

IDENTIFICAÇÃO E ANÁLISE “IN SILICO” DE RECEPTORES LRR-RLK E
NB-LRR DE CANA-DE-AÇÚCAR INDUZIDOS DURANTE A
COLONIZAÇÃO POR BACTÉRIAS ENDOFÍTICAS

LEANDRO DE MATTOS PEREIRA

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY RIBEIRO
CAMPOS DOS GOYTACAZES – RJ

Dezembro- 2009

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

IDENTIFICAÇÃO E ANÁLISE “IN SILICO” DE RECEPTORES LRR-RLK
E NB-LRR DE CANA-DE-AÇÚCAR INDUZIDOS DURANTE A
COLONIZAÇÃO POR BACTÉRIAS ENDOFÍTICAS

LEANDRO DE MATTOS PEREIRA

Dissertação apresentada ao Centro de Biociências e Biotecnologia da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Mestre em Biociências e Biotecnologia.

Orientador: Prof. Gonçalo Apolinário de Souza Filho

CAMPOS DOS GOYTACAZES – RJ

Dezembro de 2009

IDENTIFICAÇÃO E ANÁLISE “IN SILICO” DE RECEPTORES LRR-RLK E
NB-LRR DE CANA-DE-AÇÚCAR INDUZIDOS DURANTE A
COLONIZAÇÃO POR BACTÉRIAS ENDOFÍTICAS

LEANDRO DE MATTOS PEREIRA

Dissertação apresentada ao Centro de Biociências e Biotecnologia da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Mestre em Biociências e Biotecnologia.

Aprovada em 14 de dezembro de 2009.

Comissão Examinadora:

Prof. José Roberto da Silva (D.Sc. Biociências e Biotecnologia) – UFRJ

Prof. Fábio Lopes Olivares (D. Sc. Agronomia) - UENF

Prof. Victor Martin Quintana Flores(D. Sc. Engenharia Genética) - UENF

Prof. Gonçalo Apolinário de Souza Filho (PhD em Genética) - UENF
Orientador

AGRADECIMENTOS

À Deus pelo dom da vida e pela imensa dádiva pelo qual me presenteou:
à curiosidade e a busca insaciável por conhecimento.

Ao meu orientador Gonçalo, pela oportunidade, pelo aprendizado no âmbito profissional e pessoal, pela amizade, pelos conselhos, por toda experiência passada.

Aos meus amados pais, Waldelírio e Leni. Pelo exemplo de vida e apoio. Agradeço de maneira especial a minha mãe por inconscientemente me inspirar na busca incansável pelo conhecimento, e por depositar em mim a educação que nunca teve.

Á meu irmão Leonardo Mattos pelos incentivos, conselhos, amizade e por sempre acreditar nos meus projetos e sonhos.

As amigas Beatriz Ferreira e Aline Intorne pelo apoio durante o curso, pelos ensinamentos, pelos incentivos, pela convivência e cumplicidade.

Ao meu amigo Carlos Eduardo Monteiro, pela convivência, amizade, companheirismo e cumplicidade.

Ao meu amigo Ronam Facini e Emílio C. Miguel pela convivência diária, amizade, pelo apoio nos momentos de dificuldades e pelos momentos de alegria proporcionados.

Aos amigos, Bruno Esteves, Wellington Campos, David Gitirana, Rafael Jardim, Edson de Jesus pela amizade e pelos momentos compartilhados.

A amiga Hérica Chagas pela amizade e apoio nos momentos de dificuldades durante o curso.

Aos amigos Saulo Pireda, Pablo e Ricardo, pela amizade, companheirismos e momentos de alegria compartilhados.

Aos meus amigos de infância Liomar, Claudiney, Rodolfo, Glécio (*em memória*), meus amados tios e tias (José Miguel, Regina, José Luís, Lemir, Waldeli, Élbio e Inês), primos e primas amados (Fernanda, Fabiano, Fabiana, Lucas, João Paulo, Luciano, Juliana, Flaviana, Michelle), por fazerem parte da minha história e pelos momentos de felicidade e aprendizado o qual me proporcionaram.

Aos amigos da UENF Ciro Morais, André Guimarães, José Paulo Mota, Felipe Chaves, Márcio Chaves e amigas Fernanda Bueno, Mariana Barduco, Vanessa Iglesias, Larissa Rangel e Viviane pela convivência, amizade e alegrias compartilhadas durante o curso.

Aos meus colegas de Laboratório Janice, Roberta, Juliana, Patrícia, Valéria, Cristina, Freed, Nathália, Bruno (“maisena”), Sílvio, Maria Clara, Edson, Fabrício, Rafael, Rívea, Telma e Adão pela presença, convivência e ajuda no dia-a-dia.

Aos professores Vanildo, Medina, Victor, Marília, e Tânia, pela amizade e ensinamentos. Aos professores Fábio Olivares, Vitor Flores, José Roberto da Silva e Leonardo Gomez da Silva por aceitarem o convite de fazer parte da banca examinadora, pelas correções, críticas e sugestões. Ao revisor da dissertação e suplente interno da banca Dr. André O. Carvalho pelas correções, sugestões, críticas e disponibilidade. A todos os amigos e professores, que embora eu não tenha citado contribuíram de maneira fundamental para minha formação profissional e realização deste trabalho.

A fonte financiadora Capes pela bolsa concedida durante o curso, que foi indispensável para o apoio logístico.

Quadro 1: Motivos conservados presentes na região NBS de proteínas CC-NB-LRR de <i>A. thaliana</i>.....	13
Quadro 2 : Lista descritiva das Bibliotecas de cDNA do SUCEST.....	28
Quadro 3: Relação dos grupos correspondentes aos tecidos de cana-de-açúcar obtidos da reunião das bibliotecas de cDNA de cana-de-açúcar do Projeto SUCEST.....	30
Quadro 4: Proteínas LRR-RLKs selecionadas no NCBI.....	31
Quadro 5: Domínios característicos de proteínas LRR-RLKs.....	32
Quadro 6: Proteínas NB-LRR selecionada no banco de dados NCBI.....	33
Quadro 7: Domínios característicos de proteínas NB-LRR.....	34
Quadro 8- Transcritos similares à proteína LRR-LRKs encontrados em cana-de-açúcar e proteínas similares em <i>Arabidopsis thaliana</i>.....	39
Quadro 9: Transcritos identificados em cana-de-açúcar que codificam proteínas LRR-RLKs e proteínas similares em <i>Arabidopsis thaliana</i>.....	41
Quadro 10: Similaridade entre as sequências ScPGIPs like 1 de cana-de-açúcar e as sequências PGIP 1 caracterizadas em outras espécies vegetais.....	42
Quadro 11: Comparação entre a sequência em aminoácidos do receptor ScMAR 1 de cana-de-açúcar e as sequências similares identificadas em outras espécies vegetais.....	53
Quadro 12: Comparação entre a sequência em aminoácidos de ScFLS2 de cana-de-açúcar as sequências similares identificadas em <i>Sorgum bicolor</i>, <i>Arabidopsis thaliana</i> e <i>Oriza sativa</i>.....	59
Quadro 13: Comparação entre as sequências ScSERK like 1 e ScSERK like 2 identificadas cana-de-açúcar com membros da família SERK presentes em diferentes espécies vegetais.....	64
QUADRO 14: Sequências NBS-LRR-ARC encontradas em cana-de-açúcar.....	67
Quadro 15: Nível de similaridade obtido entre as sequências NB-LRRs de cana-de-açúcar e a sequências similares em <i>Arabidopsis thaliana</i>.....	68

Quadro 16: Nível de similaridade obtido através de alinhamento no programa ClustalW 2 entre as sequências ScADR1 e as sequências similares identificadas em outras espécies vegetais.....72

LISTA DE FIGURAS

Figura 1: Relação filogenética das proteínas CC-NB-LRR encontrados no genoma de *Arabidopsis thaliana*12

Figura 2: Análise da expressão de genes que codificam proteínas LRR-RLKs através de “Northen eletrônico” 44

Figura 3: Análise da expressão de genes que codificam proteínas LRR-RLKs através de “Northen eletrônico” 49

Figura 4: Motivos conservados no domínio cinase de MAR1 em diferentes espécies..... 56

Figura 5: Relações filogenéticas das proteínas ScFLS2, ScSERK - like 1 e ScSERK - like 2 e ScMAR1..... 58

Figura 6. Alinhamento múltiplo dos domínios tirosina cinases dos homólogos de ScFLS2 em diferentes espécies..... 61

Figura 7: Relações filogenéticas da família SERK, incluindo ScSERK-like 1 e ScSERK-like 2..... 65

Figura 8: Análise da expressão de genes que codificam proteínas CC-NB-LRR em cana-de-açúcar através de “Northen eletrônico”71

Figura 9: Alinhamento Múltiplo da região N-terminal na família da proteína ADR1 em *Arabidopsis thaliana* e seus homólogos em outras espécies vegetais, incluindo cana-de-açúcar..... 74

Figura 10: Relação filogenética da família ADR1 em plantas, incluindo ScADR1..... 79

LISTA DE ABREVIATURAS

ACC oxidase	<i>1-aminocyclopropane-1-carboxylic acid oxidase</i>
ADP	5' difosfato de adenosina
Apaf1	<i>apoptotic peptidase activating factor 1</i>
ATP	5' trifosfato de adenosina
BR1	<i>brassinosteroid insensitive 1- associated receptor kinase 1</i>
BAK1	<i>BR1- associated receptor kinase</i>
CC	<i>coiled coils</i>
CED4	<i>cell death abnormality 4</i>
DNA	acido desoxiribonucléico
DAMPs	<i>damage associated molecular pattern molecules</i>
EFR	<i>ef-tu receptor</i>
EFTu	<i>elongation factor-tu</i>
ENODS	<i>early nodulins</i>
EST	<i>expressed sequence tag</i>
FLS2	<i>flagellin-sensitive 2</i>
GTP	5' trifosfato de guanosina
HR	<i>hypersensitive response</i>
LRR-RLK	<i>leucine-rich repeat receptor-like kinase</i>
LYSM	<i>lysin</i>
LYK3	<i>lysM receptor kinase 3</i>
LYK4	<i>lysM receptor kinase 4</i>
MAMPs	<i>microbe associated molecular patterns</i>
mRNA	RNA mensageiro
N ₂	di-nitrogênio
NH ₃	amônia
NFP	<i>nod factor perception</i>
NFR1	<i>nod-factor receptor 1</i>
NFR5	<i>nod-factor receptor 5</i>
NSP1	<i>nodulation signaling pathway 1</i>

NSP5	<i>nodulation signaling pathway 5</i>
NB-LRRs	<i>nucleotide binding leucine-rich repea</i>
NB-ARC	<i>nucleotide binding-apoptosis repressor with card</i>
RIN4	<i>RPM1 interacting protein 4</i>
RNA	ácido ribonucléico
ROS	espécies reativas de oxigênio
RPS2	<i>resistant to Pseudomonas syringae 2</i>
RPS4	<i>resistant to Pseudomonas syringae 4</i>
RPS5	<i>resistant to Pseudomonas syringae 5</i>
RPP8	<i>recognition of Peronospora parasítica 8</i>
RPP13	<i>recognition of Peronospora parasítica 13</i>
RT-PCR	<i>reverse transcription-polymerase chain reaction</i>
RT-qPCR	<i>quantitative real time polymerase chain reaction</i>
PRR	<i>pattern recognition receptor</i>
PCR	<i>polimerase chain reaction</i>
SERK	<i>somatic embryogenesis receptor kinase</i>
SUCEST	<i>sugarcane est project</i>
SYMRK	<i>symbiosis receptor-like kinase</i>
SMART	small modular architecture research tool

Resumo

Bactérias endofíticas têm sido isoladas de diferentes plantas, entre elas: arroz, milho e cana-de-açúcar. Essas bactérias são capazes de promover o crescimento vegetal através de fixação biológica de nitrogênio, produção de fitormônios e outros mecanismos. Em cana-de-açúcar o estudo da interação com as bactérias *G. diazotrophicus* e *H. rubrisubalbicans* tem recebido especial atenção. Entretanto, os mecanismos envolvidos nesta associação ainda não foram elucidados. As plantas são capazes de reconhecer moléculas associadas a microrganismos e ativar seu sistema de defesa. Esse reconhecimento é mediado por receptores LRR-RLK e proteínas CC-NB-LRR. Contudo, microrganismos patogênicos têm a habilidade de subverter esse sistema e favorecer a colonização da planta. Análises “in silico” evidenciam a ativação de genes de defesa em cana-de-açúcar durante a colonização por *G. diazotrophicus*, mas sintomas de doença não são observados. O presente trabalho teve como objetivo identificar em cana-de-açúcar, proteínas NB-LRRs e LRR-RLKs, e avaliar sua ativação durante a associação com endófitos. Como resultado, análises “in silico” permitiram identificar as proteínas receptoras, que foram denominadas ScSERK-like 2, ScMAR1 e ScLRR-RLK2. Tais proteínas são induzidas em plantas inoculadas com os endófitos supracitados. O alinhamento de seqüências e análises filogenéticas indicam que ScSERK-like 2 é um membro da família SERK. Os transcritos da proteína ScMAR1 são altamente induzidos por *H. ubrisubalbicans*, e seu ortólogo em *A. thaliana* está envolvido na resposta de defesa. Transcritos da proteína ScLRR-RLK2 são preferencialmente induzidos por *G. diazotrophicus*, apresentando similaridade com os receptores LTRK1 (*Z. mays*) e SFR7(*A. thaliana*) envolvidos na divisão e morfogênese celular, respectivamente. No âmbito das proteínas NB-LRRs, foi identificada a proteína ScADR1 (CC-NB-LRR) que é fortemente induzida por *G. diazotrophicus*. Tal proteína é altamente conservada em diferentes espécies, e seu ortólogo (ADR1) em *A. thaliana* medeia a resistência contra microrganismos patogênicos. Juntos, os dados permitiram sugerir a ocorrência de ativação dos mecanismos de defesa da planta, através da transcrição de receptores, durante a colonização por bactérias endofíticas. Tal resposta mostrou-se específica para cada espécie bacteriana analisada.

ABSTRACT

Endophytic bacteria have been isolated from different plants, including rice, corn and sugarcane. These bacteria are able to promote plant growth through nitrogen fixation, production of hormones and other mechanisms. In sugarcane the study of the interaction with the bacteria *G. diazotrophicus* and *H. rubrisubalbicans* has received special attention. The mechanisms involved in this association have not been elucidated yet. Plants are able to recognize molecules associated with microorganisms and activate its defense system. This recognition is mediated by receptors LRR-RLK and CC-NB-LRR. However, pathogenic microorganisms have the ability to subvert the system defense and promote the colonization of plants. Analyses “in silico” show the activation of defense genes in sugarcane during the colonization by *G. diazotrophicus*, but disease symptoms are not observed. This study aimed to identify in sugarcane proteins LRRs-RLKs and NB-LRR and evaluate their activation during the association with endophytes. As a result, analysis “in silico” allowed identify the receptors proteins, which were named ScSERK-like 2, ScMAR1 and ScLRR-RLK2. These proteins are induced in plants inoculated with endophytes. Alignment and phylogenetic tree indicate that ScSERK-like 2 is a member of family SERK. Transcripts of the protein ScMAR1 are highly induced by *H. rubrisubalbicans*, its orthologo in *A. thaliana* is involved in defense response. Transcripts of protein ScLRR-RLK2 are preferentially induced by *G. diazotrophicus*, showing similarity with the receptors LTRK1 (*Z. mays*) and SFR7 (*A. thaliana*) involved in cell division and morphogenesis, respectively. Within the NB-LRR, the protein ScADR1 was identified (CC-NB-LRR) which is highly induced by *G. diazotrophicus*. The protein ScADR1 is highly conserved in different species, the orthologo (ADR1) in *A.thaliana* mediates resistance against pathogenic microorganisms. Together, the date suggests the occurrence of activation of plant defense mechanisms, through of the transcription of receivers during colonization by endophytic bacteria. This response was shown to be specific for each species of bacteria analyzed.

Índice	
Página de capa	I
Página de rosto	II
Página de aprovação	III
Agradecimentos	IV
Lista de Quadros	VI
Lista de Figuras	VII
Lista de Abreviaturas	VIII
RESUMO	X
ABSTRACT	XI
Índice	XII
1-INTRODUÇÃO	01
2- REVISÃO DA LITERATURA	03
2.1 - Sistema imune em plantas.....	03
2.2– Subversão da resposta imune por bactérias patogênicas.....	06
2.3 - Teoria gene a gene.....	07
2.4- As proteínas NB-LRRs e mecanismos de sinalização	09
2.5 – Classificação de proteínas NB-LRRs e motivos conservados.....	11
2.6 – Os receptores LRR-RLKs em plantas	14
2.7 – Interação entre plantas e microorganismos simbiotes.....	15
2.8 – A cana-de-açúcar e endófitos.....	19
2.9- O Projeto SUCEST.....	22
3- Banco de Dados	23
3.1- “Northen Eletrônico e bancos de dados de expressão gênica.....	25
4 – OBJETIVO GERAL	26
4.1 - Objetivos específicos.....	26
5– MATERIAL E MÉTODOS	27
5.1 – Utilização das bibliotecas de cDNA do SUCEST.....	28

5.2 – Geração de Dados de expressão gênica.....	29
5.3 – Identificação de genes de cana-de-açúcar que codificam proteínas RLKs induzidas durante a interação com endófitos.....	31
5.4 - Identificação de genes de cana-de-açúcar que codificam proteínas NB-LRR reguladas por endófitos.....	33
5.5 - Comparações entre a sequência dos receptores LRR-RLKs identificados em cana-de-açúcar com sequências similares em outras espécies vegetais.....	35
5.6 – Análises filogenéticas dos receptores LRR-RLKs identificados em cana-de-açúcar.....	35
5.7 – Identificação de sequências similares à ADR1 em diferentes espécies vegetais e análises filogenéticas.....	36
5.8- Análise de domínios presentes nas proteínas e busca de motivos.....	37
6 - RESULTADOS	38
6.1- “Clusters” que codificam proteínas “LRR-RLKs” identificados no banco de dados SUCEST.....	38
6. 2 – “Northen Eletrônico” dos receptores LRR-RLKs identificados em cana-de-açúcar.....	43
6. 3 - Comparação entre as sequências de aminoácidos das proteínas LRR-RLKs de cana-de-açúcar com sequências similares identificadas em outras espécies vegetais.....	53
6. 3.1 – Comparação entre a sequência de ScMAR1 de cana-de-açúcar com sequências similares em outras espécies vegetais.....	53
6.4 – Análise comparativa entre a sequência de ScFLS2 e as sequências similares em outras espécies vegetais.....	60
6. 5 – Análise comparativa das sequência em aminoácidos de ScSERK like 1 e ScSERK like 2 com as proteínas SERKs encontradas em outras espécies vegetais.....	63
6. 6 - “Clusters” que codificam proteínas com domínios CC-NB-LRR induzidos por <i>Gluconacetobacter diazotrophicus</i> e/ou <i>Herbaspirillum rubrisubalbicans</i> ..	66

6.7 - “Northen Eletrônico” das proteínas CC-NB-LRR identificadas no SUCEST.....	69
6.8 – Análise comparativa entre a sequência de aminoácidos de ScADR1 e as sequências similares em outras espécies vegetais.....	72
6.9 - Relação filogenética da proteína ScADR1 com sequências similares em outras espécies vegetais	78
7 – DISCUSSÃO.....	80
7.1 – Identificação e análise de proteínas LRR-RLKs em cana-de-açúcar.....	80
7.2 – Genes que codificam proteínas CC-NB-LRR induzida por <i>G. diazotrophicus</i> ou <i>Herbaspirillum rubrisubalbicans</i>	84
8- CONCLUSÕES.....	87
9 - REFERÊNCIAS BIBLIOGRÁFICAS.....	88

1- INTRODUÇÃO

A cana-de-açúcar é amplamente cultivada em diversos países do mundo, dentre os quais o Brasil destaca-se como o maior produtor. Esta gramínea possui grande importância econômica devido a sua utilização na fabricação de álcool, açúcar e outros produtos.

Com base na importância econômica da cana-de-açúcar, alguns pesquisadores têm se dedicado ao estudo intensivo dos aspectos biológicos desta espécie vegetal. Em 1958 Döbereiner e Ruschel identificaram bactérias fixadoras de nitrogênio associadas à cana-de-açúcar.

A partir de 1958, o estudo da Fixação biológica nitrogênio (FBN) ganhou impulso e posteriormente várias bactérias fixadoras de nitrogênio foram identificadas, tais como: *Herbaspirillum seropedicae* (Baldani et al 1994), *Herbaspirillum rubrisubalbicans* (Baldani et al., 1996), *Herbaspirillum frisingense* (Kirchhof et al., 2001), *Azospirillum doebereineriae* (Eckert et al., 2001) e *Burkholderia tropica* (Baldani et al., 2005).

Estudos com balanço de nitrogênio e quantificação do nitrogênio assimilado por uso de fertilizante marcado com ^{15}N mostraram que variedades brasileiras de cana-de-açúcar podem obter até 70% do N necessário para seu desenvolvimento, por meio da FBN (Lima et al., 1987; Urquiaga et al., 1992).

Dentre as bactérias fixadoras de nitrogênio, *Herbaspirillum* spp. e *Gluconacetobacter diazotrophicus* parecem ser as principais responsáveis pela alta taxa de FBN observadas nas variedades da cana-de-açúcar brasileiras (James et al., 1997). *G. diazotrophicus*, é conhecida por seus vários efeitos promotores do crescimento vegetal. Dentre estes, se podem ressaltar a fixação de nitrogênio, a mobilização de nutrientes, a proteção contra patógenos e a produção de fitormônios.

A bactéria Gram-negativa fixadora de nitrogênio *H. seropedicae* é encontrada em associação com raízes, caule e folhas de arroz, cana-de-açúcar e sorgo, promovendo o crescimento destas espécies vegetais através da FBN (Roncato-Mancari et al., 2003). Quanto ao gênero *Herbaspirillum*, a presença de ácido indolacético e giberelinas A1 e A3 foi constatada em culturas de *H. seropedicae* (Bastián et al., 1998) e indoís em estirpes de *H. rubrisubalbicans* (Radwan et al.,

2004). Este gênero é formado pelas espécies *H. rubrisubalbicans*, *H. seropedicae*, *H. frisingense* e *H. lusitana* (Radwan et al., 2005). O gênero *Herbaspirillum* é considerado endofítico obrigatório, por apresentar baixa sobrevivência no solo (Olivares et al., 1996).

A bactéria fixadora de nitrogênio *H. rubrisubalbicans* é capaz de colonizar plantas de cana-de-açúcar, sorgo, arroz e capim elefante. Estudos de caracterização ecológica demonstraram a ocorrência de *Herbaspirillum rubrisubalbicans* em raízes, caules e folhas de cana-de-açúcar (Olivares et al., 1996). Interessantemente, plantas de cana-de-açúcar do cultivar B4362, sorgo e capim elefante apresentaram sintomas de doença quando colonizadas por *H. rubrisubalbicans* (Olivares et al 1997)

O entendimento da associação entre cana-de-açúcar e endófitos é de importância considerável para agricultura. Os mecanismos que possibilitam bactérias endofíticas colonizarem seu hospedeiro de um modo endofítico e não patogênico ainda não foram elucidados. Ainda não foi mostrado como essas bactérias são reconhecidas pelas plantas.

Em plantas, o sistema de percepção de microrganismos e ativação das respostas de defesa é mediado pelos receptores ricos em leucina com domínio cinase intracelular (LRR-RLKs) presentes na membrana plasmática da célula hospedeira ou por proteínas citoplasmáticas com domínios de ligação a nucleotídeos (NBS) e regiões ricas em leucinas (LRR), denominadas NB-LRRs. Este sistema de sobrevivência em plantas é denominado sistema imune. Entretanto, as bactérias patogênicas desenvolveram mecanismos que bloqueiam ou suprimem as respostas de defesa ativadas por esses receptores.

No ano de 2000 pesquisadores do Estado de São Paulo criaram o projeto SUCEST, que teve como objetivo seqüenciar clones de bibliotecas de DNA complementar (cDNA) de cana-de-açúcar, de vários órgãos e tecidos em diferentes estágios de desenvolvimento e condições fisiológicas. Dentre estas análises, está contida a biblioteca de cDNA obtida de plantas inoculadas com a bactéria endofítica *G. diazotrophicus*.

Através da utilização de bancos de dados e programas de bioinformática podemos identificar e comparar um gene ou proteína contra genes ou proteínas de outros bancos de dados e inferir sobre a funcionalidade destes genes ou proteínas.

Assim, genes ou proteínas podem ser identificados baseados na significância estatística da similaridade e a função não conhecida de proteínas pode ser inferida baseada na função do homólogo (Sjölander, K., et al., 2004).

Através da utilização do banco de dados SUCEST e ferramentas de bioinformática disponíveis na web, este trabalho teve como objetivo identificar receptores LRR-RLKs e proteínas NB-LRRs diferencialmente regulados por *G. diazotrophicus* e/ou *H. rubrisubalbicans*. Com isso, esse trabalho pretende contribuir para elucidação do papel das proteínas LRR-RLKs e NB-LRRs em cana-de-açúcar durante a colonização por endófitos.

2- REVISÃO DA LITERATURA

2.1-Sistema imune em plantas

Todas as espécies vegetais estão expostas a microrganismos causadores de doenças. Contudo, embora existam estas interações não-harmônicas entre plantas e microrganismos, sabe-se que o estabelecimento da vida vegetal na terra foi facilitado por interações simbióticas com fungos, sugerindo que as plantas têm co-evoluído com microrganismos desde seu primeiro aparecimento sobre a terra (Chisholm et al., 2006, Shan et al., 2007).

No primeiro contato da interação entre plantas e microrganismos, estes se deparam com barreiras físicas e químicas pré-existentes na célula vegetal, tais como espessura e dureza da parede celular, ceras, espessura da cutícula, exsudados, saponinas, fitoalexinas, lectinas e outros. Estas barreiras dificultam a penetração direta do patógeno (Mysore et al., 2004). No entanto, as bactérias patogênicas acabam tendo acesso à célula vegetal, penetrando através de aberturas naturais como estômatos, hidatódios, poros usados para trocas gasosas, ou feridas (Chisholm et al., 2006). O mecanismo molecular pelo qual bactérias entram através destas aberturas não é conhecido. Acredita-se que estas aberturas funcionem como portas passivas de entrada para esses microrganismos (Melloto et al., 2006). Uma vez que as barreiras físicas e químicas da parede celular são transpassadas, a penetração do microrganismo atinge a membrana plasmática, onde se encontram

vários receptores de reconhecimento padrão (PRRs), que reconhecem padrões moleculares associados a microrganismos patogênicos e não patogênicos (MAMPs) (Chisholm et al., 2006, Boller et al., 2009). Deste modo, apesar das plantas não possuírem células imunes especializadas e imunidade inata somática adaptativa (Jonas e Dangl, 2006), elas são capazes de reconhecer e responder aos MAMPs e a eliciadores endógenos que são denominados DAMPs (danos associados a padrões moleculares) (Boller et al., 2009).

Atualmente, os MAMPs melhor caracterizados são: flagelina, fator de alongação EFtu, hairpin (Hrpz), lipossacarídeos, quitina e proteínas indutoras de necrose (NPP) (He et al., 2007, Boller et al., 2009). Os exemplos clássicos de DAMPs são oligogalacturonídeos, monômeros de quitina e sisteminas (Boller et al., 2009). A percepção destes MAMPs e DAMPs pelos PRRs induz respostas de defesa que contribuem para imunidade inata de plantas contra microrganismos patogênicos e não patogênicos (He et al., 2007, Boller et al., 2009). A imunidade inata em plantas é a primeira linha de resposta do hospedeiro, caracterizada como um sistema de sobrevivência, que detecta a presença e a natureza da infecção (Ryan et al., 2007). Em 2006, He e et alli mostraram através da análise de expressão de genes marcadores que *Pseudomonas syringae* DC3000 interfere com a imunidade inata suprimindo a resposta ativada por MAMPs.

Em mamíferos, os MAMPs são reconhecidos por um número limitado de PRRs (Toll e NOD). Tais receptores detectam componentes distintos dos microrganismos, ativando a resposta imune (He et al., 2007, Boller et al., 2009). Já em plantas, existem centenas de genes que codificam receptores transmembrana do tipo LRR-RLKs (“Receptor Like Kinase”) e LRR-RLP (“Receptor Like proteins”). Alguns destes desempenham o papel na percepção de diferentes MAMPs (Boller et al., 2009).

O MAMP melhor caracterizado foi isolado de fitobactérias, e denominado flagelina, um componente protéico constituinte da estrutura do flagelo de bactérias gram-negativas. A região central desta molécula é variável, todavia as porções N-amino e C-carboxi terminal são altamente conservadas (Chisholm et al., 2006). Estudos têm mostrado que um epítipo de 22 aminoácidos (Fgl22) da região N-terminal da flagelina é reconhecido pelo receptor FLS2 de *Arabidopsis thaliana*. Este

possui regiões extracelulares ricas em leucina (LRR) e um domínio intracelular do tipo serina/treonina/cinase (RLK) (Gómez-Gómez e Boller, 2000). Após a ligação de flg22 com FLS2, este dímero induz auto-fosforilação e múltiplas respostas de defesa são ativadas. Entre estas respostas pode-se destacar: a) influxo de cálcio, b) cascatas sinalizadas por proteínas cinases mitogênicas ativadas (MAPKS) e fatores de transcrição, c) produção de espécies reativas de oxigênio (ROS), d) alcalinização do meio de cultivo, e) produção de componentes antimicrobianos, f) deposição de calose e indução de genes de resposta à defesa (Abramovitch et al., 2006, Navarro et al., 2004, He et al., 2006, Ryan et al., 2007).

Neste sentido, plantas de *A. thaliana* que tiveram o gene que codifica FLS2 mutado, tornaram-se mais suscetíveis à infecção pela bactéria patogênica *Pseudomonas syringae* pv. *tomato*. Portanto, o receptor FLS2 medeia resistência a *P. syringae* pv *tomato* através do reconhecimento da flagelina em *A. thaliana* (He et al., 2006). Em 2008 Takai et al. identificaram e caracterizaram o ortólogo do receptor da flagelina (AtFLS2) em arroz que foi denominado OsFLS2. A superexpressão de OsFLS2 em células de arroz causou morte celular programada.

Além da flagelina, outro MAMP bastante conhecido é a proteína EFTu, responsável pela elongação dos ribossomos durante a tradução. Os primeiros 18 aminoácidos desta proteína são reconhecidos em *A. thaliana* pelo receptor rico em leucina (LRR) do tipo cinase (RLK) EFR, que é uma proteína pertencente à subfamília de receptores LRR-RX. Após o reconhecimento de EFTu, uma resposta de defesa semelhante a Flg22 é ativada (Ryan et al., 2007). Ambos os receptores FLS2 e EFR descritos anteriormente são membros do clado LRR XII, que inclui ainda proteínas de arroz do tipo LRR-RLKs (“Receptor Like Kinase”) (Shiu et al., 2004).

Em arroz, o gene Xa21, que codifica para um receptor do tipo LRR-RLK é induzido após a infecção por *Xantomonas orizae*, o que ocasiona a ativação de respostas de defesa, tais como: ROS, morte por hipersensibilidade (HR) e ativação de genes (Shiu et al. 2004). No entanto, em monocotiledôneas ainda não existe nenhuma evidência que proteínas do tipo LRR-RLKs reconheçam o fator de elongação EFTu (Morillo e Tax, 2006). Existem poucos trabalhos na literatura mostrando receptores LRR-RLK mediando resistência à doença em monocotiledôneas.

2.2-Subversão da resposta imune de plantas por bactérias patogênicas

As plantas são capazes de reconhecer MAMPs ativando suas respostas de defesa, e muitas vezes essas culminam com a resistência ao invasor. Contudo, algumas bactérias, ao longo da evolução, desenvolveram mecanismos para suprimir tal resistência. Dentre estes mecanismos, o sistema de secreção tipo III (TTSS III) é bastante utilizado por bactérias patogênicas para suprimir a defesa do hospedeiro. Durante a infecção, bactérias patogênicas utilizam este sistema para introduzir na célula vegetal moléculas efetoras que suprimem a resposta iniciada por receptores PRR (Shan et al., 2007).

Os patógenos de animais podem secretar um número limitado de efetores no interior das células hospedeiras. Entretanto, patógenos de plantas, tal como *Pseudomonas syringae*, podem secretar de 20 a 30 moléculas efetoras diferentes durante a infecção (Büttner et al., 2006).

As moléculas efetoras podem alterar a fisiologia da planta quanto à capacidade de colonização pela bactéria, podendo sustentar o crescimento de um patógeno. Algumas destas moléculas possuem atividade enzimática, como cisteínas proteases, proteases ubiquitinas simile e proteínas fosfatases, e podem modificar proteínas do hospedeiro para aumentar a virulência do patógeno ou a sua evasão da resposta imune (Chisholm et al., 2006, Janjusevic et al., 2006). Outras delas, tais como a família de proteínas efetoras de *Xanthomonas* (AvrBs3, AvrXa10, e AvrXa7), contém um sinal de localização nuclear (NLS) na região C-terminal e um domínio ácido de ativação transcricional (AAD). Estas características sugerem que moléculas efetoras podem alterar a transcrição gênica durante o processo de infecção, provavelmente inibindo os genes de defesa (Chisholm et al., 2006).

