13, Hrt, Bs2, GPA2, RX2	SCRUFL1024D04.g	1050	350	4.0 e^{-41}	166	+1	40253400	*Oryza sativa*
	RPP8, RPM1, Pib, Sw5, Pi-ta, R1 GPA2	SCCCCL3080A08.g	762	254	8.0 $^{-35}$	145	+1	22208466	*Sorghum bicolor*
	RPP8, RPM1, Hero, I2, Pib, Sw5, Hrt, Bs2, RP1	SCJLRZ1019B10.g	1107	369	4.0 e^{-33}	140	-2	22652528	*Oryza sativa*
Class IV LRR	*Cf5,Cf9*	SCJLRT1020F05.g	1842	614	5.0 e^{-68}	256	+1	50934027	*Oryza sativa*
	Cf4,Cf5,Cf9	SCCCLR1001A03.g	2976	992	1.0 e^{-58}	224	+3	18390097	*Sorghum bicolor*
	Cf5,Cf9	SCJLRT1020F05.g	1842	614	5.0 e^{-49}	192	+1	50934027	*Oryza sativa*
	Cf5	SCCCLR1001A03.g	2976	992	7.0 e^{-48}	189	+3	18390097	*Sorghum bicolor*
	Cf4,Cf5,Cf9	SCJLRT1020F05.g	1842	614	3.0 e^{-45}	180	+1	50934027	*Oryza sativa*
Class V LRR KIN	*Xa21*	SCCCLR1001A03.g	2970	990	1.0 e^{-93}	341	+1	18390097	*Sorghum bicolor*
	Xa21	SCSGLR1045B05.g	1173	391	3.0 e^{-70}	263	+3	34915440	*Oryza sativa*
	Xa21	SCVPRT2081F09.g	798	266	1.0 e^{-66}	351	-1	7434424	*Oryza sativa*
	Xa21	SCCCRZ2C01E06.g	2184	728	1.0 e^{-65}	248	+2	34898088	*Oryza sativa*
	Xa21	SCQGAM2108A06.g	708	236	7.0 e^{-61}	232	+2	34915440	*Oryza sativa*

Table 2 - Comprehensive inventory of the organisms identified in Genbank as best alignment to each of the 196 identified sugarcane clusters related to known resistance genes. The organisms are grouped by gene class (I to V), taxonomic affiliation (class, family and species). Numbers in parenthesis indicate amount of members in each taxonomic group or species.

Gene Class	Higher Taxonomic Affiliation	Family	Species
I KINASE	Monocots (70)	Poaceae (70)	*Oryza sativa* (64)
			Zea mays (2)
			Sorghum bicolor (4)
	Dicots (22)	Brassicaceae (19)	*Arabidopsis thaliana* (19)
		Leguminosae (1)	*Glycine Max* (1)
		Solanaceae (2)	*Nicotiana tabaccum* (2)
II and **III** CC/TIR LRR-NBS	Monocots (60)	Poaceae (60)	*Oryza sativa* (41)
			Zea mays (5)
			Sorghum bicolor (6)
			Hordeum vulgare (2)
			Saccharum hybrid (1)
			Triticum aestivum (2)
			Triticum monococcum (3)
	Dicots (2)	Brassicaceae (2)	*Arabidopsis thaliana* (2)
IV LRR	Monocots (25)	Poaceae (25)	*Oryza sativa* (22)
			Zea mays (1)
			Sorghum bicolor (2)
	Dicots (2)	Brassicaceae (2)	*Arabidopsis thaliana* (2)
V LRR-KINASE	Monocots (14)	Poaceae (14)	*Oryza sativa* (10)
			Oryza longistaminata (2)
			Sorghum bicolor (2)
	Dicots (1)	Apiaceae (1)	*Daucus carota* (1)
			Class / Number
Grouped by taxonomic affiliation			Dicots 27
			Monocots 169

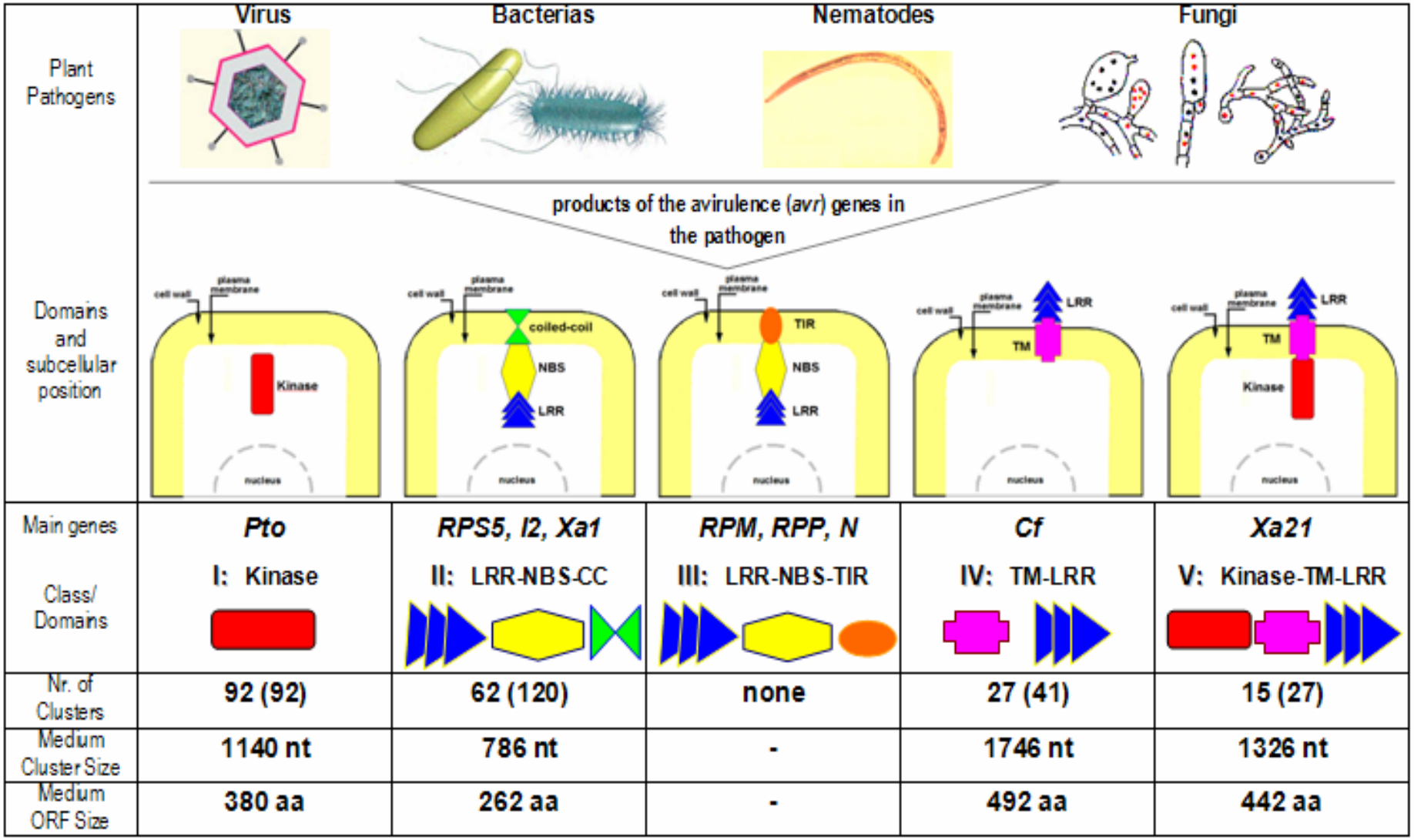


Figure 1 - Representation of main *R*-genes classes considering the presence of conserved domains, from SUCEST database. For each class domains are represented with their putative subcellular position, as well as number of clusters, size range of sequence in nucleotides (nt) and of ORF in amino-acids (aa). Given cluster numbers were selected using the e^{-20} cut-off for tBLASTn alignments, while numbers between parentheses indicate clusters obtained using a less stringent (e^{-10}) cut-off with the same tool. Abbreviations: **LRR:** Leucine-Rich-Reats; **NBS**: Nucleotide-Binding-Site. **Class I:** KINASE; **Class II:** CC-NBS-LRR; **Class III**: TIR-NBS-LRR; **Class IV:** TM-LRR; **Class V:** LRR-TM-KINASE.

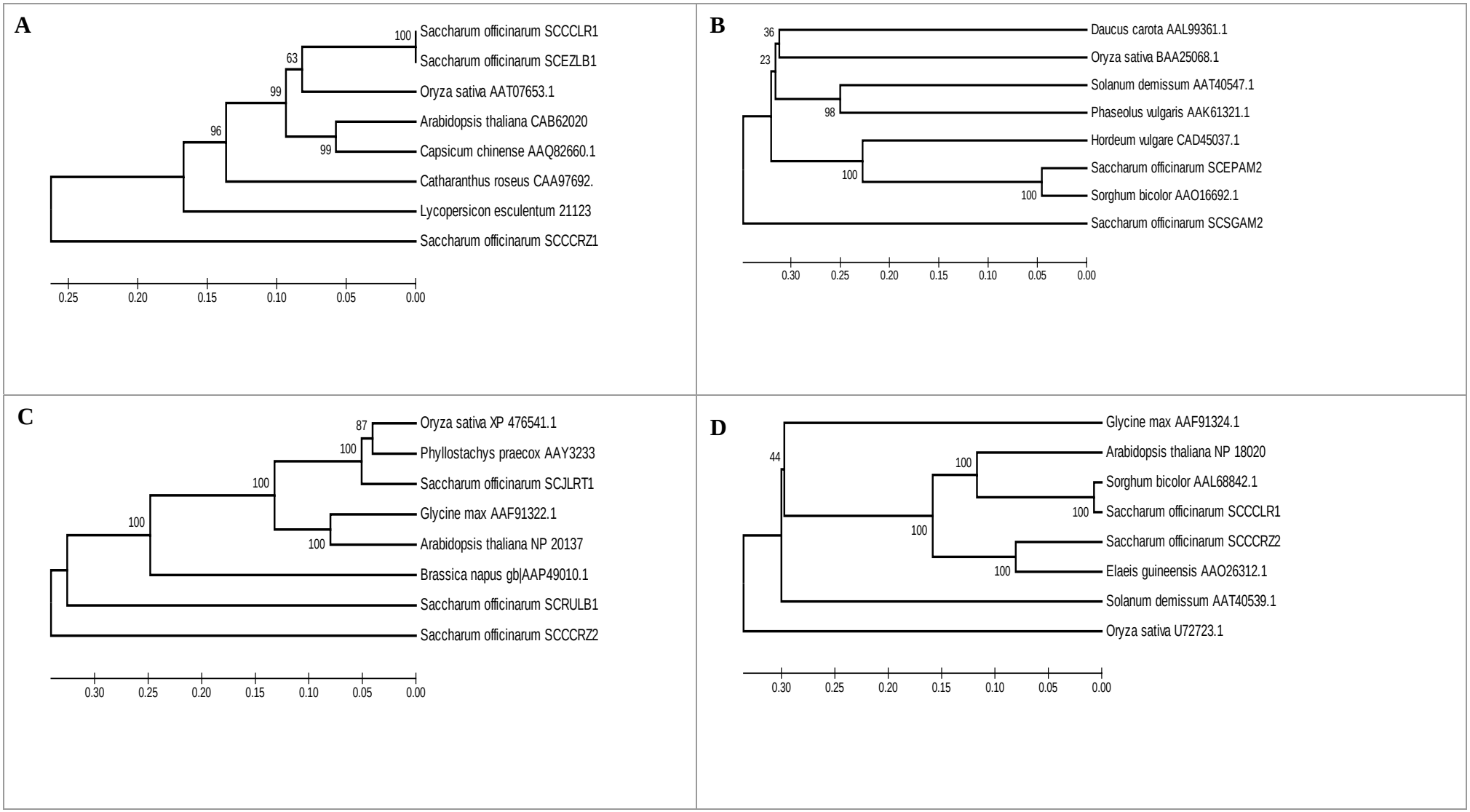


Figure 2. UPGMA analysis of selected dendrograms generated by MEGA 3 program using bootstrap function (1000 replicates). Numbers in the base of clades mean bootstrap values and bar means genetic distance. Codes after taxonomic identities refer to the accession number in NCBI or cluster identity in the case of sugarcane. **A**: *Pto* gene; **B:** *Xa1* gene; **C:** *Cf9* gene and **D:** *Xa21* gene.

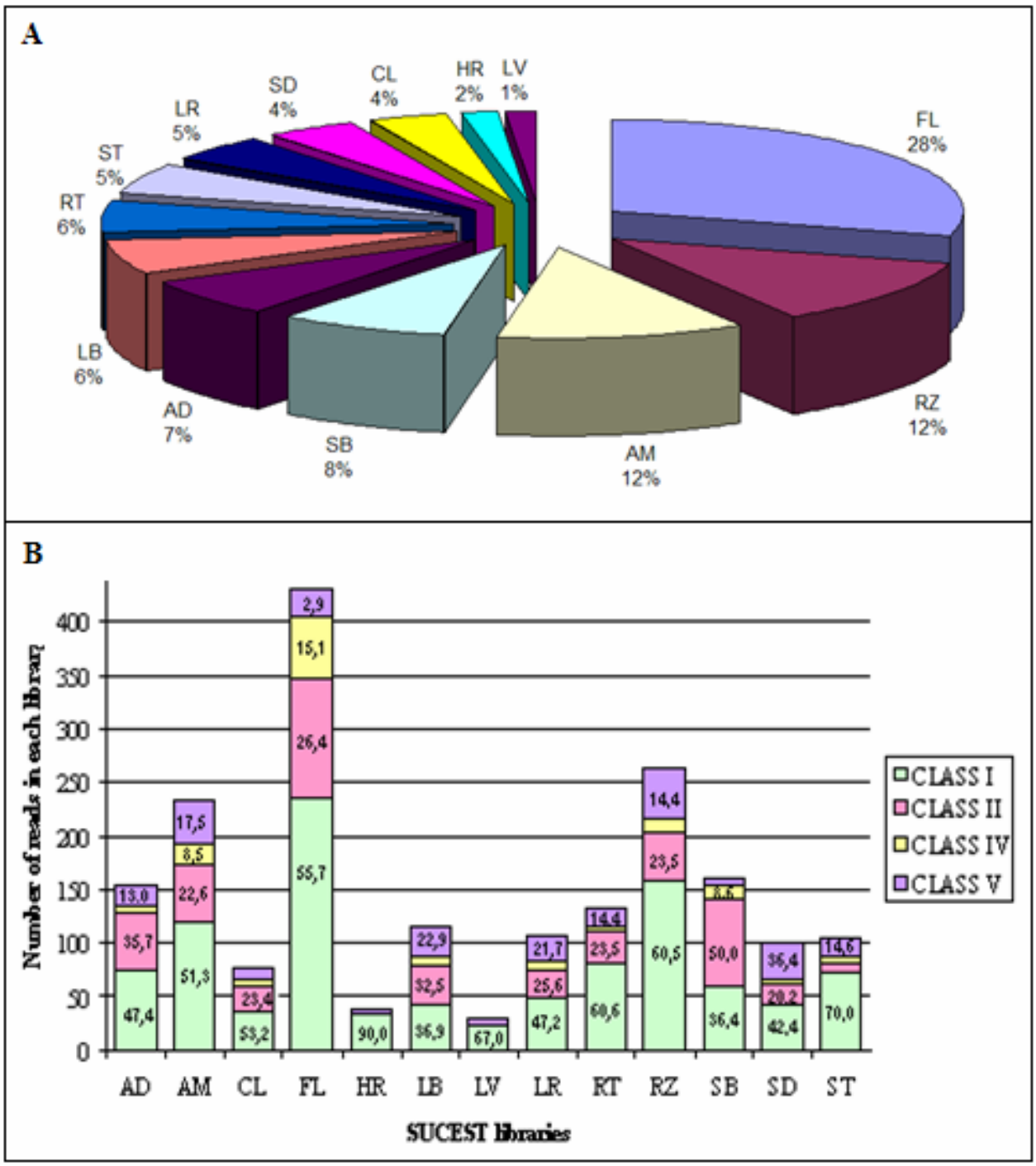


Figure 3. Prevalence of sugarcane *R*-genes in the SUCEST libraries. (**A**) General distribution of *R*-gene clusters in the SUCEST libraries; (**B**) prevalence of gene classes in the libraries. Numbers inside columns refer to the percentage of reads in each library for each *R*-gene class. Library code: **AD:** tissues infected by *Gluconacetobacter diazotroficans*, **AM:** Apical meristem; **CL:** Callus; **FL:** Flower; **HR:** tissues infected with *Herbaspirillum rubrisubalbicans*; **LB:**Lateral Bud; **LR:** Leaf Roll ; **LV:** Leaves; **RT**: Root; **RZ:** Stem-Root transition; **SB:** Stalk Bark; **SD:** Seeds;**ST:**Stem.

