UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE QUÍMICA

**Identificação e análise funcional de genes
relacionados à transdução de sinais na cana-de-
açúcar**

Flávia Riso Rocha

Tese de doutorado

Orientação: Dra. Glaucia Mendes Souza

SÃO PAULO

Agosto/2006

Aos meus pais, Solange e Cacildo, pelo carinho, pela compreensão, pelo suporte e por terem sempre acreditado no meu potencial.

Agradecimentos

À Dra. Gláucia Mendes Souza, pela orientação ao longo de todos esses anos, pelo incentivo e por todo o apoio,

À Flávia Stal Papini Terzi, que me acompanhou durante grande parte do meu projeto, pela amizade e por todos os ensinamentos,

Ao Ricardo Vencio, pelo suporte estatístico, pela prontidão em sempre ajudar e por todos os ensinamentos de análise de *microarrays*,

Aos professores Dr. Sérgio Verjovski de Almeida, Dra. Aline Maria da Silva e Dr. Hugo Armelin, que viabilizaram nossa participação no Projeto CAGE-lab,

Aos colaboradores do meu projeto: Dr. Marcelo Menossi, Dr. Eugênio Ulian, Dra. Adriana Hemerly, Dr. Márcio de Castro Silva Filho, Dra. Sônia Marli Zingaretti, Dr. Antônio Vargas de Oliveira Figueira, Juliana Felix, Renato Vicentini, Rodrigo Drummond, Fabiano Vinagre, Maria Cristina Falco, Sabrina Chabregas, Carla Barsalobres, Ane Marckbart de Medeiros e Fabiana Aparecida Rodrigues, que foram fundamentais para que todo o projeto se realizasse,

Ao pessoal do laboratório: Luciana Mantzouranis, Milton Yutaka, Humberto Tambor, Sarah Cavalcanti, Érica Bandeira, Alessandro Waclawovski, Rodrigo Fandino, Leonardo Fabri, Alessandra Paiva, Paulo Schlogl, Josélia Marques, Raquel Bagattini, Fernando Ogata, Kátia Oliveira, Rodrigo Mascarenhas. Agradeço por toda a ajuda e por terem deixado todos os meus dias no laboratório muito mais alegres,

Ao pessoal do CAGE, especialmente para a Adriana Matsukuma e a Denise Yamamoto, por todo o suporte e companheirismo ao longo desses vários anos de experimentos de *microarrays*,

A todo o pessoal do Bloco 12, o bloco mais animado desse Instituto,

Aos meus pais, que sempre estiveram ao meu lado, me apoiaram e por todo o amor,

Ao Luiz, que foi meu grande companheiro nos anos que se passaram,

À FAPESP, pelo suporte financeiro.

Índice

1. Abreviaturas	7
2. Resumos	10
2.1. Resumo	10
2.2. Abstract.....	12
3. Introdução.....	14
3.1. A cana-de-açúcar e os projetos SUCEST e SUCAST	14
3.2. Transdução de sinais em plantas	16
3.3. A tecnologia de microarranjos de cDNA.....	22
3.4. Estímulos bióticos e abióticos em plantas	28
3.4.1. Fitormônios: ácido jasmônico e ácido abscísico	29
3.4.2. Herbivoria	32
3.4.3. Interação com bactérias endofíticas	34
3.4.4. Deficiência em fosfato	35
3.4.5. Resposta à seca	37
3.5. O desenvolvimento de cultivares de cana-de-açúcar mais ricos em sacarose.....	38
4. Objetivos	42
5. Material e Métodos	44
5.1. Prospecção dos dados, anotação e análise de domínios conservados	44
5.2. Análise filogenética de proteínas quinases e RLKs.....	44
5.3. Material biológico para a extração de RNA.....	45
5.3.1. Amostras de órgãos de cana-de-açúcar	45
5.3.2. Amostras para a avaliação da variabilidade individual de plantas de cana-de-açúcar no campo	47
5.3.3. Amostras de órgãos para os ensaios de PCR em tempo real	48
5.3.4. Ensaios para a análise do perfil de expressão em resposta a estímulos ambientais e a tratamentos com fitormônios	49
5.3.5. Amostras de genótipos de cana-de-açúcar de alto e de baixo brix	52
5.4. Extração de RNA	53
5.5. Manutenção dos clones bacterianos.....	54
5.6. Amplificação dos insertos de DNA dos clones bacterianos	54
5.7. Purificação dos produtos de PCR para confecção dos microarranjos	55
5.8. Eletroforese em gel de agarose	55
5.9. Deposição dos produtos de PCR em lâminas de vidro	56
5.10. Síntese e marcação dos alvos de cDNA	56
5.11. Hibridização dos alvos marcados nos microarranjos de cDNA.....	58
5.12. Extração e análise dos dados	59
5.13. Reações de transcrição reversa	61
5.14. Reações de PCR em tempo real.....	62
5.15. Reações de sequenciamento	64

5.16. Avaliação dos níveis de açúcares em colmos de cana-de-açúcar por HPLC.....	65
6. Resultados.....	67
6.1. Prospecção de SAS relacionados a transdução de sinais em cana-de-açúcar.....	67
6.1.1. O Catálogo SUCAST	67
6.1.2. O quinoma de cana-de-açúcar	69
6.2. Análise do perfil de expressão gênica utilizando a tecnologia de microarranjos de cDNA.....	85
6.2.1. O microarranjo SUCAST	85
6.2.2. Análise da variabilidade individual.....	90
6.3. Perfil de expressão em seis órgãos de cana-de-açúcar	90
6.4. Perfil de expressão frente a fitormônios e a estímulos ambientais.....	101
6.5. Perfil de expressão em populações de alto e baixo brix	117
6.6. O Banco de Dados SUCAST	124
7. Discussão	128
7.1. Análise <i>in silico</i> do quinoma de cana-de-açúcar.....	128
7.2. Padronização das condições para análise dos dados de microarranjos.....	132
7.3. Análise de perfis de expressão em seis órgãos de cana-de-açúcar	136
7.4. Análise da expressão gênica em resposta a estímulos ambientais e a fitormônios	144
7.4.1. Visão geral dos genes responsivos a estímulos ambientais e a fitormônios em cana-de-açúcar	144
7.4.2. Componentes do quinoma de cana-de-açúcar responsivos a fitormônios e a estímulos ambientais.....	153
7.5. Análise da expressão gênica em genótipos contrastantes para acúmulo de sacarose.....	156
8. Conclusões & Perspectivas	166
9. Colaboradores.....	169
10. Referências.....	170

Índice das Figuras

Figura 1	70
Figura 2	72
Figura 3	74
Figura 4	78
Figura 5	88
Figura 6	91
Figura 7	93
Figura 8	96
Figura 9	99
Figura 10	100
Figura 11	104
Figura 12	111
Figura 13	118
Figura 14	120
Figura 15	125

Índice das Tabelas

<u>Tabela 1</u>	68
<u>Tabela 2</u>	82
<u>Tabela 3</u>	83
<u>Tabela 4</u>	84
<u>Tabela 5</u>	86
<u>Tabela 6</u>	89
<u>Tabela 7</u>	94
<u>Tabela 8</u>	95
<u>Tabela 9</u>	97
<u>Tabela 10</u>	102
<u>Tabela 11</u>	106
<u>Tabela 12</u>	109
<u>Tabela 13</u>	112
<u>Tabela 14</u>	113
<u>Tabela 15</u>	114
<u>Tabela 16</u>	115
<u>Tabela 17</u>	116
<u>Tabela 18</u>	119
<u>Tabela 19</u>	122
<u>Tabela 20</u>	123
<u>Tabela 21</u>	125

1. Abreviaturas

AB	Alto brix
ABA	Ácido abscísico
ACC oxidase	Ácido 1-aminociclopropano-1-carboxílico oxidase
ACC sintase	Ácido 1-aminociclopropano-1-carboxílico sintase
ADP	Adenosina difosfato
AMPc	Adenosina 3',5'-monofosfato cíclico
AOC	Aleno óxido ciclase
AOS	Aleno óxido sintase
ATP	Adenosina 5'-trifosfato
BB	Baixo brix
BCCCenter	Centro brasileiro de coleção de clones (<i>Brazilian Clone Collection Center</i>)
BLAST	Ferramenta de busca de alinhamento local (<i>Basic Local Alignment Search Tool</i>)
BNF	Fixação biológica de nitrogênio (<i>Biological Nitrogen Fixation</i>)
BRI1	Receptor de brassinosteróides
CCOMT	Cafeoil CoA 3- <i>o</i> -metiltransferase
CCaMK	Proteína quinase dependente de cálcio e calmodulina (<i>Calcium and calmodulin-dependent protein kinase</i>)
cDNA	DNA complementar
CDK	Proteína quinase dependente de ciclina (<i>Cyclin-Dependent Kinase</i>)
CDPK	Proteína quinase dependente de cálcio (<i>Calcium-Dependent Protein Kinase</i>)
CIPK	Proteína quinase que interage com proteína semelhante à calcineurina B (<i>Calcineurin B-like Interacting Protein Kinase</i>)
CLV1	Receptor CLAVATA 1
COMT	Ácido caféico-3- <i>o</i> -metiltransferase
CR4	Receptor CRINKLY 4
Ct	Número de ciclos para atingir o limite (<i>threshold cycle number</i>)
CTC	Centro de Tecnologia Canaveira
CYP	Citocromo P450 monooxigenase
dATP	Desoxiadenosina 5'-trifosfato
dCTP	Desoxicitosina 5'-trifosfato
DEPC	Dietilpirocarbonato
dGTP	Desoxiguanosina 5'-trifosfato
DO	Densidade ótica
dsDNA	DNA dupla fita
DTT	Ditiotreitol
dTTP	Desoxitimidina 5'-trifosfato
EDTA	Ácido etilenodiaminotetraacético
ESTs	Etiquetas de sequências expressas (<i>Expressed Sequence Tags</i>)
GAPDH	Gliceraldeído-3-fosfato desidrogenase
Genes-R	Genes de resistência
GDP	Guanosina difosfato
GMP	Guanosina monofosfato
GPCR	Receptor acoplado à proteína G (<i>G Protein-Coupled Receptor</i>)
GSK3	Glicogênio sintase quinase (<i>Glycogen Synthase Kinase 3</i>)
GTP	Guanosina trifosfato
HMM	Cadeias de Markov (<i>Hidden Markov Model</i>)
HPLC	Cromatografia líquida de alto desempenho (<i>High Performance Liquid Chromatography</i>)
IAA	Ácido indolacético
JA	Ácido jasmônico
LOWESS	Análise de regressão linear localmente ponderada (<i>Locally Weighted Linear Regression Analysis</i>)
LOX	Lipoxigenase
LRR	Repetições ricas em leucina (<i>Leucin-Rich Repeats</i>)
LTI	Proteínas induzidas pelo frio (<i>Low Temperature Induced Proteins</i>)

LTK1	Proteína quinase transmembrânica com LRR 1 (<i>Leucin-Rich Transmembrane Kinase 1</i>)
LysM	Motivo de lisina (<i>Lysin Motif</i>)
MADs	Desvios absolutos da mediana (<i>Median Absolute Deviations</i>)
MAPK	Proteína quinase ativada por mitógeno (<i>Mitogen-Activated Protein Kinase</i>)
MAPKK	Proteína quinase de MAPK
MAPKKK	Proteína quinase de MAPKK
MeJA	Metiljasmonato
MGED	Sociedade de Dados de Expressão Gênica de Microarranjos (<i>Microarray Gene Expression Data Society</i>)
MIAME	Informação mínima a respeito de um experimento de microarranjo (<i>Minimum Information About a Microarray Experiment</i>)
MOPS	Ácido 3-[N-morfolino] propanossulfônico
mRNA	RNA mensageiro
MTM	Média filtrada com base na mediana (<i>Median-Based Trimmed Mean</i>)
NADPH	Nicotinamida adenina dinucleotídeo fosfato (forma reduzida)
NBS (ou NB-ARC ou NB)	Domínio de ligação de nucleotídeo (<i>Nucleotide Binding Site</i>)
NJ	Agrupamento de vizinhos (<i>neighbour-joining</i>)
OPDA	Ácido 12-oxofitodienólico
ORF	Fase de leitura aberta (<i>Open Reading Frame</i>)
PCA	Análise de Componente Principal (<i>Principal Component Analysis</i>)
PCR	Reação em cadeia da polimerase (<i>Polymerase Chain Reaction</i>)
Pfam	Banco de dados de alinhamentos e HMMs de famílias de proteínas (<i>Protein Families Database of Alignments and HMMs</i>)
Pi	Ortofosfato
PIP	Proteínas intrínseca da membrana plasmática (<i>Plasma membran Intrinsic Protein</i>)
PKC	Proteína quinase C
<i>Pkinase</i>	Domínio de proteína quinase (<i>protein kinase domain</i>)
PPCK	Fosfoenolpiruvato carboxilase quinase (<i>Phosphoenolpyruvate Carboxylase Kinase</i>)
PR	Proteína relacionada à patogênese (<i>Pathogenesis-Related protein</i>)
Projeto SUCAST	Projeto de Transdução de Sinal da Cana-de-açúcar (<i>Sugarcane Signal Transduction Project</i>)
Projeto SUCEST	Projeto de ESTs da Cana-de-açúcar (<i>The Sugar Cane EST Project</i>)
PSK	Fitossulfoquina
RLCK	Proteína quinase citoplasmática similar a receptor (<i>Receptor-Like Cytoplasmic Kinase</i>)
RLK	Proteína quinase similar a receptor (<i>Receptor-Like Kinase</i>)
ROS	Espécie reativa de oxigênio (<i>Reactive Oxygen Specie</i>)
rpm	Rotações por minuto
rRNA	RNA ribossomal
Rubisco	Ribulose-1,5-bisfosfato carboxilase
SA	Ácido salicílico
SAS	Grupo(s) de sequências agrupadas de cana-de-açúcar (<i>Sugarcane Assembled Sequences</i>)
SDS	Dodecil sulfato de sódio (<i>Sodium dodecyl sulfate</i>)
Ser	Serina
SERK	Receptor quinase de embriogênese somática (<i>Somatic Embryogenesis Receptor Kinase</i>)
SMART	Ferramenta simples de buca de arquitetura modular (<i>Simple Modular Architecture Research Tool</i>)
SNF1	Não fermentador de sacarose (<i>sucrose non-fermenting</i>)
SnRK	Proteína quinase semelhante à SNF1 de levedura (<i>SNF1-related kinase</i>)
SOM	Mapas auto-organizáveis (<i>Self-Organizing Maps</i>)
SSC	Tampão de cloreto de sódio e citrato de sódio
TAE	Tampão Tris-acetato-EDTA
TIP	Proteínas intrínsecas do tonoplasto (<i>Tonoplast Intrinsic Proteins</i>)
TPR	Repetições tetratricopeptídicas (<i>Tetratricopeptide Repeats</i>)
Thr	Treonina

Tyr	Tirosina
UDP	Uridina difosfato
UPGMA	Método de agrupamento de pares não ponderado com média aritmética (<i>Unweighted Pair Group Method with Arithmetic mean</i>)
UPK	Proteína quinase não-classificada (<i>Unclassified Protein Kinase</i>)
URLK	RLK não-classificado (<i>Unclassified RLK</i>)
U	Unidades
UV	Ultravioleta
VOC	Composto orgânico volátil (<i>Volatile Organic Compound</i>)
WAK	Proteína quinase associada à parede celular (<i>Wall-Associated Kinase</i>)

2. Resumos

2.1. Resumo

Diversos processos envolvidos no crescimento, desenvolvimento e adaptação a variações ambientais são regulados por vias de transdução de sinal. O projeto SUCAST (*Sugarcane Signal Transduction Project*) tem como objetivos a identificação e caracterização funcional dos componentes de transdução de sinais da cana-de-açúcar (Souza *et al.*, 2001). O presente trabalho insere-se dentro do Projeto SUCAST e teve como foco a identificação de componentes de transdução de sinais, a categorização desses elementos utilizando-se ferramentas de bioinformática e a avaliação de seus perfis de expressão pela tecnologia de microarranjos de cDNA. Genes relacionados à transdução de sinais foram buscados entre as seqüências armazenadas no banco de dados SUCEST - *The Sugar Cane EST Project* (Vettore *et al.*, 2003) - pela ferramenta de BLAST, o que permitiu a categorização de mais de 3.500 genes relacionados a vias de sinalização em cana-de-açúcar. A categoria de proteínas quinases foi também analisada quanto à presença de domínios conservados e classificada por agrupamento filogenético do domínio catalítico predito. Com a análise filogenética, foram obtidos seis grupos característicos para proteínas quinases e quatro grupos principais de receptores do tipo Ser/Thr quinases. Um total de 527 genes do quinoma de cana foi analisado.

Os padrões de expressão de componentes do catálogo SUCAST foram avaliados em seis órgãos (raiz, flor, folha, gema lateral, primeiro e quarto entrenós), em resposta a diferentes estímulos (tratamentos com os fitormônios ácido abscísico e metiljasmonato, seca, interação com bactérias diazotróficas endofíticas, ataque por *Diatraea saccharalis* e deficiência em fosfato) e em populações contrastantes para acúmulo de sacarose.

As análises dos dados de microarranjos mostraram 217 genes diferencialmente expressos em pelo menos um dos órgãos analisados e 153 genes considerados de expressão ubíqua. Para os experimentos de respostas a tratamentos com fitormônios e estímulos ambientais, 179 genes diferencialmente expressos em pelo menos uma das condições analisadas foram identificados.

Cinquenta e um genes mostraram-se diferencialmente expressos ao se comparar amostras (folhas +1 e entrenós em diferentes fases de maturação) de duas populações contrastantes para acúmulo de sacarose. Os perfis de expressão foram também analisados ao longo do tempo para os experimentos de deficiência em fósforo, seca e tratamentos com fitormônios, através de agrupamentos por *Self-Organizing Maps* (SOM).

Resultados de PCR em tempo real confirmaram os dados de expressão para 72% de 25 genes selecionados para comparações entre diferentes órgãos, e 80,5% de 36 perfis de expressão selecionados para experimentos de resposta a tratamentos com hormônios e a estímulos ambientais. Adicionalmente, reações de PCR em tempo real foram realizadas para se avaliar os níveis de expressão de cinco genes selecionados em indivíduos contrastantes para acúmulo de sacarose e 57% dos resultados obtidos mostraram-se de acordo com os dados de microarranjos.

Todos os dados obtidos foram integrados e compilados no banco de dados SUCAST (<http://www.sucest-fun.org>). O conhecimento gerado com o projeto representa um progresso significativo na compreensão da função de alguns dos genes identificados no projeto SUCEST, indicando alvos a serem explorados no programa de melhoramento da cana-de-açúcar.

2.2. Abstract

A diversity of processes related to growth, development and adaptation to environmental conditions are regulated by signal transduction pathways. The SUCAST Project (*Sugarcane Signal Transduction*) (Souza *et al.*, 2001) aims to identify, characterize and associate putative functions to sugarcane signal transduction components using bioinformatic tools and to evaluate their expression profiles using cDNA microarrays. BLAST searches conducted on sequences stored in the SUCEST databank - *The Sugar Cane EST Project* (Vettore *et al.*, 2003) - revealed more than 3,500 putative genes related to signaling in sugarcane. The protein kinases were analyzed for the presence of conserved domains and classified according to a phylogenetic analysis of the predicted catalytic domain. The phylogenetic approach indicated six characteristic groups for protein kinases and four main groups for receptor-like kinases.

The expression patterns of SUCAST components were evaluated in six organs (root, flower, leaf, lateral bud, first and fourth internodes), in response to different stimuli (treatment with the phytohormones abscisic acid and methyljasmonate, drought, endophytic bacteria interaction, attack by *Diatraea saccharalis* and phosphate deficiency) and in sugarcane cultivars contrasting for sucrose accumulation.

The tissue profiling experiment showed 217 differentially expressed genes in at least one of the six organs analyzed and 153 genes of ubiquitous expression. A total of 179 differentially expressed genes were obtained in response to phytohormones and environmental stimuli, in at least one of the conditions analyzed were obtained. Comparisons between high and low sucrose (brix) cultivars led to the detection of 51 differentially expressed genes in at least one of the samples analyzed (leaves +1 and internodes at different maturing stages). The expression patterns were also analyzed in time-course experiments of phosphate deficiency, drought and phytohormone treatments by *Self-Organizing Maps* (SOM) clusterization.

Quantitative PCR results confirmed 72% of the microarray expression data in sugarcane organs for 25 selected genes and 80.5% of the 36 expression profiles selected in response to phytohormones and environmental stimuli. Additionally, real-time PCR reactions were carried out to

evaluate the expression levels of five genes in individuals contrasting for sucrose accumulation and 57% of the results obtained were in agreement with the microarray data.

All data generated were integrated and compiled into the SUCAST databank (<http://www.sucest-fun.org>). The knowledge generated adds to the comprehension of the function of SUCAST components, indicating targets to be explored in the improvement of sugarcane cultivars.

3. Introdução

3.1. A cana-de-açúcar e os projetos SUCEST e SUCAST

A cana-de-açúcar é uma monocotiledônea C4 originada do cruzamento de diversas espécies do gênero *Saccharum*, nativo da Ásia. Os cultivares comerciais disponíveis atualmente são híbridos provenientes de cruzamentos entre *S. officinarum*, *S. barberi*, *S. robustum*, *S. spontaneum* e *S. edule*, bem como de gêneros relacionados que são capazes de cruzar com *Saccharum*, como *Erianthus*, *Miscanthus*, *Narenga* e *Sclerostachya*. Dentro do gênero *Saccharum*, os genótipos de *S. spontaneum* são os de distribuição geográfica mais abrangente. A origem de *S. spontaneum* não é clara, mas acredita-se que essa espécie tenha se originado da introgressão de *Miscanthus*, *Erianthus* e *Sclerostachya* (Roach & Daniels, 1987). *S. officinarum* originou-se a partir de *S. robustum* por seleção natural e/ou humana em Nova Guiné e caracteriza-se pela elevada capacidade de acúmulo de sacarose. *S. spontaneum* e *S. officinarum* são as espécies mais largamente empregadas em programas de melhoramento genético clássico (Perez *et al.*, 1997). Os cultivares modernos de cana-de-açúcar possuem um genoma extremamente complexo, com elevado nível de aneuploidia e poliploidia e contêm de 100 a 130 cromossomos. Apenas 15 a 25% dos seus cromossomos são herdados de *S. spontaneum* (Grivet & Arruda, 2001).

Devido à sua capacidade única de acúmulo de sacarose nos colmos, com concentrações de até 500 mM no meio apoplástico de entrenós maduros (Moore, 1995), a cana-de-açúcar é uma das plantas de maior valor econômico no mundo. Aproximadamente 60% da produção mundial de açúcar deriva da cana-de-açúcar, além de constituir também importante matéria-prima para a produção de etanol (Grivet & Arruda, 2001). Cada tonelada de cana-de-açúcar tem o potencial energético equivalente ao de 1,2 barril de petróleo (<http://www.portalunica.com.br/>). A agroindústria da cana-de-açúcar do Brasil é um setor muito ativo e de grande relevância econômica, social, estratégica e ambiental. Econômica, porque é uma atividade com grande participação no Produto Interno Bruto (PIB). O Brasil é o maior produtor mundial de cana-de-açúcar, sendo que o Estado de São Paulo contribui com cerca de 60% dessa produção (<http://www.portalunica.com.br/>). Social,

porque é uma das atividades que mais gera emprego na área rural, além de empregos indiretos através de uma variada cadeia de produtos e serviços nas cidades. Estratégica, porque o setor energético se vale da cana para reduzir a vulnerabilidade do país devido à instabilidade do mercado de petróleo no mundo e também porque contribui para melhorar a balança de pagamentos externos do país. E, finalmente, ambiental, porque o álcool combustível polui menos do que a gasolina, e, conseqüentemente, sua adição à gasolina contribui para diminuir a poluição. A cana-de-açúcar pode contribuir em um futuro próximo com diversos outros benefícios, tais como a co-geração de eletricidade, produção de açúcares não calóricos, plásticos biodegradáveis, e de outros produtos de grande valor agregado que utilizem metabólitos secundários. Em vista disto, a busca por variedades de cana-de-açúcar melhoradas, as quais, por exemplo, acumulem mais açúcar, sejam capazes de ser cultivadas em climas secos ou sejam menos suscetíveis a pragas agrícolas, é extremamente desejável. O melhoramento clássico é um processo extremamente lento e acaba sendo incompatível com as necessidades atuais do mercado altamente competitivo. Por outro lado, há a possibilidade do melhoramento pela alteração direta de genes de interesse, que pode levar à obtenção de variedades melhoradas muito mais rapidamente e com especificidade para o objetivo desejado. É imprescindível, para isso, que sejam disponibilizados conhecimentos sólidos da biologia do sistema a ser trabalhado, de forma que os alvos (genes) a serem manipulados possam ser determinados com precisão. Os genes relacionados a componentes de transdução de sinais são fortes candidatos a serem alvos para o melhoramento genético da cana-de-açúcar, pois podem permitir que determinadas vias de interesse sejam ativadas ou inibidas, de acordo com o objetivo do melhoramento em questão.

O projeto SUCAST foi iniciado com o objetivo de identificar, catalogar e caracterizar as ESTs (*Expressed Sequence Tags*) da cana-de-açúcar que estão envolvidas em vias de transdução de sinal, bem como a organização de todas essas informações em um banco de dados que facilite as análises funcionais (Souza *et al.*, 2001). Essas ESTs foram selecionadas a partir das sequências geradas pelo projeto SUCEST (*The Sugar Cane EST Project*, Vettore *et al.*, 2003). O Projeto SUCEST sequenciou 237.954 ESTs a partir das extremidades 5' e 3' de clones de 26 bibliotecas de cDNA de cana-de-açúcar, preparadas a partir de 11 órgãos diferentes e de plantas

submetidas a estresse térmico e à interação com bactérias fixadoras de N₂. Estima-se que essa ampla amostragem de órgãos possibilitou o sequenciamento de mais de 90% do genoma expresso de cana-de-açúcar (Vettore *et al.*, 2003). As ESTs obtidas foram agrupadas em 43.141 SAS (*Sugarcane Assembled Sequences*) (Vettore *et al.*, 2001, Vettore *et al.*, 2003). Para efeito dos trabalhos aqui relatados, cada SAS identificado pelo projeto SUCEST será tratado como um gene de cana-de-açúcar, embora se saiba que o agrupamento de ESTs em SAS pode ser sujeito a erros.

Dentre os mais de 43.000 SAS determinados estão genes envolvidos na resistência a pragas, estresses abióticos, metabolismo de açúcares, absorção de nutrientes, regulação do ciclo celular, assimilação de nitrogênio e tolerância ao alumínio. Adicionalmente, aproximadamente 30% das ESTs não têm nenhum correspondente em bases de dados públicas, logo, compreendem um excelente recurso para descoberta de novos genes (Vettore *et al.*, 2003).

As etapas do projeto SUCAST compreendem, de forma sucinta, a prospecção, a anotação e a caracterização funcional de componentes de transdução de sinais utilizando-se ferramentas de bioinformática e a análise de expressão gênica através da tecnologia de microarranjos de cDNA. No presente projeto, as análises conduzidas *in silico* focaram a categoria de quinases de proteínas, que inclui a maior parte dos receptores em plantas e também alguns dos genes de resistência (genes-R), além de outras proteínas quinases. O foco nessa categoria específica justifica-se pelo fato dos receptores e genes-R representarem os primeiros componentes de cascatas de transdução de sinais, agindo como “sensores” para estímulos (tanto endógenos quanto exógenos) e permitindo o desencadeamento de uma resposta pela célula frente a esse estímulo. Além disso, as proteínas quinases compreendem uma das categorias mais numerosas dentro da transdução de sinais e podem constituir pontos regulatórios chaves para o controle de muitas vias.

3.2. Transdução de sinais em plantas

Todas as células vivas utilizam vias de transdução de sinais para responder a estímulos endógenos e exógenos. A partir da detecção de um estímulo, cascatas de transdução de sinais

são ativadas, permitindo que o sinal detectado seja amplificado, integrado com sinais provenientes de outros estímulos e que respostas específicas sejam geradas. Diversos sinalizadores são comuns na transdução de sinal de vias diferentes, como é o caso de Ca^{2+} , fosfolípidos de inositol, proteínas-G, nucleotídeos cíclicos, proteínas quinases e proteínas fosfatases. Geralmente, as respostas a um único estímulo dependem da ativação de não apenas uma, mas de várias vias de transdução de sinais que interagem entre si ou mesmo com vias disparadas por outros estímulos. A interação de diversas vias de sinalização entre si permite que, utilizando um arsenal relativamente pequeno de mensageiros, as células sejam capazes de responder a diversos estímulos apropriadamente.

Dentre os mensageiros secundários encontrados em plantas, os íons Ca^{2+} são de particular importância, pois participam nas vias de sinalização a diversos estímulos diferentes, tais como respostas a hormônios, à seca, à salinidade, ao estresse térmico, à luz e a patógenos (Clark *et al.*, 2001). De fato, nenhum mensageiro, até o momento, se mostrou mais responsivo a uma diversidade de estímulos do que Ca^{2+} citoplasmático (Sanders *et al.*, 1999). A principal questão a respeito da sinalização por Ca^{2+} é como os componentes de vias de sinalização celular conseguem discriminar oscilações nos níveis de Ca^{2+} disparadas em consequência da detecção de estímulos tão diferentes entre si. Sabe-se que essa discriminação dos fluxos de Ca^{2+} depende da magnitude, duração, oscilações de fluxos e localização subcelular das alterações nas concentrações desse íon (Sanders *et al.*, 1999). Os íons Ca^{2+} são detectados por proteínas sensoras específicas, dentre as quais podem ser citadas as proteínas quinases dependentes de Ca^{2+} (CDPKs), a proteína quinase C (PKC), as proteínas quinases dependentes de calmodulina, a proteína fosfatase 2B calcineurina, a calmodulina e as anexinas (Sanders *et al.*, 1999, Clark *et al.*, 2001, Sanders *et al.*, 2002).

Em mamíferos, nucleotídeos cíclicos, principalmente o AMPc, mediam as respostas a diversos hormônios e neurotransmissores e são capazes de regular ao menos uma enzima em cada via metabólica conhecida. Apesar de inicialmente ter se acreditado que vias mediadas por nucleotídeos cíclicos não estavam presentes em plantas, atualmente já existem evidências da presença de nucleotídeos cíclicos, nucleotidil ciclases, fosfodiesterases de nucleotídeos cíclicos e

quinases com domínio de ligação a nucleotídeos cíclicos nesses organismos (Newton & Smith, 2004).

Os receptores de plantas constituem um vasto grupo e são tipicamente classificados em receptores histidino-quinase, receptores acoplados à proteína G (GPCRs), fotorreceptores, receptores com repetições ricas em leucina (LRRs) de resistência contra determinados patógenos e receptores Ser/Thr quinases (McCarty & Chory, 2000).

Em *Arabidopsis*, 11 genes codificam para histidino-quinases do tipo duplo-componente. O sistema de duplo componente é constituído por uma histidino-quinase e pelo regulador de resposta, com o domínio de recepção, que possui um resíduo característico de aspartato no N-terminal, que é fosforilado. A análise genética e bioquímica indica que essas proteínas funcionam como receptores de etileno, de citocininas e também como osmosensores em plantas (Urao *et al.*, 2000, Urao *et al.*, 2001).

Receptores do tipo GPCRs, de sete domínios transmembrânicos, constituem uma numerosa família gênica em mamíferos. Esses receptores interagem com proteínas G heterotriméricas, associadas à face citossólica da membrana. As proteínas G heterotriméricas apresentam-se na forma inativada quando ligadas com GDP. Ao interagir com um ligante, o receptor do tipo GPCR altera sua conformação e atua sobre a proteína G associada, levando à substituição de GDP por GTP. A proteína G, assim ativada, dissocia-se em subunidade α , carregando o GTP, e em um dímero $\beta\gamma$. Essas subunidades são, então, capazes de atuar sobre proteínas, que, ao serem estimuladas, levam à produção de mensageiros secundários. A subunidade α possui atividade GTPásica intrínseca, o que leva à substituição de GTP por GDP e à reconstituição do heterotrímero no estado inativado.

Poucos exemplos de GPCRs já foram descritos em plantas. Buscas no genoma já seqüenciado de *Arabidopsis thaliana* apontam para a existência de apenas um GPCR nesse organismo, denominado GCR1, com aproximadamente 20% de identidade de aminoácido com o receptor CAR1 de AMPc de *Dictyostelium discoideum* (Josefsson & Rask, 1997). Apesar de nenhum ligante ter sido identificado para GCR1, estudos apontam para o envolvimento desse receptor em vias de sinalização do fitormônio ácido abscísico (ABA) e no processo de germinação

de sementes regulado pelos hormônios giberelina e brassinosteróides (Chen *et al.*, 2004, Pandey *et al.*, 2006).

Genes para as subunidades α , β e γ de proteínas G em plantas já foram identificados. Tanto em *Arabidopsis thaliana* quanto em arroz, foram identificadas um gene para $G\alpha$, um para $G\beta$ e dois para $G\gamma$, além de um gene para regulador de sinalização por proteína G, o gene de *Arabidopsis thaliana* RGS1 (Perfus-Barbeoch *et al.*, 2004). Buscas no banco de dados SUCEST mostram a existência de três genes para a subunidade α , 12 para a subunidade β e 1 para a subunidade γ , revelando um número relativamente alto desses elementos em cana-de-açúcar (Souza *et al.*, 2001).

As plantas detectam a luz através de múltiplos fotorreceptores, que incluem os fitocromos e criptocromos. Os criptocromos são flavoproteínas que participam de muitas respostas à luz azul, incluindo o fototropismo e o ritmo circadiano. Os fitocromos funcionam como Ser/Thr quinases e potencialmente interagem com os criptocromos no seu processo de sinalização em resposta à luz vermelha (Ahmad, 1999).

A maior parte dos receptores encontrados em plantas está dentro da categoria dos receptores Ser/Thr quinases, também denominados RLKs (*Receptor-Like Kinases*). Cinco grupos principais de RLKs foram descritos em plantas: RLKs com LRRs, que são característicos de alguns receptores já bem descritos como Clavata 1 (CLV1), receptor de brassinosteróide (BRI1) e ERECTA; RLKs do tipo CR4 (CRINKLY4), os quais contêm domínios extracelulares similares ao do tipo TNFR (domínios encontrados em receptores do fator de necrose tumoral, do inglês, *tumor necrosis factor receptors*); RLKs com domínio-S, que contêm domínios extracelulares similares ao domínio extracelular de glicoproteínas do *locus*-S; RLKs com domínio semelhante ao de fator de crescimento epidérmico (EGF-like, de *epidermal growth factor-like*, também denominado de tipo WAK, *wall-associated kinase*) e RLKs com domínio de lectina (McCarty & Chory, 2000). Os RLKs são definidos pela presença de uma sequência sinal, região transmembrânica no domínio N-terminal e um domínio catalítico na porção carboxi-terminal da molécula (Walker, 1994, Torii, 2000).

Em *Arabidopsis*, os RLKs representam uma das maiores famílias gênicas, com 400 membros, representando 2,5% dos genes codificadores para proteínas e 60% de todas as proteínas quinases encontradas em *Arabidopsis* (Shiu & Bleecker, 2003). Desses, menos de 2% possuem funções conhecidas e ainda menos é conhecido sobre seus componentes de sinalização ou ligantes (Shiu & Bleecker, 2001a). As famílias dos RLKs podem ser definidas com base nas relações filogenéticas entre os domínios catalíticos de seus membros (Shiu & Bleecker, 2001b) e são posteriormente subclassificadas, com base na presença ou não de domínio extracelular e região transmembrânica, em RLKs e RLCKs (*Receptor-like Cytoplasmic Kinases*), essas últimas representando os membros citoplasmáticos dessa família. Ainda não se sabe muito a respeito das RLCKs; algumas delas podem ser simplesmente formas ancestrais de proteínas quinases (Shiu & Bleecker, 2001b). Para algumas delas, entretanto, uma função já foi identificada, como para a proteína NAK de *Arabidopsis thaliana*, que parece estar relacionada com o crescimento e desenvolvimento das plantas (Moran & Walker, 1993).

A família de RLKs em arroz foi analisada e comparada com as sequências conhecidas de RLKs em *Arabidopsis* (Shiu *et al.*, 2004). Essa análise mostrou um maior número (aproximadamente o dobro) de RLKs em arroz em relação ao número encontrado em *Arabidopsis thaliana*, sendo que o maior número de eventos de duplicação ocorreu para RLKs envolvidos com estresse e defesa.

Receptores do tipo proteína quinase têm sua atividade catalítica ativada quando um ligante interage com o seu domínio extracelular. Com essa ativação, ocorre a fosforilação do domínio citoplasmático via autofosforilação, permitindo que o receptor interaja com uma proteína-alvo e a ative. Entretanto, como já dito acima, poucos RLKs de plantas já tiveram os seus ligantes identificados e os mecanismos pelos quais esses receptores têm a sua atividade regulada ainda não são conhecidos, com exceção de alguns poucos RLKs. Em um trabalho recente, Johnson & Ingram (2005) discutem diversos mecanismos pelos quais os RLKs de plantas podem ser ativados e regulados. Esses mecanismos incluem hetero-oligomerização quando da interação RLK-ligante (como já demonstrado para o receptor de brassinosteróides, BRI1, que forma oligômeros com a proteína BAK1), a probabilidade dos diferentes RLKs dimerizarem em diferentes combinações,

bem como a inibição da sinalização citoplasmática ativada por RLKs através da ação de proteínas fosfatases, de inibidores de fosforilação, por degradação mediada por ubiquitinação e via endocitose dos RLKs. Em face do que já foi estudado para os mecanismos de sinalização por RLKs até o momento, esses autores concluem que, provavelmente, não exista um mecanismo único de regulação dos RLKs, e sim diversos mecanismos que diferem de receptor para receptor.

Além das RLCKs e dos RLKs, diversas outras subfamílias de proteínas quinases em plantas foram identificadas: CDPKs, proteínas quinases relacionadas a SNF1 (SnRKs), MAPK, MAPKK, MAPKKK, quinases dependentes de ciclinas (CDKs), caseína quinase I, caseína quinase II, quinases semelhantes a GSK3/shaggy e outras (Hardie, 1999). As proteínas quinases participam de um grande número de vias de transdução de sinal em células eucarióticas, incluindo a resposta a mensageiros extracelulares e a estímulos ambientais de estresse, defesa contra patógenos e controle do ciclo celular. Trabalhos recentes descreveram o quinoma de levedura, *Drosophila*, *C. elegans* e o humano (Manning *et al.*, 2002a, Manning *et al.*, 2002b). Com relação às proteínas quinases de plantas, é interessante notar que, apesar de seus papéis críticos em transdução de sinal, muitas delas ainda permanecem pouco caracterizadas. Há diversos tipos de proteínas quinases, mas todas possuem em comum a capacidade de adicionar um grupo fosfato, suprido pela hidrólise de ATP em ADP, a resíduos de aminoácidos específicos de uma proteína-alvo. As proteínas quinases possuem um domínio catalítico, constituído por 250 a 300 aminoácidos organizados em 12 subdomínios já bem caracterizados com relação aos resíduos importantes para a atividade catalítica (Hanks *et al.*, 1988, Hanks & Quinn, 1991, Hanks & Hunter, 1995). As proteínas quinases podem ser classificadas, de acordo com os resíduos de aminoácidos que são fosforilados na proteína-alvo, em Tyr quinases, Ser/Thr quinases e quinases de especificidade dupla, essas últimas com capacidade de fosforilar tanto resíduos de tirosina quanto resíduos de Ser/Thr (Lewin, 2001). Ao contrário do que ocorre em células de mamíferos, Tyr quinases são raras em células vegetais e os RLKs de plantas possuem atividade de Ser/Thr quinase (Trewavas, 2000).

A compreensão das vias de transdução de sinal ativadas durante o processo de patogênese em plantas é de extrema importância, visto que esse conhecimento pode permitir a

geração, através da engenharia genética, de cultivares resistentes a doenças, o que pode levar ao controle de pragas, aumento da produtividade agrícola e diminuição do uso de pesticidas. A indução dos mecanismos de defesa é possível pelo reconhecimento do patógeno pela planta (hospedeiro), seguido pela transdução de sinal por vias relacionadas a patogênese. Frequentemente ocorre a resposta de hipersensibilidade (HR, de *Hypersensitive Response*) que consiste em morte celular induzida no sítio de infecção pelo patógeno e é responsável por limitar a disseminação da infecção pela planta. Os produtos de genes de resistência (genes-R) são receptores para ligantes específicos, codificados pelos genes de avirulência do patógeno (*Avr*), e esse reconhecimento desencadeia a resposta de hipersensibilidade do hospedeiro. Dentre as categorias de genes-R, os que codificam para proteínas com domínio do tipo NBS-LRR correspondem à categoria mais numerosa. Entretanto, alguns genes-R codificam para proteínas quinases, como é o caso de *Xa21* e *Rpg1* (Song *et al.*, 1995, Brueggeman *et al.*, 2002) e há também genes-R que codificam para glicoproteínas ancoradas na membrana com LRRs extracelulares, como é o caso dos genes-R do locus *Cf* de tomate (Thomas *et al.*, 1998).

3.3. A tecnologia de microarranjos de cDNA

A tecnologia de arranjos de DNA, aplicada à análise da expressão gênica em larga escala, permite a exploração dos dados do sequenciamento de genomas e a associação de possíveis funções aos genes identificados. Os arranjos de DNA compreendem suportes sólidos (de náilon, no caso de macroarranjos e de vidro no caso de microarranjos) onde sequências completas ou parciais de genes, aqui denominadas como sondas, são fixadas, permitindo o estudo da expressão de centenas a milhares de genes em paralelo e em diferentes condições (diferentes tratamentos, órgãos ou intervalos de tempo, por exemplo).