Em *P. syringae*, o efector Avrpt0 foi o primeiro a ser caracterizado como supressor da resposta imune inata. Neste contexto, foi verificado que plantas transgênicas de *A. thaliana* expressando Avrpt0, AvrRpt2 ou AvrRpm1 são comprometidas em sua habilidade para depositar calose quando infectadas por *P. syringae* mutantes para TTSSIII (Nomura et al. 2005). Estudos posteriores mostraram que os efetores Avrpt0 e Avrpt0 B interceptam as respostas de sinalização mediadas pelos receptores FLS2, EFR e BAK1, impedindo a ativação de

cascadas MAPKs, deposição de calose e outras respostas de defesa (Boller et al., 2009a).

Os efetores também podem manipular as respostas de defesa mediadas pelos hormônios ácido salicílico, etileno e ácido jasmônico. Foi verificado que os efetores de *P. syringae* AvrB, AvrRpt2, AvrPphB, HopPtoK e AvrPphEpto podem manipular a via do ácido jasmônico em *A. thaliana*. Também os efetores AvrPto e AvrptoB de *P. syringae* pv. *tomato* aumentam a expressão da enzima formadora do etileno (ACC oxidase) em plantas de tomate. Isto ocasiona morte celular, podendo melhorar o acesso do patógeno aos nutrientes da planta ou a disseminação no ambiente (Abramovitch et al., 2006).

Algumas proteínas efetoras de *P. syringae* interferem nos eventos que induzem a ativação da HR pela planta hospedeira. Por exemplo, as proteínas efetoras Avrpt2 e AvrPtoB de *P. syringae* suprimem a imunidade baseada em HR (Abramovitch et al., 2003). A proteína efetora VirPphA de *P. syringae* pv. *phaseolicola* (também conhecida como HopAB1) também modula a resistência baseada na HR, permitindo o patógeno escapar da resposta imune. A atividade enzimática de várias proteínas efetoras indica que tais proteínas podem modificar ou degradar proteínas hospedeiras associadas com a HR (Abramovitch et al., 2003).

2.3-Teoria gene a gene

Os microrganismos têm co-evoluído com seus hospedeiros desde os primórdios da vida na Terra. Assim, a evolução de proteínas efetoras secretadas pelos patógenos conduziu ao surgimento de proteínas em plantas, que reconhecem esses efetores de fungos, bactérias e vírus (Chisholm et al. 2006). A associação que descreve o reconhecimento de proteínas efetoras por proteínas de resistência (R) no interior da célula hospedeira foi caracterizada inicialmente por Flor em 1955 (Flor 1955 apud Chisholm et al., 2006) e foi descrita como resistência gene a gene (Nimchuk et al. 2003). Na presença de uma associação entre uma proteína cognata R e um efetor, a resistência é ativada, resultando na sinalização de defesa e resistência do hospedeiro. Geralmente, a resistência se manifesta como morte celular localizada e inibição do crescimento do patógeno. Ao contrário, na ausência

desta interação proteína efetora e proteína R, o patógeno não é detectado pela planta hospedeira, o que resulta na sua proliferação dentro da célula vegetal e no aparecimento de sintomas de doença (Chisholm et al. 2006).

Interessantemente, algumas plantas possuem resistência a vários microrganismos invasores, tal fenômeno é denominado resistência não hospedeira ou resistência à espécie. As plantas resistentes para todos os isolados de um dado patógeno são referidas como plantas não hospedeiras. Essa resistência é determinada por mecanismos de resistência múltiplas que compreendem barreiras físicas, reações induzidas (por exemplo: síntese de fitoalexinas) e proteínas R (Ürnberger, et al., 2005).

As plantas possuem uma grande variedade de proteínas R, que podem ser classificadas em duas classes principais: proteínas intracelulares com domínios NBS-LRR (sítio de ligação a nucleotídeo-região repetida rica em leucina) e proteínas extracelulares LRR (eLRR) (Chisholm et al. 2006). As proteínas intracelulares NBS-LRR podem ter os seguintes domínios N-terminais: a) um suposto domínio “coiled coil” (CC) b) um domínio que compartilha homologia com o domínio citoplasmático de *Drosophila* TOLL e receptores de mamíferos IL-1 (TIR: receptor interleucina toll) (Nimchuk et al. 2003).

As proteínas NBS-LRR desempenham um papel importante na resistência a doença em plantas. Esta classe de proteína é um componente central do sistema imune em plantas. A primeira característica comum destas proteínas é que elas compartilham similaridades de seqüência com regiões NB de indutores apoptóticos tais como CED4 de *Caenorhabditis elegans* e Apaf-1 de humanos. Com isto, sugere-se que atividade de proteínas R, pode estar associada à ligação ou hidrólise de ATP. A região LRR tem tamanho de 20-30 aminoácidos de comprimento. Este domínio tem sido elucidado a partir da infecção por vírus em eucariotos e pode estar relacionado à interação proteína-proteína (Chisholm et al. 2006).

A segunda classe de genes R codifica para proteínas extracelulares LRR (eLRR). Essa classe inclui as seguintes proteínas: RLP (receptores proteínas símile; domínios extracelulares LRR ou transmembrana), RLK (LRR extracelular, domínio transmembrana e cinase citoplasmática) e PGIP (proteína inibidora de galacturonase; região rica em leucina, na parece celular) (Chisholm et al. 2006). A

maioria das proteínas de resistência de plantas está agrupada nestas duas classes descritas anteriormente.

2.4- As proteínas NB-LRR e mecanismos de sinalização

Existem uma grande quantidade de proteínas NB-LRR encontradas em plantas. Em *A. thaliana* foram identificadas 150, em *O. sativa*, aproximadamente 600 genes codificam para tais proteínas. No entanto, apenas algumas foram caracterizadas (Rafiqi et al., 2009).

A principal função das proteínas NB-LRR já caracterizadas é o reconhecimento de moléculas efetoras de patógenos. Contudo, algumas proteínas NB-LRRs são componentes da vias de sinalização de imunidade, atuando “downstream” de outros receptores (RLPs, NB-LRRs). Por exemplo, a proteína NCR1 (NB-LRR) é necessária para ativar a defesa iniciada pelo receptor Cf4, um receptor do tipo RLP (receptor proteína símile) que reconhece a molécula efetora avr4 de *Cladosporium fulvum*. A proteína NCR1 induz a resposta de defesa mediada por diversos receptores (Cf-4, Cf-9, LeEIX2, Rx, Prf e Mi). Este tipo de sinalização é similar a função da proteína humana NOD2 (NB-LRR), que participa das cascatas de sinalização do sistema imune, porém não atua como receptor do sistema imune (Rafiqi et al., 2009).

As proteínas NB-LRR, podem também associar-se para induzir uma resposta imune contra um determinado patógeno. Em *A. thaliana*, as proteínas RPP2A (TIR-NB-LRRs) e RPP2B (TIR-NB-LRR- “like”) interagem para estabelecer a imunidade contra *Peronospora parasitica*. Contudo, a imunidade pode ser estabelecida independentemente da associação.

As proteínas NB-LRR podem reconhecer moléculas efetoras indiretamente (denominado Modelo guarda), através da percepção de modificações bioquímicas em proteínas do hospedeiro que são produzidas por moléculas Avr; ou diretamente pela associação com a proteína Avr (denominado Modelo receptor ligante) (DeYoung e Innes, 2006). A forma mais simples de reconhecimento ocorre através da associação física direta da molécula efetora com o receptor do sistema imune R (DeYoung e Innes, 2006).

O primeiro exemplo de reconhecimento direto foi elucidado para a proteína Pita (CC-NB-LRR) de *O. sativa*. O domínio LRR de Pita interage diretamente com a molécula efetora AVR-Pita, secretada pelo fungo *Magnaporthe grisea* (Jia et al., 2000). Uma única substituição de aminoácidos no domínio LRR de Pita pode abolir a interação, resultando na perda da resistência (Bryan et al., 2000).

O modelo de reconhecimento indireto (Modelo Guarda) foi proposto inicialmente para explicar os mecanismos em que a proteína Avrpto de *P. syringae* pv *tomato* é reconhecida pelas proteínas Pto (RLP) e Prf (NB-LRR) (Rafiqi et al., 2009). A proteína Prf é encontrada em associação com a proteína Pto. Esta associação mantém a proteína Prf em um estado inativo (Rafiqi et al., 2009). Porém, Prf é ativada na interação Avrpto-Pto, e a resposta de defesa é induzida. Neste modelo de reconhecimento as proteínas NB-LRR são chamadas de guardas, que monitoram o estado de outras proteínas hospedeiras (guardas) ativáveis por moléculas efetoras secretadas por patógenos (Rafiqi et al., 2009). O Modelo guarda suporta ou explica o fato de um pequeno número de proteínas NB-LRRs reconhecerem uma enorme quantidade de proteínas efetoras de patógenos (Rafiqi et al., 2009).

Outros exemplos que apoiam a hipótese guarda são as proteínas RIN4, RPM1 e RPS2. A proteína RIN4 (regulador negativo da defesa em plantas), é alvo de três proteínas Avr de *P. syringae*, Avrpm1, AvrB e Avrpt2 e guardadas por duas proteínas CC-NB-LRR (RPM1 e RPS2) (Caplan et al., 2008). A proteína RIN4 é clivada ou fosforilada sob interação com Avrpm1, AvrB ou Avrpt2. Esta modificação resulta no distúrbio de sua ação inibitória sobre RPM1 ou RPS2. Outro exemplo de hipótese guarda é ilustrado pela proteína PBS1 que é degradada pela proteína efetora HopAR1 de *P. syringae* (Rafiqi et al., 2009). Assim, a proteína RPS5 é capaz de perceber esta degradação para ativar mecanismos de resistência. Interessantemente, RIN4 e PBS1 (os guardas), associam-se com o domínio CC de seus respectivos guardas, RPM1 e RPS5 (Rafiqi et al., 2009).

2.5- Classificação das proteínas NB-LRRs e motivos conservados

As proteínas NBS-LRR possuem distintos domínios que são compostos de motivos característicos. O domínio NBS é necessário para ligação a molécula de ATP ou GTP, esta habilidade foi demonstrada para as proteínas I2 e Mi (NB-LRRs) de tomate (Tameling et al., 2002). Os motivos conservados no domínio NBS podem ser usados para classificar as sequências em subgrupos com funções discretas (Meyers et al., 2003). As proteínas NB-LRR apresentam LRR (regiões ricas em leucina) de tamanho variáveis. As proteínas NBS-LRR podem apresentar na sua região N terminal os motivos “coiled-coil (CC) ou regiões de homologia para os receptores TIR (Meyers et al., 2003).

Em *A. thaliana*, foram identificadas 150 proteínas NB-LRR, que foram subdivididas nos grupos: a) CC-NBS-LRR, contendo o domínio CC N-terminal, b) TIR-NBS-LRR, contendo o domínio TIR. Nestes subgrupos 55 proteínas apresentam os domínios CC-NBS-LRR e 94 domínios TIR-NBS-LRR (Meyers et al., 2003). A análise filogenética da região NBS das proteínas CC-NB-LRR de *A. thaliana* revelou que tais proteínas formam quadro grados denominados CNL-A, CNL-B, CNLC, CNLD (Figura 1) (Meyers et al., 2003).

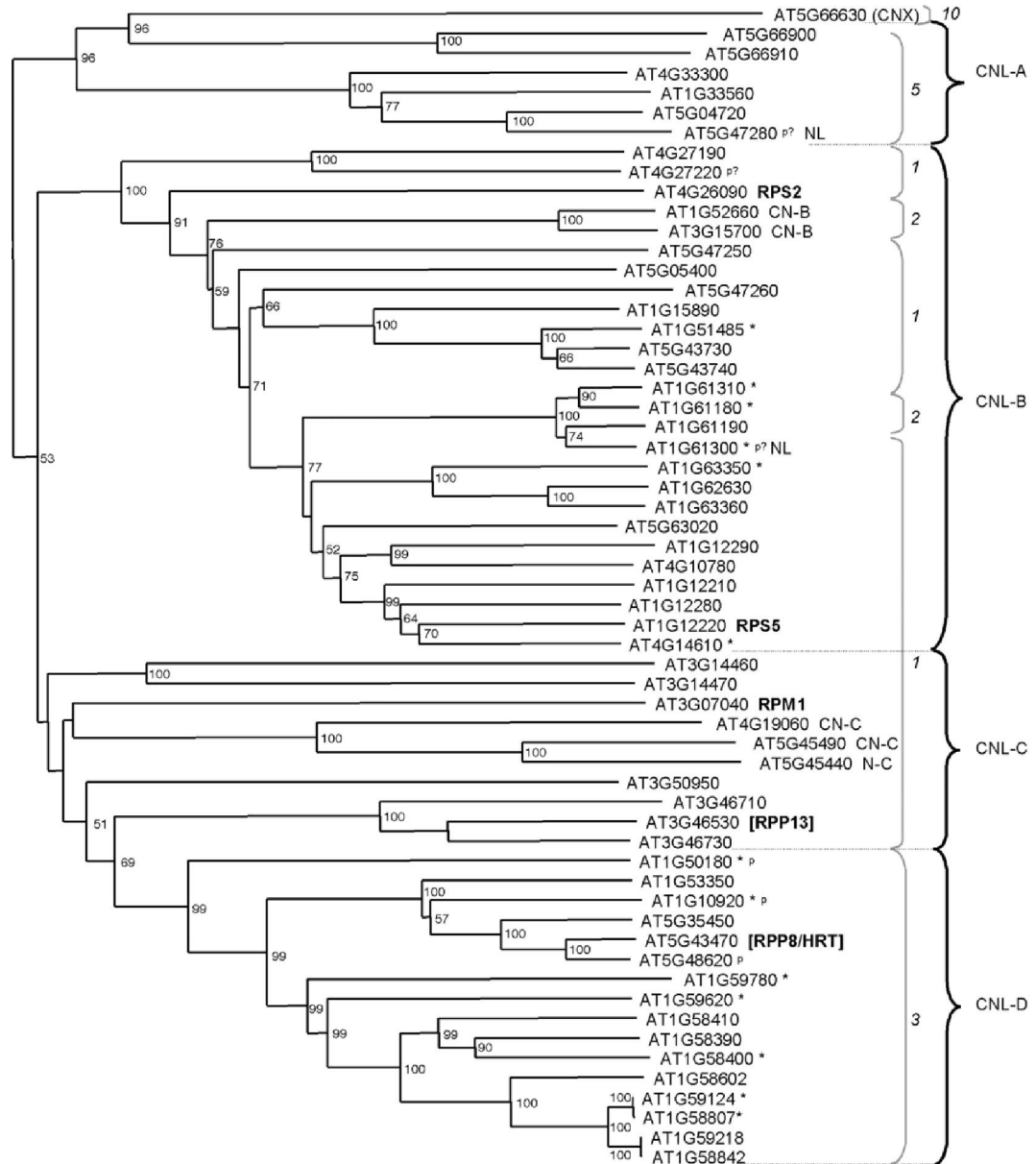


Figura 1- Relação filogenética das proteínas CC-NB-LRR encontrados no genoma de *A.thaliana*

Árvore filogenética “neighbor-joining” (NJ) da matriz de distância construída de acordo os parâmetros do método de Kimura (1980), usando o resultado de alinhamento da região NBS. Os comprimentos dos braços são proporcionais as distâncias genéticas. Os nomes dos genes conhecidos que pertencem aos grados (A, B, C e D) estão identificados (negrito/preto). Os números indicam a porcentagem de “bootstrap” de 1000 réplicas que suporta o nodo adjacente. Os valores >50% de “bootstrap” não foram informados. As chaves (}) à direita indicam o nome do subgrupo baseado na filogenia e posição/número do íntron, sendo CNL-A (grado A), CNLB (grado B), CNL-C (grado C) e CNLD (grado D). Fonte: adaptado de Meyers et al., 2003.

Em 2003, Meyers et alli analisaram os domínios NBS (N-terminal) e LRR (C-terminal) quanto à presença de motivos no programa MEME (“Motif-based sequence analysis tools”). No domínio N-terminal CC (Coiled coils) foram identificados 8 motivos conservados, pareados a subgrupos definidos pela posição definida do íntron.

Dentre os motivos encontrados, MHDV está presente na região NBS de 156 proteínas NB-LRRs de *A. thaliana*. Os outros 7 motivos são: P-loop, RBNSA, kinase 2, RBNSB, RBNSC, RBNSD e GLPL. As sequências correspondentes a estes motivos são apresentadas no QUADRO 1 (Meyers et al., 2003).

Quadro 1: Motivos conservados presentes na região NBS de proteínas CC-NB-LRR de *A. thaliana*. Fonte: adaptado de Meyers et al., 2003.

Motivo	Sequência
P-loop	VGIYGMGGVGKTTLARQIF
RBNSA	VKxGFDIVWVVVSQEFTLKKIQDILEK
RBNSB	NGCKVLFTRSEEV
RBNSC	KVECLTPEEAWELFQRKV
RBNSD	CFLYCALFPEDYEIxKEKLIDYWIAEGFI
KINASE 2	KRFLVLDDIW
GLPL	EVAKKCGGLPLALKVI

Estes motivos são conservados no domínio NBS de plantas e animais (DeYoung e Innes, 2006). Embora as funções bioquímicas destes motivos ainda não serem bem definidas, estes são importantes, pois substituições nestes domínios afetam a função das proteínas NB-LRRs (DeYoung e Innes, 2006). Dentre os 8 motivos conservados, a família ADR1 apresenta os motivos: P-loop, kinase-2, RNBS-A, GLPL, RNBS-D e MHDV (Grant et al. 2003). Além destes, um novo motivo denominado TVS foi encontrado unicamente para a família de proteínas ADR1 (Chini e Loake, 2005). Os outros membros do grupo A, At5g66900 e At5g66910 não possuem este motivo (Chini e Loake, 2005).

O domínio NBS geralmente termina com o motivo MHDV (Ashfield et al. 2004), enquanto o domínio LRR começa aproximadamente 40-65 aminoácidos C-terminal desse motivo (Chini e Loake, 2005). Contudo, uma variação desse motivo está presente no clado A, no qual o aminoácido metionina (M) é substituído pelo

aminoácido glutamina(Q) no terceiro resíduo do motivo MHDV (Meyers et al., 2003). Entre a região NBS e LRR há um interdomínio denominado NL. Este é conservado entre as diferentes classes de NBS-LRR mas varia entre as classes (Meyers et al., 2003). O interdomínio NL na família ADR1 é composto de 78 resíduos de aminoácidos, e um novo motivo denominado PKAE, que foi identificado pelas análises no programa MEME. Dentre os quatro grupos de proteínas CC-NB-LRR, na família ADR1 este motivo na região NL foi o mais conservado. Este motivo foi mais conservado do que os outros motivos presentes na região NBS, com exceção do P-loop e Kinase 2. Outros membros do clado A como At5g66900 e At5g66910 não apresentam este motivo (Chini e Loake, 2005).

2.6- Os Receptores RLKs em plantas

Todos os sistemas vivos são capazes de reconhecer estímulos e processar a informação ao nível celular. Este sistema de percepção do ambiente é mediado por várias classes de receptores na superfície celular, dentre os quais estão os receptores RLKs (Shiu e Bleecker, 2001., Shiu e Bleecker, 2003). Em *A. thaliana*, RLKs formam uma família monofilética que contém mais de 610 membros, que incluem receptores cinases e receptores citoplasmáticos que se assemelham à cinases (RLCK) (Shiu e Bleecker, 2003). Em arroz foram identificados aproximadamente 1.131 genes que codificam receptores que se assemelham aos RLK sugerindo uma expansão da família em monocotiledôneas (Shiu et al., 2004). Na família de genes RLKs de *A. thaliana*, mais de 400 membros possuem um domínio de configuração que se assemelham a receptores transmembranas (Shiu e Bleecker, 2001) e são denominados receptores cinases transmembranas de plantas (PRKs) (Diévert e Clark, 2004).

Os receptores transmembranas são definidos como proteínas que contêm um domínio extracelular, um domínio unipasso na membrana e um domínio citoplasmático serina/treonina cinase (Diévert e Clark, 2004). Com isso, sugere-se que tais receptores estejam envolvidos na percepção de sinais na superfície celular da planta e ligantes. As proteínas PRKs podem ser classificadas com base no seu domínio extracelular (Shiu e Bleecker, 2001). As proteínas LRR-RLK representam

um grande número de PRKs no genoma de *A. thaliana*, com 216 membros. A principal característica dos LRR-RLKs são seu domínio extracelular compostos de motivos ricos em leucina organizados em seqüência e o domínio cinase intracelular (Yin, et al., 2002).

As proteínas LRR-RLK em *A. thaliana* representam 13 subfamílias (LRR I a XIII) e são classificadas de acordo com seu domínio extracelular LRR (Diévert e Clark, 2004). Todas LRR-RLKs analisadas em plantas apresentaram atividade funcional no seu domínio cinase (Diévert e Clark, 2004). Contudo apesar do grande número de LRR-RLK em plantas, apenas 10 têm função conhecidas.

Outros receptores LRR-RLK têm funções conhecidas como podemos citar: CLAVATA1 (Diévert et al., 2004), ERECTA (Godiard et al., 2003), BRI (Bishop e Kong, 2002), SERKs (Albrecht et al., 2008), HAESA (Jin et al., 2000), FLS2 (Gómez-Gómez e Boller, 2000), EFR (Zipfel, et al., 2006) SYMRK (Markmann et al., 2008).

2.7- Interações entre plantas e microrganismos simbiotes

As plantas podem interagir com microrganismos formando associações simbióticas, endofíticas ou patogênicas. Durante o desenvolvimento destas associações, o sistema de defesa da planta é estritamente induzido, determinando a compatibilidade da interação (Mithöfer, 2002).

Em uma interação incompatível, a planta hospedeira resistente estabelece um conjunto de mecanismos de defesa diretamente contra o patógeno. Dentre estes se destaca a fortificação da parede celular, a geração e o acúmulo de espécies reativas de oxigênio (ROS), a produção de fenilpropanóides e a expressão de proteínas relacionadas à patogênese (PR) (Mithöfer, 2002).

Em contraste, uma associação simbiótica desenvolvida entre plantas do tipo leguminosas e bactérias do gênero *Rhizobium*, leva ao desenvolvimento de órgãos vegetais especializados denominados nódulos que são formados nas raízes. Nos nódulos, estas bactérias convertem N_2 em NH_3 . Esta, por sua vez, é assimilada pela planta hospedeira. Em contrapartida, a bactéria é suprida com componentes de carbono fornecidos pela planta. A enzima responsável por essa conversão é denominada nitrogenase, que é bem caracterizada e sabidamente inibida por

oxigênio, embora a bactéria seja aeróbica obrigatória. Assim, uma das funções dos nódulos é regular o nível de oxigênio para permitir a atividade da nitrogenase e a viabilidade bacteriana (Oldroyd, 2001).

O estágio inicial desta associação envolve uma complexa interação entre planta e bactéria. Os compostos flavonóides são um dos primeiros sinais liberados pelas raízes da planta hospedeira. As bactérias compatíveis presentes na rizosfera percebem tais sinais e ativam os genes bacterianos *nods* que estão envolvidos na produção dos fatores Nod. Estes fatores bacterianos Nods são moléculas que possuem uma estrutura central com resíduos de *N*-acetil-d-glucosamina unidos por ligações β -1,4, que pode diferir em número dentro da mesma espécie ou entre espécies diferentes (Jones et al., 2007). Acredita-se, que tais fatores sejam reconhecidos pela célula vegetal por RLK que possuem um domínio LysM, um peptídeo sinal, um segmento transmembrana e um domínio serina/treonina cinase (Jones et al., 2007). Esses receptores estão agrupados em uma família com três membros em *Medicago truncatula* (NFP, LYK3 e LYK4) e dois membros em *Lotus japonicus* (NFR1 e NFR5) (Arrighi et al., 2006). Adicionalmente, outros supostos receptores envolvidos na interação são os RLKs com motivos repetidos ricos em leucina, tais como: DMI1, DMI2, DMI3 (“Doesn’t Make Infections”), NSP1 e NSP2 em *Medicago truncatula* (Arrighi et al., 2006).

Após o reconhecimento dos fatores Nod bacterianos por seus receptores específicos na célula vegetal, ocorrem os seguintes eventos: sinalização por fosfolipase C, aumento intracelular do nível de cálcio seguida por sua forte oscilação, fluxos de íons, alteração do citoesqueleto da raiz e ativação da expressão de genes de nodulação na planta (ENODS) (Geurts et al., 2005, Oldroyd e Downie, 2008). Estes eventos resultam na indução de mitose nas células corticais das raízes e na formação dos nódulos, que formam um cordão de infecção permitindo o movimento das bactérias contidas na rizosfera para o córtex das raízes. Após o estabelecimento dos nódulos, *Rhizobium* cresce dentro da raiz e são liberadas no interior das células corticais envolvidas por membranas através de um processo endocítico (Peleg-Grossman et al., 2007).

Durante a colonização da planta hospedeira por *Rhizobium*, a bactéria não elicia reações do sistema de defesa da planta, normalmente induzidas pelos

microrganismos invasores. Contudo, em outras circunstâncias, várias reações de defesa podem ser ativas na interação entre leguminosas e bactérias noduladoras. Estudos têm demonstrado que durante a interação entre *Sinorhizobium meliloti* e alfalfa (*Medicago sativa*) a planta controla a infecção e o número de nódulos através da ativação do sistema de defesa. Neste tipo de resposta foi observada a formação de células necróticas concomitantemente com o acúmulo de proteínas relacionadas à patogênese, componentes fenólicos e ROS (Mithöfer, 2002).

Trabalhos mostram que na interação entre *Rhizobium* e *Medicago truncatula* vários genes relacionados à defesa foram induzidos durante o processo de nodulação. Entre estes podemos citar genes que codificam: proteínas relacionadas à defesa (MtN1 e MtN13), enzimas envolvidas na biossíntese de fenilpropanóides, fitoalexinas, proteínas envolvidas na modificação da parede celular (β - 1, 3 glucanase e peroxidase), PR-10 e quitinase (Mithöfer, 2002; Stacey et al., 2006). Em alguns casos mais pronunciados de reações de defesa, não foi possível a associação entre leguminosas e *Rhizobium*, ocorrendo apenas a formação de pseudo-nódulos (Mithöfer, 2002). Um fato interessante é que alguns destes genes de defesa foram induzidos nos estágios iniciais de nodulação e, posteriormente, ocorreu a diminuição da expressão nos nódulos. Portanto, estes resultados indicam que reações de defesa são induzidas em plantas de legumes colonizadas por *Rhizobium* e, para que ocorra o sucesso na simbiose, possivelmente, tais bactérias suprimem ou evitam as respostas de defesa (Jones et al., 2007).

Embora o sistema de defesa seja ativado, uma das questões básicas a serem respondidas é se as bactérias *Rhizobium* são reconhecidas como patógenos ou não (Mithöfer, 2002). Durante tal interação, alguns fatores microbianos ativam a resposta imune e outros fatores suprimem esta resposta. No entanto, ainda não está completamente elucidado como a planta percebe e responde aos padrões moleculares associados à *Rhizobium* (Jones et al., 2007).

Muitas plantas são capazes de reconhecer uma seqüência conservada do epítipo da flg 22 e ativar a defesa basal. Contudo, *Rhizobium*, juntamente com outras proteobactérias como *Agrobacterium tumefaciens* e *Brucella* spp, não possui o epítipo da flg22 na sua seqüência. Além da flagelina, outros dois epítipos que

induzem a resposta imune basal, o elf18 e o EFTu, são conservados em *Rhizobium*, mas leguminosas não respondem a estes epítomos (Jones et al., 2007).

É conhecido que durante a infecção de leguminosas por *Rhizobium*, os fatores bacterianos Nods, em contato com a planta, induzem a produção de ROS. Esta resposta se assemelha a resposta de defesa basal de plantas quando colonizadas por patógenos (Jones et al., 2007). Assim, possivelmente, a reação de defesa induzida pelos fatores Nods é atenuada ou modificada por outros efetores de *Rhizobium*, como os lipossacarídeos (LPS), para que haja sucesso na interação e a bactéria consiga colonizar a planta hospedeira. O componente lipídico de *S. meliloti* pode suprimir ROS e a expressão de genes de defesa em culturas de células hospedeiras de *M. trunculata* e *M. sativa*. Contudo, freqüentemente, LPS de bactérias patogênicas e não hospedeiras, induzem respostas de defesa que resultam em ROS e ativação de genes de defesa (Jones et al., 2007).

As proteínas efetoras são injetadas por patógenos na célula vegetal através do TTSS causando HR e morte celular programada. O TTSSIII foi identificado em plasmídeos de várias bactérias do gênero *Rhizobium*, normalmente ligados a genes de simbiose. Contudo, estes efetores estão ausentes nas espécies *S. meliloti* e *R. leguminosarum* bv. *viciae* e em algumas espécies de *M. loti* (Soto et al., 2006).

Em *Rhizobium*, o sistema TTSSIII também participa na modulação da interação com hospedeiros suscetíveis, assim como em patógenos. Estudos têm mostrado que o bloqueio deste sistema leva a vários efeitos na associação entre leguminosas e *Rhizobium*, que são específicos na relação planta hospedeiro (Soto et al. 2006). O bloqueio do TTSSIII pode melhorar a nodulação ou converter pseudo-nódulos para nódulos fixadores de nitrogênio. No entanto, alguns efeitos negativos também foram observados como uma forte resposta de defesa do hospedeiro, interferindo com a nodulação (Soto et al. 2006).

2.8- A cana-de-açúcar e endófitos

A cana-de-açúcar é uma gramínea pertencente à família Poacea do gênero *Saccharum*. Acredita-se, que foi originada no Sudoeste da Ásia, na Nova Guiné (Zambrano et al., 2003).

As variedades cultivadas de cana-de-açúcar são resultados da hibridização interespecífica envolvendo *Saccharum officinarum*, *S. barberi*, *S. sinense*, e as duas espécies selvagens *S. spontaneum* and *S. robustum*. É uma das espécies vegetais que possuem um dos mais complexos genomas, carregando um número variável de cromossomos (geralmente $2n = 70-120$) (Lu et al., 1994).

A cana-de-açúcar é uma das mais importantes fontes de açúcar e álcool no mundo, sendo cultivada em regiões tropicais e subtropicais (Wanderley-Nogueira et al., 2007). O Brasil possui a maior área cultivada com cana-de-açúcar e destaca-se como o maior produtor mundial. Na indústria, a cana-de-açúcar é usada para fabricação de açúcar, etanol, aguardente, dextranos, plásticos, papel e alguns produtos farmacêuticos (Enriquez-Obregón et al., 1998).

Em 2000, Reis Jr. et alli fizeram um levantamento da presença de bactérias diazotróficas em diferentes genótipos de cana-de-açúcar, no qual constatou-se a presença das seguintes bactérias: *Azospirillum lipoferum*, *A. brasilense*, *A. amazonense*, *Herbaspirum* spp. e *G. diazotrophicus*. Dentre elas, as duas últimas parecem ser as principais responsáveis pela alta taxa de FBN observada nas culturas da cana-de-açúcar brasileiras (James et al., 1997).

A bactéria endofítica fixadora de nitrogênio *G. diazotrophicus* é Gram-negativa, e foi isolada primeiramente a partir de raízes, caules e folhas de variedades brasileiras cana-de-açúcar (Cavalcante e Döbereiner, 1988). Sua ocorrência na rizosfera, apesar de já constatada, é menos enfatizada devido às dificuldades de isolamento no solo ou a sua baixa sobrevivência nesta condição (Saravanan et al., 2007). Atualmente, esta bactéria já foi isolada de várias outras espécies de plantas, incluindo de arroz e café (Perin et al., 2004), que também possuem elevada importância econômica (Jimenez-Salgado et al., 1997).

A classificação de *G. diazotrophicus* agrupa esta bactéria no filo Proteobacteria, seção Alfa-Proteobacteria, ordem Rhodospirillales e família

Acetobacteraceae (Kerstens, 2006). Atualmente, esta família contém três gêneros de bactérias fixadoras de nitrogênio, compreendendo sete espécies: *Acetobacter nitrogenifigens*, *Gluconacetobacter kombuchae*, *Gluconacetobacter johannae*, *Gluconacetobacter azotocaptans*, *G. diazotrophicus*, *Swaminathania salitolerans* e *Acetobacter peroxydans* (Saravanan et al., 2007).

A bactéria *G. diazotrophicus* pode ser adicionada à lista de bactérias promotoras de crescimento vegetal devido a suas seguintes características: fixação de nitrogênio, solubilização de minerais zinco e fósforo, aumento da internalização de nutrientes, produção de fitormônios (auxinas e giberilinas), vitaminas e supressão de desenvolvimento de patógenos (Saravanan et al., 2007).