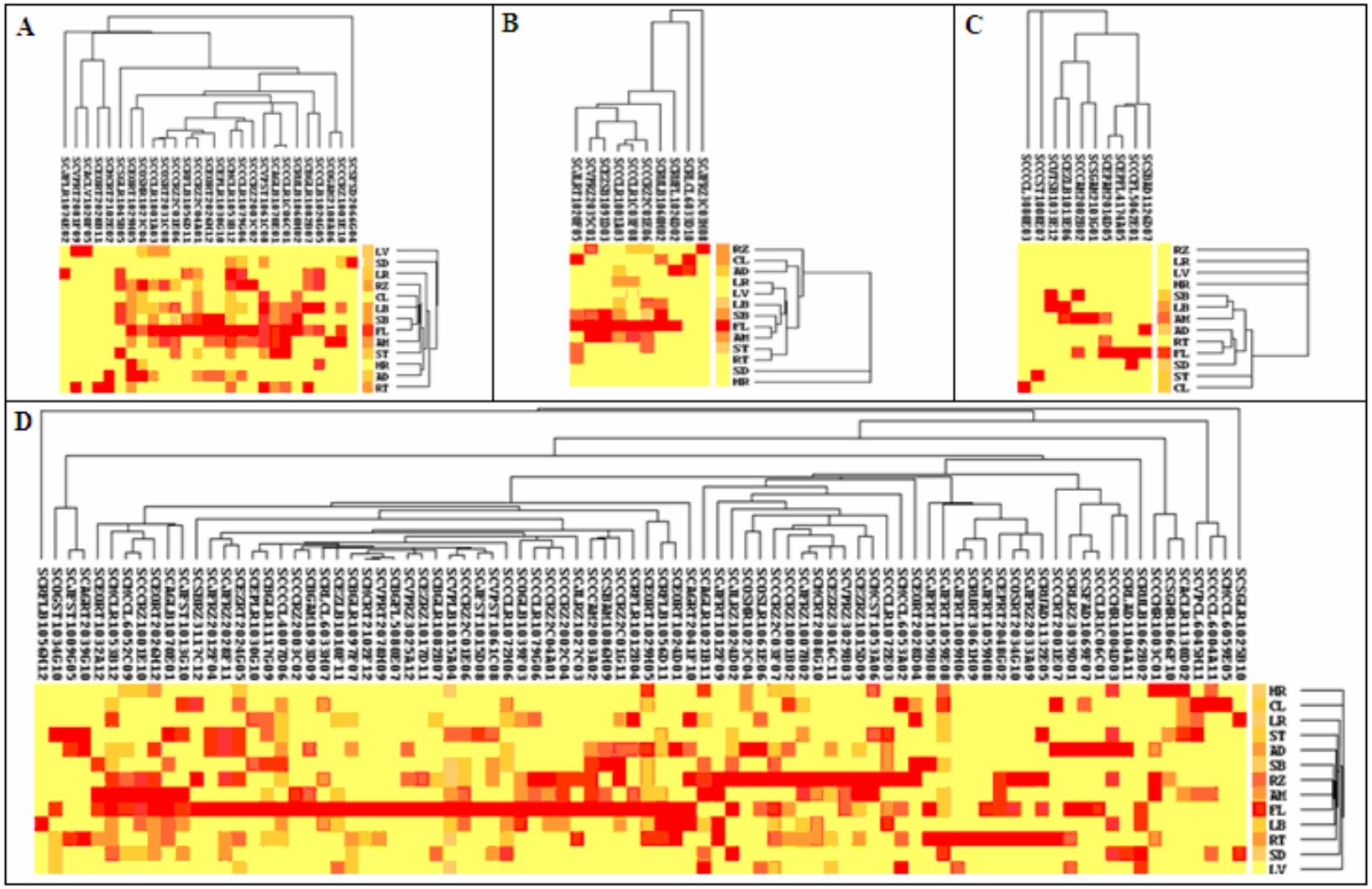


Figure 4 - Differential display of sugarcane clusters representing selected *R*- gene classes. **A:** Graphic representing of expression in *Xa1;* **B:** *Cf9*; **C:** *Xa21* and **D:** *Pto*. Yellow means no expression and red means all levels of expression. Library codes: **AD:** tissues infected by *Gluconacetobacter diazotroficans*, **AM:** Apical meristem; **CL:** Callus; **FL:** Flower; **HR**: tissues infected with *Herbaspirillum rubrisubalbicans*; **LB:** Lateral Bud; **LR:** Leaf Roll ; **LV:** Leaves; **RT**: Root; **RZ:** Stem-Root transition; **SB:** Stalk Bark; **SD:** Seeds; **ST:** Stem.

5.0- ANEXO
5.1- Artigo publicado na revista *Genetics and Molecular Biology*

Genetics and Molecular Biology, 28, 3 (suppl), 562-574 (2005)

www.sbg.org.br

Research Article

In silico survey of resistance (*R*) genes in *Eucalyptus* transcriptome

Adriano Barbosa-da-Silva, Ana C. Wanderley-Nogueira, Raphaela R.M. Silva, Luiz C. Berlarmino, Nina M. Soares-Cavalcanti and Ana M. Benko-Iseppon

Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Departamento de Genética, Laboratório de Genética e Biotecnologia Vegetal, Recife, PE, Brazil.

Abstract

A major goal of plant genome research is to recognize genes responsible for important traits. Resistance genes are among the most important gene classes for plant breeding purposes being responsible for the specific immune response including pathogen recognition, and activation of plant defence mechanisms. These genes are quite abundant in higher plants, with 210 clusters found in *Eucalyptus* FOREST database presenting significant homology to known R-genes. All five gene classes of *R*-genes with their respective conserved domains are present and expressed in *Eucalyptus*. Most clusters identified (93) belong to the LRR-NBS-TIR (genes with three domains: Leucine-rich-repeat, Nucleotide-binding-site and Toll interleucine 1-receptor), followed by the serine-threonine-kinase class (49 clusters). Some new combinations of domains and motifs of *R*-genes may be present in *Eucalyptus* and could represent novel gene structures. Most alignments occurred with dicots (94.3%), with emphasis on *Arabidopsis thaliana* (Brassicaceae) sequences. All best alignments with monocots (5.2%) occurred with rice (*Oryza sativa*) sequences and a single cluster aligned with the gymnosperm *Pinus sylvestris* (0.5%). The results are discussed and compared with available data from other crops and may bring useful evidences for the understanding of defense mechanisms in *Eucalyptus* and other crop species.

Key words: serine-threonine kinase, nucleotide binding site, leucine-rich repeats, gene-for-gene interaction.

Received: May 28, 2004; Accepted: March 28, 2005.

Introduction

Pathogen attack can severely affect crop production, with losses that can achieve 80% of the production especially in tropical countries. At the global level, losses have been estimated to accomplish around 12% of the world crop production (James *et al.*, 1990). The most important group of genes that has been used by breeders for disease control is the plant resistance (*R*) genes: single determinant of an effective and specific resistance that can often be characterized by localized necrosis at attempted infection sites (Rommens and Kishore, 2000).

It is proposed that pathosystems are usually highly specific, with a matching *R*-gene on vegetal cell that recognizes elicitor proteins (called Avr-effector) of each infective pathogen. Plant will be resistant and the growth of the pathogen will be arrested only when both genes, *R* and *Avr*, are present (Ellis *et al.*, 2000a). So, for each *R*-gene a correspondent *Avr* gene co-exists: this is the basis of the gene-for-gene concept, suggested by Flor (1956, 1971).

Send correspondence to Ana Maria Benko-Iseppon, Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Departamento de Genética, Laboratório de Genética e Biotecnologia Vegetal, Av. Prof. Moraes Rego s/ n., 50732-970 Recife, PE, Brazil. E-mail: celisep@hotlink.com.br.

Avirulence gene products actually described do not comprise a defined family of related proteins, since no sharing similar motifs or domains could be found. On the opposite, *R*-gene products are separated into distinct but related protein classes, according to their conserved structural domains. Conserved domain function identified for R proteins suggests two fundamental mechanisms during pathogenic infection: (I) the pathogen recognition, conducted mainly by leucine-rich repeats (LRR) regions, which play a direct role in protein-protein specific recognition event; and (II) signaling of pathogen presence in order to activate defense related genes (Richter and Ronald, 2000).

The TIR (Toll interleukine 1-receptor) and CC (coiled coil) regions are involved in signal transduction during many cell processes (Martin *et al.*, 2003), while the NBS (Nucleotide Biding Site) usually signalizes for programmed cell death in animal cells (van der Biezen and Jones, 1998). Additionally, a kinase catalytic region is present in some *R*-genes. This domain plays a direct role in both signaling processes and pathogen effectors. Additionally the NBS region contains not only the three motifs involved in nucleotide binding but additional motifs as well. This extended region of homology is referred to as the NB-ARC domain (Richter and Ronald, 2000). Sometimes this do-

main contains a distinct predicted nucleoside triphosphatase (NTPase) domain known as NACHT, common in animal, fungal and bacterial proteins, implicated with apoptosis induction and transcription activation (Koonin and Avarind, 2000).

Resistance genes are members of a very large multigene family, are highly polymorphic and have diverse recognition specificities. They are commonly clustered in the genome, often in tandem direct repeats, what is consistent with the theory that they originated through gene duplication and that they are continuously evolving through unequal exchange (Song *et al.*, 1997).

Most of the resistance genes that have been cloned and characterized resemble components involved in signal transduction. These can be classified into five categories based on their predicted protein structure (Song *et al.*, 1997, Ellis and Jones, 1998).

The first class is represented by the *Pto* gene of tomato, which encodes a protein with a catalytic serine-threonine kinase (ser-thre-kinase) and a myristoylation motif in his amino terminal region (Martin *et al.*, 1993).

The second class comprises many proteins that present a region rich in repetitions of leucine (LRR, Leucine-rich repeats), a Nucleotide Binding Site (NBS) and a leucine zipper (LZ) or a coiled-coil (CC) sequence. Many genes encode proteins of this class: *I2* (Ori *et al.*, 1997), *Mi* (Milligan *et al.*, 1998) and *Sw5* (Brommonschenkel *et al.*, 2000) from tomato; *RPM1* (Grant *et al.*, 1995), *RPP8* (McDowell *et al.*, 1998), *RPS2* (Mindrinos *et al.*, 1994) and *RPP13* (Bittner-Eddy *et al.*, 2000) from *Arabidopsis thaliana*; *Pib* (Wang *et al.*, 1999), *Pi-ta* (Bryan *et al.*, 2000) and *Xa1* (Yoshimura *et al.*, 1998) from *Oryza sativa* (rice); *Gpa2* (Van der Vossen *et al.*, 2000), *Hero* (Ernst *et al.*, 2002), *R1* (Ballvora *et al.*, 2002), *Rx1a* (Bendahmane *et al.*, 1995) and *Rx2* (Bendahmane *et al.*, 2000) from potato; *Rp1* from maize (Collins *et al.*, 1999); *Mla* from barley (Halterman *et al.*, 2001) and *Dm3* from lettuce (Meyers *et al.*, 1998).

The third class includes similar proteins as described for class II, presenting a toll receptor for interleukine-I (IL-1R) instead of a CC sequence at the amino terminal region (Meyers *et al.*, 1999). This class is referred as TIR-NBS-LRR, including the genes *L* (Lawrence *et al.*, 1995), and *P* (Dodds *et al.*, 2001) of flax; *RPP1* (Botela *et al.*, 1998), *RPP4* (van der Biezen *et al.*, 2002), *RPP5* (Parker *et al.*, 1997) and *RPS4* (Gassmann *et al.*, 1999) of *A. thaliana* and *N* (Whithan *et al.*, 1996) of tobacco. This class (also present in animals) is supposed to be absent in monocotyledonous plants (Ellis and Jones, 1998), being present in all dicotyledonous taxa actually studied.

The proteins encoded by the three classes of genes previously cited do not present a transmembrane sequence and are therefore classified as intracellular *R*-proteins (Martin *et al.*, 2003).

The fourth class of resistance genes belongs to the tomato *Cf*-family, encoding similar proteins with an extracellular LRR and a short cytoplasmatic tail, but no NBS or any further recognizable domain (Dixon *et al.*, 1996). Member of this family are *Cf-2* (Dixon *et al.*, 1998), *Cf-4* (Joosten *et al.*, 1994; Thomas *et al.*, 1997), *Cf-5* (Dixon *et al.*, 1998) and *Cf-9* (Jones *et al.*, 1994).

The fifth class includes a single gene, *Xa21* from rice that presents an extracellular LRR, a transmembrane region (TM) and a cytoplasmatic ser-thre-kinase. Thus, the structure of *Xa21* indicates an evolutionary link between different classes of plant disease resistance genes (Song *et al.*, 1997).

There is still a sixth class that presents genes with no conserved domains, as described for the previous five classes. This group comprises the gene *Hm1* from maize, a reductase that confers resistance to the fungus *Cochliobolus carbonum* (Johal and Briggs, 1992); *Mlo* from barley, a putative regulator of defense against *Blumenaria graminis* (Piffanelli *et al.*, 2002) possibly associated to the plasma membrane (Buschges *et al.*, 1997); and *RPW8* from *A. thaliana*, that confers non-specific resistance to the fungus *Erysyphe chicoracearum* (Xiao *et al.*, 2001).

Due to its qualities as high level of adaptability, fast growing capacity and wood quality, *Eucalyptus* plantations are carried out in all tropical areas in diverse continents. *Eucalyptus* is the most widely used tree for delivering raw material for the paper industry used in the production of cellulose and to regenerate degraded areas. Over the past 50 years large-scale planting of fast growing exotic *E. grandis*, *E. urophyla*, *E. saligna* and many hybrids (particularly *grandis* x *urophyla*) has occurred in Brazil aiming to reforest some regions and to create an adequate supply of wood, timber and fuel for different purposes (McNabb, 2002). In the late 2001s growing areas reached 138.132 ha, generating more than 7,398 direct employments (BRACELPA, 2004).

The advance of plantations to hot and humid areas resulted in favourable conditions to the development of diseases especially in young individuals that are often severely attacked by fungal (*e.g. Mycosphaerella cryptica*, *Dichomera versiformis*, *Cylindrocladium* spp. and *Phaeophleospora epicpccoides*) and bacterial pathogens (Barber *et al.*, 2003, Mafia and Alfenas, 2003).

Eucalyptus Genome Sequencing Consortium (FOREST) aimed to identify over 15,000 expressed genes from 100,000 sequenced EST from 19 libraries from specific tissues and stages.

The present work aimed to perform a data mining-based identification of plant disease *R*-genes in FOREST database, by using well known *R*-genes sequences as template, comparing the identified sequences with known *R*-genes deposited in public DNA and protein databases.

Materials and Methods

Amino-acid sequences of known genes have been used as query in the search for *R*-gene homologues and analogs in *Eucalyptus* transcriptome database. Accession numbers at NCBI (National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov) of sequences used are shown in Table 1, together with sequences features and accession numbers. They are grouped according to the conserved domains previously described. Members of the sixth class (reductases and other *R*-genes with no recognizable conserved domains) have not been included in the present evaluation.

All *Eucalyptus* sequences used during this work were obtained from FOREST project and derived from cDNA libraries specific to different tissues, organs or conditions of growth from the species *E. grandis*, *E. globulus*, *E. saligna* and *E. urophylla*. For detailed information see https://forests.esalq.usp.br/Librariesinfo.html.

Reverse alignments were realized on 'FOREST EG_Clusters' database using the program TBLASTN (Altschul *et al.*, 1990), the e-value cutoff adopted was $1e^{-23}$. Matching clusters to query sequences were then annotated on a local database called 'non-redundant' made with aid of the Microsoft Access® program. Cluster name was adopted as primary key in order to prevent data redundancy regarding clusters aligning with more than one query sequence. In the few cases when this occurred the name of both queries has been also annotated for the respective cluster.

The clusters frame of the TBLASTN alignment was used to predict the Open Reading Frames (ORFs) for each searched cluster. For this purpose, the Expasy Translate Tool (bo.expasy.org/tools/dna.html) was used, which predicts the correct ORF for a DNA sequence in the corresponding amino acid FASTA sequence. The obtained ORFs were subsequently submitted to a Reverse Position Specific BLAST (RPS-BLAST) against Conserved Domain Database (Marchler-Bauer *et al.*, 2002) aiming to identify patterns or motifs in predicted cluster products.

Reciprocal alignments were conducted for ORFs by downloading the nr databank and stand alone BLAST package from NCBI ftp site for local use at our server (Laboratório de Genética e Biotecnologia Vegetal, UFPE) performing a high-throughput alignment approach. Matched sequences were annotated for latter comparison.

Predictions of subcellular localization have been inferred by using TargetP program available at CBS (Center for Biotechnology Sequence Analysis) Prediction Servers

Table 1 - Classification and features of *R*-genes used as query against the FOREST database. The used genes are grouped in five *R*-gene classes (**I**: Kinase; **II**: LRR+NBS; **III**: LRR+NBS+TIR; **IV**: only LRR; **V**: LRR+Kinase) with respective accession number at NCBI, source species, gene name and domain range (in amino-acids).