Em um experimento de microarranjo, ocorre a hibridização de alvos fluorescentes provenientes de duas amostras sobre os arranjos de DNA, que podem ser constituídos por produtos de PCR purificados ou sequências de oligonucleotídeos, por exemplo. Para a obtenção desses alvos, parte-se de RNA total ou RNA mensageiro, que serve de molde para síntese de

cDNA marcado com o fluoróforo (Cy3 para uma das amostras e Cy5 para a outra, por exemplo). Ambos os alvos são hibridizados competitivamente em uma mesma lâmina. Pelo fato do DNA no suporte estar em excesso, a quantidade de alvo hibridizado ao DNA serve como medida da abundância do transcrito em uma população de mRNA. Dessa maneira, pode-se determinar o padrão de expressão de um determinado gene ou EST, comparando-se a intensidade do seu sinal nos diversos tratamentos.

A imagem gerada em experimentos de microarranjos é analisada por programas próprios para a extração de seus dados, em que a intensidade da fluorescência emitida pelos alvos marcados hibridizados em cada amostra de DNA depositado pode ser quantificada e subtraída da intensidade de fundo. Posteriormente, esses valores são normalizados e utilizados para o cálculo da razão de expressão de cada gene entre as duas amostras (condições) comparadas entre si no experimento.

Os dados extraídos devem ser inicialmente normalizados com o objetivo de minimizar variações sistemáticas nos níveis de expressão gênica medidos entre as duas amostras de mRNA co-hibridizadas, bem como permitir a comparação dos níveis de expressão entre diferentes lâminas. Como exemplo de variações sistemáticas que ocorrem em um experimento de microarranjo podem ser citadas as diferenças nos níveis de fluorescência entre os fluoróforos Cy3 e Cy5 (devido às propriedades intrínsecas de cada um dos fluoróforos), diferenças na eficiência de incorporação dos fluoróforos, variações intensidade-dependentes dos valores de razão, variações locais (geradas por inconsistências entre as agulhas utilizadas para o processo de deposição de DNA nas lâminas, por variabilidade nas superfícies das lâminas ou mesmo diferenças locais no processo de hibridização) e os parâmetros adotados no *scanner* para se obter as imagens (Yang *et al.*, 2002a, Uchida *et al.*, 2005). Todos esses fatores levam a diferenças entre as intensidades para o canal de Cy3 e de Cy5 que não refletem expressão diferencial. Há diferentes formas de se proceder à normalização dos dados, como revisado por Quackenbush (2001 e 2002). De uma forma geral, os diferentes métodos de normalização sempre assumem que a maior parte dos elementos do microarranjo possui em média uma razão de expressão entre as duas amostras igual a um, ou seja, não apresenta expressão diferencial. Com isso, calcula-se um fator de

normalização, usado pra ajustar os dados, compensando o efeito da variabilidade experimental e equilibrando a intensidade dos sinais de fluorescência entre os dois canais. Um dos métodos mais simples de normalização dos dados é pela intensidade total. Esse método é deficiente por não levar em consideração muitos dos erros sistemáticos que ocorrem com os dados. Já outro tipo de normalização é o emprego do *Dye-Swap* (inversão dos fluoróforos), onde alíquotas do mesmo RNA são marcadas separadamente com Cy3 e Cy5 (para ambas as amostras a serem hibridizadas entre si) e as hibridizações são feitas de modo a permitir que em um experimento dada amostra esteja marcada com Cy3 e, na réplica, essa amostra esteja marcada com Cy5. Todavia, esse modo de normalização implica que para cada combinação de amostras de RNA a serem hibridizadas entre si, duas hibridizações devem ser feitas. Com isso, para um grande número de experimentos planejados, o número de hibridizações a serem realizadas aumenta muito, aumentando o custo do projeto. Outras alternativas de normalização seriam, por exemplo, o emprego do método de LOWESS ou a normalização com base nos dados de controles internos, como é o caso do sistema *Scorecard* (GE Healthcare). Esse sistema compreende amostras de DNA sintético que permitem a obtenção de controles de calibração, controles da razão entre sinal de Cy3 e Cy5 e controles negativos (sem DNA). O método de LOWESS é um método de normalização não-linear e baseia-se no fato da variabilidade dos valores de razão para determinado gene ser dependente dos seus respectivos valores de intensidade. O algoritmo de LOWESS é, então, capaz de remover os efeitos dependentes da intensidade do sinal (Yang *et al.*, 2002b).

Os padrões de expressão obtidos para diferentes genes em diferentes amostras podem ser comparados pela escolha de um algoritmo de agrupamento desses dados. Dentre os diferentes algoritmos de agrupamento disponíveis, pode-se citar o agrupamento hierárquico, agrupamento por *k-means* e *Self-Organizing Maps* (SOM), dentre outros (Quackenbush, 2001).

Outro aspecto de importância, dado o montante de informação gerada nos experimentos de microarranjos, é a disponibilidade de um banco de dados, no qual todos os resultados obtidos possam ser depositados e compartilhados por todos os participantes do projeto.

A tecnologia de microarranjos pode potencialmente indicar genes co-regulados que participam de uma mesma via de transdução de sinal e, por isso, a aplicação dessa tecnologia pode ser extremamente poderosa para evidenciar novas vias ou mesmo a relação de diferentes vias entre si. Experimentos de microarranjos são comumente usados para comparação de classes, com o objetivo de se identificar genes diferencialmente expressos ao se comparar diferentes tipos de amostras, ou para predição de classes (Golub *et al.*, 1999). O perfil de expressão gênica pode ser indicativo da função de novos genes e o emprego de microarranjos permite fazer essa análise para milhares de genes ao mesmo tempo.

Diversos projetos centrados na tecnologia de microarranjos já foram ou estão sendo desenvolvidos para plantas. Exemplos de trabalhos com microarranjos para plantas compreendem, por exemplo, a análise da expressão gênica frente a estresse salino (Kawasaki *et al.*, 2001), frio e seca (Seki *et al.*, 2001), mecanismos de defesa em plantas (Schenk *et al.*, 2000), tratamento com nitrato (Wang *et al.*, 2000) e o perfil de expressão de diferentes órgãos (Ruan *et al.*, 1998).

Os resultados obtidos com o uso de microarranjos são influenciados por diversos fatores, como o processo de obtenção das lâminas, a qualidade do RNA utilizado, a eficiência na marcação dos alvos, as condições da hibridização e o protocolo de extração, normalização e análise dos dados. Além disso, cada laboratório segue um protocolo próprio para a condução dos experimentos de microarranjos, bem como para análise dos dados gerados. A Sociedade de Dados de Expressão Gênica de Microarranjos (*MGED Society*) foi criada com os objetivos de estabelecer os padrões para anotação e representação de dados de microarranjos, facilitando a criação de bancos de dados e o intercâmbio de informação de dados de microarranjos entre laboratórios. Para isso, a MGED determinou o denominado protocolo MIAME (*Minimum Information About a Microarray Experiment*), que corresponde à mínima informação a respeito de um experimento de microarranjo necessária para que os dados gerados possam ser interpretados e reproduzidos, bem como à padronização da ontologia para a anotação de experimentos de microarranjos (Brazma *et al.*, 2001).

Em face das numerosas fontes de variação inerentes à técnica de microarranjos, é interessante que os resultados obtidos por essa técnica sejam validados. Uma das técnicas

utilizada para essa validação é a técnica de *Northern Blot*. Entretanto, o uso de *Northern Blot* tem algumas desvantagens, como o uso de material radioativo e a dificuldade de realização dessa metodologia para um número maior de genes. Além disso, essa técnica baseia-se no mesmo princípio dos microarranjos de DNA, que é o pareamento específico de sondas em determinada sequência-alvo. Uma alternativa ao *Northern Blot* como forma de validação dos dados de microarranjos é a utilização da técnica de PCR em tempo real. O PCR em tempo real baseia-se na detecção do dsDNA sintetizado à medida que ocorre uma reação de PCR. Os princípios usados na detecção de PCR em tempo real são baseados na detecção de fluorescência, que é proporcional à quantidade de produto de PCR presente no meio reacional. O modo mais simples e mais barato de detecção para PCR em tempo real é pelo uso de um fluoróforo capaz de intercalar em dsDNA. Apesar da facilidade dessa tecnologia, há o risco de ocorrer também a detecção de produtos inespecíficos e, por isso, a especificidade do produto de PCR gerado deve ser sempre garantida pela análise de curvas de dissociação (Klein, 2002).

Há dois métodos empregados para analisar os dados de PCR em tempo real: quantificação absoluta e quantificação relativa. A quantificação absoluta é utilizada para determinar o número de cópias de determinado gene ou de seu transcrito em uma amostra, o que é feito relacionando-se o sinal de PCR a uma curva-padrão, construída com base na leitura do sinal para diferentes diluições de uma referência (RNA, cDNA, DNA genômico ou plasmidial). A quantificação relativa refere-se à alteração de nível de expressão de determinado gene relativo a alguma referência ou controle e o resultado final obtido é dado em quantas vezes um gene está mais ou menos expresso em uma amostra em relação à outra. O método de quantificação relativa para dados de PCR em tempo real mais empregado utiliza o cálculo do $2^{-\Delta\Delta C_t}$, baseado na determinação empírica de um limite (ponto em que todas as reações alcançam intensidade de fluorescência acima do ruído de fundo) situado na fase exponencial da reação e a comparação entre os valores de C_t (número de ciclos necessários para que a curva referente a determinada amostra alcance o limite definido) nas diferentes reações (Livak & Schmittgen, 2001). Apesar da praticidade, esse método é limitado por assumir uma eficiência igual a 100% na reação de PCR para todas as amostras, o que nem sempre ocorre, além da dificuldade de se definir um único valor de limite que seja adequado para

todas as amostras. Com o intuito de minimizar essas limitações, outros métodos mais sofisticados de análise foram desenvolvidos (Pfaffl 2001, Liu & Saint 2002a, Liu & Saint 2002b, Peirson *et al* 2003).

Um dos grandes desafios do emprego da tecnologia de PCR em tempo real é a escolha do gene a ser utilizado como normalizador. O normalizador age como um controle interno e tem como função eliminar o erro associado a eventuais diferenças na quantidade de cDNA adicionada a cada uma das reações. Para isso, o normalizador deve ser um gene que apresente níveis de expressão iguais nas diferentes amostras testadas. Há diversos genes que são classicamente utilizados como normalizadores, como actina, albumina, tubulina, gliceraldeído-3-fosfato desidrogenase (GAPDH) e ciclofilina, além dos ribossomais, amplamente utilizados. Entretanto, diversos estudos demonstraram que nem sempre esses genes são realmente expressos de forma homogênea em diferentes condições e órgãos testados (Zhong & Simons 1999, Schmittgen & Zakrajsek, 2000, Kim *et al* 2003), levando à conclusão de que não há um normalizador universal e sim aquele que é o mais adequado para dada condição experimental.

Para concluir esse tópico, é oportuno citar a discussão que vem sendo levada à tona atualmente pela comunidade científica acerca da necessidade ou não de se validar os resultados de hibridizações contra microarranjos. Em princípio, a adoção de uma técnica de validação é altamente recomendável, já que a técnica de microarranjos é passível de muitas variações e erros, como, por exemplo, falso-positivos oriundos de hibridizações cruzadas, genes com variantes de processamento cujas expressões podem não estar sendo analisadas de modo separado pelo microarranjos, incertezas com relação às sequências de cDNA depositadas nas lâminas ou mesmo erros decorrentes da cinética do processo em si, pois provavelmente nem todas as reações de hibridização de um experimento de microarranjos ocorrem em condições idênticas (Rockett & Hellmann, 2004). Por outro lado, dependendo do estudo conduzido, o uso de reações de validação é dispensável, como, por exemplo, quando há limitação na disponibilidade de amostras de RNA, quando o objetivo do estudo é avaliar mudanças gerais no padrão de expressão gênica celular, sem focar a discussão em transcritos individualmente, ou quando um grande número de réplicas dos experimentos de microarranjos foi conduzido (Rockett & Hellmann, 2004). Apesar desse tema

ainda estar sendo discutido e não ser um consenso entre os pesquisadores, o que prevalece atualmente é o bom senso no momento de se decidir ou não pela validação, bem como na escolha de quantos perfis de expressão validar e quais genes selecionar para ensaios de validação (Rockett & Hellmann, 2004).

3.4. Estímulos bióticos e abióticos em plantas

Estresses de origem ambiental constituem os principais responsáveis por perdas na agricultura e podem reduzir a produtividade média de 65% a 87%, dependendo da planta (Bray *et al.*, 2000). As plantas são capazes de reagir a alterações ambientais através de diversas respostas celulares ativadas por estímulos de estresses, levando à defesa da planta e/ou ao ajuste às condições adversas. Plantas com capacidade melhorada de sobreviver em condições extremas ou a estresses bióticos vêm sendo selecionadas ao longo dos anos por programas de melhoramento genético. Contudo, o desenvolvimento de cultivares melhorados por técnicas clássicas de melhoramento genético é um processo extremamente lento. Nesse cenário, a utilização de técnicas moleculares é altamente promissora, já que permite modificações direcionadas ao gene-alvo.

As alterações fisiológicas disparadas por sinais externos podem ser moduladas por regulação transcricional, resultando em indução ou repressão gênica. Diversos estudos foram conduzidos em larga escala a fim de se determinar alterações nos perfis de expressão de plantas submetidas a condições estressantes (Schenk *et al.*, 2000, Kawasaki *et al.*, 2001, Seki *et al.*, 2001, Cheong *et al.*, 2002, Cooper *et al.*, 2003, Hazen *et al.*, 2003, Mahalingam *et al.*, 2003, Oono *et al.*, 2003, Oono *et al.*, 2006). Esses estudos mostram que a expressão gênica de componentes de vias de transdução de sinais é alterada em resposta ao estresse possivelmente levando a alterações no crescimento e desenvolvimento e ao ajuste a condições ambientais. Entretanto, poucos estudos foram conduzidos no que diz respeito às respostas da cana-de-açúcar a estresses bióticos e abióticos. Como exemplo de estudos desse tipo, podemos citar os conduzidos por Nogueira *et al.* (2003) e de Rosa *et al.* (2005), que avaliaram alterações no transcriptoma de cana-de-açúcar

ativadas pelo frio e pelo tratamento com metiljasmonato, respectivamente. Esses estudos são necessários para que se possa desvendar a intrincada rede de sinalização que ocorre nas células em respostas a alterações no meio, mas é importante ressaltar que, ao contrário das condições comumente empregadas experimentalmente, as plantas estão sempre submetidas a condições de estresses múltiplos quando no campo. Além disso, cada órgão da planta responde aos estresses de forma diferente.

A seguir, os estímulos bióticos e abióticos cujas respostas em nível transcricional foram analisadas e que foram selecionados por alterarem a produtividade (positivamente ou negativamente) da cana são descritos. Adicionalmente, os papéis de ABA e MeJA na regulação das respostas fisiológicas relacionadas a estresses, crescimento e desenvolvimento são apresentados.

3.4.1. Fitormônios: ácido jasmônico e ácido abscísico

As respostas a estresses em plantas são reguladas por múltiplas vias de sinalização e há uma sobreposição considerável entre os padrões de expressão induzidos nas plantas em resposta a diferentes estresses, mostrando a existência de redes regulatórias complexas que permitem que a planta possa responder de forma ótima a um ambiente em constante mudança. Essa sobreposição na resposta a diferentes tipos de estresses já foi demonstrada por trabalhos como Schenk *et al.* (2000), Seki *et al.* (2001) e Chen *et al.* (2002). Além de outras funções fisiológicas, diversos fitormônios participam das vias de defesa das plantas contra estresses bióticos e abióticos, como é o caso dos jasmonatos e do ABA.

O ácido jasmônico (JA) e outras oxilipinas são produzidos de modo característico em resposta à patogênese e à herbivoria, sendo que os jasmonatos são responsáveis por mediar as respostas de defesa contra insetos, fungos e bactérias (Farmer *et al.*, 2003). Além de respostas de defesa, os jasmonatos também estão relacionados a outros eventos fisiológicos das plantas, como florescimento, amadurecimento dos frutos, crescimento das raízes, senescência e respostas a estresses ambientais, como seca, frio e salinidade (Creelman & Mullet, 1997, Wasternack &

Parthier, 1997, Turner *et al.*, 2002, Cheong & Choi, 2003). Os jasmonatos são sintetizados nas plantas pela via octadecanóica. Determinados estímulos ativam fosfolipases, levando à liberação de ácido linolênico da membrana. Através da ação das enzimas lipoxigenase (LOX), aleno óxido sintase (AOS) e aleno óxido ciclase (AOC), o ácido linolênico liberado é convertido em ácido 12-oxofitodienóico (OPDA), a partir do qual é sintetizado o ácido jasmônico (JA) (Cheong & Choi, 2003, Devoto & Turner, 2003). O metabólito OPDA também é biologicamente ativo, capaz de induzir resistência de amplo espectro mesmo na ausência de JA (Stintzi *et al.*, 2001). Metiljasmonato (MeJA) é sintetizado a partir de JA através da ação da enzima JA carboxil metil transferase (JMT). Essa enzima é expressa em órgãos vegetativos e em flores em desenvolvimento e é induzida quando plantas são submetidas a ferimentos ou tratamento com MeJA (Tuner *et al.*, 2002). O metiljasmonato é capaz de mediar muitas das respostas do JA e, sendo volátil, pode mediar tanto respostas locais quanto sistêmicas, além de poder agir na comunicação entre plantas (Seo *et al.*, 2001).

Pouco se sabe ainda sobre as vias de sinalização dos jasmonatos. Presume-se que os jasmonatos interajam com receptores de membrana, o que ativa uma via de sinalização que resulta em alterações na transcrição, tradução e outras respostas. Até o momento, nenhum receptor foi identificado para JA, mesmo com as análises de mutantes identificados para a via de transdução de sinais de JA tendo sido conduzidas à exaustão, o que leva à hipótese de que deve haver redundância entre os receptores de JA (Devoto & Turner, 2003). Dentre os genes identificados como sendo importantes para a via de jasmonatos, estão *COI1* e *JAR1*. *COI1* codifica para uma proteína com domínios F-box e LRRs (Ruegger *et al.*, 1998, Xie *et al.*, 1998). Proteínas do tipo F-box são componentes de complexos SCF (SKP1, CDC53p, CUL1, proteína F-box), que agem como receptores para marcar proteínas específicas para proteólise mediada pela ubiquitina (Bai *et al.*, 1996, Glickman & Ciechanover, 2002). Recentemente, uma proteína do tipo F-box (TIR1) foi identificada como sendo o receptor do fitormônio auxina (Dharmasiri *et al.*, 2005, Kepinski & Leyser, 2005).

Diversos estudos de expressão gênica relacionados ao tratamento com MeJA já foram conduzidos (vide, por exemplo, Schenk *et al.*, 2000, Sasaki *et al.*, 2001). Dentre os genes

observados como sendo diferencialmente expressos em decorrência do tratamento com jasmonatos estão incluídos aqueles relacionados com a biossíntese de jasmonato, metabolismo secundário, formação de parede celular, germinação das sementes, desenvolvimento de flores e frutos, genes codificadores para proteínas envolvidas com vias de resposta a estresses e defesa e genes relacionados com o processo de fotossíntese (Creelman & Mullet, 1997, Cheong & Choi, 2003). A repressão de genes de fotossíntese pelos jasmonatos pode estar relacionada com o processo de senescência ou com a redução da capacidade da planta de assimilação de carbono em situações de excesso de luz ou de carbono, com a finalidade de equilibrar as capacidades de absorção de energia e de utilização de mesma (Creelman & Mullet, 1997).

Contudo, vale lembrar que, como já apontado por Cheong & Choi (2003), JA e MeJA têm sido usados em muitos experimentos sem se considerar qual é o papel específico de cada tipo de jasmonato nas respostas observadas. Ou seja, ainda há necessidade de se diferenciar claramente as funções dos membros da família dos jasmonatos.

O fitormônio ABA está envolvido em diferentes processos fisiológicos das plantas: maturação e dormência de sementes, adaptação dos órgãos vegetativos a estresses abióticos (principalmente frio e seca) e desenvolvimento (Himmelbach *et al.*, 2003). Como ocorre com os demais fitormônios, sabe-se que as vias desencadeadas por ABA podem interagir tanto de forma sinérgica quanto antagônica com a de outros fitormônios. Entretanto, é também interessante realçar que o ABA possui muitos efeitos similares e antagônicos aos efeitos das vias reguladas por açúcares e relacionados com desenvolvimento de sementes, germinação e crescimento (Finkelstein & Gibson, 2002).

O ABA pertence a uma categoria de metabólitos conhecidos como isoprenóides, ou terpenóides. Esses compostos derivam de um precursor de cinco carbonos, o isopentenil. A molécula de ABA contém 15 átomos de carbono (C₁₅) e é sintetizada a partir da quebra de carotenóides C₄₀ produzidos a partir do ácido mevalônico (Nambara & Marion-Poll, 2005). Ainda não estão bem determinados quais são os locais de síntese de ABA em situação de estresse, mas sabe-se que, em condições de não-estresse, o ABA é sintetizado principalmente nos tecidos vasculares e transportado para diversos órgãos após sua síntese, podendo também ser sintetizado

nas células-guarda. O catabolismo de ABA ocorre via reações de hidroxilação, com participação de citocromo P450 monooxigenases, e de conjugação, onde grupos hidroxila e carboxila do ABA ou de seus derivados oxidados são conjugados com glicose (Nambara & Marion-Poll, 2005).

Diversas vias de sinalização ativadas pelo fitormônio ABA já foram elucidadas. Dentre os mensageiros intracelulares envolvidos com as vias do ABA estão cálcio, ácido fosfatídico, inositol 1,4,5-trifosfato, óxido nítrico, peróxido de hidrogênio e GMP cíclico (Himmelbach *et al.*, 2003). Importantes reguladores das vias iniciadas por ABA incluem proteína G, NADPH oxidase, fosfatases da classe das PP2Cs, proteínas quinases dependentes de cálcio e proteínas quinases do tipo SnRKs (Himmelbach *et al.*, 2003). O ABA apresenta um envolvimento direto na resposta a estresses abióticos, já que algumas das enzimas da via de biossíntese desse fitormônio são induzidas frente à seca e ao estresse osmótico (Nambara & Marion-Poll, 2005). Como um importante avanço nos estudos das vias de transdução de ABA, destaca-se o trabalho recente de Osakabe *et al.* (2005), que caracterizaram o receptor RPK1, de *Arabidopsis thaliana*, um receptor do tipo Ser/Thr quinase com domínio extracelular com LRRs e que pode estar relacionado com a percepção desse fitormônio.

3.4.2. Herbivoria

Dentre as inúmeras pragas que atacam as lavouras de cana-de-açúcar, destaca-se a broca da cana (*Diatraea saccharalis*). As lagartas de *D. saccharalis*, após a eclosão, alimentam-se do parênquima das folhas e convergem para as bainhas das mesmas. Posteriormente, as lagartas da primeira muda penetram nas gemas laterais e abrem galerias dentro dos colmos, onde completam o desenvolvimento larval. As lagartas, quando atacam as canas novas, causam a morte da gema apical, cujo sintoma é conhecido por "coração morto". Em cana adulta também ocorre perda de peso, brotação lateral, enraizamento aéreo, quebra do colmo e atrofia dos entrenós. Os orifícios deixados pelas lagartas representam potenciais portas de entrada para fungos que ocasionam a podridão vermelha, determinando queda no rendimento industrial pela inversão da

sacarose, diminuição da pureza do caldo e problemas de contaminações no processo de fermentação alcoólica (Barsalobres, 2004). Dado os prejuízos que tal praga acarreta para as lavouras de cana-de-açúcar, o estudo das vias de sinalização desencadeadas pelo ataque por esse inseto é de grande interesse biotecnológico.

As defesas das plantas contra herbívoros podem ser classificadas como defesas diretas e indiretas e podem ser tanto constitutivas quanto induzidas. As defesas diretas visam a impedir que os herbívoros se alimentem e incluem barreiras físicas ou químicas (via produtos de metabolismo secundário, como fenilpropanóides, terpenóides e alcalóides, ou via proteínas de defesa especializadas, como é o caso dos inibidores de proteases). Já as defesas indiretas compreendem a liberação de compostos voláteis no ambiente e a oferta de fonte de alimento (via secreção de néctar extrafloral) que atraem inimigos naturais dos herbívoros (Arimura *et al.*, 2005).

O estímulo de herbivoria é capaz de ativar diversas vias de sinalização em plantas. Dentre as vias ativadas podem ser citadas as vias de sinalização por cálcio, vias de sistemina, etileno, ácido salicílico e JA, vias de MAPKs, produção de espécies reativas de oxigênio (ROS), dentre outras (Falco *et al.*, 2001). O sinal para a ativação dessas vias pode ser o ferimento gerado nas plantas em decorrência da herbivoria e a presença de substâncias presentes na secreção oral dos insetos (Arimura *et al.*, 2005).

Os jasmonatos desempenham um grande papel na resistência a insetos e doenças. Em situação de herbivoria (bem como em resposta a outros estresses, como patógenos e estresses abióticos), as plantas são capazes de liberar compostos orgânicos voláteis (VOCs) na atmosfera. Os VOCs sinalizam para as demais plantas a presença do herbívoro em plantas vizinhas e é capaz de ativar alterações no transcriptoma, proteoma e metaboloma, com ativação de vias de defesa (Baldwin *et al.*, 2006). A constituição dos VOCs varia muito e pode conter, dentre outros compostos, terpenóides, metanol, etileno, metilsalicilato e metiljasmonato. Dentre as funções exercidas pelos jasmonatos nas respostas de herbivoria, destacam-se a indução da expressão de inibidores de proteases e proteínas antifúngicas, síntese de fitoalexinas, compostos fenólicos e aldeídos voláteis e modulação da síntese de parede celular (levando à síntese de barreiras contra

a infecção) (Creelman & Mullet, 1997). McConn e colaboradores (1997) mostraram, através de estudos com mutantes de *Arabidopsis* deficientes para a biossíntese de jasmonatos, a necessidade desse hormônio para que a planta desencadeie a resposta de defesa contra insetos. Além da participação dos jasmonatos na via de sinalização de herbivoria, outros fitormônios podem desempenhar importantes papéis, como ácido salicílico, etileno, ABA e auxina (Farmer *et al.*, 2003, Arimura *et al.*, 2005).

3.4.3. Interação com bactérias endofíticas

É sabido que determinados genótipos de cana-de-açúcar podem ser cultivados sem a necessidade de elevadas concentrações de fertilizantes e tal fenômeno deve-se à associação dessas plantas com bactérias endofíticas diazotróficas, que são bactérias com capacidade de fixação de nitrogênio e que passam a maior parte do seu ciclo de vida dentro das plantas sem causar sintomas de doenças ao hospedeiro (Reinhold-Hurek & Hurek, 1998). Um exemplo de interação entre plantas e microorganismos fixadores de nitrogênio já bem estudado é a formação de nódulos nas raízes de leguminosas, que constituem um microambiente com condições ótimas para o crescimento de bactérias fixadoras de nitrogênio. As gramíneas, entretanto, não possuem estruturas especializadas para a interação com microorganismos fixadores de nitrogênio. Por outro lado, bactérias como, por exemplo, *Gluconacetobacter diazotrophicus* e *Herbaspirillum seropedicae*, presentes nos órgãos de cana-de-açúcar, possuem a capacidade de penetrar nas plantas através das raízes e de se espalhar sistemicamente, provavelmente via vasos do xilema (Reinhold-Hurek & Hurek, 1998). Essas bactérias vivem no interior dos órgãos das plantas, um local de reduzido teor de oxigênio, onde podem adquirir suprimento de carbono das plantas e realizar uma eficaz fixação de nitrogênio.

A bactéria fixadora de nitrogênio *Herbaspirillum seropedicae* foi isolado de milho, sorgo e arroz em 1986 (Baldani *et al.*, 1986) ao passo que *Herbaspirillum rubrisubalbicans* foi descrita apenas em 1996 (Baldani *et al.* 1996). Já a bactéria *Gluconacetobacter diazotrophicus* foi isolada

de raízes e colmos de cana-de-açúcar, em 1988 (Cavalcante & Döbereiner, 1988). As bactérias endofíticas colonizam os espaços intercelulares e tecidos vasculares da maior parte dos órgãos da cana, sem causar sintomas de doenças em cultivares de associação eficiente. Essas bactérias endofíticas sintetizam IAA (ácido indolacético) e giberelina, o que contribui para o crescimento das plantas com as quais elas se associam (Reinhold-Hurek & Hurek, 1998).

Ainda há muito a se estudar com relação aos mecanismos moleculares que promovem essa associação eficiente entre endofitas e plantas. De forma geral, os genes envolvidos na interação planta-endofíticos podem ser classificados como pertencentes às categorias de metabolismo de nitrogênio, de crescimento (incluindo genes de ciclo celular e genes de resposta à auxina, giberelina e etileno), de sinalização celular e de genes homólogos à nodulina (Nogueira *et al.*, 2001). É interessante notar que o genótipo da planta desempenha um papel importante nessa associação, já que há genótipos de cana-de-açúcar com diferentes taxas de BNF (fixação biológica de nitrogênio) (Baldani & Baldani, 2005). Dentre alguns dos avanços recentes com relação à interação endofitas-planta, podemos citar o trabalho de Blanco e colaboradores (2005), que encontraram receptores na parede celular de *Gluconacetobacter diazotrophicus* capazes de reconhecer e de se ligar a glicoproteínas da cana-de-açúcar. Mais recentemente, um RLK de cana-de-açúcar (SHR5) foi descrito como sendo reprimido em plantas associadas com bactérias diazotróficas endofíticas e também relacionado à eficiência dessa associação (Vinagre *et al.*, 2006).

3.4.4. Deficiência em fosfato

O fósforo é um nutriente essencial para todos os organismos vivos, já que exerce várias funções biológicas básicas. Além de ser elemento estrutural de moléculas como ácidos nucleicos, fosfolípidos e ATP, possui papel regulador na transferência de energia, ativação de proteínas e metabolismo de carbono e de aminoácidos (Marschner, 1995). Com isso, a disponibilidade de fosfato possui consequências diretas no crescimento e na fisiologia das plantas. A deficiência em

fosfato é um dos principais fatores limitantes da produtividade agrícola em todo o mundo. Após o nitrogênio, o fósforo é o nutriente de menor disponibilidade tanto nos ecossistemas terrestres quanto nos aquáticos (Abel *et al.*, 2002) e também o elemento de menor acessibilidade para as plantas. Isso ocorre porque as plantas preferencialmente adquirem o fosfato a partir dos solos sob a forma de ortofosfato (Pi) e mais de 80% do ortofosfato dos solos está imobilizado, não prontamente disponível para as raízes (Holford, 1997). Desse modo, a disponibilidade de fósforo nos solos nativos raramente é adequada para o crescimento ótimo das plantas. Para enfrentar tal carência, as plantas desenvolveram diversos mecanismos adaptativos, incluindo alterações morfológicas, bioquímicas, fisiológicas e de regulação da expressão gênica.

As alterações morfológicas envolvem basicamente alterações na arquitetura radicular, que levam ao aumento da razão superfície da raiz/volume de solo e aumentam a capacidade da planta em acessar o fosfato do solo (Franco-Zorrilla *et al.*, 2004). As alterações bioquímicas que ocorrem em resposta à deficiência em fosfato possuem duas funções básicas: aumento da disponibilidade de Pi endógeno e do solo e adaptações metabólicas à carência nutricional, levando ao uso de vias metabólicas alternativas que não requeiram Pi. Para isso, são desencadeadas respostas como a liberação de prótons e ácidos orgânicos e o acúmulo de antocianinas (devido a limitações nas reações de fotossíntese) (Franco-Zorrilla *et al.*, 2004). Também ocorrem alterações na expressão gênica de transportadores de Pi (Raghothama, 2000), RNases (Bariola *et al.*, 1999), enzimas mobilizadoras de fosfato como fosfatase ácida, fitase e ribonuclease (Bosse & Köck, 1998), fosfodiesterases (Abel *et al.*, 2000) e enzimas envolvidas no catabolismo da glicose (Dumont *et al.*, 1993). Ainda com relação ao controle da expressão gênica, foi mostrado que muitos dos genes induzidos precocemente pela carência de Pi são também induzidos pela deficiência em K e Fe, apontando para potencial interação entre as vias de sinalização de deficiência em diferentes nutrientes (Wang *et al.*, 2002).

3.4.5. Resposta à seca

Entre os estresses que afetam a produtividade agrícola, a resposta ao estresse hídrico destaca-se como o fator mais limitante ao crescimento das plantas. A disponibilidade de água doce diminuiu mais de seis vezes entre 1900 e 1995, mais do que o dobro da razão do crescimento populacional (Kates & Parris, 2003). A agricultura, principalmente a irrigação, é responsável por 70% do uso de água, o que aponta para uma necessidade de se aumentar a eficiência na produção agrícola (Kates & Parris, 2003). Dada a escassez de água prevista para o futuro em muitas áreas, o sustento da indústria canavieira no Brasil e no mundo irá depender do desenvolvimento de cultivares de cana-de-açúcar tolerantes ao estresse hídrico e com capacidade de uso dos recursos hídricos com maior eficiência.

As respostas à seca variam dependendo da intensidade ou duração do estímulo e compreendem alterações rápidas no fluxo de íons para otimizar a pressão osmótica celular, fechamento dos estômatos para reduzir a perda por transpiração, produção de substâncias osmoprotetoras, como açúcares, prolina e proteínas e, no caso de seca prolongada, alterações no padrão e crescimento das plantas (Riera *et al.*, 2005).

Genes induzidos por estresses abióticos (como seca, salinidade e frio) incluem os que codificam para proteínas que agem na proteção celular ao estresse, como é o caso de chaperonas, osmotina, proteínas que se ligam a mRNA, enzimas para biossíntese de osmólitos, canais de água, transportadores de açúcar e de prolina, enzimas de detoxificação, enzimas de metabolismo de ácidos graxos, inibidores de proteases, proteínas LEA (*Late Embryogenesis Abundant proteins*), ferritina e proteínas de transferência de lipídeos (Yamaguchi-Shinozaki & Shinozaki, 2006). Adicionalmente, são induzidos genes envolvidos na regulação de vias de transdução de sinais e reguladores da expressão de genes relacionados a respostas a estresses. Dentro dessa categoria, encontram-se proteínas quinases, proteínas fosfatases, enzimas do metabolismo de fosfolipídeos, proteínas de ligação com calmodulina e, principalmente, fatores de transcrição (Yamaguchi-Shinozaki & Shinozaki, 2006). Diversos genes responsivos à seca no nível de transcrição já estão catalogados (Bray, 1993, Bohnert *et al.*, 1995, Ingram & Bartels,

1996; Seki *et al.*, 2001, Oono *et al.*, 2003), podendo ser divididos nos grupos dependente e independente de ABA, já que esses estímulos levam à síntese desse fitormônio (Riera *et al.*, 2005).

O estímulo de seca tem grande impacto sobre as vias celulares de transdução de sinais. Esse estímulo induz fluxos transientes de cálcio para o citoplasma, seguidos pela ativação de sensores de cálcio (como, por exemplo, CDPKs e a família SOS3 de sensores de cálcio), estimula o acúmulo de ROS, ativa vias de MAPKs e induz a expressão de determinados fatores de transcrição, como, por exemplo, DREB2A e DREB2B (Xiong *et al.*, 2002). Muitos dos genes responsivos à seca têm em comum o fato de possuírem em seus promotores elementos responsivos à desidratação do tipo DRE/CRT, reconhecidos por fatores de transcrição do tipo DREB/CBF (Yamaguchi-Shinozaki & Shinozaki, 2006).

3.5. O desenvolvimento de cultivares de cana-de-açúcar mais ricos em sacarose

Devido à sua capacidade única de armazenar sacarose nos colmos, a cana-de-açúcar representa um modelo interessante para estudos de síntese, transporte e acúmulo de açúcares. Devido ao elevado acúmulo de sacarose em colmos de cana-de-açúcar, as células desse órgão devem sofrer adaptações ao longo de sua maturação, permitindo que sejam mantidas a integridade e a homeostasia celular. A qualidade do caldo da cana-de-açúcar é determinada pela sua concentração de sacarose, que deve ser elevada, e pela concentração de outros componentes, como outros tipos de açúcares, ácidos orgânicos e sais minerais, que deve ser baixa. Os cultivares de cana-de-açúcar diferenciam-se entre si pela máxima concentração de sacarose alcançada no caldo, bem como a época do ano em que a planta alcança essa concentração máxima (Lingle, 1997).

Os entrenós de cana-de-açúcar amadurecem progressivamente em direção à base do colmo, com um aumento correspondente nos níveis de sacarose. Enquanto um entrenó maduro atinge teores de sacarose correspondentes a cerca de 20% do seu peso fresco (cerca de 80% de seu peso seco), os entrenós mais jovens (mais apicais) contêm menores níveis de sacarose e

podem conter glicose e frutose em níveis de aproximadamente 5% do peso fresco. Outros açúcares, além de glicose, frutose e sacarose também são encontrados em cana-de-açúcar, embora em níveis reduzidos. Os níveis de amido nos cultivares modernos geralmente não ultrapassam 0,01% do peso fresco do colmo (Glasziou & Gayler, 1972).

O teor de sacarose é uma característica fenotípica selecionada em programas de melhoramento genético. Uma forma simples de se inferir o teor de sacarose no caldo da cana-de-açúcar é pela medição de brix, que avalia o teor de sólidos solúveis em determinada amostra, através de um refratômetro. Dentre os constituintes sólidos solúveis do caldo da cana-de-açúcar, a maior parte (70 a 91%) corresponde à sacarose. Como mencionado anteriormente, os cultivares comerciais modernos derivam principalmente de cruzamentos e retrocruzamentos de duas espécies, *Saccharum officinarum* e *Saccharum spontaneum*. Esses cruzamentos se iniciaram a partir do final do século XIX, com o processo de “nobilização da cana-de-açúcar”. Nesse processo, foram conduzidos programas de cruzamentos interespecíficos envolvendo essencialmente genótipos de *S. officinarum* e *S. spontaneum*. As progênies híbridas geradas foram, então, repetidamente retrocruzadas com *S. officinarum* de forma a minimizar os efeitos negativos do parental selvagem (Lu *et al.*, 1994). *S. spontaneum* caracteriza-se por ter colmos finos, com muito pouco açúcar e resistente a condições adversas. Já *S. officinarum* apresenta colmos suculentos com alto teor de açúcar (Grivet *et al.*, 2004).

As células do mesófilo de folhas representam o órgão fotossintético principal (“fonte”) e o produto da fotossíntese, principalmente sob a forma de sacarose, é transportado para meristemas e órgãos em desenvolvimento. Em cana-de-açúcar, folhas jovens em crescimento e o colmo constituem os órgãos importadores (“dreno”) de carboidratos mais importantes. Órgãos fonte e dreno devem ser coordenadamente regulados no nível de expressão gênica e de atividade enzimática de forma a produzir crescimento rápido e acúmulo de sacarose eficiente. A sacarose, sintetizada nas folhas, é transportada por transportador específico pelos vasos do floema até os colmos, onde passa para o espaço apoplástico. Nesse espaço, a sacarose, através da ação de invertases extracelulares, é convertida a glicose e frutose e esses monossacarídeos entram nas células do parênquima através de transportadores específicos. Dentro dessas células, frutose e

glicose combinam-se novamente para originar sacarose-fosfato (através da ação da enzima sacarose fosfato sintase), que é transportada para o interior dos vacúolos. A sacarose armazenada nos vacúolos pode ser mobilizada via ação de invertases, que levam à sua quebra e geração de frutose e glicose. Esses açúcares difundem-se, então, para o citoplasma das células do parênquima (Moore, 1995, Roitsch, 1999).

Jansen e Nap (2001) introduziram o termo “genômica genética” para se referir à abordagem de emprego de técnicas de análise de expressão gênica em larga escala para populações segregantes. Com isso, genes associados com características de interesse como, por exemplo, acúmulo de sacarose, podem ser identificados e direcionados para a melhoria dos cultivares comerciais atuais. Diversos estudos baseados na análise da expressão gênica em larga escala em cana-de-açúcar, objetivando a detecção de genes relacionados ao acúmulo de sacarose, já foram realizados (Carson & Botha, 2002; Carson *et al.*, 2002; Casu *et al.*, 2003, 2004 e 2005). Entretanto, esses estudos focam simplesmente a comparação entre os entrenós de cana-de-açúcar em diferentes estágios de desenvolvimento e não na comparação de cultivares contrastantes para acúmulo de sacarose entre si.

Por último, é importante citar que os açúcares agem também como moléculas sinalizadoras em plantas, afetando uma variedade de processos fisiológicos, como fotossíntese, metabolismo em órgãos fonte e dreno e respostas de defesa (Smeekens, 2000, Rolland *et al.*, 2002, Koch, 2004). Entretanto, os mecanismos moleculares envolvidos na percepção dos açúcares e a transmissão intracelular desses sinais ainda não são bem conhecidos. Sabe-se, contudo, que a sinalização por açúcares pode ser em parte desencadeada pela hexoquinase, enzima essa com capacidade de agir como sensor dos níveis intercelulares de glicose, e levar a alteração nos níveis dos transcritos de genes codificando, por exemplo, para ciclinas, genes relacionados a processos de resposta à patogênese e a estresses, genes relacionados a vias de metabolismo de carbono, de nitrogênio e de metabolismo secundário (Rolland *et al.*, 2002). As vias de sinalização por açúcares apresentam respostas cruzadas com outras vias, incluindo as de respostas a hormônios, de controle de ciclo celular e de respostas a nitrogênio (Halford & Paul, 2003). Dentre essas respostas cruzadas, é importante ressaltar a participação do fitormônio ABA na sinalização de

açúcares dependente de hexoquinase. Ainda não está claro se os açúcares são capazes de aumentar a sensibilidade celular ao ABA ou se os açúcares levam ao aumento na concentração desse hormônio. As interações entre vias mediadas por ABA e por açúcares incluem a regulação por esses dois estímulos de genes de órgãos drenos, como esporamina e β -amilase, a promoção do desenvolvimento do tubérculo de batata, a ativação por ABA de vias que inibem a germinação de sementes e a ação desse hormônio em inibir o carregamento de sacarose no floema, tendo um papel direto na regulação celular dos níveis de açúcares (Smeekens, 2000).