Deste modo, plantas de cana-de-açúcar inoculadas com a bactéria mostram aumento significativo no seu crescimento em condições limitantes de N no solo. Plantas de cana-de-açúcar inoculadas com *G. diazotrophicus* fixam acima de 150 kg N/hectare na estação de crescimento, o que representa 50% do nitrogênio total fixado necessário para o cultivo (Fischer et al., 2005). Por isso, o entendimento da associação entre cana-de-açúcar e *G. diazotrophicus* é de importância considerável para agricultura (Fischer et al., 2005).

A colonização de *G. diazotrophicus* em cana-de-açúcar não se assemelha à simbiose de leguminosas e *Rhizobium*, onde as bactérias são restritas aos nódulos. *G. diazotrophicus* coloniza os espaços intercelulares e tecidos vasculares de muitos órgãos da planta, sem formar nódulos (Vinagre et al., 2006). Possivelmente, para não causar danos e reações de defesa do hospedeiro, a bactéria modula as vias de defesa da cana-de-açúcar. No entanto, os mecanismos que permitem a bactéria colonizar deste modo endofítico e não patogênico ainda não estão elucidados (Vinagre et al., 2006).

Estudos têm mostrado que plantas de cana-de-açúcar cultivar Ja60-5 inoculadas com *G. diazotrophicus* apresentaram acúmulo de polissacarídeos e taninos nas células parenquimáticas, ao redor do metaxilema (Dong et al., 1997; Lambais, 2001). No cultivar SP70-1143, não foi observada HR após a inoculação de folhas e caule com *G. diazotrophicus*, embora uma matriz extracelular se acumule ao redor da bactéria no protoxilema e xilema (Lambais, 2001). Com isso, há indícios que sugerem a ativação do sistema de defesa durante a interação com esta bactéria.

Lambais et alli em 2001, através de análises “in silico” no banco de dados do SUCEST, mostraram que vários genes de cana-de-açúcar relacionados à defesa vegetal foram induzidos ou suprimidos na presença de *G. diazotrophicus*. Entre os transcritos diferencialmente expressos podemos citar a enzima óxido nítrico sintase (NOS1), uma isoforma da catalase (CAT3), a peroxidase (POX9), a fenilalanina amônia liase e um fator de transcrição WRKY.

Neste sentido, outro aspecto que pode ser considerado é que genótipos de cana-de-açúcar distintos têm taxas de FBN diferente (Vinagre et al., 2006). Isto sugere que fatores genéticos podem estar controlando o processo de reconhecimento, colonização e fixação de nitrogênio. Com isso, possivelmente as respostas de defesa da cana-de-açúcar a *G. diazotrophicus* são influenciadas pelo genótipo (Vinagre et al., 2006).

A bactéria Gram-negativa fixadora de nitrogênio *H. seropedicae* também é encontrada em associação com raízes, caule e folhas de plantas economicamente importantes, como arroz, cana-de-açúcar e sorgo, promovendo o crescimento destas espécies vegetais (Roncato-Manccari et al., 2003). Este gênero *Herbaspirillum* é formado pelas seguintes espécies: *H. seropedicae*, *H. rubrisubalbicans*, *H. frisingense* e *H. lusitanum* (Radwan et al., 2005). No entanto, dentre estas, é importante destacar que nem todas estabelecem uma interação benéfica com seus hospedeiros. *H. rubrisubalbicans*, diferentemente de *H. seropediceae*, é causadora de doença em cana-de-açúcar e sorgo (Roncato-Manccari et al., 2003). A descrição do gênero inclui bactérias em forma de bastonetes curvos, gram negativas, vibróides, móveis, com 1 a 3 flagelos inseridos em 1 ou 2 pólo e micro-aeróbicas quando fixam nitrogênio (Krieg, 1984).

A espécie *Herbaspirillum rubrisubalbicans* é a nova reclassificação de *Pseudomonas rubrisubalbicans*, reconhecida como agente causal da doença estria mosqueada em cana-de-açúcar (Baldani et al., 1997). Pimentel et al. em 1991 avaliaram a patogenicidade de diversas estirpes de *H. rubrisubalbicans* e *H. seropedicae* em cana-de-açúcar (cv. B 4362), sorgo (*Sorghum bicolor*) e capim Napier (*Pennisetum purpureum*) e observaram que a inoculação artificial dos isolados de *H. rubrisubalbicans* causava sintomas nas três plantas usadas no ensaio (Baldani et al., 1997).

Posteriormente, James et al. (1997) e Olivares et al. (1997) testaram a patogenicidade das estirpes de *H. rubrisubalbicans* e compararam com estirpes de *H. seropedicae* através de injeções de suspensões de células diretamente no limbo foliar. *H. rubrisubalbicans* desenvolveu sintomas de estrias mosqueadas em sorgo, capim elefante e na variedade sensível de cana-de-açúcar B-4362 (Baldani et al., 1997). Já a variedade resistente de cana-de-açúcar, a SP 70-1143, não apresentou nenhum sintoma da doença (Baldani et al., 1997).

2.9- O projeto SUCEST

Projeto SUCEST foi fundado no estado de São Paulo, sob a coordenação da FAPESP, com objetivo de seqüenciar os genes expressos em cana-de-açúcar em diferentes tecidos e condições fisiológicas (Vettore, 2001). O projeto gerou 237.954 ESTs (Marcas de sequências expressas), e as sequências foram reunidas em 43.141 cluster (Menossi, 2008). As seqüências do SUCEST foram obtidas a partir de clones de 37 bibliotecas de cDNA construídas a partir de diferentes tecidos ou condições fisiológicas da planta (Vettore, 2001). Inicialmente, os transcritos RNA mensageiro (RNAm) em um tecido ou condição foram submetidos a transcrição reversa para obtenção de moléculas de cDNA, os cDNAs obtidos foram clonados em vetores plasmidiais (Vettore, 2001). A partir desse ponto seguiu-se o seqüenciamento, as seqüências obtidas são denominadas “Expressed Sequence Tags” (ESTs) (Vettore, 2001). Os clones de cDNA seqüenciados são denominados “Reads” e são agrupados de acordo com seu grau de homologia em um mesmo arquivo o qual é denominado Cluster (Ferreira, 2002). Os clones (Reads) armazenados em um mesmo cluster são inferidos como transcritos de um mesmo gene ou, no do mesmo alelo (Vettore, 2003). Posteriormente, os clusters passaram por um processo de anotação que consiste na identificação de suas seqüências através de alinhamento no banco de dados “National Center for Biotechnology Information” (NCBI).

As seqüências do projeto SUCEST foram armazenadas em um banco de dados privado (<http://sucest.lad.dcc.unicamp.br/en/>) e agrupadas de acordo com seu tecido e condição fisiológica. Dentre as condições fisiológicas analisadas, cabe

ressaltar as bibliotecas de cDNA (denominadas AD1, HR1) obtidas de plantas inoculadas com as bactérias endofíticas fixadoras de nitrogênio *G. diazotrophicus* e *H. rubrisubalbicans* (Vettore et al., 2003). O acesso ao banco é restrito aos grupos que participaram em uma das etapas do projeto como sequenciamento ou anotação.

Através de alinhamento de seqüências dentro do próprio banco de dados do SUCEST utilizando a ferramenta “Basic Local Alignment Search Tool” (BLAST) podemos fazer uma busca de seqüências em cana com similaridades às proteínas ou domínios de proteínas já caracterizadas na literatura. A ferramenta BLAST no SUCEST permite identificar uma suposta seqüência em cana-de-açúcar similar a uma proteína alvo de interesse.

3-Banco de dados

Um grande conjunto de seqüências de projetos Genomas, Transcriptomas e Proteomas estão disponíveis na rede (“World Wide Web”), armazenados em bancos de dados. Dentre estes o “GenBank” é um banco de dados público construído pelo NCBI uma divisão da biblioteca nacional de medicina (NLM) localizada em Bethesda, Estados Unidos no campus do Instituto Nacional de Saúde. O “GenBank” hospedado no site do NCBI incorpora as seqüências submetidas para o banco de dados do “European Molecular Biology Laboratory” (EMBL) - Reino Unido e “DNA Data Bank of Japan” (DDBJ) (Benson, et al., 2003). Através de colaboração internacional entre os três bancos de dados, as informações são trocadas diariamente para assegurar uma coleção uniforme de seqüências (Benson, et al., 2003). O “Genbank” prove o acesso para mais de 370 genomas de microrganismos, >2.450 seqüências genômicas de vírus, >1.050 seqüências referentes organelas e mais de 20 genomas de eucariotos (Wheeler et al., 2006). Além dessas o NCBI contém diversas outras ferramentas e dados biológicos (estruturas de proteínas e literatura biomédica).

O NCBI disponibiliza um programa de alinhamento denominado BLAST, através do qual podemos utilizar seqüências de aminoácidos, RNAm ou DNA de uma espécie para procurar seqüências similares armazenadas no “Genbank”. O programa é baseado em um algoritmo heurístico que promove alinhamentos globais, a

detecção da relação entre as seqüências e avaliação das regiões similares entre elas (Wheeler, 2007). Para cada alinhamento retornado pelo BLAST é atribuído um valor ou medida de significância estatística denominada “Expectation Value” (e-value) (Wheeler et al., 2006). O alinhamento recuperado pelo BLAST pode ser limitado pelo “e-value” atribuído no início da análise (Wheeler et al., 2006).

Além do NCBI existem outros bancos de dados públicos tais como TIGR Gene Índices e “Munich Information Centre for Protein Seqüência” (MIPs). O banco de dados MIPs e “TIGR Gene Índices” são excelentes ferramentas para identificação de genes em plantas.

O banco de dados MIPs é hospedado em Neuherberg na Alemanha, e provê acesso aos genomas de *Z. mays*, *M. truncatula*, *L. japonicus*, *O. sativa*, *L. esculentum*, *S. bicolor* e *A. thaliana*. O MIPs desenvolve e mantém automaticamente anotado genomas de outros bancos de dados e provê ferramentas para análises de seqüências de proteínas (Mewes et al., 2000).

O banco de dados “TIGR Gene Indices” contém uma coleção de ESTs de mais de 77 espécies que incluem animais, plantas e protistas. O “TIGR Gene Indices” baseia-se em um protocolo altamente refinado para analisar os genes e seqüências EST na tentativa de identificar e caracterizar os transcritos expressos, e assim, permitir o acesso destes na rede (Lee et al., 2005). O “TIGR Gene Índices” contém 47 bancos de dados EST de plantas nos quais estão incluídas monocotiledôneas e dicotiledôneas.

Através dos bancos de dados e ferramentas de bioinformática disponíveis na rede podemos identificar e comparar seqüências homólogas (aminoácidos ou nucleotídeos) de diferentes organismos e assim responder diferentes perguntas biológicas.

As proteínas homólogas compartilham um ancestral comum e podem ser caracterizadas como ortólogas ou parálogas. Os ortólogos surgiram a partir de um ancestral comum por um evento de especiação, e os parálogos são relacionados por um evento de duplicação. As proteínas ortólogas em geral possuem funções similares ou idênticas, enquanto os parálogos podem adquirir novo papel funcional (Vicentini et al., 2009).

Na análise genômica comparativa, o agrupamento de ortólogos provê uma maneira de investigar a conservação de famílias de genes e processos biológicos. As sequências conservadas podem ser usadas para inferir a história evolutiva, e homologies implicam funções bioquímicas conservadas (Vicentini et al., 2009).

3.1- “Northern” Eletrônico e bancos de dados de expressão gênica

Após a identificação de uma sequência em cana-de-açúcar, podemos analisar a ocorrência da seqüência de cDNA (correspondente a um RNAm para um referido gene) nas 37 bibliotecas de cDNA do projeto SUCEST, específicas para determinado tecido ou condição (Ferreira, 2002). Essa abordagem é conhecida como análise da expressão gênica “in silico” ou “Northern” Eletrônico e baseia-se na contabilização do número de vezes que cada ESTs aparece nas diferentes bibliotecas de cDNA (Rafalski et al., 1998). Através dessa abordagem podemos estimar o padrão de expressão de um determinado gene e correlacionar sua ocorrência com seu papel funcional. Os dados de “Northern” Eletrônico podem ser confirmados por diferentes metodologias tais como “Microarray”, RT-PCR, RT-qPCR (PCR em tempo real), entre outras.

Alem do banco de dados SUCEST, um grande conjunto de dados de microarray de diferentes espécies vegetais tem sido disponibilizado na rede através de banco de dados, oferecendo uma excelente abordagem para um estudo comparativo de genômica funcional em plantas. Entre esses podemos citar os bancos de *A. thaliana*: “Nottingham Arabidopsis Stock Centre's microarray” (NASCArrays), AtgenExpress e “Arabidopsis thaliana trans-factor and cis-element prediction database” (ATTED-II). Esses bancos de dados possibilitam análises “in silico” sobre o padrão de expressão de genes de *A. thaliana* em diversos tecidos ou condições fisiológicas. Através desses bancos de dados pode-se selecionar genes em *A. thaliana* com perfil de expressão associado às respostas a estresses bióticos, abióticos ou genes com padrão de expressão diferencial na presença de compostos tais como hormônios ou eliciadores de defesa. Posteriormente, com auxílio de programas de bioinformática e bancos de dados é possível identificar genes com homologia em outras espécies vegetais, como por exemplo, cana-de-açúcar.

Portanto, o acesso e o estudo de informações depositadas em bancos de dados de microarray através de ferramentas de bioinformática, consistem em um poderoso método para identificação e caracterização preliminar de genes (Rensink e Buell, 2005). O estudo das análises de transcrição de genes “in silico” pode viabilizar à descoberta de funções biológicas para uma proteína.

4.0-OBJETIVO GERAL

O presente trabalho teve como objetivo identificar, em cana-de-açúcar, receptores do tipo NB-LRRs e LRR-RLKs e avaliar sua ativação durante a associação entre cana-de-açúcar e endófitos.

4.1- OBJETIVOS ESPECÍFICOS

- Identificar proteínas LRR-RLKs e NB-LRRs em cana-de-açúcar ativadas durante a interação com bactérias endofíticas através de análises “in silico” junto ao banco de dados SUCEST.
- Analisar “in silico” o padrão de expressão dos receptores LRR-RLKs ou NB-LRR identificados, nas diferentes bibliotecas do SUCEST.
- Alinhar a seqüência dos receptores (NB-LRRs e LRR-RLKs) identificados em cana-de-açúcar com seqüências homólogas ou similares em outras espécies, visando identificar as regiões conservadas e o nível de similaridade.
- Realizar análises filogenéticas dos receptores identificados em cana-de-açúcar e seqüências similares em outras espécies vegetais visando inferir sobre o grau de proximidade e a funcionalidade destes genes.

5- MATERIAIS E MÉTODOS

5.1- Utilização das bibliotecas de cDNA do SUCEST

Os receptores em cana-de-açúcar foram identificados utilizando os dados de ESTs (“Expressed Sequence Tags”) armazenados no SUCEST (<http://sucest.lad.ic.unicamp.br/en/>). O SUCEST possui em seu banco de dados 37 bibliotecas de cDNA, sendo cada biblioteca correspondente a um determinado tecido ou órgão da cana-de-açúcar sob uma dada condição fisiológica. Das 37 bibliotecas de cDNA existentes no SUCEST, 13 bibliotecas têm um número de ESTs menor que 5000, e foram excluídas das análises, no presente trabalho. Assim foram utilizadas apenas 24 bibliotecas de cDNA para a realização das análises dos dados: AD1, AM1, AM2, CL6, FL1, FL3, FL4, FL5, FL8, HR1, LB1, LB2, LR1, LV1, RT1, RT2, RT3, RZ2, RZ3, SB1, SD1, SD2, ST1, ST3 (Quadro 2).

Quadro 2 - Lista descritiva das Bibliotecas de cDNA do SUCEST

Código das Bibliotecas SUCEST	Descrição da biblioteca	Número de "Reads"
AD1	Plantas "in vitro" inoculadas com <i>Gluconacetobacter diazotrophicus</i>	18137
AM1	Meristema apical de folhas jovens e caules de canasmaduras(var. SP80-3280)	12480
AM2	Meristema Apical de folhas jovens e caules e canas maduras(var. SP80-3280)	15648
CL6	Calos submetidos a estresse luz(escuras) e estresse por temperatura 4°C e 37°C	7392
FL1	Flor de cana com base medindo 1cm de comprimento (var. SP8087432)	18523
FL3	Flor de cana com base medindo 5cm de comprimento (var. PB5211 PB57150-4)	13056
FL4	Pedúnculo floral com 50 cm de comprimento (var. PB5211PB57150-4)	16885
FL5	Pedúnculo floral com 20cm de comprimento (var. PB5211PB57150-4)	10080
FL8	Pedúnculo floral com 10cm de comprimento (var. PB5211PB57150-4)	5167
HR1	Plantas "in vitro" inoculadas com <i>Herbaspirillum rubrisubalbicans</i> (var. SP70-1143)	12000
LB1	Broto lateral de plantas de cana (var. SP80-3280)	7488
LB2	Broto lateral de plantas de cana adultas	10559
LR1	"Leaf roll" ou cartucho foliar de plantas de cana-de-açúcar (var. SP80-3280)	14036
LV1	Folhas estioladas de plântulas (SP83-5077, SP80-185,SP87-396, SP80-3280 e SP803280 X SP81-5441)	6432
RT1	Raiz de cana-de-açúcar crescidas no solo (var. SP80-3280)	8640
RT2	Raiz de cana-de-açúcar crescidas no solo (var. SP80-3280)	12288
RT3	Raiz de plantas adultas de cana-de-açúcar crescidas no solo	10559
RZ2	Zona de transição entre raiz/folha de plantas de cana (var. SP80-3280)	5760
RZ3	Zona de transição entre raiz/folha de plantas de cana imaturas (var. SP80-3280)	15168
SB1	Cascas de plantas de cana crescidas em casa de vegetação (var. SP80-3280)	16318
SD1	Sementes de cana-de-açúcar em diferentes estágios de desenvolvimento (var. RB805028)	11040
SD2	Sementes de cana-de-açúcar em diferentes estágios de desenvolvimento (var. RB805028)	10366
ST1	Primeiro interno de plantas de cana-de-açúcar (var. SP803280)	8448
ST3	Quarto interno de plantas de cana-de-açúcar (var. SP803280)	11930

5.2 - Geração dos dados de expressão gênica

A frequência dos “Reads” nos tecidos/condições em cana-de-açúcar referentes a cada proteína foi analisada em um programa denominado proteínas “on line” desenvolvido pela Doutoranda Verônica Silva, linguagem de programação Java. Para tanto os códigos dos READs referentes a cada cluster foram obtidos do banco de dados SUCEST e salvos em arquivos txt. O programa faz a leitura do arquivo txt (bloco de nota) para avaliar a frequência dos reads nos diversos tecido/condição e gera um gráfico correspondente à frequência desses transcritos.

Para comparar a frequência de um dado clone (“cluster”) em diferentes tecidos/condições e bibliotecas, o programa faz a normalização dos números de “Reads”; com as seguintes etapas:

Passo 1: Calcula fator tecido específico

Fator tecido específico= [Número de “Reads” da maior biblioteca / número de total de “Reads” da biblioteca analisada]

Passo 2: Calcula fator da biblioteca

Fator = FB= [Fator tecido específico X número total de “Reads” da biblioteca analisada]

Passo 3: Número de clones

Número de clones= soma dos Fatores das bibliotecas que sejam similares (Ex: FL= soma do fator de FL1 à FL8).

Após a normalização do número de “Reads” representado em cada tecido/condição e em cada biblioteca é gerado um gráfico nos planos (x, y) que correlaciona o tecido/condição ou as bibliotecas de cDNA do SUCEST, e o número de clones (soma dos fatores das bibliotecas similares) (Quadro 3).

Quadro 3 - Relação dos grupos correspondentes aos tecidos de cana-de-açúcar obtidos da reunião das bibliotecas de cDNA de cana-de-açúcar do Projeto SUCEST

Tecido	Descrição
AD	Clones das bibliotecas de cDNA de AD1
AM	União dos clones das bibliotecas de cDNA AM1 e AM2
CL	Clones das bibliotecas de cDNA CL6
FL	União dos clones das bibliotecas de cDNA FL1 A FL8
HR	Clones das bibliotecas de cDNA HR1
LB	União dos clones das bibliotecas de cDNA LB1 e LB2
LR	Clone da biblioteca de cDNA LR1
LV	Clones da bibliotecas de cDNA LV1
RT	União dos clones das bibliotecas de cDNA RT1 a RT3
RZ	União dos clones das bibliotecas de cDNA RZ2 e RZ3
SB	Clones da biblioteca de cDNA SB1
SD	União dos clones das bibliotecas de cDNA SD1 e SD2
ST	União dos clones das bibliotecas de cDNA ST1 e ST2

5.3- Identificação de genes de cana-de-açúcar que codificam proteínas LRR-RLKs induzidos durante a interação com endófitos.

Para identificação e seleção de proteínas LRR-RLKs reguladas por endófitos no SUCEST foram utilizadas duas abordagens.

5.3.1- ETAPAS DA ABORDAGEM 1:

- 1) Seleção de proteínas LRR-RLKs caracterizadas por participarem da resposta de defesa e percepção de patógenos. Para tanto, foram selecionadas três proteínas LRR-RLKs caracterizadas como receptores de MAMPs, as proteínas FLS2 e EFR. Estes são receptores de flagelina e eflu MAMPs encontrados em diversas bactérias patogênicas e não patogênicas incluindo *G. diazotrophicus* e *H. rubrisubalbicans*;
- 2) Obtenção da sequência em aminoácidos das proteínas (OsFLS2, FLS2 e EFR) através do banco “Gen bank” (NCBI). O número de acesso das proteínas citadas são apresentados no Quadro 4:

Quadro 4: Proteínas LRR-RLKs selecionadas no NCBI

Espécie	ID/NCBI	Proteína
<i>O. sativa</i>	NP_001053894.1	OsFLS2
<i>A. thaliana</i>	NP_199445.1	FLS2
<i>A. thaliana</i>	NP_197548	EFR

- 3) Obtenção do domínio tirosina cinase das proteínas citadas no Quadro 4 através do programa SMART (<http://smart.embl-heidelberg.de/>);
- 4) Para identificação de sequências similares aos receptores descritos no Quadro 4. A ferramenta BLAST, especificamente “tblastn” localizada dentro do banco de dados do SUCEST (<http://sucest.lad.ic.unicamp.br/en/>) foi

utilizada. Para tanto, os domínios tirosina cinase desses receptores (Quadro 4) foram utilizados como chave de entrada no tblastn;

- 5) Os clusters que apresentam “Reads” nas bibliotecas AD (plantas inoculadas com *G.diazotrophicus*) e/ou HR (plantas inoculadas com *H.rubrisubalbicans*) foram selecionados;
- 6) Obtenção da sequência em aminoácidos do cluster selecionado no banco de dados SUCEST;
- 7) Análise da sequência em aminoácidos quanto à presença dos seguintes domínios:

Quadro 5: Domínios característicos de proteínas LRR-RLKs

Regiões	Domínios presentes em proteínas LRR-RLKs
LRR	Regiões ricas em leucina
TM	Transmembrana
cinase	Serina e/ou treonina

- 8) As proteínas de cana-de-açúcar que apresentam os domínios característicos de LRR-RLKs foram selecionadas para análises.

5.3.2- ETAPAS DA ABORDAGEM 2:

- 1) Obtenção da regiões ricas em leucina das proteínas citadas no Quadro 4 através do programa SMART(<http://smart.embl-heidelberg.de/>);
- 2) Para identificação de sequências similares aos receptores descritos no QUADRO 4. A ferramenta BLAST, especificamente “tblastn” localizada dentro do banco de dados do SUCEST foi utilizada. Para tanto, as regiões ricas em leucina

desses receptores (Quadro 4) foram utilizados como chave de entrada na ferramenta tblastn;

- 3) Repetição das etapas 5 à 8 do item 5.3.1.

5.4- Identificação de genes de cana-de-açúcar que codificam proteínas NB-LRR reguladas por endófitos.

Para identificação de proteínas CC-NB-LRR ou TIR-NB-LRR reguladas por endófitos no SUCEST foram utilizadas duas abordagens.

5.4.1- ETAPAS DA ABORDAGEM 1:

- 1) Seleção de proteínas NB-LRRs caracterizadas por participarem da resposta de defesa contra patógenos. Para tanto, foram selecionadas 4 proteínas NB-LRRs, as proteínas RPS2, RPM1, ADR1 e RPS4. Estas foram selecionadas pois são proteínas de resistência induzidas em resposta à patógenos;
- 2) Obtenção da sequência em aminoácidos das proteínas NB-LRRs (RPS2, RPM1, ADR1 e RPS4) no banco de dados “gen bank” (NCBI). O número de acesso correspondente a tais proteínas são:

Quadro 6: Proteínas NB-LRR selecionada no banco de dados NCBI

Espécie	ID/NCBI	Nome da Proteína
<i>A. thaliana</i>	NP_174620	ADR1
<i>A. thaliana</i>	NP_199338.1	RPS4 (■)
<i>A. thaliana</i>	ABU44507.1	RPS2
<i>A. thaliana</i>	NP_187360.1	RPM1

Proteína com domínio TIR-NB-LRR (■). Espécie- *A.thaliana*, ID- número de acesso no NCBI.

3) Obtenção da região NB-ARC ou TIR das proteínas citadas no Quadro 6 através do programa SMART;

4) Para identificação de sequências similares aos receptores descritos no Quadro 6. A ferramenta BLAST, especificamente “tblastn” localizada dentro do banco de dados do SUCEST foi utilizada. Para tanto, os domínios NB-ARC ou TIR das proteínas citadas no Quadro 6 foram utilizados como chave de entrada no tblastn;

5) Os clusters que apresentavam “Reads” nas bibliotecas AD (plantas inoculadas com *G.diazotrophicus*) e/ou HR(plantas inoculadas com *H.rubrisubalbicans*) foram selecionados;

6) Obtenção da sequência em aminoácidos dos clusters selecionados no banco SUCEST;

7) Análise da sequência em aminoácidos quanto à presença dos seguintes domínios:

Quadro 7: Domínios característicos de proteínas NB-LRR

Domínios presentes em proteínas NBS-LRR	Domínios presentes em proteínas NBS-LRR
LRR	Regiões ricas em leucinas
NB-ARC	ATPase e “nucleotide-binding state”
TIR	Toll- receptor interleucina

8) As proteínas que apresentavam os domínios descritos na Quadro 7 foram selecionadas para outras análises.

5.4.2- ETAPAS DA ABORDAGEM 2:

- 1) Obtenção das proteínas citadas no Quadro 6 através do banco de dados NCBI (<http://www.ncbi.nlm.nih.gov/>);
- 2) Para identificação de sequências similares aos receptores descritos no Quadro 6. A ferramenta BLAST, especificamente “tblastn” localizada dentro do banco de dados do SUCEST foi utilizada. Para tanto, as proteínas (Quadro 6) foram utilizados como chave de entrada na ferramenta tblastn;
- 3) Repetição das etapas 5 à 8 item 6.1.

5.5- Comparações entre a sequência dos receptores LRR-RLKs identificados em cana-de-açúcar com sequências similares em outras espécies vegetais

Sequências similares às proteínas LRR-RLKs de cana-de-açúcar foram identificadas em outras espécies vegetais através dos bancos de dados MIPS (mips.gsf.de/projects/plants) e NCBI (<http://blast.ncbi.nlm.nih.gov/>) utilizando a ferramenta blastp (compara sequências de aminoácidos de proteínas) e no projeto gene Index (<http://compbio.dfci.harvard.edu>) utilizando a ferramenta blastn. As sequências completas ou parciais (domínio tirosina cinase) das proteínas LRR-RLKs identificadas em cana-de-açúcar foram alinhadas e comparadas com os seus homólogos em outras espécies vegetais utilizando o programa Clustalw2 ([HTTP://.ebi.ac.uK/Tools/clustalw2/](http://www.ebi.ac.uk/Tools/clustalw2/)), ferramenta Align.

5.6- Análises filogenética dos receptores LRR-RLKs identificados em cana-de-açúcar

Para construção filogenética das proteínas ricas em leucina identificados em cana-de-açúcar, as sequências em aminoácidos das proteínas codificadas por tais genes foram alinhadas no NCBI utilizando a ferramenta psi-blastp. Para tanto os seguintes parâmetros foram utilizados: “expect threshold: 1, matrix blosum e psi-threshold: 0, 001”. Sequências similares com $\text{eval} \geq 1e^{-26}$ recuperadas no NCBI,

proteínas SYMRK e sequências similares em outras espécies foram utilizadas para análise filogenética. Para tanto, os domínios tirosina cinases das proteínas LRR-RLKs foram alinhados utilizando o programa clustalX (Thompson et al., 1994, <http://align.genome.jp/>) usando a proteína S61006 de *Saccharomyces cerevisiae* como grupo externo. Os seguintes parâmetros foram utilizados no ClustalW: alinhamento “Pairwise– Gap opening, 35.0, Gap extension, 0.75; Multiple alignment – Gap opening, 15.00, Gap extension”, 0.30. O resultado do alinhamento no clustalw em formato PHYLIP foi submetido ao software PHYLM (<http://www.atgc-montpellier.fr/phylm/>). Esse software gera árvores filogenéticas pelo método “maximum likelihood” (Guindon et al., 2003). Os seguintes parâmetros foram utilizados no software: modelo de substituição JTT, 100 “bootstrapped”, “4 Number of substitution rate categories”. O resultado obtido no PHYLM em formato Newick foi visualizado no programa MEGA 4.0.

5.7- Identificação de sequências similares à ADR1 (NB-LRR) em diferentes espécies vegetais e análises filogenéticas

Sequências similares a ADR1 (NB-LRR) foram identificadas no projeto Gene-índice (<http://compbio.dfci.harvard.edu/tqi/>), banco de dados MIPs (“Munich Information Centre for Protein Sequences”, <http://mips.gsf.de/>) e NCBI. Para tanto, a sequência em aminoácidos de ADR1(At1g33560) foi alinhada no banco de dados do projeto gene índice utilizando a ferramenta tblastn e no MIPs ferramenta blastp e NCBI ferramenta psi-blast. As sequências em aminoácidos da região NBS-ARC de ADR1 e seus homólogos em diferentes espécies vegetais foram alinhadas utilizando o programa ClustalX (Thompson et al., 1994, <http://align.genome.jp/>) usando a proteína CED4 de *Caenorhabditis elegans* como grupo externo. Os seguintes parâmetros foram utilizados no ClustalW: alinhamento “Pairwise– Gap opening, 35.0, Gap extension, 0.75; Multiple alignment – Gap opening, 15.00, Gap extension”, 0.30. O resultado do alinhamento no clustalw em formato PHYLIP foi submetido ao software PHYLM (<http://www.atgc-montpellier.fr/phylm/>). Esse software gera árvores filogenéticas pelo método “maximum likelihood” (Guindon et al., 2003). Os seguintes parâmetros foram utilizados no software: modelo de substituição JTT, 100

“bootstrapped”, “4 Number of substitution rate categories”. O resultado obtido no PHYLIM em formato Newick foi visualizado no programa MEGA 4.0.

5.8 - Análise de domínios presentes nas proteínas e busca de motivos

O modelo de Markov da ferramenta do simples módulo de arquitetura (programa SMART, <http://smart.embl-heidelberg.de/>) foi usado para identificar os domínios presentes nas proteínas e obter a seqüência em aminoácidos que corresponde a esses domínios. Foram realizadas análises de “motivos” das proteínas através do programa MEME (“Motif-based sequence analysis tools”, http://meme.sdsc.edu/meme4_3_0/cgi-bin/meme.cgi).

6-0-Resultados

6.1-“Clusters” que codificam proteínas “LRR-RLKs” identificados no banco de dados SUCEST.

Os organismos são capazes de perceber estímulos externos e processar a informação ao nível celular. Em plantas essa percepção é mediada por receptores do tipo LRR-RLKs. Nesse sentido uma das metas desse trabalho foi à identificação de receptores LRR-RLKs em cana-de-açúcar induzidos diferencialmente por *G. diazotrophicus* e/ou *H. rubrisubalbicans*. As sequências total ou parcial dos receptores FLS2, OsFLS2 e EFR foram utilizados para busca de receptores LRR-RLKs no banco de dados do SUCEST. Assim foram selecionados 79 clusters candidatos a receptores LRR-RLKs, estes apresentam “Reads” nas bibliotecas de plantas inoculadas com *G. diazotrophicus* e/ou *H. rubrisubalbicans*, exceto para ScFLS2 (Quadro 8).

Contudo, análises dos domínios das sequências em aminoácidos desses “Clusters” através do programa SMART revelaram que apenas 9 (Quadro 9) são possíveis receptores LRR-RLKs, pois apresentam regiões ricas em leucina no domínio N-terminal, região transmembrana e domínio cinase citoplasmático. Estes foram selecionados para outras análises.