Class of *R*-gene	Accession number	Source species	Gene name	Sequence size	Domain range (initial-last aa)							
					LRR		Kinase		NBS		TIR	
				(aa)	Start	End	Start	End	Start	End	Start	End
I	2112354A	*Lycopersicon esculentum*	*Pto*	321	-	-	41	236	-	-	-	-
	AF234174_1	*Arabidopsis thaliana*	*HRT*	909	579	868	-	-	150	460	-	-
	NP_172686.1	*Arabidopsis thaliana*	*Rps5*	889	540	636	-	-	140	444	-	-
II	AF118127_1	*Lycopersicon esculentum*	*I2*	1266	578	1231	-	-	154	457	-	-
	AAG31014.1	*Lycopersicon esculentum*	*Sw5*	1246	-	-	-	-	519	818		
	BAA25068.1	*Oryza sativa*	*Xa1*	1802	771	1773	-	-	283	593	-	-
	AAP81262.1	*Zea mays*	*Rp1*	1269	596	1228	-	-	145	457	-	-
	AAC72977.1	*Arabidopsis thaliana*	*RPP1*	1189	668	1011	-	-	226	505	54	184
	RP13_ARATH	*Arabidopsis thaliana*	*RPP13*	835	-	-	-	-	147	453	14	148
	AF440696_1	*Arabidopsis thaliana*	*RPP4*	1135	642	1053	-	-	185	441	15	145
	AAF08790.1	*Arabidopsis thaliana*	*RPP5*	1361	643	1151	-	-	188	465	14	148
	RPP8_ARATH	*Arabidopsis thaliana*	*RPP8*	908	577	867	-	-	149	459	15	145
	BAB11393.1	*Arabidopsis thaliana*	*Rps4*	1232	663	889	-	-	198	473	21	149
III	AAP41025.1	*Lactuca serriola*	*RGC2*	352	49	235	-	-	-	-	21	149
	AF093649_1	*Linum usitatissimum*	*L*	1294	607	1277	-	-	220	521	63	195
	T18548	*Linum usitatissimum*	*M*	1305	744	1288	-	-	235	534	78	210
	AF310960_2	*Linum usitatissimum*	*P*	1211	693	1023	-	-	205	238	23	153
	AF202179_1	*Capsicum chacoense*	*Bs2*	905			-	-	152	439	63	195
	A54810	*Nicotiana glutinosa*	*N*	1144	597	908	-	-	172	447	14	147
	AF195939_1	*Solanum tuberosum*	*Gpa2*	912	561	863	-	-	119	422	14	147
	CAA61264.1	*Solanum tuberosum*	*Rx1*	248			-	-	-	-	23	153
	CAB56299.1	*Solanum tuberosum*	*Rx2*	938	561	859	-	-	138	422	78	210
	CAD29728.1	*Solanum tuberosum*	*HERO*	1283	-	-	-	-	504	811	54	184
IV	T07015	*Lycopersicon esculentum*	*Cf4*	855	81	758	-	-	-		-	-
	AAC78591.1	*Lycopersicon esculentum*	*Cf5*	968	96	855	-	-	-	-	-	-

site (http://www.cbs.dtu.dk/services/). Additionally, transmembrane helix segments were inferred with aid of the TMHMM program as well.

Results

After the TBLASTN alignments performed at FOREST EG_Clusters database, a total of 478 clusters aligned with the diverse *R*-genes (Table 1) used as query (data not showed). These clusters were, as described in section 'Material and Methods', inserted on a local database called 'non-redundant'. This procedure generated a set of 210 non-redundant clusters which have been annotated for one or more than one *R*-gene (data summarized in Figure 1 and Tables 2 and 3).

Clusters representing exclusive *R*-gene classes were: (I) serine-threonine kinase (here named KINASE): 49; (II) LRR+NBS: 21; (III) LRR+NBS+TIR: 93; (IV) Only LRR + Transmembrane (LRR+TM): 17 and (V) LRR+TM+ Kinase: 8 (Figure 1).

Regarding the sequence identity of the best alignment, 22 clusters showed equally significant similarity to two different classes of *R*-genes. From these, 18 included LRR plus LRR-Kinase here called MIX I (sequence data presented in Table 3); three included NBS-LRR plus TIR-NBS-LRR (called MIX II) and one LRR plus Kinase (called MIX III).

Sizes of *Eucalyptus* clusters aligned to *R*-genes varied from 3,316 (cluster EGEQRT3301C03 classified to group MIX-III) to 520 nucleotides. The prediction of clusters coding regions revealed that ORFs were coded in both forward and reverse reading frames, with an average of 304 amino acids (aa) in length. ORF sizes varied from 990 (cluster EGEQRT3301C03 of the LRR-KINASE class) to 134aa. Regarding the average ORF length in each *R*-gene class, we observed 417aa for KINASE, 276aa for NBS, 238aa for TIR-NBS-LRR, 247aa for LRR-TM, 352aa for LRR-KINASE, 372aa for MIX I, 343aa for MIX II and 990aa for MIX III class.

The search for conserved domains (CD-Search) revealed conserved regions (Figure 1, Table 1) in 166 of the 210 here analyzed clusters. A total of 40 clusters presented the kinase domain, 37 of them matched to *Pto* gene (class I) after the TBLASTN alignment, with only three grouping into KINASE-LRR (two of them) and MIX III (one of them) classes. These two classes also showed associated LRR segments as well. Regarding the LRR domains, these could be identified in 67 different clusters in all classes (except KINASE class I, represented by *Pto*) with a total of 442 occurrences. This number is higher than the number of clusters due to their occurrence in tandem repetitions. Sometimes these sequences are imperfect and may be difficult to recognize with available *in silico* tools, so it is possible that a larger number may be identified manually.

Twenty clusters showed the NB-ARC domain. In a specific case, this domain occurred associated to a different TIR domain as was cited above. Additionally, a NACHT domain (closed-related to NB-ARC) was identified exclu-

Literature Data			Forest Database							
Known Features of *R*-Genes			Nr. of Clusters	Number (and %) per class of clusters bearing CD					Maximal Values	
Class	Domains Main Genes Reported	Gene Architecture		KIN	LRR	NBS	TIR	TM	Size (n)	ORF (aa)
I	KINASE *Pto*		49	49 (100)	-	-	-	20 (40.8)	2575 658	847 218
II	LRR+NBS *RPS5, I2, SW5, Rp1, Xa1*		21	-	7 (35)	9 (42.8)	-	-	1775 686	468 188
III	LRR+NBS+TIR *RPP1, RPS4, L, M, P, N*		93	-	11 (11.8)	16 (17.2)	39 (40.8)	5 (5.3)	2074 520	459 134
IV	LRR+TM *Cf-Family*		17	-	16 (94.1)	-	-	7 (41.1)	1361 630	338 151
V	LRR+TM+KINASE *Xa21*		08	2 (25)	8 (100)	-	-	5 (62,5)	1711 672	570 233
MIX I	LRR, LRR+KINASE *Cf-family plus Xa21*	**Classes IV and V**	18	-	18 (100)	-	-	6 (33,3)	713 2237	210 709
MIX II	LRR+NBS, LRR+NBS+TIR *I2, RPS5, RPS4, RPP5*	**Classes II and III**	03	-	1 (33,3)	-	-	-	2109 778	149 646
MIX III	LRR, KINASE *Cf9-family plus Pto*	**Classes I and IV**	01	1 (100)	1 (100)	-	-	1 (100)	3316	990

Legend for Conserved Domains

KINASE	LRR	NBS	TIR	TM
Serine-Threonine Kinase	Leucine-Rich-Repeats	Nucleotide-Binding Site	Toll-Interleucine-Region	Transmembrane Region

Figure 1 - Representation of main *R*-genes classes considering the presence and position of conserved domains from literature data, as compared with *Eucalyptus* clusters from FOREST database. For each class the data about significant alignments to *R*-genes is given, including following information: number of clusters identified for each class (clusters aligning with more than one class are not included), number and percentage of clusters per class bearing indicated conserved domains, size range (maximal and minimum) of sequence in nucleotides (n) and of ORF in amino-acids (aa). Abbreviation: CD – Conserved domains.

Table 2 - Blast results and sequence evaluation of *Eucalyptus* R genes, including the best matches of each *R* gene and MIX classes: (I) data about the query: gene class and name, NCBI gi | -number, species and family. (II) Features and evaluation results of *Eucalyptus* clusters related to R-genes: cluster number, cluster size in nucleotides (n), ORF (Open Reading Frame) size in amino-acids (aa), e-value; score and frame.

(I) Query Information				(II) Cluster features and evaluation					
Gene class & expected domain	Gene name	NCBI gi | -nr.	Plant species and family	*Eucalyptus* cluster n.	Size (n)	ORF (aa)	E-value	Score and frame	
Class I KINASE	*Pto*	27754635	*Arabidopsis thaliana*	EGEQRT3100D07	2460	722	0.0	1036,6	2
	Pto	15235204	*Arabidopsis thaliana*	EGEQRT3104A12	2575	847	0.0	909,8	1
	Pto	18418211	*Arabidopsis thaliana*	EGUTFB1098H02	2511	616	0.0	904,0	3
	Pto	10177052	*Arabidopsis thaliana*	EGCBRT3133E11	1728	575	0.0	738,0	3
	Pto	25405628	*Arabidopsis thaliana*	EGMCRT3148C12	1705	568	0.0	709,9	1
Class IV LRR	*Cf5*	14626935	*Gossypium hirsutum*	EGEQSL5001G09	1223	321	$2.00e^{-139}$	496,1	2
	Cf5	15240263	*Arabidopsis thaliana*	EGCCRT3339F06	922	307	$3.8e^{-83}$	309,3	2
	Cf5	15239124	*Arabidopsis thaliana*	EGCBST2063A06	697	232	$2.0e^{-61}$	236,5	-2
	Cf4, Cf5	27754637	*Arabidopsis thaliana*	EGACRT3321G06	1361	338	$1.4e^{-60}$	234,6	3
	Cf4, Cf9	14269077	*Lycopersicom esculentum*	EGJMCL1299H10	682	226	$1.9e^{-53}$	209,9	3
Class V LRR KIN	*Xa21*	9651941	*Glycine max*	EGRFRT3357D01	1584	527	0.0	869,4	3
	Xa21	15239540	*Arabidopsis thaliana*	EGEQCL1200B12	1711	570	0.0	658,7	1
	Xa21	19881587	*Oryza sativa*	EGJEST2023F09	716	238	$5.6e^{-51}$	201,8	1
	Xa21	15218385	*Arabidopsis thaliana*	EGSBCL1280C05	725	241	$7.8e^{-48}$	191,4	3
	Xa21	15218385	*Arabidopsis thaliana*	EGSBCL1280C05	725	241	$3,00e^{-04}$	191.4	3
Class II NBS LRR	*Hrt, I2, Sw5, Xa1, Rp1, R1*	18652501	*Oryza sativa*	EGUTRT3110A12	1041	346	$1.2e^{-62}$	241,5	3
	I2, Xa1, Rpm1, Rp1, R1, Pib, Mi1	28300299	*Manihot esculenta*	EGJFSL4202E08	876	291	$5.3e^{-55}$	215,7	3
	Bs2, Gpa2, I2, Rx1, Rx2, Sw5, Xa1, Rp1Mi1	15487949	*Theobroma cacao*	EGEQCL1001F08	934	311	$2.1e^{-52}$	207,2	1
	Gpa2, Rx2, Rpm1, R1, Pib, I2	28300299	*Manihot esculenta*	EGJERT3026C12	804	267	$7.7e^{-50}$	198,4	2
	Gpa2, Hrt, Rpp13, Rpp8, Rx2, Sw5, Rpm1R1, Pi-Ta, Pib	22775643	*Oryza sativa*	EGCECL1282E03	779	231	$1.1e^{-48}$	194,1	1
Class III TIR NBS LRR	*L, M, N, P, Rpp1, Rpp4, Rpp5, Rps4*	7488903	*L. usitatissimum*	EGJMFB1107C10	1395	445	$1.7e^{-83}$	311.2	1
	L, M, N, P, Rpp1, Rpp4, Rpp5, Rps4	9965103	*Glycine max*	EGMCLV2264D03	1155	329	$7.4e^{-75}$	282,0	3
	L, M, N, P, Rpp1, Rpp4, Rpp5, Rps4	12056928	*Glycine max*	EGJEST2234G10	1270	420	$8.8e^{-74}$	278,9	-1
	L, M, N, P, Rpp1, Rpp4, Rpp5, Rps4	27764536	*Glycine max*	EGJMST6019E06	1155	351	$3.0e^{-69}$	263,5	1
	L, M, N, P, Rpp1, Rpp5, Rps4	23477203	*Populus balsamifera*	EGCBRT6029A01	1227	378	$2.6e^{-66}$	253,8	1
MIX I (LRR and LRR-KIN)	*Cf4, Cf5, Cf9, Xa21*	25287710	*Arabidopsis thaliana*	EGBMRT3129F10	2129	709	0.0	705,3	2
	Cf4, Cf5, Cf9, Xa21	21391894	*Lycopersicon peruvianum*	EGCEST2256F04	1687	561	0.0	704,1	3
	Cf4, Cf5, Cf9, Xa21	15240215	*Arabidopsis thaliana*	EGUTFB1136E01	1911	595	$5.00e^{-172}$	605,9	2
	Cf4, Cf5, Cf9, Xa21	15240528	*Arabidopsis thaliana*	EGEZST2207A10	2237	465	$2.00e^{-143}$	510,8	1
	Cf5, Xa21	15230539	*Arabidopsis thaliana*	EGEQRT3201E07	1229	376	$4.00e^{-127}$	455,7	1
MIX II TIR-NBS-LRR and NBS-LRR	*Rpp5, Rps4, I2*	15218365	*Arabidopsis thaliana*	EGEZRT3006B12	2109	646	$1.1e^{-50}$	203,0	3
	Rpp5, Rps5	15221252	*Arabidopsis thaliana*	EGEQST6001H02	778	234	$8.9e^{-33}$	141,4	1
	Rpp5, Rps5	15487963	*Theobroma cacao*	EGJECL1208G03	871	149	$5.8e^{-24}$	110,5	1
MIX III KINASE and LRR	*Cf5, Cf9, Pto*	26450791	*Arabidopsis thaliana*	EGEQRT3301C03	3316	990	0.0	1293,9	1

sively in two TIR-NBS-LRR related clusters (EGCCCL1328B05.g and EGSBRT3118H01).

Most of the 44 clusters with no conserved domains presented shorter ORFs (262 aa in average), with four of them presenting a putative transmembrane region.

A graphic representation of the distribution of conserved domains as compared with class-grouped clusters is presented in Figure 2.

Considering the best matches to the 210 clusters identified, 198 were from plants of Dicotyledonous families, with emphasis on *A. thaliana*. From monocots only rice (*O. sativa*) sequences appeared as best matches (11 clusters). One of the sequences from MIX III group aligned with *Pinus silvestris* (Gymnosperm), the only non-Angiosperm included in the present study. A comprehensive inventory of all species that aligned with *Eucalyptus* with their taxonomic affiliation and habit (herbaceous or woody) is presented in Table 4.

The post-translational inferences carried out for cluster products (TargetP program) revealed a large number of

Table 3 - FOREST clusters classified in the MIX I group, resembling to genes which belong to LRR and LRR-KINASE classes, including: respective templates (query sequences), cluster number and size in nucleotides (n), ORF-size in amino-acids (aa), range of LRR domain after CD-search, identity and results of the best alignment (BLASTp) in NCBI (GI number, species, score and e-value).

Template	Cluster	Size (n)	ORF (aa)	LRR-domain		GI	Description	Score	E-value
				Start	End				
	EGBMRT3129F10.g	2129	709	139	620	25287710	*Arabidopsis thaliana*	705.3	0.0
	EGUTFB1136E01.g	1911	595	91	523	15240215	*Arabidopsis thaliana*	605.9	$5e^{-172}$
Cf4	EGEZST2207A10.g	2237	465	22	262	15240528	*Arabidopsis thaliana*	510.8	$2e^{-143}$
Cf5	EGEQST2201G12.g	2049	412	36	223	25402587	*Arabidopsis thaliana*	360.9	$1.7e^{-98}$
Cf9	EGUTRT3368G02.g	799	265	38	254	15225805	*Arabidopsis thaliana*	323.2	$2e^{-87}$
Xa21	EGCEST2256F04.g	1687	561	22	501	21391894	*Lycopersicon peruvianum*	704.1	0.0
	EGUTSL4018B05.g	1384	447	96	432	3894385	*Lycopersicon esculentum*	257.3	$3e^{-67}$
Cf5, Cf9	EGSBRT3314G03.g	1263	412	118	380	15223460	*Arabidopsis thaliana*	408.3	$9e^{-113}$
Xa21	EGBMRT3131G11.g	1155	384	2	336	15237312	*Arabidopsis thaliana*	349.4	$4.7e^{-95}$
	EGEQRT3201E07.g	1229	376	106	348	15230539	*Arabidopsis thaliana*	455.7	$4e^{-127}$
	EGABST2047C09.g	773	210	13	180	15225805	*Arabidopsis thaliana*	249.6	$1.9e^{-65}$
	EGBMSL4023G05.g	729	242	4	219	15237426	*Arabidopsis thaliana*	226.9	$1.7e^{-58}$
	EGCBST6013F02.g	808	265	32	254	18700171	*Arabidopsis thaliana*	214.5	$1.0e^{-54}$
Cf5	EGCESL5078H03.g	771	257	27	245	15237426	*Arabidopsis thaliana*	211.1	$1.1e^{-53}$
Xa21	EGBGLV3221H06.g	734	235	24	213	3894383	*Lycopersicon esculentum*	200.7	$1.2e^{-50}$
	EGCBRT6048F01.g	713	237	13	227	3641252	*Malus X domestica*	327.4	$8.9e^{-89}$
	EGEZST2003B08.g	1586	353	107	346	21952787	*Oryza sativa* (cv. *japonica*)	304.7	$1.2e^{-81}$
	EGEPSL4003G09.g	857	278	59	251	12054894	*Pinus sylvestris*	240.4	$1.9e^{-62}$

predictions (Figure 3). The reliability class (RC), which is a confidence measure for the prediction, showed that only 11 sequences were defined into RC1 (higher than 80%), and 53 for RC2 (higher than 60%) class. Most of the sequences are predicted to be located at unspecific subcellular localization (133 sequences) while 35, 20 and 19 were predicted to contain mitochondrial targeting, signal and chloroplast transit peptides, respectively (Figure 3).