A compreensão da regulação da síntese de sacarose em cana-de-açúcar em nível transcricional e a detecção de genes responsivos ao acúmulo de açúcar serão etapas de grande valor para o sucesso de programas de melhoramento genético e para a compreensão dos mecanismos de sinalização por açúcares.

4. Objetivos

Considerando-se a identificação bem sucedida de mais de 40.000 possíveis genes da cana-de-açúcar pelo projeto SUCEST e visando à obtenção de dados funcionais que beneficiem a agroindústria da cana, é oportuno avançar na prospecção da funcionalidade biológica dos genes identificados. Dessa forma serão descobertos novos genes associados a traços agrônômicos relevantes.

O presente projeto objetivou a catalogação e anotação de componentes de transdução de sinais obtidos a partir do banco SUCEST utilizando-se a ferramenta BLAST (Altschul *et al.*, 1997), com ênfase na análise *in silico* do quinoma da cana-de-açúcar. Para a melhor categorização das proteínas quinases de cana-de-açúcar, essas foram anotadas com base em buscas por BLAST e analisadas também quanto a domínios conservados e por similaridades no domínio catalítico.

Dado que a análise do padrão de expressão pode ser indicativa da função desempenhada por determinados genes, em uma segunda fase do projeto a técnica de hibridização contra microarranjos de cDNA, contendo elementos do Catálogo SUCAST, foi adotada para a comparação dos níveis de expressão em diferentes órgãos da cana-de-açúcar. Essa técnica foi também adotada para a detecção de genes responsivos a fitormônios e a estímulos que interferem diretamente com a produtividade da indústria canavieira (seca, deficiência em fosfato, herbivoria pela broca da cana e inoculação com bactérias fixadoras de nitrogênio) e de genes que possam estar relacionados ao controle de síntese e acúmulo de sacarose. Os objetivos específicos do projeto são citados abaixo:

1. Busca de genes correspondentes às várias categorias de transdução de sinais, com ênfase na categoria de proteínas quinases, utilizando-se o algoritmo BLAST (Altschul *et al.*, 1997).
2. Busca de domínios codificados nos transcritos candidatos a proteínas quinases utilizando-se o algoritmo PFAM (Sonnhammer *et al.*, 1998) e SMART (Schultz *et al.*, 1998).

3. Alinhamento e análise filogenética das famílias de proteínas quinases utilizando-se o algoritmo de agrupamento de vizinhos (*neighbour-joining*), utilizando a seqüência consenso dos SAS de cana-de-açúcar e seqüências de genes de outros organismos.
4. Seleção dos componentes SUCAST a serem representados nos microarranjos, implementação da tecnologia de microarranjos de cDNA e construção do microarranjo SUCAST.
5. Análise de variabilidade biológica entre indivíduos cultivados em campo, por meio de hibridizações utilizando-se amostras de RNA de folha +1 de diferentes indivíduos.
6. Realização de experimentos de hibridização em microarranjos de DNA utilizando amostras de RNA de diferentes órgãos de cana-de-açúcar (raiz, folha, flor, gema lateral, primeiro e quarto entrenós), de plantas submetidas a tratamentos com MeJA (1, 6 e 12 horas após o tratamento), ABA (30 min, 1, 6 e 12 horas após o tratamento), herbivoria pela broca da cana (30 minutos e 24 horas), associação com endofitas (inoculação com mistura das bactérias endofíticas *Herbaspirillum seropedicae*, estirpe HRC54, e *Herbaspirillum rubrisubalbicans*, estirpe HCC103 ou inoculação com *Gluconacetobacter diazotrophicus*), deficiência em fosfato (6, 12, 24 e 48 horas de deficiência nutricional), seca (24, 72 e 120 horas após a supressão de rega), bem como de cultivares contrastantes para teor de brix, provenientes de introgressão entre as espécies *S. spontaneum* e *S. officinarum*.
7. Padronização do modo de normalização dos dados de expressão gerados, bem como de detecção de genes diferencialmente expressos.
8. Análise dos dados de expressão gerados, com detecção de genes diferencialmente expressos e agrupamento dos perfis de expressão.
9. Utilização da tecnologia de PCR em tempo real para avaliar os níveis de expressão para genes selecionados como sendo diferencialmente expressos para os diversos experimentos de microarranjos e avaliação do nível de validação alcançado para os dados de microarranjos comparando-se os resultados provenientes das duas técnicas entre si.

5. Material e Métodos

5.1. Prospecção dos dados, anotação e análise de domínios conservados

Sequências de proteínas relacionadas a proteínas quinases e receptores (principalmente de *Arabidopsis thaliana*, arroz e milho) foram selecionadas a partir de buscas na literatura e usadas em buscas através do algoritmo TBLASTN (Altschul *et al.*, 1997) contra o banco de dados do SUCEST (<http://sucest.lad.ic.unicamp.br/en/Services/services.html>). Todos os SAS obtidos com *E value* igual ou menor do que 10^{-15} foram analisados e anotados de acordo com similaridades a outras sequências. As sequências de proteínas preditas pelo ESTScan (Iseli *et al.*, 1999) para os SAS selecionados a partir das buscas por BLAST foram inspecionadas quanto à presença de domínios conservados usando-se o banco de dados Pfam (Sonnhammer *et al.*, 1998).

5.2. Análise filogenética de proteínas quinases e RLKs

Para a construção das árvores filogenéticas, as sequências dos domínios de *pkinase* (conforme predito pelas buscas contra o banco de dados Pfam) das proteínas quinases e RLKs de cana-de-açúcar foram alinhadas usando-se o programa ClustalW (Thompson *et al.*, 1994) com os parâmetros no modo padrão. O alinhamento foi manualmente ajustado de acordo com os subdomínios conservados para proteínas quinases de eucariotos (Hanks *et al.*, 1988, Hanks & Quinn, 1991, Hanks & Hunter, 1995) utilizando-se o editor de alinhamento de sequências Se-Al (Rambaut, 1996). Regiões entre os subdomínios conservados foram descartadas por compreenderem lacunas na maior parte das sequências alinhadas. As sequências que apresentaram menos de 50% do alinhamento total foram descartadas da análise. O alinhamento gerado foi analisado com o programa PAUP (Swofford, 1999) usando-se o algoritmo de agrupamento de vizinhos e distância como critério. Como teste de confiança das árvores geradas, uma análise de *bootstrap* com 500 replicatas foi realizada, com NJ/UPGMA como tipo de busca.

Domínios catalíticos de *pkinese* de outros organismos foram incluídos na análise. Esses foram selecionados principalmente a partir do banco de dados PlantsP (Gribskov *et al.*, 2001).

5.3. Material biológico para a extração de RNA

5.3.1. Amostras de órgãos de cana-de-açúcar

A variedade comercial SP80-3280 (CTC, Brasil) foi utilizada para a coleta de todos os órgãos, com exceção das amostras de flores, para as quais a variedade SP87-342 (CTC, Brasil) foi utilizada, tanto para as amostras para experimentos de microarranjos quanto para as amostras de PCR em tempo real. As amostras foram coletadas no Centro de Tecnologia Canavieira (CTC) em Piracicaba, sob supervisão do Dr. Eugênio Ulian.

- FL-1: Diversas inflorescências imaturas com comprimentos variando de 5 a 30 cm foram coletadas e armazenadas a -80°C. As amostras foram processadas em pequenos pedaços contendo todos os órgãos, congeladas em nitrogênio líquido e armazenadas a -80°C.
- FL-2: Inflorescências imaturas com aproximadamente 50 cm de comprimento foram coletadas, processadas em pedaços de diferentes partes ao longo do comprimento, congeladas em nitrogênio líquido e armazenadas a -80°C.
- LV-1: Porções diferentes ao longo da folha +1 (folha mais jovem com inserção da bainha com o limbo foliar visível) foram coletadas de uma planta de 12 meses crescida em campo, seccionadas em pedaços de 1-2 cm, congeladas em nitrogênio líquido e armazenadas a -80°C.

- LV-2: Porções diferentes ao longo da folha +1 foram coletadas de uma planta de 14 meses crescida em campo, seccionadas em pedaços de 1-2 cm, congeladas em nitrogênio líquido e armazenadas a -80°C.
- LB-1: toletes com apenas um nó (contendo uma gema lateral) foram extraídos de plantas de 12 meses cultivadas no campo. Os explantes foram tratados com fungicida Benlate (Benomyl) 0,6 g/l e inseticida Decis (Deltametrina) e germinados no escuro sobre papel absorvente úmido por 10 dias a 25°C. As gemas desenvolvidas foram coletadas, congeladas em nitrogênio líquido e armazenadas a -80°C.
- LB-2: toletes com apenas um nó (contendo uma gema lateral) foram extraídos de plantas cultivadas no campo durante 14 meses e plantados em copos plásticos de 200 mL com areia branca úmida sem nenhum tratamento. As gemas laterais foram coletadas após 12 dias, congeladas em nitrogênio líquido e armazenadas a -80°C.
- RT-1: toletes com apenas um nó (contendo uma gema lateral) foram extraídos de plantas de 12 meses cultivadas no campo. Os explantes foram tratados com fungicida Benlate (Benomyl) 0,6 g/l e inseticida Decis (Deltametrina) e germinados no escuro sobre papel absorvente úmido por 10 dias a 25°C. As raízes adventícias que se desenvolveram foram coletadas, congeladas em nitrogênio líquido e armazenadas a -80°C.
- RT-2: toletes com apenas um nó (contendo uma gema lateral) foram extraídos de plantas cultivadas no campo durante 14 meses e plantados em copos plásticos de 200mL com areia branca úmida sem nenhum tratamento. As raízes adventícias que se desenvolveram após 12 dias foram coletadas, congeladas em nitrogênio líquido e armazenadas a -80°C.
- IN1-1 (entrenó 1): o primeiro e o segundo entrenós formados a partir do ápice da planta (visíveis após a remoção das folhas jovens) foram coletados de planta de 12 meses

cultivada em campo. A região do nó foi descartada por ser muito fibrosa e os entrenós foram seccionados, congelados e armazenados a -80°C.

- IN1-2 (entrenó 1): o primeiro e o segundo entrenós formados a partir do ápice da planta (visíveis após a remoção das folhas jovens) foram coletados de planta de 14 meses cultivada em campo. A região do nó foi descartada e os entrenós foram seccionados, congelados e armazenados a -80°C.
- IN4-1 (entrenó 4): o quarto entrenó formado a partir do ápice da planta (visível após a remoção das folhas jovens) foi coletado de planta de 12 meses cultivada em campo. O material foi seccionado, congelado e armazenado a -80°C.
- IN4-2 (entrenó 4): o quarto entrenó formado a partir do ápice da planta (visível após a remoção das folhas jovens) foi coletado de planta de 14 meses cultivada em campo. O material foi seccionado, congelado e armazenado a -80°C.
- *Ref Pool*: uma mistura equimolar de RNA das amostras acima descritas (FL-1, LV-1, LB-1, RT-1, IN1-1) foi preparada após a extração individual de cada RNA (para maiores detalhes, vide a seção **Discussão**).

5.3.2. Amostras para a avaliação da variabilidade individual de plantas de cana-de-açúcar no campo

Para a avaliação da variabilidade individual de plantas no campo, cinco amostras de folhas, provenientes de cinco indivíduos, foram coletadas (LV-1, LV-2, LV-a, LV-b, LV-c). A folha +1 foi coletada de uma planta de 12 meses para a amostra LV-1, de uma planta de 14 meses para a amostra LV-2 (ambas plantadas em 2001 e descritas no item anterior) e de plantas de 12 meses de idade para as amostras LV-a, LV-b e LV-c (plantadas em 2002). Todas as folhas usadas nos

ensaios de avaliação de variabilidade individual foram coletadas da variedade SP80-3280. Porções ao longo da folha +1 foram cortadas em pedaços pequenos de 1-2 cm, congeladas em nitrogênio líquido e armazenadas a -80°C.

5.3.3. Amostras de órgãos para os ensaios de PCR em tempo real

- **Raiz:** toletes com apenas um nó (contendo uma gema lateral) foram germinados em temperatura ambiente e colocados sobre papel absorvente úmido, após tratamento com benzamidazol 60%, por 20 dias. As raízes adventícias que se desenvolveram foram coletadas, congeladas em nitrogênio líquido e armazenadas a -80°C.
- **Folha +1 :** a amostra de folha +1 foi coletada de uma planta de 12 meses, cultivada em campo. Diferentes porções da folha +1 foram amostradas, cortadas em pedaços pequenos, congeladas em nitrogênio líquido e armazenadas a -80°C.
- **Entrenó 1:** o primeiro e o segundo entrenós formados a partir do ápice da planta (visíveis após a remoção das folhas jovens) foram coletados de uma planta de 12 meses cultivada em campo. A região do nó foi descartada e os entrenós foram seccionados, congelados e armazenados a -80°C.
- **Entrenó 4:** o quarto entrenó formado a partir do ápice da planta (visível após a remoção das folhas jovens) foi coletado de uma planta de 12 meses cultivada em campo. Os entrenós foram seccionados, congelados e armazenados a -80°C.
- **Gema lateral:** toletes com apenas um nó (contendo uma gema lateral) foram germinados durante oito dias no escuro, em temperatura ambiente e colocados sobre papel absorvente úmido, após tratamento com benzamidazol 60%. As gemas desenvolvidas foram coletadas, congeladas em nitrogênio líquido e armazenadas a -80°C.

Devido à dificuldade de se obter amostras de flor de cana-de-açúcar na época do ano de coleta dessas amostras, essa amostra de RNA foi a mesma usada anteriormente para os experimentos de microarranjos (amostra FL-2). Todas as outras amostras constituem amostras biológicas provenientes de uma terceira coleta de material.

5.3.4. Ensaios para a análise do perfil de expressão em resposta a estímulos ambientais e a tratamentos com fitormônios

Experimentos com cana-de-açúcar foram realizados pelos grupos de pesquisa do Dr. Marcelo Menossi (Universidade Estadual de Campinas, UNICAMP)¹, da Dra. Adriana Hemerly (Universidade Federal do Rio de Janeiro, UFRJ)², do Dr. Marcio de Castro Silva Filho (Universidade de São Paulo, USP)³, da Dra. Sônia Marli Zingaretti (Universidade Estadual Paulista, UNESP)⁴ e do Dr. Antônio Vargas de Oliveira Figueira (USP)⁵. Todos os cultivares de cana-de-açúcar aqui utilizados foram cedidos pelo Dr. Eugênio Ulian, do CTC. Para cada um dos experimentos foram conduzidas três réplicas biológicas, sendo que duas réplicas experimentais foram utilizadas para os experimentos de microarranjos e a terceira réplica foi utilizada para reações de PCR em tempo real. Amostras de RNA foram extraídas de plantas submetidas a diferentes tratamentos, como descrito a seguir:

- **¹MeJA:** mini-toletes de cana-de-açúcar (variedade SP80-3280, CTC, Brasil) foram plantados em copos plásticos de 200 mL em meio comercial (Plantmax, Eucatex) e cultivados em casa de vegetação por 20 dias. As plântulas foram aclimatadas por 48 horas em câmara de crescimento com irradiância aproximada de $70 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, fotoperíodo de 16 horas de luz e temperatura de 26°C. Uma solução de MeJA 100 μM (Bedoukian Research Inc.) foi pulverizada na parte aérea, de forma a cobrir ambas as faces foliares. Para a coleta de órgão foliar, padronizou-se a tomada da segunda folha emitida pela planta. As amostras foram coletadas após 0, 1, 6 e 12 horas de tratamento. Para cada ponto experimental de cada amostra biológica, folhas de seis

plântulas foram coletadas. Após a coleta, as amostras foram submersas em nitrogênio líquido e mantidas a -80° C para posterior extração do RNA total.

- ¹**ABA:** Plântulas da variedade SP80-3280 de cana-de-açúcar (CTC, Brasil), derivadas de culturas axênicas, foram cultivadas *in vitro* com irradiância aproximada de 70 $\mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, fotoperíodo de 16 horas de luz e temperatura de 26°C em meio líquido. Após quatro semanas, a solução de cultura foi acrescida de ácido abscísico (Sigma Chem. Co) para uma concentração final de 100 μM . Amostras de folhas de seis plantas (para cada uma das réplicas biológicas) foram coletadas após 0 h (controle), 30 minutos e 1, 6 e 12 horas de tratamento. Após a coleta, as amostras foram submersas em nitrogênio líquido e mantidas a -80° C para posterior extração do RNA total.

- ²**Interação com endofíticos:** Plantas da variedade SP 70-1143 (CTC, Brasil), de associação eficiente com bactérias diazotróficas endofíticas e elevada BNF (Urquiaga *et al.*, 1992), foram obtidas livres de microrganismos através de cultura de meristemas de acordo com o método de Hendre *et al.* (1983). As matrizes foram multiplicadas *in vitro* e induzidas a enraizar (aproximadamente 15 dias). Após enraizamento, as plantas foram inoculadas com uma mistura das bactérias *Herbaspirillum seropedicae* (estirpe HRC54) e *Herbaspirillum rubrisubalbicans* (estirpe HCC103) ou com *Gluconacetobacter diazotrophicus* (estirpe PAL5) como descrito (James *et al.*, 1994) com 0,1 mL de suspensão de bactérias a 10^6 - 10^7 células/mL. Os controles foram inoculados somente com meio. Todas as plantas foram mantidas a 30°C com irradiância de 60 $\mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, fotoperíodo de 12 horas de luz. Um dia após a inoculação, o material foi esterilizado superficialmente e avaliado quanto ao grau de colonização pelas bactérias diazotróficas endofíticas pela estimativa do Número Mais Provável, de acordo com o método de Reis *et al.* (1994). Para cada amostra biológica, cinco plântulas foram coletadas. Após a coleta, as amostras foram submersas em nitrogênio líquido e mantidas a -80° C para posterior extração do RNA total.

- ³**Herbivoria com lagartas (*Diatraea saccharalis*):** Plantas da variedade comercial SP80-3280 (CTC, Brasil), suscetível a *Diatraea saccharalis* (Barsalobres, 2004), foram obtidas através de propagação vegetativa de entrenós e cresceram em copos plásticos por dois meses em estufa. Lagartas de *D. saccharalis* com 10 dias de idade (2º ínstar) (gentilmente cedidas pelo Dr. Enrico Arrigoni, CTC) foram alimentadas com dieta artificial à base de farelo de soja e germe de trigo e mantidas a 25°C, 60±10% de umidade e fotoperíodo de 14/10 horas de claro/escuro até 18 horas antes da montagem do experimento, quando as lagartas foram transferidas para um frasco, onde permaneceram em jejum. A montagem do ensaio biológico foi realizada em câmaras climáticas, a 28°C, 14/10 horas de claro/escuro e 70% de umidade relativa. Com base em resultados anteriores (Barsalobres, 2004), optou-se pela colocação de duas lagartas de 2º ínstar em cada planta em plantas de dois meses de idade de cana-de-açúcar. A coleta do material foi realizada nos tempos de 30 minutos e 24 horas após o início do experimento. Após a transferência às plantas, as larvas foram observadas por um período de duas horas para a observação dos primeiros sinais de herbivoria (folhas ou bainhas raspadas e presença de fezes). Os controles foram mantidos sem serem atacados pelas lagartas. Decorrido o tempo de cada tratamento, as plantas foram cortadas na zona de transição raiz/colmo. Toda a parte aérea de cada planta foi coletada, congelada em nitrogênio líquido e armazenada a -80°C, até o momento de sua utilização para a extração do RNA total. Material de dois indivíduos foi utilizado para cada ponto experimental de cada uma das réplicas biológicas.

- ⁴**Seca:** O experimento foi instalado em casa de vegetação com condições controladas (28°C, 14/10 horas de claro/escuro e 56±4% de umidade relativa). Foram utilizados toletes da variedade SP90-1638 de cana-de-açúcar (CTC, Brasil), suscetível à seca (de acordo com o Relatório Técnico Interno -RTB, CTC, 2002). Os toletes foram cultivados, durante quinze dias, em bandejas com areia grossa estéril e regas periódicas com água. Após esse período, as plântulas foram transplantadas para vasos de dez litros contendo areia grossa peneirada esterilizada e irrigadas com solução nutritiva de Hogland (Hoagland & Arnon, 1950) por 90 dias em condições de rega controladas por um atmômetro (atmômetro de Livingstone), aparelho que simula a

evapotranspiração das plantas e permite o controle da quantidade adequada de água para irrigação (Broner & Law, 1991). A parte aérea das plantas de cana-de-açúcar foi coletada após 24, 72 e 120 horas de supressão de rega. Para cada um dos pontos experimentais foram coletadas também amostras referentes aos controles (plantas com rega continuada). Como controle do experimento, amostras de solo de cada vaso foram coletadas e o peso úmido do solo foi comparado com seu peso seco para a certificação da perda de água nos solos das plantas experimentais. As amostras foram congeladas imediatamente, em nitrogênio líquido, e depois armazenadas em freezer (-80°C) até o processamento. Material de seis indivíduos foi utilizado para cada ponto experimental de cada uma das réplicas biológicas.

- **⁵Deficiência em fosfato:** O material vegetal foi originado a partir de culturas *in vitro* de meristemas de cana de açúcar da variedade SP80-3280 (CTC, Brasil). As plântulas com estabelecimento do sistema radicular foram aclimatadas pelo cultivo inicial em solução nutritiva de Hoagland & Arnon (1950) 1/20x. A solução nutritiva foi trocada a cada sete dias, aumentando-se a concentração dos nutrientes até a concentração de 1/4x em 3 semanas. As plantas foram então transferidas para potes de 2,8 L com solução nutritiva 1/4x. Após uma semana, as plantas foram divididas em dois grupos: o grupo controle recebeu solução nutritiva fresca com 250 µM Pi, enquanto o grupo experimental foi transferido para solução nutritiva desprovida de fosfato, onde permaneceu durante todo o experimento, sendo H₂PO₄ substituído por H₂SO₄ (Muchhal *et al.*, 1996). Amostras de raiz de dois indivíduos por ponto experimental (por réplica biológica) foram coletadas após 6, 12, 24 e 48 horas (indivíduos diferentes para cada coleta), tanto para o grupo controle quanto para o grupo experimental.

5.3.5. Amostras de genótipos de cana-de-açúcar de alto e de baixo brix

Os cruzamentos que deram origem aos genótipos de alto e baixo brix foram realizados no CTC no período de 1994 a 1998. Dois cruzamentos intra-específicos foram conduzidos inicialmente, um entre 21 genótipos de *Saccharum officinarum* e outro combinando 13 genótipos

de *Saccharum spontaneum*. Posteriormente, as progênies desses cruzamentos independentes foram avaliadas quanto ao acúmulo de sacarose pela medição de brix e os indivíduos mais extremos foram cruzados entre si. Uma série de eventos de recombinação e seleção foi realizada para selecionar oito genótipos de alto brix (AB) e oito com baixo brix (BB) a partir da progênie F3 sempre selecionando de cada geração os oito indivíduos extremos. Quinhentos indivíduos tiveram seu teor de sacarose avaliado em cada geração. Os 16 genótipos extremos em brix da F3 foram plantados no campo em 2001. As leituras de brix e as coletas de folhas +1, entrenós 1, 5 e 9 ocorreram em março de 2002, quando as plantas estavam com 10 meses de idade. Para as leituras de brix, um coletor de caldo dos colmos foi utilizado. O caldo foi coletado fazendo-se um buraco no quinto entrenó visível a partir do topo após a remoção das folhas +1. Algumas gotas do caldo foram colocadas em um refratômetro manual (N1, ATAGO) e a leitura direta de brix foi obtida.

5.4. Extração de RNA

As extrações de RNA das amostras de órgãos, para avaliação da variabilidade individual, de genótipos de alto e de baixo brix, de plantas tratadas com fitormônios, submetidas à seca e à deficiência em fósforo foram realizadas pelo grupo do Dr. Marcelo Menossi (UNICAMP). As amostras de RNA de plantas submetidas à herbivoria foram extraídas pelo grupo do Dr. Márcio de Castro Silva Filho (USP). RNA das amostras de plantas inoculadas com bactérias diazotróficas endofíticas foram extraídas pelo grupo da Dra. Adriana Hemerly (UFRJ).

Amostras de órgão congelado (2-2,5 g) foram pesadas e pulverizadas a pó fino, em nitrogênio líquido, usando-se cadinho e pistilo previamente resfriados. O pó foi transferido para um tubo de 50 mL e homogeneizado com 5 mL de Trizol (Life Technologies) por grama de tecido, e o protocolo de extração foi feito de acordo com as instruções do fabricante.

Os *pellets* de RNA foram ressuspensos em 20 µL de água DEPC morna (previamente aquecida a 65°C), vortexando-se gentilmente no decorrer de 15 minutos. As amostras de RNA foram quantificadas mediante a leitura de uma alíquota de 2 µL da amostra diluída em 1 mL de

água tratada com DEPC no comprimento de onda de 260 nm (onde 1 DO equivale a 40 µg de RNA). As amostras de RNA foram analisadas por eletroforese em gel de agarose 1,5% contendo 16% de formaldeído e 10% de tampão MOPS 10x (MOPS 0,2 M, acetato de sódio 20 mM, EDTA 10 mM, pH 7,0). A eletroforese foi corrida em tampão MOPS 1x a 80 V. Para a análise por eletroforese em gel de agarose, às amostras de RNA (3 a 5 µg) foram adicionados 5 µl de tampão de amostra (formamida 64,4%, MOPS 1,3x, formaldeído 2,8%, azul de bromofenol 0,35%, 50 µg/mL de brometo de etídeo em água tratada com DEPC). Após a corrida, o gel foi fotodocumentado com o sistema ImageMaster VDS (Pharmacia Biotech) e as amostras analisadas quanto à quantidade e integridade.

5.5. Manutenção dos clones bacterianos

Os clones de bactérias contendo os cDNAs do projeto SUCEST foram fornecidos pelo BCCCenter. As placas de 96 poços foram armazenadas em freezer a -80°C. Réplicas para manutenção do estoque foram feitas em 100 µL de meio TB (*Terrific Broth*) (bacto-triptona 1,2%, extrato de levedura 2,4%, glicerol 0,4%, KH₂PO₄ 0,017M, K₂HPO₄ 0,072M) e incubadas a 37°C por 14 horas. Após o crescimento, igual volume de glicerol 30% foi adicionado ao meio. Para as amplificações por PCR foram utilizadas as placas originais recebidas do BCCCenter, com meio de cultura contendo 8% de glicerol.

5.6. Amplificação dos insertos de DNA dos clones bacterianos

Os fragmentos de DNA foram amplificados por PCR em placas de 96 poços (Fisher Scientific International Inc.) em um volume total de reação de 100µL. A mistura de reação, contendo MgCl₂ 2 mM, 1 µL de Taq Buffer 10x (Bioline Ltd) (KCl 500 mM, Tris-HCl 100 mM pH 8,8, MgCl₂ 15 mM, Triton 100 1%), dATP 0,2 mM, dGTP 0,2 mM, dCTP 0,2 mM, dTTP 0,2 mM, 1 U de Taq polimerase Biolase (Bioline Ltd.), 40 pmol de *primer* T7 (Life Technologies), 40 pmol de *primer* SP6 (Life Technologies), foi preparada e distribuída nos poços. Utilizando-se o transferidor de

colônias (aplicador manual), duas transferências de clones foram realizadas (cerca de 2 µL) de uma cultura em glicerol previamente descongelada diretamente para o meio reacional distribuído em microplacas de 96 poços para PCR.

A reação de PCR foi feita em um Termociclador DNA Engine PTC-200 (MJ Research) com um programa de 40 ciclos de desnaturação a 95°C por 45 segundos, anelamento a 51°C por 30 segundos e polimerização a 72°C por 2 minutos, finalizado por incubação a 72°C por 10 minutos. Após a reação de PCR, as placas foram armazenadas seladas a -20°C até a purificação dos produtos de PCR.

5.7. Purificação dos produtos de PCR para confecção dos microarranjos

Dez microlitros de *PCR Binding Buffer* - guanidina-HCl 7 M, tampão MES (ácido 2-[N-morfolino]-etanossulfônico) 200 mM, pH 5,6 - foram adicionados em cada poço de uma placa Multiscreen Millipore de 96 poços e a placa foi centrifugada a 3.000 rpm por 1 minuto (Centrífuga Eppendorf 5810R, rotor de placas). Cem microlitros de *PCR Binding Buffer* foram adicionados aos produtos de PCR e os 200 µL obtidos foram transferidos para a placa Multiscreen Millipore previamente umedecida. A placa foi centrifugada a 3.000 rpm por 1 minuto a 25°C. As lavagens foram coletadas em uma placa de 96 poços com fundo em U. Os poços foram lavados duas vezes utilizando-se 200 µL de etanol 80% em cada lavagem. Na primeira lavagem com etanol a placa foi centrifugada a 3.000 rpm por 1 minuto e na segunda, a 3.000 rpm por 10 minutos. Todas as lavagens ocorreram a 25°C. A placa em U foi trocada por uma nova e os produtos de PCR foram eluídos com 50 µL de Tris-HCl 10mM, pH 8,0, com centrifugação a 3.000 rpm por 5 minutos, a 25°C.

As placas contendo os produtos de PCR purificados foram seladas e armazenadas a -20°C até o rearranjo e deposição dos produtos de PCR nas lâminas de vidro.

5.8. Eletroforese em gel de agarose

Uma alíquota de 4 μ L da reação de PCR ou 2 μ L após purificação foram adicionados a 2 μ L ou 1 μ L, respectivamente, de tampão de corrida (azul de bromofenol 2,5 mg/mL, xileno cianol 2,5 mg/mL, glicerol 30%) e a amostra foi submetida à eletroforese em gel de agarose 1% em TAE 1x (Tris-acetato 40 mM, EDTA 1mM, pH 8) com 0,5 μ g/mL de brometo de etídio, a 80 V. O tamanho dos fragmentos foi estimado com base no marcador de tamanho *1 kb DNA ladder* (Life Technologies). Como estimador de concentração foram aplicados 4 μ L (400 ng) de vetor pUC18 pré-digerido com *ScaI*, *NdeI* e *HindIII*, sendo que o fragmento de 692 pares de bases corresponde a aproximadamente 100 ng. O tampão usado para a eletroforese foi TAE 1x. O gel foi fotografado e uma planilha com os resultados das reações de PCR foi preparada com base na eficiência de amplificação e no aparecimento ou não de bandas inespecíficas.

5.9. Deposição dos produtos de PCR em lâminas de vidro

Para a confecção dos arranjos de DNA, os produtos de PCR purificados foram manualmente rearranjados das placas de 96 poços para placas de 384 poços. No processo de rearranjo, quinze microlitros de produto de PCR purificado foram misturados com igual volume de Reagente D (GE Healthcare). Os produtos de PCR foram depositados em lâminas espelhadas tipo 7 Star (GE Healthcare) silanadas, usando o equipamento Generation III Microarranjos Spotter (GE Healthcare). Após a deposição dos fragmentos de DNA, as lâminas foram submetidas a 50 mJ de luz UV e armazenadas em dessecador. Cada produto foi depositado pelo menos duas vezes na mesma lâmina, sendo uma vez em cada uma das metades da lâmina. No processo de deposição dos produtos de PCR em lâminas de vidro, foram também depositadas as amostras do *Lucidea Universal ScoreCard* (GE Healthcare).

5.10. Síntese e marcação dos alvos de cDNA

Para a síntese e marcação dos alvos de cDNA, utilizou-se o *CyScribe Post-Labeling kit* (GE Healthcare).

A reação de anelamento dos *primers* ao molde de RNA foi realizada com 7-15 µg de RNA total, 4 µL de oligo-dT ancorado e água mili-Q tratada com DEPC em volume suficiente para completar o volume de reação para 10,5 µL. No tubo contendo a amostra a ser marcada com o fluoróforo Cy3 foi adicionado 0,5 µL do RNA Cy3 Referência diluído 1:3 e na outra amostra foi adicionado 0,5 µL do RNA Cy5 Teste, também diluído 1:3, ambos do sistema *Lucidea Universal ScoreCard* (GE Healthcare). As amostras foram incubadas a 70°C por 5 minutos, seguidos por 10 minutos em temperatura ambiente.

Para a reação de transcrição reversa, foram misturados 4 µL de tampão CyScript 5x, 2 µL de DTT 0,1M, 1 µL da mistura de nucleotídeos, 1 µL de aminoalil-dUTP e 1 µL de transcriptase reversa CyScript (100 U/µL) em cada um dos tubos contendo as diferentes amostras de RNA. Todos esses reagentes são fornecidos com o *CyScribe Post-Labeling kit* (GE Healthcare). As amostras foram incubadas por 10 minutos em temperatura ambiente, seguindo-se incubação a 42°C por 3 horas.

O mRNA foi hidrolisado pela adição de 2 µL de hidróxido de sódio 2,5 mM aos tubos de reação seguindo-se de incubação a 37°C por 15 minutos e adição de 10 µL de HEPES (ácido 4-(2-hidroxietil)-1-piperazina etanossulfônico) ácido livre 2M.

O cDNA sintetizado foi purificado adicionando-se 120µL de tampão de ligação (acetato de potássio 150 mM, guanidina 5,3 M) e transferindo-se para placa de purificação de 96 poços (*Millipore Multiscreen filter plate*). A placa foi centrifugada a 1.000 g por 1 minuto a 25°C (centrífuga Eppendorf 5810R, rotor de placas) utilizando-se uma placa de 96 poços com fundo em U para coletar as lavagens. Cada poço foi lavado quatro vezes com etanol 80%, com centrifugação a 1.000 g por 1 min a 25°C, exceto a última lavagem, que foi realizada com centrifugação a 3.500 rpm por 5 minutos. O filtrado foi descartado e a placa de coleta trocada por uma nova. O cDNA sintetizado foi eluído duas vezes com 45 µL de tampão Tris-HCl 10mM, pH 8,0, com centrifugação de 3.000 rpm por 5 minutos e totalmente seco em *speed-vac* a 30°C. Posteriormente, foi realizada a reação de marcação com os fluoróforos Cy3 e Cy5 e, a partir desse ponto, todos os procedimentos foram realizados sem a incidência de luz direta. O cDNA aminoalil purificado e seco

foi ressuspenso em 15 µL de água mili-Q tratada com DEPC. Uma alíquota de Cy3 (contendo 40 nmol do fluoróforo) e outra de Cy5 do kit foram ressuspenso em 15 µL de bicarbonato de sódio 0,1 M, pH 9, cada uma. Os fluoróforos ressuspenso foram misturados ao cDNA respectivo e o meio foi incubado em temperatura ambiente por uma hora. Para os experimentos de comparação de cultivares contrastantes para teor de brix e para experimentos com amostras de plantas submetidas a estímulos bióticos ou abióticos, esse protocolo sofreu uma pequena modificação. Nesses casos, o cDNA aminoalil seco e purificado foi ressuspenso em 40 µL de bicarbonato de sódio 0,1M, pH 9, e foi utilizado diretamente para solubilizar as alíquotas de Cy3 e de Cy5 incubando-se em temperatura ambiente por duas horas.

Foram adicionados 15 µL de hidroxilamina 4M a cada reação incubando-se em temperatura ambiente por 15 minutos. Os alvos marcados foram purificados em placa de purificação de 96 poços (*Millipore Multiscreen filter plate*) como descrito acima, totalmente secos em *speed-vac* a 30°C e mantidos a -20°C até o momento da hibridização nas lâminas de vidro.

5.11. Hibridização dos alvos marcados nos microarranjos de cDNA

Os alvos marcados e secos foram ressuspenso com 6,75 µL de água mili-Q tratada com DEPC. O alvo marcado com Cy3 foi misturado com o marcado com Cy5, totalizando um volume de 13,5 µL ao qual foram adicionados 13,5 µL de tampão de hibridização (*Microarray Hybridization Buffer* - GE Healthcare) e 27 µL de formamida. As soluções foram incubadas a 92°C por 2 minutos e colocadas em gelo. Os 54 µL de solução foram aplicados em uma das extremidades da lâmina. Uma lamínula foi colocada por cima da lâmina de vidro e o sistema foi incubado a 42°C por 16 horas. Após a hibridização, a lâmina foi lavada uma vez em 1xSSC (NaCl 0,3 M, citrato de sódio 0,03 M, pH 7), 0,2% SDS, por 10 minutos a 55°C, duas vezes em 0,1xSSC, 0,2% SDS (10 minutos a 55°C), uma vez em 0,1xSSC por 1 minuto em temperatura ambiente, mergulhada quatro vezes em água mili-Q e seca imediatamente com jato de nitrogênio. A leitura dos sinais foi feita utilizando-se o *scanner* confocal *Generation III Scanner* (GE Healthcare) com voltagem para PMT

(tubo fotomultiplicador) de 700 ou 750 volts para ambos os canais. Os arquivos obtidos foram nomeados de acordo com o código de barras das lâminas e a PMT utilizada.

5.12. Extração e análise dos dados

A extração dos dados foi feita no programa *ArrayVision* (Imaging Research Inc.) utilizando-se a medida de densidade MTM e subtraindo-se o valor da intensidade de fundo. Foram descartados *pixels* que estavam fora da faixa de ± 4 MADs (desvios absolutos da mediana). Todos os dados e imagens dos experimentos de comparação de seis órgãos de cana-de-açúcar foram armazenados em um banco de dados implementado pelo grupo do Dr. Marcelo Menossi (UNICAMP) na plataforma BASE (Saal *et al.*, 2002). Os dados referentes aos demais experimentos foram armazenados em um banco de dados local.

A normalização dos valores brutos de intensidades foi realizada pelo algoritmo de LOWESS (Yang *et al.*, 2002b). Pontos referentes a clones com falha ou inespecificidade nas reações de PCR, pontos do microarranjo com baixa intensidade ou com sinais saturados, bem como pontos visualmente inspecionados quanto à morfologia alterada ou com sujeira foram retirados da análise.

Os dados normalizados foram visualizados em um gráfico do tipo M x S, onde:

$$M = \log_2(\text{Cy3}/\text{Cy5})$$

$$S = \log_2[(\text{Cy3} + \text{Cy5})/2]$$

sendo que Cy3 e Cy5 correspondem, respectivamente, aos valores normalizados de intensidade para os canais de Cy3 e Cy5 para determinado ponto. Sempre que a amostra referência (controle) foi marcada com Cy3, o cálculo de M foi feito como sendo $M = \log_2(\text{Cy5}/\text{Cy3})$. Dessa forma, valores negativos de M sempre denotam expressão maior na amostra referência e valores positivos, expressão maior na amostra experimental. Por esse motivo, o valor de M também pode ser escrito como:

$$M = \log_2(I_E/I_C)$$

onde I_E e I_C correspondem, respectivamente, aos valores normalizados de intensidade para determinado ponto nas amostras experimental e controle, respectivamente. No caso das comparações entre populações contrastantes para acúmulo de sacarose, as amostras referentes a populações de baixo brix foram tidas como controles. Para as réplicas técnicas (repetições do mesmo cDNA na mesma lâmina), calculou-se a mediana dos valores da razão obtidos.

A determinação de genes diferencialmente expressos foi feita pelo método HT-self (Vencio & Koide, 2005). A partir de hibridizações homotípicas da amostra referência contendo uma mistura equimolar dos RNAs de órgãos, determinaram-se, para diferentes intervalos de intensidade (S), oito intervalos de valores para a razão da intensidade entre os dois canais (M) integrando-se a função densidade de probabilidade para 99,5%. Para cada hibridização de uma amostra de órgão contra a amostra referência (*Ref Pool*), genes com pelo menos 70% dos pontos-réplicas acima ou abaixo dos intervalos de razão determinados foram considerados como diferencialmente expressos, ao passo que genes com 55% dos pontos-réplicas dentro dos intervalos de razão em todas as amostras foram considerados como ubíquos e de expressão homogênea entre os seis órgãos em estudo. Para os experimentos de comparação de populações contrastantes para teores de brix, os genes diferencialmente expressos foram determinados a partir do método HT-self com intervalos de credibilidade de 98% (Vencio & Koide, 2005) usando-se dados de hibridizações homotípicas, como descrito acima, e, similar ao que foi realizado para os dados referentes aos órgãos de cana-de-açúcar, genes com pelo menos 70% dos pontos-réplicas acima ou abaixo dos intervalos de razão determinados foram considerados como mais ou menos expressos na população de alto em relação à de baixo brix, respectivamente. Para os dados de expressão referentes a plantas submetidas a tratamentos com fitormônios ou a estímulos ambientais, a detecção de genes diferencialmente expressos foi feita pelo método de detecção de valores extremos (*outliers*) (Koide *et al.*, 2006) adotando-se intervalos de credibilidade de 80% e genes com pelo menos 60% dos pontos-réplicas acima ou abaixo dos intervalos de razão foram considerados como diferencialmente expressos.

Para a avaliação da variabilidade individual, coeficientes de correlação de Pearson foram calculados entre diferentes amostras de folhas utilizando-se os valores de razão de expressão normalizados das hibridizações das amostras de folha contra *Ref Pool* para 1.280 genes.

A comparação direta dos níveis de expressão entre dois órgãos foi inferida a partir das razões virtuais entre pares de órgãos, utilizando a amostra *Ref Pool* como denominador comum. As razões virtuais, para todos os elementos dos microarranjos, foram calculadas e os dados obtidos foram usados para a construção de um agrupamento hierárquico (UPGMA, distância Euclideana) no programa Spotfire 7.1 software (Spotfire Inc.).

No caso dos experimentos de tratamentos com fitormônios (ABA e MeJA), seca e deficiência em fósforo, os dados de expressão foram agrupados pelo algoritmo de *Self-Organizing Maps*, SOM (Tamayo *et al.*, 1999). Para isso, para cada um desses experimentos, foram selecionados os SAS com coeficiente de correlação igual ou superior a 0,7 entre os perfis de expressão ao longo do tempo obtidos para as duas réplicas biológicas. SAS com pelo menos um valor de M da série temporal inválido foram descartados da análise. Aos valores das medianas das razões de expressão para cada réplica isoladamente foi aplicado o método de normalização com centralização pela média (*mean-centring*), subtraindo-se cada valor da média de expressão ao longo do tempo, e foram calculadas, para cada ponto experimental, as médias entre as duas réplicas biológicas dos valores de expressão normalizados. Esses valores foram utilizados para gerar um agrupamento pelo algoritmo de SOM no programa Spotfire 7.1 software (Spotfire Inc.). A fim de se estimar o número de grupos para serem determinados nas análises de SOM, realizou-se uma Análise de Componente Principal (PCA) previamente à análise por SOM.