Tais “Clusters” codificam proteínas LRR-RLKs que apresentam similaridades com proteínas LRR-RLKs de *A.thaliana*. Dentre estes, 6 “Clusters”(Quadro 9) codificam proteínas que apresentam similaridades com proteínas LRR-RLKs já caracterizadas, tais como: FLS2, SERK 1, SERK 2 e SRF7 (*sclrr-rlk 2* e *sclrr-rlk3*), destacados em (▲) no Quadro 9.

Quadro 8- Transcritos similares à proteína LRR-LRKs encontrados em cana-de-açúcar e proteínas similares em *A.thaliana*

TAIR – número de acesso no banco de dados TAIR (“The Arabidopsis Information Resource”), SUCEST- número de acesso no banco de dados SUCEST, I - Nível de Identidade obtido no blastp/NCBI, “e-value” - probabilidade de ser a sequência esperada, Total- número de “Reads” totais correspondente ao cluster analisado, HR- número de “Reads” na biblioteca HR, AD- número de “Reads” na biblioteca AD, soma- soma de “Reads” da biblioteca AD com HR, %HR- porcentagem de “Reads” induzidos na biblioteca HR, %AD- porcentagem de “Reads” induzidos na biblioteca AD, % total- soma das porcentagens de AD com HR, Fator= soma total de “Reads” (soma de AD com HR) multiplicado pela porcentagem total (soma de %AD com %HR). LRR- região rica em leucina, TM- região transmembrana, (●) – proteínas inibidoras de galacturonase, (■) – proteínas LRR-RLKs

TAIR	SUCEST	Similaridade SUCEST/NCBI	I	Evalue	Total	HR	AD	soma	%HR	%AD	% -total	Fator
Domínios identificados: LRR e peptídeo sinal												
At3g56100	SCBGAD1027E01.g	Meristematic receptor like cinase	49%	2,00E-65	2	0	1	1	0	50	50	50
At1g79620	SCBFLR1060F03.g	Leucine-rich repeat transmembrane protein	61%	1,00E-52	4	1	0	1	25	0	25	25
At5g06860	SCVPLR2019B03.g	Polygalacturonase inhibiting protein 1	37%	6,00E-50	72	6	6	12	8,333	8,333	16,667	200
At3g49750	SCCCLR2C01A11.g	Leucine-rich repeat family protein	57%	2,00E-63	6	4	0	4	66,67	0,0	66,7	267
Domínio identificado: LRR												
AT2G26330	SCCCLR1001A03.g	ERECTA	67%	0,0	46	0	2	2	0	4,348	4,3478	9
At2g24230	SCBGLR1095F05.g	Leucine-rich repeat transmembrane	38%	3,00E-46	2	0	1	1	0	50	50	50
At1g67510	SCSFST3076D08.g	Leucine-rich repeat family protein	56%	7,00E-42	3	0	1	1	0	33,33	33,333	33
At3g22800	SCQSLB1052C11.g	Leucine-rich repeat family protein	59%	3,00E-99	3	0	1	1	0	33,33	33,333	33
At5g21090	SCEPLR1030B03.g	Leucine-rich repeat protein	63%	4,00E-75	31	2	0	2	6,452	0	6,4516	13
At4g13340	SCCCCL4009B01.g	Leucine-rich repeat family protein	62%	2,00E-124	12	1	1	2	8,333	8,333	16,667	33
At4g06744	SCACHR1038G07.g	Leucine-rich repeat family protein	43%	1,00E-70	9	1	1	2	11,11	11,11	22,222	44
At4g03390	SCSGAD1005E03.g	Strubbelig-receptor gamily 3	35%	7,00E-36	1	0	1	1	0	100	100	100
At1g48480	SCCCLR1C05E11.g	Receptor-like kinase 1	42%	1,00E-58	15	0	2	2	0	13,33	13,333	27
At5g06860	SCQGHR1012F03.g	Polygalacturonase inhibiting protein 1 (●)	34%	1,00E-36	2	1	0	1	50	0	50	50
At5g06860	SCSFAD1125C08.g	Polygalacturonase inhibiting protein 1 (●)	39%	9,00E-43	3	0	1	1	0	33,33	33,333	33
At5g06860	SCRUAD1063C06.g	Polygalacturonase inhibiting protein 1 (●)	43%	3,00E-56	4	1	2	3	25	50	75	225
At1g28340	SCJFRZ2033D09.g	Leucine-rich repeat family protein	65%	5,00E-139	7	0	1	1	0	14,29	14,286	14
At3g13380	SCRCLL6033D10.g	Brassinosteroid insensitive like 3	58%	9,00E-120	3	0	1	1	0	33,33	33,333	33
At5g10290	SCEQZR3020C02.g	Leucine-rich repeat family protein	38%	1,00E-39	8	0	1	1	0	12,5	12,5	13
At5g21090	SCAGLR1043F01.g	Leucine-rich repeat protein	82%	6,00E-94	65	0	1	1	0	1,538	1,5385	2
At4g20940	SCEZHR1086F04.g	Leucine-rich repeat family protein	65%	3,00E-58	2	1	0	1	50	0	50	50
At4g20140	SCSGCL6070B09.g	Leucine-rich repeat transmembrane protein kinase	47%	3,00E-58	3	1	0	1	33,33	0	33,333	33
Domínios identificados: LRR-TM												
At3g24480	SCCCCL4002E02.g	Leucine-rich repeat family protein	57%	1,00E-108	39	1	4	5	2,564	10,26	12,821	64
At5g25930	SCMCRZ3065H07.g	Leucine-rich repeat family protein	40%	8,00E-54	3	1	0	1	33,33	0	33,333	33
At5g10290	SCJFRZ2028F11.g	Leucine-rich repeat family protein	70%	0,0	23	0	3	3	0	13,04	13,043	39
At1g53730	SCJLST1027G05.g	Strubbelig Receptor Family 7	48%	1,00E-62	5	0	1	1	0	20	20	20
At4g03390	SCJLRZ1021D12.g	Strubbelig Receptor Family 3	48%	3,00E-61	12	0	2	2	0	16,67	16,667	33
At4g13340	SCCCCL5001D06.g	leucine-rich repeat family protein	62%	8,00E-78	5	0	1	1	0	20	20	20
Domínios identificados: LRR-TM-cinase												
At5g46330	SCBFRT3095E07.g	Flagellin sensitive 2	50%	2,00E-83	2	0	0	0	0	0	0	0
At1g71830	SCEQRT1029H05.g	Somatics embryogenesis receptor like kinase 1 (■)	79%	0,0	16	3	2	5	18,75	12,5	31,25	156
At1g34210	SCMCLR1053B12.g	Somatics embryogenesis receptor like kinase 2 (■)	82%	0,0	31	0	1	1	0	3,226	3,2258	3
At4g36180	SCQGHR1012G04.g	ScMAR1 (■)	56%	2,00E-117	13	9	0	9	69,23	0	69,231	623
At5g10020	SCVPCL6041G02.g	ScLRR-RLK1 (■)	43%	2,00E-163	24	1	0	1	4,167	0	4,1667	4
At3g14350	SCVPST1062F08.g	ScLRR-RLK2 (■)	47%	3,00E-98	6	0	3	3	0	50	50	150
At3g14350	SCCCLR1072F06.g	ScLRR-RLK3 (■)	61%	0,0	45	0	4	4	0	8,889	8,8889	36
At3g03770	SCSGHR1067C04.g	ScLRR-RLK4 (■)	50%	0,0	20	2	0	2	10	0,0	10,0	20
At4g23740	SCSFRT2068B02.g	ScLRR-RLK5 (■)	54%	1,00E-166	6	0	2	2	0	33,33	33,333	67
At1g63430	SCSBHR1051B09.g	ScLRR-RLK6 (■)	48%	5,00E-136	7	1	1	2	14,29	14,29	28,571	57
Domínios identificados: LRR-cinase												
At1g67510	SCCCST1003H06.g	Leucine-rich repeat transmembrane	41%	3,00E-142	13	4	2	6	30,77	15,38	46,154	277
At1g31420	SCCCLB1023E12.g	Leucine-rich repeat transmembrane	67%	0,0	35	5	0	5	14,29	0	14,286	71
At1g48480	SCCCRZ1C01H05.g	Receptor-like kinase 1	55%	3,00E-171	27	0	1	1	0	3,704	3,7037	4
At1g07650	SCJFRZ2027C06.g	Leucine-rich repeat transmembrane protein -SKG6	55%	4,00E-152	5	0	1	1	0	20	20	20
Domínios identificados: TM-cinase												
At5g48380	SCQSHR1023C04.g	Leucine-rich repeat family protein	59%	0,0	15	1	5	6	6,667	33,33	40	240
At5g65700	SCBGAM1093D09.g	Barely any meristem 1	73%	0,0	5	0	1	1	0	20	20	20
At5g63930	SCCCLR1066D09.g	Leucine-rich repeat transmembrane protein	69%	8,00E-132	5	0	1	1	0	20	20	20
At3g59110	SCCCCL4007D06.g	Protein kinase family protein	61%	3,00E-168	23	0	3	3	0	13,04	13,043	39
At3g59110	SCRFRL1012B04.g	Protein kinase family protein	67%	2,00E-155	14	1	2	3	7,143	14,29	21,429	64
At5g25930	SCCCCL3001G07.b	Leucine-rich repeat family protein	45%	4,00E-116	15	1	1	2	6,667	6,667	13,333	27
At5g58300	SCSGLR1084G09.g	Leucine-rich repeat family protein	61%	1,00E-126	17	1	2	3	5,882	11,76	17,647	53
At3g47570	SCMCRZ2102E02.g	Leucine-rich repeat transmembrane	40%	6,00E-58	2	0	1	1	0	50	50	50
At2g27060	SCCCAM2C04G07.g	ATP binding / protein serine	45%	8,00E-122	14	0	4	4	0	28,57	28,571	114
At5g25930	SCEQRT1033H06.g	Leucine-rich repeat family protein (peptídeo sinal)	42%	1,00E-86	9	3	0	3	33,33	0	33,333	100
Domínio identificado: cinase												
At2g42800	SCSGHR1070E01.g	Leucine-rich repeat family protein (peptídeo sinal)	52%	1,00E-41	2	1	0	1	50	0	50	50
At4g23740	SCEZHR1047A01.g	Leucine-rich repeat transmembrane protein kinase	54%	1,00E-53	2	2	0	2	100	0	100	200
At1g66150	SCCCRZ1001E10.g	Transmembrane cinase 1	59%	0,0	18	1	0	1	5,556	0	5,5556	6
At4g20140	SCQSST3116D02.g	Leucine-rich repeat transmembrane protein kinase	57%	1,00E-105	6	0	1	1	0	16,7	16,7	17
At5g65700	SCRUFL1024D02.g	Big apical meristem 1	50%	3,00E-114	4	0	2	2	0	50	50	100
At5g65700	SCSBST3101B03.g	Barely any meristem 1	75%	9,00E-153	12	0	2	2	0	16,67	16,667	33
At1g17230	SCCCLR1C06C01.g	Protein binding / protein kinase	62%	5,00E-120	11	0	1	1	0	9,091	9,0909	9
At1g56130	SCAGLR1024B11.g	Leucine-rich repeat family protein	57%	0,0	20	3	0	3	15	0	15	45
At4g03230	SCCCLR2C02A04.g	S-locus lectin protein kinase	59%	2,00E-115	11	2	0	2	18,18	0	18,182	36
At1g26150	SCBGLR1097F07.g	Protein kinase	62%	2,00E-119	18	0	2	2	0	11,11	11,111	22
At4g02010	SCRFLB1056D11.g	Protein kinase family protein	79%	0,0	16	0	1	1	0	6,25	6,25	6
At4g06744	SCVPRZ2039B05.g	leucine-rich repeat family protein / extensin	49%	4,00E-79	10	1	0	1	10	0	10	10
At1g25570	SCMCAM1100B05.g	Leucine-rich repeat protein-related	49%	4,00E-75	11	0	1	1	0	9,091	9,0909	9
At2g02220	SCRCLAD1104A11.g	Phytosulfokin receptor 1	64%	1,00E-93	3	0	2	2	0	66,67	66,667	133
At3g02130	SCQSAD1057C09.g	Receptor-like kinase 1	40%	1,00E-25	1	1	1	1	0	100	200	0

Quadro 9 – Transcritos identificados em cana-de-açúcar que codificam proteínas LRR-RLKs e proteínas similares em *A.thaliana*

TAIR	SUCEST	Similaridade SUCEST/NCBI	I	Evalue	Total	HR	AD	soma	%HR	%AD	%-total	Fator
		Domínios identificados: LRR-transmembrana-cinase										
At5g46330	SCBFRT3095E07.g	<i>flagellin sensitive 2</i> (▲)	50%	2.00E-83	2	0	0	0	0	0	0	0
At1g34210	SCEQRT1029H05.g	<i>somatics embryogenesis receptor like kinase 1</i> (▲)	79%	0.0	16	3	2	5	18.75	12.5	31.25	156
At1g71830	SCMCLR1053B12.g	<i>somatics embryogenesis receptor like kinase 2</i> (▲)	82%	0.0	31	0	1	1	0	3.226	3.2258	3
At4g36180	SCQGHR1012G04.g	<i>scmar1</i> (■)	56%	2.00E-117	13	9	0	9	69.23	0	69.231	623
At5g10020	SCVPCL6041G02.g	<i>scrr-rlk1</i> (▲)	43%	2.00E-163	24	1	0	1	4.167	0	4.1667	4
At3g14350	SCVPST1062F08.g	<i>scrr-rlk 2</i> (▲)	47%	3.00E-98	6	0	3	3	0	50	50	150
At3g14350	SCCCLR1072F06.g	<i>scrr-rlk 3</i> (▲)	61%	0.0	45	0	4	4	0	8.889	8.8889	36
At3g03770	SCSGHR1067C04.g	<i>scrr-rlk 4</i>	50%	0.0	20	2	0	4	10	0.0	10.0	40
At4g23740	SCSFRT2068B02.g	<i>scrr-rlk5</i>	54%	1.00E-166	6	0	2	2	0	33.33	33.333	67
At1g63430	SCSBHR1051B09.g	<i>scrr-rlk6</i>	48%	5.00E-136	7	1	1	2	14.29	14.29	28.571	57

TAIR – número de acesso no banco de dados TAIR (“The Arabidopsis Information Resource”), SUCEST- número de acesso no banco de dados SUCEST, I - Nível de Identidade obtido no PSI-BlastP/NCBI, “e-value” - probabilidade de ser a sequência esperada, Total- número de “Reads” totais correspondente ao cluster analisado, HR- número de “Reads” na biblioteca HR, AD- número de “Reads” na biblioteca AD, soma- soma de “Reads” da biblioteca AD com HR, %HR- porcentagem de “Reads” induzidos na biblioteca HR, %AD- porcentagem de “Reads” induzidos na biblioteca AD, % total- soma das porcentagens de AD com HR, Fator= soma total de “Reads” (soma de AD com HR) multiplicado pela porcentagem total (soma de %AD com %HR), (▲)- proteínas LRR-RLKs similares a proteínas caracterizadas na literatura, (■)- proteína LRR-RLK altamente induzida na biblioteca HR.

Interessantemente, um dos “Clusters” LRR-RLK em cana-de-açúcar é altamente induzido durante a colonização por *H. rubrisubalbicans*, que corresponde a 69, 23% do total de “Reads”. Contudo, na biblioteca de plantas inoculadas com *G. diazotrophicus* o “Cluster” não é induzido (Quadro 9, (■)). Esse “Cluster” foi denominado *scmar1* (“Sugarcane Microbe Activated Receptor” 1). Conforme apresentado no Quadro 9, dentre os 10 receptores LRR-RLKs identificados o gene *scmar 1* é o que apresenta maior indução por *H. rubrisubalbicans*. O resultado do alinhamento de ScMAR1 no programa Blastp/NCBI mostrou que a proteína de cana-de-açúcar (monocotiledônea) apresenta 56% de identidade com a proteína At4g36180 de *A. thaliana* (dicotiledônea) (Quadro 9, (■)).

Pode ser observado, nos QUADROS 8 (●) e 10, que alguns dos demais “Clusters” identificados codificam proteínas similaridades à PGIP (proteína inibidora de galacturonase; região rica em leucina na parece celular) (Chisholm et al. 2006) e apresentam “Reads” em AD e/ou HR (Quadro 8).

Quadro 10: Similaridade entre as sequências ScPGIPs like 1 de cana-de-açúcar e as sequências PGIP 1 caracterizadas em outras espécies vegetais

Proteína em <i>Saccharum</i> sp.	SUCEST	Nome da proteína em outras sps.	Similaridade entre as sequências ScPGIPs like 1 X PGIP 1 de outras sps.			
			OsPGIP 1 (309 Aa)	TaPGIP 1 (337 Aa)	AtPGIP 1 (330 Aa)	GmPGIP 1(336 Aa)
ScPGIPs like 1 (435 Aa)	SCVPLR2019B03.g	Polygalacturonase inhibiting protein 1	36.3%	42.2%	40.4%	41.3%
ScPGIP like 1 (237 Aa)	SCQGHR1012F03.g	Polygalacturonase inhibiting protein 1	37.5%	38.6%	40.3%	38.9%
ScPGIP like 1 (393 Aa)	SCSFAD1125C08.g	Polygalacturonase inhibiting protein 1	32.4%	37.9%	35.3%	34.9%
ScPGIP like 1 (445 Aa)	SCRUAD1063C06.g	Polygalacturonase inhibiting protein 1	35.9%	36.5%	33.2%	29.1%

SUCEST- número de acesso, Aa : número de aminoácidos, Os: *O. sativa*, Ta: *T. aestivum*, At: *A. thaliana*, Gm: *G. max*. Número de acesso no NCBI: CAJ55691(OsPGIPs), CAJ55695.1 (TaPGIP1), BAB11144.1 (AtPGIP 1), CAI99393.1 (GmPGIP 1). Alinhamento no programa CLUSTALW2 (Larkin et al. 2007).

As proteínas PGIPs são membros da segunda classe de genes R (resistência) que correspondem proteínas com domínios LRR extracelular, tais como RLPs e LRR-RLKs (Chisholm et al. 2006) e desempenham papel na resposta imune contra fungos (Ferrari et al., 2006).

6.2- “Northen Eletrônico” das proteínas LRR-RLKs identificadas em cana-de-açúcar.

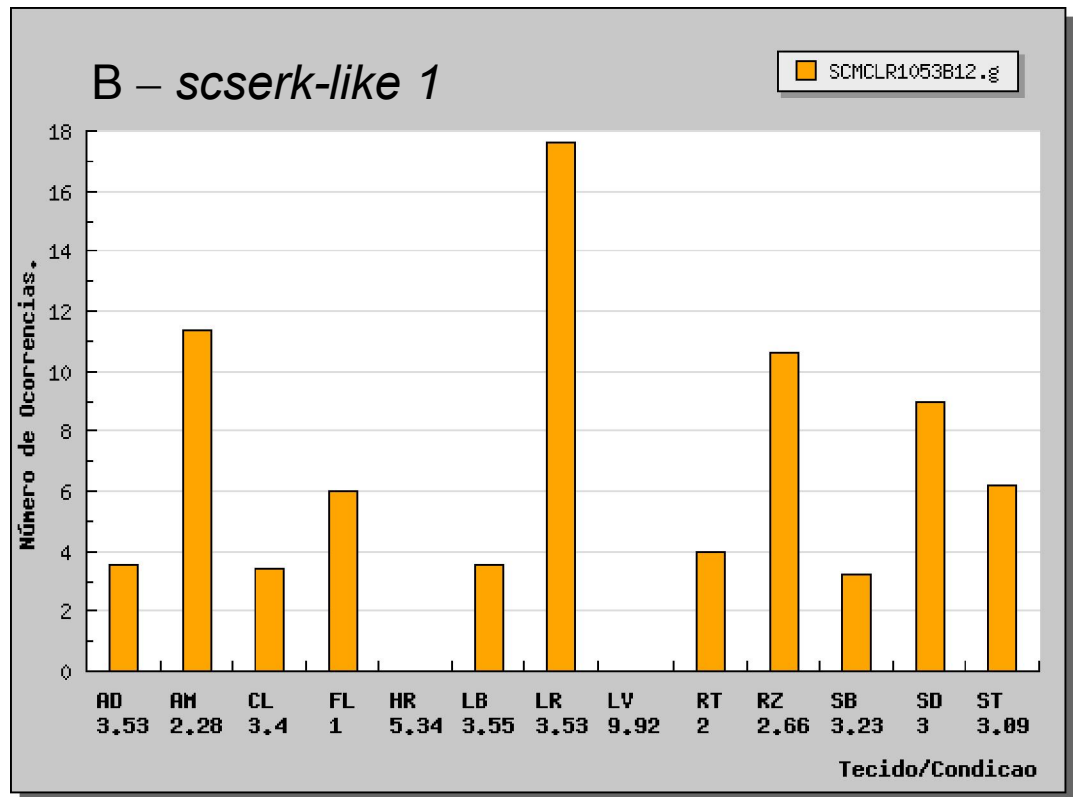
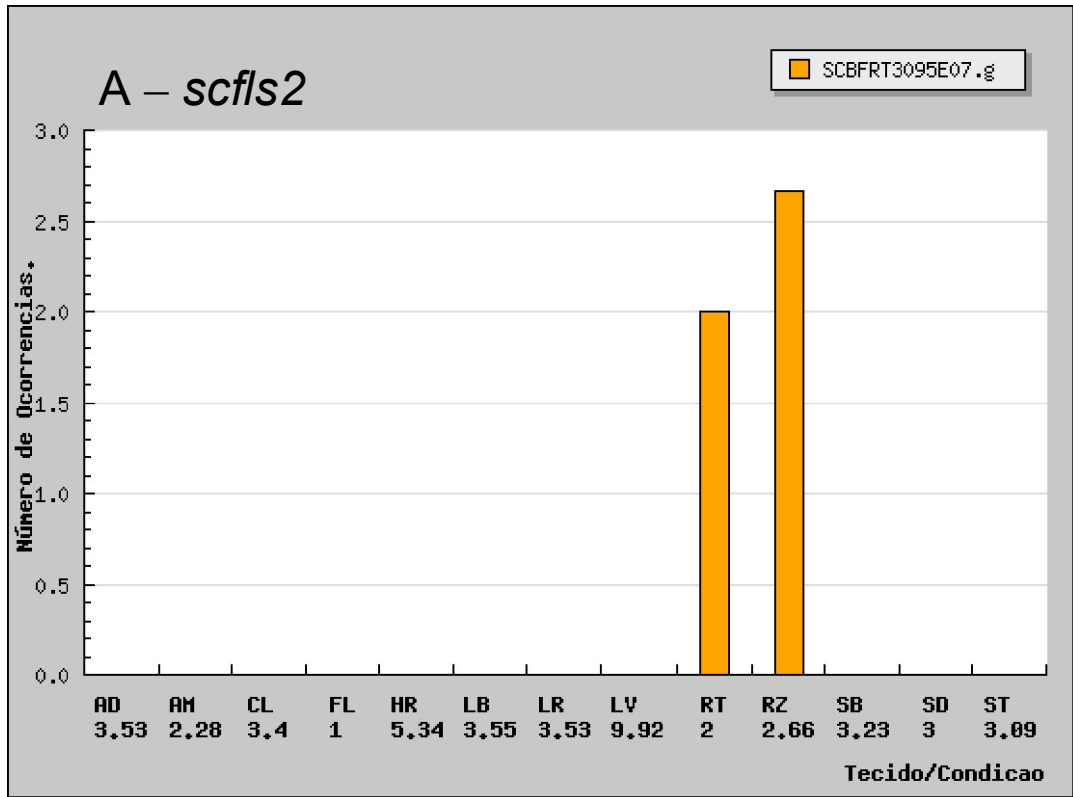
Análises de expressão “in silico” de proteínas LRR-RLKs através de “Northen” eletrônico, utilizando o banco de dados do SUCEST. Os genes selecionados para tais análises foram: *scfls2*, *scmar1*, *scserk-like 1* e *scserk-like 2* (Figura 2). Podemos observar que não há ocorrência de *scfls2* (Figura 2: A) nas bibliotecas de plantas inoculadas com *G. diazotrophicus* (AD) e *H. rubrisubalbicans* (HR). Contudo, podemos observar uma fraca ocorrência em raízes de plantas de cana-de-açúcar crescidas no solo (RT) e zona de transição entre raiz e folha (RZ).

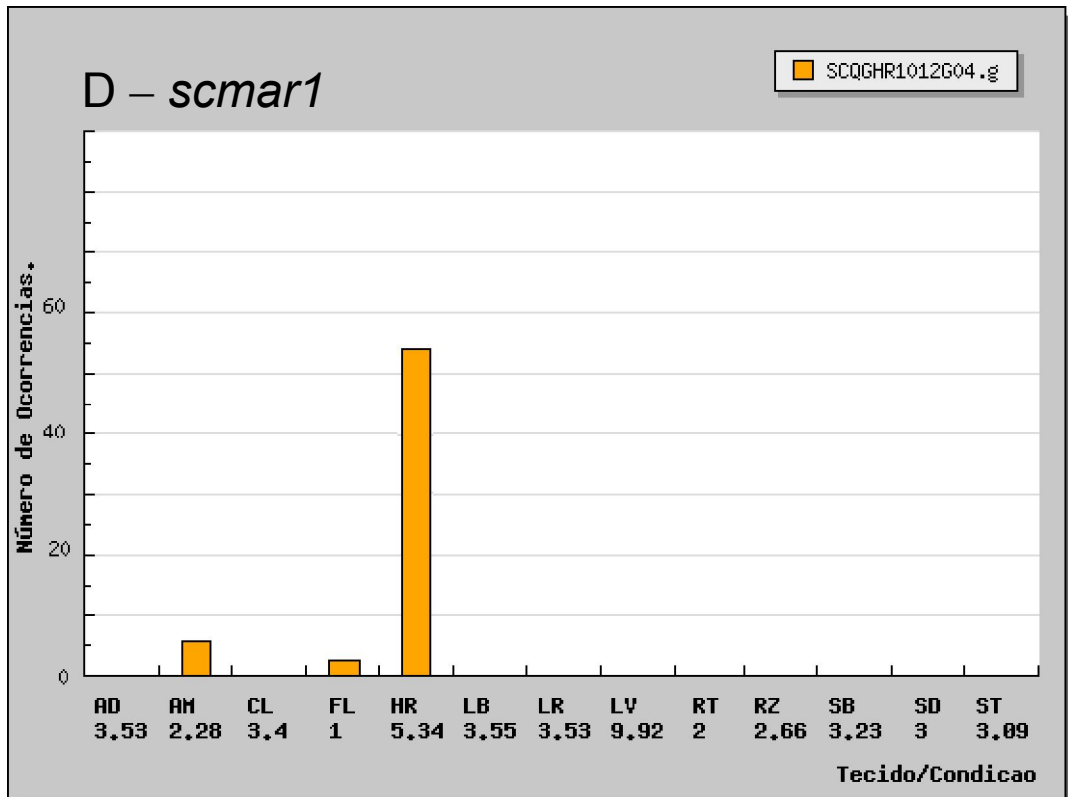
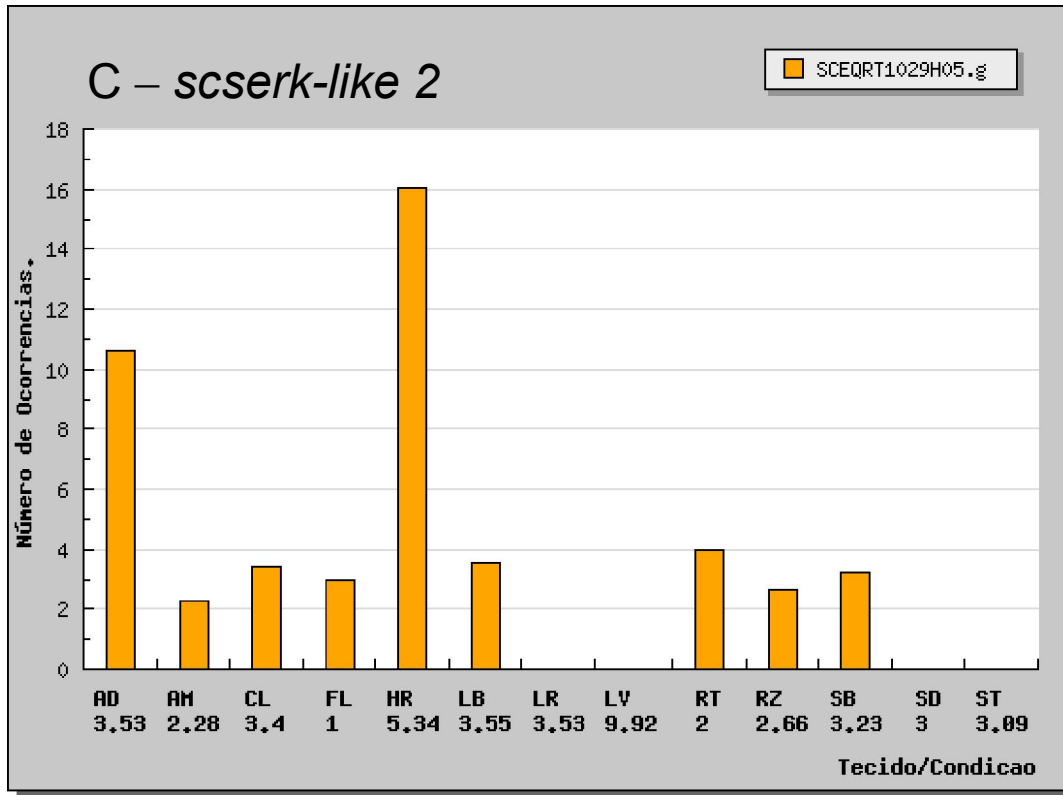
As seqüências em cana-de-açúcar similares aos membros da família SERK descritos em outras espécies, foram denominadas *scserk-like 1* e *scserk-like 2*. Podemos observar que transcritos *scserk-like 1* (Figura 2: B) são encontradas em plantas inoculadas com *G. diazotrophicus* (AD), meristema apical (AM), calos (CL), pedúnculo floral (FL), broto lateral de plantas (LB), cartucho foliar (LR), raiz de cana-de-açúcar (RT), zona de transição entre raiz e folha (RZ), cascas de plantas (SB), sementes (SD) e primeiro interno (ST). Porém, os tecidos onde ocorre maior indução são AM contendo 11 “Reads”, LR 18 “Reads”, RZ 10 “Reads” e SD “9 Reads”, nos demais tecidos a expressão é baixa. Na biblioteca de plantas inoculadas com *H. rubrisubalbicans* (HR) e folhas estioladas de plantas (LV) não há ocorrência de transcritos.

O gene *scerk-like 1* (Figura 2: B) não é induzido por HR e é fracamente expresso em AD. O gene *scserk-like 2* (Figura 2: C) é induzido nas bibliotecas de plantas inoculadas com *G. diazotrophicus* (AD) e *H. rubrisubalbicans* (HR), onde são encontrados 11 e 16 “Reads”, respectivamente. Nos demais tecidos analisados (AM, CL, FL, LB, RT, RZ e SB), são encontrados “Reads”, mas a expressão é baixa.

Figura 2: Análise da expressão de genes que codificam proteínas LRR-RLKs através de “Northen eletrônico”.

No eixo X as duas letras (Ex: AD, *G. diazotrophicus*) indica a sigla da biblioteca de cDNA obtidas de cana-de-açúcar nos diferentes tecidos ou condições. Os números abaixo das siglas indicam os fatores de correção para cada biblioteca. O eixo Y corresponde ao número de transcritos para os genes nas bibliotecas. Os nomes dos genes em cana-de-açúcar estão no topo de cada gráfico, sendo A – *scfls2*, B – *scserk-like 1*, C – *scserk-like 2*, D – *scmar1*. Foram considerados induzidos os genes cuja média em “Reads” em AD e/ou HR foram maiores em relação aos demais tecidos.





Interessantemente, os resultados do “Northen” eletrônico para *scmar1* (Figura 2: D) mostram uma grande frequência de transcritos ou indução para esse gene na biblioteca de plantas inoculadas com *H. rubrisubalbicans* (HR), onde são observados aproximadamente 55 “Reads”. Já meristema apical de folhas jovens e caules (AM) e pedúnculo floral (FL) apresentam 5 “Reads”, a expressão é baixa. Pode-se ressaltar que não há ocorrência de transcritos na biblioteca de plantas inoculadas com *G. diazotrophicus* (AD) e nas demais bibliotecas.

Conforme apresentado na Figura 3 foram realizados “Northen eletrônico” para outros 6 “Clusters” de cana-de-açúcar que codificam receptores transmembranas que se assemelham a cinases. A função desses ainda não é conhecida em outras espécies, exceto para *sclrr-rlk 2* e *sclrr-rlk 3*, cujas sequências em aminoácidos apresentam similaridades com o receptor Strubbelig 7 de *A. thaliana*. A proteína ScLRR-RLK3 apresenta também 86, 2% (Clustalw2/toosl/align) de similaridade com o receptor LTK 1 de *Z. mays* e ScLRR-RLK2 apresenta 40% de similaridade (Clustalw2/toosl/align).