After evaluation with the TargetP program, sequences with motifs specific for transmembrane anchoring could be identified in 44 of all analyzed sequences. From these 19 belonged to LRR or LRR-KINASE-related sequences and, unexpectedly, five showed to be TIR-NBS-LRR and 20 to be KINASE-related sequences.

Discussion

The reverse alignment (TBLASTN) strategy (Altschul *et al.*, 1997) adopted by our group identified a set of 210 clusters similar to the major classes of disease *R*-genes in the current version of the FOREST database,

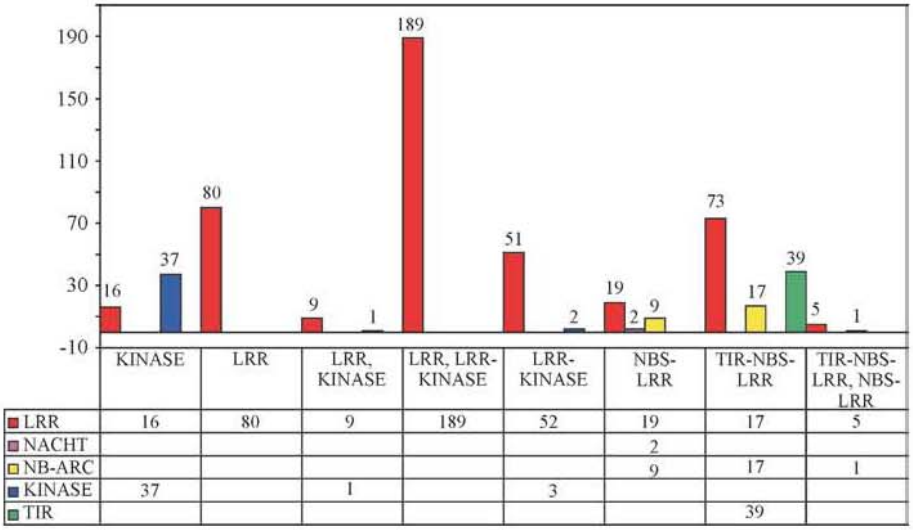


Figure 2 - Graphic representation of the distribution of conserved domains against class-grouped clusters. Values on the base after each domain indicate the number of clusters of each class presenting the indicated domain (also represented in the corresponding columns). Abbreviations: LRR – Leucine-rich-repeats; NB-ARC – Nucleotide-binding-site and additional motifs; NACHT – NB-ARC related domain, including an NTPase implicated in apoptosis and MHC transposition activation.

Table 4 - Inventory of the organisms that appeared as best alignment to each of the 210 here identified *Eucalyptus* clusters related to known resistance genes. The organisms are grouped by gene class (I to V and MIX I to III), taxonomic affiliation (class, subclass, family and species) and habit (herbaceous or woody). Numbers in parenthesis indicate amount of gene members in each taxonomic group or species.

Gene class	Higher taxonomic affiliation	Family	Species	Habit
I KINASE	Dicots (46)	Brassicaceae (43)	*Arabidopsis thaliana* (42)	H
			Brassica napus (1)	H
		Curcubitaceae (1)	*Cucumis melo* (1)	H
		Salicaceae (1)	*Populus nigra* (1)	W
		Solanaceae (2)	*Capsicum annuum* (1)	H
			Nicotiana tacabum (1)	H
	Monocots (2)	Poaceae (2)	*Oryza sativa* (2)	H
II LRR-NBS	Dicots (14)	Asteraceae (1)	*Lactuca sativa* (1)	H
		Brassicaceae (6)	*Arabidopsis thaliana* (6)	H
		Euphorbiaceae (3)	*Manifot esculenta* (3)	W
		Leguminosae (2)	*Glycine max* (1)	H
			Phaseolus vulgaris (1)	H
		Stercullariaceae (2)	*Theobroma cacao* (2)	W
	Monocots (7)	Poaceae (7)	*Oryza sativa* (7)	H
III NBS-LRR-TIR	Dicots (93)	Brassicaceae (10)	*Arabidopsis thaliana* (10)	H
		Curcubitaceae (5)	*Cucumis melo* (5)	H
		Leguminosae (10)	*Glycine max* (10)	H
		Asteraceae (14)	*Helianthus annuus* (14)	H
		Linaceae (34)	*Linum usitatissimum* (34)	H
		Euphorbiaceae (1)	*Manihot esculenta* (1)	W
		Salicaceae (12)	*Populus balsamifera* (10)	W
			Populus tremula (2)	W
		Solanaceae (7)	*Lycopersicon esculentum* (1)	H
			Solanum tuberosum (6)	H
IV LRR	Dicots (17)	Brassicacea (7)	*Arabidopsis thaliana* (7)	H
		Leguminosae (1)	*Glycine max* (1)	H
		Malvaceae (1)	*Gossypium hirsutum* (1)	W
		Solanaceae (8)	*Lycopersicon esculentum* (3)	H
			Lycopersicon hirsutum (2)	H
			Nicotiana tabacum (1)	H
			Petunia X hybrida (1)	H
			Solanum tuberosum (1)	H
V LRR-KINASE	Dicots (7)	Brassicaceae (6)	*Arabidopsis thaliana* (6)	H
		Leguminosae (1)	*Glycine max* (1)	H
	Monocot (1)	Poaceae (1)	*Oryza sativa* (1)	H
MIX I	Dicots (16)	Brassicaceae (12)	*Arabidopsis thaliana* (12)	H
		Solanaceae (3)	*Lycopersicon esculentum* (2)	H
			Lycopersicon peruvianum (1)	H
		Rosaceae (1)	*Malus X domestica* (1)	W
	Monocot (1)	Poaceae (1)	*Oryza sativa* (1)	H
	Gymnosperm (1)	Pinaceae (1)	*Pinus sylvestris* (1)	W
MIX II	Dicots (3)	Brassicaceae (2)	*Arabidopsis thaliana* (2)	H
		Stercullariaceae (2)	*Theobroma cacao* (1)	W
MIX III	Dicot 1	Brassicaceae (1)	*Arabidopsis thaliana* (1)	H

Synopsis regarding features of aligned species

		N.	%
Grouped by taxonomic affiliation	Dicots	198	94,3
	Monocots	11	5,2
	Gymnosperm	1	0,5
Grouped by habit	Herbaceous	187	89,0
	Woody	23	10,9

what comprises 0.63% of the actually generated clusters. This approach allowed the identification of a large set of candidate sequences by using various representative genes per class, while some recent works employed few genes (Koczyk and Chelkowski, 2003). Using several previously described and sequenced *R*-genes as template was a useful and low-time consuming strategy in the search for *R*-genes candidates in plants. In this approach it was expected that some similar genes grouped at the same class should cause some level of redundancy (Meyers *et al.*, 1999). The strategy of generating a local database (called non-redundant) by adopting the cluster number as a primary key register was very effective in the solution of this problem. Additionally, this approach was useful in the identification of the respective *R*-gene class for each *Eucalyptus* cluster.

The number of *R*-genes here identified is quite high, especially considering that none of the 19 libraries were obtained under pathogen stress condition. By the other hand, when additional ESTs are generated especially under infection by pathogen, many of the identified clusters may be united in larger clusters of *R*-genes that may include more domains.

Evidences have shown that *R*-genes are quite abundant in higher plants, but the most functionally defined *R*-genes belong to the supergene LRR-NBS family. After completing the whole genome sequencing of the model plant *A. thaliana* a total of 85 TIR-NBS-LRR have been identified (The Arabidopsis Genome Initiative, 2000), less than the number of clusters (93) actually identified in *Eucalyptus*. Especially genes containing NBS-LRR domains were estimated to be in number of ca.166 for *A. thaliana* and ca.600 for rice (*O. sativa*) by Richly *et al.* (2002), but this later number is still not confirmed.

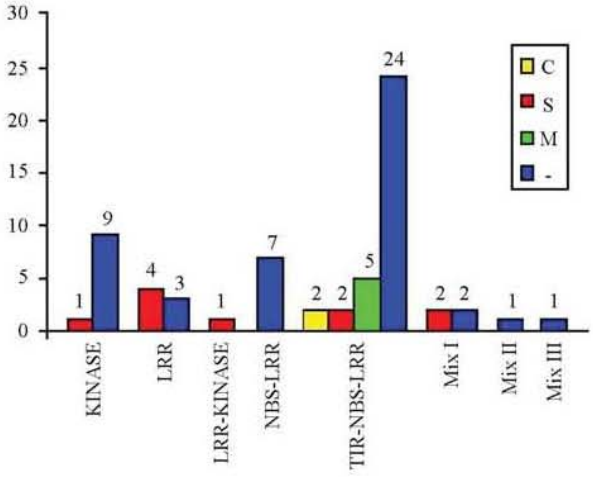


Figure 3 - Subcellular prediction for each class of analyzed *R* genes in *Eucalyptus* transcriptome, considering the predictions whit RC1 and RC2 of Target-P program. Legend: C – Chloroplast transit peptides; S – Secretory pathway; M – Mitochondrial targeting; '-' – No specific localization; LRR – Leucine-rich-repeat; NBS – Nucleotide-binding-site; TIR – Toll-interleucine-1 receptor.

A recent work reevaluated and reannotated all NBS-LRR encoding genes in *A. thaliana* genome database, revealing 149 genes of this class (including 94 TIR and 55 non-TIR sequences) in the genome of *A. thaliana* (Meyers *et al.*, 2003). In our evaluation of FOREST database we found 114 clusters (93 and 21, respectively) of this class. It is interesting to note that in the evaluation of Meyers *et al.* (2003) not only the presence of the TIR or of the CC motif was determinant for the grouping of both distinct classes. Also the NBS-LRR domains co-evolved and were determinant in the divergent evolution of the two groups, with the CC-bearing sequences forming four subgroups and the TIR-bearing sequences forming eight subgroups, regarding the size, composition and order or introns and exons.

Pan *et al.* (2000) compared tomato and *Arabidopsis* sequences of this class by systematically amplifying the tomato genome using a variety of primer pairs based on ubiquitous NBS motifs, generating 70 sequences, from which 10% were putative pseudogenes. The sequences were also used in mapping approaches, revealing a clustering *R*-gene homologues between tomato and potato (*Solanum tuberosum*, also from the Solanaceae family). Clustering of *R*-genes was also detected in *A. thaliana*, with most of the genes located in chromosomes 1 (49) and 5 (55), confirming the initial hypothesis that these genes are clustered in few chromosomes (The Arabidopsis Genome Initiative, 2000). This fact was also observed in other crops, as chickpea (*Cicer arietinum*; Benko-Iseppon *et al.*, 2003). In this last case, with some synteny and colinearity within this species and *Arabidopsis*. The clustering of *R*-genes in specific chromosomes and the existence of conserved domains have allowed the establishment of interesting strategies for identification, mapping and breeding directed to the incorporation of such genes from wild relatives. Considering the number of genes from this group in this last species, it is to expect that they are also clustered in *Eucalyptus*, what can also be valuable for the establishment of *Eucalyptus* breeding strategies in the future, especially considering the previous existence of mapping populations for this crop.

Overall annotation revealed that *Arabidopsis* also carries homologues of other *R*-gene classes, including 174 genes encoding LRR-kinases (*Xa21* group), but many of which are likely to play a role in development rather than defense (Jones, 2001). The present work revealed only eight clusters with significant homology to *Xa21* but this number can increase if only the kinase sequence is used as template, since the LRR may be quite variable between rice and *Eucalyptus*. Exceptional *R*-genes have proven to provide durable disease control, due to the fast evolving pathogen genome that breaks resistance. The *Xa21* gene is an important exception to this rule that reveals the full potential of *R*-genes for breeding purposes (Rommens and Kishore, 2000). This may be very valuable especially considering the possibility of pyramidization of such genes in important crops, increasing the potentiality of an effective specific *R-Avr* intection.

Another abundant family of *R*-genes in plants is the ser-thr-kinase with about 50 genes in *Arabidopsis* encoding protein kinases that are strongly homologous to tomato's *Pto* gene (Jones, 2001). In *Eucalyptus* we found almost the same number (49) of clusters also with high homology to the *Pto* sequence.

Regarding *R*-gene classes identified in *Eucalyptus*, an interesting phenomenon was observed in the present work: *R*-genes pertaining to different classes were able to align significantly to the same cluster on *Eucalyptus* database. This can be explained by the evidences that known *R*-genes combine a limited number of related functional domains (Ellis *et al.*, 1999, 2000a). Then, similar motifs would be present in different *R*-genes, and it is possible that a gene resembling to a determined class may search another belonging to a different class by local similarity at the site of the conserved motif. But in the practice, previous works do not speculate this possibility, once that the genes identified for specific *R*-genes are directly assigned to its own class as shown by evidences raised from works previously reported (Ronald, 1997; Jones, 2001; Romeis, 2001).

The MIX class one (MIX I) included 18 clusters resembling to genes which belong to both LRR and LRR-KINASE classes. These clusters were searched basically by using Cf (Jones *et al.*, 1998) and Xa21 (Song *et al.*, 1995) amino acid sequences as queries. In this case, the most plausible explanations would be the presence of the LRR domain, common to both classes, being responsible for the alignment and grouping of some clusters in both classes. By the other hand, LRRs are referred as fast evolving sequences and are in some cases quite imperfect, making manual annotation necessary. Often their amino-acid sequences are quite specific to their gene group (Dixon *et al.*, 1998; Ellis *et al.*, 1999). For example, using the LRR of *Xa21* against GenBank database will reveal significant alignments only to *Xa21* genes of rice (and some other Poaceae) and less significantly to *Arabidopsis*, but no sequence including other gene classes align significantly. A similar approach to the present work was used for the analysis of SUCEST (Sugarcane EST project, also running in Brazil) database (Morais, 2003) with no similar results. Song *et al.*, (1997) suggested that the structure of *Xa21* (here referred as class V) itself indicates an evolutionary link between different classes (I and IV) of plant disease resistance genes. May this be the case of this cluster that present a new link between two classes and can represent a new gene for Angiosperms?

Another surprising result was obtained by analyzing the unique cluster with both domains LRR and KINASE. It would be expected to find both domains in genes resembling *Xa21* but this cluster (EGEQRT3301C03.g) showed itself similar to both *Pto* (class I, described by Martin *et al.*, 1993) and *Cf* (Class IV, described by Jones *et al*, 1994)

genes. This double similarity occurred on different motifs. The *Pto* gene is known to encode a ser-thre-kinase protein (Martin *et al.*, 1993) and it was at this motif that the cluster showed similarity to this gene. On the other hand, *Cf* genes encode extracellular LRRs and it was at the LRR motif that the similarity was found. This cluster could be grouped in the LRR-KINASE class. So, why did it not align with *Xa21*, the single known gene with both LRR-KINASE domains? It should be answered by analyzing the KINASE-related clusters. Despite of the conservation of this region (Romeis, 2001), none of the *Pto* (KINASE) or *Xa21* (LRR and a receptor-KINASE) related clusters were mixed (aligned together) during the annotation process. This shows that the kinase segment is less-redundant than LRR at least during our *in silico* gene prediction, once that the kinase CD is present in both *Pto* and *Xa21* genes, they do not caused the mixture of their matching clusters on a mixed class.

The last case of mixture occurred to MIX class II including the motif TIR-NBS-LRR. Two of the three clusters pertaining to this mixed class (EGEQST6001H02.g and EGJECL1208G03.g) were searched at the FOREST database by the genes *RPP5* (TIR-NBS-LRR; Parker *et al.*, 1997) and *RPS5* (NSB-LRR; Noel *et al.*, 1999). The third cluster (EGEZRT3006B12.g) was obtained through search using *RPP5* and *RPS4* (both TIR-NBS-LRR; Gassmann *et al.*, 1999) and *I2* (NBS-LRR; Simmons *et al.*, 1998) queries. We initially supposed that the redundancy was due to the presence of NB-ARC (NBS) conserved motif. However, the first two clusters did not show any motif after *in silico* CD-search and, again, the region that apparently caused the mixture of the classes was the LRR motif, once that it was predicted in cluster EGEZRT3006B12.g.

In view of the results discussed above, could we speculate that *Eucalyptus* bears some new classes of *R*-genes? Before taking further conclusions and in order to solve the questions raised by the present work, we intend to evaluate these groups of clusters in regard to their domain and interdomain structure and organization, evaluating also the clusterization process, before taking further conclusions.

The conserved domains (CDs) identified during our investigation showed that most of the *Eucalyptus* predicted sequences possess the same motifs shared by disease *R*-genes. The CD with the higher level of sampling was LRR, which was present in all classes (except KINASE class I, represented by *Pto*) with a total of 442 occurrences. The other frequent domain shared by *R*-genes, the NB-ARC, was observed in 27 sequences, notably in TIR-NBS-LRR and NBS-LRR predicted clusters. This motif is commonly found in such sequences, and it is proposed that NB-ARC plays a role in activation of downstream effectors (van der Biezen and Jones, 1998) by their sequence similarity to mammalian CED-4 and APAF-1 proteins which are involved in apoptosis (Chinnaiyan *et al.*, 1997). In plants the TIR motif is found only associated to NBS regions of dicotyledones, being possibly absent in monocotyledones (Meyers *et al.*, 1999). In *Eucalyptus* (a eudicot genus of the Myrtaceae family) TIR domains were quite abundant, as expected, being found in 39 clusters (all from TIR-NBS-LRR-class).