5.13. Reações de transcrição reversa

Dois a cinco µg de RNA total foram tratados com a enzima *DNase Amplification Grade* (Life Technologies) nos experimentos de comparação de níveis de expressão em seis órgãos de cana-de-açúcar ou DNaseI (Promega) para os demais experimentos, de acordo com as instruções do fabricante e uma alíquota de 7,5 µL do RNA tratado foi utilizada para a reação de transcrição

reversa usando-se o kit *SuperScript First-Strand Synthesis System for RT-PCR* (Life Technologies). Cada reação de transcrição reversa de 20 µL contém, além do RNA tratado com DNase, 2 µL de tampão de transcrição reversa 10x (Tris-HCl 200 mM pH 8,4, KCl 500 mM), dATP, dGTP, dCTP e dTTP (0,5 mM cada), 50 ng de *primers* hexâmeros randômicos, 0,25 µg de *primer* oligo(dT), MgCl₂ 5 mM, DTT 10 mM, 40 U de *RNase OUT*, 50 U de *SuperScript II Reverse Transcriptase* e água tratada com DEPC em quantidade suficiente para 20 µL. Todos esses reagentes são fornecidos pelo *kit* supracitado. O RNA, hexâmeros randômicos, dNTPs e oligo(dT) foram misturados primeiro, incubados a 70°C por 5 minutos e colocados em gelo. Posteriormente, os demais componentes, exceto a enzima *SuperScript II Reverse Transcriptase*, foram adicionados à reação e a mistura foi aquecida a 25°C por 10 minutos e depois, a 42°C por 2 minutos. A enzima foi, então, adicionada a cada tubo e a reação foi incubada a 42°C por 90 minutos, seguida por 72°C a 10 minutos e imediatamente resfriada em gelo. Uma reação idêntica, mas sem a transcriptase reversa foi feita em paralelo para confirmar a ausência de DNA genômico nas reações de PCR em tempo real (*No Amplification Control*, NAC). O produto de cDNA foi tratado com 2 U de RNaseH (Life Technologies) por 30 minutos a 37°C, seguidos por 10 minutos a 72°C e posteriormente armazenado a -20°C.

5.14. Reações de PCR em tempo real

As reações de PCR em tempo real foram realizadas utilizando-se o *SYBR Green PCR Master Mix* (Applied Biosystems) em um equipamento *GeneAmp 5700 Sequence Detection System* (Applied Biosystems). Os *primers* específicos foram desenhados usando o *Primer Express 2.0 Software* (Applied Biosystems).

As reações de PCR em tempo-real foram realizadas utilizando-se 2 µL do cDNA sintetizado diluído 1:10, *primer* R (*reverse*) 600 nM, *primer* F (*forward*) 600 nM, 12,5 µL de *SYBR Green PCR Master Mix* e água milli-Q autoclavada para um volume final de 25 µL. O ciclo para as reações de PCR em tempo real foram: 50°C por 2 minutos, 95°C por 10 minutos e 40 ciclos de desnaturação a 95°C por 15 segundos seguido de anelamento e extensão a 60°C por 1 minuto. A especificidade

dos produtos amplificados foi analisada pelas curvas de dissociação geradas pelo equipamento. Foram utilizados controles negativos sem adição de cDNA sintetizado para confirmar a ausência de qualquer tipo de contaminante no meio reacional (*No Template Control*, NTC), bem como controles negativos contendo, ao invés de cDNA, o produto da reação de NAC (*No Amplification Control*), diluído 1:10. Para a validação dos dados de resposta a fitormônios, estresses e estímulos ambientais, as reações de PCR em tempo real foram otimizadas para volume total de 20 µL, sendo que as quantidades de todos os seus componentes foram alteradas de forma a manter as proporções descritas acima.

Para a validação dos dados de expressão entre seis órgãos de cana-de-açúcar, a razão entre a quantidade relativa de determinado gene e o gene referência (normalizador) foi determinada utilizando-se o método de $2^{-\Delta\Delta Ct}$ (Livak & Schmittgen, 2001) com modificações. O nível de expressão normalizado foi calculado como

$$L = 2^{-\Delta Ct}$$

sendo $\Delta Ct = C_{T, \text{ alvo}} - C_{T, \text{ referência}}$, para cada órgão. Dois normalizadores foram utilizados: um gene da proteína 14-3-3 e um de poliubiquitina, sendo que a normalizações dos dados para a amostra de raiz usando a 14-3-3 como normalizador e os dados para a amostra de flor tendo a poliubiquitina como normalizador foram descartados (conforme explicado abaixo, no item “Discussão”). Para classificar a distribuição dos níveis de expressão de cada gene entre os diversos órgãos, foi utilizada a medida de entropia (MacKay, 2003):

$$H_{\text{gene}} = \sum p_i \cdot \log_2(1/p_i)$$

Onde $p_i = L_i / \sum L_i$, L_i representa o nível de expressão de um gene no órgão “i”, e as somas são feitas para os seis órgãos.

No caso das comparações dos perfis de expressão entre genótipos de alto e de baixo brix, o mesmo gene de poliubiquitina usado anteriormente foi adotado como normalizador e os níveis de expressão normalizados foram também calculados como:

$$L = 2^{-\Delta Ct}$$

Já para experimentos com amostras de plantas tratadas com fitormônios ou submetidas a diferentes estímulos ambientais, os níveis de expressão foram calculados exatamente como descrito em Livak & Schmittgen (2001).

Para a avaliação da significância estatística das razões de expressão determinadas, um modelo log-normal foi assumido e a probabilidade P da razão de expressão entre amostra experimental e controle ser maior do que um e a probabilidade P dessa razão ser menor do que um foram calculadas para os experimentos de comparação de genótipos contrastantes para acúmulo de sacarose, de tratamentos com fitormônios e de submissão a diferentes estímulos ambientais.

5.15. Reações de sequenciamento

Todos os produtos de PCR purificados e depositados nos microarranjos foram resequenciados para terem suas identidades confirmadas. As reações de seqüenciamento foram realizadas com o Sistema *BigDye Terminator* versão 3.1 (Applied Biosystems), de acordo com as instruções do fabricante. Cada reação foi realizada com 1 µL do produto de PCR purificado e 10 pmol de *primer* T7 (Life Technologies). A reação de PCR foi feita em um Termociclador DNA Engine PTC-200 (MJ Research) segundo um programa de desnaturação inicial a 94°C por 3 minutos seguidos de 35 ciclos de desnaturação a 94°C por 1 minuto, anelamento a 51°C por 1 minuto, polimerização a 72°C por 2 minutos e finalizado por incubação a 72°C por 10 minutos. Após a reação de PCR, os produtos obtidos foram purificados mediante precipitação com etanol. Para isso, a cada uma das reações, contidas em placas de 96 poços (Fisher Scientific International Inc.), foram adicionados 25 µL de etanol absoluto, 1 µg de glicogênio e 1 µL de acetato de sódio (3M, pH 5,2). A reação foi incubada em gelo por 15 minutos seguindo-se de centrifugação a 25°C, 4.000 rpm por 30 minutos (Centrífuga Eppendorf 5810R, rotor de placas). Posteriormente, o sobrenadante foi desprezado e a placa de 96 poços contendo os produtos de PCR foi centrifugada duas vezes (700 rpm) por aproximadamente cinco segundos, na posição invertida, com papel absorvente em baixo, para permitir a retirada de solução do meio. Cinquenta microlitros de etanol

absoluto foram adicionados a cada uma das reações, seguido de centrifugação a 4.000 rpm (25°C, 30 minutos). O sobrenadante foi desprezado e duas centrifugações com a placa de 96 poços invertida foram novamente realizadas, como descrito acima. A placa contendo os produtos de PCR purificados foi deixada aberta, por aproximadamente uma hora, para permitir a evaporação do etanol. Os produtos das reações de sequenciamento purificados e secos foram, então, enviados ao Serviço de Seqüenciamento de DNA no IQ-USP (SSDNA), que utiliza o equipamento ABI PRISM 3100 Genetic Analyzer (Applied Biosystems).

5.16. Avaliação dos níveis de açúcares em colmos de cana-de-açúcar por HPLC

O doseamento de açúcares em colmos de genótipos de cana-de-açúcar provenientes da população de alto e de baixo brix analisadas no presente estudo foi feito por HPLC (cromatografia líquida de alto desempenho), pelo técnico Paulo Rezende de Gouvêa, do CTC.

Como solução padrão, foi utilizada uma amostra contendo 0,05 g de rafinose (Riedel-de Haen), 0,70 g de sacarose (Merck), 0,35 g de glicose (Sigma-Aldrich), 0,35 g de frutose (Sigma-Aldrich), 0,15 g de glicerol (Sigma-Aldrich), 0,70 g de etanol (Merck) e 48 mL de água milli Q, filtrada em membrana de acetato/nitrato de celulose de 0,45 µm e 47 mm de diâmetro (Schleicher & Schuell) e conservada em geladeira até o momento de uso. Todos os reagentes utilizados na preparação da solução padrão são de grau analítico. Como fase móvel, foi utilizada solução de NaOH 0,0005 N, filtrada em membrana de acetato/nitrato de celulose de 0,45 µm e 47 mm de diâmetro (Schleicher & Schuell).

A cromatografia ocorreu em Sistema HPLC constituído de bomba LC-10AD (Shimadzu), injetor manual 7125 (Rheodyne), detector de índice de refração RID-6A (Shimadzu), forno para colunas CTO-10A (Shimadzu), sistema de aquisição de dados (EZChrom, versão 6.7), coluna analítica KS-801 (8 x 300 mm) e pré-coluna KS-801 (6 x 50 mm) (Shodex). Como parâmetro das condições da HPLC, foi utilizado um fluxo de solvente de 0,4 mL/min, 20 µL de volume de injeção, 8x de atenuação do detector e temperatura da coluna de 40°C.

A solução padrão foi injetada, os cromatogramas foram obtidos e as áreas de cada pico foram determinadas. Para cada um dos componentes da solução padrão, foram calculadas as médias das áreas dos picos e o respectivo fator de resposta pela fórmula:

$$Fr_i = \frac{A_i}{C_i}$$

sendo Fr_i o fator de resposta do componente i , A_i a média das áreas dos picos do componente i e C_i , a concentração do componente i na solução padrão.

O entrenó 9 contado a partir do ápice da planta (visível após a remoção das folhas jovens) foi coletado, seccionado, congelado em nitrogênio líquido e armazenado a -80°C . Um grama desse material foi pesado e pulverizado utilizando-se cadinho e pistilo previamente resfriados. O pó obtido foi, então, mantido em água Milli-Q durante a noite e, no dia seguinte, a solução aquosa obtida foi coletada. Essa amostra foi diluída 16 vezes (para a determinação de sacarose), duas vezes (para determinação de glicose, frutose e outros sacarídeos) ou quatro vezes (para determinação de glicerol e etanol) com água milli-Q. A amostra diluída foi filtrada em membrana de acetato/nitrato de celulose de $0,45\ \mu\text{m}$ e 25 mm de diâmetro, com pré-filtro de 22 mm (Schleicher & Schuell), injetada no cromatógrafo e as áreas dos picos obtidos com o cromatograma foram determinadas. A concentração de cada componente na amostra foi calculada como:

$$C_i = \frac{A_i \times f}{Fr_i}$$

sendo que C_i representa a concentração do composto i na amostra (em % g/g), A_i é a área do pico do composto i na amostra; Fr_i representa fator de resposta do composto i e f , o fator de diluição da amostra.

6. Resultados

6.1. Prospecção de SAS relacionados a transdução de sinais em cana-de-açúcar

6.1.1. O Catálogo SUCAST

A resposta das plantas a sinais exógenos e endógenos é orquestrada por inúmeros componentes celulares de detecção e transdução através do citoplasma e núcleo, onde disparam respostas metabólicas apropriadas. As vias de sinalização coordenam o crescimento e o desenvolvimento, bem como respostas de defesa a estresses e a patógenos. O projeto SUCAST surgiu com objetivo de criar um catálogo da transdução de sinal para a cana-de-açúcar, bem como de caracterizar os componentes identificados (Souza *et al.*, 2001; Papini-Terzi *et al.*, 2005). Partindo-se das sequências geradas pelo Projeto SUCEST (Vettore *et al.*, 2003) foram catalogados genes de cana-de-açúcar possivelmente relacionados com transdução de sinais utilizando-se a ferramenta BLAST (Altschul *et al.*, 1997). Para isso, sequências de outros organismos relacionados a componentes de transdução de sinais, como receptores, adaptadores, proteínas G, pequenas GTPases, membros do sistema de duplo-componente, proteínas quinases, proteínas fosfatases e elementos da maquinaria de ubiquitinação foram selecionados para as buscas por BLAST contra o banco SUCEST. Aproximadamente 1.500 SAS codificando para proteínas de transdução de sinais e também 616 fatores de transcrição foram anotados no Catálogo SUCAST. Adicionalmente, 758 SAS potencialmente envolvidos em processos disparados por estresses e patógenos ou que possam ter um papel em crescimento e desenvolvimento foram também catalogados. A **Tabela 1** resume as categorias do Catálogo SUCAST, que totaliza, atualmente, 3.521 componentes, representando cerca de 8% dos SAS no banco de dados SUCEST. O Catálogo inclui também 75 SAS para biossíntese de hormônios e 555 SAS sem similaridades a proteínas conhecidas ou sem

Classificação SUCAST	# SAS
<u>Categorias de proteínas:</u>	
Receptores	481
Adaptadores	12
Proteínas G	17
Pequenas GTPases	74
Sistema de duplo componente	20
Ciclase	1
Metabolismo de cálcio	69
Metabolismo de inositol	46
Proteínas fosfatases	108
Proteínas quinases	508
Ubiquitinação	107
Fatores de transcrição	616
Biossíntese de hormônios	75
Proteínas relacionadas a vias de hormônios	22
<u>Categorias funcionais:</u>	
Desenvolvimento	33
Ciclo celular	34
Estresse	307
Patogenicidade	384
<u>Proteínas desconhecidas e genes sem similaridades nos bancos de dados (no matches)</u>	
"No matches"	416
Proteínas desconhecidas	139
<u>Outros*</u>	52
TOTAL	3521

Tabela 1. Categorias do Catálogo SUCAST. A tabela mostra o número de SAS para as categorias que compõem o Catálogo SUCAST.

* Essa categoria inclui genes não relacionados à transdução de sinal, mas que estão também presentes no banco SUCAST.

homologias (*no matches*) nos bancos de dados públicos, categorias essas selecionadas em virtude do interesse de se associar função a novos genes.

Com base em análise de sequências, infere-se que 13% dos genes de *Arabidopsis* estão envolvidos em transcrição ou transdução de sinais (Arabidopsis Genome Initiative, 2000). A anotação automática dos dados do SUCEST indicaram que 13,6% dos SAS pertencem a essas categorias (Vettore *et al.*, 2003).

6.1.2. O quinoma de cana-de-açúcar

As proteínas quinases constituem uma família de grande importância, já que muitos componentes de cascatas de sinalização celular são regulados mediante reações de fosforilação/desfosforilação. Representantes do quinoma de cana-de-açúcar são encontrados nas categorias “Proteínas quinases”, “Receptores” e também “Patogenicidade”, visto que algumas proteínas quinases atuam como genes de resistência, tendo um importante papel para o reconhecimento de patógenos. No total, o Catálogo SUCAST identificou 1.031 proteínas quinases. A categoria de RLKs é a mais abundante entre os receptores, representando 93,5% de todos os receptores encontrados.

Devido ao alto nível de similaridade nos domínios catalíticos de famílias de proteínas quinases, uma análise da presença, arquitetura e distribuição de domínios adicionais é necessária já que esses frequentemente determinam a que família pertence determinada proteína. Sendo assim, as sequências de proteína preditas para os 1.031 SAS referentes ao quinoma de cana-de-açúcar foram utilizadas para se conduzir buscas contra o banco de dados Pfam (Sonnhammer *et al.*, 1998) e SMART (Schultz *et al.*, 1998). A **Figura 1** indica a distribuição de domínios (ou combinação de domínios) Pfam e SMART para o quinoma de cana-de-açúcar. Os números no gráfico referem-se ao número de ocorrências para cada caso. A **Figura 1** mostra o predomínio de SAS contendo domínio de *pkinese* (domínio catalítico de proteína quinase), de SAS com LRRs e com EF-hand. Para 77 SAS nenhum domínio Pfam foi indicado na busca. Alguns dos SAS

Figura 1. Domínios (ou combinação de domínios) Pfam e SMART encontrados para as sequências de proteína preditas pelo ESTScan de SAS representando o quinoma da cana-de-açúcar. Os números no gráfico referem-se ao número de ocorrências para cada caso.

anotados, embora tenham resultados de BLAST que permitem a sua anotação como proteína quinase, não têm ORF (fase aberta de leitura) predita pelo ESTScan (Iseli *et al.*, 1999), ferramenta empregada pelo Projeto SUCEST para a predição da sequência de proteína a partir da sequência consenso do SAS. Por isso, esses SAS não estão contabilizados na análise de busca de domínios conservados. É importante notar também que a anotação baseou-se em primeira instância em similaridades via BLAST. Sendo assim, é possível, por exemplo, que existam SAS anotados como proteínas quinases e que não tenham o domínio de *pkinese* (domínio catalítico de proteína quinase) detectado pelo Pfam. A sequência, nesses casos, apesar de incompleta, possui homologia suficiente com proteínas quinases já caracterizadas em outras regiões além do domínio catalítico.

No processo de anotação das proteínas quinases, não foi possível uma categorização por similaridade com membros de famílias conhecidas através de buscas por BLAST ou de domínios conservados para 39% dos SAS anotados nessa categoria. Sendo assim, esses SAS foram anotados como proteína quinase ou RLKs indefinidos. Para melhor categorizar o quinoma da cana-de-açúcar, especialmente para as proteínas quinases e RLKs indefinidos, uma abordagem filogenética foi adotada para os componentes catalogados que apresentavam domínio de *pkinese* predito pelo Pfam. Proteínas quinases de cana-de-açúcar (277) e RLKs (250) contendo um domínio de *pkinese* predito e proteínas quinases e RLKs de outros organismos (156) foram selecionados, alinhados e categorizados através do algoritmo de agrupamento de vizinhos. Como é sabido que os RLKs e as RLCKs de plantas formam uma família monofilética em relação às demais proteínas quinases (Shiu & Bleeker, 2001 b), duas árvores filogenéticas foram construídas: uma para as proteínas quinases, incluindo apenas alguns representantes da categoria de RLKs e RLCKs e outra apenas para os RLKs e RLCKs. Os homólogos aos genes-R Xa21 (Song *et al.*, 1995) e Rpg1 (Brueggeman *et al.*, 2002), por serem genes-R com domínio de *pkinese* altamente semelhante ao dos RLKs, foram também incluídos nessa análise.

A **Figura 2** mostra as árvores de agrupamento de vizinhos obtidas para as proteínas quinases e para os RLKs/RLCKs. As proteínas quinases se apresentaram distribuídas em seis

Figura 2. Análise filogenética de proteínas quinases e RLKs/RLCKs de cana-de-açúcar. Árvores esquemáticas obtidas para proteínas quinases (**A**) e exclusivamente para RLKs/RLCKs (**B**). Os domínios catalíticos preditos pelo Pfam foram alinhados e usados para a construção de árvores pelo algoritmo de agrupamento de vizinhos. Apenas alguns representantes da categoria de RLKs/RLCKs estão incluídos na árvore mostrada em **A**. Sequências de proteínas quinases, RLKs e RLCKs de outros organismos foram também incluídas nas análises. Os principais componentes de cada grupo são KA: RLKs e RLCKs, KB: ATN1, CTR1, EDR1 e similar a Raf, KC: caseína quinase I, KD: MAPKKK, Ste20 e MAPKK, KE: CDPK, PPCK, SnRK e S6 quinase, KF: GSK3/shaggy, caseína quinase II, MAPK, CDK e LAMMER, RA: RLCKs e SERK, RB: receptores-S e RLKs com domínio de L-lectina, RC: WAK e RLKs com LysM, RD: CLV1, Erecta, BRI1 e LTK1.

grupos principais, denominados KA a KF, sendo que um deles (KA) é composto por RLKs e RLCKs. Os principais componentes dos outros grupos são: ATN1, CTR1, EDR1 e similar a Raf para grupo KB, caseína quinase I para o grupo KC, MAPKKK, STE20 e MAPKK para o grupo KD, CDPK, PPCK, SnRK e S6 quinase para o grupo KE, GSK3/shaggy, caseína quinase II, MAPK, CDK e LAMMER para o grupo KF. No caso da árvore filogenética construída para RLKs e RLCKs, quatro grupos foram determinados, sendo alguns de seus componentes: RA: RLCKs e SERK, RB: Receptores-S e RLKs com domínio de L-lectina, RC: WAK, RLKs com LysM, RD: CLV1, ERECTA, BRI1 e LTK1. Os grupos obtidos nessas análises estão em concordância com os previamente definidos por Shiu & Bleecker (2001b) e pelo banco de dados de PlantsP para o quinoma de outras plantas, ainda que o modo de análise seja diferente do método aqui apresentado (Gribskov *et al.*, 2001). As **Figuras 3 e 4** indicam todos os SAS representados nos grupos acima.

Todas as quinases e RLKs catalogados no banco SUCAST foram nomeados, tendo o prefixo “cane” e uma numeração contínua dentro de cada subcategoria (por exemplo, caneMAPK6, para a sexta MAPK de cana-de-açúcar que foi anotada). No caso das proteínas quinases que permaneceram indefinidas, essas foram nomeadas como UPKs (*unclassified protein kinases*) e os RLKs indefinidos, como URLKs (*undefined RLKs*). As proteínas quinases indefinidas e categorizadas pela análise filogenética foram nomeadas de acordo com o grupo filogenético a que pertencem (por exemplo, “caneRLK B1”, para o primeiro receptor indefinido do grupo RB). Todas as famílias filogenéticas constituídas por dois ou mais SAS indefinidos e suportadas por um valor de *bootstrap* superior a 50% foram nomeadas com algarismos romanos e os SAS incluídos nessas famílias foram nomeados de acordo (por exemplo, “canePK-BIII3 para a terceira proteína quinase da família III do grupo KB). As **Tabelas 2, 3 e 4** listam todos os SAS e respectivas nomenclaturas para as categorias de proteínas quinases, RLKs e RLCKs, respectivamente. Com esses critérios, foi possível definir seis famílias constituídas inteiramente por proteínas quinases indefinidas e 32 famílias de RLKs/RLCKs indefinidos. A análise filogenética permitiu a classificação de 32 proteínas quinases indefinidas. O grupo KB contém 16 dessas 32 proteínas quinases. Trinta e quatro RLCKs e 117 RLKs indefinidos foram incluídos na árvores construída para RLKs/RLCKs. O grupo RA

Domínios

ank + pkinase

pkinase + ACT

pkinase + cNMP binding

pkinase + efhand

pkinase + pkinase C

pkinase + UBA + KA1

pkinase

Anotação

Aur	Aurora quinase
AMP	Proteína quinase ativada por AMPc
ATMR_	Similar a AtMRK1
ATN1_	Similar a ATN1
Cas_	Caseína quinase I
Cas2_	Caseína quinase II
CIPK_	CIPK
cdc2_	CDC2
cdk_	CDK
cdpk_	CDPK
CTR1_	CTR1
CYCG_	Quinase associada à ciclina G
EDR1_	EDR1
Ire_	Transdutor do estresse do retículo endoplasmático
G11	Similar à quinase G11A
GSK3_	Similar a GSK3
KIPK_	Quinase que interage com KCBP (KCBP = <i>Kinesin heavy Chain-like calmodulin-Binding Protein</i>)
LAM_	LAMMER quinase
MAPK_	MAPK
MAP2_	MAPKK
MAP3_	MAPKKK
MEK_	MEK1
MHK_	Quinase homóloga à MAK (MAK = <i>Male germ cell-Associated Kinase protein</i>)
NPH1_	NPH1 (fotorreceptor)
NPK1_	NPK1
NRK_	NRK1
OSA_	Quinase ativada por estresse osmótico (<i>Osmotic Stress-Activated kinase</i>)
PDK_	Quinase dependente de fosfatidilinositol 3-fosfato
PKABA_	Quinase induzida por ABA
PPCK_	Fosfoenolpiruvato carboxilase quinase
PRP4_	PRP4 (<i>Pre-mRna-Processing 4</i>)
REK_	REK
S6K_	S6 quinase
shag_	Similar à <i>shaggy</i> quinase
SNF1_	Similar à SNF1 (SnRK1)
SRPK_	Quinase de fatores de <i>splicing</i> da família SR
STE_	STE20
Tous_	Tousled
und_	Quinase indefinida
yakA_	Similar a yakA

Figura 3 (legenda na última página da figura)

KF

Figura 3 (continuação)

KE

KD

Figura 3 (continuação)

KC

KB

Figura 3. Árvore filogenética para proteínas quinases de cana-de-açúcar. Os domínios catalíticos preditos pelo Pfam foram alinhados e usados para a construção de uma árvore pelo algoritmo de agrupamento de vizinhos utilizando-se o programa PAUP. Como teste de confiança das árvores geradas, uma análise de *bootstrap* com 500 replicatas foi realizada. Apenas alguns representantes da categoria de RLKs/RLCKs estão incluídos nessa árvore (em cinza). Sequências de proteínas quinases de outros organismos foram também incluídas nas análises (em itálico). As proteínas quinases indefinidas estão destacadas em vermelho. As famílias formadas por duas ou mais proteínas quinases indefinidas e suportadas por valor de *bootstrap* superior a 50% receberam uma nomenclatura específica. A cor de cada ramo da árvore denota os domínios Pfam encontrados para a sequência de proteína predita para o SAS indicado (como consta no quadro “Domínios”). O nome de cada SAS é precedido por um prefixo, que mostra a sua anotação, como indicado no quadro acima (“Anotação”).

Figura 4 (legenda na última página da figura)

RC

Figura 4 (continuação)

RB

Figura 4 (continuação)

Figura 4. Árvore filogenética para RLKs e RLCKs de cana-de-açúcar. Os domínios catalíticos preditos pelo Pfam foram alinhados e usados para a construção de uma árvore pelo algoritmo de agrupamento de vizinhos utilizando-se o programa PAUP. Como teste de confiança das árvores geradas, uma análise de *bootstrap* com 500 replicatas foi realizada. Sequências de RLKs e RLCKs de outros organismos foram também incluídas nas análises (em *itálico*). Os RLKs e RLCKs indefinidos estão destacadas em vermelho. As famílias formadas por dois ou mais RLKs ou RLCKs indefinidos e suportadas por valor de *bootstrap* superior a 50% receberam uma nomenclatura específica. A cor de cada ramo da árvore denota os domínios Pfam encontrados para a sequência de proteína predita para o SAS indicado (como consta no quadro “Domínios”). O nome de cada SAS é precedido por um prefixo, que mostra a sua anotação, como indicado no quadro acima (“Anotação”).

SAS	Nomenclatura	SAS	Nomenclatura	SAS	Nomenclatura	SAS	Nomenclatura
SCCCLR2003E06.g	cane aurora kinase-1	SCACHR1040D12.g	caneCIPK-15	SCJLRT1014C12.g	caneMAPK-19	SCQSRT1035E03.g	caneShaggy-like kinase-4
SCJLHR1026C05.g	cane aurora kinase-2	SCBFSB1046D04.g	caneCIPK-16	SCEQRT1030A09.g	caneMAPK-2	SCRURT3061E12.b	caneShaggy-like kinase-5
SCQSSST3113H01.g	cane Calcium/calmodulin-regulated-1	SCEZLB1012C07.g	caneCIPK-17	SCRLRZ3040C05.g	caneMAPK-20	SCVPAM2067F05.g	caneShaggy-like kinase-6
SCJLST1025H09.g	cane cAMP dependent-1 (catalytic subunit)	SCACLR1036B06.g	caneCIPK-18	SCCCLB1002F04.g	caneMAPK-21	SCVPLR1049A12.g	caneShaggy-like kinase-7
SCVPCL6042B07.g	cane cyclin G-associated kinase-like protein-1	SCJLRT1023G09.g	caneCIPK-19	SCCCLB1C01H11.g	caneMAPK-22	SCQSLR1040B05.g	caneSnRK1-1
SCEPRZ1009C10.g	cane osmotic stress-activated protein kinase-1	SCUTLR1037E04.g	caneCIPK-2	SCEQRZ3022F11.g	caneMAPK-23	SCJFRZ2032G01.g	caneSnRK1-2
SCQGHR1011E10.g	cane osmotic stress-activated protein kinase-10	SCCCLR2C01G07.g	caneCIPK-20	SCEZAM2031C06.g	caneMAPK-24	SCCCLR1070D11.g	caneSnRK1-3
SCCCST1006B11.g	cane osmotic stress-activated protein kinase-2	SCMORT2103B04.g	caneCIPK-21	SCMCLV1029A11.g	caneMAPK-25	SCCCLR1079G08.g	caneSnRK1-4
SCCCLR1068F10.g	cane osmotic stress-activated protein kinase-3	SCAGLB2046F06.g	caneCIPK-22	SCSBHR1056F04.g	caneMAPK-26	SCJFRZ1007C06.g	caneSnRK1-5
SCEPRZ3044B07.g	cane osmotic stress-activated protein kinase-4	SCBGLB2011A02.g	caneCIPK-23	SCSBSB1095G08.g	caneMAPK-27	SCCCCL3001D04.b	caneSRPK-1
SCSGFL4031C01.g	cane osmotic stress-activated protein kinase-5	SCCCLR1002D12.g	caneCIPK-24	SCVPRT2081F06.g	caneMAPK-28	SCSFSB1072E02.g	caneSRPK-2
SCJFLR1074F04.g	cane osmotic stress-activated protein kinase-6	SCCCSD1001H06.g	caneCIPK-25	SCVPRT2084F07.g	caneMAPK-29	SCUTST3092B09.g	caneSRPK-3
SCCCST1004A07.g	cane osmotic stress-activated protein kinase-7	SCEQRT2030G04.g	caneCIPK-26	SCCCRZ2C04G10.g	caneMAPK-3	SCCCRZ1001B03.g	caneSte20-1
SCEZRZ1013F09.g	cane osmotic stress-activated protein kinase-8	SCSBAD1051C09.g	caneCIPK-27	SCSBAM1084E01.g	caneMAPK-4	SCCCLR1C06E11.g	caneSte20-2
SCMCRZ2088D10.g	cane osmotic stress-activated protein kinase-9	SCSGAM2077H04.g	caneCIPK-28	SCVPCL6042B11.g	caneMAPK-5	SCJFAD1011C07.b	caneSte20-3
SCJLRT2049D02.g	cane tousled-1	SCSGHR1070F12.g	caneCIPK-29	SCCCHR1003D05.g	caneMAPK-6	SCEQRT1028B04.g	caneSte20-4
SCCCLR1079F12.g	cane Wee1 kinase-1	SCCCLR1C05B07.g	caneCIPK-3	SCCCRZ1003E02.g	caneMAPK-7	SCSGCL6072G02.g	caneSte20-5
SCEQRT2099H09.g	caneATMRK1-1	SCBFRZ2018E02.g	caneCIPK-30	SCCCLB1C05H07.g	caneMAPK-8	SCACHR1035F02.g	caneUPK-1
SCCCST2002B08.g	caneATMRK1-2	SCSGRT2066D03.g	caneCIPK-31	SCCCRZ2C04F05.g	caneMAPK-9	SCBGLR1098G03.g	caneUPK-10
SCRULR1020B06.g	caneATN1-1	SCEQRT2027H03.g	caneCIPK-32	SCEZAM2061E04.g	caneMAPKK-1	SCCCL3005C10.b	caneUPK-11
SCCEPAM1020A03.g	caneATN1-2	SCEPAM2012F04.g	caneCIPK-5	SCCCLR1072C05.g	caneMAPKK-2	SCCCCL7038E04.g	caneUPK-12
SCQGLR2017D05.g	caneCDC2-1	SCCORT1001C10.g	caneCIPK-6	SCCCCL4002H05.g	caneMAPKK-3	SCCCFL4001A09.g	caneUPK-13
SCQSAM2099F03.g	caneCDC2-2	SCCCLR1075E10.g	caneCIPK-7	SCACLR2029C10.g	caneMAPKK-4	SCCCLB1C04H10.g	caneUPK-14
SCEPLB1044E07.g	caneCDC2-3	SCEQLB2019B08.g	caneCIPK-8	SCJFRT2058E01.g	caneMAPKK-5	SCCCLB2004H03.g	caneUPK-15
SCVPRT2074E06.g	caneCDC2-4	SCJLJRZ1023H04.g	caneCIPK-9	SCEQAM1041G09.g	caneMAPKKK-1	SCCCLR1C03D05.g	caneUPK-16
SCJFRZ2011A04.g	caneCDC2-5	SCCCRZ2001B02.g	caneCKI-1	SCJFRZ2027G02.g	caneMAPKKK-10	SCCCRZ2001F02.g	caneUPK-17
SCACLR1057G03.g	caneCDC2-6	SCJLJR1073E05.g	caneCKI-10	SCAGCL6013F06.g	caneMAPKKK-11	SCCCRZ2C01F06.g	caneUPK-18
SCUTSB1031H05.g	caneCDC2-7	SCVPAM1055A12.g	caneCKI-11	SCCCLR1069C08.g	caneMAPKKK-12	SCCCST1004B03.g	caneUPK-19
SCCCAM2004F12.g	caneCDC7-1	SCCCCL3001A02.g	caneCKI-12	SCEPRZ1009H06.g	caneMAPKKK-13	SCAGFL1085G05.g	caneUPK-2
SCEPRZ3086D11.g	caneCDK-1	SCCCRZ1001E02.g	caneCKI-13	SCEQRT2025G09.g	caneMAPKKK-14	SCEPAM1018E02.g	caneUPK-20
SCBFRZ1065B09.g	caneCDK-10	SCCCRZ2002F06.g	caneCKI-14	SCSGAM1094H06.g	caneMAPKKK-15	SCEPAM1024G05.g	caneUPK-21
SCCCLR1022H07.g	caneCDK-11	SCEZLR1031H01.g	caneCKI-15	SCQSLR1089G03.g	caneMAPKKK-2	SCEPRT2044B06.g	caneUPK-22
SCCCLR2003F06.g	caneCDK-12	SCEZLR1052F12.g	caneCKI-16	SCBGLR1117D09.g	caneMAPKKK-3	SCEQFL5045G05.g	caneUPK-23
SCQGRZ1042H04.g	caneCDK-13	SCEZRZ1015G02.g	caneCKI-17	SCCCRZ2C01D08.g	caneMAPKKK-4	SCEQHR1081B11.g	caneUPK-24
SCRFRRZ3057H04.g	caneCDK-14	SCUTSD1028F08.g	caneCKI-18	SCJFRT1008D02.g	caneMAPKKK-5	SCEQLB2020F07.g	caneUPK-25
SCACAM2043G01.g	caneCDK-15	SCEQRT2096B11.g	caneCKI-2	SCMCLF5011G03.g	caneMAPKKK-6	SCEQRT1029H06.g	caneUPK-26
SCACHR1037C05.g	caneCDK-16	SCCCLR1C04G08.g	caneCKI-3	SCJFLR1017A05.g	caneMAPKKK-7	SCEQRT2027D07.g	caneUPK-27
SCAGAM2016A11.g	caneCDK-17	SCCCCL3004A09.b	caneCKI-4	SCJLRT1014C03.g	caneMAPKKK-8	SCEQRZ3090A11.g	caneUPK-28
SCBGLR1096C08.g	caneCDK-18	SCCCLR1079B11.g	caneCKI-5	SCRLAM1014B03.g	caneMAPKKK-9	SCEQSB1018H07.g	caneUPK-29
SCCCAM1004C06.g	caneCDK-19	SCVPRZ2038C09.g	caneCKI-6	SCCCST2002E08.g	caneMEKK-1	SCBFFL1144A02.g	caneUPK-3
SCJLAM2094D05.g	caneCDK-2	SCSFHR1042G09.g	caneCKI-7	SCAGFL3021H01.g	caneMHK-1	SCEZAM2059E11.g	caneUPK-30
SCCCLR1075G01.g	caneCDK-20	SCJFRZ2031E11.g	caneCKI-8	SCBGBL5081D07.g	caneMHK-2	SCEZHR1049E12.g	caneUPK-31
SCEPRT2046F01.g	caneCDK-21	SCUTRZ2023G06.g	caneCKI-9	SCCORT2002G11.g	caneMHK-3	SCEZLB1012F09.g	caneUPK-32
SCEQAM2036G12.g	caneCDK-22	SCQSLR1089A12.g	caneCKII-1	SCBGLR1023A01.g	caneMHK-4	SCEZRZ3053B06.g	caneUPK-33
SCJLLR1108G11.g	caneCDK-23	SCCCCL4017C03.g	caneCKII-2	SCEZLR1031G10.g	caneMHK-5	SCJFHR1C05B06.b	caneUPK-34
SCRULR1042C09.g	caneCDK-24	SCQSLR1090A06.g	caneCKII-3	SCVPST1060B01.g	caneNIMA-1	SCJFHR1C07C03.g	caneUPK-35
SCRUSB1062H04.b	caneCDK-25	SCVPAM1058A04.g	caneCKII-4	SCBGLR1023A11.g	caneNIMA-2	SCJFLR1017F04.g	caneUPK-36
SCSBSZ3098A02.g	caneCDK-26	SCVPFL1067B04.g	caneCKII-5	SCEZLR1009B06.g	caneNPK1-1	SCJFRT2056F09.g	caneUPK-37
SCSFAD1072G12.g	caneCDK-27	SCBFLR1083G06.g	caneCTR1-1	SCRFRRZ3056D04.g	caneNPK1-2	SCJFRZ2030B04.g	caneUPK-38
SCUTST3090G10.g	caneCDK-28	SCCCLB1001F06.g	caneCTR1-2	SCCCLR1C01H09.g	caneNPK1-3	SCJFST1011B11.g	caneUPK-39
SCVPRT2081G05.g	caneCDK-3	SCJLRT2049F12.g	caneCTR1-3	SCCCRZ3004E09.g	caneNPK1-4	SCBFFL4113A06.g	caneUPK-4
SCRFLR1055G12.g	caneCDK-4	SCMCSB1108H04.g	caneCTR1-4	SCEPEL6020C02.g	caneNPK1-5	SCJFST1011H06.g	caneUPK-40
SCJFRZ2015D01.g	caneCDK-5	SCCORT1001F08.g	caneCTR1-5	SCMCRZ3067A10.g	caneNPK1-6	SCJLRT1023B10.g	caneUPK-41
SCRLFL1012B10.g	caneCDK-6	SCAGRT2041C12.g	caneCTR1-6	SCEPRZ3045E10.g	caneNRK-1	SCMCAM2081D11.g	caneUPK-42
SCRUAD1131C11.b	caneCDK-7	SCAGRT3047H09.g	caneCTR2-1	SCVPRT2042C11.g	canePK-1	SCMCL6055F11.g	caneUPK-43
SCJFLR1073B02.g	caneCDK-8	SCAGLR1043E12.g	caneEDR1-1	SCJFRZ2007A03.g	canePEPC kinase-1	SCMCLR1125E04.g	caneUPK-44
SCACLR1057B10.g	caneCDK-9	SCCCNR1001B12.g	caneEDR1-2	SCVPRZ2036G11.g	canePEPC kinase-2	SCMCRZ2089F11.g	caneUPK-45
SCRULR1020G01.g	caneCDPK-1	SCJLRT1017B11.g	caneEDR1-3	SCACLR2007G02.g	canePKABA-1	SCMCSZ2061D05.g	caneUPK-46
SCEQRT1033A11.g	caneCDPK-10	SCCCCL6002B11.g	caneEDR1-4	SCRFLR2034A09.g	canePKABA-1	SCMCST1050G05.g	caneUPK-47
SCJLLR1011H04.g	caneCDPK-11	SCRFLR1050F12.g	caneEDR1-5	SCRFLR1034G06.g	canePKABA1-3	SCQCAD1065F02.g	caneUPK-48
SCCCLR1C04F08.g	caneCDPK-12	SCJFRT2055E09.g	caneEDR1-6	SCEPLB1043F10.g	canePKABA1-4	SCQAM2026G09.g	caneUPK-49
SCEPRZ1008C06.g	caneCDPK-13	SCJFRT2060G05.g	caneEDR1-7	SCRFAD1022A03.g	canePK-B1 (with ankryrin repeat)	SCBFFL4115F07.g	caneUPK-5
SCEPRZ1009A12.g	caneCDPK-14	SCRLAM1014D07.g	caneEDR1-8	SCCCCL3001H11.b	canePK-B1I	SCQGFL3059A12.g	caneUPK-50
SCCCCL4011A09.g	caneCDPK-15	SCSBLB1032B03.g	caneEDR1-9	SCCCST1007C11.g	canePK-BI2	SCQGFL4078G11.g	caneUPK-51
SCVPRT2075A12.g	caneCDPK-16	SCJLLR1101B02.g	caneG11A kinase-1	SCCCLR1072F12.g	canePK-BI3	SCQGRZ1039C03.g	caneUPK-52
SCCCLB1C03G07.g	caneCDPK-17	SCEQRT1024B02.g	caneG11A kinase-2	SCCCFL1002E10.g	canePK-BII1	SCQGSB1083A03.g	caneUPK-53
SCMCLR1122G02.g	caneCDPK-18	SCQGLB1029B05.g	caneG11A kinase-3	SCRFLR2021H11.g	canePK-BII2	SCQSLB1049H10.g	caneUPK-54
SCCCLR1C04E07.g	caneCDPK-19	SCCCAM1073G01.g	caneG11A kinase-4	SCJFHR1C03E01.b	canePK-BII3	SCQSRT2036C06.g	caneUPK-55
SCJFL1048C06.g	caneCDPK-2	SCMCLF5011B04.g	caneG11A kinase-5	SCEPRZ1011A12.g	canePK-BII4	SCRFAD1024B11.g	caneUPK-56
SCSGAM2075E08.g	caneCDPK-20	SCEZRZ1012E04.g	caneG11A kinase-6	SCEQAM1041C05.g	canePK-BIIII	SCRFSD1022C09.g	caneUPK-57
SCCCRZ1002A06.g	caneCDPK-21	SCCCLR2C03C04.g	caneGSK3-1	SCJFLR1073E06.g	canePK-BIIII2	SCRLV1026E08.b	caneUPK-58
SCCCLR1C01H12.g	caneCDPK-22	SCEZLB1007B11.g	caneGSK3-10	SCEZLB1005C02.g	canePK-BIIII3	SCRLSB1040B12.g	caneUPK-59
SCACAD1037G07.g	caneCDPK-23	SCJFRZ2025A07.g	caneGSK3-11	SCSGST1068H05.g	canePK-BIIII4	SCBFLB2092A06.g	caneUPK-6
SCBFRZ2019F05.g	caneCDPK-24	SCRLAD1104C08.g	caneGSK3-12	SCJFLR2036E03.g	canePK-BIIII5	SCRUAD1061B11.b	caneUPK-60
SCBGSZ2049D08.g	caneCDPK-25	SCRLFL4029H04.g	caneGSK3-13	SCQSST1038C02.g	canePK-BIIII6	SCRUAD1061G09.g	caneUPK-61
SCEQRT1028A04.g	caneCDPK-26	SCCCCL4001E05.g	caneGSK3-2	SCJLLR1054C03.g	canePK-BIIII7	SCRUAD1133D02.g	caneUPK-62
SCEQRT2099H01.g	caneCDPK-27	SCAGLR1021H12.g	caneGSK3-3	SCQGHR1013B02.g	canePK-BIIII8	SCRULF1023G11.g	caneUPK-63
SCJFLR1013F07.g	caneCDPK-28	SCJLRZ1027E02.g	caneGSK3-4	SCVPCL6042B12.g	canePK-C1	SCRULF4023A02.g	caneUPK-64
SCJFRZ2025A01.g	caneCDPK-29	SCVPRZ2036C02.g	caneGSK3-5	SCJLRT1014H06.g	canePK-C2	SCRURT2012A03.g	caneUPK-65
SCCCRZ2001D07.g	caneCDPK-3	SCCCRZ1C01G06.g	caneGSK3-6	SCBFLR1039G06.g	canePK-D1	SCRURT3063D03.g	caneUPK-66
SCJFRZ2032C05.g	caneCDPK-30	SCACLR2007E01.g	caneGSK3-7	SCCCCL3001B01.g	canePK-D1	SCRURZ3082F01.g	caneUPK-67
SCJLLR1108E02.g	caneCDPK-31	SCCCCL3003G11.b	caneGSK3-8	SCCCCL4005F05.g	canePK-E1	SCSBLF1042D04.g	caneUPK-68
SCJLRZ1019A04.g	caneCDPK-32	SCCCST1002C02.g	caneGSK3-9	SCCCST1008F06.g	canePK-E2	SCSBLF5022F03.g	caneUPK-69

Tabela 2. SAS anotados como codificadores para proteínas quinases em cana-de-açúcar. A tabela lista todos os SAS anotados como pertencentes à categoria de proteínas quinases (excetuando-se os RLKs e as RLCKs) e a nomenclatura atribuída a esses SAS, exatamente como consta no banco de dados SUCAST.