Nossas análises mostram que os genes *sclrr-rlk 2* (Figura 3: B) e *sclrr-rlk 5* (Figura 3: E) são induzidos na biblioteca de plantas de cana-de-açúcar inoculadas com *G. diazotrophicus* (AD). Podemos observar que *sclrr-rlk 2* (Figura 3: B) apresenta 11 “Reads” em AD, 6 “Reads” em ST e 1 “Read” em FL, em outros tecidos o gene não é expresso.

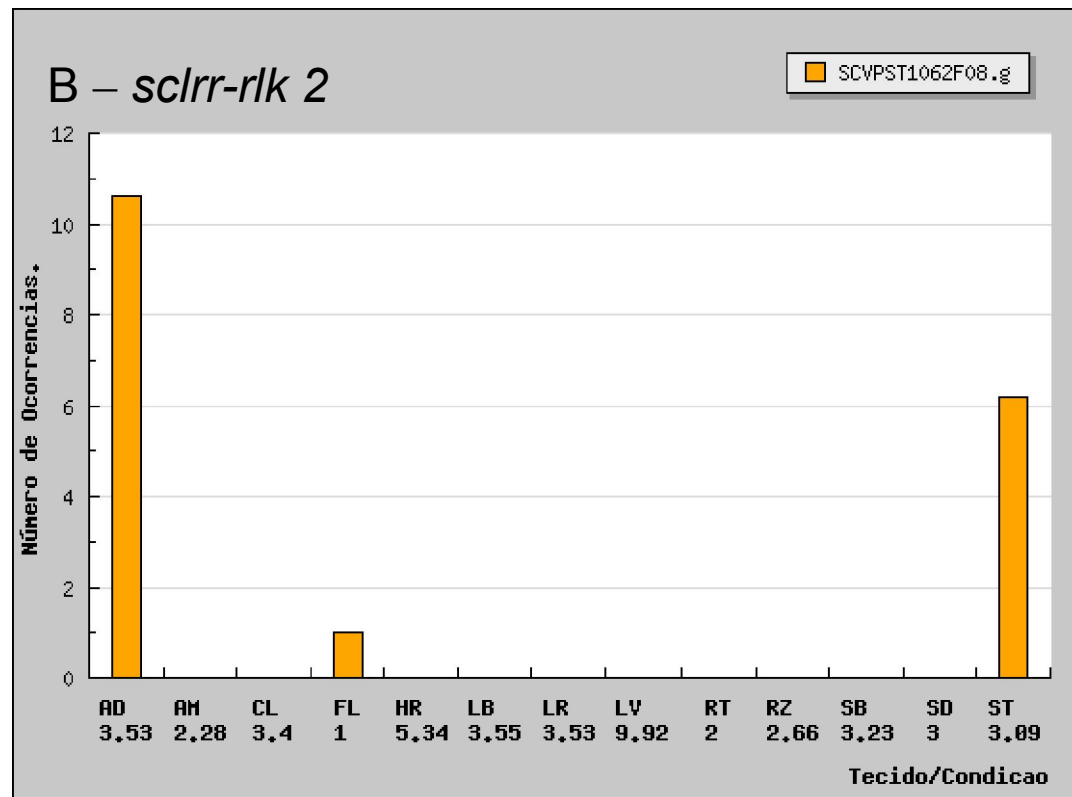
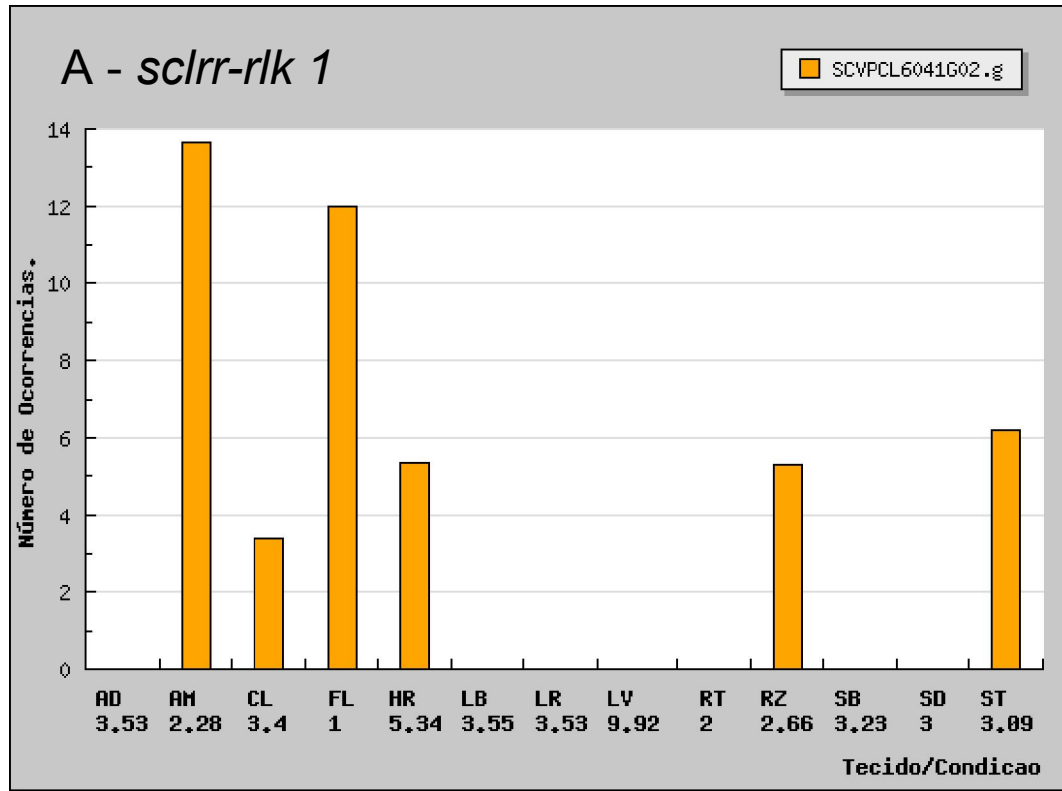
O gene *sclrr-rlk 3* (Figura 3: C) é expresso em vários tecidos/ou condições, com exceção de CL, LV, SD e ST. Os tecidos onde são encontrados maiores ocorrências são: LR aproximadamente 25 “Reads”, LB 21 “Reads”; SB aproximadamente 19 “Reads”, AM 18 “Reads”, AD 14 “Reads”, RZ 13 “Reads”. Foram também encontrados “Reads” em FL, HR e RT, porém, a expressão nestes tecidos é baixa.

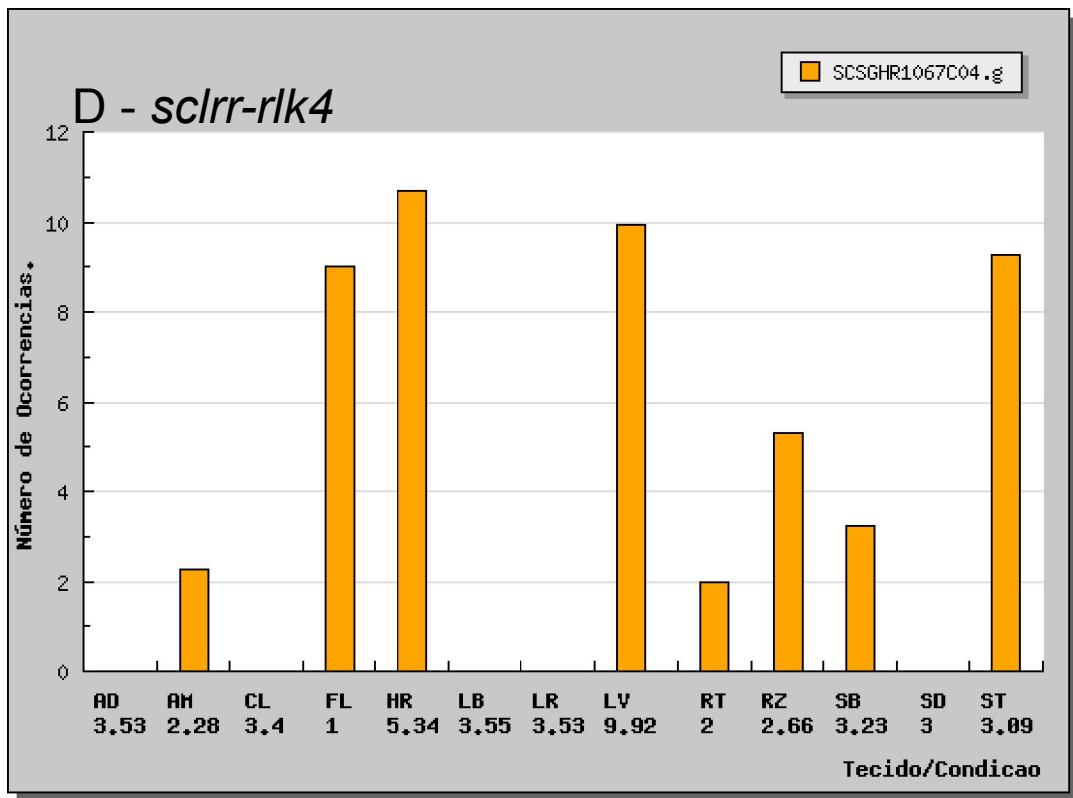
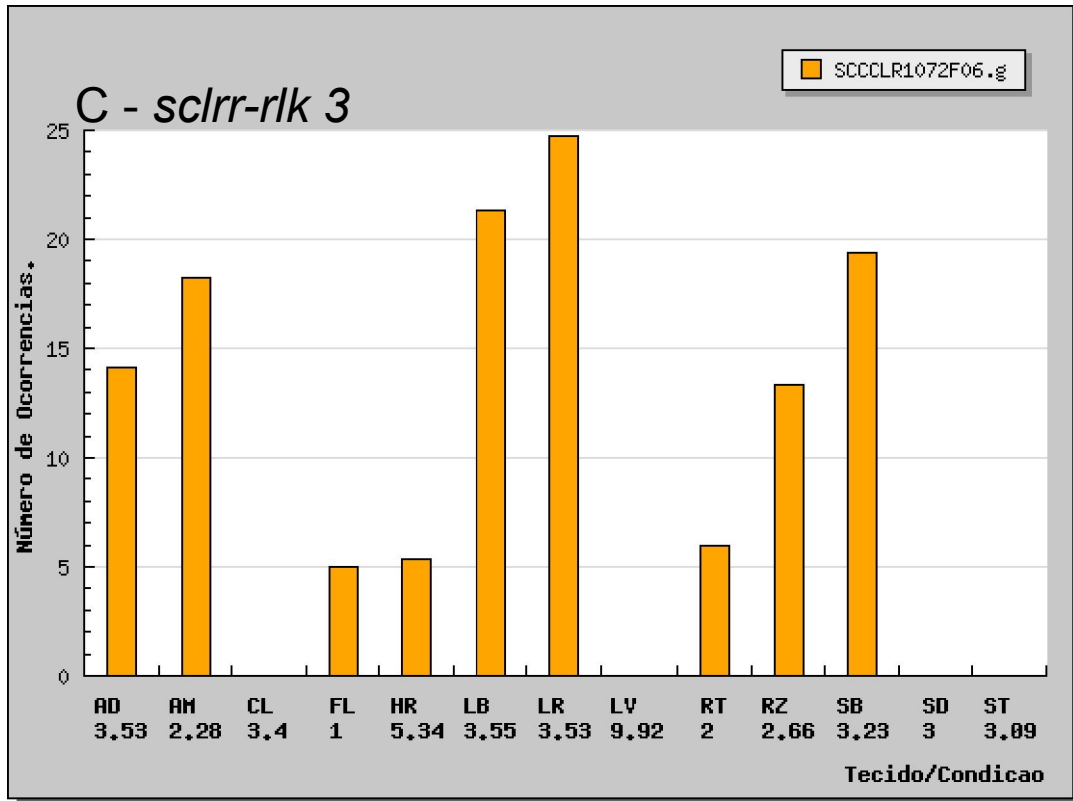
O gene *sclrr-rlk 4* (Figura 3: D) é expresso em vários tecidos/ou condições, com exceção de AD, CL, LB, LR e SD. Os tecidos/condições onde foram observadas as maiores ocorrências de “Reads” são HR, onde são encontrados aproximadamente 11 “Reads”, em LV 10 “Reads”, ST aproximadamente 9 “Reads”, FL 9 “Reads” e RZ aproximadamente 5 “Reads”. Foram também encontradas ocorrências de “Reads” em AM, RT e SB, porém a expressão nestes tecidos é baixa.

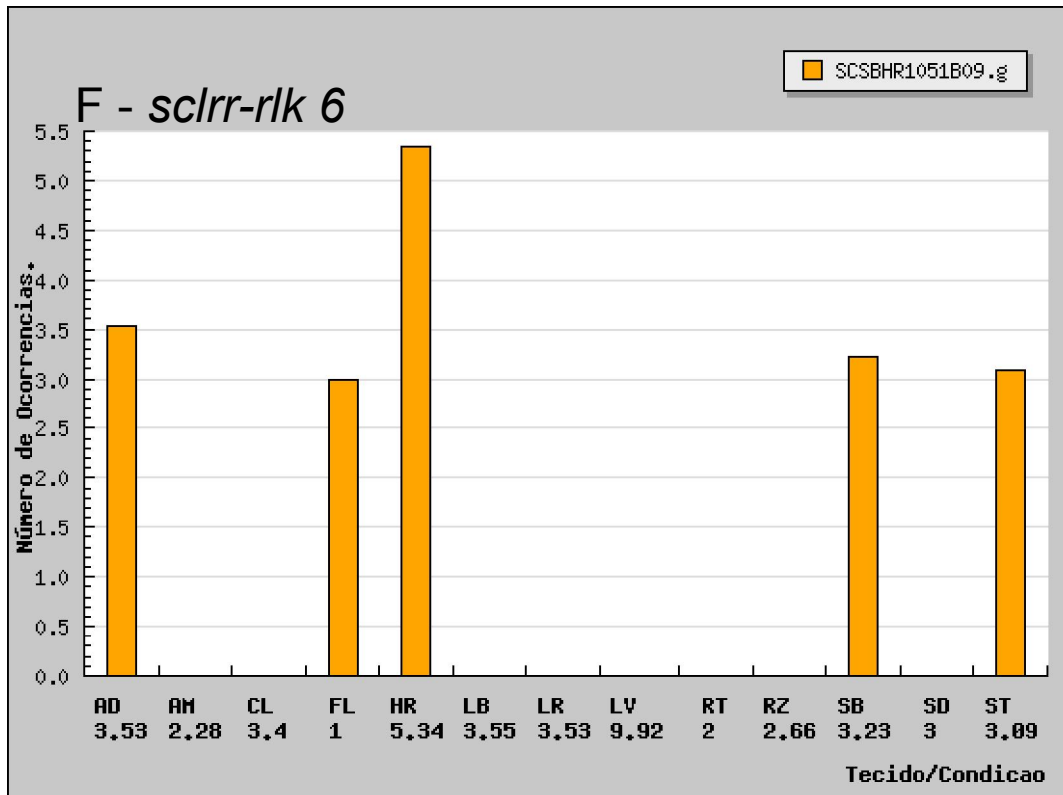
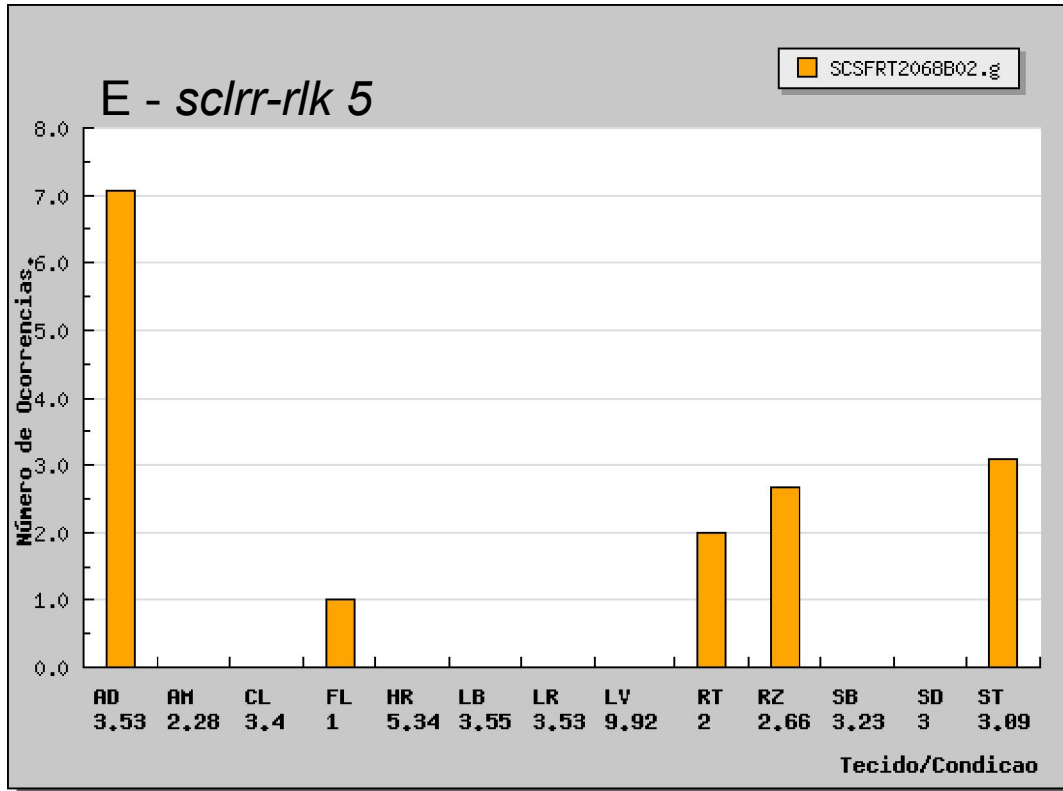
O gene *sclrr-rlk 5* (Figura 3: E) apresenta 7 “Reads” em AD, porém, em outros tecidos a expressão é baixa. O gene *sclrr-rlk6* (Figura 3: F) apresenta aproximadamente 5 “Reads” em HR e 3,5 em AD, nos demais tecidos (FL, SB e ST) a ocorrência de “Reads” é menor. Os genes *sclrr-rlk 2* (Figura 3: B) e *sclrr-rlk 5* (Figura 3: E) parecem ser induzidos por *G. diazotrophicus*, pois a expressão nestes tecidos é maior que a média observada em outros tecidos analisados. A sequência similar a proteína ScLRR-RLK 5 em *A.thaliana* (At4g23740) ainda não foi caracterizada. Nossas análises indicam que a proteína em cana-de-açúcar pode estar envolvida na percepção de *G. diazotrophicus*.

Figura 3: Análise da expressão de genes que codificam proteínas LRR-RLKs através de “Northen eletrônico”.

No eixo X as duas letras (Ex: AD, *G.diazotrophicus*) indica a sigla da biblioteca de cDNA obtidas de cana-de-açúcar nos diferentes tecidos ou condições. Os números abaixo das siglas indicam os fatores de correção para cada biblioteca. O eixo Y corresponde ao número de transcritos para os genes nas bibliotecas. Os nomes dos genes em cana-de-açúcar estão no topo de cada gráfico, sendo A – *sclrr-rlk 1*, B – *sclrr-rlk 2*, C – *sclrr-rlk 3*, D – *sclrr-rlk 4*, *sclrr-rlk 5*, *sclrr-rlk 6*. Foram considerados induzidos os genes cuja média em “Reads” em AD e/ou HR foram maiores em relação aos demais tecidos.







6.3- Comparação entre as sequências de aminoácidos das proteínas LRR-RLKs identificadas em cana-de-açúcar com sequências similares identificadas em outras espécies vegetais.

6.3.1- Comparação entre a proteína ScMAR1 (LRR-RLK) de cana-de-açúcar com sequências similares em outras espécies vegetais

A sequência ScMAR1 (LRR-RLK) foi alinhada no programa CLUSTALW2 (Larkin et al. 2007) com as sequências similares em outras espécies vegetais para verificar o nível de similaridade e conservação dos domínios presentes nas sequências. Posteriormente, estas sequências alinhadas foram analisadas com o programa SMART e observou-se que a proteína ScMAR1 e seus homólogos possuem os seguintes domínios: LRR (região rica em leucina), TM (região transmembrana) e TKC (tirosina cinase). Estes são domínios característicos de receptores LRR-RLKs. Contudo, variações no número de regiões ricas em leucina são observadas; onde as sequências de *Saccharum* sp, *O. sativa*, *T. aestivum*, e *Z. mays* apresentam um número menor de regiões ricas em leucina que as sequências de *S. bicolor* e *A. thaliana* (Quadro 11).

QUADRO 11: Comparação entre a sequência em aminoácidos do receptor ScMAR 1 de cana-de-açúcar e as sequências similares identificadas em outras espécies vegetais.

Espécie	Nome	Acesso			Similaridade cana X outras sps	
		NCBI/TIGRS/SUCEST	Aas	LRRs	domínio TKC	Proteína
<i>Saccharum</i> sp	ScMAR1	SCQGHR1012G04.g	368	1	100%	100%
<i>O.sativa</i>	OsMAR1	Os04g0576900	622	4	64.6%	56.5
<i>T.aestivum</i>	TaMAR1	TC412967	366	1	89.1%	71.3%
<i>Z.mays</i>	ZmMAR1	TC412967	1255	7	48.2%	23.5%
<i>S.bicolor</i>	SbMAR1	Sb06g026090.1	1165	12	63.4%	31.0%
<i>A.thaliana</i>	At4g36180	At4g36180	1135	14	52.8%	24.0%
<i>A.thaliana</i>	At1g75640	At1g75640	1140	17	57.9%	23.1%

O tamanho das sequências, em aminoácidos, das proteínas similares também difere (Quadro 11). O alinhamento no programa Clustalw2 revelou que as sequências em *T. aestivum*, *O. sativa* e *Saccharum sp* apresentam menor número de aminoácidos (Quadro 11). Isto pode sugerir que apenas partes destas sequências foram depositadas no banco de dados.

No Quadro 11 podemos observar o nível de similaridade entre a sequência de aminoácidos de ScMAR1 e as sequências similares identificadas em outras espécies vegetais; o nível de similaridade com ênfase ao domínio tirosina cinase também é apresentado. O domínio tirosina cinase de ScMAR1 apresenta alta similaridade (89%) com TaMAR (Quadro 11) e uma baixa similaridade com ZmMAR1 (48,2%). Um fato interessante é que tal domínio de ScMAR1 apresenta maior similaridade com as sequências At4g36180 e At1g75640 de *A. thaliana* que com a sequência de *Z. mays*, embora, *A. thaliana* seja uma espécie vegetal mais distante em termos evolutivos.

Análises no programa MEME revelaram três motivos (A, B, C) conservados no domínio cinase da sequência de aminoácidos de ScMAR1 (Figura 4). Os motivos A e B presentes em ScMAR1 possuem 100% de identidade com os motivos presentes na sequência de *S. bicolor*, *O. sativa* e *T. aestivum* (Figura 4). A sequência similar à ScMAR1 em *Z.mays* apresenta os motivos A e C, porém o motivo B está ausente (Figura 4). Em *A. thaliana* (At4g36180 e At1g75640) os motivos A, B, C estão presentes, porém, apresentam diferenças quanto à mudança de alguns aminoácidos. Em A, B e C (Figura 4) são mostrados os aminoácidos conservados nas sequências similares à ScMAR1 de outras espécies vegetais.

Análises filogenéticas da sequência de aminoácidos de ScMAR1 foram realizadas utilizando as seqüências dos domínios cinases das proteínas LRR-RLKs similares recuperadas pela ferramenta psi-blastp do NCBI, as proteínas SYMRK envolvidas na simbiose entre plantas e *Rhizobium* foram adicionadas na análise.

O resultado apresentado na Figura 5 mostra que ScMAR1 não possui relação de parentesco com os receptores já caracterizados, tais como CLAVATA1 (Clark et al., 1996) associado no controle da proliferação apical do meristema, ERECTA (Van Zanten et al., 2009) envolvido com iniciação, alongação dos órgãos e defesa, SERKs (Albrecht et al., 2005, Song et al., 2008) marcador de embriogênese somática e

papel na defesa, HAESA (Jin et al., 2000) envolvido com abscisão floral, Xa21 (Song et al., 1997) medeia resistência para *Xanthomonas orizae*, FLS2 (Gómez-Gómez et al., 2000) receptor da flagelina e receptores SYMRK (Gherbi et al. 2008) envolvidos com simbiose.

\

Figura 4: Motivos conservados no domínio cinase de MAR1 em diferentes espécies.

O motivo 1 pode ser observado em A, o motivo 2 em B e motivo 3 em C. Cada letra corresponde ao aminoácido, “a”, aromáticos (F, Y, W, H); “l”, alifáticos (I, V, L); “h”, hidrofóbicos (aromáticos, alifáticos, A, G, M, C, K, R, T); “+”, positivos (H, K, R); “-”, negativos (D, E); “p”, polar (positivo, negativo, Q, N, S, T, C). A cor de cada letra é referente às características físico químicas dos aminoácidos. Aminoácidos com a mesma cor possuem as mesmas características físico químicas.

A		
Nome	Início	
SbMAR1	118	GLAFLRQSGVVHGDVVKPQNILFDADFEPLSDF
ZmMAR1	118	GLAFLRQSGVVHGDVVKPQNILFDADFEPLSDF
OsMAR1	118	GLAFLRQSGVVHGDVVKPQNILFDADFEPLSDF
ScMAR1	118	GLAFLRQSGVVHGDVVKPQNILFDADFEPLSDF
TaMAR1	118	GLAFLRQSGVIHGDVVKPQNILFDADFEPLSDF
At4g36180	112	GLGFLRQSNMVIHGDIKPQNVLFDAADFEALISDF
At1g75640	112	GLSFLRSLSIHGDLPKQNVLFDAADFEALSEF
consenso		GL-F L H - - - - - HGD - KP QN- L FDA DFE - H - S - F

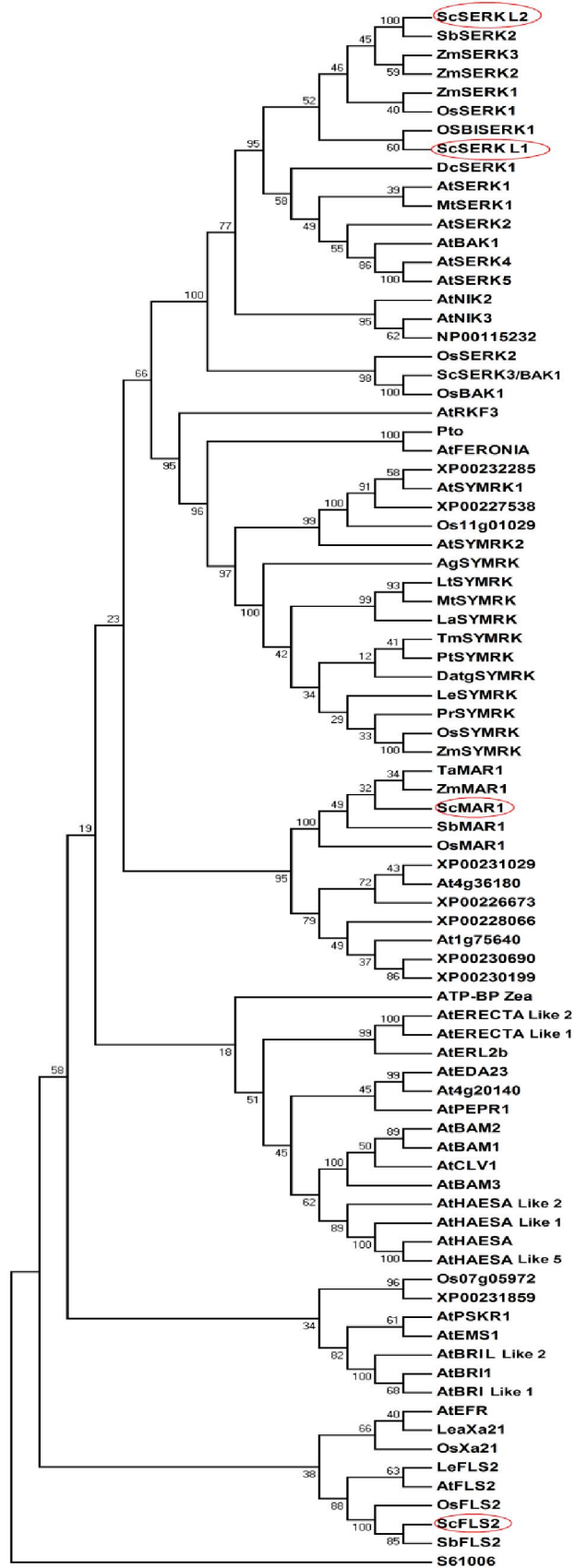
B		
Nome	Início	
SbMAR1	80	YDYPNGNLATLLQEASHQDGHILNWPMR
OsMAR1	80	YDYPNGNLATLLQEASHQDGHILNWPMR
TaMAR1	80	YDYPNGNLATLLQEASHQDGHILNWPMR
ScMAR1	80	YDYPNGNLATLLQEASHQDGHILNWPMR
At4g36180	74	YDYPNGNLATLLQEASHQDGHVNLNWPMR
At1g75640	74	YDYPNGNLSTLLQEASHQDGHVNLNWPMR
consenso		YD YMPNGNL - TLL QEASHQDGH - LNWPMR

C		
Nome	Início	
SbMAR1	0	FDEENVLSRGRHGLVFKACYS DGTVLAIQRLEPSTSSDGAVV
ScMAR1	0	FDEENVLSRGRHGLVFKACYS DGTVLAILRLEPSTSSDGAVV
OsMAR1	0	FDEENVLSRGRHGLVFKACYS DGTVLAILRLEPSTSSDGAVV
ZmMAR1	0	FDEENVLSRGRHGLVFKACYS DGTVLAILRLEPSTSSDGAVV
TaMAR1	0	FDEENVLSRGRHGLMFKACYS EGTVLAILRLEPSTSSDGAVV
At4g36180	0	FDEENVLSRTRYGLLFKANYNDGMVLSIRRLPENGSLLENL
At1g75640	0	FDEENVLSRGRYGLVFKATFRDGMVLSVRRRLMDGASITDAT
consenso		F DE ENVLS R- R - GL- FKA - - - - G- V L - - - RL - - - - -

e

Figura 5: Relações filogenéticas das proteínas ScFLS2, ScSERK-like 1 e ScSERK-like 2 e ScMAR1.

O domínio cinase das proteínas similares às sequências ScFLS2, ScSERK-like 1, ScSERK-like 2 e ScMAR1 recuperadas do banco de dados NCBI com “e-value” $\geq 1e^{-26}$, foram utilizadas para alinhamento no programa ClustalX usando HsERK1 S61006 da proteína *Saccharomyces cerevisiae* como grupo externo. Os seguintes parâmetros foram utilizados: “Pairwise alignment” – “Gap opening”, 35.0, “Gap extension”, 0.75; “Multiple alignment” – “Gap opening”, 15.0, “Gap extension”, 0.30. Os resultados de alinhamento no formato “Phylip” foram submetidos para o programa disponibilizado na rede “PHYML” para gerar uma árvore “maximum likelihood bootstrapped”. As proteínas ScFLS2, SbFLS2, ScMAR1, ScSERK-like1 e ScSERK-like 2 estão circuladas. Os “scores” de “bootstrap” correspondem aos números.



6.4 - Análise comparativa entre a sequência de aminoácidos de ScFLS2 e as sequências similares em outras espécies vegetais

Conforme, observado no Quadro 12, a sequência em aminoácidos da proteína ScFLS2 foi comparada com as sequências similares em *L. esculentum*, *S. bicolor*, *A. thaliana* e *O. sativa*. O domínio tirosina cinase de ScFLS2 apresentou alta similaridade com OsFLS2, SbFLS2, LeFLS2 e com o receptor FLS2 caracterizado em *A. thaliana* (QUADRO 12). Análises de alinhamento no programa CLUSTALW2 (Larkin et al. 2007) de ScFLS2 com outras espécies sugerem que apenas parte da seqüência dos transcritos ScFLS2 foi depositada no SUCEST. O cluster SCBFRT3095E07. g codifica uma sequência contendo 368 aminoácidos. Adicionalmente, o programa SMART revelou que essa região corresponde apenas o domínio tirosina cinase da proteína. O alinhamento do domínio tirosina cinase de ScFLS2 com as sequências similares em outras espécies vegetais revelou que essa região é altamente conservada (Figura 6). Análises no programa MEME revelaram três motivos (A, B, C) no domínio tirosina cinase de ScFLS2, OsFLS2, SbFLS2, LeFLS2 e AtFLS2 (Figura 6). Nas espécies monocotiledôneas (*Saccharum sp.*, *O. sativa* e *S. bicolor*) os motivos A, B e C são altamente conservados. Nas dicotiledôneas (*L. esculentum* e *A. thaliana*) esses motivos apresentam alguns aminoácidos diferentes. Análise filogenética de ScFLS2, SbFLS2 e LeFLS2 sugerem que essas proteínas são ortólogas a proteína AtFLS2 e OsFLS2 (Figura 5).