Another very common motif present in two classes of disease *R*-genes is the kinase domain. This motif is shared by *Pto* (ser-thre-kinase) and *Xa21* (receptor-kinase) genes, members of the KINASE and LRR-KINASE classes, respectively. We found that all kinase domains found were associated to the classes KINASE, LRR-KINASE and MIX III. As commented here, despite of its conservation, this domain generally does not cause redundancy while searching in databases.

Transmembrane motifs were found only in 44 of all analyzed sequences. Of these clusters five TM were, unexpectedly, found in TIR-NBS-LRR-related sequences (a group of *R*-genes that acts at the intracellular level), while the remaining 19 were as expected LRR or LRR-KINASE-related sequences.

Information regarding the localization of disease resistance proteins in plant cells is still scarce (Martin, 1999). Spatial organization is usually variable among distinct gene classes and tissues affected, and there are no strong evidences in favor of conserved correspondence between R and Avr products spatial occurrence (Bonas and Lahaye, 2002). However, immunocytochemistry approaches allowed the subcellular localization of some Avr and R components (Boyes *et al.*, 1998). Here, we adopted an *in silico* approach which uses neural network-based methods to predict the topology (*i.e.* localization) of protein sequences of the selected clusters. In spite of the large number of predictions obtained, only 11 sequences were defined into RC1 (reliability class 1 $\geq$ 80%), and 53 for RC2 ($\geq$ 60%). Of these significant predictions, we observed that neural network was able to predict the localization of only a small number of proteins (29.62%) compared to the total sample of *Eucalyptus R*-genes. This percentage of representation is much lower than the 80% obtained for plant test sets carried out by Emanuelsson *et al.* (2000) with the same approach. It is important to note that these predictions are based on the N-terminal information available for sequences. Thus, this low number of predictions can be explained by the fact that the FOREST database was obtained from expressed sequence tags, an approach that usually do not include N-termini for many EST generated.

Our *Eucalyptus* transcriptome cDNA sequence analysis revealed that there are 210 clusters with significant alignment to major classes of plant disease *R*-genes. Differentially from the other genomic efforts, as *O. sativa* (Goff *et al.*, 2002) we used a redundant set of well described *R*-genes to screen for RGAs (Resistance Genes Analogs) on FOREST database. This proved to be a very sensitive approach, since best matches in NCBI present sometimes annotation mistakes and we also observed during the present work that

some of the best GenBank matches to *Eucalyptus R*-clusters presented no conclusive description of function. This was also the case also of the first annotation of *Arabidopsis* genome sequences, as pointed out by Meyers *et al.*, (2003). After reannotation of NBS-LRR sequences a total of 56 of the *A. thaliana R*-genes had to be corrected from earlier evaluations on GenBank (Meyers *et al.*, 2003). These results show how important procedures as annotation and detailed evaluation of generated sequences are. These evidences bring to reflections about the strategic design of many genome and transcriptome projects, considering that the data mining is not expensive (normally only fellowships are needed) but still receive few investments from financing agencies, diminishing the final impact of the results.

The comparison of our results regarding the number (and maybe the organization) of identified *Eucalyptus* clusters was mainly with *A. thaliana*, especially due to the lack of open databases for other plant species with EST projects. Many differences considering the here analyzed *R*-related sequences can be explained by using diverse arguments: (i) The larger genome of *Eucalyptus* (e.g. *E. grandis* with 640 Mbp; Myburg *et al.*, 2003) in contrast with the small and "compact" genome of *A. thaliana* (120 Mbp) (ii) The distant taxonomic position: both are dicots, but distantly related families (Brassicaceae and Myrtaceae) and finally (iii) the different levels of complexity: *Eucalyptus* is a wood perennial plant species and *Arabidopsis* is an annual herb. Herbaceous species are often regarded as faster evolving than woody species considering different morphological and genetic aspects (Bennet, 1972, Enrendorfer, 1982, Morawetz 1984, 1986, Bennet and Leitch, 1995, 2000).

Considering these evidences we observed that most of the information regarding *R*-genes available in databases refer to herbaceous (not woody) crop plants (few wild plants), maybe because most identified and sequenced *R*-genes were consequence of mapping approaches that are very time consuming in woody plants and difficult to realize in open pollinated species. The larger number of sequences from *A. thaliana* representing best alignments to *Eucalyptus* does not represent a higher similarity to this plant species, moreover it reflects the large number of sequences of this model plant deposited in GenBank. In our evaluation, only 23 woody species appeared as best matches for the clusters studied, including 22 species from different dicotyledonous families and one Gymnosperm species (*Pinus sylvestris*). This may justify some of the surprising results obtained in the present work and suggest that identification of *R*-genes in a larger number of taxonomic groups may be a very promissory approach to understand the natural evolution of these sequences when not affected by the influence of man. Regarding the actual knowledge of R-gene structure and diversity, some authors suggested that this gene class evolves faster than other genes (Ellis *et al.*, 2000b) what should be evaluated in a larger number of taxonomic entities including wild species and also primitive taxa.

Concluding Remarks

Using bioinformatic tools it was possible to identify classify and verify the actually sequenced *R*-genes in *Eucalyptus* transcriptome. No previous sequences of this type could be found in protein or nucleotide databases for this crop. The identified sequences will be valuable resources for the development of markers for molecular breeding and identification of RGAs (resistance gene analogs) in *Eucalyptus* and other related species. The identified clusters constitute also excellent probes for physical mapping of genes in this species, giving support to genetic mapping programs and synteny studies. Considering the size of some clusters, they may also be used for fluorescent *in situ* hybridization (FISH) on *Eucalyptus* chromosomes, helping also in the comparison of different parental species and the respective hybrids.

The present work on *Eucalyptus*, based on FOREST database brought some light to the existing *R*-gene group in this important crop species and also regarding resistance response in higher plants, leading to the following conclusions:

- All five gene classes of *R*-genes with their respective conserved domains are present and expressed in *Eucalyptus*.
- Some new combinations of domains and motifs of *R*-genes may be present in *Eucalyptus* and could represent novel *R*-gene structures, what should be analyzed in detail.
- Despite the lack of libraries from tissues ellicitated by pathogens a high number of *R*-genes was found in different libraries of FOREST project. This may suggest, that the identified clusters are expressed constitutively but also leads to the supposition that a higher number of *R*-genes may be present in *Eucalyptus* under other experimental conditions.

Besides the detailed analysis of different groups of genes and domains we intend to evaluate the expression of the selected clusters in the different libraries of the project. Furthermore, some additional efforts may be necessary to complete some sequences of *R*-genes, especially considering that their size vary between 321 (in case of *Pto*) and 1802 amino-acids (in case of *Xa1* gene) and many identified sequences possibly present incomplete domains.

Further *in silico*, *in vitro* and *in vivo* evaluations of *Eucalyptus* genome may be a very promissory approach. Manipulation of the expression of these genes in economically important woody plant species aiming to improve disease resistance is necessary. Despite of the challenge that this mission may represent, some reports indicate that this strategy is feasible.

Acknowledgements

The present authors thank Ms. David Anderson de Lima Morais and Dr. Valdir Queiroz Balbino for interesting discussions and instructions about some of the pro-

grams and tools used in the present work. To Dr. Reginaldo de Carvalho and Claudete Maria Marques da Silva we thank for valuable technical support. We thank also CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the concession of a fellowship to the last author (Grant no. 478895/2003).

References

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

Altschul SF, Madden TL, Schaffer AA, Zhang J, Mille W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

Ballvora A, Ercolano MR, Weiss J, Meksem K, Bormann CA, Oberhagemann P, Salamini F and Gebhardt C (2002) The *R1* gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. Plant J 30:361-371.

Barber PA, Smith IW and Keane PJ (2003) Foliar diseases of *Eucalyptus* spp. grown for ornamental cut foliage. Austral Plant Pathol 32:109-111.

Bendahmane A, Kohn BA, Dedi C and Baulcombe DC (1995) The coat protein of potato virus X is a strain-specific elicitor of Rx1-mediated virus resistance in potato. Plant J 8:933-941.

Bendahmane A, Querci M, Kanyuka K and Baulcombe DC (2000) *Agrobacterium* transient expression system as a tool for the isolation of disease resistance genes: Application to the Rx2 locus in potato. Plant J 21:73-81.

Benko-Iseppon AM, Winter P, Huettel B, Staginnus C, Muehlbauer FJ and 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. Theor Appl Genet 107:379-386.

Bennett MD (1972) Nuclear DNA content and minimum generation time in herbaceous plants. Proc R Soc Lond Bot 181:109-135.

Bennet MD and Leitch IJ (1995) Nuclear DNA Amounts in Angiosperms. Ann Bot 76:113-176.

Bennet MD and Leitch, IJ (2000) Variation in nuclear DNA amount (C-value) in monocots and its significance. In: Wilson KL and Morrison DA (eds) Monocots: Systematics and Evolution. 1st edition. CSIRO Publishers, Sydney, pp 137-146.

Bittner-Eddy PD, Crute IR, Holub EB and Beynon JL (2000) RPP13 is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. Plant J 21:177-88.

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.

Botella MA, Parker JE, Frost LN, Bittner-Eddy PD, Beynon JL, Daniels MJ, Holub EB and Jones JD (1998) Three genes of the *Arabidopsis* RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. Plant Cell 10:1847-1860.

Boyes DC, Nam J and Dangl JL (1998) The *Arabidopsis thaliana RPM1* disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. Proc Natl Acad Sci USA 95:15849-15854.

BRACELPA (2004), Associação Brasileira de Celulose e Papel. Brazil. Available from World Wide Web: http://www.bracelpa.org.br, release date 20/March/2004, cited 25/April/2004.

Brommonschenkel SH, Frary A and Tanksley SD (2000) The broad-spectrum tospovirus resistance gene *Sw-5* of tomato is a homolog of the root-knot nematode resistance gene *Mi*. Mol Plant-Microbe Interact 13:1130-38.

Bryan GT, Wu KS, Farrall L, Jia Y, Hershey HP, McAdams SA, Faulk KN, Donaldson GK, Tarchini R and Valent B (2000) A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant Cell 12:2033-2046.

Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, Topsch S, Vos P, Salamini F and Schulze-Lefert P (1997) The barley *Mlo* gene: A novel control element of plant pathogen resistance. Cell 88:695-705.

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

Collins N, Drake J, Ayliffe M, Sun Q, Ellis J, Hulbert S and Pryor T (1999) Molecular characterization of the maize Rp1-D rust resistance haplotype and its mutants. Plant Cell 11:1365-1376.

Dixon MS, Jones JGD, Keddie JS, Thomas CM, Harisson K and Jones JGD (1996) The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell 84:451-459.

Dixon MS, Hatzixanthis K, Jones DA, Harisson K and Jones JGD (1998) The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell 10:1915-1925.

Dodds P, Lawrence G and Ellis J (2001) Six amino acid changes confined to the leucine-rich repeat beta-strand/beta-turn motif determine the difference between the P and P2 rust resistance specificities in flax. Plant Cell 13:163-78.

Ehrendorfer F (1982) Speciation patterns in woody angiosperms of tropical origin. In: Barigozzi C (ed) Mechanisms of Speciation. Alan R. Liss. Inc., New York, pp 479-509.

Ellis J and Jones D (1998) Structure and function of proteins controlling strain-specific pathogen resistance in plants. Curr Opin Plant Biol 1:288-293.

Ellis JG, Lawrence GJ, Luck JE and Dodds N (1999) Identification of regions in alleles of the flax rust resistance gene *L* that determines differences in gene-for-gene specificity. Plant Cell 11:495-506.

Ellis J, Dodds P and Pryor T (2000a) The generation of plant disease resistance genes specificities. Trends Plant Sci 5:373-379.

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

Emanuelsson O, Nielsen H, Brunak B and von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300:1005-1016.

Ernst K, Kumar A, Kriseleit D, Kloos DU, Phillips MS and Ganal MW (2002) The broad-spectrum potato cyst nematode resistance gene (Hero) from tomato is the only member of a large gene family of NBS-LRR genes with an unusual amino acid repeat in the LRR region. Plant J 31:127-136.

Flor HH (1956) The complementary genetic systems in flax and flax rust. Adv Genet 8:29-54.

Flor HH (1971) Current status of the gene-for-gene concept. Annu Rev Plant Pathol 9:275-296.

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.

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A and Briggs S (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92-100.

Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW and Dangl JL (1995) Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. Science 269:843-846.

Halterman D, Zhou F, Wei F, Wise RP and Schulze-Lefert P (2001) The MLA6 coiled-coil, NBS-LRR protein confers AvrMla6-dependent resistance specificity to *Blumeria graminis* f. sp. *hordei* in barley and wheat. Plant J 3:335-348.

James WC, Teng PS and Nutter FW (1990) Estimated losses of crops from plant pathogens. In: Pimentel D (ed) CRC Handbook of Pest Management, CRC Press, Boca-Raton, pp 15-50.

Johal GS and Briggs SP (1992) Reductase activity encoded by the *HM1* disease resistance gene in maize. Science 158:958-987.

Jones DA, Thomas CM, Hammond-Kosac KE, Balint-Kurti J and Jones JGD (1994) Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266:789-793.

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

Jones JB, Stall RE and Bouzar H (1998) Diversity among xanthomonads pathogenic on pepper and tomato. Ann Rev Phytopathol 36:41-58.

Joosten MH, Cozijnsen TJ and De Wit PJ (1994) Host resistance to a fungal tomato pathogen lost by a single base-pair change in an avirulence gene. Nature 367:384-386.

Koczyk G and Chelkowski J (2003) An assessment of the resistance gene analogues of *Oryza sativa* ssp. *japonica*, their presence and structure. Cell Mol Biol Lett 8:963-972.

Koonin EV and Aravind L (2000) The NACHT family - A new group of predicted NTPases implicated in apoptosis and MHC transcription activation. Trends Biochem Sci 25:223-224.

Lawrence GJ, Finnegan EJ, Ayliffe MA and Ellis JG (1995) The *L6* gene for flax rust resistance is related to *Arabidopsis* bacterial resistance gene *RPP2* and tobacco viral gene *N*. Plant Cell 7:1195-1206.

Mafia RG and Alfenas AC (2003) Diferenciação sintomatológica de manchas foliares em *Eucalyptus* spp. causadas por patógenos fúngicos e bacterianos. Fitopatol Bras 28:688-688.

Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY and Bryant SH (2002) CDD: A database of conserved domain alignments with links to domain three-dimensional structure. Nucl Acids Res 30:281-283.

Martin GB (1999) Functional analysis of plant disease resistance genes and their downstream effectors. Curr Opin Plant Biol 2:273-279.

Martin GB, de Vicente MC and Tanksley SD (1993) Hight resolution linkage analysis and physical characterization of the Pto bacterial locus in tomato. Mol Plant-Microbe Interact 6:26-34.

Martin GB, Bogdanove AJ and Sessa G (2003) Understanding the functions of plant disease resistance proteins. Annu Rev Plant Physiol Plant Mol Biol 54:23-61.

McDowell JM, Dhandaydham M, Long TA, Aarts MG, Goff S, Holub EB and Dangl JL (1998) Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of *Arabidopsis*. Plant Cell 10:1861-1874.

McNabb K (2002) Clonal propagation of *Eucalyptus* in Brazilian nurseries. In: Dumroese RK, Riley LE and Landis TD (eds) National Proceedings: Forest and Conservation Nursery Associations. USDA Forest Service, Rocky Mountain Research Station, Ogden, pp 165-168.

Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z and Michelmore RW (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10:1817-32.

Meyers BC, Diekcman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW and Young ND (1999) Plant disease resistance genes encodes members of an ancient and diverse protein family within the nucleotide-biding superfamily. Plant J 20:317-332.

Meyers BC, Kozik A, Griego A, Kuang H and Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15:809-834.

Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I, Zabel P and Williamson VM (1998) The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. Plant Cell 10:1307-19.

Mindrinos M, Katagiri F, Yu GL and Ausubel FM (1994) The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-biding site and leucine-rich repeats. Cell 78:1089-1099.

Morais DAL (2003) Análise bioinformática de genes de resistência a patógenos no genoma da cana-de-açúcar. Master Dissertation, Universidade Federal de Pernambuco, Recife.

Morawetz W (1984) How stable are genomes of tropical woody plants? Heterozygosity in C-banded Karyotypes of *Porcelia* as compared with *Annona* (Annonaceae) and Drymys (Winteraceae). Pl Syst Evol 145:29-39.

Morawetz W (1986) Remarks on karyological differentiation patterns in tropical woody plants. Pl Syst Evol 152:49-100.

Myburg AA, Griffin AR, Sederoff RR and Whetten RW (2003) Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. Theor Appl Genet 107:1028-1042.

Noel L, Moores TL, van Der Biezen EA, Parniske M, Daniels MJ, Parker JE and Jones JD (1999) Pronounced intraspecific haplotype divergence at the RPP5 complex disease resistance locus of *Arabidopsis*. Plant Cell 11:2099-2112.

Ori N, Eshed Y, Paran I, Presting G, Aviv D, Tanksley S, Zamir D and Fluhr R (1997) The I2C family from the wilt disease resistance locus I2 belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. Plant Cell 9:521-532.

Pan Q, Liu YS, Budai-Hadrian O, Sela M, Carmel-Goren L, Zamir D and Fluhr R (2000) Comparative genetics of nucleotide binding site leucine-rich repeat resistance gene homologues in the genomes of two dycotyledons: Tomato and *Arabidopsis*. Genetics 155:309-322.