SAS	Nomenclatura	SAS	Nomenclatura	SAS	Nomenclatura	SAS	Nomenclatura
SCACLR2022D11.g	cane PSK receptor-1	SCRULB1060H02.g	caneLRPKm1-2	SCBFLR1046A03.g	caneRLK-DX1 (with LRR)	SCQSLB1049G10.g	caneURLK-44
SCRLLAD1104A11.g	cane PSK receptor-2	SCSZRZ1012E03.g	caneLTK1-1	SCRRFZ3055C09.g	caneRLK-DX2	SCQSR1037E01.g	caneURLK-45
SCRUAD1132E05.g	cane PSK receptor-3	SCCCLR1072F06.g	caneLTK1-10	SCACRZ3035F10.g	caneRLK-DX3	SCQSRZ3038H03.g	caneURLK-46
SCCCRZ1002B07.g	cane RLK with DUF26 domain-1	SCJLST1027G05.g	caneLTK1-11	SCRFFL1030A11.g	caneRLK-DX4	SCQSS1036G03.g	caneURLK-47
SCJFRZ22015H10.g	cane RLK with lectin domain-1	SCQGLF801B604.g	caneLTK1-12	SCJLST1020H09.g	caneRLK-DX11	SCRFAD1119G09.g	caneURLK-48
SCRLS20211B08.g	cane RLK with lectin domain-10	SCQSS1040H09.g	caneLTK1-13	SCCCRZ1002F10.g	caneRLK-DX12 (with LRR)	SCRRFZ3058E06.g	caneURLK-49
SCEQRT2091H06.g	cane RLK with lectin domain-11	SCRFST3142E06.g	caneLTK1-14	SCSGSB1009F02.g	caneRLK-DX13	SCBFRT1071B01.g	caneURLK-50
SCRSLD1010A04.g	cane RLK with lectin domain-12	SCSGLR1045E07.g	caneLTK1-15	SCVPRZ2044C03.g	caneRLK-DX14	SCRLAM1011A01.g	caneURLK-51
SCACSB1123C03.g	cane RLK with lectin domain-13	SCVPRZ3029G09.g	caneLTK1-16	SCCCCL3001G07.b	caneRLK-DX111 (with LRR)	SCRCL6303C05.g	caneURLK-52
SCAGLV1040D10.g	cane RLK with lectin domain-14	SCVPST1062F08.g	caneLTK1-17	SCSFAD1115H03.g	caneRLK-DX12	SCRRLZ3113C11.g	caneURLK-53
SCEPSD1073A02.g	cane RLK with lectin domain-15	SCMCLR1053F05.g	caneLTK1-2	SCVPRZ2043F09.g	caneRLK-DX13	SCRRLZ3115C10.g	caneURLK-54
SCEQRT2028E11.g	cane RLK with lectin domain-16	SCJLRZ1027A04.g	caneLTK1-3	SCQSSB1077D06.g	caneRLK-DX14	SCRUAD1064A08.b	caneURLK-55
SCEQRT2029A12.g	cane RLK with lectin domain-17	SCSFRT2068D10.g	caneLTK1-4	SCCCLB1C03C05.g	caneRLK-DX111	SCRURH1073B02.g	caneURLK-56
SCZLZB2068H05.g	cane RLK with lectin domain-18	SCQSAM2100F06.g	caneLTK1-5	SCCCRT3003D07.g	caneRLK-DX112	SCRURT2005G03.g	caneURLK-57
SCJFAM1069A05.g	cane RLK with lectin domain-19	SCCCCL3002E01.b	caneLTK1-6	SCCCLB1023E12.g	caneRLK-DX1V1 (with LRR)	SCRUSB1064D06.g	caneURLK-58
SCCCST2004D11.g	cane RLK with lectin domain-2	SCCCLR1068G01.g	caneLTK1-7	SCSBAM1086H09.g	caneRLK-DX1V2	SCSFFL4086D11.g	caneURLK-59
SCJFHR1C06H07.b	cane RLK with lectin domain-20	SCCCAM2C04G08.g	caneLTK1-8	SCACCL6011A10.g	caneRLK-DXV1	SCSFRT2067F09.g	caneURLK-60
SCRJLT1014C10.g	cane RLK with lectin domain-21	SCQGLR2010E05.g	caneLTK1-9	SCJFRZ3C07E06.g	caneRLK-DXV2	SCSGAD1027C12.g	caneURLK-61
SCJLRT2051B10.g	cane RLK with lectin domain-22	SCCCRZ3001D06.g	caneNPH1-1	SCCCRT2004A04.g	caneRLK-DXV3	SCSFSB1069E12.g	caneURLK-62
SCRUF1018F09.g	cane RLK with lectin domain-23	SCJFST1017H03.g	caneNPH1-2	SCVPAM2064F02.g	caneRpg1-1	SCSGAD1009D03.g	caneURLK-63
SCRUF11023D11.g	cane RLK with lectin domain-24	SCSBFL1107D07.g	caneNPK15-1	SCJFRZ3C01D07.g	caneRpg1-10	SCSGAM1096E08.g	caneURLK-64
SCRURH1074H12.g	cane RLK with lectin domain-25	SCACAM1070H08.g	canePERK1-1	SCAGLR1064D03.g	caneRpg1-11	SCSGFL4195H09.b	caneURLK-65
SCSBFL1106B05.g	cane RLK with lectin domain-26	SCVPFL1137C07.g	canePERK1-2	SCJFRT1061C08.g	caneRpg1-12	SCSGRT2065D10.g	caneURLK-66
SCSBFL4064G11.b	cane RLK with lectin domain-27	SCMCST1053A06.g	canePERK1-3	SCJFRT1061E08.g	caneRpg1-13	SCSGRZ3059B07.g	caneURLK-67
SCSBSB1096G12.g	cane RLK with lectin domain-28	SCCCRZ2002C04.g	canePERK1-4	SCEQRT2095F09.g	caneRpg1-14	SCSGRZ3062A12.g	caneURLK-68
SCSFAD1106E11.g	cane RLK with lectin domain-29	SCJFRZ1007B02.g	canePERK1-5	SCCCLB1024F02.g	caneRpg1-15	SCUTAM2009A05.g	caneURLK-69
SCVPST1061C08.g	cane RLK with lectin domain-3	SCBGLR1097F07.g	canePERK1-6	SCSBSB1054B08.g	caneRpg1-16	SCUTLR1058A04.g	caneURLK-70
SCSFST1067H02.g	cane RLK with lectin domain-30	SCEQRT2028D04.g	canePERK1-7	SCJLLB2077H11.g	caneRpg1-17	SCVPAM2064B04.g	caneURLK-71
SCSGFL4196H05.g	cane RLK with lectin domain-31	SCAGAM2017G02.g	canePhyA-1	SCSBL1032H07.g	caneRpg1-18	SCBGLR1044E04.g	caneURLK-72
SCSGRZ3059D07.g	cane RLK with lectin domain-32	SCCCCL3080H06.g	canePhyA-2	SCAGAD1077D08.g	caneRpg1-19	SCVPRZ3027C05.g	caneURLK-73
SCVPL6041F01.g	cane RLK with lectin domain-33	SCQSLR1040D12.g	canePhyB-1	SCJLAM1062C02.g	caneRpg1-2	SCBFLB2093C12.g	caneURLK-74
SCEQRT1029G10.g	cane RLK with lectin domain-4	SCCCLR1065C10.g	canePhyC-1	SCCCCL6C05H09.g	caneRpg1-20	SCBFLR1060F03.g	caneURLK-75
SCBGLF4053G06.g	cane RLK with lectin domain-5	SCVPRT2081G11.g	canePhyC-2	SCCCRT2001G03.g	caneRpg1-21	SCBGLB2072D02.g	caneURLK-76
SCSGAM2076H02.g	cane RLK with lectin domain-6	SCRCL6303H07.g	caneRLK-AX1	SCRRFSB1024C11.g	caneRpg1-22	SCBGLR1095F05.g	caneURLK-77
SCBGLB2071E02.g	cane RLK with lectin domain-7	SCJFRT1012F09.g	caneRLK-AX2	SCRLLAD1044C10.g	caneRpg1-23	SCCCCL3001G07.g	caneURLK-78
SCBGR13012C05.g	cane RLK with lectin domain-8	SCBGLR1082B07.g	caneRLK-AX3	SCUTFL1057B05.g	caneRpg1-24	SCCCFL1100F03.g	caneURLK-79
SCEPSD1005C08.g	cane RLK with lectin domain-9	SCCCCL4007D06.g	caneRLK-AX1	SCRFAM2132D06.g	caneRpg1-3	SCCCFL5057E05.g	caneURLK-80
SCCCCL3120C09.g	cane RLK with LysM-1	SCRFLR1012B04.g	caneRLK-AX2	SCRLLAM1005D08.g	caneRpg1-4	SCCCBL1003E03.g	caneURLK-81
SCCCRZ20C0803.g	cane RLK with LysM-2	SCCCLR1072E03.g	caneRLK-AX3	SCACAD1036D05.g	caneRpg1-5	SCCCLR1C05E11.g	caneURLK-82
SCEZAM2033G12.g	cane RLK with LysM-3	SCEPLR1030G10.g	caneRLK-AX4	SCJFST1012A03.g	caneRpg1-6	SCCCCL6005H02.g	caneURLK-83
SCVPRZ2078H09.g	cane RLK with LysM-4	SCQSAM2047D12.g	caneRLK-B1	SCJFST1017E10.g	caneRpg1-7	SCCCLR2C01A11.g	caneURLK-84
SCVPRZ2073H09.g	cane RLK with LysM-5	SCSGST1072B02.g	caneRLK-B2	SCSFHR1043A00.g	caneRpg1-8	SCCCRT3007E07.g	caneURLK-85
SCACAM2044B03.g	cane RLK with LysM-6	SCSZRZ3051D07.g	caneRLK-B3	SCVPL6038A02.g	caneRpg1-9	SCCCRT3010F04.g	caneURLK-86
SCAGLR1021B11.g	cane SHR5 receptor-1	SCJFRT2053D10.g	caneRLK-B4	SCJFRT1009H06.g	caneSERK-1	SCCCST1006D01.g	caneURLK-87
SCSGLR1025B10.g	cane SHR5 receptor-2	SCCCLR1001B10.g	caneRLK-B5	SCJLRZ1027C03.g	caneSERK-10	SCCCST3C01D11.g	caneURLK-88
SCMCAM2082C07.g	cane SHR5 receptor-3	SCEQRT1024D01.g	caneRLK-B6	SCUTLR1037A06.g	caneSERK-11	SCEPAM1051A02.g	caneURLK-89
SCAGLR1070B09.g	cane S-receptor-1	SCJFRZ1006H05.g	caneRLK-B7	SCBGLF5081A12.g	caneSERK-12	SCPEPL3084B12.g	caneURLK-90
SCEPRT2048G02.g	cane S-receptor-10	SCQSHR1022C10.g	caneRLK-B8 (with DUF26 domain)	SCQGLF4075B06.g	caneSERK-13	SCEPRZ3045G05.g	caneURLK-91
SCEZAM1080D04.g	cane S-receptor-11	SCAGLR1070E01.g	caneRLK-B11	SCAGRT2039G10.g	caneSERK-2	SCEQRT1028C09.g	caneURLK-92
SCSGSB1005D04.g	cane S-receptor-12	SCCCLR1C05G04.g	caneRLK-B12	SCSBRZ3117C12.g	caneSERK-3	SCEQRT3020C02.g	caneURLK-93
SCMCRT2088G10.g	cane S-receptor-13	SCJFRZ2033A09.g	caneRLK-B13	SCCCRZ2003C02.g	caneSERK-4	SCCCFL4094H12.g	caneURLK-94
SCEZLB1010B01.g	cane S-receptor-14	SCQSLR1061E06.g	caneRLK-B14	SCJFRZ2028F11.g	caneSERK-5	SCEZHR1086F04.g	caneURLK-95
SCSFAM1076A05.g	cane S-receptor-15	SCVPRZ3028E05.g	caneRLK-B15	SCMCCL6059E05.g	caneSERK-6	SCEZLR1016G11.g	caneURLK-96
SCBFRT1063C06.g	cane S-receptor-16	SCVPRZ3029B03.g	caneRLK-B111	SCMCLR1053B12.g	caneSERK-7	SCEZSB1091D03.g	caneURLK-97
SCMCRT2017E06.g	cane S-receptor-17	SCSZRZ1017D11.g	caneRLK-B12	SCEQRT1029H05.g	caneSERK-8	SCJFLR1073G01.g	caneURLK-98
SCVPST1061D08.g	cane S-receptor-18	SCQGR10140D05.g	caneRLK-B13	SCQGST1031D02.g	caneSERK-9	SCJFRT1009B11.g	caneURLK-99
SCCCLR2C02A04.g	cane S-receptor-19	SCEPRZ3085A05.g	caneRLK-B111	SCCCRZ1001E10.g	caneTMK1-1	SCJFRT2055F11.g	caneURLK-100
SCEZLB1007H09.g	cane S-receptor-2	SCMCST1057H06.g	caneRLK-B112	SCCCLR1C08E01.g	caneTMK1-2	SCJFRZ2027C06.g	caneURLK-101
SCCCRZ2C04A01.g	cane S-receptor-20	SCMCCL6052C09.g	caneRLK-C1	SCCCFL5060E06.g	caneTMKL1-1	SCJFRZ22033D09.g	caneURLK-102
SCACAM1071F11.g	cane S-receptor-21	SCQSRT2031C08.g	caneRLK-C10 (with TPR+U-box)	SCJFFL3C02F01.g	caneURLK-1	SCJFST1010G04.g	caneURLK-103
SCACSD2015H12.g	cane S-receptor-22	SCSBAM1086F04.g	caneRLK-C11	SCCCFL5065A07.g	caneURLK-10	SCJLRZ1021D12.g	caneURLK-104
SCBFSB1048C08.g	cane S-receptor-23	SCEQRT1033H06.g	caneRLK-C2 (with LRR)	SCJLRZ1023H11.g	caneURLK-100 (with LRR)	SCBGLF4055E08.g	caneWAK-1
SCBGLR1115B08.g	cane S-receptor-24	SCRFFL1034D06.g	caneRLK-C3	SCJLST1022A12.g	caneURLK-101 (with LRR)	SCEQAM1042D12.g	caneWAK-2
SCCCAD1001H12.g	cane S-receptor-25	SCQGSB1082E12.g	caneRLK-C4	SCMCAM1099G06.g	caneURLK-102 (with LRR)	SCEZLB1013C12.g	caneWAK-3
SCCCAD1002H08.g	cane S-receptor-26	SCBFAD1045F11.g	caneRLK-C5	SCMCCL6027F11.g	caneURLK-103 (with LRR)	SCMCRT2085F04.g	caneWAK-4
SCCCCL4012A11.g	cane S-receptor-27	SCJFFL3C02F01.b	caneRLK-C6	SCMCRT2085H02.g	caneURLK-104 (with LRR)	SCSGAD1006F09.g	caneWAK-5
SCCCCL4013C01.g	cane S-receptor-28	SCQGLB1029E11.g	caneRLK-C7	SCMCRT3065H07.g	caneURLK-105 (with LRR)	SCVPRZ3029B10.g	caneWAK-6
SCCCRT2002E11.g	cane S-receptor-29	SCVPLB1015A04.g	caneRLK-C8 (with LRR)	SCMCST1054C11.g	caneURLK-106 (with LRR)	SCRURT3061H09.b	caneWAK-15
SCEZLB1010H12.g	cane S-receptor-3	SCJLFL1054A01.g	caneRLK-C9 (with U-box)	SCQGR1045G11.g	caneURLK-107 (with LRR)	SCEPSD2070D10.g	caneWAK-16
SCCCRZ1001H12.g	cane S-receptor-30	SCBFRT3091C08.g	caneRLK-C11	SCQGST1031F08.g	caneURLK-108 (with LRR)	SCBFRT1065C06.g	caneWAK-17
SCEQHR1081E05.g	cane S-receptor-31	SCVPHR1095F07.g	caneRLK-C12	SCQSRT2032B03.g	caneURLK-109 (with LRR)	SCBFRZ2016E08.g	caneWAK-18
SCEQRT2094F06.g	cane S-receptor-32	SCCCHR1004D03.g	caneRLK-C111	SCCCHR1004C04.g	caneURLK-11	SCCCRT3007H04.g	caneWAK-19
SCEZAM2035H01.g	cane S-receptor-33	SCRFLB1056H12.g	caneRLK-C12	SCQSSB1058E04.g	caneURLK-110 (with LRR)	SCCCSD2002D01.g	caneWAK-20
SCJFAD1013F09.b	cane S-receptor-34	SCQSRT2034G10.g	caneRLK-C13	SCQSSB1058F07.g	caneURLK-111 (with LRR)	SCCCRZ3004H02.g	caneWAK-21
SCJFAD1013F09.g	cane S-receptor-35	SCRULB1062B02.g	caneRLK-C14	SCRFFL1033E01.g	caneURLK-112 (with LRR)	SCEPAM1017E02.g	caneWAK-22
SCJFHR1033H11.g	cane S-receptor-36	SCSGHR1066F10.g	caneRLK-C111	SCRLLAM1010E01.g	caneURLK-113 (with LRR)	SCEPAM2053C12.g	caneWAK-23
SCJFRT2058F08.g	cane S-receptor-37	SCEPSD1071A12.g	caneRLK-C112	SCRLLAM2051G05.g	caneURLK-114 (with LRR)	SCEPLB1043G11.g	caneWAK-24
SCMCRT2088C01.g	cane S-receptor-38	SCEZLB1010F11.g	caneRLK-C113	SCRLLF1004B03.g	caneURLK-115 (with LRR)	SCEQRT2027G09.g	caneWAK-25
SCMCSB1108G01.g	cane S-receptor-39	SCCCLR1024A02.g	caneRLK-C114	SCRLLF1005D09.g	caneURLK-116 (with LRR)	SCEQRT2098A02.g	caneWAK-26
SCEQAD1018F07.g	cane S-receptor-4	SCCCLR1072H06.g	caneRLK-C115	SCRLLF4025D06.g	caneURLK-117 (with LRR)	SCEQRT3090D03.g	caneWAK-27
SCMCST1057G07.g	cane S-receptor-40	SCEQRT2097B11.g	caneRLK-C1V1	SCRULB1057B11.g	caneURLK-118 (with LRR)	SCEZSB1092G10.g	caneWAK-28
SCQGHR1013H03.g	cane S-receptor-41	SCJLRT1020H05.g	caneRLK-C1V2 (with U-box)	SCSBLB1035F03.g	caneURLK-119 (with LRR)	SCQGHR1010D01.g	caneWAK-29
SCQSLF3032E04.g	cane S-receptor-42	SCCCLR1048F02.g	caneRLK-D1	SCCCLR1070C04.g	caneURLK-12	SCRUAD1131A12.g	caneWAK-30

Tabela 3. SAS anotados como codificadores para RLKS em cana-de-açúcar. A tabela lista todos os SAS anotados como pertencentes à categoria de RLKs e a nomenclatura atribuída a esses SAS, exatamente como consta no banco de dados SUCAST.

SAS	Nomenclatura	SAS	Nomenclatura
SCJFRT1010C12.g	cane putative SIK protein-1	SCVPCL6063F08.g	canePti1-18
SCBGLR1082H10.g	caneAPK1-1	SCJFRT1059E08.g	canePti1-2
SCVPLR1049G06.g	caneAPK1-10	SCCCCL6004A11.g	canePti1-3
SCBGLR1097F03.g	caneAPK1-2	SCQGAM2111A02.g	canePti1-4
SCEPSD1007E02.g	caneAPK1-3	SCJFRT1059H08.g	canePti1-5
SCJLRT2049A05.g	caneAPK1-4	SCACRZ3109D03.g	canePti1-6
SCMCCCL6056F07.g	caneAPK1-5	SCBFAD1088C06.g	canePti1-7
SCR FAM2071A12.g	caneAPK1-6	SCEPLB1041A07.g	canePti1-8
SCRFL1006D03.g	caneAPK1-7	SCEPRZ3131G08.g	canePti1-9
SCRURT2005C04.g	caneAPK1-8	SCEQRT2026H12.g	caneRLCK-A1
SCSFHR1045D05.g	caneAPK1-9	SCEZRT2024G05.g	caneRLCK-A2
SCBFRZ2046D07.g	caneNAK-1	SCJFST1009G05.g	caneRLCK-A3
SCRULB1058D08.g	caneNAK-10	SCSFST3073C11.g	caneRLCK-A4
SCUTFL3072G07.g	caneNAK-11	SCJFRT1059B08.g	caneRLCK-A5
SCMCCCL6053A02.g	caneNAK-2	SCQSRT2032F03.g	caneRLCK-A6
SCAGHR1019F07.g	caneNAK-3	SCEZAM1079G04.g	caneRLCK-A7
SCBFAM2025E08.g	caneNAK-4	SCRFLB1056D11.g	caneRLCK-AI1
SCCCCL5003G07.g	caneNAK-5	SCCCLR1079G06.g	caneRLCK-AI2
SCJFST1009F12.g	caneNAK-6	SCMCRT2102F12.g	caneRLCK-AI3
SCJFST1048E03.g	caneNAK-7	SCJFST1015D08.g	caneRLCK-AII1
SCJLRT1006H03.g	caneNAK-8	SCJLRZ1026F03.g	caneRLCK-AII2
SCQGRT1039C07.g	caneNAK-9	SCCCHR1003C01.g	caneRLCK-AIII1
SCQGLB1039F03.g	caneNPK15-2	SCMCRT2108E08.g	caneRLCK-AIII2
SCJFRZ3C06G01.b	caneNPK15-3	SCJFST1013G10.g	caneRLCK-AIV1
SCJLRZ1024D02.g	canePBS1-1	SCCCRZ2C01G11.g	caneRLCK-AIV2
SCSGLV1008E10.g	canePBS1-10	SCQGST1034G10.g	caneRLCK-AIV3
SCVPCL6045H11.g	canePBS1-2	SCEZRZ3016C11.g	caneRLCK-AV1
SCCCAM2003A02.g	canePBS1-3	SCBGLR1117G09.g	caneRLCK-AV2
SCEQRT1032A12.g	canePBS1-4	SCSFAD1069F07.g	caneRLCK-AV3
SCR LRZ3039D01.g	canePBS1-5	SCSFBS1071H07.g	caneRLCK-AVI1
SCVPRZ3025A12.g	canePBS1-6	SCACLR1130D02.g	caneRLCK-AVI2
SCEQLB1064E08.g	canePBS1-7	SCCCRT2001E07.g	caneRLCK-AVII1
SCEQRT1027F04.g	canePBS1-8	SCCCCL5071E03.g	caneRLCK-AVII2
SCQSAM2047C10.g	canePBS1-9	SCEZRZ1015D09.g	caneRLCK-AVIII1
SCCCRZ2C03F07.g	canePti1-1	SCJLRT1021D01.g	caneRLCK-AVIII2
SCEQRT1033C04.g	canePti1-10	SCACLR1036G11.g	caneRLCK-D1
SCEZHR1086B05.g	canePti1-11	SCBFFL4117E02.g	caneRLCK-D2
SCEZRT2018D01.g	canePti1-12	SCRLLR1038D11.g	caneRLCK-D3
SCJFRZ2031D10.g	canePti1-13	SCJFRT2059H09.g	caneRLCK-DII1
SCRFFL1027E12.g	canePti1-14	SCSGST1068A11.g	caneRLCK-DII2
SCRLLV1026E08.g	canePti1-15	SCJLLR1101D07.g	caneRLCK-DII3
SCSGAD1009G12.g	canePti1-16	SCMCST1049H07.g	caneRLCK-DIX1
SCSGAM2103B05.g	canePti1-17	SCEZRT2019B07.g	caneRLCK-DIX2

Tabela 4. SAS anotados como codificadores para RLCKS em cana-de-açúcar. A tabela lista todos os SAS anotados como pertencentes à categoria de RLCKs e a nomenclatura atribuída a esses SAS, exatamente como consta no banco de dados SUCAST.

contém a maioria das RLCKs indefinidas (76%) e o grupo RD, a maioria dos RLKs indefinidos (57%).

6.2. Análise do perfil de expressão gênica utilizando a tecnologia de microarranjos de cDNA

6.2.1. O microarranjo SUCAST

Os microarranjos SUCAST foram construídos a partir de clones de bactérias fornecidos pelo BCCCenter. Para as hibridizações apresentadas no presente projeto, duas versões de microarranjos foram utilizadas. Para os experimentos de comparação da expressão gênica entre órgãos de cana-de-açúcar, foi utilizada a versão #1. Essa versão foi construída a partir de produtos de PCR derivados de 1.632 clones de cDNA selecionados a partir do banco SUCAST e representativos de todas as categorias. Para aproximadamente 22% desses clones, não foi possível obter resultados satisfatórios nas reações de PCR (reação de PCR falha ou presença de produtos inespecíficos). Os dados referentes a esses clones foram descartados das análises. Com isso, a versão #1 do microarranjo foi constituída por 1.280 elementos válidos.

Para os experimentos de comparação da expressão gênica entre populações contrastantes para brix e para avaliação da regulação gênica frente a estímulos ambientais e a hormônios, um segundo microarranjo foi construído (versão #2). Essa segunda versão dos microarranjos possui 1.228 elementos válidos em comum com a versão utilizada anteriormente e 317 elementos adicionais, dos quais 229 são representantes do quinoma de cana-de-açúcar. Inicialmente, 1.830 genes únicos foram representados nessa segunda versão dos microarranjos. Para 1.545 desses genes foi possível obter sucesso nas reações de PCR e seus respectivos dados foram levados em consideração nas análises.

Os produtos de PCR depositados nas lâminas de vidro foram seqüenciados para a confirmação das sequências depositadas em cada uma das posições dos microarranjos. Tanto a versão #1 quanto a #2 dos microarranjos tiveram 90% dos fragmentos depositados com sucesso nas reações de sequenciamento. Para os demais fragmentos, a reação de sequenciamento falhou e, por isso, as identidades dessas posições dos microarranjos não puderam ser confirmadas.

Classificação SUCAST	versão #1 de microarranjo	versão #2 de microarranjo
<u>Categorias de proteínas:</u>		
Receptores	181	231
Adaptadores	9	9
Proteínas G	10	8
Pequenas GTPases	36	36
Sistema de duplo componente	8	11
Metabolismo de cálcio	31	30
Metabolismo de inositol	21	20
Proteínas fosfatases	36	41
Proteínas quinases	65	235
Ubiquitinação	41	43
Fatores de transcrição	175	179
Biossíntese de hormônios	30	31
Proteínas relacionadas a vias de hormônios	13	13
<u>Categorias funcionais:</u>		
Desenvolvimento	13	18
Ciclo celular	10	11
Estresse	119	119
Patogenicidade	104	112
<u>Proteínas desconhecidas e genes sem similaridades nos bancos de dados (no matches)</u>	294	295
<u>Outros*</u>	84	103
TOTAL	1280	1545

Tabela 5. Categorias presentes nos microarranjos SUCAST (versão #1 e #2). A tabela mostra o número de SAS para cada uma das categorias presentes nas versões #1 e #2 dos microarranjos.

* Essa categoria inclui também componentes do banco de dados SUCAMET incluídos nessas versões dos microarranjos SUCAST.

A **Tabela 5** apresenta as diferentes categorias de componentes de transdução de sinais presentes na primeira e na segunda versão dos microarranjos. Além de componentes catalogados no banco SUCAST, os microarranjos utilizados também contêm genes relacionados a vias metabólicas. Esses SAS fazem parte de um outro banco de dados, denominado SUCAMET (*Sugarcane Metabolism Database*). Contudo, 93% dos elementos na versão #1 e 93% dos elementos na versão #2 correspondem a categorias do Catálogo SUCAST. É importante ressaltar que, para cada um dos SAS selecionados para serem representados nos microarranjos, optou-se pelo uso do clone SUCEST que gerou o *longest read* (a EST que cobre a maior parte da sequência consenso do SAS selecionado) para representar o SAS nos microarranjos.

Todos os elementos dos microarranjos estão depositados nas lâminas de vidro no mínimo duas vezes. A maior parte dos elementos está depositada quatro vezes em cada lâmina, o que representa um elevado número de réplicas técnicas internas, aumentando a confiabilidade dos resultados obtidos.

Como parte do processo de implementação da tecnologia de microarranjos de cDNA para cana-de-açúcar, foram avaliados dois modos de normalização: o método de LOWESS (Yang *et al.*, 2002b) e de normalização com base nos dados do *Scorecard*, que funcionam como controles internos para a normalização. A comparação entre esses dois tipos de normalização é mostrada na **Figura 5**. Os dados dos experimentos de microarranjos são visualizados em um gráfico do tipo M x S, onde M representa o valor logarítmico (na base 2) da razão de expressão normalizada entre a amostra experimental e a controle e S representa o valor logarítmico (na base 2) da média dos valores de intensidade normalizados nos dois canais (Cy3 e Cy5) para cada um dos pontos dos microarranjos. A normalização pelo *Scorecard* (**Figura 5A**) mostrou-se deficiente, já que os pontos dos microarranjos de um modo geral sofrem um deslocamento em direção a um dos canais, como se a grande maioria dos genes se mostrasse mais expressa em uma das amostras em relação à outra. Já a normalização por LOWESS (**Figura 5B**) não apresentou essa distorção, além de levar em consideração o efeito dependente dos valores de intensidade na dispersão dos pontos (Yang *et al.*, 2002b), como será discutido no item “Discussão”. Sendo assim, optou-se pela adoção do método de LOWESS para a normalização dos dados de microarranjos.

Figura 5. Comparação entre os métodos de normalização. **A-** Normalização tendo o *Scorecard* como referência. **B-** Normalização pelo algoritmo de LOWESS. O gráfico aponta a intensidade (S) do sinal no eixo x e a razão de expressão (M) no eixo y, em escala logarítmica ($y=0$ equivale à razão 1). Os pontos referentes ao *Scorecard* estão destacados em verde. Os pontos dentro do círculo em vermelho são referentes aos pontos de razão 1:1 entre os *spikes* do sistema *Scorecard*
 $M = \log_2(I_E/I_C)$, $S = \log_2[(I_E + I_C)/2]$

A. Hibridizações das amostras de órgãos x referência comum (Ref Pool)

Flor			Gema lateral			Folha			Raiz			1º entrenó			4º entrenó		
Cy3	Cy5		Cy3	Cy5		Cy3	Cy5		Cy3	Cy5		Cy3	Cy5		Cy3	Cy5	
Pool	vs.	FL-1	Pool	vs.	LB-1	Pool	vs.	LV-1	Pool	vs.	RT-1	Pool	vs.	IN1-1	Pool	vs.	IN4-1
Pool	vs.	FL-1	Pool	vs.	LB-1	Pool	vs.	LV-1	Pool	vs.	RT-1	Pool	vs.	IN1-1	IN4-1	vs.	Pool
FL-1	vs.	Pool	LB-1	vs.	Pool	LV-1	vs.	Pool	RT-1	vs.	Pool	IN1-1	vs.	Pool	Pool	vs.	IN4-2
FL-2	vs.	Pool	LB-2	vs.	Pool	Pool	vs.	LV-2	RT-2	vs.	Pool	IN1-2	vs.	Pool			
						LV-2	vs.	Pool									
						LV-a	vs.	Pool									
						LV-b	vs.	Pool									
						LV-c	vs.	Pool									

B. Hibridizações das amostras para avaliação das respostas a tratamento com hormônios e a estímulos ambientais

Tratamento com MeJA			Tratamento com ABA			Herbivoria			Deficiência de fosfato			Gluconacetobacter			Herbaspirillum			Seca		
Cy3		Cy5	Cy3		Cy5	Cy3		Cy5	Cy3		Cy5	Cy3		Cy5	Cy3		Cy5	Cy3		Cy5
0h	vs.	1h	0h	vs.	30min	30min E	vs.	30min C	6h E	vs.	6h C	E	vs.	C	E	vs.	C	24h C	vs.	24h E
0h	vs.	6h	0h	vs.	1h	24h E	vs.	24h C	12h E	vs.	12h C	C	vs.	E	C	vs.	E	72h C	vs.	72h E
0h	vs.	12h	0h	vs.	6h	30m C	vs.	30min E	24h E	vs.	24h C							120h C	vs.	120h E
1h	vs.	0h	0h	vs.	12h	24h C	vs.	24h E	48h E	vs.	48h C							24h E	vs.	24h C
6h	vs.	0h	30min	vs.	0h				6h C	vs.	6h E							72h E	vs.	72h C
12h	vs.	0h	1h	vs.	0h				12h C	vs.	12h E							120h E	vs.	120h C
			6h	vs.	0h				24h C	vs.	24h E									
			12h	vs.	0h				48h C	vs.	48h E									

C. Hibridizações das amostras de genótipos contrastantes para teor de brix

Folha		1º entrenó		5º entrenó		9º entrenó	
Cy3	Cy5	Cy3	Cy5	Cy3	Cy5	Cy3	Cy5
AB	vs. BB	AB	vs. BB	AB	vs. BB	AB	vs. BB
BB	vs. AB	BB	vs. AB	BB	vs. AB	BB	vs. AB

Tabela 6. Hibridizações realizadas com microarranjos de cDNA. As hibridizações das amostras de órgãos (A) foram realizadas contra uma referência comum, constituída por uma mistura de amostras de RNA provenientes dos órgãos em estudo (*Ref Pool*). Já as hibridizações para respostas a estímulos bióticos e abióticos (B) foram conduzidas comparando-se diretamente amostras de plantas controle (para plantas não tratadas ou referentes ao ponto 0 h de tratamento) versus experimental. Em "C" são mostradas as hibridizações para comparações de genótipos contrastantes para teor de brix, onde uma amostra correspondente a RNA de uma população de oito indivíduos de alto brix (amostra AB) foi hibridizada contra uma amostra correspondente a RNA de uma população de oito indivíduos de baixo brix (BB). As hibridizações em "A", referentes às amostras de órgãos de cana-de-açúcar, foram conduzidas com a versão #1 do microarranjo SUCAST, as demais foram realizadas com a versão #2. Laranja e amarelo referem-se às duas amostras biológicas utilizadas em cada experimento. As hibridizações em azul são referentes às amostras de folhas utilizadas somente para a avaliação da variabilidade individual em campo. C = controle, E = experimental

A **Tabela 6** mostra o desenho experimental conduzido para a análise do perfil de expressão dos genes em seis órgãos de cana-de-açúcar, avaliação da expressão gênica frente a diferentes estímulos e comparação de populações contrastantes para acúmulo de sacarose.

6.2.2. Análise da variabilidade individual

Com o propósito de se avaliar a variabilidade da expressão gênica entre indivíduos cultivados em campo, cinco amostras de folhas +1 foram coletadas, sendo três delas referentes a três plantas cultivadas em condições rigorosamente idênticas. Como a propagação da cana-de-açúcar ocorre vegetativamente, espera-se que a variabilidade genética entre os indivíduos seja baixa se as condições ambientais e de crescimento forem iguais.

As folhas foram coletadas em maio de 2002 (LV-1), julho de 2002 (LV-2) e maio de 2003 (LV-a, LV-b e LV-c). Alvos sintetizados a partir dessas amostras foram hibridizados contra a referência comum (*Ref Pool*) para órgãos. Os cálculos de correlação de Pearson par a par mostram uma elevada correlação entre as três amostras individuais coletadas ao mesmo tempo ou dentro de um pequeno intervalo de tempo (0,84 a 0,88) e uma correlação menor entre indivíduos coletados em anos diferentes (0,61 a 0,64) (**Figura 6**). Os resultados implicam baixa variação individual dentro de um mesmo evento de amostragem e entre eventos próximos e indicam que a amostragem de um elevado número de indivíduos para representar uma subpopulação não é necessária.

6.3. Perfil de expressão em seis órgãos de cana-de-açúcar

Para os experimentos de análise de expressão gênica em diferentes órgãos, optou-se por realizar as hibridizações sempre contra uma mesma referência (*Ref Pool*), constituída por uma mistura equimolar de RNA de diferentes órgãos (raiz, flor, folha, entrenó e gema lateral), representativa da “expressão média” dos órgãos estudados. Para cada órgão analisado, foram

Figura 6. Determinação da variabilidade individual. Os perfis de expressão em amostras de folhas foram comparados entre si, par a par, através de correlação de Pearson. Os gráficos mostram os valores de M entre as amostras de folhas LV-1, LV-2, LV-a, LV-b e LV-c. Os valores de correlação estão mostrados em azul. $M = \log_2(I_E/I_C)$.

realizadas pelo menos três réplicas experimentais, sendo uma delas técnica e a outra utilizando uma outra amostra biológica do mesmo tecido, nem sempre coletada em condições idênticas.

Para a determinação dos parâmetros a serem utilizados na identificação dos genes diferencialmente expressos, adotou-se o método HT-self (Vencio & Koide, 2005). Para isso, oito intervalos de M dependentes da intensidade foram determinados baseados em hibridizações homotípicas da amostra *Ref Pool*, para um valor de P igual a 99,5%. Genes com no mínimo 70% de seus pontos-réplicas fora (acima ou abaixo) das faixas de corte determinadas em ambas as amostras biológicas de um mesmo órgão foram considerados como diferencialmente expressos. Por outro lado, genes com mais de 55% de suas réplicas dentro dos intervalos de corte para todas as amostras biológicas utilizadas foram considerados ubíquos. A **Figura 7** indica como foram feitas a determinação dos intervalos de corte e a análise da reprodutibilidade dos dados. Um resumo dos dados obtidos encontra-se na **Tabela 7**. Os maiores níveis de expressão preferencial em dado órgão foram encontrados para ambas as amostras de folhas (8,84% and 7,36%), na amostra de raiz RT-2 (11,78%) e na amostra de entrenó IN4-2 (8,26%). Similarmente, uma elevada proporção de genes mostrou expressão reduzida para determinadas amostras. As amostras de flores apresentaram a maior porcentagem de genes não-expressos ou menos expressos em relação aos demais órgãos em estudo (8,33% and 6,87%). Em média, somente 11% dos genes apresentaram um perfil de expressão variável (com alta variação entre réplicas técnicas).