Quadro 12: Comparação entre a sequência em aminoácidos de ScFLS2 de cana-de-açúcar as sequências similares identificadas em *S.bicolor*, *A.thaliana* e *O.sativa*

Espécie	Nome	Acesso	Aas	Cana X outras sps		
				LRR	TKC	Proteína
<i>Saccharum sp.</i>	ScFLS2	SCBFRT3095E07.g	368	*	100%	100%
<i>O. sativa</i>	OsFLS2	NP001053894.1	1183	16	80,9%	21,1%
<i>S. bicolor</i>	SbFLS2	Sb06g028760.1	1188	17	91%	22,1%
<i>A. thaliana</i>	AtFLS2	At5g46330	1173	17	68,2%	18%
<i>L. esculentum</i>	LeFLS2	AC215381 11.1	1166	15	54,3%	18,1%

*regiões LRR ausentes

Figura 6. Alinhamento múltiplo dos domínios tirosina cinases dos homólogos de ScFLS2 em diferentes espécies.

As espécies apresentadas são Sb: *S. bicolor*, Sc: *Saccharum sp.*, Os: *O. sativa*, Le: *L. esculentum* e At: *A. thaliana*. Em A, B e C pode se observar os motivos conservados nas 5 espécies apresentadas. As sequências conservadas apresentam as mesmas cores. Cada letra corresponde ao aminoácido, “a”, aromáticos (F, Y, W, H); “l”, alifáticos (I, V, L); “h”, hidrofóbicos (aromáticos, alifáticos, A, G, M, C, K, R, T); “+”, positivos (H, K, R); “-”, negativos (D, E); “p”, polar (positivo, negativo, Q, N, S, T, C).

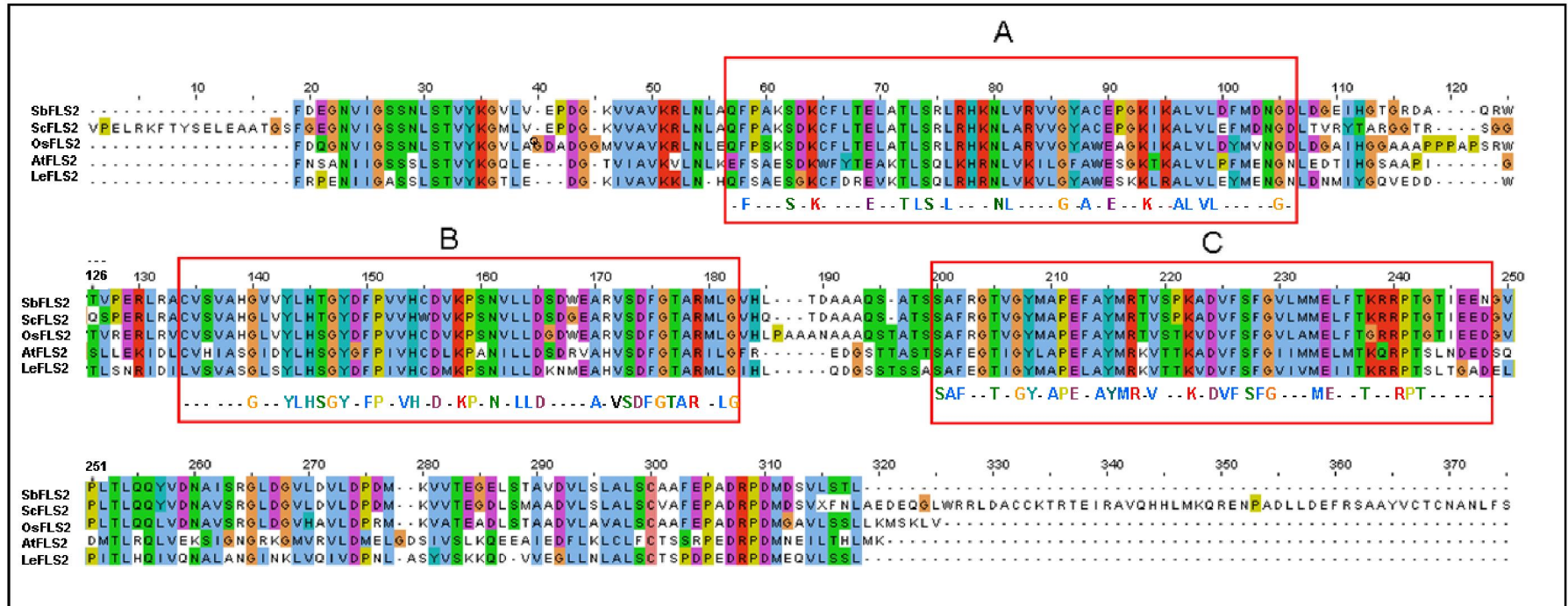


Figura 6: Alinhamento múltiplo do domínio tirosina cinase de seqüências homólogas à FLS2 (*Arabidopsis thaliana*) em diferentes espécies vegetais.

6.5 - Análise comparativa das sequências em aminoácidos de ScSERK-like 1 e ScSERK-like 2 com as proteínas SERKs encontradas em outras espécies vegetais.

As sequências ScSERK-like 1 e ScSERK-like 2 foram comparadas através de alinhamento no Clustalw2 (Larkin et al. 2007) com os membros da família SERK em plantas. Interessantemente, a proteína ScSERK-like 1 apresenta alta similaridade com os membros da família SERK apresentados no Quadro 13, b. Porém, com seu parólogo ScSERK like 2 (Quadro 13, (■)) a similaridade é baixa. A proteína ScSERK-like 1 e seu domínio tirosina cinase compartilham alta similaridade com os cinco membros (▲) da família SERK em *A. thaliana* (SERK1, SERK 2, SERK3 e SERK4 Quadro 13, B e b (●)). Já a proteína ScSERK-like 2 e seu domínio cinase apresentam uma menor similaridade com a família SERK (QUADRO 13, A e a). Adicionalmente, foi observado que a proteína ScSERK-like 2 apresenta uma menor região rica em leucina e um número menor de aminoácidos (Quadro 13, A).

Análise no programa SMART revelou que domínios catalíticos para ScSERK-like 2 parecem estar ausentes na região tirosina cinase. As análises filogenéticas (Figura 7) indicam que ScSERK-like 1 e ScSERK-like 2 apresentaram relações de parentesco com os demais membros da família SERK. A árvore filogenética revelou que ScSERK-like 1 é ortóloga à proteína OsBISERK1 e ScSERK-like 2 está mais proximamente relacionada à SbSERK2 e ZmSERK2 (Figura 7). Porém, ScSERK-like 2 e SbSERK2 agruparam-se em braços distintos dos demais membros família SERK (Figura 7). Os dados indicam que estas proteínas são novos membros da família SERK, cuja função ainda não foi elucidada. Contudo, nossos dados de “Northern” eletrônico apontam uma possível função do receptor ScSERK-like 2 (Figura 2: C) na associação entre cana-de-açúcar e bactérias endofíticas.

Quadro 13: Comparação entre as sequências ScSERK-like 1 e ScSERK-like 2 identificadas em cana-de-açúcar com membros da família SERK presentes em diferentes espécies vegetais.

Espécie	Nome	Acesso	NCBI/MIPs/SUCEST	Aas	LRRs	Similaridade cana X outras sps			
						Domínio TKC		Proteína	
						ScSERK-L2	ScSERK-L1	ScSERK-L2	SERK-L1
						A	B	a	b
<i>Saccharum sp.</i>	ScSERK-like 2	SCEQRT1029H05.g	480	2	100%	56.4%	100%	38.9% (■)	
<i>Saccharum sp.</i>	ScSERK-like 1	SCMCLR1053B12.g	685	4	56.4%	100%	38.9%	100%	
<i>Saccharum sp.</i>	ScBAK1	ABX83681.1	619	3	35.4%	51.9%	26.1%	72.2%	
<i>O.sativa</i>	OsBAK1	ABA94685	608	4	39.0%	57.0%	31.0%	75.5%	
<i>O.sativa</i>	OsSERK1	AAU88198	628	5	58.8%	79.5%	38.9%	96%	
<i>O.sativa</i>	OsBiSERK1	BAD37288	616	5	44.1%	63.6%	32.4	79%	
<i>O.sativa</i>	OsSERK2	BAD22198	607	4	39.0%	59.5%	32.7%	82.9%	
<i>S.bicolor</i>	SbSERK2	Sb04g023810.1	626	2	61.2%	81.2%	38.5%	96.4%	
<i>A.thaliana</i>	AtSERK1 (▲)	AAK82463	625	5	54.6%	79.2% (●)	38.2%	96% (●)	
<i>A.thaliana</i>	AtSERK2 (▲)	AAK68073	628	5	54.3%	80% (●)	37.8%	95.3% (●)	
<i>A.thaliana</i>	AtSERK3 (▲)	AAK68074	615	3	52.9%	74.8% (●)	37.8%	95.6% (●)	
<i>A.thaliana</i>	AtSERK4 (▲)	NP_178999	620	5	49.9%	71.2% (●)	36.7%	92.7% (●)	
<i>A.thaliana</i>	AtSERK5 (▲)	NP_179000	601	3	45.9%	69.1% (●)	34.5%	90.5% (●)	
<i>M.trucatula</i>	MtSERK1	AAN64293	627	2	54.2%	79.9%	37.8%	95.6%	
<i>Z.mays</i>	ZmSERK1	CAC37638	621	5	58.2%	79.8%	36.7%	92.7%	
<i>Z.mays</i>	ZmSERK2	CAC37639	626	2	61.4%	81.7%	39.3%	96.0%	
<i>Z.mays</i>	ZmSERK3	CAC37642	541	5	62.8%	71.1%	41%	89.8%	
<i>D.carota</i>	DcSERK1	AAB61708	553	*	45.4%	72.2%	37.2%	94.2%	

*regiões LRR ausentes

Aas - número de aminoácidos da proteína, LRR - número de regiões ricas em leucina, (▲) – membros da família SERK em *A. thaliana*, (●) – similaridade entre proteína ScSERK- L1 versus família SERK em *A. thaliana*.

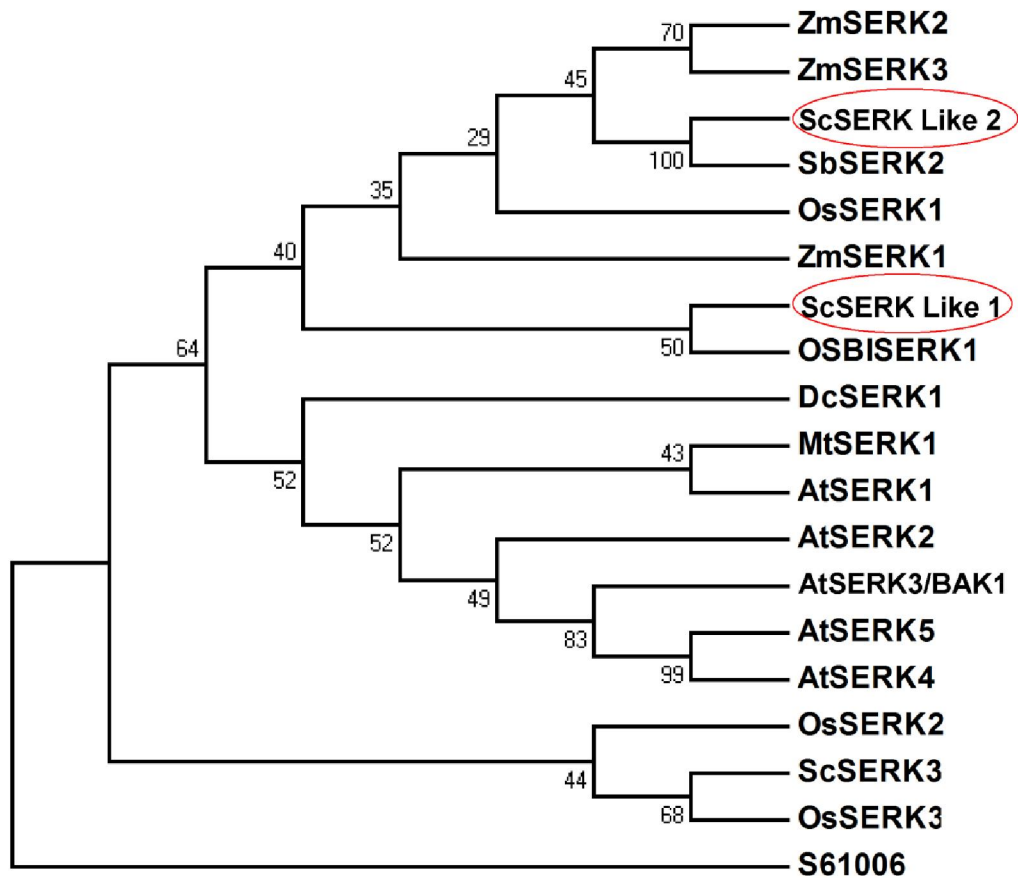


Figura 7: Relações filogenéticas da família SERK, incluindo ScSERK-like 1 e ScSERK-like 2.

A sequência S61006 da proteína *Saccharomyces cerevisiae* foi utilizada como grupo externo. O domínio cinase de proteínas da família SERKs foram alinhadas no programa ClustalW. Os resultados do alinhamento no formato "Phylip" foram submetidos para o programa disponibilizado na rede "PHYML" para gerar uma árvore "maximum likelihood bootstrapped". As proteínas ScSERK-like 1 e ScSERK-like 2 estão circuladas. Os "scores" de "bootstrap" correspondem aos números.

6.6 – “Clusters” que codificam proteínas com domínios CC-NB-LRR induzidos por *Gluconacetobacter diazotrophicus* e/ou *Herbaspirillum rubrisubalbicans*.

Um grande número de genes de resistência em plantas é representado por proteínas contendo um sítio “nucleotide binding” (NB) e regiões ricas em leucina. As proteínas NB-LRR podem apresentar na região N-terminal os domínios coiled-coil (CC) ou Toll-interleukin-1 receptor (TIR). Foram realizadas buscas de seqüência em cana-de-açúcar que codificam proteínas com domínios CC-NB-LRR e TIR-NB-LRR induzidas nas bibliotecas AD e/ou HR. Os resultados são apresentados no Quadro 14, onde 16 clusters (▲, ■) codificam proteínas que apresentam similaridades com proteínas que possuem os domínios NB-LRR ou NB-ARC.

Porém, para os clusters representados por (■), os domínios NB-LRR ou NB-ARC não foram identificados no programa SMART. Possivelmente apenas parte das seqüências foram depositadas no banco de dados. Assim, estas não foram classificadas como NB-LRRs. Para tanto, os genes que codificam tais proteínas devem ser seqüenciados para uma re-análise dos domínios estruturais destas proteínas. Cabe ressaltar, que estas proteínas apresentam transcritos na biblioteca de plantas inoculadas com *G. diazotrophicus* e/ou *H. rubrisubalbicans*. Como exemplo, o cluster SCVPRZ3028C05. g que apresenta 4 “Reads” na biblioteca de plantas inoculadas com *H. rubrisubalbicans*, correspondendo a 50% do total de “Reads”. Outros 14 clusters, Quadro 14, (●), codificam proteínas com outros domínios estruturais e não são classificadas como proteínas NB-LRRs. Neste trabalho não foram encontradas seqüências com domínios TIR induzidas por *G. diazotrophicus* e *H. rubrisubalbicans* no SUCEST.

Para a classe CC-NB-LRR 6 clusters identificados, Quadro 14, (▲), codificam proteínas que possuem os domínios NB-LRRs ou NB-ARC confirmados no programa SMART e estas apresentam similaridade com proteínas (ADR1, RPM1, RPP8 e RPP13) já caracterizadas. Tais proteínas foram utilizadas para as demais análises. O nível de similaridade (S) obtido no clustalw2 entre as seqüências NB-LRRs de cana-de-açúcar e a seqüência similar em *A. thaliana* é apresentado no Quadro 15.

QUADRO 14: Sequências NBS-LRR-ARC encontradas em cana-de-açúcar.

TAIR	SUCEST	Domínios identificados	Similaridade SUCEST/NCBI	I	Evalue	Total	HR	AD	soma	%HR	%AD	%-total	Fator
At1g33560	SCSGAD1006H08.g	NB-ARC (▲)	ScADR1	56%	8,00E-179	15	0	8	8	0	53,3	53,3	427
At3g07040	SCSFSB1103A12.g	NB-ARC (▲)	ScRPM1	30%	5,00E-76	6	0	2	2	0	33,3	33,3	67
At3g07040	SCJFRZ2014A03.g	NB-ARC (▲)	ScNB-LRR1/RPM1	42%	8,00E-105	4	0	1	1	0	25	25	25
At3g46530	SCRUFL1024D04.g	NB-ARC (▲)	ScNB-LRR2/RPP13	36%	1,00E-40	3	0	1	1	0	33,3	33,3	33
At5g43470	SCRLAD1098H03.g	NB-ARC (▲)	ScNB-LRR3/RPP8	31%	1,00E-09	1	0	1	1	0	100	100	100
At3g07040	SCVPHR1091E12.g	NB-ARC (▲)	ScNB-LRR4/RPM1	32%	8,00E-23	1	1	0	1	100	0	100	100
At1g50180	SCRFAD1119E10.g	Não identificado (■)	Disease resistance protein	28%	2,00E-12	1	0	1	1	0	100	100	100
At5g63020	SCBFAD1089A09.g	Não identificado (■)	Disease resistance protein	38%	3,00E-16	1	0	1	1	0	100	100	100
At3g50950	SCSBAD1126D07.g	Não identificado (■)	Disease resistance protein	31%	1,00E-08	2	0	1	1	0	50	50	50
At1g63350	SCJFRZ3C03F12.g	Não identificado (■)	Disease resistance protein	27%	7,00E-09	6	0	1	1	0	16,7	16,7	17
At1g63350	SCCCAM1004E10.g	Não identificado (■)	Disease resistance protein	35%	3,00E-08	2	0	1	1	0	50	50	50
At5g44510	SCVPRZ3028C05.g	Não identificado (■)	Disease resistance protein-Rp3	34%	2,00E-15	8	4	0	4	50	0	50	200
At2g34930	SCRUAD1063E09.g	Não identificado (■)	Disease resistance protein	42%	2,00E-16	1	0	1	1	0	100	100	100
At3g14470	SCJFRZ2028H04.g	Não identificado (■)	Disease resistance protein	24%	9,00E-06	4	2	0	2	50	0	50	100
At1g61190	SCJLRT1016E09.g	Não identificado (■)	Disease resistance protein	34%	3,00E-14	5	0	2	2	0	40	40	80
At3g14470	SCEQAD1016E12.g	Não identificado (■)	Disease resistance protein	31%	2,00E-10	2	0	1	1	0	50	50	50
At1g15740	SCJFRT1009F02.g	LRR (●)	Leucine-rich repeat family protein	64%	0,0	20	0	0	1	0	0	0	0
At4g33210	SCCCCL4007A08.g	LRR e FBOX (●)	F-box family protein (FBL15)	51%	0,0	29	0	2	2	0	6,9	6,9	14
At2g26330	SCCCLR1001A03.g	LRR e STKC (●)	ERECTA	67%	0,0	46	0	2	2	0	4,35	4,3	9
At4g35470	SCUTAD1033C08.g	LRR (●)	Leucine-rich repeat family protein	56%	1,00E-126	8	0	2	2	0	25	25	50
At5g21090	SCEPLR1030B03.g	LRR (●)	Leucine-rich repeat protein	63%	4,00E-75	31	2	0	2	6,5	0	6,5	13
At3g59140	SCEPLR1008E12.g	ABC membrana e AAA (●)	ATPase	71%	0,0	8	1	0	1	13	0	12,5	13
At1g67940	SCQSLB1051C05.g	AAA (●)	ATNAP3/transporter	70%	1,00E-96	9	1	0	1	11	0	11	11
At3g61140	SCJLLR2013G12.g	COP9 e PINT (●)	Fusca 6	68%	6,00E-180	20	2	4	6	10	20	30	180
At5g47040	SCMCST1049D08.g	AAA e LonC (●)	Lon Protease 2	81%	0,0	15	1	1	2	6,7	6,67	13	27
At3g13772	SCEZRZ1015B08.g	Cytochrom B _C eTM9SF (●)	Endomembrane protein 70	78%	0,0	37	0	3	3	0	8,11	8	24
At5g64220	SCSFRZ2071A12.g	CG-1 domains (●)	Calmodulin-binding protein	69%	1,00E-57	3	1	0	1	33	0	33	33
At5g65700	SCQGST1031F08.g	Não identificado (●)	Barely any Meristem 1	48%	6,00E-38	5	0	2	2	0	40	40	80
At4g08850	SCCCRZ1001F12.g	Não identificado (●)	Kinase	39%	1,00E-28	2	0	1	1	0	50	50	50
At1g23870	SCCCCL2001H04.b	Trehalose e Hydrolase 3 (●)	ATTPS9/transferase	61%	4,00E-127	9	0	2	2	0	22,2	22,2	44

TAIR – número de acesso no banco de dados TAIR, SUCEST- número de acesso no banco de dados SUCEST, I - Nível de Identidade obtido no Psi-blastp/NCBI, evalue - probabilidade de ser a sequência esperada, Total-número de “Reads” totais correspondente ao cluster analisado, HR- número de “Reads” na biblioteca HR, AD- número de “Reads” na biblioteca AD, soma- soma de “Reads” da biblioteca AD com HR, %HR- porcentagem de “Reads” na biblioteca HR,%AD- porcentagem de “Reads” na biblioteca AD, % total- soma das porcentagens de AD com HR, Fator= soma total de “Reads” (soma de AD com HR) multiplicado pela porcentagem total (soma de %AD com %HR), (▲) – proteínas de cana-de-açúcar similares à proteínas NB-LRR caracterizadas na literatura, (■) - proteínas de cana-de-açúcar que apresentam similaridades com proteínas NB-LRRs, porém, com domínios NB-ARC não preditos no programa SMART, (●) – proteínas não classificadas como NB-LRR.

Quadro 15: Nível de similaridade obtido entre as sequências NB-LRRs de cana-de-açúcar e a sequências similares em *A. thaliana*

TAIR	SUCEST	Similaridade SUCEST/NCBI	S	ScNB-LRRs/Aas	AtNB-LRRs/Aas
At1g33560	SCSGAD1006H08.g	ScADR1	49%	569 (■)	787 (●)
At3g07040	SCSFSB1103A12.g	ScRPM1	36,3%	610 (■)	926 (●)
At3g07040	SCJFRZ2014A03.g	ScNB-LRR1/RPM1	16,9%	430 (■)	926 (●)
At3g46530	SCRUFL1024D04.g	ScNB-LRR2/RPP13	11%	208 (■)	835 (●)
At5g43470	SCRLAD1098H03.g	ScNB-LRR3/RPP8	10,1%	208 (■)	908 (●)
At3g07040	SCVPHR1091E12.g	ScNB-LRR4/RPM1	12,6%	190 (■)	926 (●)

TAIR – número de acesso no banco de dados TAIR (“The Arabidopsis Information Resource”), SUCEST- número de acesso no banco de dados SUCEST, I - Nível de Identidade obtido no blastp/NCBI, S- Nível de similaridade (%) entre a sequência de cana-de-açúcar e a sequência de *A.thalian*, ScNB-LRR/Aas- Número de aminoácidos presentes na sequência em cana-de-açúcar, AtNB-LRR/Aas- números de aminoácidos presentes na sequência em *A.thaliana*. Alinhamento no ClustalW 2 (Larkin et al. 2007)

Podemos observar que o nível (%) similaridade entre as sequências de cana-de-açúcar e as de *A. thaliana* é muito baixa, exceto para ScRPM1 e ScADR1. O tamanho em aminoácidos das sequências ScNB-LRR1, ScNB-LRR2, ScNB-LRR3 e ScNB-LRR4 (Quadro 15, (▲)) é bem menor em comparação com a sequência similar em *A. thaliana*, (●), Quadro 15. Isto indica que apenas parte das sequências foram depositadas no banco de dados SUCEST, embora o domínio NB-ARC ou NB-LRR tenha sido identificado no programa SMART.

Assim, as proteínas foram nomeadas de ScADR1 (“Sugarcane activated disease resistance 1”), ScRPM1, ScNB-LRR1/RPM1, ScNB-LRR2/RPP13, ScNB-LRR3/RPP8 e ScNB-LRR4/RPM1. Em plantas inoculadas com *G. diazotrophicus* (Quadro 14) ocorre um grande número de “Reads” para ScADR1 (Quadro 14), correspondendo a 53,3% do total.

Em cana-de-açúcar a proteína ScADR1 ainda não foi caracterizada. Nossos dados sugerem que a proteína participe do processo de associação entre cana-de-açúcar e *G. diazotrophicus*.

6.7 - “Northen Eletrônico” das proteínas CC-NB-LRR identificadas no SUCEST

Análises de expressão “in silico” através de “Northen eletrônico” foram realizadas em cana-de-açúcar (SUCEST) para as sequências *scadr1*, *scrpm1* e *scnb-lrr2* em 13 bibliotecas de cana-de-açúcar. Podemos observar (Figura 8) que os transcritos para a proteína ScADR1 são altamente induzidos por *G. diazotrophicus* (AD), sendo também transcritos em Flor (FL), broto lateral (LB), raízes crescidas no solo (RT) e cascas de plantas de cana-de-açúcar crescidas em casa de vegetação. Porém nestes tecidos a expressão é muito baixa. Interessantemente, o gene *scadr1* não é induzido por *H. rubrisubalbicans*. O gene *scrpm1* mostrou expressão na biblioteca de plantas inoculadas com *G. diazotrophicus* (AD), zona de transição entre raízes e folhas (RZ), em cascas (SB) e em sementes (SD). Os transcritos de *scnb-lrr2/rpp13* foram induzidos na biblioteca de cDNA de plantas de cana-de-açúcar inoculadas com *G. diazotrophicus* (AD) e em folhas (FL). Porém cabe ressaltar que os números de transcritos para *scrpm1* e *scnb-lrr2* são relativamente baixos.

Em nossas análises foram encontrados 6 “Clusters” (Quadro 14) que codificam proteínas com os domínios CC-NB-LRR, dentre estes, 5 (SCSGAD1006H08.g, SCSFSB1103A12.g, SCJFRZ2014A03.g, SCRUF11024D04.g e SCRLAD1098H03.g) apresentam transcritos na biblioteca de plantas inoculadas por *G. diazotrophicus*. O gene *scnb-lrr4* apresenta transcrito apenas na biblioteca de plantas inoculadas com *H. rubrisubalbicans*, porém a ocorrência é baixa. Portanto, dentre os “Clusters” identificados que codificam proteínas NB-LRRs, o gene que codifica ScADR1 é fortemente induzido por *G. diazotrophicus*. Para os demais genes a expressão é baixa em AD ou HR.

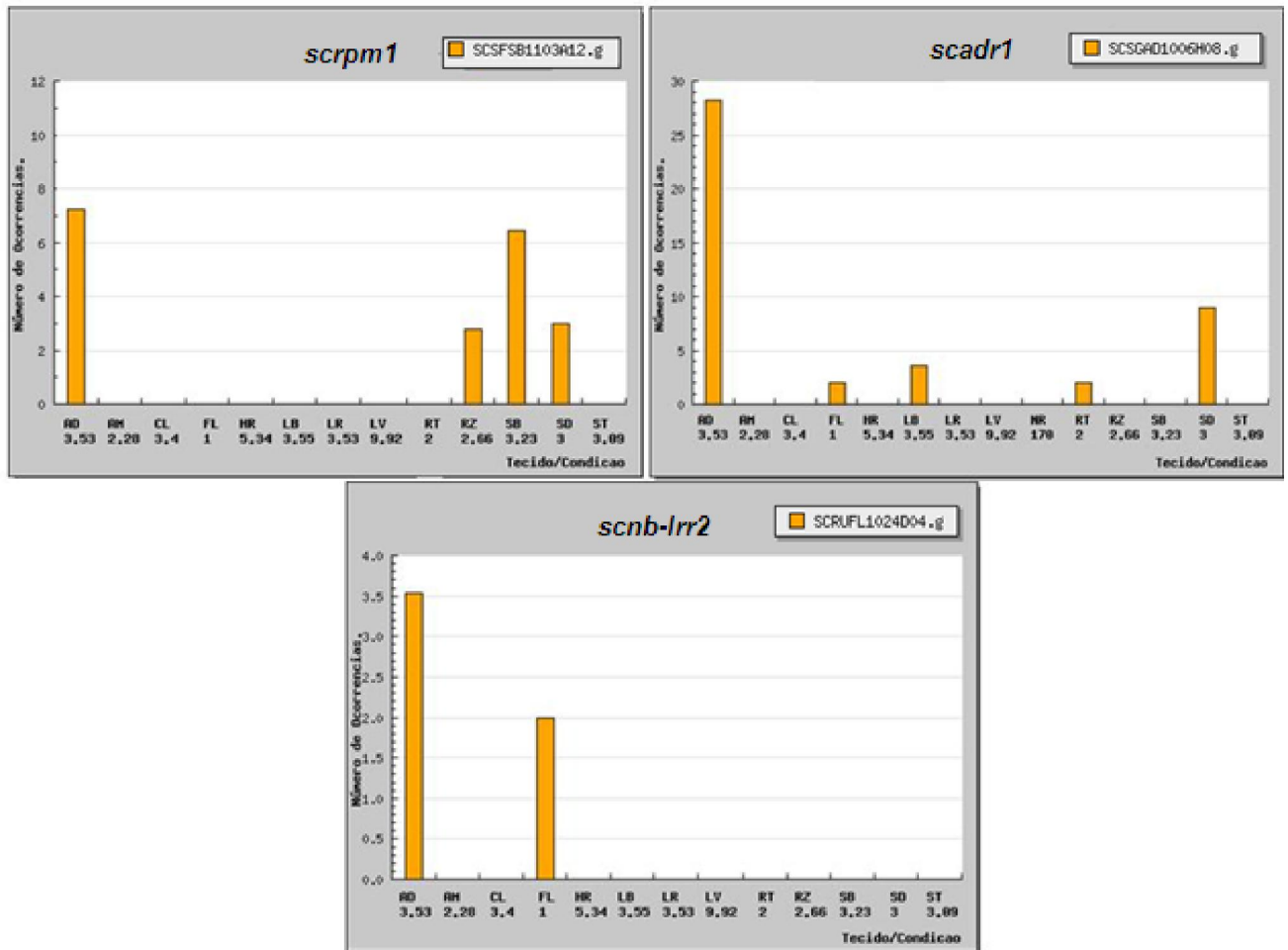


Figura 8: Análise da expressão de genes que codificam proteínas CC-NB-LRR em cana-de-açúcar através de “Northen eletrônico”.

No eixo X as duas letras (Ex: AD, *G.diazotrophicus*) indica a sigla da biblioteca de cDNA obtida de cana-de-açúcar nos diferentes tecidos ou condições. Os números abaixo das siglas indicam os fatores de correção para cada biblioteca. O eixo Y corresponde ao número de transcritos para o gene nas bibliotecas. Os nomes dos genes em cana-de-açúcar estão no topo de cada gráfico. Foram considerados induzidos os genes cuja média em “Reads” em AD e/ou HR foram maiores em relação aos demais tecidos.

6.8 - Análise comparativa entre a sequência de aminácidos de ScADR1 com sequências similares em outras espécies vegetais

O resultado do alinhamento (Clustalw2) apresentado no QUADRO 16 indica que ScADR1 apresenta alta similaridade com as proteínas de *A. thaliana* At1g33560(ADR1), At4g33300 (ADR1-L1), At5g04720(ADR1-L2) e At5g47280, (ADR1-L3) e com os homólogos em diversas espécies vegetais. A proteína ScADR1 apresenta maior similaridade com o membro ADR1-L3 em *A.thaliana* (Quadro 16).

Quadro 16: Nível de similaridade obtido através de alinhamento no programa ClustalW 2 entre as sequências ScADR1 e as sequências similares identificadas em outras espécies vegetais.