Parker JE, Coleman MJ, Dean C and Jones JGD (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.

Piffanelli P, Zhou F, Casais C, Orme J, Jarosch B, Schaffrath U, Collins NC, Panstruga R and Schulze-Lefert P (2002) The barley MLO modulator of defense and cell death is responsive to biotic and abiotic stress stimuli. Plant Physiol 129:1076-1085.

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

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

Romeis T (2001) Protein kinases in the plant defense response. Curr Opin Plant Biol 4:407-414.

Rommens CM and Kishore GM (2000) Exploiting the full potential of disease resistance genes for agricultural use. Curr Opin Biotechnol 11:120-125.

Ronald PC (1997) The molecular basis of disease resistance in rice. Plant Mol Biol 35:179-186.

Simmons G, Groenendijk J, Wijbrandi J, Reijans M, Groenen J, Diergaarde van der Lee T, Bleeker M, Onstenk J, De Both M, Haring M, Mes J, Cornelissen B, Zabeau M and Vos P (1998) Dissection of the fusarium *I2* gene cluster in tomato reveals six homologs and one active gene copy. Plant Cell 10:1055-1068.

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

The *Arabidopsis* Genome Iniciative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796-815.

Thomas CM, Jones DA, Parniske M, Harrison K, Balint-Kurti PJ, Hatzixanthis K and Jones JD (1997) Characterization of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in *Cf-4* and *Cf-9*. Plant Cell 9:2209-2224.

van der Biezen EA and Jones JGD (1998) The NB-ARC domais: A novel signaling motif shared by plant resistance gene products and regulators of cell death in animals. Curr Biol 8:R226-R227.

van der Biezen EA, Freddie CT, Kahn K, Parker JE and Jones JD (2002) *Arabidopsis RPP4* is a member of the *RPP5* multigene family of TIR-NB-LRR genes and confers downy mildew resistance through multiple signaling components. Plant J 29:439-51.

van der Vossen EA, van der Voort JN, Kanyuka K, Bendahmane A, Sandbrink H, Baulcombe DC, Bakker J, Stiekema WJ and Klein-Lankhorst RM (2000) Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: A virus and a nematode. Plant J 23:567-576.

Wang ZX, Yano M, Yamanouchi U, Iwamoto M, Monna L, Hayasaka H, Katayose Y and Sasaki T (1999) The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant J 19:55-64.

Whithan S, McCormick S and Baker B (1996) The *N* gene of tobacco confers resistance to tobacco mosaic virus in transgenic tomato. Proc Natl Acad Sci USA 93:8776-81.

Xiao S, Ellwood S, Calis O, Patrick E, Li T, Coleman M and Turner JG (2001) Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by RPW8. Science 291:118-20.

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang ZX, Kono I, Kurata N, Iwata N and Sasaki T (1998) Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc Natl Acad Sci USA 95:1663-1668.

Associate Editor: Marcio de Castro Silva Filho

5.2 Capítulo do livro "Estresses ambientais: danos e benefícios em plantas"

GENES ASSOCIADOS A ESTRESSES BIÓTICOS E ABIÓTICOS EM FEIJÃO-CAUPI [*VIGNA UNGUICULATA* (L.) WALP.] E OUTRAS ANGIOSPERMAS

Benko-Iseppon, A.M.; Soares-Cavalcanti, N.M.; Wanderley-Nogueira, A.C.; Berlarmino, L.C.; Silva, R.R.M.; Almeida, P.M.L.; Brunelli, K.R.; Kido, L.M.H.& Kido, E.A.

Lab. de Genética e Biotecnologia Vegetal & Lab. de Genética Molecular, Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, 50732-970, Recife, PE, Brasil. E-Mail: celisep@hotlink.com.br

Resumo - O presente estudo apresenta os principais grupos de genes relacionados às vias de estresse abiótico (com ênfase em salinidade e seca), bem como os fatores genéticos determinantes da resistência específica (genes R) previamente descritos em feijão-caupi em outras plantas, avaliando seqüências disponíveis em bancos de dados restritos e públicos com ferramentas de bioinformática. Tal levantamento servirá como base para comparação com novas seqüências que estão sendo geradas no âmbito do projeto "Genômica Funcional, Estrutural e Comparativa de Feijão-Caupi (*Vigna unguiculata*)", que ora se inicia, incluindo bibliotecas desenhadas para a identificação de genes relacionados a estresses bióticos e abióticos nesta importante cultura. O levantamento efetuado revela que, comparativamente a outras plantas cultivadas, incluindo leguminosas, ainda são escassos os estudos com feijão-caupi (e com outras espécies do gênero *Vigna*). Seqüências depositadas em bancos genéticos restringem-se em sua maioria a marcadores moleculares seqüenciados (em geral sem domínios conservados completos) ou a proteínas derivadas estudos de expressão voltados para seca e salinidade. Face à sua capacidade de adaptação a condições adversas, pode ser considerado como uma fonte ímpar de genes (e características) de resistência a estresses ambientais e bióticos, como os reinantes na região nordeste do Brasil.

Palavras-Chave: Genes R, Salinidade, Seca, Interação gene-a-gene, Estresse.

Abstract - This work presents a survey of the main gene groups related to abiotic stress pathways in cowpea and other plants with emphasis to salinity and drought. It also discusses the main gene groups responsible for the specific pathogen resistance in plants (*R* genes). The evaluations included the use of bioinformatic tools applied to data available on public and restrict plant databases, being an important start for comparison to sequences generated on the ongoing project "Functional, structural and comparative genomics in cowpea (*Vigna*

unguiculata)", a project that will analyse sequences expressed under biotic and abiotic stresses in this important crop. The results revealed that there is still little information regarding response to biotic and abiotic stress in cowpea (and remaining *Vigna* species), as compared to other cultivated plants, including Leguminosae. The identified sequences refer mainly to sequenced molecular markers (normally with incomplete conserved domains) or proteins expressed under salinity and drought conditions. The possibilities are promising, especially considering the unique capacity of adaptation that cowpea presents in various abiotic and biotic conditions, as it is the case of Brazilian North-eastern region.

Keywords: R genes, Salinity, Drought, Gene-by-gene interaction, Stress.

Introdução

Dentre as plantas cultivadas e usadas pela população do Nordeste Brasileiro, nenhuma tem tantas características vantajosas como o feijão-de-corda (*Vigna unguiculata* (L.) Walp). Conhecido também como feijão-caupi, inclui o popular feijão-de-corda, fradinho ou macassar e o feijão-verde (Freire-Filho *et al.*, 1999). Chamado de "*southern pea*" ou "ervilha do sul" nos Estados Unidos, país no qual constitui-se em produto para exportação e consumo interno, na forma de laticínios (do tipo seleta de legumes). Algumas variedades e espécies próximas são usadas na obtenção de *moyashi* (broto-de-feijão), uma cara iguaria da cozinha asiática, já difundida no mundo todo, também comercializada em laticínios ou congelados. Além de produzir os grãos citados, está entre as plantas mais versáteis, sendo muito valiosa como forrageira, com uma capacidade ímpar de fertilizar solos pobres, mesmo em condições de calor, seca e salinidade. Devido à sua adaptabilidade, o feijão-caupi foi escolhido pela NASA (*North American Space Agency*) como uma das poucas plantas propícias para cultivo em estações espaciais (Ehlers & Hall, 1997).

A capacidade do feijão-caupi de crescer sob condições de solo e clima 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, solo argiloso) e no nordeste (seca, solo arenoso, por vezes salino, muito sol), crescendo também em locais menos estressantes e com clima temperado, como no estado norte-americano da Califórnia (Ehlers & Hall, 1997; Freire-Filho *et al.*, 1999). Também se apresenta muito promissora como fonte de importantes genes para o melhoramento de outras leguminosas, como a soja, o feijão-comum, a ervilha, a lentilha, o amendoim, o grão-de-bico, etc. (Benko-Iseppon, 2001). Comparado a estas leguminosas o feijão-caupi apresenta características superiores para a recuperação de áreas degradadas, uma vez que em associação com rizóbio nativo torna-se eficiente fixador de nitrogênio sob as mais adversas condições (Ehlers & Hall, 1997; 1998),

sendo um excelente candidato para uso em biorremediação (recuperação de ambientes degradados), transformando regiões em processo de desertificação em áreas novamente férteis e produtivas (Benko-Iseppon, 2001).

Estresse abiótico em plantas

O crescimento vegetal pode ser severamente afetado por estresses abióticos, como seca, alta salinidade, bem como baixas ou altas temperaturas. Tais estresses abióticos se constituem em fatores severamente limitantes da produção vegetal. Quando tais condições abióticas adversas ocorrem, várias respostas bioquímicas e fisiológicas são induzidas em plantas, de forma a propiciar a tolerância ou aumentar suas chances de sobrevivência.

As reações moleculares ao estresse hídrico em plantas vêm sendo analisadas, revelando vários genes de resposta à seca, alta salinidade e frio em nível transcricional. Os produtos identificados podem ser classificados em dois grupos principais. Aqueles que agem diretamente na proteção aos estresses ambientais e aqueles que regulam a expressão e a transdução de sinais direcionados à resposta ao estresse (Thomashow, 1999; Hasegawa *et al.*, 2000).

Estudos de microarray visando monitorar a expressão de 7000 genes de indivíduos de *Arabidopsis thaliana* submetidos à seca, a altos níveis de salinidade e a frio revelaram expressão aumentada (em mais de cinco vezes comparativamente aos controles não-estressados) de 277, 194 e 53 genes, respectivamente. Interessante foi observar que 22 genes tinham sua expressão aumentada em todas as três situações de estresse abiótico. Na definição da resistência e/ou susceptibilidade aos estresses citados, torna-se importante não apenas observar quais genes são ativados, mas *a velocidade* com que determinados fatores são transcritos, observando-se acentuadas diferenças entre indivíduos resistentes e suscetíveis, sob este aspecto. Entre os três grupos de estresses, 11 genes imediatamente indutíveis na fase pós-estresse (entre poucos minutos até duas horas após o tratamento) foram observados, revelando fatores ideais para inclusão em estudos visando identificar fontes gênicas de resistência a estresses abióticos (Seki *et al.*, 2002; Oono *et al.*, 2003).

Estresse biótico em plantas

Embora os fatores abióticos sejam determinantes para o sucesso da atividade agrícola, fatores bióticos, como pragas e microrganismos, também podem comprometer os níveis de produtividades esperados para as culturas. Resistência a estes agentes tem merecido atenção especial dentro dos programas de melhoramento genético. Ao longo da co-evolução planta-patógeno, as plantas vêm desenvolvendo

mecanismos sofisticados de proteção contra o ataque destes organismos (Tang *et al.*, 1999). Em vegetais, genes de resistência (*R*) estão entre os mais importantes, sendo responsáveis pela resposta imune específica. Para cada gene conferindo resistência no hospedeiro, existirá um gene correspondente no patógeno, designado como gene de avirulência (*avr*), o qual determina a patogenicidade. A planta será resistente e o crescimento do patógeno interrompido apenas quando ambos os genes *R* e *avr* estiverem presentes e forem compatíveis, desencadeando assim a reação de hipersensibilidade (HR, *Hipersensitive Response*; Bonas & Lahaye, 2002).

De acordo com a natureza de seus produtos, os genes *R* podem ser classificados em cinco diferentes grupos, de acordo com sua estrutura e com os domínios que apresentam (vide Figura 1). Sugere-se que essa estrutura conservada seja determinante da especificidade patógeno-hospedeiro, ou seja, se a planta será ou não resistente a um determinado vírus, bactéria, fungo ou nematóide (Richter & Ronald, 2000). Deve-se destacar que o mesmo gene pode determinar resistência a diferentes patógenos, dependendo da espécie em que se expressa. Por exemplo, o gene de resistência *N* de *Nicotiana tabacum* L. determina a resistência ao vírus do mosaico no tabaco. Porém, este gene foi responsável pela aquisição de resistência a uma bactéria (*Xanthomonas campestris* pv. *vesicatoria*) quando expresso em tomate (Tai *et al.*, 1999).

A identificação destes genes é de grande relevância para o melhoramento genético de plantas, constantemente expostas ao ataque dos mais diversos patógenos.

Material e Métodos

Seqüências de nucleotídeos e amino-ácidos foram identificadas no GenBank (NCBI, *National Center for Biotechnology Information*; http://www.ncbi.nlm.nih.gov). No caso de genes *R*, membros das classes II a V previamente descritas foram buscados através de seus domínios conservados (Figura 1). Membros da primeira classe de genes *R* (redutases) que não apresentam domínios conservados, não foram incluídos no estudo.

As seqüências selecionadas foram avaliadas com auxílio do grupo de ferramentas BLAST, sendo traduzidas no *Expasy Translate Tool* (bo.expasy.org/tools/dna.html), onde também foram feitas predições de ORFs, as quais foram submetidas ao RPS-BLAST (*Reverse Position Specific BLAST*) contra o banco de dados de domínios conservados (*Conserved Domain Database*). Dados obtidos e melhores alinhamentos (com o respectivo *e-value*) foram anotados em planilhas para análise e comparação com dados disponíveis em fontes de literatura.

Resultados e Discussão

Genes de Resistência a Patógenos

Ainda são escassos os trabalhos prévios envolvendo genes de resistência em feijão-caupi, como fica evidente na Figura 1 que compara o número de genes preditos para planta-modelo *Arabidopsis thaliana* com aqueles disponíveis nos bancos genéticos públicos para os gêneros *Vigna* e seu gênero irmão *Phaseolus*.

As seqüências pré-existentes para o feijão-caupi consistem na sua maioria de produtos de PCR com primers desenvolvidos para domínios conservados destes genes, os chamados análogos de genes de resistência (RGAs, *Resistance Gene Analogs*), quase em sua totalidade relacionados com a classe III de genes *R* (NBS-LRR-TIR ou NBS-LRR-Coiled-Coil). Tais marcadores foram gerados por Timko (2000), justificando-se a predominância da classe III pelos domínios usados para desenho de primers, fato que justifica um maior número de seqüências deste grupo também em feijão *Phaseolus*.

No caso do feijão *Phaseolus*, o estudo revelou a existência de 35 seqüências para a classe II (que apresenta o domínio quinase). Para esta classe não foram até o momento geradas seqüências para o gênero *Vigna*, embora genes desta classe provavelmente sejam mais abundantes e diversificados em feijão-caupi que na planta-modelo *A. thaliana*. A forma como as seqüências foram obtidas (PCR com primers ancorados em regiões conservadas dos domínios) explicam a existência de seqüências parciais, com ORFs e domínios conservados incompletos, ressaltando a necessidade de mais estudos para um melhor entendimento da estrutura e evolução desses genes tanto no feijão-caupi como no feijão *Phaseolus*.

Genes Relacionados à Seca e à Salinidade

Em vegetais, os mecanismos genéticos de percepção que vêm sendo relacionados às respostas aos estresses hídrico e salino incluem: (I) atividades de quinases de histidina, envolvendo a proteína sensora EnvZ (*Envelope Z*) e o fator de transcrição OmpR (*Outer Membrane Protein*; Proteína Externa da Membrana; Shinozaki *et al.*, 1999); (II) as quinases da classe MAPK (*Mitogen-Activated Protein Kinase;* Quinase Ativada por Mitogen; Jonak *et al.*, 1996); (III) as quinases dependentes de cálcio (Urao *et al.*, 1994); (IV) a fosfolipase C, atuando no metabolismo de produção de inositol di e tri-fosfato, com função nas liberações de cálcio no citoplasma, onde o cálcio e calmodulina funcionariam como uma chave molecular na rota de transmissão do sinal de estresse (Munnik, *et al.*, 1998); (V) as proteínas DREB (*Dehydration Responsive Elements Binding Proteins*; Elementos Protéicos de Ligação em Resposta à Desidratação), que aderem às seqüências de

DNA presentes em regiões promotoras de genes expressos durante a desidratação (Kasuga *et al.*, 1999) e (VI) seqüências promotoras responsivas ao ácido abscísico (Abe *et al.*, 1997), entre outras.