Seiscentos e oitenta e oito genes mostraram expressão diferencial em pelo menos uma das amostras. Contudo, um gene foi considerado como diferencialmente expresso apenas quando, além de satisfazer todos os requisitos citados no item anterior, apresentou um resultado condizente nas diferentes amostras biológicas para o mesmo órgão. Com isso, foram computados 217 genes diferencialmente expressos em pelo menos um dos órgãos em estudo (**Tabela 8**). A **Figura 8** mostra o agrupamento hierárquico dos padrões de expressão desses genes. Podem ser observados grupos com expressão característica de folhas, raízes, entrenós, bem como grupos menores de genes enriquecidos em flores e gemas laterais. Além dos 217 genes diferencialmente expressos em pelo menos um dos órgãos, o perfil de expressão nos diferentes órgãos para os elementos dos microarranjos revela um total de 153 genes ubíquos e homogeneamente expressos

Figura 7. A. Determinação dos limites de corte para detecção de genes diferencialmente expressos. Dados de quatro hibridizações homotípicas de *Ref Pool* foram computados para se estabelecer os limites de variação randômica nos experimentos de microarranjos. Os gráficos mostram os valores normalizados de M *versus* valores normalizados de S. Oito intervalos foram definidos com valor de credibilidade de 99,5% para diferentes níveis de intensidades. As barras horizontais indicam os limites para valores de razão, ao passo que as linhas verticais delimitam os intervalos de intensidade.

B. Reprodutibilidade dos dados e critério de expressão diferencial. Os gráficos mostram os valores normalizados de M *versus* os valores normalizados de S. Os gráficos mostram exemplos da distribuição dos dados das réplicas técnicas de um gene considerado com expressão órgão-enriquecida (a), ubíquo, ou seja, sem expressão diferencial (b), de baixa reprodutibilidade (c) ou de expressão reduzida em um determinado órgão em particular (d). As barras indicam os limites de corte intensidade-dependentes.

C. Distribuição global dos dados de microarranjos para experimentos de comparação de seis órgãos de cana-de-açúcar. A posição de cada ponto no gráfico refere-se, para cada uma das amostras, à porcentagem das réplicas acima ou abaixo dos limites de corte determinados para cada um dos genes representados nos microarranjos. As áreas “a”, “b”, “c” e “d” contêm dados exemplificados nos gráficos mostrados em B.

$$M = \log_2(I_E/I_C), S = \log_2[(I_E + I_C)/2]$$

Amostras		% de genes acima dos intervalos	% de genes dentro dos intervalos	% de genes abaixo dos intervalos	% de genes com perfil variável	Total de SAS analisados
flores	FL-1	3,19%	78,59%	8,33%	9,89%	1224
	FL-2	5,00%	80,15%	6,87%	7,97%	1179
gemas laterais	LB-1	1,18%	90,10%	3,47%	5,25%	1182
	LB-2	4,61%	77,30%	5,36%	12,73%	1194
folhas	LV-1	8,84%	72,50%	9,98%	8,67%	1222
	LV-2	7,36%	76,05%	9,13%	7,46%	1073
raízes	RT-1	5,52%	80,46%	7,31%	6,71%	1177
	RT-2	11,78%	61,33%	10,46%	16,43%	1205
entrenós	IN1-1	3,91%	86,34%	4,58%	5,16%	1201
	IN1-2	6,20%	69,31%	5,80%	18,69%	1225
	IN4-1	4,40%	88,33%	1,91%	5,36%	1045
	IN4-2	8,26%	54,57%	10,77%	26,40%	1235
Média		5,87%	76,02%	7,05%	11,05%	1180

Tabela 7. Distribuição de genes diferencialmente expressos e ubíquos entre as amostras de órgãos de cana-de-açúcar. Genes com pelo menos 70% de seus pontos-réplicas acima ou abaixo dos limites de corte em ambas as amostras biológicas de um mesmo órgão foram considerados diferencialmente expressos, ao passo que genes com mais de 55% de seus pontos-réplicas dentro dos limites de corte em todas as amostras foram considerados ubíquos. A porcentagem de genes com expressão acima, dentro ou abaixo dos limites de corte está indicada para cada uma das amostras.

Figura 8. Agrupamento hierárquico dos genes de acordo com os seus perfis de expressão em diferentes órgãos de cana-de-açúcar. Duzentos e dezessete elementos com expressão diferencial e consistente nas duas amostras biológicas de pelo menos um dos órgãos em estudo foram agrupados por UPGMA (distância Euclidean). Cada coluna representa uma das amostras biológicas de cada um dos órgãos em estudo. Os valores usados para o agrupamento foram os da mediana de M de cada gene, em cada amostra, em relação à amostra *pool*. Verde indica valor de $M < -1$ (pelo menos 2x mais expresso na amostra referência em relação ao órgão indicado), vermelho indica valor de $M > 1$ (pelo menos 2x mais expresso no órgão indicado do que na amostra referência), preto indica que não houve expressão diferencial. Os componentes de alguns grupos são mostrados como exemplos.

$M = \log_2(I_e/I_c)$

Categoria	#
Ciclo celular	1
Proteínas G	1
Biossíntese de hormônios	1
Vias de duplo componente	1
Metabolismo de inositol	3
Pequenas GTPases	3
Metabolismo de cálcio	4
Ubiquitinação	5
Proteínas quinases	6
Proteínas desconhecidas	9
Respostas a estresses	10
Proteínas fosfatases	11
Patogenicidade	15
Receptores	17
Fatores de transcrição	20
<i>No matches</i>	35
Outros*	11
TOTAL	153

Tabela 9. Categorias de genes identificados como de expressão ubíqua dentre os seis órgãos analisados.

Um SAS foi identificado como de expressão ubíqua quando mais de 55% de seus pontos-réplicas localizaram-se dentro dos limites de corte determinados para todas as amostras analisadas.

* Essa categoria inclui genes não-relacionados à transdução de sinais e genes incluídos no banco SUCAMET.

nos órgãos em estudo. A **Tabela 9** indica as categorias representadas nesse grupo. Os genes de expressão ubíqua e homogênea nos seis órgãos em estudo serão simplificaradamente denominados no decorrer do texto simplesmente como “genes ubíquos”.

A comparação dos níveis de expressão para dado gene entre dois órgãos foi inferida a partir das razões obtidas experimentalmente calculando-se as “razões virtuais” entre pares de órgãos, utilizando a amostra *Ref Pool* como denominador comum. As “razões virtuais” para todas as combinações possíveis entre os órgãos, para todos os elementos dos microarranjos, foram calculadas e os dados obtidos foram usados para a construção de um agrupamento hierárquico. Essa matriz de perfis de expressão é mostrada na **Figura 9** com destaque para a categoria de genes aparentemente mais expressos nas duas amostras de flores analisadas.

Para a validação de alguns dos resultados obtidos para as hibridizações de amostras de diferentes órgãos de cana-de-açúcar, implementou-se a tecnologia de PCR em tempo real. Todas as amostras de RNA de órgãos utilizadas nas reações de PCR em tempo real são diferentes das utilizadas para os experimentos com os microarranjos (exceto para amostras de flores). Foram realizadas buscas utilizando o algoritmo BLAST contra o banco SUCEST utilizando as sequências dos *primers* com o intuito de se verificar a especificidade do anelamento dos *primers* selecionados. Adicionalmente, após as reações de PCR em tempo real, são geradas, para cada uma das reações, curvas de dissociação, sendo que a especificidade do produto de PCR gerado é averiguada pela existência de um pico único nas curvas obtidas.

O grande desafio para as reações de PCR em tempo real para validação dos resultados em seis órgãos de cana-de-açúcar foi o estabelecimento de qual gene adotar como normalizador (referência interna). Inicialmente, foram testados dois normalizadores clássicos: um gene de actina e um de tubulina de cana-de-açúcar. Esses dois genes não estão representados nos microarranjos e foram selecionados por apresentarem uma boa representatividade de suas ESTs nas diversas bibliotecas adotadas pelo Projeto SUCEST, o que poderia ser um indicativo de expressão homogênea nos diferentes órgãos. No entanto, nenhum desses mostrou ser um bom normalizador, já que as curvas de amplificação (log fluorescência x ciclos) geradas a partir de amostras dos diferentes órgãos não se mostraram com boa sobreposição. A **Figura 10A** mostra as curvas

A

Fator de transcrição - Nin
 Adaptador - Proteína 14-3-3
 Fator de transcrição - MADS
 Receptor - caneRLK-BH1
 Resposta à seca e ao frio - Precursor de cisteína protease RD19A
 No match
 Fator de transcrição - MADS *
 Fator de transcrição - MADS *
 No match
 No match
 No match
 Fator de transcrição - MADS *
 Gene- R com NBS/LRR
 Fator de transcrição - MADS **
 Proteína desconhecida
 No match
 Receptor EXS
 Gene- R com NBS/LRR
 Receptor - caneRLK-DVIII1 (com LRR)
 Chaperonina 10
 No match
 Proteína quinase - caneMAPKKK-12
 Fator de transcrição - Similar a Alfin
 Receptor - caneTMK1-2
 Fator de transcrição - Similar a Alfin
 Metabolismo de inositol - 1-Fosfatidilinositol 4-quinase
 Provável proteína do locus S
 No match (possível transcrito não codificador)
 Gene- R com NBS/LRR
 Fator de transcrição - Tub
 Ciclo celular - Relacionado a retinoblastoma
 Receptor - caneRLK-DIV3
 Gene-R - Rpg1
 No match
 Tradução - Pumilio
 Proteína quinase - cane tousled-1
 Metabolismo de inositol - 1-Fosfatidilinositol 3-quinase

B

Figura 9. Matriz de perfis de expressão derivados da comparação de seis órgãos de cana-de-açúcar par a par. A. Visão geral da matriz de expressões virtuais. As razões de expressão normalizadas provenientes das hibridizações de cada amostra *versus* a referência comum (*Ref Pool*) foram usadas para o cálculo das razões virtuais entre cada par de órgãos para todos os genes incluídos nos microarranjos SUCAST. Os perfis de expressão foram agrupados hierarquicamente (UPGMA, distância Euclideana). Adicionalmente, os resultados dos valores de razão obtidos diretamente das hibridizações contra *Ref Pool* são mostrados nas 12 últimas colunas. As áreas em verde indicam expressão menor no órgão indicado do que nos demais ($\leq 0,5x$), as áreas em vermelho indicam expressão maior no órgão indicado ($\geq 2x$) e as áreas em amarelo indicam expressão similar em ambos os órgãos comparados. As áreas em cinza indicam dados de baixa qualidade e que foram excluídos da análise. Cada linha da matriz corresponde a um gene e cada coluna, aos dados de comparação de um par de amostras de órgãos entre si (por exemplo, RT1/FL1). Todas as combinações possíveis entre pares de amostras foram calculadas. **B. Porção ampliada da matriz de razões virtuais indicando genes com expressão preferencial nas amostras de flores.** Os componentes da área ampliada são mostrados, realçando-se o enriquecimento de fatores de transcrição do tipo MADS, como descrito no item “Discussão”. Os quatro fatores de transcrição MADS marcados com asterisco referem-se a fatores selecionados como expressos em menor grau nas raízes ou outros órgãos em relação à referência, sendo que apenas um deles (marcado com **) mostrou-se também expresso preferencialmente em flor de acordo com o método HT-self.

Figura 10. Validação dos dados de expressão em seis órgãos de cana-de-açúcar por PCR em tempo real. A a C – Os valores brutos do $\log_{10}$ fluorescência foram plotados contra o número de ciclos para um gene de tubulina (a), um de poliubiquitina (b) e um gene de proteína 14-3-3 (c). Cada órgão analisado está representado por uma cor diferente. Todas as reações foram realizadas em paralelo e cada reação foi realizada em duplicata. D a K – Exemplos de resultados de PCR em tempo real para alguns dos genes selecionados para a validação. Os histogramas mostram os níveis de mRNA do gene de interesse em relação aos níveis de mRNA do gene de poliubiquitina (PUB) usado como normalizador. Resultados similares foram obtidos utilizando-se o gene de proteína 14-3-3 como referência. RT = raiz, LV = folha, FL = flor, LB = gema lateral, IN1 = 1º entrenó, IN4 = 4º entrenó. O valor de entropia (H) calculado para cada uma das distribuições está indicado na figura. As barras de erro foram calculadas como descrito em Livak & Schmittgen (2001)

obtidas para o gene da tubulina. Como uma alternativa na busca de normalizadores, foram selecionados, a partir dos dados de microarranjos, alguns SAS que se mostraram ubíqua e homogeneamente expressos nos diferentes órgãos. Foram desenhados *primers* para esses SAS e esses foram testados quanto à sua capacidade de servir como gene normalizador para ensaios de PCR em tempo real. Os melhores resultados foram obtidos usando-se *primers* para um gene de poliubiquitina (**Figura 10B**) e para um gene de proteína 14-3-3 (**Figura 10C**). Mesmo esses genes não se mostraram perfeitos como normalizadores: o gene da proteína 14-3-3 parece ser expresso de forma igual em todos os órgãos, exceto nas raízes, onde é menos expresso. Já o gene da poliubiquitina parece ser mais expresso nas flores em relação aos demais órgãos em análise. Contudo, esses genes aparentaram ser melhor normalizadores do que a actina e a tubulina testadas. Com base nisso, foi decidido fazer uso de dois normalizadores concomitantemente, no caso, o gene da proteína 14-3-3 e da poliubiquitina, sendo que a normalizações dos dados para raiz usando a 14-3-3 como normalizador e os dados de flor tendo a poliubiquitina como normalizador são descartados.

Vinte e cinco genes foram selecionados para a validação dos dados de expressão em órgãos, sendo 18 diferencialmente expressos e sete ubíquos. Os genes selecionados, bem como os *primers* desenhados para esses genes encontram-se na **Tabela 10**. Um gene só foi considerado como tendo o seu perfil de expressão validado quando os resultados de PCR em tempo real usando tanto o gene de 14-3-3 quanto o gene de poliubiquitina como referências foram concordantes entre si e com os dados obtidos nos experimentos de microarranjos. Além disso, um gene foi considerado como tendo o seu perfil de expressão validado quando apresentou valor de entropia maior ou igual a 0,9 para genes ubíquos, ou abaixo desse valor para os considerados órgão-enriquecidos. De acordo com esses critérios, 72% dos resultados de PCR em tempo real validaram os dados de expressão obtidos por microarranjos de cDNA. O perfil de expressão para algum desses genes é mostrado na **Figura 10D a 10K**.

6.4. Perfil de expressão frente a fitormônios e a estímulos ambientais

SAS	Categoria	Primer forward	Primer reverse	Valor de entropia (H)
SCACLR1126E09.g	No match	GCTCCTGAACCCGGCCTCTAC	CACACGATAATTTACCTCAGCAA	0,09
SCQGLR1085F11.g	Induzido por seca - deidrina	TCACACCACACAAAAAGTGAAA	GCCATGAATAAGTGTCCCTATCCT	0,41
SCEQRT2100B02.g	Aquaporina (PIP2)	TCCAGTACATGACGCCCTGTT	CACGTGCTTGACACAACACTAACG	0,42
SCBGF1508E07.g	Receptor quinase EXS	TCCGTGTGCTGCGCAT	TCCAGCATAGTTGGCCTCTTC	0,55
SCCCLR2001E10.g	No match	AGCGGCATCGTCAAAAG	CGATCCTCTCCCCAACCAA	0,68
SCBGLR1023G07.g	No match	CAGCGGATCCAATGAACCTT	AGAGTAGCTGTTCACCCAAA	0,77
SCRFLR1012F12.g	Ácido caféico 3- <i>o</i> -metiltransferase	CGGGTTCAAGGCCACCTA	AGGTGCGGTATTTACTTGATGAAT	0,78
SCCCLRT2001H11.g	Pequena GTPase - Arf	GCAGAAACGGAGCCATGGT	GCCAAATTTACGCGAAACGA	0,78
SCRLLR1038H01.g	Biossíntese de ABA - zeaxantina epoxidase	TGCAACTGATGAAGAACGAGTTTC	GACCTTTCCCCCAATTAATAGTAGGT	0,80
SCCCLRT1001E01.g	Biossíntese de jasmonatos- LOX	CGTGTGCTTGGCGAAGAAAG	CGTCGCCAACTCATGGATCT	0,80
SCRFLB2056A06.g	No match	CACCTCGCCGTCATTCTG	AAATCCAGTCCGGCACCAT	0,81
SCCLPAM1024H12.g	Histona desacetilase (similar a HOS1)	GAGAAAGAGGAGACATGGACAA	CATAAGCTCCTCCGCTCCATA	0,81
SCBGLR1023D05.g	Vias de transdução de sinal de genes-R - LSD1	CAAATCAAGTTGCGCATGTAAAC	TCACAGACCTAGCTCCATACTGGTA	0,87
SCCCLR1066D10.g	Pequena GTPase - Arf	GCACTGGTACATCCAGAGCACTT	TGTTGTTGAAAGCCAGTCAAG	0,89
SCEQRT1024E12.g	Biossíntese de SA - fenilalanina amônia-lase	CTTCCAGGGCACTCCCAT	GAGAACTGCGGAAACATGAG	0,91
SCCCRZ1003G02.g	Proteína G - subunidade beta	CCAAAACCACAGGCTTCCA	CAATTCAAATGGCTGTAATGAGGAT	0,92
SCQGLR1085B04.g	Proteína fosfatase - PP2A / subunidade catalítica	CCATCCATTGAGACTCTTTGACAAC	ATTGGACCTTCATGAGGAACTTCT	0,92
SCEQRT1028H06.g	Biossíntese de auxina - nitrilase	GATTGCGTGCTTGCAATTTGT	GGGAGCGGGCAAGATGT	0,94
SCRFLR2034D09.g	No match	TGCTATTTTCAATTTGACATGGAAGT	ACTCTCTGACCTGCTGCTTTATAATGG	0,94
SCCCCL4002C06.g	Proteína fosfatase - PP6 / subunidade catalítica	CCCGGCAAACTAGCTCGATA	AGCAGGTTGGCTGTTTGGTT	0,95
SCCCCL3120C09.g	Receptor com domínio de LysM	CAGCGCCACGACTATGGAT	TCCTGGCAAGGGCGGTGA	0,97
SCEPAM1022H12.g	Citocromo P450 monooxigenase - CYP71A	CCAAACGAAATCAAGGCTATCATC	TTCCATCGCGTGTTGT	0,97
SCEPLB1042D02.g	Fator de transcrição - NAM	CACGTGGCTGCAACACCAT	CGAAACGAAAGAAAGGCAGAA	0,99
SCCCST2001G02.g	Poliubiquitina*	CCGGTCTTTAAACCAACTCAGT	CCCTCTGGTGTAACCTCCATTGG	0,97
SCCCLR1048F12.g	Proteína 14-3-3*	GGTGGCGGCTTGGG	TTTGTTTCGGTTTCAAGTCGATAA	0,97
SCCCRZ1002H03.g	Tubulina	CGACATTGAGAGGCCCAACT	GGGAAGCAGTAAGTGATGAGATCA	
SCCCLR1069D05.g	Actina	CCAGTTCCATTGTACACAAACAAAG	TCTCCGGAATCCGTAGCAAA	

Tabela 10. Genes de cana-de-açúcar órgão-enriquecidos ou ubíquos selecionados para validação por PCR em tempo real, *primers* desenhados e respectivos valores de entropia (H). Os genes de poliubiquitina e de 14-3-3 (marcados com asterisco) foram utilizados como referências para a normalização dos dados. Os genes de tubulina e actina mostrados na tabela foram também testados como normalizadores. Buscas de BLAST foram conduzidas contra o banco de dados SUCEST de forma a garantir a especificidade dos *primers* desenhados. Os valores de entropia aqui mostrados foram calculados utilizando-se o gene de poliubiquitina como normalizador. Resultados similares foram obtidos utilizando-se o gene de 14-3-3 como normalizador (dados não mostrados). Os valores de entropia também são aqui exemplificados para os genes de referência. O valor de entropia para o gene de poliubiquitina foi calculado tendo o gene de 14-3-3 como referência para a normalização dos dados.

As respostas das plantas a estímulos externos dependem da ativação de cascatas de redes moleculares envolvidas na percepção do estresse e na transdução dos sinais, que irão ativar respostas de controle transcricional de genes responsivos a esses estímulos. Para a identificação de genes regulados transcricionalmente em resposta a fatores bióticos e abióticos, plantas de cana-de-açúcar foram submetidas a uma variedade de condições que afetam a produtividade negativamente (seca, deficiência em fósforo, herbivoria por *Diatraea saccharalis*) ou positivamente (interação com bactérias diazotróficas endofíticas *Herbaspirillum* spp. ou *Gluconacetobacter diazotrophicus*). Dado que os fitormônios ABA e MeJA já foram descritos como envolvidos na regulação de respostas das plantas a estresses (Berger, 2002, Howe & Schilmiller, 2002, Devoto & Turner, 2003, Himmelbach *et al.*, 2003, Verslues & Zhu, 2005), a resposta da cana-de-açúcar a esses fitormônios foi também avaliada.

Para contrastar os padrões de expressão gênica e identificar genes diferencialmente expressos, fragmentos amplificados referentes a 1.545 genes foram hibridizados com alvos fluorescentes gerados a partir de amostras de plantas controles e tratadas. As hibridizações foram conduzidas conforme descrito na **Tabela 6**. Como critério para se definir genes diferencialmente expressos, adotou-se o método de busca de valores extremos, os *outliers* (Koide *et al.*, 2006). Esse método utiliza os dados de cada experimento isoladamente para definir curvas de corte intensidade-dependentes. Assim, é possível detectar os genes que são mais distantes da distribuição do que a maioria dos pontos. Esse método foi adotado para esses experimentos já que, para essas hibridizações, notou-se que os pontos do gráfico M x S podem estar deslocados mais para a direita ou mais para a esquerda no eixo x. (**Figura 11**). Além disso, a análise da **Figura 11** também deixa evidente que a dispersão dos pontos ao longo do eixo y (M) pode ser maior ou menor dependendo do experimento. Com isso, conclui-se que o uso de intervalos de razão estáticos, definidos a partir de hibridizações homotípicas e aplicados da mesma forma para todos os experimentos, levaria a erros no momento de se definir quais genes são considerados diferencialmente expressos. Pelo método de busca de valores extremos, inicialmente, os dados M x S referentes a determinada lâmina foram usados para gerar intervalos de razão dependentes da intensidade, com nível de credibilidade de 80%. Após a definição desses intervalos de razão, os

Figura 11. Dispersão dos pontos no gráfico M x S para alguns dos experimentos de microarranjos. O gráfico M x S é mostrado para quatro experimentos, evidenciando perfis diferentes quanto às dispersões dos pontos nos eixos x e y. Os exemplos aqui mostrados são referentes aos experimentos: inoculação com *Gluconacetobacter*, experimento #2 (**A**), tratamento com MeJA por 1 h, experimento #1 (**B**), tratamento com ABA por 1 h, experimento #2 (**C**), tratamento com ABA por 12 h, experimento #1 (**D**). As curvas dependentes de intensidade foram ajustadas lâmina a lâmina, com critério de $P=0,8$. Essas curvas foram adotadas para a definição de genes diferencialmente expressos.
 $M = \log_2(I_E/I_C)$, $S = \log_2[(I_E + I_C)/2]$

mesmos foram usados para determinar quais são os genes que possuem 60% ou mais de seus pontos-réplicas acima ou abaixo desses intervalos. A **Figura 11** mostra exemplos da definição das curvas intensidade-dependentes. Um gene foi considerado diferencialmente expresso quando obteve resultado consistente nas duas réplicas biológicas analisadas para cada experimento.

No total, 179 genes foram detectados como diferencialmente expressos em pelo menos um dos tratamentos. Noventa e três genes foram responsivos à seca, 31 genes ao tratamento com ABA, 36 ao tratamento com MeJA e 14 foram regulados pela deficiência nutricional. Dos estímulos bióticos, foram obtidos nove genes regulados por herbivoria e 34 pela inoculação com bactérias diazotróficas endofíticas. Dos 179 genes diferencialmente expressos, 29 mostraram-se diferencialmente expressos em resposta a dois ou mais estímulos. A maioria desses genes (18) foi responsiva a fitormônios e à seca. Esse dado aponta para um envolvimento dos fitormônios nas respostas à seca.

A **Tabela 11** mostra os resultados de expressão para os 179 genes diferencialmente expressos observados, bem como suas respectivas categorias. Essa tabela inclui também os dados de expressão obtidos anteriormente para esses SAS nos seis órgãos de cana-de-açúcar, para os genes que se mostraram também diferencialmente expressos nesses experimentos. As alterações na expressão gênica foram mais aparentes nos pontos experimentais tardios do que nos precoces. Oitenta e oito por cento dos genes responsivos à seca mostraram-se diferencialmente expressos exclusivamente após 72 ou 120 horas de supressão de rega. No caso de tratamento com fitormônios, essa tendência também foi vista já que 58% e 50% dos genes regulados por ABA e MeJA, respectivamente, foram regulados exclusivamente após 12 horas de exposição ao fitormônio. Entretanto, a maioria dos genes regulados pela deficiência nutricional (78%) foi detectada como diferencialmente expressa precocemente (após 6 horas de tratamento).

A variedade SP90-1638, sensível à seca, foi a variedade adotada para os ensaios de estresse hídrico. A supressão de rega foi a condição na qual a maioria dos genes diferencialmente expressos foi observada. Cinquenta e dois por cento dos 179 genes diferencialmente expressos

Tabela 11. SAS com expressão diferencial frente a fatores bióticos e abióticos. A tabela lista todos os SAS que se mostraram transcricionalmente regulados em ambas as réplicas biológicas de cada ponto experimental. O sinal + indica que a gene foi induzido e o sinal – indica que a expressão do SAS foi reprimida para dado estímulo. O asterisco indica que a identidade do SAS não foi confirmada por resequenciamento. A tabela mostra também o perfil de expressão desses SAS nos seis órgãos de cana-de-açúcar, resultados esses obtidos nos experimentos de microarranjos descritos anteriormente. As quatro últimas colunas mostram em qual grupo da análise de SOM esses SAS foram incluídos, no caso de terem sido selecionados para essa análise também.

SAS	Categoria	Subcategoria #1	Subcategoria #2	Seca_24h	Seca_72h	Seca_120h	ABA_30min	ABA_1h	ABA_6h	ABA_12h	Herbasprillum spp.	Glucosebacter diazotrophicus	MeJA_1h	MeJA_6h	MeJA_12h	Herbivoria_30min	Herbivoria_24h	Deficiência de fosfato_6h	Deficiência de fosfato_12h	Deficiência de fosfato_24h	Deficiência de fosfato_48h	Flor	Gema lateral	Folha	Raiz	Entrenó 1	Entrenó 4	SOM ABA	SOM MeJA	SOM Deficiência de fosfato	SOM Seca
SCVPAM1055A12.g	Proteínas quinases	Caseína quinases	caneCKI-11																												
SCEZRZ1015G02.g	Proteínas quinases	Caseína quinases	caneCKI-17																												
SCVPRZ2036C02.g	Proteínas quinases	GSK3/Shaggy	caneGSK3-5																												
SCEZLB1005C02.g	Proteínas quinases	Indelíndas	canePK-BIII3																												
SCJFRT1010C12.g	Proteínas quinases	Provável RLCK	caneSIK-1																												
SCACLR1130D02.g	Proteínas quinases	Provável RLCK	caneRLCK-AV12																												
SCJLLR1101D07.g	Proteínas quinases	Provável RLCK	caneRLCK-DI3																												
SCEQRT1032A12.g	Proteínas quinases	RLCK	canePBS1-4																												
SCCCST1006B11.g	Proteínas quinases	SnRKs	cane osmotic stress-activated protein kinase-2																												
SCCCCLR1076F09.g	Proteínas quinases	SnRKs	caneCIPIK-13																												
SCJFRZ2032C08.g	Proteínas quinases	SnRKs	caneCIPIK-14																												
SCACLR1036B06.g	Proteínas quinases	SnRKs	caneCIPIK-18																												
SCMCR2103B04.g	Proteínas quinases	SnRKs	caneCIPIK-21																												
SCAGLB2046F06.g	Proteínas quinases	SnRKs	caneCIPIK-22																												
SCFOLB2019B08.g	Proteínas quinases	SnRKs	caneCIPIK-8																												
SCJLJR1023H04.g	Proteínas quinases	SnRKs	caneCIPIK-9																												
SCCCLB1003E11.g	Proteínas quinases	SnRKs	caneREK-1																												
SCEQRT1027E02.g	Proteínas quinases	Vias de cálcio	caneCDPK-9																												
SCQSAM2099A11.g	Proteínas quinases	Vias de MAPKs	caneMAPK-1																												
SCCCCHR1003D05.g	Proteínas quinases	Vias de MAPKs	caneMAPK-6																												
SCCCCR23001D06.g	Receptores	Fotorreceptores	caneNPH1-1																												
SCVPR172081G11.g	Receptores	Fotorreceptores	canePhyC-2																												
SCRUAD1133D10.b	Receptores	Fotorreceptores	Receptor de luz azul - cry1																												
SCBFSR1046C08.g	Receptores	Ser/Thr quinases	cane receptor-S-23																												
SCCCCL4007D06.g	Receptores	Ser/Thr quinases	caneRLK-AX1																												
SCRFRL1012B04.g	Receptores	Ser/Thr quinases	caneRLK-AX2																												
SCBFAD1045F11.g	Receptores	Ser/Thr quinases	caneRLK-C5																												
SCMCGCL6027F11.g	Receptores	Ser/Thr quinases - não classificado	caneURLK-13 (com LRR)																												
SCSFFL4017B07.g	Receptores	Ser/Thr quinases - não classificado	caneURLK-131																												
SCRLLR1038D05.g	Relacionado a vias de hormônios	Ácido jasmônico	Responsivo a jasmonatos																												
SCCCLR2002F08.g	Relacionado a vias de hormônios	Auxina	Reprimido por auxina																												
SCEZLB1009A09.g	Relacionado a vias de hormônios	Similar à proteína BLE1																													
SCCCAD1004H02.g	Respostas a estresses	Catalases																													
SCEQRT1031H03.g	Respostas a estresses	Citocromo P450 monooxigenases	CYP71																												
SCCCAM2001B04.g	Respostas a estresses	Citocromo P450 monooxigenases	CYP71E																												
SCAGLR1043E04.g	Respostas a estresses	Citocromo P450 monooxigenases	CYP74A																												
SCSGFL5008D07.g	Respostas a estresses	Citocromo P450 monooxigenases	CYP78																												
SCBGLR1099E12.g	Respostas a estresses	Citocromo P450 monooxigenases	CYP86																												
SCSGHR1069F04.b	Respostas a estresses	Citocromo P450 monooxigenases																													
SCAGLR1089C03.g	Respostas a estresses	Glutathione S-transferase																													
SCJLRT1021D12.g	Respostas a estresses	Chalcona sintase																													
SCVPLR2027D02.g	Respostas a estresses	Chalcona sintase																													
SCJLRT1016G06.g	Respostas a estresses	Ribonuclease																													
SCCCLR2001F06.g	Respostas a estresses	Induzido por ferimento																													
SCQGLR1085F11.g	Respostas a estresses	Induzido por frio																													
SCVPFL3045B09.g	Respostas a estresses	Metalotioneína																													
SCCCAD1001C08.g	Respostas a estresses	Peroxidases	P7X																												
SCCCST3001H12.g	Respostas a estresses	Aquaporina (TIP)																													
SCQGLR1062E12.g	Respostas a estresses	Glioxalase																													
SCCCLR1022B11.g	Respostas a estresses	Precursor de cisteino-protease RD19																													
SCEPRZ3087C08.g	Respostas a estresses	Proteína LTI																													
SCUTST3084F06.g	Respostas a estresses	Proteína LTI																													
SCJFLR1074E09.g	Respostas a estresses	Proteína LTI																													
SCACCL6008H06.g	Respostas a estresses	Proteína LTI																													
SCCCLR2001H09.g	Respostas a estresses	Tiorredoxina																													
SCJFLR1073B06.g	Respostas a estresses	Tiorredoxina																													
SCRFRL1056D04.g	Respostas a estresses	Tiorredoxina																													
SCCCLR1077F09.g	Sistema de duplo componente	Regulador de resposta de vias de duplo componente (similar a ARR)																													
SCACLR1057C07.g	Sistema de duplo componente	Regulador de resposta de vias de duplo componente (similar a ARR)																													
SCCCLR1067A05.g	Transportadores	Transportador de açúcar																													
SCAOLR2014E12.g	Ubiquitinação	Enzima conjugadora de ubiquitina E2																													
SCCCLR1C07A06.g	Ubiquitinação	Enzima conjugadora de ubiquitina E2																													
SCCCRZ2C01F09.g	Ubiquitinação	Enzima conjugadora de ubiquitina E2																													

Tabela 11. (continuação)

foram responsivos à condição de seca. Para 21 genes, a resposta detectada após 72 horas persistiu no ponto experimental de 120 horas. Por outro lado, o menor nível de regulação gênica dentre os genes representados no microarranjo SUCAST foi observado para os experimentos de herbivoria. As diferenças observadas quanto à regulação gênica nos diferentes experimentos devem ser, no entanto, observadas com atenção, já que diferentes cultivares foram utilizados em virtude de seus históricos já documentados de resposta aos tratamentos aplicados. No caso dos experimentos de herbivoria por *Diatraea saccharalis*, foi utilizada a variedade SP80-3280, reconhecidamente sensível a esse estresse (Barsalobres, 2004). A variedade SP80-3280 também foi adotada para os demais experimentos, exceto para os experimentos de seca, como já dito anteriormente, e para os experimentos de inoculação com bactérias diazotróficas endofíticas, para os quais a variedade SP70-1143, com elevada BNF e associação eficiente com bactérias diazotróficas endofíticas (Urquiaga *et al.*, 1992) foi escolhida.

De forma complementar à análise de genes diferencialmente expressos, os dados de expressão para os experimentos de tratamento com fitormônios, seca e deficiência em fósforo foram utilizados para uma análise por agrupamento pelo algoritmo de SOM (Tamayo *et al.*, 1999). Para essa análise foram selecionados apenas os SAS que mostraram um perfil de expressão altamente concordante (avaliado pelo índice de correlação) entre as duas réplicas biológicas analisadas para cada experimento. Adicionalmente, SAS com pelo menos um valor de M nulo ou inválido foram descartados da análise. De acordo com esses critérios, 158 genes foram selecionados para o agrupamento de perfis de expressão em resposta a ABA, 68 em resposta ao tratamento com MeJA, 146 para deficiência em fósforo e 485 para seca. Previamente ao agrupamento por SOM, os perfis de expressão de cada SAS foram normalizados por *mean-centring* para enfatizar similaridades quanto aos desvios do valor médio. O número de grupos definido para o algoritmo de SOM foi previamente definido pela análise de PCA (Quackenbush, 2001), cujos resultados são mostrados na **Tabela 12**. Os resultados obtidos mostram que o estabelecimento de quatro a seis grupos para esses dados é suficiente para representar as principais fontes de variabilidade existentes entre os genes selecionados. As geometrias de 2x2 e 2x3 foram, então, testadas ao se gerar o agrupamento por SOM desses dados. Com isso, concluiu-

# grupos	Tratamento com ABA	Tratamento com MeJA	Deficiência de fosfato	Seca
2	10,70	3,51	4,47	100,02
3	5,23	4,91	2,21	62,18
4	7,44	1,88E-17	3,21	1,02E-14
5	7,77E-17	1,38E-16	4,81E-16	2,09E-15
6	1,46E-16	3,15E-18	5,85E-17	6,59E-16
7	1,05E-16	1,51E-17	1,26E-16	5,71E-16
8	1,03E-16	1,56E-17	2,74E-17	1,77E-16
9	2,40E-17	1,75E-17	1,29E-17	2,65E-16
10	3,88E-17	5,04E-17	2,50E-17	1,05E-16
11	1,31E-17	3,11E-18	2,94E-17	2,15E-16

Tabela 12. Análise de PCA para a definição do número de grupos da análise de SOM. A diferença entre os auto-valores dos componentes principais para cada tratamento (colunas) estimam quanto grupos são necessários para cada análise de SOM. Uma diferença próxima de zero significa que o aumento do número de grupos definidos no SOM não leva a informação adicional substancial.

se que a geometria de 2x2 no caso de tratamento com MeJA e 2x3 no caso de tratamento com ABA, deficiência em fosfato e seca resultaram em grupos com pouca variação interna. Os agrupamentos gerados são mostrados na **Figura 12**. As **Tabelas 13, 14, 15 e 16** mostram os componentes dos grupos gerados por SOM para tratamento com ABA, MeJA, deficiência nutricional e seca, respectivamente. Na **Tabela 11** são também mostrados os grupos da análise por SOM dos genes detectados como diferencialmente expressos pelo método de busca dos valores extremos.

Vinte e dois genes foram selecionados dentre os diferencialmente expressos para terem os seus perfis de expressão validados por PCR em tempo real. Os *primers* desenhados para esses genes são mostrados na **Tabela 17**. O perfil de expressão para esses genes ao longo do tempo foi avaliado por PCR em tempo real para os experimentos de tratamento com fitormônios. Para os demais experimentos, as reações de PCR em tempo real foram realizadas exclusivamente para os pontos experimentais para os quais o gene foi detectado como sendo diferencialmente expresso.

Como referência para normalização, foram adotados os seguintes genes: um gene de poliubiquitina para o tratamento com ABA e dados de seca, um gene de GAPDH para os dados de herbivoria e tratamento com MeJA, o gene de 25S rRNA para interação com bactérias endofíticas e um gene de proteína 14-3-3 para os dados de deficiência nutricional. As diferentes referências foram selecionadas após a avaliação de suas respectivas curvas (log fluorescência x ciclos) obtidas com os ensaios de PCR em tempo real e após avaliar os seus perfis de expressão no espaço M x S para cada experimento em particular, como feito anteriormente para a escolha de normalizadores para órgãos de cana-de-açúcar. Os genes da poliubiquitina e da 14-3-3 aqui adotados foram os mesmos adotados previamente como referências para a avaliação dos níveis de expressão nos seis órgãos de cana-de-açúcar. Os genes de GAPDH e 25S rRNA foram descritos (Iskandar *et al.*, 2004) como bons genes de referência para órgãos e genótipos de cana-de-açúcar. No caso dos *primers* para 25S rRNA, os *primers* aqui adotados foram os denominados como 25S rRNA1F e 25S rRNA1R. Além disso, experiências prévias dos grupos do Dr. Márcio de Castro Silva Filho (USP) e da Dra. Adriana Hemerly (UFRJ) mostram que GAPDH e rRNA são

Figura 12. Agrupamento por SOM dos perfis de expressão para dados de tratamento com ABA (A), tratamento com MeJA (B), deficiência de fosfato (C) e seca (D). Os SAS foram selecionados baseados na concordância dos seus perfis de expressão entre as duas réplicas biológicas avaliadas para cada experimento. Os valores da mediana das razões de expressão para cada gene em cada réplica foram normalizados por *mean-centring* e os valores normalizados utilizados para o cálculo das médias entre as duas réplicas biológicas. Esses valores médios foram utilizados para o agrupamento por SOM. A geometria foi escolhida com base na análise de PCA. Os gráficos representam a média dos valores normalizados por *mean-centring* de M (eixo y) ao longo do tempo experimental (eixo x). Os número entre parênteses nos gráficos indicam o número de genes incluídos em cada grupo.