Espécie	Acesso		Similaridade cana X outras sps
	SUCEST/NCBI/TIGRS/MIPs	Aas	
<i>Saccharum sp.</i>	ScADR1	569	100%
<i>A.thaliana</i>	ADR1	787	49.0
<i>A.thaliana</i>	ADR1-L1	816	49.3%
<i>A.thaliana</i>	ADR1-L2	811	50.3%
<i>A.thaliana</i>	ADR1-L3	623	63.7%
<i>V.vinifera</i>	TC100421 (♦)	611	71.1%
<i>Coffea sp.</i>	TC2260 (■)	347	48.2%
<i>Gossypium sp.</i>	TC126938 (■)	453	43.0%
<i>S.tuberosum</i>	TC180809 (■)	232	33.2%
<i>G.max</i>	TC260722 (■)	312	41.5%
<i>T.aestivum</i>	TC307454 (■)	349	52.4%
<i>P.trichocarpa</i>	TC122773 (■)	671	63.9%
<i>M.truncatula</i>	TC117547 (■)	824	51.4%
<i>L.japonicus</i>	TC33560 (■)	313	41.2%
<i>O.sativa</i>	Os12g0586000 (▲)	601	91.8%
<i>Z.mays</i>	NP001140526.1	531	44.8%
<i>S.bicolor</i>	Sb08g019690.1 (●)	872	65.1%

A proteína ScADR1 apresenta alta similaridade com a proteína de *V. vinifera* (♦) (71.1%) (Quadro 16). Pode-se observar, no mesmo QUADRO, que em monocodilêneas a sequência ScADR1 apresenta alta similaridade com a sequência de *O. sativa* (▲). O tamanho em aminoácidos das sequências das proteínas da família ADR1 difere nas espécies vegetais apresentadas. A maior sequência protéica é do homólogo em *S. bicolor* (●), contendo 872 aminoácidos (Quadro 16). Análises no programa SMART e o alinhamento no ClustalW2 indica

que para as espécies vegetais *Coffea sp.* (347 aminoácidos), *Gossypium sp.* (453 aminoácidos), *S. tuberosum* (232 aminoácidos), *G. max* (312 aminoácidos), *T. aestivum* (349 aminoácidos) e *L. japonicus* (313 aminoácidos) no Quadro 16, (■), apenas parte das sequências dos transcritos correspondente as proteína foram depositadas no banco de dados TIGRs/gene índices.

Nos homólogos de ADR1 alinhados (Figura 9) pelo programa Clustalw foram feitas buscas dos motivos TVS, QHDV, RBNS-D e PKAE (localizados na região NBS N terminal) em ScADR1 (*Saccharum sp.*). Os resultados mostram que o motivo TVS não está presente em ScADR1 (Figura 9), embora, seja encontrado em *A. thaliana*, *V. vinifera*, *Gossypium sp.*, *Poplar sp.*, *M. truncatula*, *L. japonicus*. Os motivos conservados QHDV, RBNS-D, PKAE foram identificados em homólogos de ADR1 nas diferentes espécies, incluindo cana-de-açúcar (Figura 9).

Figura 9: Alinhamento Múltiplo da região N-terminal na família da proteína ADR1 em *A.thaliana* e seus homólogos em outras espécies vegetais, incluindo cana-de-açúcar.

Para identificação dos motivos conservados TVS (A), QHDV (B), RBNS-D (C) e PKAE (D) a região N-terminal dos homólogos ADR1 em diferentes espécies foram alinhadas. Pode-se observar o alinhamento das sequências na região onde os motivos estão presentes. Pode-se observar em A o motivo TVS, em B o motivo QHDV, em C o motivo RBNS-D e D o motivo PKAE. As espécies apresentadas são *L. japonicus*, *M. trucatula*, *G. max*, *Poplar sp.*, *Gossypium sp.*, *S. bicolor*, *S. officinarum sp.*, *O. sativa*, *Z. mays*, *T. aestivum*, *S. tuberosum*, *L. esculentum*, *Coffe*, *V. vinifera* e *A. thaliana*. As sequências conservadas apresentam as mesmas cores. Cada letra corresponde ao aminoácido, “a”, aromáticos (F, Y, W, H); “l”, alifáticos (I, V, L); “h”, hidrofóbicos (aromáticos, alifáticos, A, G, M, C, K, R, T); “+”, positivos (H, K, R); “-”, negativos (D, E); “p”, polar (positivo, negativo, Q, N, S, T, C).

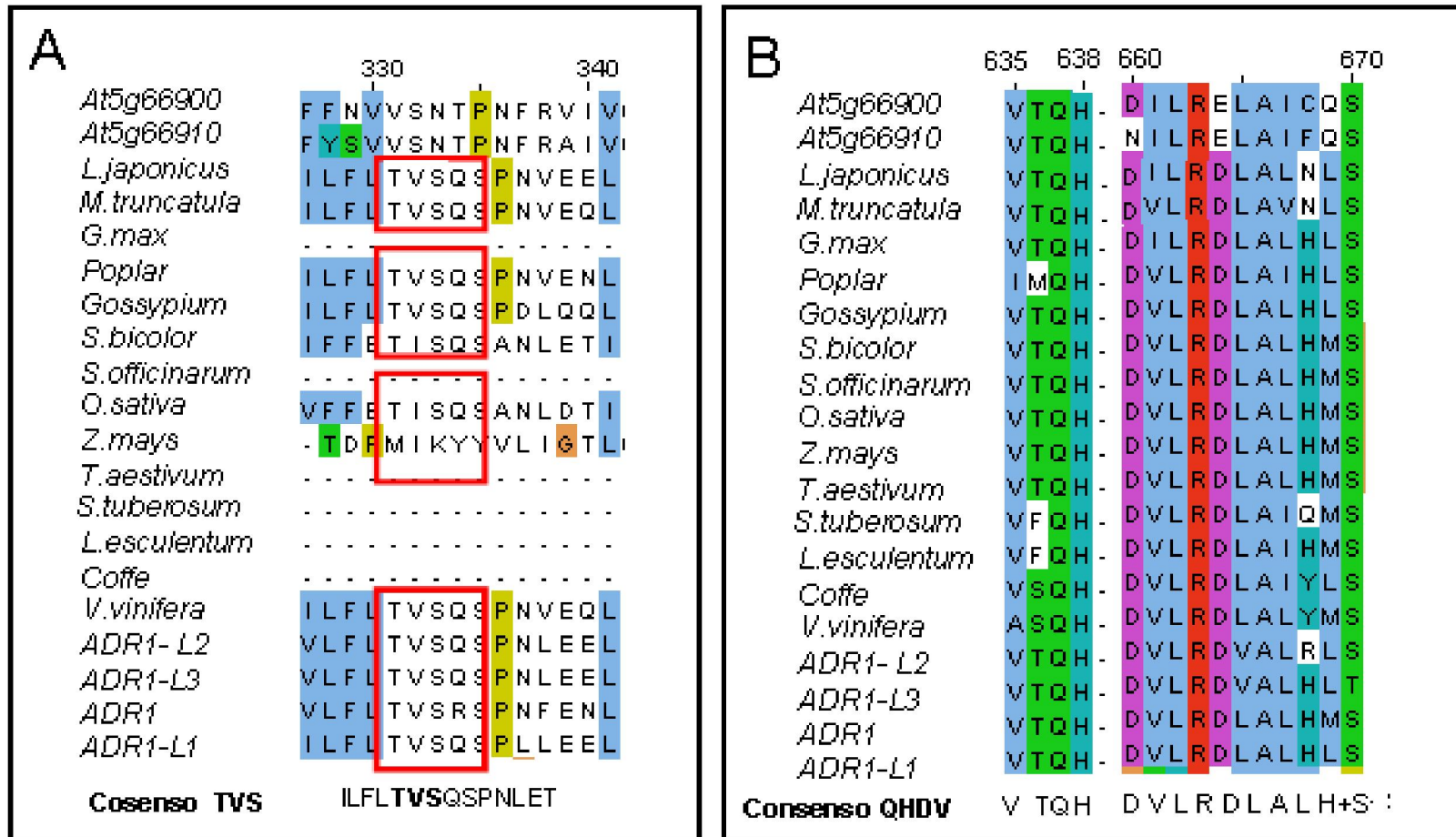


Figura 9 A e B: Alinhamento dos motivos TVS e QHDV, respectivamente identificados nas sequências da família ADR1 em diferentes espécies vegetais.

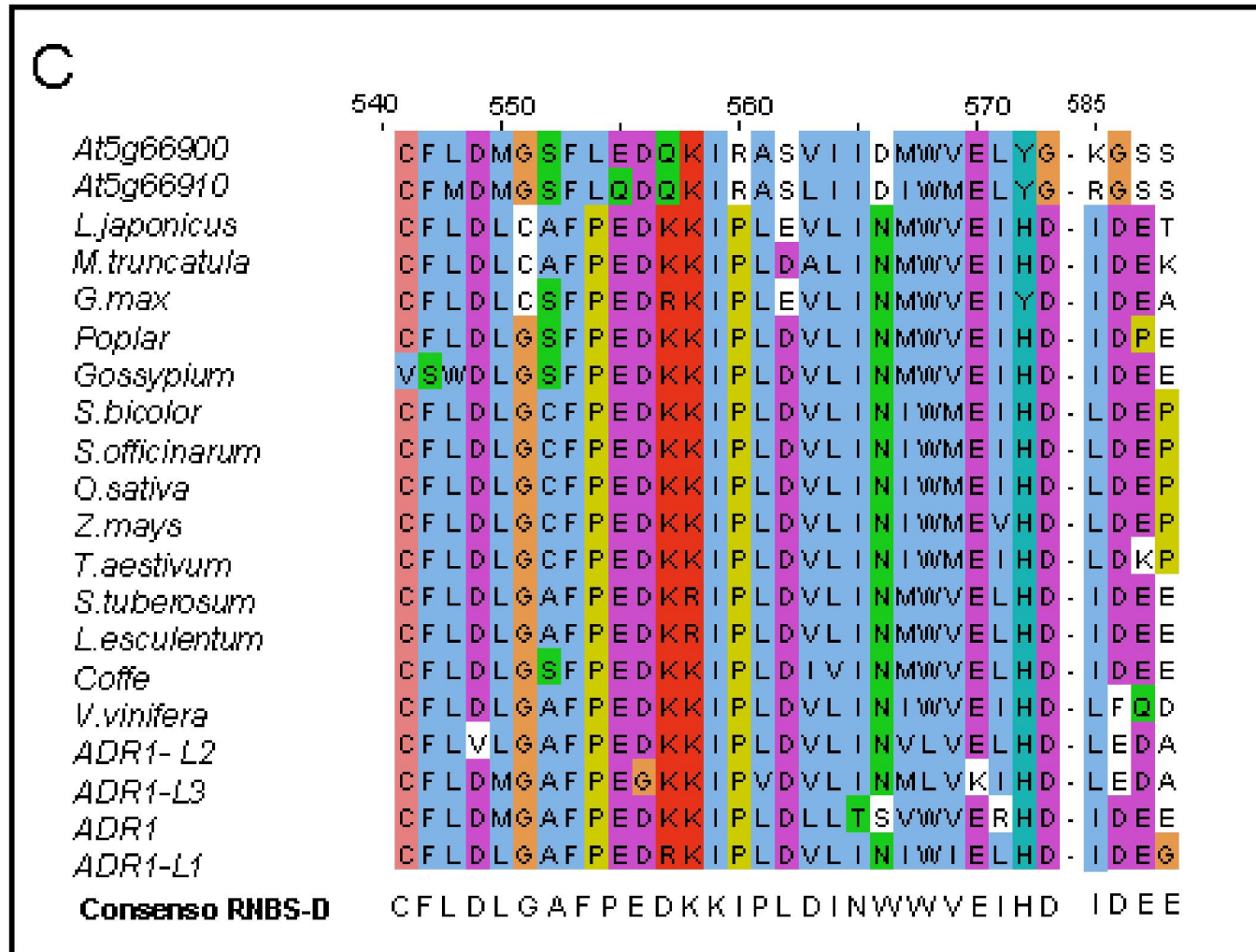


Figura 9 C: Alinhamento dos motivo RBNSD identificados nas seqüências da família ADR1 em diferentes espécies vegetais.

6.9 – Relação filogenética da proteína ScADR1 com sequências similares em outras espécies vegetais

Análises filogenéticas foram efetuadas com os domínios NB-ARC dos diferentes homólogos de ADR1, incluindo a sequência de cana-de-açúcar ScADR1 (Figura 10). Os genes At5g66900 e At5g66910 foram incluídos como controle do Clado A. As proteínas RPS2, RPM1, RPP13 e RPM1, RPP8 são controles representantes dos clados CNL-B, CNL-C, CNL-D respectivamente, e RPP1 é controle da família TNL (TIR). Como resultado, as sequências homólogas em 14 espécies vegetais, incluindo cana-de-açúcar se agruparam com os 4 membros da família ADR1 em *A.thaliana*. Os controles CC-NB-LRR (AtRPM1, AtRPP8, RPS2) e TIR(RPP1) agruparam em braços distantes da árvore. Os dados indicam que a sequência ScADR1 e as 13 sequências de diferentes espécies vegetais são homólogas à família ADR1 de *A.thaliana*.

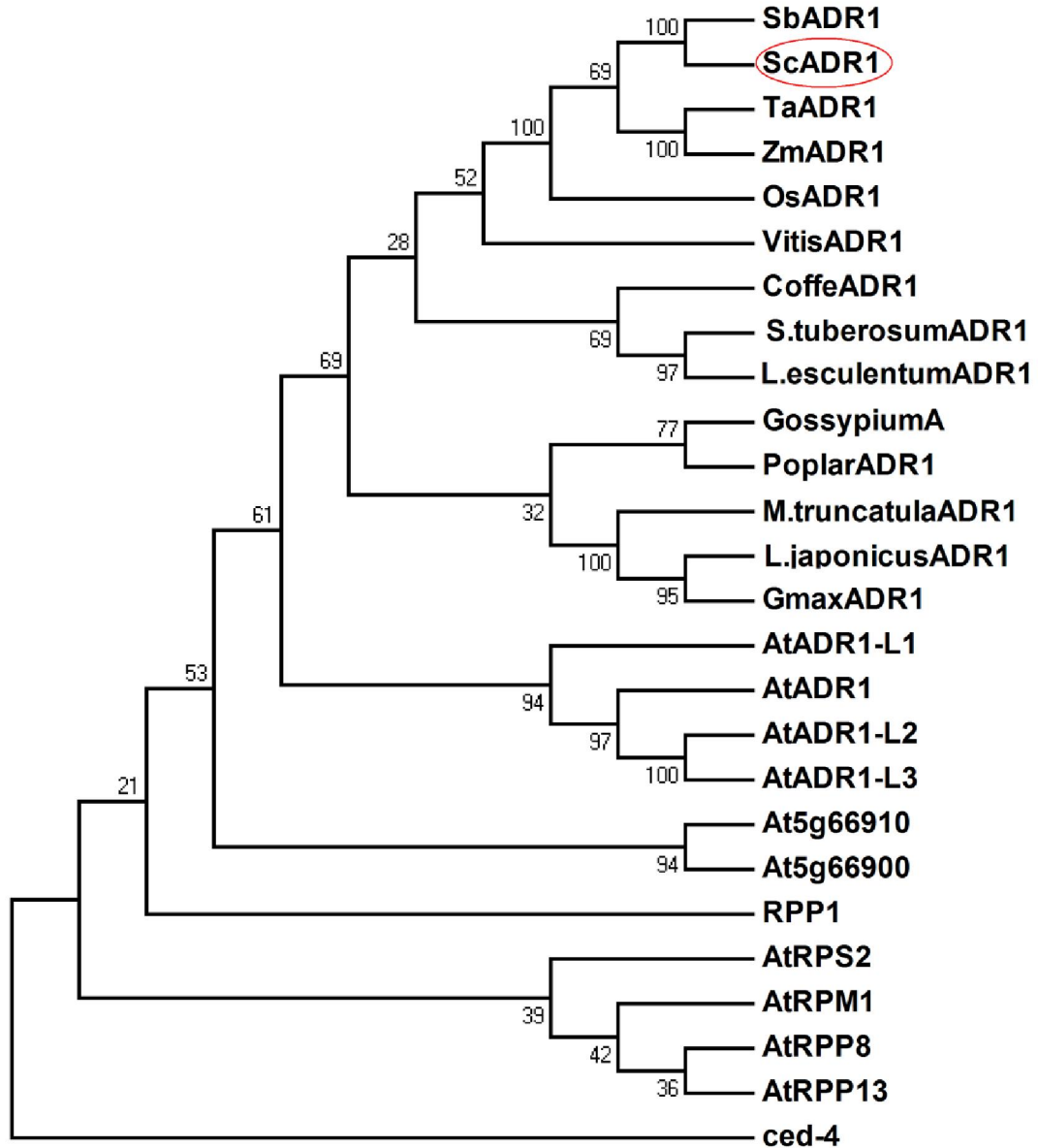


Figura 10: Relação filogenética da família ADR1 em plantas, incluindo ScADR1.

A sequência Ced4 da proteína de *C.elegans* foi utilizada como grupo externo. A sequência de aminoácidos das regiões NBS-ARC foram alinhadas no programa ClustalW. Os resultados de alinhamento (<http://align.genome.jp/>) no formato "Phylip" foram submetidos para o programa online PHYML (<http://atgc.lirmm.fr/phyml/>) para gerar uma árvore "maximum likelihood bootstrapped". A proteína ScADR1 está circulada. Os "scores" de "bottstrap" correspondem aos números.

7-DISCUSSÃO:

7.1- Identificação e análises de proteínas LRR-RLKs em cana-de-açúcar

Em 2008, Takai e colaboradores mostraram que, em arroz, o receptor OsFLS2 detecta flagelina de linhagens incompatíveis de bactérias patogênicas do gênero *Pseudomonas avenae*. Entretanto, linhagens compatíveis da bactéria não induzem forte resposta de defesa. Os dados na literatura sugerem que a via de percepção FLS2-flg22 seja conservada em monocotiledôneas e dicotiledôneas (Takai et al., 2008). No presente trabalho, foram identificados transcritos em cana-de-açúcar que codificam a seqüência similar à proteína FLS2. O domínio tirosina cinase de ScFLS2 apresenta alta similaridade com seus homólogos em outras espécies vegetais. Em cana-de-açúcar o transcrito *scfls2* não apresenta acúmulo nas bibliotecas de plantas inoculadas com os endófitos. O gene parece não ser regulado nessas condições. Contudo, a expressão ocorre em raízes e na zona transição entre raiz e folha. Análises funcionais de FLS2 poderão ser executadas em cana-de-açúcar visando sua caracterização.

O gene *serk* foi caracterizado inicialmente em *Daucus carota* como marcador de embriogênese somática. Os genes *serk1* e *serk2* têm sido considerados marcadores de competência para embriogênese somática em *A. thaliana*. Sua superexpressão aumenta a habilidade de células em suspensão e calos de sofrerem embriogênese somática (Colcombet et al. 2005; Hecht, 2001). Em milho, seqüências similares as proteínas SERK1, SERK2 e SERK3 foram encontradas. A expressão *zmserk2* ocorre em tecidos embriogênicos e não embriogênicos, porém a expressão de *zmserk1* é encontrada apenas em calos e tecidos associados à reprodução (Baudino et al., 2001). A expressão de *zmserk2* é uniforme em todos tecidos implicando funções adicionais a esse receptor. Os membros da família SERK em plantas estão envolvidos em vias de desenvolvimento e defesa atuando em resposta a esteróides e peptídeos ligantes (Heese et al., 2007).

Em 2005, Hu et alli mostraram que *osserk1* é induzido por patógenos e por moléculas de sinalização de defesa tais como: ácido salicílico, ácido jasmônico e ácido abscísico. O gene *osbiserk1* é induzido por benzotriazol (BTH) e interações

não compatíveis com *M. grisea* (Song et al., 2008). Em *A. thaliana* o gene *serk1* é induzido por ácido diclorofenoxiacético (2,4D). Em *M.truncatula* o gene *serk1* é induzido durante o desenvolvimento da planta e é altamente expresso durante a formação de nódulos na associação com *Rhizobium* (Nolan et al., 2009).

Em nossas análises foram identificados transcritos que codificam a proteína ScSERK like 1, que é ortóloga à OSBISERK1 e apresenta alta similaridade com todos os membros da família SERK em diferentes espécies, incluindo AtSERK3/BAK1. O transcrito *scserk-like 1* não é induzido na biblioteca de plantas inoculadas com endófitos, mas é induzido em tecidos meristemáticos, tais como broto lateral (LR) e meristema apical (LR). Cabe ressaltar que, dada a alta similaridade entre ScSERK-like 1 e todos os membros SERKs, não podemos descartar a hipótese de que esta proteína possa desempenhar papel funcional na sinalização de defesa, embriogênese ou sinalização de brassinosteróides. Para tanto, análises de expressão em diferentes tratamentos e/ou condições poderão ser realizadas para avaliar o papel funcional de ScSERK-like 1.

Em *A. thaliana*, o receptor SERK 3/BAK1 é um regulador central da imunidade inata, atuando como co-receptor de FLS2 para mediar a resposta de defesa, após o reconhecimento da flagelina de *P. syringae* pv *tomato*. Este receptor tem também papel na via de sinalização de brassinosteróides, atuando na interação com o receptor “BR-INSENSITIVE 1” (BRI1) (Heese et al., 2007).

Em 2007, Chinchilla e colaboradores mostraram que mutantes para AtBAK1 apresentam sensibilidade parcial para os MAMPs flagelina e EF-tu. Com isto, acredita-se que outras proteínas da família SERKs possam substituir BAK1 para mediar a resposta imune. As diferentes proteínas SERKs podem possuir funções redundantes, como informado para SERK1 e SERK2 (Albrecht et al., 2005). Em 2008, foi demonstrado que o receptor SERK1 pode associar com BRI (receptor brassinosteróide) e mediar a resposta de sinalização a brassinosteróides, sugerindo que BAK1 pode ser substituído por SERK1. Os dados indicam que apenas SERK1 e BAK1 estão envolvidos na via de sinalização de brassinosteróides (Albrecht et al., 2008). Adicionalmente foi demonstrado que o receptor SERK4 é parcialmente redundante com SERK3, pois este também é capaz de mediar respostas de defesa

(Albrecht et al., 2008). Plantas mutantes para SERK4 são mais suscetíveis a colonização por patógenos (Albrecht et al., 2008).

Em 2009, Vicentini e colaboradores identificaram o gene *scbak1* em cana-de-açúcar ortólogo ao gene *Atbak1*. O gene *scbak1* é induzida em folhas maduras de plantas contendo alto conteúdo de sacarose. O papel deste gene parece estar associado à síntese de sacarose e processos específicos nas células da bainha de folhas maduras (Vicentini, et al., 2009). Ainda não foi demonstrado se ScBAK1 atua na resposta imune inata em cana-de-açúcar.

Neste trabalho, foi identificado o gene *scserk-like 2*, parálogo do gene *scbak1*. O gene *scserk-like 2* é induzido por *G. diazotrophicus* e *H. rubrisubalbicans*, sendo também expresso em outros tecidos. Na biblioteca de plantas inoculadas com *H.rubrisubalbicans* foi verificado o maior nível de ocorrência para *scserk-like 2*. Os dados indicam que *scserk-like 2* está envolvido no processo de associação entre endófitos e cana-de-açúcar. Com isso, podemos especular que a proteína ScSerk like-2 pode estar associado à percepção ou resposta à microrganismos, atuando como um co-receptor, interagindo com outros receptores LRR-RLKs.

O gene *shr5* (LRR-RLK) de cana-de-açúcar está associada ao controle do estabelecimento da associação benéfica entre cana-de-açúcar e *G.diazotrophicus* (Vinagre et al. 2006). A expressão do gene *shr5* é diminuída durante a colonização por *G.diazotrophicus* em cultivares de cana-de-açúcar que apresentam alta FBN, e nenhuma diferença significativa é observada quando as plantas são infectadas por outros microrganismos (Vinagre et al., 2006).

No presente trabalho foram identificados transcritos que codificam uma proteína LRR-RLK, denominada ScMAR1. Tais transcritos são altamente induzidos por *H. rubrisubalbicans*, porém não são induzidos por *G. diazotrophicus*. A sequência da proteína ScMAR1 é conservada em monocodilêneas (*T. aestivum*, *Z. mays*, *O. sativa* e *S. bicolor*) e apresenta similaridade com as proteínas LRR-RLKs, At4g36180 e At1g75640 de *A.thaliana*. Os dados na literatura apontam o promotor do gene At4g36180 como um possível alvo do fator de transcrição WRKY53 (Laun et al., 2004).

Os fatores de transcrição WRKY compõem uma família de 74 membros em *A. thaliana*, nos quais 43 respondem a infecção por *P. syringae* pv *tomato* ou ácido

salicílico e são componentes importantes da via de sinalização de defesa em *A. thaliana* (Murray et al., 2007). Em *A. thaliana*, o fator de transcrição WRKY53 é um regulador positivo da senescência foliar e da resistência basal contra *P. syringae pv tomato* (Murray et al., 2007). Plantas mutantes para WRKY53 apresentam aumento na suscetibilidade para *P. syringae pv tomato*. O fator de transcrição WRKY53 é induzido por ácido salicílico e reprimido pela sinalização por ácido jasmônico e etileno (Li et al., 2004, Yu et al. 2001). Em arroz (Nipponbare), *oswrky53* é induzido por oligômeros de quitina e pelo fungo *M. grisea* estirpe 007 (Chujo et al., 2007).

Nossas análises filogenética e de expressão (por meio de “Northern eletrônico”) sugerem que *scmar1* codifica um novo receptor induzido em resposta à microrganismos em cana-de-açúcar. Porém, análises da expressão gênica durante a interação com endófitos e patógenos devem ser realizadas para confirmar nossas evidências. Assim, poderemos esclarecer se o gene *scmar1* é específico da resposta a endófitos e se o gene *scmar1* atua na resposta imune inata contra microrganismos.

Além dos genes descritos anteriormente, foram encontradas duas seqüências cujas seqüências em aminoácidos foram denominadas ScLRR-RLK 2 e ScLRR-RLK 3. Tais proteínas apresentam similaridades com a proteína LRR-RLK SRF7 (Strubbelig receptor family 7) de *A. thaliana* e com a proteína LRR-RLK LTK1 de *Z. mays*. Strubbelig 7 é um membro de uma família de proteínas LRR-RLKs composta por 9 membros (SRF1-SRF9) em *A. thaliana* (Eyüboğlu et al., 2007), e está relacionado a família LRR-RLK LTK de *Z. mays* (Li e Wurtel, 1988). Os membros desta família desempenham função no controle do desenvolvimento e morfogênese celular (Eyüboğlu et al., 2007, Chevallier 2005). Algumas evidências apontam que a proteína SRF7 (LRR-RLK), em *A. thaliana*, tenha papel na biossíntese da parede celular e processos que requerem a enzima celulase (Eyüboğlu et al., 2007). Interessantemente, a proteína SFR6 (LRR-RLK) em *A. thaliana* é induzido pelo fungo arbuscular micorriza (Eyüboğlu et al., 2007). A proteína LTK1 em *Z. mays* é induzida nas raízes, folhas, endosperma, embrião e óvulos não fertilizados, e está associada com o controle da divisão celular nestes tecidos (Li e Wurtzel, 1998).

Nossas análises mostram que o transcrito *sclrr-rlk 2* é induzido por *G. diazotrophicus*. Em cana-de-açúcar tal gene pode estar sendo induzida por uma

cascata a “downstream” da percepção de *G. diazotrophicus*, e pode estar associada a morfogênese celular ou divisão celular durante a colonização por *G. diazotrophicus*.

O gene *sclrr-rlk3* (similar à SFR7 e LTK1) não parece ser induzido por endófitos, porém é expresso em vários tecidos, tais como: cartucho foliar (LR), broto lateral (LB), meristema apical (AM) e outros. Sua função nestes tecidos pode estar associada com o controle da morfogênese celular e proliferação celular nestes tecidos.

A proteína ScLRR-RLK5 apresenta similaridade com a proteína At4g23740 (LRR-RLK) de *A. thaliana*, porém esta ainda não foi caracterizada. Em cana-de-açúcar o transcrito *sclrr-RLK5* parece ser induzido por *G. diazotrophicus*. Análises da expressão gênica deste gene deverão ser realizadas em cana-de-açúcar durante a associação com patógenos e endófitos para avaliar seu papel durante a interação.

7.2 – Genes que codificam proteínas com domínios CC-NB-LRR induzidos por *Gluconacetobacter diazotrophicus* e *Herbaspirillum rubrisubalbicans*

Em 2003, Grant e colaboradores mostraram que plantas mutantes de *A.thaliana* com expressão constitutiva de *adr1* adquirem resistência contra os patógenos biotróficos *E. cichoracearum* (UED1), *P. parasitica* (Noco2) e contra o vírus do mosaico do pepino, mas não para o fungo necrotrófico *Botrytis cinerea*. As plantas também apresentavam um alto acúmulo de ROS, ácido salicílico e transcritos associados à defesa (PR1). Foi observada também a acumulação de *adr1* em folhas 1,5 horas após a infecção com *P. syringae* pv *tomato* DC3000 (Grant et al., 2003). Segundo Grant 2003, a proteína ADR1 pode atuar na transdução de sinais durante a resistência local e no estabelecimento da imunidade sistêmica. Além do seu papel na defesa, em 2004, Chini e colaboradores mostraram que o aumento da expressão de *adr1* confere resistência à seca em 100% das plantas submetidas ao estresse hídrico e induz a expressão do fator de transcrição DREB2A (elemento responsivo a desidratação). Este fator de transcrição é capaz de se ligar ao promotor de genes de resposta a seca, em uma via independente de ácido abscísico (ABA).

Em cana-de-açúcar foi identificada um “cluster” que codifica uma sequência homóloga à proteína ADR1, o transcrito *scadr1* é altamente induzida por *G.*

diazotrophicus. A proteína ScADR1 é um membro da classe CC-NB-LRR. Tal classe de proteínas está envolvida na detecção de patógenos e ativação do sistema de defesa através do reconhecimento direto ou indireto de moléculas secretadas por esses microrganismos.

A expressão de *scadr1* sugere que o sistema de defesa em cana-de-açúcar é ativado durante a colonização por *G. diazotrophicus*. Possivelmente, a expressão da proteína ScADR1 pode participar no controle da população de *G. diazotrophicus* nos tecidos de cana-de-açúcar.

G. diazotrophicus possui os sistemas de secreção do tipo II (Arrieta et al., 2004) e IV (blast/ncbi). Porém, ainda não foi demonstrado se *G. diazotrophicus* secreta moléculas efetoras com intuito de subverter o sistema de defesa do hospedeiro. As proteínas NB-LRRs podem atuar como receptores do sistema imune ou “downstream” de outros receptores (exemplo: NB-LRRs, RLPs) durante a imunidade (Caplan et al., 2008). Plantas de *N. benthamiana*, para prover resistência contra vírus do mosaico do tabaco, necessitam do receptor do sistema imune N e da proteína CC-NB-LRR denominada NRG1 (gene N requerimento) (Gabriels et al., 2007). É importante ressaltar que ScADR1 (CC-NB-LRR) é uma proteína citoplasmática, assim, seu modo de atuação/reconhecimento é citoplasmático. Com isso, podemos especular que ADR1 pode estar funcionando como um receptor do sistema imune, ou como um co-receptor de outros receptores, tais como proteínas LRR-RLKs, RLPs ou outras proteínas NB-LRRs.

Em 2006, Arencibia et al. mostraram que durante a co-inoculação de cana-de-açúcar com *G. diazotrophicus* e *Xanthomonas albilinea*, *G. diazotrophicus* confere um efeito protetor contra o patógeno. Adicionalmente, observou-se a indução de transcritos associados à defesa (Arencibia et al., 2006). O efeito protetor mediado por *G. diazotrophicus* é similar à resistência sistêmica induzida (ISR) (Arencibia et al., 2006). Conforme citado anteriormente, em *A. thaliana* transcritos para a proteína ADR1 foram encontrados em tecidos não tratados, após infecção com *P. syringae* pv *tomato* (Grant et al., 2003). Com isso, podemos especular que ScADR1 participe do efeito protetor produzido durante a colonização por *G. diazotrophicus*.

Conforme mostrado neste trabalho, as proteínas da família ADR1 em plantas são conservadas. Este nível de conservação pode refletir na função dentro ou fora do

contexto da doença, tal como tolerância a seca em plantas de cana-de-açúcar quando colonizadas por *G. diazotrophicus*.

Para analisar o papel da proteína ScADR1 na interação com microrganismos, ensaios experimentais devem ser executados para verificar sua expressão durante a colonização por outros endófitos ou patógenos de cana-de-açúcar. Caso seja confirmada a especificidade de indução e resposta de ScADR1 a presença de *G. diazotrophicus*, essa proteína poderá ser aplicada com um marcador da associação entre *G. diazotrophicus* e cana-de-açúcar.

Em 2005, Chini e Loake identificaram sequências homólogas à proteína ADR1 em 11 espécies vegetais. Segundo, Chini e Loake 2005, o motivo TVS está presente nas sequências homólogas em *A. thaliana*, *M. truncatula* e *O. sativa*. Neste trabalho, o motivo TVS foi identificado em *A. thaliana*, *V. vinifera*, *Gossypium sp.*, *Poplar sp.*, *M. truncatula* e *L. japonicus*, mas está ausente em *Saccharum sp.* (ScADR1). Já os motivos QHDV, RBNS-D, PKAE foram identificados nas sequências ScADR1 e nos homólogos encontrados em diferentes espécies vegetais. Portanto, as análises de motivos, comparação de sequências por alinhamento e árvore filogenética indicam que a proteína ScADR1 pertence à família ADR1.

Além de ScADR1, foram identificadas 5 sequências NB-LRRs similares aos receptores RPM1, RPP8 e RPP13, porém o nível de similaridade entre as sequências é baixo. Os transcritos *scrpm1*, *scnb-lrr1/rpm1*, *scnb-lrr2/rpp13*, *scnb-lrr3/rpp8* e *scnb-lrr4/rpm1* não parecem ser induzidos por endófitos. Contudo, experimentos deverão ser realizadas para avaliar se tais genes de resistência desempenham papel na defesa em cana-de-açúcar contra microrganismos patogênicos.