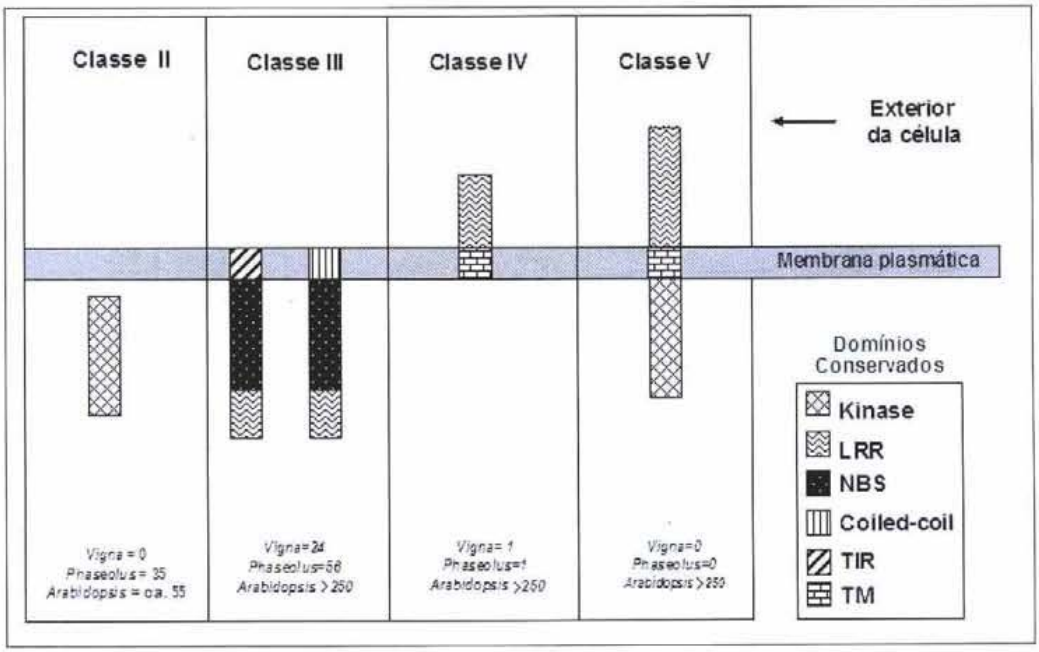


Figura 1: Painel dos principais genes de resistência (genes *R*) com seqüências (em geral parciais) depositadas no GenBank para o gêneros *Vigna*, *Phaseolus*, comparativamente à planta-modelo *A. thaliana*. Abreviações dos domínios conservados: LRR=*Leucine-Rich- Repeat*; NBS=*Nucleotide Biding Site*; TIR=*Toll Interleukine Region* e TM= *Transmembrane Region*. O esquema mostra, além dos domínios conservados que compõem cada classe de gene *R*, também sua posição subcelular mostrando que, de acordo com o domínio existente, a ação de reconhecimento gene-a-gene pode ocorrer em nível intra-celular (classe II), intra-celular com um domínio TIR ou *coiled-coil* associado à membrana (classe III), ou ainda com um LRR extracelular, associado a um domínio transmembrana (TM) associado (classe V) ou não (classe IV) a um domínio quinase.

A Figura 2 apresenta os principais grupos de genes considerados importantes na resposta aos estresses de seca/salinidade, indicando o número de seqüências de proteínas e nucleotídeos de feijão-caupi depositados no GenBank, comparativamente a *A. thaliana*.

Embora existam muitas classes gênicas com poucas (eventualmente nenhuma) seqüências de caupi descritas, observam-se outros grupos, onde há um número

significativo de seqüências disponíveis (como no caso dos osmoprotetores, com 15 membros), principalmente seqüências de amino-ácidos, derivadas de ensaios de expressão sob estresse de seca ou salinidade.

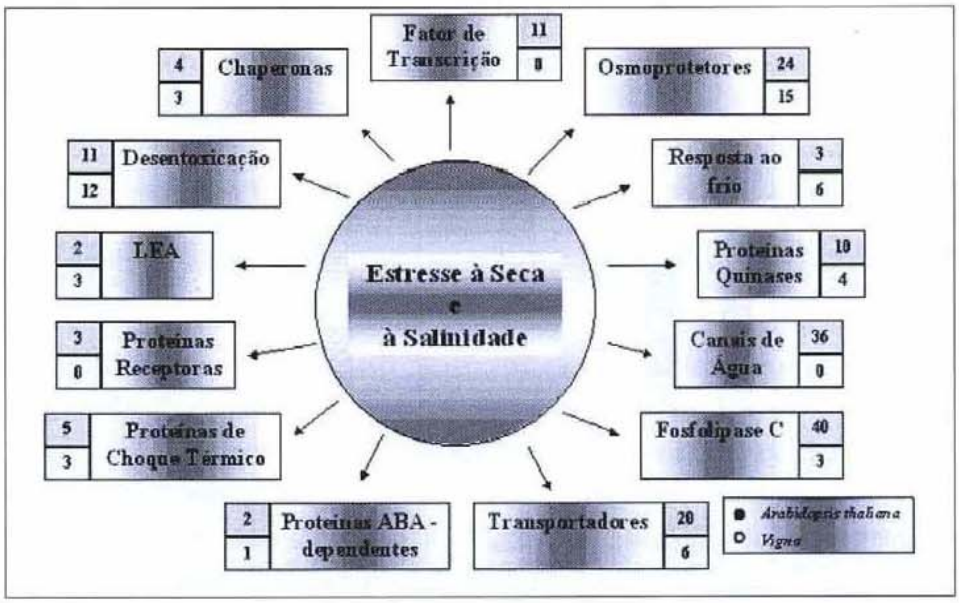
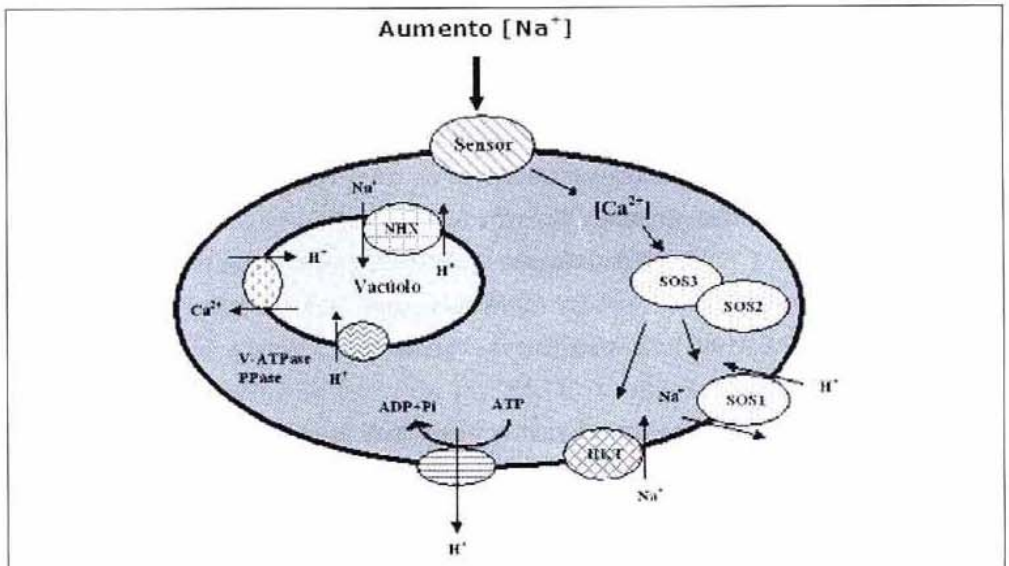


Figura 2. Principais classes de seqüências relacionadas ao metabolismo de sais ou estresse salino, de acordo com sua atividade ou mecanismo. Os valores na linha superior referem-se ao número de seqüências descritas em *A. thaliana*, comparativamente a espécies do gênero *Vigna* (em sua maioria *V. radiata*) avaliadas (linha inferior de cada classe).

Em plantas tolerantes ou sensíveis à salinidade o estresse inicia um complexo de respostas, começando com a percepção do estresse, a qual desencadeia uma cascata de eventos moleculares que induz vários níveis de respostas fisiológicas, metabólicas e de crescimento. Os genes da via SOS (*Salt Overly-Sensitive*), de *A. thaliana*, estão envolvidos na manutenção da homeostase celular. O estresse sódico inicia um sinal de cálcio que estimula o complexo SOS3/SOS2, que ativa o gene *SOS1*, o qual codifica uma proteína de antiporte Na^+/H^+, permitindo a exclusão do íon Na^+ e minimizando os efeitos da salinidade sobre a planta (Zhu, 2003).

Os componentes da via SOS são relativamente bem conhecidos em *A. thaliana*, sendo também encontrados no banco de dados do SUCEST (genoma expresso da cana de açúcar, *Saccharum officinarum*), com exceção do gene codificante da proteína NKH (vide Figura 3). Especificamente no caso das proteínas relacionadas à

salinidade, um maior número de informações sobre espécies do gênero *Vigna* (especialmente *V. radiata*), deve-se à existência de estudos de expressão e seqüenciamento de proteínas (especialmente em tecidos de raiz) sob estresse salino, uma vez que alguns genótipos do grupo são considerados como bons modelos de robustez e resistência sob condições adversas, mantendo por vezes a capacidade de nodulação mesmo em condições de salinidade (Elahi *et al.*, 2004). Apesar dos estudos de expressão e seqüenciamento protéico, com mais de 50 seqüências disponíveis, destaca-se a carência de dados em nível genômico ou relativos a ESTs (*Expressed Sequence Tags*) para o gênero.



Organismo/Gene	*SOS1*	*SOS2*	*SOS3*	*ATPase*	*HKT*	*NHX*	*Trocador H^+/Ca^{2+}*	*V-ATPase*	*V-PPase*
Arabidopsis thaliana	X	X	X	X	X	X	X	X	X
Saccharum officinarum	X	X	X	X	-	X	X	X	X
Vigna radiata	-	-	X	-	-	-	X	X	X

Figura 3. Representação esquemátca dos principais componentes da via SOS, indicando membros identificados em *A. thaliana*, *S. officinarum* e *V. radiata* (indicados com um X). Seqüências não encontradas em bancos genéticos públicos (GenBank para *Arabidopsis* e *V. radiata*) ou privados (SUCEST para cana), são indicadas por um "-".

No conjunto os dados demonstram haver ainda muito pouco conhecimento sobre os processos genéticos da resposta ao estresse hídrico/salino, bem como da resistência a patógenos em feijão-caupi. Considerando-se suas características

ímpares neste particular, espera-se que o seqüenciamento de bibliotecas geradas de indivíduos resistentes e susceptíveis na presença e na ausência de estresse forneçam informações valiosas para a cultura e para o melhoramento de outras leguminosas.

Agradecimentos:

Os autores agradecem ao programa Renorbio/MCT/BNB, à Facepe e à Fundação Volkswagen pelo apoio financeiro. Ao CNPq e à CAPES agradecemos pela concessão de bolsas de estudos.

Referências Bibliográficas

Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, Shinozaki K (1997) Role of *Arabidopsis* MYC and homologs drought-abscisic acid-regulated gene expression. Plant Cell 9:1859-1868

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

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

Ehlers JD, Hall AE (1997) Cowpea (*Vigna unguiculata* L. Walp.). Field Crops Res. 53:187-204

Ehlers JD, Hall AE (1998) Heat tolerant cowpea lines in short and long days. Field Crops Res. 55:11-21

Elahi NN, Mustafa M, Mirza JI (2004) Growth and nodulation of mungbean (*Vigna radiata* [L.] Wilczek) as affected by sodium chloride. J. Res. Sci. 15:139-143

Freire-Filho FR, Ribeiro VQ, Barreto PD, Santos CAF (1999) Melhoramento genético do caupi (*Vigna unguiculata* (L.) Walp.) na região nordeste. In: Queirós MA de, Goedert CO, Ramos SRR (eds), Recursos genéticos e melhoramento de plantas para o Nordeste Brasileiro. EMBRAPA, CPATSA, on line: http://www.cpatsa.embrapa.br (ISBN 85-7405-001-6), Petrolina

Hasegawa PM, Bressan RA, Zhu JK, Bohnert HJ (2000) Plant cellular and molecular responses to high salinity. Annu. Rev. Plant Physiol. Mol. Biol. 51:463-499

Jonak C, Kiegerl M, Ligterink W, Barker PJ, Huskisson NS, Hirt H (1996) Stress signaling in plants: a MAP kinase pathway is activated by cold and drought. Proc. Natl. Acad. Sci. USA 93:11274-11279

Kasuga M, Liu Q, Miura S, Yamaguchi- Shinozaki K, Shinozaki K (1999) Improving plant drought, salt and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nature Biotech. 17:287-291

Munnik T, Irvine RF, Musgrave A (1998) Phospholipid signaling in plants. Biochim. Biophys. Acta 1389:222-272

Oono Y, Seki M, Nanjo T, Narusaka M, Fujita M, Satoh R, Satou M, Sakurai T, Ishida J, Akiyama K, Lida K, Maruyama K, Satoh S, Yamaguchi-Shinozaki K, Shinozaki K (2003) Monitoring expression profiles of *Arabidopsis* gene expression during rehydration process after dehydration using ca 7000 full-length cDNA microarray. Plant J. 34:868-87

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

Seki M, Ishida J, Narusaka M, Fujita M, Nanjo T, Umezawa T, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K (2002) Monitoring the expression pattern of around 7,000 *Arabidopsis* genes under ABA treatments using a full-length cDNA microarray. Funct. Integr. Genomics 2:282-91

Shinozaki K, Yamaguchi-Shinozaki K, Liu Q, Kasuga M, Ichimar K, Mizoguchi T, Urao T, Miyata S, Nakashima K, Shinwari ZK, Abe H, Sakuma Y, Ito T, Seki M (1999) Molecular responses to drought stress in plants: regulation of gene expression and signal transduction. In: Smallwood MF, Calvert CM, Bowles DJ (eds), Plant Responses to Environmental Stress. Pp.133-143. BIOS Scientific Publisher, Oxford

Tai TH, Dahlbeck D, Clark ET, Gajiwala P, Pasion R, Whalen MC, Stall RE, Staskawicz BJ (1999) Expression of the Bs2 pepper gene confers resistance to bacterial spot disease in tomato. Proc. Natl. Acad. Sci. USA 96:14153-14158

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

Thomashow MF (1999) Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. Annu. Rev. Plant Physiol. Plant Mol. Biol. 50:571-599

Timko MP (2000) 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, Tamò M (eds), Challenges and opportunities for enhancing sustainable cowpea production. Pp.197-212. International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria

Urao T, Katagiri T, Mizoguchi T, Yamaguchi-Shinozaki K, Hayashida N, Shinozaki K (1994) Two genes that encode Ca^{2+}-dependent protein kinase are induced by drought and high-salt stress in *Arabidopsis thaliana*. Mol. Gen. Genet. 224:331-334

Zhu JK (2003) Regulation of ion homeostasis under salt stress. Curr Opin. Plant Biol. 6:441-445

6.0-CONCLUSÕES

- Tanto no transcriptoma da cana-de-açúcar quanto no do eucalipto, estão representadas as cinco classes principais de Genes de Resistência.

- Em cana-de-açúcar foram encontrados mais representantes da classe quinase, enquanto em eucalipto a classe NBS-LRR foi a mais abundante.

- Como previsto, genes da subclasse que inclui o domínio TIR não foram encontrados no transcriptoma da cana-de-açúcar, só foram encontrados em seqüências de candidatos a genes de resistência em Eucalipto.

- Um grande número de candidatos a genes *R* pode ser encontrado tanto em cana-de-açúcar quanto em eucalipto, mesmo com o baixo número de bibliotecas construídas com tecidos de plantas infectadas nos projetos SUCEST e FOREST; tal fato sugere que estes genes possuam uma expressão constitutiva, mesmo que em baixos níveis.

- No transcriptoma do eucalipto podem ser observadas combinações de domínios característicos de diferentes classes em determinados candidatos a genes de resistência, novos para a literatura, indicando que análises adicionais em grupos não amostrados (com ênfase para plantas lenhosas) podem revelar novas classes de genes e combinações de domínios.

- Organismos que pertencem à mesma família tendem a se agrupar nos dendrogramas gerados através do alinhamento de seus ortólogos, mesmo que a análise fenética considere apenas aspectos de similaridade e não de evolução.

- Os candidatos a genes *R* de cana-de-açúcar alinham melhor com monocotiledêneas, enquanto em eucalipto alinham melhor com genes de outras Dicotiledêneas.

- As seqüências identificadas neste trabalho representam uma ferramenta valiosa para o desenvolvimento de marcadores moleculares e também para a identificação de RGAs nas espécies estudadas bem como em outras plantas nativas e cultivadas.

7.0- APÊNDICE

7.1. Instruções para Autores

Genetics and Molecular Research (GMR) publishes research articles, research reports, technical notes,scientific commentaries, news, views, and review articles on Genetics, Evolution and Molecular Biology. It is an exclusively online journal.

Genetics and Molecular Research is maintained by a not-for-profit scientific foundation Ribeirão PretoFoundation for Scientific Research (FUNPEC-RP). There has been no institutional support for publicationsince 2003, so we are obliged to recover part of the costs of publishing. The Journal will bill authors for all "papers" accepted and submitted after February 1, 2005. The cost per accepted submission will be R$500.00 for Brazilian authors and US$250.00 for authors outside Brazil. This fee covers part of the expenses for language and technical revision, and for page setup, for publishing online in pdf and html.

Authors of all papers are expected to pay "page charges", which will be billed to the author forcorrespondence at the time the galley proofs of the paper are sent. Payment, both from within or outside Brazil, should be made by bank deposit or by credit card (Visa or Master Card). Please contact the editorial office (gmr@funpecrp.com.br) if you have any questions.

Contributions should be sent either by e-mail as attachments to gmr@funpecrp.com.br, or on disk by post to:
Prof. Dr. Francisco A. Moura Duarte, Editor
Av. Presidente Vargas, 2627 2º andar - Itamarati
14020-260 Ribeirao Preto, SP
BRAZIL

It is a fundamental condition that submitted manuscripts have not been published and will not be simultaneously published elsewhere. With the acceptance of a manuscript for publication, the publishers acquire full and exclusive copyright for all languages and countries.

The use of registered names, trademarks, etc., in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

An initial evaluation of the language will be made upon receipt of each manuscript. Those that are considered inadequate will be returned or sent out for correction, at the discretion of the author. The manuscript will be considered officially received when the corrected version is ready to be sent to the referees.

Before final acceptance, a submission letter with the title of the article and names and signatures of all the authors should be posted to the above address or faxed to the journal at 55 16 3621 1991.

GMR articles have no rigid length restrictions. They should contain sufficient technical detail for an expert reader to understand and assess the methods and results. There is no page limit for GMR articles. You should still be concise, however, for two main reasons. Firstly, our electronic refereeing system relies on e-mail, and very large files occasionally cause problems. Referees also tend to dislike lengthy manuscripts, and they take longer to process. Readers of electronic journals often print articles to read them. Remember that a 10,000- word article takes up around eleven pages. How many pages would you be willing to read on-screen? What length of article would you be prepared to print out?