SAS	Anotação	Grupo	SAS	Anotação	Grupo
SCSBAM1085B06.g	Ômega-3 dessaturase	A1	SCCCHR1003D05.g	Proteína quinase (caneMAPK-6)	A4
SCEZHR1047G06.g	Biossíntese de auxinas - nitrilase	A1	SCCCRZ1003E02.g	Proteína quinase (caneMAPK-7)	A4
SCCCRZ1002H11.g	Indol-3-glicerol fosfato sintase	A1	SCCCLR1024A02.g	Receptor (caneRLK-CIII4)	A4
SCPEAM2013C04.g	Provável proteína com repetição pentatricopeptídica	A1	SCCCCL3002B05.b	Proteína responsiva à auxina	A4
SCVPCL6038A02.g*	Gene-R (caneRpg1-9)	A1	SCCCRZ2C04A05.g	Regulador de resposta (tipo ARR)	A4
SCJFAM1065F08.g	No match	A1	SCEQLB1067E03.g	Regulador de resposta (tipo ARR)	A4
SCCCSB1001F06.g	Aminoidrolase	A1	SCEZAM2032C05.g	Pumílio	A4
SCJLRZ1019B10.g	Gene-R (NBS-LRR)	A1	SCJFAD1013B07.g	No match	A4
SCUTSB1033E12.g	Gene-R (NBS-LRR)	A1	SCACLR1126E09.g	No match	A4
SCEZRZ3052A09.g	Proteína fosfatase - Família PPM (PP2A/subunidade B)	A1	SCPEAM2015C06.g	No match	A4
SCCCLR1072C05.g	Proteína quinase (caneMAPKK-2)	A1	SCCFHR1034H08.g	No match	A4
SCEPSB1136D04.g	Proteína quinase (caneYakA-1)	A1	SCCCLR1022D05.g	Proteínas 14-3-3	A5
SCBFSB1048C08.g	Receptor (cane S-receptor-23)	A1	SCCCRZ1001D02.g	Proteínas 14-3-3	A5
SCSFRT2071G08.g	Receptor (caneURLK-121)	A1	SCQGHR1010A04.g	Hidroxiperoxido liase	A5
SCVPST1060B03.g	Transcrição - Nin	A1	SCVPLR2012A10.g	Biossíntese de etileno (ACC oxidase)	A5
SCUTAD1030E06.g	Transcrição - Alfin	A1	SCEZHR1087F06.g	Citocromo P450 monooxigenase (CYP84)	A5
SCEQRT2091D04.g	Ubiquitinação - proteínas F-box	A1	SCSBST3093A10.g*	Citocromo P450 monooxigenase	A5
SCEQRT1028H06.g	Biossíntese de auxinas - nitrilase	A2	SCCCCL3120G07.g	HSP70 (proteína de choque térmico)	A5
SCCCRZ2C04A07.g	Citocromo P450 monooxigenase (CYP71E)	A2	SCEZRZ1017E02.g	Trealase ácida	A5
SCAGRT3046D01.g	Citocromo P450 monooxigenase (CYP88)	A2	SCCCCL3002E04.b	Trealose 6-fosfato sintase	A5
SCRLAD1102F02.g	Citocromo P450 monooxigenase	A2	SCJFRZ2009G01.g	Proteína ribossomal	A5
SCSBHR1052E03.g	Induzido por ABA e estresses	A2	SCCCLR1C01H06.g	Proteína similar à topoisomerase	A5
SCSBRZ3119G01.g	Aminoidrolase	A2	SCCCRZ2001H03.g	Provável hematóglutina streptocócica	A5
SCEZRZ1015H12.g	Histona deacetilase (HOS-1)	A2	SCEQLR1029B03.g	Gene-R (NBS-LRR)	A5
SCJFLR1013A09.g	Precursor de proteinase de cisteína RD19A	A2	SCCCLR1066D10.g	GTPase - Arf	A5
SCJFLR1074E10.g	Precursor de proteinase de cisteína RD19A	A2	SCCCLR1066D10.g	GTPase - Arf	A5
SCCCST3001H12.g	Aquaporin (TIP)	A2	SCCCLR1C01G03.g	GTPase - Arf	A5
SCCCLR2001H09.g	Tiorredoxina	A2	SCSBST3097D10.g	Proteína ativadora de Ran GTPase	A5
SCCCRZ1C01H06.g	Proteínas que se ligam à calmodulina (Apirase)	A2	SCFRBL2056B08.g	Proteína desconhecida	A5
SCCCRZ2001F03.g	Frutose-bisfosfato aldolase	A2	SCSGLR1025H02.g	Proteína desconhecida	A5
SCCCRZ2002C12.g	Proteína ribossomal S19	A2	SCCCCL3005D01.b	Proteína fosfatase - Família PPM (PP2C/subunidade catalítica)	A5
SCCCLR1022F10.g	Glicina hidroximetiltransferase	A2	SCCCCL4005E10.g	Proteína fosfatase - Família PPM (PP4/subunidade catalítica)	A5
SCCCLR1C08E09.g	GTPase - Rab	A2	SCEQHR1078G03.g	Proteína fosfatase - Família PPM (Proteína fosfatase 6)	A5
SCBGLR1117A05.g	GTPase - Rab	A2	SCRFLR1034G06.g	Proteína quinase (canePKABA1-3)	A5
SCCCLR1080C06.g	Proteína desconhecida	A2	SCMCCL6052C09.g	Receptor (caneRLK-C1)	A5
SCRFLR2038D12.g	Proteína desconhecida	A2	SCJLRT1013A09.g*	Receptor (caneRLK-DI4)	A5
SCVPLR1049H11.g	Proteína desconhecida	A2	SCACLR2022D09.g	Transcrição - Aux/IAA	A5
SCAGLR1021H12.g	Proteína quinase (caneGSK3-3)	A2	SCCCAM2001D09.g	Transcrição - Homeobox (knotted homeobox)	A5
SCEQRT1024B02.g	Proteína quinase (caneG11A kinase-2)	A2	SCSFAD1124E07.g	Transcrição - MYB	A5
SCJLRT1021D01.g	Proteína quinase (caneRLCK-AVIII2)	A2	SCJLHR1027E08.g	Transcrição - NAM	A5
SCSGLR1025B10.g	Receptor (cane SHR5 receptor-2)	A2	SCSBFL1039F08.g	Transcrição - NAM	A5
SCSGHR1067C04.g*	Receptor (caneRLK-DV3)	A2	SCJFRT2060G09.g	Transcrição - Tub	A5
SCJFFL3C02F01.g*	Receptor (caneURLK-1)	A2	SCUTRZ2024G05.g	Provável transportador de vesícula v-SNARE	A5
SCUTRZ2023E07.g	Transporte de auxina	A2	SCVPLR2005B09.g	No match	A5
SCQGLB1030E04.g	Proteína regulada por etileno	A2	SCBGLR1112B03.g*	No match	A5
SCCCFL4091A07.g	Proteína responsiva à giberelina	A2	SCJFHR1033B03.g	No match (não codificador)	A5
SCEZLB1009A09.g	Similar à proteína BLE1	A2	SCSBFL1036A08.g	No match (não codificador)	A5
SCJFAM1066B05.g	Transcrição - HIT (tríade de histidina)	A2	SCEQRT1025D06.g	Proteínas 14-3-3	A6
SCRURT2007E11.g	Transcrição - MYB	A2	SCCCAD1004H02.g	Catalase	A6
SCACLR1130H08.g	Transcrição - Proteínas dedo de zinco (C2C2 / YABBY)	A2	SCRFHR1009G06.g	Proteína de resposta a estresses	A6
SCACLR1036E09.g	Transcrição - WRKY	A2	SCAGFL1089C03.g	Glutathione S-transferase	A6
SCEPLB1043A10.g	Provável transportador de açúcar	A2	SCCCLR2C01F06.g	Induzido por fermentos	A6
SCCCRZ2C01F09.g	Ubiquitinação - E2	A2	SCCCLR1022B11.g	Precursor de proteinase de cisteína RD19A	A6
SCSBSD2058D04.g	Ubiquitina	A2	SCMCLR1123E10.g	Proteína do complexo T (chaperonina)	A6
SCSBFL1044A08.g	No match	A2	SCJFRT1059D05.g	GTPase - Rab	A6
SCUTSD2085H06.g	No match	A2	SCUTAM2115C12.g	Proteína desconhecida	A6
SCCCLR1C03G01.g	Delta-12 oleato dessaturase	A3	SCMCAM1100C01.g	Proteína desconhecida	A6
SCEQRT1024E12.g	Biossíntese de SA (fenilalanina amônia-liase)	A3	SCCCAM1001D05.g	Proteína fosfatase - Família PPM (PP2C/subunidade catalítica)	A6
SCSGAM1094D05.g	Biossíntese de SA (fenilalanina amônia-liase)	A3	SCEZRZ1015G02.g*	Proteína quinase (caneCKI-17)	A6
SCQGLR1062E12.g	Glioxalase	A3	SCJLRT1014C12.g	Proteína quinase (caneMAPK-19)	A6
SCRULR1020D11.g	Frutose-bisfosfato aldolase	A3	SCCCCL5002B10.g	Proteína quinase (caneUPK-87)	A6
SCCCLR1C02F07.g	Mio-inositol-1-fosfato sintase	A3	SCEQRT1026B06.g	Proteína quinase (caneUPK-88)	A6
SCEQRT1028C03.g	Proteína PR	A3	SCVPRT2074E06.g	Proteína quinase (caneCDC2-4)	A6
SCQSRT1036D03.g	Proteína PR	A3	SCRFRT3057H04.g	Proteína quinase (caneCDK-14)	A6
SCBGLR1023D05.g	Vias de transdução de genes-R - LSD1	A3	SCJLRT1023G09.g*	Proteína quinase (caneCIPK-19)	A6
SCCCRZ2C03B08.g	Proteína desconhecida	A3	SCMCLR1053E09.g	Receptor (caneRLK-DVI2)	A6
SCSBSD2029F05.g	Proteína desconhecida	A3	SCCCLB1023E12.g	Receptor (caneRLK-DXIV1)	A6
SCSGSB1009D11.g	Proteína desconhecida	A3	SCJFRT2055F11.g	Receptor (caneURLK-95)	A6
SCUTLR1058C02.g	Proteína desconhecida	A3	SCCCLR2002F08.g	Proteína reprimida por auxina	A6
SCVPRT2074D04.g	Proteína desconhecida	A3	SCJFRT1062G05.g	Transcrição - CCAAT	A6
SCBFSB1047C02.g	Proteína desconhecida	A3	SCRLFL4029G02.g	Transcrição - PCF3	A6
SCCCAM2004F11.g	Proteína quinase (canePK-FII2)	A3	SCCCCL4003D08.g	Transcrição - Proteínas dedo de zinco (C3H)	A6
SCCCRT3003D07.g	Receptor (caneRLK-DXIII2)	A3	SCSBST3101E05.g	Transcrição - Proteínas dedo de zinco (C3H)	A6
SCCCST3C01D11.g	Receptor (caneURLK-84)	A3	SCSFFL4082E12.g	Ubiquitinação - E1	A6
SCACAM1071D04.g	Transcrição - HSF	A3	SCBFSB1035H11.g	No match	A6
SCRURT2008F06.g	Transcrição - Proteínas dedo de zinco (C2H2 / Wrep)	A3	SCCCCL4007A05.g	No match	A6
SCACLR2014E12.g	Ubiquitinação - E2	A3	SCEPFL4174B08.g	No match	A6
SCBFSB3136A06.g	No match	A3	SCRFLL4006H10.g	No match	A6
SCEPAM1106G01.g	Proteína ativadora de Ran GTPase	A4	SCRLFL1008C11.g	No match	A6

Tabela 13. Componentes dos grupos SOM mostrados na Figura 12, para dados de resposta ao tratamento com ABA. Os SAS marcados com asterisco são aqueles que não tiveram a identidade confirmada por resequenciamento.

SAS	Anotação	Grupo
SCCCLB1003A09.g	Proteína associada à senescência	B1
SCJLRT1020A09.g	Glutathione S-transferase	B1
SCJLHR1028C12.g	Histona H4	B1
SCCCLR1068G11.g	Histona H2B	B1
SCCCLR2002D04.g	Histona H5	B1
SCEQLR1091A10.g	Proteína ribossomal L24	B1
SCAGSD1041C06.g	No match	B1
SCCCST1001C04.g	No match	B1
SCCCRZ2C04C03.g	No match	B1
SCMCAM2084G01.g	No match (não codificador)	B1
SCRLLR1111D02.g	No match (não codificador)	B1
SCBGLR1023D05.g	Vias de transdução de gene-R (LSD1)	B1
SCEPLR1030B03.g	Semelhante à proteína LRP de tomate	B1
SCSBS2029F05.g	Proteína desconhecida	B1
SCVPLR1049B12.g	Proteína desconhecida	B1
SCCCLR2003C11.g	Proteína fosfatase - Família PPP (PP1 subunidade catalítica)	B1
SCJLRT1021D01.g	Proteína quinase (caneRLCK-AVIII2)	B1
SCCCCL3120C09.g	Receptor (cane RLK with LysM-1)	B1
SCCCST3C01D11.g	Receptor (caneURLK-84)	B1
SCEPRT2048E07.g	Regulador de resposta (tipo ARR)	B1
SCCCRZ1001G10.g	Transcrição - Aux/IAA	B1
SCJFAM1066B05.g	Transcrição - HIT	B1
SCQGRZ3011D06.g	Transcrição - Proteína tipo Alfin	B1
SCBGLR1003D06.g	Ubiquitinação - E2	B1
SCCCLR1C04E03.g	Ubiquitinação - E2	B1
SCACLR2014E12.g	Ubiquitinação - E2	B1
SCEQRT1031D02.g	Proteínas 14-3-3	B2
SCCCRZ1001D02.g	Proteínas 14-3-3	B2
SCVPRZ2038F04.g	Biossíntese de JÁ - ácido linoléico dessaturase	B2
SCJLFL3015H10.g	Proteína associada à senescência	B2
SCEZST3149E09.g	Glutathione S-transferase	B2
SCCCLR1075A11.g	Proteína induzida por ferimentos	B2
SCEPSD2071B08.g	Inibidor de quimiopripsina	B2
SCAGLR1043C02.g	Ascorbato peroxidase	B2
SCEPAM1024H12.g	Histona deacetilase (HOS-1)	B2
SCCCLR2C03D05.g	Superóxido dismutases - Cu/Zn	B2
SCEPLB1043G09.g	No match	B2
SCVPLR2005B09.g	No match	B2
SCCCLR2C01F04.g	No match	B2
SCCCLR2001E09.g	No match	B2
SCCCCL3001F04.g	No match (não codificador)	B2
SCCCCL4008D12.g	Proteína com provável domínio de ligação a metila	B2
SCSBLB1034C04.g	Gene-R (NBS-LRR)	B2
SCQSRT1036D03.g	Proteína PR	B2
SCCCLR1066D10.g	GTPase - Arf	B2
SCCORT2001H11.g	GTPase - Arf	B2
SCBGFL4051G01.g	Proteína desconhecida	B2
SCCCRZ2C03B08.g	Proteína desconhecida	B2
SCCCRZ1C01G06.g	Proteína quinase (caneGSK3-6)	B2
SCCCLR1C03E03.g*	Proteína quinase (caneLAMMER-2)	B2
SCEZRZ1015D09.g	Proteína quinase (caneRLCK-AVIII1)	B2
SCSGLR1025B10.g	Receptor (cane SHR5 receptor-2)	B2
SCSGHR1067C04.g*	Receptor (caneRLK-DV3)	B2
SCRRLR2041C03.g	Transcrição - GARP	B2
SCCCLR1079H06.g	Transcrição - Proteína tipo Alfin	B2
SCCCRZ2001H09.g	Ubiquitinação - E2	B2
SCCCAD1004H02.g	Catalase	B3
SCBFSD1035H11.g	No match	B3
SCACLR1057C07.g	Regulador de resposta (tipo ARR)	B3
SCCCLR2C01F06.g	Proteína induzida por ferimentos	B4
SCUTHR1063G08.g	No match	B4
SCRLLR1006E08.g	No match	B4
SCRFHR1006H10.g	No match	B4
SCUTSD2085H06.g	No match	B4
SCQSFL3034B08.g	No match	B4
SCEZRZ3098G10.g	Inibidor de proteases - taumatina	B4
SCEQAM1037D01.g	Proteína de resistência indefinida	B4
SCJLRZ1019D02.g	Receptor histidina quinase - receptor de etileno	B4

Tabela 14. Componentes dos grupos SOM mostrados na Figura 12, para dados de resposta ao tratamento com MeJA. Os SAS marcados com asterisco são aqueles que não tiveram a identidade confirmada por resequenciamento.

SAS	Anotação	Grupo	SAS	Anotação	Grupo
SCVPC6041E07.g	Citocromo P450 monooxigenase (CYP98A)	C1	SCCCRZ2002E06.g	Fosfolipase C	C5
SCJFLR1073B06.g	Tiorredoxina	C1	SCCCLR1078F05.g	Calmodulina	C5
SCBFLR1039B05.g	Xiloglucana endotransglicosilase	C1	SCCCLR1070F12.g	Calnexina	C5
SCRULB2065G10.g	No match	C1	SCEPRZ1008D06.g	Canal iônico regulado por nucleotídeo cíclico e calmodulina	C5
SCJLHR1029D03.g	No match	C1	SCSGFL1078H06.g	Canal iônico regulado por nucleotídeo cíclico e calmodulina	C5
SCRLST3166F11.g	No match (não codificador)	C1	SCCCLR1075G05.g	Provável fator de <i>splicing</i>	C5
SCSBST3097D10.g	Proteína de ativação de Ran GTPase	C1	SCRLFL3006E04.g	No match	C5
SCRFLR2038D12.g	Proteína desconhecida	C1	SCSBFL1107H01.g	No match	C5
SCEPRZ1010E06.g	Proteína fosfatase - Família PPM (tipo PP2C)	C1	SCEPLB1043G09.g	No match	C5
SCAGST3138B05.g	Fotorreceptor (receptor de luz azul, cry1)	C1	SCRUFL1015G07.g	No match	C5
SCQSST1038C07.g	Receptor (caneCLV1-5)	C1	SCAGLR2026G12.g	No match	C5
SCJFRZ1007B02.g	Receptor (canePERK1-5)	C1	SCCCLR2001F04.g	No match	C5
SCSBLB2040E07.g	Regulador de resposta (tipo ARR)	C1	SCBGLF4055H04.g	No match	C5
SCVPAM1055F07.g	Transcrição - WRKY	C1	SCEPAM1020E03.g	No match	C5
SCUTLR2023G02.g	Transcrição - WRKY	C1	SCRUFL1019D03.g	No match	C5
SCEQRT2090E04.g	Calreticulina	C2	SCEZLB1013D01.g*	No match (não codificador)	C5
SCEQLR1093F09.g	Provável beta-D-glucana exoidrolase	C2	SCSGCL6071D03.g	No match (não codificador)	C5
SCRUAD1064B08.g	No match	C2	SCJLLR1105A08.g	Proteína relacionada a fosfolípido/glicerol aciltransferase	C5
SCAGLR1043D11.g	No match	C2	SCJFRZ2005E03.g	Provável associada à microtúbulo	C5
SCCCCL3004C02.b	No match	C2	SCVPC6038A02.g*	Gene-R (caneRpg1-9)	C5
SCCCSD1003H09.g	No match	C2	SCVPRT2081F09.g	Gene-R (caneXa21-5)	C5
SCMCST1051F08.g	Proteína fosfatase de especificidade dupla	C2	SCRFAM2131G06.g	Gene-R (NBS-LRR)	C5
SCRFLR1012B04.g	Receptor (caneRLK-AX2)	C2	SCCCCL4009G05.g	GTPase - Arf	C5
SCEZHR1047A01.g	Receptor (caneRLK-DV11)	C2	SCQSST1038E01.g	Inibidor de dissociação de GDP de Rho	C5
SCEPLR1051A10.g	Transcrição - Homeobox	C2	SCQSST1037B07.g	Proteína de ativação de Rho GTPase	C5
SCCCCL4007B12.g	Ubiquitinação - E2	C2	SCRFLR1012C09.g	GTPase - Rab	C5
SCQGRT1045D01.g	Ubiquitinação - E3	C2	SCACLR1036D11.g	GTPase - Rab	C5
SCEPAM1022H12.g	Citocromo P450 monooxigenase (CYP71A)	C3	SCCCLR2002H12.g*	GTPase - Ran	C5
SCEPSD2071B08.g	Inibidor de quimi tripsina	C3	SCRFAM1025A12.g	Proteína desconhecida	C5
SCJFLR1074E09.g	Resposta à seca e ao frio (proteína LTI)	C3	SCEQFL5046E01.g	Proteína desconhecida	C5
SCJFRZ3C07D05.g	No match	C3	SCBGLF4051G01.g	Proteína desconhecida	C5
SCJFFL3C03C02.g	No match	C3	SCCCCL4004C06.g	Proteína desconhecida	C5
SCVPRZ2042A04.g	No match	C3	SCJLFL3017H03.g	Proteína desconhecida	C5
SCQSFL3034B08.g	No match	C3	SCJFRZ2029F09.g	Proteína desconhecida	C5
SCRUFL1117G06.g	No match	C3	SCCCAM2C01D08.g	Proteína desconhecida	C5
SCVPC6042D07.g	No match (não codificador)	C3	SCCCLR1070B11.g	Proteína desconhecida	C5
SCACAD1036D05.g	Gene-R (caneRpg1-5)	C3	SCSFAD1106A08.g	Proteína desconhecida	C5
SCAGLR1043E12.g	Proteína quinase (caneEDR1-1)	C3	SCRLAM1014C02.g	Proteína fosfatase - Família PPM (PP2A/Subunidade B)	C5
SCCCLR1079F12.g	Proteína quinase (cane Wee1 kinase-1)	C3	SCCCRZ2004B09.g	Proteína fosfatase - Família PPM (PP2A/Subunidade B)	C5
SCCCLR2002G01.g	Proteína quinase (caneCDPK-8)	C3	SCCCCL3005D01.b	Proteína fosfatase - Família PPM (PP2C/subunidade catalítica)	C5
SCJFRZ2015D01.g	Proteína quinase (caneCDK-5)	C3	SCQGLB1030C05.g	Proteína fosfatase - Família PPM (PP7/subunidade catalítica)	C5
SCEPRZ1008A04.g	Transcrição - AP2/EREBP	C3	SCJLRZ1021C12.g	Proteína tirosina fosfatase	C5
SCAGFL1089C03.g	Glutathione S-transferase	C4	SCCCNR1001B12.g	Proteína quinase (caneEDR1-2)	C5
SCCCCL3002C09.b	Glutathione S-transferase	C4	SCJLRT1017B11.g	Proteína quinase (caneEDR1-3)	C5
SCCCLR1048A06.g	Superóxido dismutase - Mn	C4	SCEPRZ1011A12.g	Proteína quinase (canePK-BII4)	C5
SCBGLR1115D10.g	No match (não codificador)	C4	SCQGRT1042H04.g	Proteína quinase (caneCDK-13)	C5
SCVPLB2087A07.g	Gene-R (caneXa21-4)	C4	SCSGAM2105E02.g	Receptor (caneURLK-124)	C5
SCJLAM1060A06.g	Gene-R (Mio)	C4	SCSBST3101B03.g	Receptor (caneCLV1-4)	C5
SCCCLR1066D10.g	GTPase - Arf	C4	SCQSRT2034G10.g	Receptor (caneRLK-CII3)	C5
SCCCLR1068D03.g	GTPase - Rab	C4	SCSGHR1067C04.g*	Receptor (caneRLK-DV3)	C5
SCQGLR1019A10.g	GTPase - Ran	C4	SCCCLR1065B12.g	Receptor (caneRLK-DVI3)	C5
SCBFRZ2045D11.g	Proteína desconhecida	C4	SCCCRZ1C01F12.g	Receptor (caneRLK-DVIII2)	C5
SCCCFL4003D09.g	Proteína desconhecida	C4	SCJFLR1035B03.g	Regulador de resposta (tipo ARR)	C5
SCCCAM1001D05.g	Proteína fosfatase - Família PPM (PP2C/subunidade catalítica)	C4	SCCCLR2001A01.g	Transcrição - AP2/EREBP	C5
SCVPC6042B07.g	Proteína quinase (cane cyclin G-associated kinase-like protein-1)	C4	SCRLAM1014G08.g	Transcrição - AP2/EREBP	C5
SCEPSD1005C08.g	Receptor (cane RLK with lectin domain-9)	C4	SCCCRZ2003H05.g	Transcrição - Aux/IAA	C5
SCBGLF5080E07.g	Receptor (caneBR11-6)	C4	SCQGST3154C02.g	Transcrição - NAM	C5
SCCCLR1066G08.g	Transcrição - HGM	C4	SCVPLR2027A05.g	Transcrição - Fator responsivo a auxinas	C5
SCEPRZ1008F02.g	Transcrição - LIM	C4	SCCCCL7001A04.g	Transcrição - Proteína dedo de zinco (C2C2 / YABBY)	C5
SCQGRT1041A07.g	Transcrição - Nin	C4	SCVPRZ2043E04.g	Transcrição - SBP	C5
SCCCLR1C04E03.g	Ubiquitinação - E2	C4	SCSBFL1106D09.g	Transcrição - Tipo alfin	C5
SCJLAM2091A04.g	Biossíntese de giberelina (GA2 oxidase)	C5	SCSGHR1066H10.g	Transcrição - WRKY	C5
SCRFAM1026H03.g	Ciclina lb	C5	SCQSFL1125C02.g	Ubiquitinação - E3	C5
SCEQRT1030A03.g	ARC1 (arm)	C5	SCCCLR1024E11.g	Superóxido dismutase - Cu/Zn	C6
SCUTAM2088C05.g	Proteína associada à senescência-13	C5	SCCCLR2C03D05.g	Superóxido dismutase - Cu/Zn	C6
SCCCCL4009H01.g	Citocromo P450 monooxigenase (CYP73)	C5	SCEPSD1069D02.g	Calreticulina	C6
SCAGAM2125G04.g	Citocromo P450 monooxigenase (CYP86)	C5	SCEZLR1009F06.g	Piruvato desidrogenase	C6
SCUTAM2005B03.g	Citocromo P450 monooxigenase (CYP9)	C5	SCMCRZ3064C08.g	No match	C6
SCSGRT2066D05.g	Citocromo P450 monooxigenase	C5	SCSBFL1035B08.g	No match	C6
SCSBRZ3119G01.g	Aminoidrolase	C5	SCAGSB1087E03.g	Gene-R (NBS-LRR)	C6
SCEPAM1024H12.g	Histona deacetilase (HOS-1)	C5	SCUTST3129E01.g	Proteína desconhecida	C6
SCBGHR1059B03.g*	Transportador ABC	C5	SCVPRZ2039D11.g	Proteína fosfatase - Família PPM (PP2C/subunidade catalítica)	C6
SCRUFL4020G01.g	Mioinositol monofosfatase	C5	SCQSST1036G03.g	Receptor (caneURLK-47)	C6

Tabela 15. Componentes dos grupos SOM mostrados na Figura 12, para dados de resposta à deficiência em fosfato. Os SAS marcados com asterisco são aqueles que não tiveram a identidade confirmada por resequenciamento.

Tabela 16. Componentes dos grupos SOM mostrados na Figura 12, para dados de resposta à seca. Os SAS marcados com asterisco são aqueles que não tiveram a identidade confirmada¹⁵ por resequenciamento.

SAS	Anotação	Sequências dos <i>primers</i>	Pontos experimentais:	Valor P:
SCRULR1020D11.g	Metabolismo de carboidrato - Frutose-bisfosfato aldolase	TGACACCATGGACCATGCAT	Seca (72h) Seca (120h) Tratamento com ABA (12h)	1,0 (reprimido) 1,0 (reprimido) 1,0 (induzido)
SCCQRT1001E01.g*	Biossíntese de ácido isomônico - LOX	CGTGTG03TTGGGGAGAAG	Tratamento com MeJA (6h)	1,0 (induzido)
SCEQRT1024E12.g*	Biossíntese de SA - Fenilalanina amônia-lyase	CTTCCAGGGCACTCCCATTT	Tratamento com ABA (12h) Tratamento com MeJA (6h) Tratamento com MeJA (12h)	1,0 (reprimido) 1,0 (reprimido) 0,99 (reprimido)
SCAGLR2026G12.g	No match	CAGTTTCAGATCAGGCTCTTCGAT	Tratamento com ABA (1h) Tratamento com ABA (6h) Tratamento com ABA (12h)	0,92 (reprimido) 1,0 (reprimido) 1,0 (reprimido)
SOMCLR1123E10.g	Proteína de complexo T (chaperonina)	GCAGCTTCATCGGATACAACAAG	Deficiência de fosfato (6h)	1,0 (reprimido)
SCCSD1003E02.g	Patogenicidade - Inibidor de protease - I-aumalina	GCAGTGTCTGCAGGAGAAGT	Herbivoria (24h)	1,0 (induzido)
SCEQRT1028C03.g	Patogenicidade - Inibidor de protease - proteína PR	GAATTCGAGCCCTCTCCTTCGT	<i>Herbaspirillum</i> spp.	1,0 (induzido)
SCEQRT2098H06.g	Patogenicidade - Gene-R com NBS-LRR	GCAGAGAGCTCGCCAGAGA	<i>Glucanacetobacter diazotrophicus</i> <i>Herbaspirillum</i> spp.	1,0 (induzido) 0,0 (induzido)
SCBGLR1023D05.g*	Patogenicidade - Vias de transdução de genes-R - LSD1	CAAAATCAAGTTGCGCATGTAAAC	Tratamento com MeJA (12h)	1,0 (induzido)
SCCCHLR1003D05.g	Proteína quinase - canaMAPK-6	CACCAAAAGTCAGTTTATGACAGAA	Tratamento com ABA (12h)	0,93 (reprimido)
SCCCLRL1003E11.g	Proteína quinase - SnRKs- caneREK-1	CTCGGCAGGTTCTTCAGGAA	<i>Glucanacetobacter diazotrophicus</i> <i>Herbaspirillum</i> spp.	1,0 (induzido) 1,0 (induzido)
SCEPRZ1010E06.g	Proteína fosfatase - Ser/Thr fosfatase - similar a PP2C	CGTCTTTCTAGCCTCTGGAACAC	Seca (72h) Seca (120h)	1,0 (induzido) 1,0 (induzido)
SCAGLR1043E04.g	Resposta a estresses - Citocromo P450 monooxigenases - CYP74A	TGAAGCGGACGAATTTGAGTAG	Tratamento com ABA (12h)	1,0 (reprimido)
SCQGLR1062E12.g	Resposta de seca e frio - Glioxalase	CAAAAGAGCGCATCTTAGTGTGA	Tratamento com ABA (12h)	1,0 (induzido)
SCQGLR1085F11.g*	Resposta a estresses - Deidrina	TCACACCACACAAAAGTGGAAA	Seca (72h) Seca (120h)	1,0 (induzido) 1,0 (induzido)
SCJLHR1028G12.g	Metabolismo de DNA - Histona H4	AACTGGTTGATTTGTCATGTTGTGT	Tratamento com MeJA (12h)	1,0 (induzido)
SCCCLR2001F06.g	Resposta a estresses - Induzido por ferimento	CACCAACCAGAGCGCAAAAT	Tratamento com MeJA (1h) Tratamento com MeJA (6h) Tratamento com MeJA (12h)	1,0 (reprimido) 1,0 (reprimido) 1,0 (reprimido)
SCCGRZ1001H05.g	Fator de transcrição - HLH	CATCCTGCTGCTTTTACTCCATT	<i>Glucanacetobacter diazotrophicus</i>	1,0 (induzido)
SCCGRZ3002D03.g	Fator de transcrição - LIM	CGAAGCTCCGTGCTGCTGTG	Herbivoria (30min)	0,41 (induzido)
SCCCLR1005B03.g	Fator de transcrição - MYB	GGCAACAATGGTCCCTCAT	Deficiência de fosfato (6h)	1,0 (reprimido)
SCCCLR2003E10.g	Fator de transcrição - NAC	CATCTTCTCCCACTCGTCTTCTT	<i>Glucanacetobacter diazotrophicus</i>	0,0 (reprimido)
SCEZHR1088E02.g	Proteína fosfatase de especificidade dupla	CAGCTGGTCAAGAAACCATCA	Seca (24h) Seca (120h)	0,6 (reprimido) 0,05 (reprimido)

Tabela 17. Genes de cana-de-açúcar responsivos a estímulos bióticos e abióticos selecionados para validação por PCR em tempo real e *primers* desenhados. Os *primers* foram desenhados utilizando-se o Primer Express 2.0 Software (Applied Biosystems) e buscas de BLAST foram conduzidas contra o banco de dados SUCEST de forma a garantir a especificidade dos *primers* desenhados. A coluna “Pontos experimentais” refere-se aos pontos em que o SAS foi detectado como sendo diferencialmente exposto pelo método de busca de valores extremos. As sequências dos primers dos normalizadores são as mesmas utilizadas para PCR em tempo real com amostras de órgãos de cana-de-açúcar ou foram baseadas em Iskandar *et al.*, 2004.

* Esses *primers* são os mesmos utilizados para experimentos comparativos entre os seis órgãos de cana-de-açúcar. A coluna “valor P” denota a probabilidade P da razão de expressão na amostra experimental ser maior do que 1 para genes induzidos ou menor do que 1 para genes reprimidos. O perfil de expressão foi considerado validado para P maior ou igual a 0,99.

bons normalizadores para dados provenientes de experimentos de herbivoria e de interação com bactérias diazotróficas endofíticas, respectivamente.

Alguns dos genes selecionados foram avaliados em mais de um ponto experimental. Um total de 36 resultados para expressão diferencial foi avaliado. Desses, 80,5% tiveram resultados nos ensaios de PCR em tempo real consistentes com os observados pelos experimentos de microarranjos, considerando-se que o perfil de expressão foi validado para $P \geq 0,99$ (**Tabela 17**). Os resultados validados por PCR em tempo real são mostrados na **Figura 13**. É importante enfatizar que as amostras de RNA utilizadas para os experimentos de PCR em tempo real são provenientes de uma terceira réplica biológica. Os resultados discrepantes obtidos entre as duas técnicas podem corresponder a variações biológicas ou mesmo a limitações técnicas dos métodos.

6.5. Perfil de expressão em populações de alto e baixo brix

Atualmente, pouco é conhecido a respeito dos mecanismos regulatórios da síntese e acúmulo de sacarose em cana-de-açúcar. Diversos cultivares contrastantes para síntese e/ou acúmulo de sacarose estão disponíveis e a comparação desses genótipos entre si no nível molecular pode apontar para genes relacionados ao teor de sacarose. A identificação de genes associados a características fenotípicas de interesse comercial é altamente desejável para a indústria canavieira, uma vez que esses genes podem vir a ser utilizados como marcadores moleculares para a seleção assistida ou na produção de plantas transgênicas melhoradas. Com o intuito de selecionar plantas com elevado teor de brix, uma série de cruzamentos envolvendo genótipos de *Saccharum officinarum* e *Saccharum spontaneum* foram realizados no CTC. Inicialmente cruzamentos intra-específicos foram realizados entre 21 genótipos de *S. officinarum* e 13 genótipos de *S. spontaneum* (**Tabela 18**). Posteriormente, as progênies desses cruzamentos independentes foram avaliadas quanto a seus teores de açúcares e os indivíduos mais extremos foram cruzados entre si. Eventos de recombinação e seleção foram promovidos, então, para selecionar duas populações contrastantes para níveis de acúmulo de açúcar (**Figura 14**).

A. Tratamento com ABA

B. Tratamento com MeJA

C. Inoculação com bactéria endofítica

D. Seca

E. Deficiência de fosfato

F. Herbivoria

Figura 13. Avaliação da expressão por PCR em tempo real de genes selecionados como diferencialmente expressos frente a tratamentos com hormônios ou a estímulos ambientais. A figura mostra os resultados obtidos em acordo com a detecção de genes diferencialmente expressos pelo método de busca de valores extremos para tratamento com ABA (A), tratamento com MeJA (B), interação com bactérias endofíticas (C), seca (D), deficiência de fosfato (E) e herbivoria (F). O eixo y refere-se à razão de expressão relativa calculada entre as amostras experimentais *versus* a amostra controle (não tratada). No caso dos experimentos com fitormônios, as razões foram calculadas em relação à amostra 0 h e os níveis de expressão dos genes selecionados foram avaliados para todos os pontos experimentais. Para os demais tratamentos, as reações de PCR em tempo real foram conduzidas exclusivamente para os pontos experimentais em que o gene foi detectado como sendo diferencialmente expresso. Apenas os resultados validados são aqui mostrados. As amostras de RNA são provenientes de uma terceira réplica biológica. Todas as reações foram realizadas em paralelo e cada reação foi realizada em triplicata. As análises e cálculo das barras de erro foram calculados como descrito em Livak & Schmittgen (2001).

<i>S. officinarum</i>	<i>S. spontaneum</i>
Caiana Fita	IN8458
IK76108	IN8488
Lahaina	Krakatau
MZ151	SES 147b
MZ151 roxa	US56158
Sabura	US7440
Salangor	US851008
Sinimbu	UM721
NG213	UM691
Fiji 47	SES 194
Hinahina 18	IK7686
Manjri Red	US56193
Muntok Java	US571723
NG77142	
Soff 8268	
SS601	
Sylva	
NG2880	
Vae Vae Ula	
IJ76315	
IN8425	

Tabela 18. Genótipos de *S. officinarum* e *S. spontaneum* usados para os cruzamentos que geraram os indivíduos de alto e baixo brix analisados no presente estudo.

Figura 14. Cruzamentos realizados entre genótipos de *S. officinarum* e *S. spontaneum* para a obtenção de populações híbridas com alto e baixo teor de sacarose. Os gráficos mostram a frequência de indivíduos da progênie F3 correspondente a cada uma das classes de brix. As medidas de brix foram realizadas utilizando-se caldo de 500 indivíduos. Os tecidos para os 8 indivíduos extremos da população de alto brix e 8 para a população de baixo brix foram coletados e misturados para as análises por microarranjos. Para a quantificação por PCR em tempo real, amostras de RNA foram extraídas independentemente para cada indivíduo.

Oito genótipos de alto brix (AB) e oito com baixo brix (BB) foram selecionados a partir da progênie F3. A **Tabela 19** mostra os valores de brix e os resultados de quantificação de frutose, glicose e sacarose em entrenós desses indivíduos. Amostras de folha +1 e de entrenós 1, 5 e 9 foram coletadas. Amostras provenientes de oito indivíduos de AB foram comparadas com uma amostra de oito indivíduos BB através de hibridizações contra microarranjos de cDNA contendo 1.545 elementos válidos. A **Tabela 6** mostra como foram conduzidas essas hibridizações.

Empregando a análise de HT-self (Vencio & Koide, 2005) e considerando que um gene é considerado como diferencialmente expresso quando possui pelo menos 70% de seus pontos-réplicas acima ou abaixo das curvas intensidade-dependentes determinadas, foram encontrados 51 genes com expressão diferencial entre amostras de plantas de AB e as de BB. A **Tabela 20** mostra o perfil de expressão desses genes também nas amostras de órgãos analisadas.

Dos 51 genes detectados como diferencialmente expressos em pelo menos uma das amostras analisadas, 52% foram detectados nas amostras de folha. Nesse caso, os transcritos para três genes foram significativamente mais expressos nas amostras de AB: um fator de transcrição do tipo DREB1, um gene sem similaridades nos bancos de dados públicos e um gene codificando para uma proteína citocromo P450 monooxigenase. Já nas amostras de folhas de plantas de BB, 23 genes mostraram-se mais abundantes em relação às plantas de AB, sendo cerca da metade desses relacionados com patogenicidade ou respostas a estresses.

Nas amostras de entrenós, 15 genes foram encontrados como sendo enriquecidos na população AB. Um desses, um gene sem similares nos bancos de dados públicos (SCCCCL3001F04) mostrou-se diferencialmente expresso no entrenó imaturo (In1) e também no entrenó 5. Adicionalmente, foram identificados 13 genes mais expressos nas amostras da população de BB em pelo menos uma das amostras de entrenós. Esses transcritos pertencem a diferentes categorias funcionais, como ubiquitinação, respostas a estresses e transcrição. Genes sem homólogos nos bancos de dados públicos e genes de função desconhecida foram também identificados como diferencialmente expressos entre amostras de entrenós de AB e de BB.

Como forma de se validar alguns dos resultados de expressão encontrados pelas hibridizações em microarranjos de cDNA, foram conduzidas reações de PCR em tempo real

			% m / m					
	Genótipo	Brix (°Bx)	outros sacarídeos	Sacarose	Glicose	Frutose	fructose+glicose	Sacarose/monossacarídeos
Alto Brix	CTC98-241	18	0,021	7,311	1,322	0,988	2,31	3,165
	CTC98-242	18,6	0,016	9,183	1,430	1,014	2,444	3,757
	CTC98-243	19,2	0,021	10,956	0,649	0,602	1,251	8,758
	CTC98-244	14,6	0,024	11,161	0,633	0,646	1,279	8,726
	CTC98-246	18,8	0,019	10,974	0,709	0,545	1,254	8,751
	CTC98-252	18	0,011	6,370	0,840	0,579	1,419	4,489
	CTC98-253	19,6	0,018	11,120	0,660	0,643	1,303	8,534
	CTC98-258	18	0,012	6,739	1,116	0,865	1,981	3,402
Baixo Brix	CTC98-261	7	0,005	1,137	0,878	0,755	1,633	0,696
	CTC98-262	7,4	<0,003	1,368	0,968	0,823	1,791	0,764
	CTC98-265	6,4	<0,003	0,491	1,200	1,090	2,29	0,214
	CTC98-268	4,8	0,005	0,701	0,342	0,326	0,668	1,049
	CTC98-271	6	<0,003	0,923	1,098	0,992	2,09	0,442
	CTC98-272	6,8	0,003	1,069	0,725	0,632	1,357	0,788
	CTC98-277	7,4	0,004	1,576	0,774	0,716	1,49	1,058
	CTC98-279	7,8	0,006	1,735	1,318	1,066	2,384	0,728

Tabela 19. Medição de brix e de teores de açúcares para os 16 indivíduos selecionados para os ensaios de expressão gênica comparativo entre genótipos de alto e de baixo acúmulo de sacarose. A avaliação de teores de açúcares foi feita com amostras de entrenós 9 desses indivíduos, por HPLC.