Neste trabalho foi constatado que endófitos induzem a ativação de transcritos que codificam proteínas LRR-RLKs e NB-LRRs. Isto indica que durante interações endófiticas o sistema de defesa é ativado e viabiliza o controle da população destes microrganismos. Porém, tais proteínas não impedem a colonização por endófitos. A próxima etapa consistirá na análise da expressão gênica dos receptores identificados visando confirmar os dados “in silico” e as evidências obtidas. O estudo dos receptores identificados possibilitará uma maior compreensão do papel de proteínas do sistema de defesa neste tipo de interação.

8- Conclusões:

8.1. Transcritos que codificam proteínas LRR-RLKs são acumulados em plantas de cana-de-açúcar colonizadas pelas bactérias endofíticas *G. diazotrophicus* e *H. rubrisubalbicans*. Dentre eles, podemos destacar ScSERK-like 2, ScMAR1 e ScLRR-RLK2.

8.2. A forte indução de transcritos da proteína ScMAR1 em resposta à *H. rubrisubalbicans*, sugere a ativação de resposta de defesa distinta daquela observada para *G. diazotrophicus*

8.3. Proteínas do tipo NB-LRR estão envolvidas na percepção e ativação da resposta à endófitos em cana-de-açúcar. Dentre elas, transcritos para ScADR1 (CC-NB-LRR) são altamente induzidos por *G. diazotrophicus*.

8.4. Juntos, os dados permitiram sugerir a ocorrência de ativação dos mecanismos de defesa da planta, através da transcrição de receptores, durante a colonização por bactérias endofíticas. Tal resposta mostrou-se específica para cada espécie bacteriana analisada.

9 - Referências Bibliográficas:

Abramovitch, R.B., Anderson, J.C., Martin, G.B. Bacterial elicitation and evasion of plant innate immunity. *Nature Reviews Molecular Cell Biology* 7(8): 601- 11. 2006.

Abramovitch, R.B., Kim, Y.J., Chen, S., Dickman, M.B., Martin, G.B. Pseudomonas type III effector AvrPtoB induces plant disease susceptibility by inhibition of host programmed cell death. *EMBO Journal* 22(1): 60- 9. 2003

Albrecht, C., Russinova, E., Hecht, V., Baaijens E., de Vries, S. The *Arabidopsis thaliana* somatic embryogenesis receptor-like kinase 1 and 2 control male sporogenesis. *The Plant Cell* 17: 3337- 3349. 2005.

Albrecht, C., Russinova, E., Kemmerling, B., Kwaaitaal, M., de Vries, S.C. *Arabidopsis* somatic embryogenesis receptor kinase proteins serve brassinosteroid - dependent and independent signaling pathways. *Plant Physiology* 148: 611- 619. 2008.

Arencibia, A.D, Vinagre, F., Estevez, Y., Bernal, A., Perez, J., Cavalcanti, J., Santana, I., Hemerly, A.S. Gluconacetobacter diazotrophicus Elicits a Sugarcane Defense Response Against a Pathogenic Bacteria Xanthomonas albilineans. *Plant Signaling & Behavior* 1(5): 265- 73. 2006.

Arrieta, J.G., Sotolongo, M., Menéndez, C., Alfonso, D., Trujillo, L.E., Soto, M., Ramírez, R., Hernández, L. A type II protein secretory pathway required for levansucrase secretion by Gluconacetobacter diazotrophicus. *The Journal of Bacteriology* 186(15): 5031- 9. 2004.

Arrighi, J.F. Barre, A. Amor, B.B., Bersoult, A., Soriano, L.C., Mirabella, R., Niebel, F.C. , Journet, E.P., Ghérardi, M., Huguet, T., Geurts, R., Dénarié, J., Rougé , P., Gough, C. The *Medicago truncatula* lysine motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. *Plant Physiology* 142: 265- 279. 2006.

Ashfield, T., Ong, L.E., Nobuta, K., Schneider, C.M., Innes, R.W. Convergent evolution of disease resistance gene specificity in two flowering plant families. *The Plant Cell* 16: 309- 318. 2004.

Baldani, J.I., Pot, B., Kirchhof, G., Falsen, E., Baldani, V.L., Olivares, F.L., Hoste, B., Kersters, K., Hartmann, A., Gillis, M., Döbereiner, J. Emended description of *Herbaspirillum*; inclusion of [*Pseudomonas*] *rubrisubalbicans*, a milk plant pathogen, as *Herbaspirillum rubrisubalbicans* comb. nov.; and classification of a group of clinical isolates (EF group 1) as *Herbaspirillum* species 3. *International Journal of Systematic Bacteriology* 46 (3) : 802- 10. 1996

Baldani, J.I. Ocorrência e caracterização de *Azospirillum amazonense* em comparação com outras espécies deste gênero em raízes de milho, sorgo e arroz, M. Sc. Thesis, Universidade Federal Rural do Rio de Janeiro, RJ, Brasil. 1984.

Baldani, J.I., Reis, V.M., Kimura, O., Döbereiner, J. Bactérias fitopatogênicas fixadoras de N₂ em associação com plantas. Seropédica: Embrapa-CNPAB, 25p. (Embrapa-CNPAB. Documentos, 41). 1997.

Baldani, J.I., Baldano, V.L.D. History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. *Anais da Academia Brasileira Ciências* 77(3): 549- 79. 2005.

Bastián, F., Cohen, A., Piccoli, P., Luna, V., Baraldi, R., Bottini, R. Production of indole-3-acetic acid and gibberellins A1 and A3 by *Acetobacter diazotrophicus* and *Herbaspirillum seropedicae* in chemically-defined culture media. *Plant Growth Regulation* 24: p.7- 11, 1998.

Baudino, S., Hansen, S., Brettschneider, R., Hecht, V.F., Dresselhaus, T., Lörz, H., Dumas, C., Rogowsky, P.M. Molecular characterization of two novel maize LRR receptor-like kinases, which belong to the SERK gene family. *Planta*: 213(1): 1- 10. 2001.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L. GenBank: update. *Nucleic Acids Research* 32: D23- D26. 2004.

Benson, D.A., Karsch-Mizrachi I., Lipman, D.J., Ostell, J., Wheeler, D.L. GenBank. *Nucleic Acids Research* 31(1): 23- 7. 2003

Bishop, G.J., Koncz, C. Brassinosteroids and Plant Steroid Hormone Signaling. *The Plant Cell* 14: S97- S110. 2002.

Boller, T., Felix, G. A Renaissance of Elicitors: Perception of Microbe-Associated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors. *Annual Review of Plant Biology* 60:379–406. 2009a.

Boller, T., He, S.Y. Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens. *Science* 324(5928): 742- 4. 2009b.

Bryan, G.T., Wu, K.S., Farrall, L., Jia, Y., Hershey, H.P., McAdams, S.A., Faulk, K.N. Donaldson GK, Tarchini R, Valent B. tA single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene Pi-ta. *The Plant Cell* 12(11): 2033- 46. 2000.

Büttner, D, Bonas, U. Who comes first? How plant pathogenic bacteria orchestrate type III secretion. *Current Opinion in Microbiology* 9(2): 193- 200. 2006.

Caplan, J., Padmanabhan, M., Dinesh-Kumar, S.P. Plant NB-LRR immune receptors: from recognition to transcriptional reprogramming. *Cell Host Microbe* 3 (3):126- 35. 2008.

Cavalcante, V.A and Döbereiner, J. A new acid tolerant nitrogen-fixing bacterium associated with sugarcane. *Plant and Soil* 108: 23- 31. 1988.

Chevalier, D., Batoux, M., Fulton, L., Pfister, K., Yadav, R.K., Schellenberg, M., Schneitz, K. strubbelig defines a receptor kinase-mediated signaling pathway regulating organ development in Arabidopsis. Proceedings of the National Academy of Sciences U S A 102(25): 9074- 9. 2005.

Chisholm, S.T., Coaker, G., Day, B., Staskawicz, B.J. Host-microbe interactions: shaping the evolution of the plant Immune response. Cell 124(4):803- 14. 2006.

Chini, A., Grant, J.J., Seki, M., Shinozaki, K., Loake, G.J. Drought tolerance established by enhanced expression of the CC-NBS-LRR gene, ADR1, requires salicylic acid, EDS1 and ABI1. The Plant Journal 38(5):810- 22. 2004.

Chini, A., Loake, G. J. Motifs specific for the ADR1 NBS-LRR protein family in *Arabidopsis* are conserved among NBS-LRR sequences from both dicotyledonous and monocotyledonous plants. Planta 221(4):597- 601. 2005.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nürnberger, T., Jones, J.D.G., Felix, G., Boller, T. A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature 448(7152):497- 500. 2007.

Chujo, T., Takai, R., Akimoto-Tomiyama, C., Ando, S., Minami, E., Nagamura, Y., Kaku, H., Shibuya, N., Yasuda, M., Nakashita, H., Umemura, K., Okada, A., Okada, K., Nojiri, H., Yamane, H. Involvement of the elicitor-induced gene OsWRKY53 in the expression of defense-related genes in rice. Biochimica Biophysica Acta 1769(7-8):497- 505. 2007.

Clark, S.E., Jacobsen, S.E., Levin, J.Z., Meyerowitz, E.M. The clavata and shoot meristemless loci competitively regulate meristem activity in Arabidopsis. Development 122(5):1567- 75. 1996.

Colcombet, J., Boisson-Dernier, A., Ros-Palau, R., Vera, C.E., Schroeder, J.I. *Arabidopsis* somatic embryogenesis receptor kinase 1 and 2 are essential for tapetum development and microspore maturation. *The Plant Cell* 17:3350- 3361. 2005.

Deslandes, L., Olivier, J., Peeters, N., Feng, D.X., Khounlotham, M., Boucher, C., Somssich, I., Genin, S., Marco, Y. Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proceedings of the National Academy of Sciences U S A* 100:8024-8029. 2003.

DeYoung, B.J., Innes, R.W. Plant NBS-LRR proteins in pathogen sensing and host defense. *Nature Immunology*: 7(12):1243- 9. 2006.

De la Fuente van Bentem, S., Vossen, J.H., de Vries, K.J., van Wees, S., Tameling, W.I., Dekker, H.L., *et al.* Heat shock protein 90 and its co-chaperone protein phosphatase 5 interact with distinct regions of the tomato I-2 disease resistance protein. *The Plant Journal* 43:284- 298. 2005.

Diévert, A., Clark, E. LRR-containing receptors regulating plant development and defense. *Development* 131:251- 261. 2004.

Döbereiner, J., Ruschel, A.P. Uma nova espécie de *Beijerinckia*. *Review Biology* 1:261- 272. 1958.

Döbereiner J. *Azotobacter paspali* sp. nov., uma bactéria fixadora de nitrogênio na rizosfera de *Paspalum*. *Pesquisa Agropecuária Brasileira* 1:357- 365. 1966.

Dong, Z., McCully, M.E., Canny, M.J. Does *Acetobacter diazotrophicus* Live and Move in the Xylem of Sugarcane Stems? Anatomical and physiological data. *Annals of Botany* 80:147- 158. 1997.

Enríquez-Obregón, G.A., Vázquez-Padrón, R.I., Prieto-Samsonov, D.L., De la Riva, G.A., Selman-Housein, G. Herbicide-resistant sugarcane (*Saccharum officinarum* L.) plants by *Agrobacterium*-mediated transformation. *Planta* 206: 20- 27. 1998.

Eyüboğlu, B., Pfister, K., Haberer, G., Chevalier, D., Fuchs, A., Mayer, K.F.X., Schneitz, K. Molecular characterisation of the Strubbelig-receptor family of genes encoding putative leucine-rich repeat receptor-like kinases in *Arabidopsis thaliana*. *BMC Plant Biology* 7:16. 2007.

Ferrari, S., Galletti, R., Vairo, D., Cervone, F., De Lorenzo, G. Antisense expression of the *Arabidopsis thaliana* AtPGIP1 gene reduces polygalacturonase-inhibiting protein accumulation and enhances susceptibility to *Botrytis cinerea*. *Molecular Plant Microbe Interaction*:19(8):931- 6. 2006.

Ferreira, B. S. Identificação de genes envolvidos na resposta ao estresse salino em cana-de-açúcar através de mineração de dados no banco de dados do projeto SUCEST, M. Sc. Thesis, Universidade Estadual do Norte Fluminense, RJ, Brasil. 2002.

Fisher, K., Newton, W.E. Nitrogenase proteins from *Gluconacetobacter diazotrophicus*, a sugarcane-colonizing bacterium. *Biochimica Biophysica Acta* 1750(2):154- 65. 2005.

Gabriels, S.H., Vossen, J.H., Ekengren, S.K., van Ooijen, G., Abd-El-Haliem, A.M., van den Berg, G.C., Rainey, D.Y., Martin, G.B., Takken, F.L., de Wit, P.J., and Joosten, M.H. An NB-LRR protein required for HR signaling mediated by both extra- and intracellular resistance proteins. *The Plant Journal* 50:14- 28. 2007.

Geurts, R., Fedorova, E., Bisseling, T. Nod factor signaling genes and their function in the early stages of *Rhizobium* infection. *Current Opinion in Plant Biology* 8:346- 352. 2005.

Gherbi, H., Markmann, K., Svistoonoff, S., Estevan, J., Autran, D., Giczey, G., Auguy, F., Péret, B., Laplaze, L., Franche, C., Parniske, M., Bogusz, D. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Frankiacteria. *Proceedings of the National Academy of Sciences U S A* 105(12):4928- 32. 2008.

Gómez-Gómez, L., Boller, T. FLS2: An LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Molecular Cell* 5(6):1003-11. 2000.

Godiard, L., Sauviac, L., Torii, K.U., Grenon, O., Mangin, B., Grimsley, N.H., Marco, Y. ERECTA, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt. *Plant Journal* 36(3):353- 65. 2003.

Grant, J.J., Chini, A., Basu, D., Loake, G.J. Targeted activation tagging of the *Arabidopsis* NBS-LRR gene, ADR1, conveys resistance to virulent pathogens. *Molecular Plant Microbe Interaction* 16(8):669- 80. 2003.

Guindon, S., Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systems Biology* 52(5):696- 704. 2003.

He, P., Shan, L., Lin, N., Martin, G.B., Kemmerling, B., Nürnberger, T. Sheen, J. Specific bacterial suppressors of MAMP signaling upstream of MAPKKK in *Arabidopsis* innate immunity. *Cell*.125:563- 575. 2006.

He, P., Shan, L., Sheen, J. Elicitation and suppression of microbe-associated molecular pattern-triggered immunity in plant–microbe interactions. *Cellular Microbiology* 9(6):1385- 96. 2007.

Hecht, V., Vielle-Calzada, J.P., Hartog, M.V., Schmidt, E.D., Boutilier, K., Grossniklaus, U., de Vries, S.C. The *Arabidopsis* somatic embryogenesis receptor kinase 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. *Plant Physiology* 127(3):803- 16. 2001.

Heese, A., Hann, D.R., Gimenez-Ibanez, S., Jones, A.M., He, K., Li, J., Schroeder, J.I., Peck, S.C., Rathjen, J.P. The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. *Proceedings of the National Academy of Sciences U S A* 104(29):12217- 22. 2007

Hu, H., Xiong, L., Yang, Y. Rice SERK1 gene positively regulates somatic embryogenesis of cultured cell and host defense response against fungal infection. *Planta*. 2005 Sep; 222(1):107-17.

Janjusevic R, Abramovitch RB, Martin GB, Stebbins CE. A bacterial inhibitor of host programmed cell death defenses is an E3 ubiquitin ligase. *Science* 311(5758):222- 6. 2006.

James, E.K., Olivares., F.L., Baldani, J.I., Döbereiner, J. *Herbaspirillum*, an endophytic diazotroph colonizing vascular tissue in leaves of *Sorghum bicolor* L. Moench *Journal of Experimental Botany* 48(308): 785- 797. 1997.

Jimenez-Salgado, T., Fuentes-Ramires, L.E., Tapia-Hernandez, A., Mascarua-Esparza, M.A.; Martinez-Romero, E., Caballero-Mellado, J. *Coffea arabica* L., a new host plant for *Acetobacter diazotrophicus*, and isolation of other nitrogen-fixing acetobacteria. *Applied and Environmental Microbiology* 63: 3676- 3683.1997.

Jin, T-S., Stone, J.M., Walker, J.C. HAESA, an *Arabidopsis* leucine-rich repeat receptor kinase, controls floral organ abscission. *Genes & Development* 14: 108- 117. 2000.

Jones, D.A., Takemoto, D. Plant innate immunity – direct and indirect recognition of general and specific pathogen-associated molecules. *Current Opinion in Immunology*.16:48- 62. 2004

Jones, K.M., Kobayashi, H., Davies, B.W., Taga, M.E., Walker, G.C. How rhizobial symbionts invade plants: the *Sinorhizobium–Medicago* model. *Nature reviews microbiology* 5(8):619- 33. 2007.

Jones, J.D.G., Dangl, J.L. The plant immune system. *Nature* 444: 323-329. 2006.

Júnior, F.B.R., Silva, L.G., Reis, V.M., Döbereiner, J. Ocorrência de bactérias diazotróficas em diferentes genótipos de cana-de-áçúcar. *Pesquisa agropecuária Brasileira* 35(5): p.985- 994. 2000.

Kerstens, K., Lisdiyanti, P., Komagata, K., Swings, J. The Family *Acetobacteraceae*: *Thegenera Acetobacter, Acidomonas, Asaia, Gluconacetobacter, Gluconobacter* and *Kozakia*. In: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H., Stackebrandt, E. (Eds.) *The Prokaryotes* (3rd Edn.): A hand-book on the biology of bacteria: Proteobacteria: alpha and beta sub classes. Springer 5:163- 200. 2006. New York.

Kimura, M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111- 120. 1980.

Kirchhof, G., Eckert, B., Stoffels, M., Baldani, J.I., Reis, V.M. Hartmann A. *Herbaspirillum frisingense* sp. nov., a new nitrogen-fixing bacterial species that occurs in C4-fiber plants. *International Journal of Systematic and Evolutionary Microbiology* 51:157- 168. 2001.

Krieg, N.R. Aerobic/microaerophilic motile helical/vibroid gram negative bacteria, In: KRIEG, N.R.; HOLT, J.G., ed. *Bergey's Manual of Systematic Bacteriology*. Baltimore: The Willians & Wilkins 1:71-90. 1984.

Lambais, M.R. In silico differential display of defense-related expressed sequence tags from sugarcane tissues infected with diazotrophic endophytes. *Genetics and Molecular Biology*, 24 (1-4): 103- 111. 2001.

Larkin, M., A, Blackshields., G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, IM., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. ClustalW and ClustalX version 2.0. *Bioinformatics* 23(21):2947- 8. 2007.

Lee, Y., Tsai, J., Sunkara, S., Karamycheva, S., Pertea, G., Sultana R., Antonescu, V., Chan, A., Cheung, F., Quackenbush, J. The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Research* 33:D71- 4. 2005.

Lima, E., Boddey, R.M., Döbereiner, J. Quantification of biological nitrogen-fixation associated with sugar cane using a ¹⁵N aided nitrogen balance. *Soil Biology and Biochemistry* 19:165- 170. 1987.

Li, J., Brader, G., Palva, E.P. The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. *The Plant Cell* 16:319- 331. 2004.

Li, Z., Wurtzel, E.T. The ltk gene family encodes novel receptor-like kinases with temporal expression in developing maize endosperm. *Plant Molecular Biology* 37(5):749- 61. 1998.

Liorente, F., Alonso-Blanco, C., Sánchez-Rodríguez, C., Jorda, L., Molina, A. Erecta receptor-like kinase and heterotrimeric G protein from *Arabidopsis* are required for resistance to the necrotrophic fungus *Plectosphaerella cucumerina*. *The Plant Journal* 43(2):165- 80. 2005.

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

Markmann, K., Giczey G., Parniske M. Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. *PLoS Biology* 6(3):e68. 2008.

Melotto, M., Underwood, W., Koczan, J., Nomura, K., He, S.Y. Plant Stomata Function in Innate Immunity against Bacterial Invasion. *Cell* 126: 969- 980. 2006.

Mewes, H.W., Frishman, D., Gruber, C., Geier, B., Haase, D., Kaps, A., Lemcke, K., Mannhaupt, G., Pfeiffer, F., Schüller, C., Stocker, S., Weil, B. MIPS: a database for genomes and protein sequences. *Nucleic Acids Research* 28(1):37- 40. 2000.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., Michelmore, R.W. Genome-wide analysis of NBSLRR encoding genes in *Arabidopsis*. *The Plant Cell* 15:809- 834. 2003.

Mithöfer. A. Suppression of plant defence in rhizobia–legume symbiosis. *Trends in Plant Science* 7(10):440- 4. 2002.

Morillo, S.A., Tax, F.E. Functional analysis of receptor-like kinases in monocots and dicots. *Current Opinion in Plant Biology* 9:460- 469. 2006.

Murray, S.L., Ingle, R.A., Petersen, L.N., Denby, K. J. Basal resistance against *Pseudomonas syringae* in *Arabidopsis* involves WRKY53 and a protein with homology to a nematode resistance protein. *Molecular Plant Microbe Interaction* 20: 1431–1438. 2007

Mysore, K.S., Ryu, C. Nonhost resistance: how much do we know? *Trends Plant Sci* 9(2): 97- 104. 2004.

Navarro, L., Zipfel, C., Rowland, O., Keller, I., Robatzek, S., Boller, T., Jonathan, D.G. The transcriptional innate immune response to flg22. Interplay and overlap with Avr

gene-dependent defense responses and bacterial pathogenesis. *Plant Physiology* 135 (2):1113- 28. 2004.

Nimchuk, Z., Eulgem, T., Holt, B.F. 3rd., Dangl, J.L. Recognition and response in the Plant immune system. *Annual Review Genetics* 37:579- 609. 2003

Nolan, K.E., Kurdyukov, S., Rose, R.J. Expression of the somatic embryogenesis receptor-like kinase (SERK1) gene is associated with developmental change in the life cycle of the model legume *Medicago truncatula*. *Journal Experimental Botany* 60(6):1759- 1771. 2009.

Nomura, K., Melotto, M., He, S.Y. Suppression of host defense in compatible plant-*Pseudomonas syringae* interactions. *Current Opinion in Plant Biology* 8(4):361- 368. 2005.

Nürnbergger, T., LipKa, V. Non-host resistance in plants: new insights into an old phenomenon. *Molecular Plant Pathology* (3), 335–345. (2005).

Oldroyd, G.E.D. Dissecting symbiosis: developments in nod factor signal transduction. *Annals of Botany* 87(6):709- 718. 2001.

Oldroyd, G.E., Downie, J.A. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annual Review Plant Biology* 59:519- 46. 2008.

Olivares, F.L.; Baldani, V.L.D.; Reis, V.M.; Baldani, J.L.; Döbereiner, J. Occurrence of the endophytic diazotrophs *Herbaspirillum* spp. in roots, stem and leaves predominantly of graminaceae. *Biology Fertility Soils*, v. 2, p.197-200, 1996.

Olivares, F.L., James, E.K., Baldani, J.I., Döbereiner J. Infection of mottled stripe disease-susceptible and resistant sugar cane varieties by the endophytic diazotroph *Herbaspirillum*. *New Phytologist* 135: 723- 737. 1997.

Peleg-Grossman, S., Volpin, H., Levine, A. Root hair curling and *Rhizobium* infection in *Medicago truncatula* are mediated by phosphatidylinositide-endocytosis and reactive oxygen species. *Journal of Experimental Botany* 58(7):1637- 1649. 2007.

Perin, L., Baldani, J.I., Reis, V.M. Diversidade de *Gluconacetobacter diazotrophicus* isolada de plantas de cana-de-açúcar cultivadas no Brasil. *Pesquisa Agropecuária Brasileira* 39(8): 763-770. 2004.

Pimentel, J.P., Olivares, F.L., Pitard, R.M., Urquiaga, S., Akiva, F., Döbereiner J. Dinitrogen fixation and infection of grass leaves by *Pseudomonas rubrisubalbicans* and *Herbaspirillum seropedicae*. *Plant and Soil* 137: 61- 65.1991.

Radwan, T. EL-S. EL-D., Mohamed, Z.K., Reis, V.M. Efeito da inoculação de *Azospirillum* e *Herbaspirillum* na produção de compostos indólicos em plântulas de milho e arroz. *Pesquisa Agropecuária Brasileira* 39(10):987- 994. 2004.

Radwan, El-S El-D. T., Mohamed, Z.K., Reis, V.M. Aeração e adição de sais na produção de ácido indol acético por bactérias diazotróficas. *Pesquisa Agropecuária Brasileira* 40(10):997- 1004. 2005.

Rafalski, J.A., Hanafey, M., Miao, G.H., Ching, A., Lee, J.M., Dolan, M., Tingey, S. *Acta Biochim Pol.* 45(4):929-34. New experimental and computational approaches to the analysis of gene expression. *Acta Biochimica Polonica* 45(4):929- 34. 1998.

Rafiqi, M., Bernoux, M., Ellis, J.G., Dodds, P.N. In the trenches of plant pathogen recognition: role of NB-LRR proteins. *Seminars in Cell Developmental Biology*. Epub ahead of print. 2009.

Rensink, W.A., Buell, C.R. Microarray expression profiling resources for plant genomics. *Trends in Plant Science* 10(12):603- 9. 2005.

Roncato-Maccari L.D.B., Ramos, H.J.O., Pedrosa, F.O, Alquini, Y., Chubatsu, L.S., Yates, M.G., Rigo, L.U., Steffens, M.B.R., Souza, E.M. Endophytic *Herbaspirillum*

seropedicae expresses nif genes in gramineous plants. FEMS Microbiology Ecology 45:39- 47. 2003

Roncato-Maccari, L.D.B., Ramos, H.J.O., Pedrosa, F.O., Alquini, Y., Chubatsu, L.S., Yates, M.G., Rigo, L.U., Steffens, M.B.R., Souza, E.M. Endophytic *Herbaspirillum seropedicae* expresses nif genes in gramineous plants. FEMS Microbiology Ecology 45:39- 47. 2003.

Ryan, C.A., Huffaker, A., Yamaguchi, Y. New insights into innate immunity in *Arabidopsis*. Cellular Microbiology 9(8):1902- 1908. 2007.

Saravanan, V.S., Madhaiyan, M., Osborne, J., Thangaraju, M., Sa, T.M. Ecological occurrence of *Gluconacetobacter diazotrophicus* and nitrogen-fixing Acetobacteraceae members: their possible role in plant growth promotion. Microbial Ecology 55(1):130- 40. 2008.

Shan, L., He, P., Sheen, J. Endless hide-and-seek: dynamic co-evolution in plant-bacterium warfare. Journal of Integrative Plant Biology 49 (1):105- 111. 2007.

Shiu, S.H., Karlowski, W.M., Pan, R., Tzeng, Y. H., Mayer, K.F.X., Li, W.H. Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. The Plant Cell 16: 1220- 1234. 2004.

Shiu, S-H, Bleecker, A.B. Plant receptor-like kinase gene family: diversity, function, and signaling. Science STKE 113: re22. 2001.

Shiu, S.H., Bleecker, A. B. Expansion of the receptor-like kinase pelle gene family and receptor-like proteins in *Arabidopsis*. Plant Physiology 132:530- 543. 2003.

Shiu, S.H., Karlowski, W.M., Pan, R., Tzeng, Y.H., Mayer, K.F., Li, W.H. Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. The Plant Cell 16(5):1220- 1234. 2004.

Sjölande, K. Phylogenomic inference of protein molecular function: advances and challenges. *Bioinformatics* 20(2):170- 179. 2004.

Stacey, G., Libault, M., Brechenmacher, L., Wan, J., May, G. D. Genetics and functional genomics of legume nodulation. *Current Opinion in Plant Biology* 9(2):110-121. 2006.

Song, W.Y., Pi, L.Y., Wang, G.L., Gardner, J., Holsten, T., Ronald, P.C. Evolution of the rice Xa21 disease resistance gene family. *The Plant Cell* 9(8):1279- 87. 1997.

Soto, M.J., Sanjuán, J., Olivares, J. *Rhizobia* and plant-pathogenic bacteria: common infection weapons. *Microbiology* 152:3167- 3174. 2006.

Song, D., Li, G., Song, F., Zheng, Z. Molecular characterization and expression analysis of OsBISERK1, a gene encoding a leucine-rich repeat receptor-like kinase, during disease resistance responses in rice. *Molecular Biology Reporter* 35(2):275-283. 2008.

Takai, R., Isogai, A., Takayama, S., Che, F.S. Analysis of flagellin perception mediated by flg22 receptor OsFLS2 in rice. *Molecular Plant Microbe Interaction* 21(12):1635- 1642. 2008.

Thompson, J. D., Higgins, D. G., Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22(22):4673-80. 1994.

Urquiaga, S., Cruz, K.H.S., Boddey., R.M. Contribution of nitrogen fixation to sugar cane: nitrogen-15 and nitrogen-balance estimates. *Soil Science Society of America Journal* 56:105-114. 1992.

Van Zanten, M., Snoek, L.B., Proveniers, M.C., Peeters, A.J. The many functions of ERECTA. *Trends in Plant Science* 14(4):214- 8. 2009.

Vettore, A.L., da Silva F,R., Kemper, E.L., Arruda, P. The libraries that made SUCEST. *Genetics Molecular Biology* 24:1- 7. (2001).

Vettore, A.L., da Silva, F.R, Kemper, E.L., Souza, G.M., da Silva, A.M., Ferro, M.I., Henrique-Silva, F., Giglioti, E.A., Lemos, M,V., Coutinho, L.L., Nobrega, M.P., Carrer, H., França, S.C., Bacci Júnior, M., Goldman, M.H., Gomes, S.L., Nunes, L.R., Camargo, L.E., Siqueira, W.J., Van Sluys, M.A., Thiemann, O.H., Kuramae, E.E., Santelli, R.V., Marino, C.L., Targon, M.L., Ferro, J.A., Silveira, H.C., Marini, D.C., Lemos, E.G., Monteiro-Vitorello, C.B., Tambor, J.H., Carraro, D.M., Roberto, P.G., Martins, V.G., Goldman, G.H., de Oliveira, R.C., Truffi, D., Colombo, C.A., Rossi M, de Araujo PG, Sculaccio SA, Angella A, Lima MM, de Rosa Júnior VE, Siviero F, Coscrato, V.E., Machado, M.A., Grivet, L., Di Mauro, S.M., Nobrega, F.G., Menck, C.F., Braga, M.D., Telles, G.P., Cara, F.A., Pedrosa, G., Meidanis, J., Arruda, P. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Research* 13:2725- 2735. (2003).

Vicentini, R., Felix J de M., Dornelas, M.C., Menossi., M. Characterization of a sugarcane (*Saccharum* spp.) gene homolog to the brassinosteroid insensitive1-associated receptor kinase 1 that is associated to sugar content. *Plant Cell Reporter* 28(3):481- 491. 2009.

Vinagre, F., Vargas, C., Schwarcz, K., Cavalcante, J., Nogueira, E.M., Baldani, J.I., Ferreira, P.C., Hemerly, A.S. SHR5: a novel plant receptor kinase involved in plant-N₂-fixing endophytic bacteria association. *Journal of Experimental Botany* 57(3):559- 69. 2006.

Wanderley-Nogueira, A.C., Soares-Cavalcanti, N.M., Morais, D.A., Belarmino, L.C., Barbosa-Silva, A., Benko-Iseppon, A.M. Abundance and diversity of resistance genes

in the sugarcane transcriptome revealed by in silico analysis. *Genetics Molecular Research* 6(4):866- 89. 2007

Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio M., Edgar, R., Federhen, S., Geer, L.Y., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D.J, Madden, T.L., Maglott, D.R., Ostell, J., Miller, V., Pruitt, K.D., Schuler, G.D., Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner, L., Yaschenko, E. Database resources of the national center for biotechnology information. *Nucleic Acids Research* 35:D5-12. 2007.

Wheeler, D. Using GenBank. *Methods Mol Biol* 406:23- 59. 2007.

Wilgenbusch, J.C, Swofford, D. Inferring evolutionary trees with PAUP*. *Current Protocols in Bioinformatics*. Chapter 6:Unit 6.4. 2003.

Yin, Y., Wu, D., Chory, J. Plant receptor kinases: systemin receptor identified. *Proceedings of the National Academy of Sciences U S A* 99(14):9090- 9092. 2002.

Yu, D., Chen, C., Chen, Z. Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression. *The Plant Cell* 13:1527- 1540. 2001.

Zambrano, A.Y., Demey, J.R., Fuchs, M., Gonzalez., V.R., De Sousa, R.O., Gutiérrez, R.O. Selection of sugarcane plants resistant to SCMV. *Plant Science* 165(1):221- 225(5). 2003.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J.D., Boller, T., Felix, G. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. *Cell* 125(4):749- 60. 2006.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)