Editorial policies: Genetics and Molecular Research is a refereed journal. Only original manuscripts will be considered for publication. Manuscripts will be reviewed by at least two independent reviewers before a decision is made on publication. The whole process is conducted electronically to speed progress and final publication. Papers will be published (placed online) within a few days after acceptance. Papers accepted in their final form from January 1 to March 31 constitute the first issue of each volume, and so on. There are four issues per year.

Manuscripts (in U.S. English), together with a cover letter from the author responsible for all correspondence, should be submitted to the Editor in electronic format as .doc files saved in Microsoft Word 97 for Windows, or later. Do not use formatting such as Word's "Heading" or "Style Sheets". Spelling, punctuation, sentence structure, spacing, length, and consistency of usage in form and descriptions should be checked before

submission. Please also check references for accuracy. Ensure that all figures and tables are mentioned in the text, and that all references are cited in the text. Figure files (see below) should be separate.

Submission information:

Authors are required to provide the following information with their electronic submissions:
Article submitted by
Article title
Authors (full list)
Article type and session
Status of article (e.g., new, revised, etc.)
Manuscript number (provided at the first submission)
Postal address
E-mail address
Phone number
Fax number
Names and types of the files sent

* Brazilian authors should not translate their institutional addresses. These should remain in the original (Portuguese) language.

Revised Versions
If you are submitting a revised version of your article then please remember to include a list of changes, and replies to the referees. You should also send all the files for the article, not just those you have revised.

Acknowledgment of electronic submissions.
Successful receipt and processing of your submission will be acknowledged by e-mail when your submission has been checked. If you have no response within one week you should contact the editor atgmr@funpecrp.com.br.

Articles are reviewed anonymously by independent referees. Authors are encouraged to suggest names of expert reviewers, but selection remains a prerogative of the editors. To facilitate the review process, the authors can send supplementary material, such as cited accepted but not published papers, which may be important for assessment of the manuscript.

Order the sections comprising the manuscript as follows: title, running title, author, address, abstract, key words, introduction, material and methods, results, discussion, acknowledgments, and bibliography.
Title Page: The title page should include the title of the article, author's names with full first names (no degrees), and author's affiliation. The affiliation should comprise the department, institution (usually university or company), city, and state (or nation). The title page should include the name and complete mailing address, telephone number, fax number, and e-mail address of the author designated to review proofs. The title page should start below the top margin, be single-spaced, and no space left before the Summary/Abstract. Provide a running title of no more than 60 characters (including spaces).

Abstract: An abstract of fewer than 250 words, single-spaced, is required of research articles and reports andshould be arranged in one paragraph. The following information (without headings) should be included: purpose, methods, results, and conclusions. Review articles also require an abstract, which need not include the same items.

Key words: A list of key words or indexing terms (no more than 6) should be included.

Text
Format: Headings should be bold, first letters capitalized and left aligned. All text should be set in Times New Roman font, 12 point, left aligned, single spaced. Do not justify the right margin. Leave only one (1) space after periods. Paragraphs should not be indented, nor should there be any blank lines between them. Use line returns only at the end of paragraphs. Do not use tabs or spaces to create indents. Use the Symbol font for symbols and special characters. Do not use equation editors or footnoting utilities. Save equations as
images. Equations should be numbered consecutively with Arabic numerals in parentheses on the right hand side of the page.

Footnotes. Footnotes should be avoided. When their use is absolutely necessary, footnotes should benumbered consecutively using Arabic numerals and should be placed at the bottom of the page to which they refer. Place a line above the footnote, so that it is set off from the text.

Tables/Charts. Special care should be taken to ensure that all Tables are properly formatted. Scientific symbols used should be in 'Symbol' or 'Times New Roman'. Tables should be on a separate page, numbered consecutively (with Arabic numerals), referred to by number in the text and designed to fit the column or page size of the Journal. Use tables with cells to separate columns. Do not use spaces, tabs or vertical lines. Left justify the title above the table (bold font). Indicate each table's location within the manuscript.

Illustrations. Illustrations/figures (photographs, drawings, diagrams, and charts) should each be in a single file, numbered in one consecutive series of Arabic numerals in the order in which they are cited in the text. Illustrations must be submitted as separate files. All illustrations are to be supplied in JPEG (jpg) format in either color or black and white. Images must be saved as separate, stand-alone files. The image resolution should be 300 dpi. Do not embed images within the text file. The

placing of graphics in the paper should be indicated in the text and should include the captions for the figures. The authors should also send, by mail, a printed version of the figures. These should be at least 10 x 15 cm, up to size A4, so that figures can be scanned to guarantee good quality for publishing online.

Reference style: References in the text should include the name of the author and the year in parentheses, e.g. (Searle, 1961) or (King and Wilson, 1975). When a reference with more than two authors is cited, only the first author is named, e.g. (Comstock et al., 1958). The references must be cited, in the text, in chronological order, e.g., (Ideber, 2001; Uetz, 2002; Ottavai, 2004). References to "unpublished results" and "submitted papers" should appear in the text in parentheses following the individual name(s). Example: (Pereira KS, Martins PK and Silva TM, unpublished results).

Abbreviations: Try to use abbreviations sparingly. When used extensively, provide a list of all nonstandard abbreviations on a separate page before the reference section. Use the metric system for all measurements without periods (cm, mL, s). Define all symbols used in equations and formulas. Do not abbreviate the word "Figure" or "Table" in titles or text.

Acknowledgments: All acknowledgments (including those for grant and financial support) should be typed in one paragraph directly preceding the reference section. Authors of manuscripts submitted to GMR arerequested to state the source of all funding that enabled the described research to be undertaken.

Bibliography: The bibliography should include only works referred to in the text. It should be arranged in alphabetical order under the first author's last name. References should be cited as follows: journal papers -names and initials of the four first authors (after that use et al.), year, full title, journal abbreviated according to Index Medicus, volume number, first and last page numbers; books - names of authors, year, full title, edition, publishers, address (city, state and country); articles published in symposia - names of authors, year, full title of book, name(s) of editor(s) in parentheses, publisher, address (city, state and country), first and last page numbers.

Examples of Bibliographic style:
Journal article
Searle SR (1961). Variance components in the unbalanced 2-way nested design. *Ann. Math. Statist*. 32:1161-1166.

Comstock RE, Kellcher T and Morrow EB (1959). Genetic variation in an asexual species, the garden strawberry. *Genetics* 43: 634-646.

Book
Mather K (1949). *Biometrical Genetics*. 1st edn. Methuen, London, England.

Chapter in book
Rhoades MM (1968). Studies on the cytological basis of crossing over. In: *Replication and Recombination of Genetic Material* (Peacock WJ and Brock RD, eds.). Australian Academy of Science, Canberra, Australia, pp. 229-241.

Thesis abstracts
Thesis abstracts should be submitted in English and in the original language (if the author is not a native speaker). Bibliographic data (original title, year, institution, address, number of pages and major (orienting professor) should be included.
Authors using GMR may view, reproduce or store copies of the Journal providing the information is not used for profit purposes.

Published by the Ribeirão Preto Foundation for Scientific Research (FUNPEC-RP). http://www.funpecrp.com.br

This page last updated on September 2, 2005

7.2- Aceite do Artigo

Wanderley-Nogueira AC, Mota N, Lima-Morais D, Silva LCB, Silva AB, Benko-Iseppon AM (2006). **Abundance and Diversity of Resistance (R) Genes in the Sugarcane Transcriptome**. Genetics and Molecular Research. Submetido Julho/2006, aceito Abril/2007.

----- Original Message -----
From: "Roberto M Cesar-Jr" <Roberto.Cesar@vision.ime.usp.br>
To: "Ana Maria Benko Iseppon" <celisep@hotlink.com.br>
Cc: "Roberto M. Cesar-Jr" <cesar@vision.ime.usp.br>; <santoro@icb.ufmg.br>; <miguel@icb.ufmg.br>
Sent: Wednesday, April 18, 2007 3:24 PM
Subject: Final instructions: x-meeting 06 paper and GMR

> Dear author,
>
> Congratulations! Your paper has been accepted to be published in GMR.
> Send the final version of your paper by email to
> roberto.cesar@vision.ime.usp.br The journal will then contact you for
> further details.
>
> Please send the final version of your paper within at your earliest
> convenience (within the next few days, if possible).
>
> As indicated in the original CFP
> (http://biotec.icb.ufmg.br/cyberchair/html/), there is a cost defined
> by the Journal of R$500.00 for Brazilian authors and US$250.00 for
> authors outside Brazil. Please explicitly confirm by email that you
> are aware of it and that the cost regarding your paper will be payed
> to the journal.
>
> Best wishes,
>
> Marcelo Matos Santoro (santoro@icb.ufmg.br)
> Roberto M. Cesar-Jr (Roberto.Cesar@vision.ime.usp.br)
> Special Issue Editors - GMR - X-meeting 06
>
> ==========
> Roberto Marcondes Cesar Junior - http://www.ime.usp.br/~cesar
> Department of Computer Science - IME - USP - Brazil
> ==========

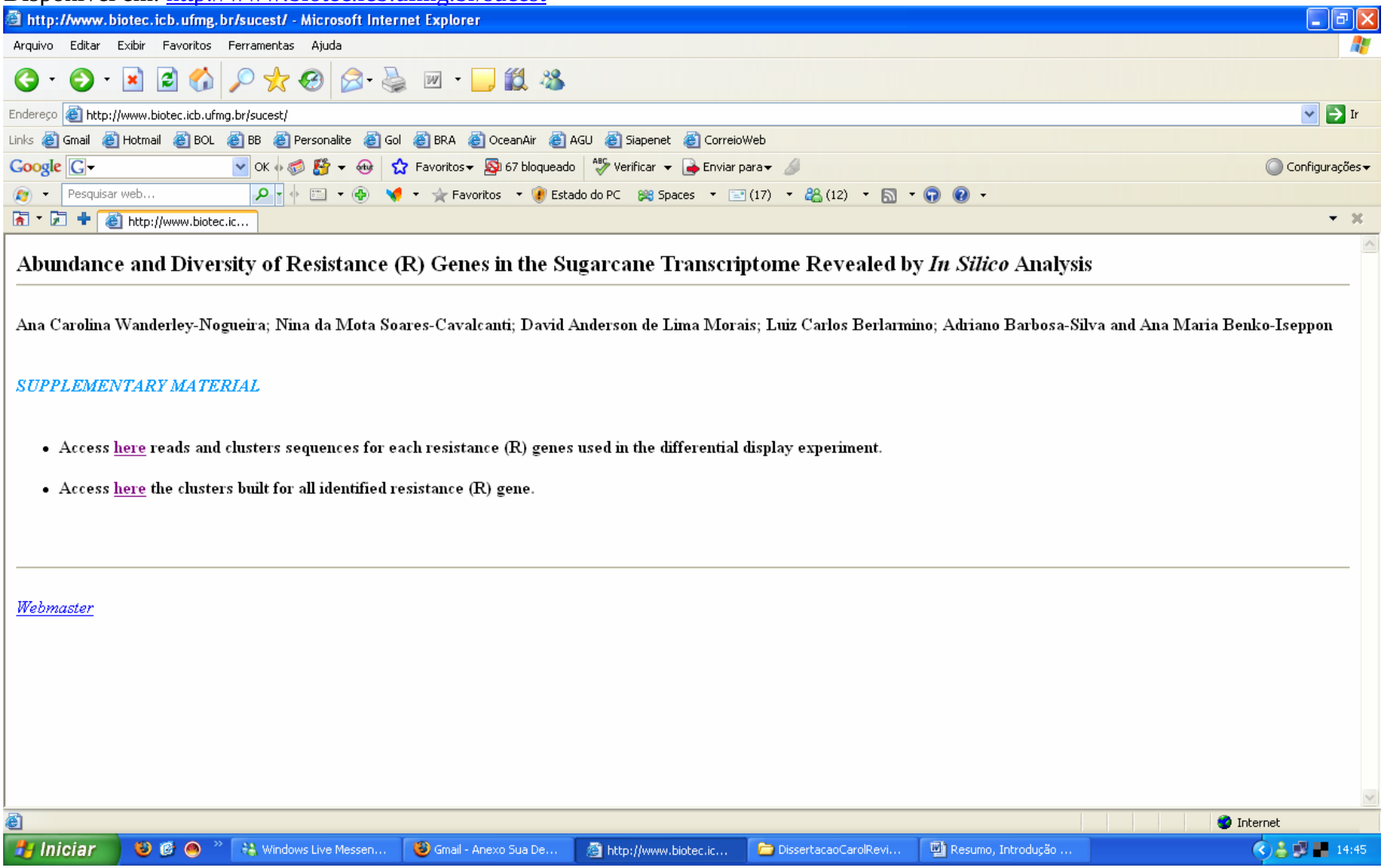
7.3 Anexo 1

Classification and features of *R*-genes seed sequences used as query against the SUCEST database. The used genes are grouped in five *R*-gene classes with respective accession number at NCBI, source species, gene name and domain range (in amino-acids).

R-Gene Class	Accession No.	Source species	Gene	Sequence size (aa)	Domain range (initial-last aa)							
					LRR		KINASE		NBS		TIR	
					Start	End	Start	End	Start	End	Start	End
I	2112354A	*Lycopersicon esculentum*	*Pto*	321	-	-	41	236	-	-	-	-
II	AF234174_1	*Arabidopsis thaliana*	*HRT*	909	579	868	-	-	150	460	-	-
	NP_172686.1	*Arabidopsis thaliana*	*Rps5*	889	540	636	-	-	140	444	-	-
	AF118127_1	*Lycopersicon esculentum*	*I2*	1266	578	1231	-	-	154	457	-	-
	AAG31014.1	*Lycopersicon esculentum*	*Sw5*	1246	-	-	-	-	519	818	-	-
	BAA25068.1	*Oryza sativa*	*Xa1*	1802	771	1773	-	-	283	593	-	-
	BAA76282	*Oryza sativa*	*Pib*	1251	930	1210	-	-	390	712	-	-
	AAK00132	*Oryza sativa*	*Pi-ta*	928	712	925	-	-	207	503	-	-
	AAP81262.1	*Zea mays*	*Rp1*	1269	596	1228	-	-	145	457	-	-
	AC016827_19	*Arabidopsis thaliana*	*RPM1*	926	559	665	-	-	173	467	-	-
III	AAC72977.1	*Arabidopsis thaliana*	*RPP1*	1189	668	1011	-	-	226	505	54	184
	RP13_ARATH	*Arabidopsis thaliana*	*RPP13*	835	-	-	-	-	147	453	14	148
	AF440696_1	*Arabidopsis thaliana*	*RPP4*	1135	642	1053	-	-	185	441	15	145
	AAF08790.1	*Arabidopsis thaliana*	*RPP5*	1361	643	1151	-	-	188	465	14	148
	RPP8_ARATH	*Arabidopsis thaliana*	*RPP8*	908	577	867	-	-	149	459	15	145
	BAB11393.1	*Arabidopsis thaliana*	*Rps4*	1232	663	889	-	-	198	473	21	149
	AAP41025.1	*Lactuca serriola*	*RGC2*	352	49	235	-	-	-	-	21	149
	AF093649_1	*Linum usitatissimum*	*L*	1294	607	1277	-	-	220	521	63	195
	T18548	*Linum usitatissimum*	*M*	1305	744	1288	-	-	235	534	78	210
	AF310960_2	*Linum usitatissimum*	*P*	1211	693	1023	-	-	205	238	23	153
	AF202179_1	*Capsicum chacoense*	*Bs2*	905			-	-	152	439	63	195
	A54810	*Nicotiana glutinosa*	*N*	1144	597	908	-	-	172	447	14	147
	AF195939_1	*Solanum tuberosum*	*Gpa2*	912	561	863	-	-	119	422	14	147
	CAA61264.1	*Solanum tuberosum*	*Rx1*	248			-	-	-	-	23	153
	CAB56299.1	*Solanum tuberosum*	*Rx2*	938	561	859	-	-	138	422	78	210
	CAD29728.1	*Solanum tuberosum*	*HERO*	1283	-	-	-	-	504	811	54	184
IV	T07015	*Lycopersicon esculentum*	*Cf4*	855	81	758	-	-	-		-	-
	AAC78591.1	*Lycopersicon esculentum*	*Cf5*	968	96	855	-	-	-	-	-	-
V	AAC80225	*Oryza longistaminata*	*Xa21*	1025	83	333	708	922	-	-	-	-

6.4 Homepage com material suplementar

Disponível em: http://www.biotec.icb.ufmg.br/sucest



Nogueira, Ana Carolina Wanderley

Caracterização de genes de resistência a patógenos em eucalipto (*Eucalyptus ssp.*), cana-de-açúcar (*Saccharum ssp.*) e feijão-caupi (*Vigna unguiculata*). / Ana Carolina Wanderley Nogueira. – Recife: O Autor, 2007.

103 folhas : il., fig. e tab.

Dissertação (mestrado) – Universidade Federal de Pernambuco. CCB. Genética, 2007.

Inclui bibliografia, anexo e apêndice.

1. Fitopatologia 2. Genes de resistência – Eucalipto 3. Genes de resistência – Cana-de-açúcar 4. Genes de resistência – Feijão-caupi I. Título.

581.2. CDU (2.ed.) UFPE

581.3 CDD (22.ed.) CCB – 2007-129