SAS	Anotação	Folha ABxBB	1º entrenó ABxBB	5º entrenó ABxBB	9º entrenó ABxBB	Flor	Gema lateral	Folha	Raiz	Entrenó 1	Entrenó 4
SCEQRT1024E12.g	Biossíntese de SA - Fenilalanina amônia-liase	-						+	-	-	
SCSBHR1050B11.g	Desenvolvimento - Proteína associada à senescência			+							
SCJFRZ2007F10.g	Desenvolvimento - Proteína com repetições arm ARC1			-							
SCRUFL1112F04.b	Estabilidade de RNA - UDP-GlcNAc			-							
SCBGLR1099G02.g	Fator de transcrição - AP2/EREBP - DREB1	+									
SCCCRZ1004H12.g	Fator de transcrição - EIL			-							
SCCCRZ1001H05.g	Fator de transcrição - HLH		+								
SCCCLR2003E10.g	Fator de transcrição - NAC			-				+			
SCEZST3147A10.g	Fator de transcrição - Proteína com dedos de zinco - C3H			+							
SCJFLR1035E04.g	Fator de transcrição - Scarecrow	-									
SCCCLR1001E04.g	Fotossíntese - Rubisco - subunidade menor	-						+	-		
SCRFLR2037F09.g	Metabolismo de cálcio - Calreticulina		-					-		-	-
SCCCAM1001A03.g	Metabolismo de cálcio - Proteína ligadora de calmodulina				+						
SCAGLR2026G12.g	No match	-							+		
SCBFST3136A06.g	No match	-						+			
SCCCCL3001F04.g	No match (transcrito possivelmente não codificador)	+	+	+							
SCBGLR1115D10.g	No match (transcrito possivelmente não codificador)				+						
SCJFST1009H11.g	No match (transcrito possivelmente não codificador)		+					-			
SCEQRT1028C03.g	Patogenicidade - Inibidor de proteases - proteína PR	-				-	-	-		-	-
SCQSRT1036D03.g	Patogenicidade - Inibidor de proteases - proteína PR	-				-	-	-		-	-
SCEZRZ3098G10.g	Patogenicidade - Inibidor de proteases - Taumatina				+						
SCCCSD2001E05.g	Patogenicidade - Inibidor de proteases - Taumatina	-				-				-	-
SCEQRT1033F01.g	Patogenicidade - Proteína com dedos de zinco - C2C2/Dof			+							
SCEPLR1030B03.g	Patogenicidade - Semelhante à proteína LRP de tomate	-									
SCCCLR1068D03.g	Pequena GTPase - Rab	-									
SCCCLR1048F03.g	Proteína desconhecida - Similar a proteína hipotética de cloroplasto			-	-						
SCVPFL3046C06.b	Proteína fosfatase (Ser/Thr fosfatase) - Subunidade catalítica de PP2C				+						
SCJLLR1054C03.g	Proteína quinase - canePK-BIII7				+						
SCBFSB1046D04.g	Proteína quinase - SnRKs - caneCIPK-16				+						
SCACLR1036B06.g	Proteína quinase - SnRKs - caneCIPK-18	-				-		+			
SCEQRT2099H01.g	Proteína quinase - vias de cálcio - caneCDPK-27	-						+			
SCCCCL3120C09.g	Receptor Ser/Thr quinase - cane RLK com LysM-1	-									+
SCVPCL6041F01.g	Receptor Ser/Thr quinase - cane RLK with L-lectin domain-33				-						
SCQGSB1082E12.g	Receptor Ser/Thr quinase - caneRLK-C4	-									
SCJFRZ2012F04.g	Receptor Ser/Thr quinase - caneRLK-DIV3				+						
SCUTAM2005B03.g	Respostas a estresses - Citocromo P450 monooxigenase - CYP724B1				+						
SCQGHR1012B09.g	Respostas a estresses - Citocromo P450 monooxigenase				+						
SCSGHR1069F04.b	Respostas a estresses - Citocromo P450 monooxigenase	+									
SCSGFL4193B05.g	Respostas a estresses - Citocromo P450 monooxigenase - CYP73	-							+		
SCAGLR1043E04.g	Respostas a estresses - Citocromo P450 monooxigenase - CYP74A	-							+		
SCEQRT1026H08.g	Respostas a estresses - Citocromo P450 monooxigenase - CYP75	-						+			
SCVPFL3045B09.g	Respostas a estresses - Metalotioneína	-									
SCUTRZ2022G04.g	Respostas a estresses - Proteína de choque térmico	-									
SCCCRZ1002E08.g	Respostas a seca e a frio - Aquaporina (PIP1)	-			-						
SCCCRZ1001F02.g	Respostas a seca e a frio - Aquaporina (TIP)				-						
SCACCL6008H06.g	Respostas a seca e a frio - Proteína LTI	-				-					+
SCACLR2014E12.g	Ubiquitinação - Enzima de conjugação com ubiquitina E2	-									+
SCBGLR1003D06.g	Ubiquitinação - Enzima de conjugação com ubiquitina E2	-									
SCCCLR1C03H09.g	Ubiquitinação - Poliubiquitina			-	-						
SCVPRZ2038C12.g	Ubiquitinação - Poliubiquitina			-	-						
SCCCCL3080A11.b	Ubiquitinação - Poliubiquitina			-	-						

Tabela 20. SAS com expressão diferencial entre genótipos de cana-de-açúcar de alto e de baixo brix. A tabela lista todos os SAS que se mostraram diferencialmente expressos em ambas as réplicas técnicas. O sinal + indica que a expressão do SAS é maior nos genótipos de alto brix. O sinal - indica que a expressão do SAS é maior na população de baixo brix. A tabela indica também os resultados de expressão diferencial entre os seis órgãos de cana-de-açúcar obtidos para esses SAS. Nesses casos, o sinal + indica expressão preferencial no órgão indicado e o sinal -, que a expressão do SAS é menor no órgão indicado do que na referência comum (*Ref Pool*) adotada para esses experimentos.

utilizando-se amostras de folhas de três genótipos de AB (CTC98-243, CTC98-246 e CTC98-253) e três genótipos de BB (CTC98-261, CTC98-265 e CTC98-272). No caso de entrenó 9, foram utilizadas amostras referentes a dois genótipos de AB (CTC98-243 e CTC98-246) e dois de BB (CTC98-261 e CTC98-265). A **Tabela 21** mostra os genes selecionados para esses ensaios e os *primers* correspondentes. O normalizador usado para essas reações foi o gene de poliubiquitina, o mesmo empregado para a avaliação dos níveis de expressão entre os seis órgãos de cana-de-açúcar. Os resultados dos ensaios de PCR em tempo real são mostrados na **Figura 15**. Como pode ser visto, os níveis de expressão podem variar muito de indivíduo para indivíduo. Sendo assim, os resultados obtidos por PCR em tempo real foram comparados entre indivíduos de alto e de baixo brix, par a par, calculando-se, para isso, o valor P, como feito para os experimentos de respostas a fatores bióticos e abióticos (**Tabela 21**). Adotando-se como critério um valor P igual ou superior a 0,97, obtêm-se as porcentagens de validação mostradas, para cada um dos genes selecionados. Computando-se todos os dados conjuntamente, um índice de validação de 57% é obtido.

6.6. O Banco de Dados SUCAST

Todos os SAS catalogados pelo Projeto SUCAST e os dados de expressão correspondentes gerados pelo presente projeto foram compilados em um banco de dados pelo aluno de mestrado Renato Vicentini (UNICAMP) e está disponível no *site* do projeto SUCEST-FUN (<http://www.sucest-fun.org>). O banco SUCAST integra as informações geradas pela categorização filogenética, as análises de agrupamento por SOM, a matriz de razões virtuais de expressão entre os seis órgãos de cana-de-açúcar e os genes diferencialmente expressos com os dados de sequência e análises dessas. É possível, através do banco de dados SUCAST, pesquisar por determinado SAS e informações a ele associadas. Para cada SAS incluído no banco de dados SUCAST, é disponibilizada a informação da sequência do seu consenso e o alinhamento das sequências correspondentes, de acordo com o agrupamento das sequências SUCEST realizada por Vettore *et al.* (2003). O banco de dados SUCAST inclui todos os dados de expressão

SAS	Categoria	Subcategoria	Comparação entre genótipos (ABxBB)	P de ser mais expresso em BB	P de ser mais expresso em AB	%validação
SCEQRT1024E12.g	Biossíntese de SA - Fenilalanina amônia-liase		243X261	0,02	0,98	0%
Primers: CTTCCAGGGCACTCCATT GAGAACTGCGCGAACATGAG			243X265	0,64	0,36	
			243X272	0,03	0,97	
			246X261	0,02	0,98	
			246X265	0,75	0,25	
			246X272	0,03	0,97	
			253X261	0,01	0,99	
			253X265	0,67	0,33	
			253X272	0,02	0,98	
SCCCRZ1002E08.g	Resposta a estresses - Provável aquaporina (PIP1)		243X261	0,99	0,01	78%
Primers: AGGCATTGGAACAACCATGA GCTTTCAGATGCCGATTCAAG			243X265	1	0	
			243X272	1	0	
			246X261	0,99	0,01	
			246X265	1	0	
			246X272	1	0	
			253X261	0,29	0,71	
			253X265	0,57	0,43	
			253X272	0,97	0,03	
SCACCL6008H06.g	Resposta a estresses - Proteína LTI		243X261	0,99	0,01	100%
Primers: AATCCCATCCATCCAAGCTAAG CGGCGGCCGATCCT			243X265	1	0	
			243X272	1	0	
			246X261	1	0	
			246X265	1	0	
			246X272	1	0	
			253X261	0,99	0,01	
			253X265	1	0	
			253X272	1	0	
SCAGLR1043E04.g	Citocromo P450 monooxigenase- CYP74A		243X261	0,97	0,03	50%
Primers: TGAAGCGGACGAATTTGAGTAG AGCTGCCATAGAGACTTGAT			243X265	0,57	0,43	
			253X261	1	0	
			253X265	0,64	0,36	
SCCCLR2003E10.g	Fator de transcrição - NAC		243X261	0,89	0,11	50%
Primers: CATCTTCTCCCACTCGTTCTTCTT AGGGATCGCTCAGCTGGAT			243X265	0,94	0,06	
			253X261	1	0	
			253X265	1	0	

Folha

Entrenó 9

Tabela 21. Genes de cana-de-açúcar selecionados dentre os diferencialmente expressos comparando-se uma população de AB com uma de BB para validação por PCR em tempo real (em amostras de folha ou entrenó 9), primers desenhados e resultados obtidos. Os primers foram desenhados utilizando-se o Primer Express 2.0 Software (Applied Biosystems) e buscas de BLAST foram conduzidas contra o banco de dados SUCEST de forma a garantir a especificidade dos primers desenhados. A tabela mostra, para cada comparação par a par entre genótipos de AB versus de BB, a probabilidade P do gene ser mais expresso em AB e a probabilidade de ser mais expresso em BB. Considerando-se como significativo somente os valores de P igual ou acima de 0,97, os valores destacados em verde na tabela referem-se aos casos onde o resultado de PCR em tempo real é concordante com o obtido por microarranjos. A porcentagem de validação para cada um desses genes isoladamente é também mostrada na tabela.

Figura 15. Avaliação da expressão por PCR em tempo real de genes selecionados como diferencialmente expressos comparando-se populações contrastantes para teor de brix. Os níveis de mRNA foram avaliados em amostras de folha ou de 9º entrenós, dependendo do caso. Três indivíduos de alto brix (AB) e três de baixo brix (BB) foram analisados no caso de folhas e dois indivíduos de AB e dois de BB foram analisados para amostras de 9º entrenós. As reações foram realizadas em triplicatas. O eixo y refere-se aos níveis de mRNA do gene de interesse em relação aos níveis de mRNA para o gene de poliubiquitina (PUB), usado como normalizador. Esse gene de poliubiquitina foi o mesmo utilizado nos experimentos de comparação de seis órgãos. O cálculo das barras de erro foram realizados como descrito em Livak & Schmittgen (2001).

referentes a respostas a estímulos ambientais e a fitormônios, bem como dados de expressão nos seis órgãos de cana-de-açúcar estudados (dados publicados em Papini-Terzi *et al.*, 2005).

Além de dados de expressão, o banco SUCAST também disponibiliza para cada SAS pesquisado informações referentes à sua anotação, resultados de buscas de BLAST contra os bancos de dados do GenBank e GEO (*Gene Expression Omnibus*) (Barrett *et al.*, 2005) e comparação com o banco de dados do Gene Ontology (Gene Ontology Consortium, 2006). É também possível obter a sequência predita de proteína codificada pela sequência consenso, bem como a presença de domínios conservados preditos pelo banco de dados Pfam para cada um dos SAS catalogados (Sonnhammer *et al.*, 1998).

Mais de 3.500 SAS estão incluídos no banco de dados SUCAST. Esses elementos estão distribuídos em 46 categorias principais, subdivididas em 1.678 subcategorias. Quatrocentos e dezesseis SAS não apresentam homólogos no banco de dados GenBank nr (usando-se a ferramenta BLASTX, valor de E para corte de 10^{-5}) e 1.083 SAS apresentaram similaridades contra sequências de proteínas incluídas no banco de dados de expressão do GEO (usando-se a ferramenta BLASTN, valor de E para corte de 10^{-5}). A comparação desses SAS com o banco de dados do Gene Ontology revelou *hits* significantes com proteínas para 3.030 SAS (usando-se a ferramenta BLASTX, valor de E para corte de 10^{-5}).

7. Discussão

7.1. Análise *in silico* do quinoma de cana-de-açúcar

A busca por componentes relacionados principalmente à transdução de sinal utilizando o algoritmo BLAST indicou mais de 3.500 componentes que foram catalogados no banco de dados SUCAST. Dentre as categorias catalogadas, os SAS referentes ao quinoma de cana-de-açúcar foram selecionados para serem melhor caracterizados através de ferramentas de bioinformática. Inicialmente, as seqüências de aminoácidos codificadas a partir das seqüências consenso dos SAS foram inspecionadas quanto a domínios conservados como predito por buscas no banco de dados Pfam (Sonnhammer *et al.*, 1998). Os resultados obtidos são mostrados na **Figura 1**. Esses SAS foram também analisados com relação a domínios conservados por buscas contra o banco de dados SMART (Schultz *et al.*, 1998) e essas análises geraram resultados semelhantes aos gerados pelas buscas no Pfam, com exceção para alguns domínios (como LysM, TPR e *U-box*, discutidos abaixo) que constam apenas no banco de dados SMART.

As proteínas quinases e receptores quinases identificados contêm uma prevalência de domínios de *pkinese*. Além desses, há um considerável número de proteínas quinases sem domínios no Pfam (77) e com a combinação dos domínios *pkinese* + *EF-hand* (17). Domínios funcionais do tipo *EF-hand* são encontrados na porção C-terminal de proteínas quinases do tipo CDPKs e CCaMKs (Hrabak *et al.*, 2003). Esses domínios são semelhantes aos *EF-hands* também encontrados em calmodulina e representam domínios de ligação com Ca^{2+} . Nenhuma quinase do tipo CCaMK foi encontrada em *Arabidopsis* (Hrabak *et al.*, 2003). De forma concordante com esse fato, nenhuma proteína quinase dessa categoria foi encontrada no banco de dados SUCEST. Sendo assim, todos os SAS encontrados com *pkinese* + EF-hand pertencem à categoria de CDPKs.

Os RLKs, além do domínio *pkinese*, apresentam um elevado número de repetições do tipo LRR. Desses, 39 possuem LRRs e *pkinese* e 89 possuem apenas LRRs e provavelmente correspondem a seqüências incompletas, sem o domínio de quinase. Domínios LRR

freqüentemente participam de interações proteína-proteína e são encontrados no domínio extracelular de vários RLKs, como CLV1, BRI1 e ERECTA (Shiu & Bleecker, 2001a). O número de repetições, a localização e a presença ou não de uma “ilha” entre repetições de LRR caracterizam diferentes subfamílias (Shiu & Bleecker, 2001a). Cento e cinquenta e um prováveis receptores de cana-de-açúcar são similares a receptores com LRRs de outros organismos, além de 82 prováveis RLKs com LRRs que não puderam ser associados a tipos específicos de RLKs com base nas buscas por BLAST. É importante notar que muitos desses SAS têm sequências incompletas e, por isso, podem não ter apresentado domínios LRR preditos pelo Pfam.

A análise *in silico* das proteínas quinases de cana-de-açúcar também se baseou em uma abordagem filogenética, que classificou essas proteínas em seis grupos (**Figura 2** e **Figura 3**). Analisando-se os RLKs/RLCKs isoladamente, quatro grupos foram obtidos para esses componentes (**Figura 2** e **Figura 4**). A análise filogenética permitiu que diversas proteínas quinases antes classificadas como indefinidas (117 RLKs, 34 RLCKs e outras 32 proteínas quinases) pudessem ser classificadas em grupos e/ou famílias. Foram obtidas seis famílias constituídas inteiramente por proteínas quinases indefinidas e 32 constituídas por RLKs/RLCKs indefinidos, todas suportadas por um valor de *bootstrap* acima de 50%. A análise indicou que muitas das proteínas quinases indefinidas agruparam com algumas sequências de proteínas já conhecidas e caracterizadas, o que pode ser indicativo de uma provável função para essas proteínas. Há, por exemplo, um SAS anotado como proteína quinase indefinida dentro do grupo KB com um provável domínio de anquirina, um dos mais importantes para a interação proteína-proteína (Bennett & Chen, 2001), predito pelo Pfam. No catálogo PlantsP (Gribskov *et al.*, 2001) encontramos também uma proteína (At5g13530) com esse tipo de domínio agrupada com as quinases ATN1/CTR1/EDR1, proteínas quinases encontradas também no grupo KB.

Seis proteínas quinases do grupo KE possuem domínios catalíticos muito similares, com uma inserção de aproximadamente 80 aminoácidos entre os subdomínios VII e VIII (como definidos em Hanks *et al.*, 1988, Hanks & Quinn, 1991, Hanks & Hunter, 1995). Essas proteínas são semelhantes à proteína quinase G11A de arroz (Lawton *et al.*, 1989). Como G11A ainda é pouco caracterizada e também contém essa sequência de inserção, será interessante investigar

futuramente alvos de quinases com essa inserção e também avaliar se a sequência de inserção possui algum papel importante no reconhecimento do substrato e/ou na reação catalítica.

Duas famílias distintas de MAP quinases foram obtidas: uma pertencente ao grupo KD (junto com MAPKKs e MAKKKs) e outra, ao grupo KF. As famílias de MAP quinases dos grupos KD e KF correspondem, respectivamente, às famílias 4.1.5 e 4.5.1 do banco de dados PlantsP, pois agruparam com sequências de MAP quinases dessas famílias de *Arabidopsis*. A categorização do PlantsP classifica as MAPKs dentro da família 4.1.5 como “prováveis MAP2Ks”, dada a alta similaridade dessas quinase também com as MAP2Ks (MAPKKs). Sendo assim, as MAPKs de cana-de-açúcar pertencentes ao grupo KD foram também denominadas como “prováveis MAP2Ks”.

Como já relatado anteriormente para RLKs de *Arabidopsis* (Shiu & Bleecker, 2001b), membros de subfamílias de RLKs tendem a ter domínios extracelulares similares, apesar da filogenia ser baseada apenas no domínio catalítico. Ou seja, similaridades estruturais no domínio de quinase muitas vezes refletem também semelhanças estruturais nos domínios extracelulares dos receptores. Como exemplo, pode-se citar a categoria de RLKs com domínio de lectina. Essa categoria de RLKs está representada na filogenia pela família “L-lectina”, dentro do grupo RB e contém 12 SAS, dos quais dois têm domínios de L-lectina preditos pelo Pfam. Interessantemente, nenhum receptor com motivos de C-lectina foi identificado em cana-de-açúcar.

Uma das famílias de receptores indefinidos do grupo RC apresenta dois membros com LysM preditos pelo Pfam. No banco SUCEST, há apenas três quinases com LysM predito pelo Pfam. LysM representa um motivo de ligação a peptidoglicana (Bateman & Bycroft, 2000) e foi postulado como sendo relacionado com a percepção de fragmentos de parede celular bacteriana (Shiu & Bleecker, 2001a). Alguns receptores com LysM foram recentemente caracterizados como sendo relacionados com a percepção de fatores Nod produzidos e secretados por microsimbiontes (Limpens *et al.*, 2003, Madsen *et al.*, 2003, Radutoiu *et al.*, 2003). Com base nas informações disponíveis até o momento, não é possível a determinação de qual seria a função exercida por esses receptores em cana-de-açúcar. Ainda dentro do grupo RC, encontra-se uma família contendo quatro RLKs indefinidos, sendo que três desses apresentam domínio de *U-box* predito

pelo banco de dados SMART (Schultz *et al.*, 1998). Esse domínio parece desempenhar um papel na ubiquitinação de proteínas (Aravind & Koonin, 2000) e é também encontrado em outras quinases de plantas ainda não caracterizadas. Um dos SAS dentro dessa família parece codificar para uma proteína de estrutura complexa, que parece conter domínios TPR, um domínio de *pkinese* e um domínio de *U-box* próximo ao C-terminal, conforme predito pelo SMART (Schultz *et al.*, 1998). Esse gene é altamente similar com outras proteínas quinases de planta, possuindo maior similaridade (74% de identidade de aminoácidos) com uma quinase de arroz (gi|55741366|) ainda não caracterizada. As repetições do tipo TPR parecem estar envolvidas com interações do tipo proteína-proteína ou com a formação de complexos macromoleculares e podem estar relacionadas a diferentes eventos, como o controle do ciclo celular, a repressão da transcrição e respostas a estresses (Goebl & Yanagida, 1991, Lamb *et al.*, 1995). O domínio *U-box* desempenha um papel no processo de ubiquitinação (Aravind & Koonin, 2000). Uma busca por proteínas contendo o domínio de *U-box* no genoma de *Arabidopsis* revelou 37 proteínas contendo esse domínio (Azevedo *et al.*, 2001). Essas proteínas foram classificadas em cinco subclasses distintas, sendo uma delas (classe IV) constituída por quatro proteínas contendo domínios de *pkinese* e de *U-box*. Contudo, a função dessas proteínas ainda não foi determinada. É importante notar que não é predita região transmembrânica pelo SMART para as proteínas codificadas para esses SAS e as análises de BLAST apontam para similaridades tanto com proteínas quinases quanto para RLKs para esses SAS. Dessa forma, os SAS codificadores para quinases com *U-box* encontrados para cana-de-açúcar podem, na verdade, ser RLCKs, e não RLKs.

Os receptores-S formaram duas famílias principais, no grupo RB. Receptores de plantas já bem caracterizados e relacionados com o crescimento e desenvolvimento, como CLV1 (Clark *et al.*, 1997), ERECTA (Torii *et al.*, 1996), BRI1 (Li & Chory, 1997) e receptor de PSK (Matsubayashi *et al.*, 2002) estão agrupados em RD, grupo que se caracteriza pelo elevado número de receptores com LRRs. Ainda no grupo RD pode ser encontrada a família RLK-DVI, composta por 11 receptores indefinidos, constituindo a maior família de quinases indefinidas identificada pela análise filogenética e suportada por um valor de *bootstrap* de 95%.

Foram detectadas, em cana-de-açúcar, RLCKs similares às quinases NAK (Moran & Walker, 1993), APK1 (Hirayama & Oka, 1992), PBS1 (Swiderski & Innes, 2001), Pti-1 (Zhou *et al.*, 1995) e NPK15 (Ito *et al.*, 1994). Algumas sequências sem regiões transmembrânicas preditas e cujos melhores similares por buscas por BLAST são proteínas quinases indefinidas agruparam com RLKs/RLCKs na análise filogenética. Por essa razão, esses SAS foram classificados como prováveis RLCKs ao invés de Proteínas Quinases Indefinidas. De fato, algumas dessas sequências podem representar novos tipos de RLCKs ainda não caracterizadas. Alternativamente, alguns desses SAS também podem codificar para RLKs, mas possuem seqüências incompletas, de tal forma que não possuem as regiões de domínio extracelular e transmembrânica seqüenciadas, que indicariam que essas quinases são receptores.

Por último, é importante ressaltar que a análise filogenética conduzida para o quinoma de cana-de-açúcar possui limitações pelo fato das sequências serem provenientes do sequenciamento de ESTs. Assim sendo, muitas proteínas quinases de cana-de-açúcar foram excluídas da análise por não apresentarem domínio catalítico predito ou por não terem pelo menos 50% do domínio catalítico sequenciado. Entretanto, os grupos obtidos estão de acordo com os grupos e famílias de proteínas quinases previamente definidos por Shiu & Bleecker (2001b) e pelo banco de dados PlantsP, apesar desse último utilizar uma metodologia diferente para a categorização das quinases (Gribskov *et al.*, 2001).

7.2. Padronização das condições para análise dos dados de microarranjos

Para determinarmos o perfil de expressão gênica dos componentes de transdução de sinal, a metodologia de escolha foi a de microarranjos de DNA, visto que essa técnica permite a análise de um número elevado de genes simultaneamente, podendo indicar genes relacionados a processos de interesse, revelando respostas a tratamentos ou específicas a determinado órgão. Entretanto, resultados de microarranjos devem ser baseados em análises estatísticas cuidadosas que minimizem a ocorrência de falso-positivos e falso-negativos.

A extração dos dados foi feita utilizando-se a medida de densidade MTM subtraída do valor da intensidade de fundo. Foram descartados *pixels* que estavam fora da faixa de ± 4 MADs. A medida da densidade de *pixels* pelo MTM é vantajosa por se basear na medida da mediana, que é menos influenciada por valores extremos do que a média. O valor de 4 MADs foi estipulado como suficiente para garantir que os *pixels* indesejáveis fossem descartados sem que se perdesse muito dos sinais obtidos.

Para a padronização do método de normalização a ser adotado, foi evitada a utilização do método de normalização pela intensidade total, já que esse não leva em consideração erros sistemáticos que ocorrem com os dados, como a maior variação dos valores de razão para pontos de menor intensidade. A técnica de normalização com base na inversão dos fluoróforos (*dye-swap*), na qual cada hibridização deve ser realizada no mínimo duas vezes, sendo que em cada hibridização uma amostra é marcada com um fluoróforo diferente, também não foi adotada, já que esse método requereria um maior número de lâminas a serem hibridizadas. As alternativas de normalização testadas foram os métodos de LOWESS e de normalização com base nos dados do *Scorecard*, que funcionam como controles internos para a normalização. Os controles fornecidos pelo kit *Lucidea Universal ScoreCard* (GE Healthcare) correspondem a controles de calibração e de razão, bem como a controles negativos. Através da adição de *spikes* às amostras de RNA utilizadas para a síntese dos alvos e da hibridização dos alvos sintetizados contra microarranjos contendo também depositados os “genes artificiais” fornecidos por esse kit, é possível ter os valores de intensidade para esses controles. Os valores de intensidade dos controles de calibração (proporção 1:1 entre as duas amostras comparadas entre si) podem ser utilizados para a normalização dos dados de microarranjos. A comparação entre a normalização por LOWESS e pelos dados de *ScoreCard* é mostrada na **Figura 5**. A normalização pelo *Scorecard* (**Figura 5A**) apresentou o efeito de gerar um deslocamento geral dos pontos em direção a um dos canais, como se a grande maioria dos genes se mostrasse mais expressa em uma das amostras em relação à outra. Esse deslocamento dos pontos é explicável, já que os *spikes* do sistema *ScoreCard* são controles exógenos às amostras de RNA em análise. Com isso, erros de pipetagem, mesmo que pequenos, no volume de *spike* adicionado à reação levam a uma distorção

geral dos pontos no momento da normalização. Já a normalização por LOWESS (**Figura 5B**) não apresentou essa distorção. Além disso, a normalização por LOWESS tem também a vantagem de ser um método baseado no fato da variabilidade dos valores de razão para determinado gene ser dependente dos seus respectivos valores de intensidade, o que é visualizado como um desvio maior de $M = \text{zero}$ para pontos de baixa intensidade (Yang *et al.*, 2002b). Os estudos de Park *et al.* (2003) mostraram que métodos de normalização intensidade-dependentes possuem melhor desempenho do que os métodos baseados em normalização global.

Além da normalização dos dados extraídos, outro ponto importante para a análise de dados de experimentos com microarranjos é a definição do critério de genes diferencialmente expressos. Comumente, opta-se pela adoção de um valor arbitrário, onde um gene é considerado diferencialmente expresso se o valor da sua razão normalizada entre os dois canais for duas vezes maior ou duas vezes menor na amostra experimental em relação à controle. Apesar de essa estratégia ser comumente empregada, sabe-se que o uso de um valor único para determinação de padrão de expressão diferencial não é o mais adequado para diferentes faixas de valores de intensidades. Com isso, um gene pode ser considerado diferencialmente expresso quando as alterações no seu valor de razão referem-se somente a flutuações aleatórias. Essa observação fica nítida quando são realizadas hibridizações homotípicas (Yang *et al.*, 2002a), ou seja, hibridizações do tipo *self-self*, onde alvos provenientes da mesma amostra de RNA, porém marcados com diferentes fluoróforos, são hibridizados entre si. Nesse caso, teoricamente espera-se que todos os pontos apresentem $M = \text{zero}$ (o que equivale a uma razão de expressão entre as duas amostras igual a um). Entretanto, a análise do gráfico $M \times S$ para esses tipos de experimentos mostra que, mesmo para hibridizações homotípicas, nem todos os pontos encontram-se precisamente na região de $M = \text{zero}$ (**Figura 7A**). O deslocamento de alguns pontos para fora da região de $M = \text{zero}$ é devido a flutuações experimentais aleatórias inerentes à metodologia de microarranjos de cDNA. Observa-se, também, que os pontos se mostram mais dispersos para valores menores de intensidade. Com isso, conclui-se que a adoção de um valor arbitrário como critério de diferencialmente expresso pode ser um valor subestimado para pontos de baixa intensidade e superestimado para pontos de intensidade alta. Sendo assim, para a determinação dos valores de

razão da intensidade entre os dois canais acima ou abaixo dos quais um gene seria considerado com diferencialmente expresso, foram realizadas hibridizações homotípicas (*self-self*, como descrito em Yang *et al.*, 2002a) e a partir dos dados obtidos foram definidos diferentes valores limite de acordo com o nível de intensidade (**Figura 7A**). Esse método de detecção de genes diferencialmente expressos (HT-self) foi descrito por Vencio & Koide (2005). Esses intervalos de corte foram utilizados para a determinação de genes diferencialmente expressos entre os seis órgãos estudados (raiz, flor, folha, gema lateral, primeiro e quarto entrenós). Levando-se em conta a reprodutibilidade dos dados, um evento foi considerado como sendo expressão diferencial quando pelo menos 70% dos pontos-réplicas mostraram valores acima ou abaixo dos valores de corte determinados para a respectiva faixa de intensidade de sinal e resultados concordantes foram obtidos para as duas amostras biológicas do mesmo órgão. Por outro lado, um gene foi considerado homogeneamente expresso em todos os órgãos quando pelo menos 55% de todos os seus pontos em todos os experimentos de hibridizações com amostras de órgãos apresentaram-se situados dentro das faixas delimitadas pelos valores de corte (**Figura 7B e 7C**).

Para a análise dos dados provenientes de comparações entre genótipos de alto e de baixo brix, foi também aplicado o método HT-self (Venio & Koide, 2005), com a diferença de, nesse caso, os limites de corte terem sido determinados como curvas contínuas, e não como faixas de valores de M dependentes da intensidade. Também foi aplicado para esses dados o conceito de reprodutibilidade mostrado na **Figura 7**, com critério de 70%.

Já para a análise dos dados gerados para hibridizações referentes a amostras de plantas submetidas a tratamentos com hormônios ou a estímulos ambientais, adotou-se a busca de valores extremos (Koide *et al.*, 2006), já que o método HT-self não pareceu adequado nesse caso (**Figura 11**). Para esses casos, os critérios para definição das curvas de corte (80%) e de reprodutibilidade (no mínimo 60% dos pontos-réplicas acima ou abaixo das curvas de corte determinadas para cada uma das amostras biológicas utilizadas) podem parecer muito pouco estridentes em relação aos adotados nos outros experimentos. Entretanto, deve-se ressaltar que, no método de busca de valores extremos, são utilizados os dados do próprio experimento para a definição das curvas de

corte, onde, ao contrário de hibridizações homotípicas, espera-se que ocorram variações nos níveis de expressão de alguns dos transcritos.

7.3. Análise de perfis de expressão em seis órgãos de cana-de-açúcar

Para os experimentos de hibridização com amostras de órgãos de cana-de-açúcar, o uso da mistura equimolar de RNA de diferentes órgãos (*Ref Pool*) como uma referência comum para todas as hibridizações de amostras de órgãos foi a solução encontrada para solucionar o problema da ausência de uma referência comum, como, por exemplo, a amostra controle (não tratada) para experimentos de respostas a estresses. A metodologia de microarranjos só permite a comparação direta entre duas amostras e realizar todas as combinações possíveis de hibridizações entre amostras de órgãos requereria um número muito elevado de lâminas a serem hibridizadas. Contudo, há o problema do resultado obtido a partir de hibridizações contra *Ref Pool* não ser um resultado direto, já que são gerados dados apenas relativos e não absolutos. Para contornar esse problema, a comparação direta dos níveis de expressão entre dois órgãos pode ser inferida a partir das razões virtuais calculadas entre pares de órgãos, utilizando a amostra *Ref Pool* como denominador comum. As razões virtuais foram assim denominadas por não representarem o resultado direto de um experimento de microarranjo. Como as razões entre dois órgãos são inferidas, ou seja, obtidas indiretamente, é de se esperar que esse resultado tenha um erro associado maior do que aquele obtido por uma hibridização direta. Speed & Yang (2002) já discutiram matematicamente como inferências indiretas sempre resultam em erros maiores do que inferências diretas. Dadas essas observações, optou-se por trabalhar sempre com a mediana das razões diretas (órgão x *Ref Pool*) para o agrupamento hierárquico dos dados e para a utilização dos critérios de expressão diferencial e os dados de razões virtuais foram utilizados de forma complementar a essas análises. A razão dos sinais obtidos para a amostra de um determinado órgão de cana-de-açúcar *versus Ref Pool* pode ser traduzida como o nível de expressão gênica nesse órgão em relação à média de expressão nos diferentes órgãos em estudo. Apesar das limitações, o uso da *Ref Pool* apresenta algumas vantagens. Em Kim *et al* (2002), o uso de

diferentes referências é discutido para hibridizações em microarranjos, sendo que o uso de uma mistura de amostras de RNA mostrou-se uma referência comum melhor do que o uso de DNA genômico para tal finalidade. Além disso, como a amostra *Ref Pool* é constituída por uma mistura equimolar de RNA de diferentes órgãos, a probabilidade de um elemento do microarranjo não se mostrar expresso em nenhum dos órgãos componentes dessa mistura é baixa. Com isso, consegue-se aumentar o índice de pontos do microarranjo com sinal significativamente maior do que o ruído de fundo (*background*), o que melhora a qualidade dos dados gerados.

É importante ressaltar que as amostras de um mesmo órgão utilizadas nos experimentos não constituem réplicas biológicas verdadeiras, já que a idade das plantas, grau de desenvolvimento do órgão, morfologia e o modo de cultivo não são necessariamente iguais entre as amostras coletadas. Com isso, adotando-se os critérios já discutidos acima, se um determinado gene é considerado enriquecido nas raízes, por exemplo, pode-se acreditar com maior confiabilidade que esse gene é mais expresso nesse órgão de uma forma geral e não apenas em uma condição específica. Essa maior segurança nos perfis de expressão obtidos é necessária, já que, como perspectiva, os genes detectados como diferencialmente expressos podem ser utilizados para a clonagem de promotores de genes órgão-específicos, o que será de grande valia para a direção da expressão de transgenes para o órgão de interesse.

Os experimentos para avaliação da variabilidade nos níveis de expressão entre indivíduos cultivados em campo mostraram que essa pode ser considerada pequena (**Figura 6**). Esse resultado é esperado, pois a propagação da cana-de-açúcar é vegetativa, sendo a variabilidade genética entre os indivíduos baixa. Com isso, conclui-se que não é necessário um elevado número de indivíduos para garantir a representatividade da amostra.

Para a maior parte dos genes presentes nos microarranjos há pouca informação em relação à distribuição dos transcritos em diferentes órgãos. Muitos dos relatos de perfil de expressão gênica em diferentes órgãos referem-se a análises para poucos genes e para vias metabólicas específicas, envolvendo famílias de genes altamente relacionados, como os fatores de transcrição MADS (Alvarez-Buylla *et al.*, 2000) e genes envolvidos em metabolismo de lipídeos (Beisson *et al.*, 2003), e freqüentemente baseiam-se apenas em dados *in silico* referentes ao

sequenciamento de ESTs. Hibridizações de microarranjos usando amostras de sete órgãos distintos de milho permitiram a elucidação de relações entre os órgãos e a detecção de expressão gênica órgão-específica (Cho *et al.*, 2002). Estudos de expressão gênica em diferentes órgãos foram também relatados para *Arabidopsis* (Obayashi *et al.*, 2004) e arroz (Ma *et al.*, 2005). Nesse último caso, o transcriptoma de arroz foi avaliado em diferentes órgãos através de hibridizações em microarranjos de oligonucleotídeos e detectou-se uma elevada porcentagem de genes de arroz com um perfil de expressão similar (76% a 88% de 6.314 genes avaliados, dependendo do órgão em questão) ao de seus respectivos ortólogos em *Arabidopsis* ao se comparar órgãos correlatos. Esse dado é particularmente interessante, pois mostra que os resultados de expressão em diferentes órgãos obtidos para cana-de-açúcar podem ser também úteis para se inferir preliminarmente o perfil de expressão de genes homólogos em outras plantas. O mapeamento espacial do transcriptoma de diversas espécies de plantas poderá ajudar a elucidar a regulação de muitas vias de desenvolvimento e de respostas a estímulos ambientais.

Os resultados de microarranjos permitiram a identificação de 217 genes diferencialmente expressos em ambas as amostras biológicas em pelo menos um dos órgãos analisados (**Tabela 8**). Esses elementos foram agrupados por UPGMA de acordo com os seus perfis de expressão, evidenciando grupos de genes com expressão pronunciada em folhas, raízes e entrenós (**Figura 8**). Grupos menores de genes com expressão preferencial em flores e gemas laterais foram também identificados.

Dentre os genes detectados como diferencialmente expressos, 24 codificam para fatores de transcrição. Dentre esses, encontram-se um gene codificando para proteína do tipo MADS-box e um codificador para a proteína YABBY, mais expressos nas amostras de flores. Já se relatou que fatores de transcrição do tipo MADS e a proteína YABBY, com dedos de zinco, possuem importantes papéis durante o desenvolvimento dos órgãos e mostraram níveis de expressão aumentados em flores de *Arabidopsis thaliana* (Henning *et al.*, 2004).

Um SAS enriquecido nas flores e que, de acordo com dados do Projeto SUCEST (Vettore *et al.*, 2003) já havia sido classificado como sendo específico de órgão floral (já que é composto por ESTs provenientes de bibliotecas de cDNA construídas a partir de amostras de flor) é muito

similar aos genes *MSP1* de arroz e *EXS/EMS1* de *Arabidopsis*. *MSP1* e *EXS/EMS1* são genes expressos em órgãos reprodutivos e, dentre outras funções, agem controlando o destino das células reprodutivas (Canales *et al.*, 2002, Zhao *et al.*, 2002, Nonomura *et al.*, 2003). Essas observações indicam que o SAS pode representar um ortólogo em cana-de-açúcar dos genes *MSP1* e *EXS/EMS1* e estar envolvido em vias de sinalização nos órgãos reprodutivos.

Doze receptores de função desconhecida mostraram-se predominantes ou com expressão diminuída em pelo menos um dos órgãos estudados. O perfil de expressão para esses receptores pode ser muito útil para a futura caracterização funcional dessa categoria. Foram obtidos cinco receptores indefinidos do grupo RD (*caneRLK-DV1*, *caneRLK-DV4*, *caneRLK-DVIII3*, *caneRLK-DXIV1*, *caneRLK-DXIV2*) com um padrão de expressão similar, geralmente caracterizado por uma menor expressão em folhas e/ou em raízes.

Dentro dos genes mais expressos em folhas, encontram-se dois codificando para CDPKs (*caneCDPK-27* e *caneCDPK-36*), indicando a presença de uma possível via dependente de cálcio envolvendo essas quinases nas folhas. Essas CDPKs podem estar envolvidas com o controle da abertura dos estômatos em células-guarda, como já descrito para uma CDPK de *Vicia faba* (Li *et al.*, 1998). É possível, inclusive, que a ativação de CDPKs em células-guarda esteja relacionada com a inibição da abertura dos estômatos mediada por ABA, que leva ao aumento de cálcio intracelular (Assmann & Shimazaki, 1999).

Ainda com relação a uma provável via dependente de cálcio nas folhas, também foi observada uma proteína quinase do tipo CIPK (*caneCIPK-18*) com expressão enriquecida nas folhas. Proteínas quinases dessa categoria agem via a interação com proteínas semelhantes à calcineurina B, que são proteínas sensoras de cálcio (Shi *et al.*, 1999, Kim *et al.*, 2000) e essa interação ocorre via um domínio de 24 aminoácidos denominado domínio NAF (Albrecht *et al.*, 2001), domínio esse que se encontra presente na sequência predita da proteína CIPK em questão.

Um SAS anotado como codificador para um provável receptor semelhante à categoria SERK (*caneSERK-4*) encontra-se expresso preferencialmente em folhas. A sequência predita de aminoácidos para esse SAS apresenta 92% de identidade com a sequência da proteína *SbRLK1* de *Sorghum bicolor*. O gene *SbRLK1* é expresso preferencialmente nas células do mesófilo e é

muito pouco expresso nas raízes. Acredita-se que esse gene possa estar envolvido com a regulação da fotossíntese e desenvolvimento das células da folha em *S. bicolor*. Assim como o referido receptor de cana-de-açúcar, esse gene de *S. bicolor* também apresenta elevada similaridade com a sequência do receptor do tipo SERK de *Daucus carota* (Annen & Stockhaus, 1999).

É também interessante notar que diversos genes relacionados à maquinaria de ubiquitinação mostraram-se mais expressos nos entrenós do que nos demais órgãos examinados. A via de ubiquitina/proteossomo 26S está relacionada à degradação específica de proteínas e é usada para controlar o nível e a atividade de proteínas em diversas vias metabólicas (Vierstra, 2003). No caso de cana-de-açúcar, uma atividade intensa de degradação de proteínas pode estar relacionada à especialização dos entrenós para o armazenamento de sacarose.

Um grupo de genes relacionados a hormônios se mostrou expresso preferencialmente em folhas e é composto por membros de vias de biossíntese de ácido salicílico (fenilalanina amônia-liase), ABA (zeaxantina epoxidase) e etileno (ACC oxidase e ACC sintase) (**Figura 8**). Um outro grupo de elementos relacionados a hormônios, incluindo quatro nitrilases e três lipoxigenases, mostrou expressão prevalente nas raízes. As lipoxigenases (LOX) constituem uma classe de dioxigenases funcionalmente diversas relacionadas a processos fisiológicos tais como crescimento, senescência e respostas a estresses e que apresentam diferentes perfis de expressão em diferentes plantas (Kolomiets *et al.*, 2001, Porta & Rocha-Sosa, 2002). As vias de biossíntese de auxinas são constituídas por rotas metabólicas diversas e redundantes, sendo que as nitrilases são enzimas da via dependente de triptofano e catalisam a conversão do composto indol-3-acetonitrila (IAN) em IAA (Ljung *et al.*, 2002). As nitrilases detectadas como mais expressas em raízes são homólogas do gene *nit4* de *Arabidopsis*, que foi caracterizado como sendo expresso predominantemente em raízes (Bartel & Fink, 1994). Ainda de forma concordante com os resultados obtidos para cana-de-açúcar, um trabalho recente mostrou, através do monitoramento da síntese de IAA em *Arabidopsis thaliana*, que as raízes constituem um importante órgão para a síntese de IAA, sendo as porções meristemáticas as mais ativas nesse processo (Ljung *et al.*, 2005). Também foi mostrado, nesse trabalho, que as respostas fisiológicas das raízes aos níveis

de auxina dependem tanto do hormônio IAA sintetizado nas raízes quanto do proveniente das partes aéreas da planta.

Um gene codificador para a enzima ácido caféico 3-*o*-metiltransferase (COMT) mostrou-se expresso principalmente no quarto entrenó. Essa enzima é envolvida com a biossíntese de lignina e, em associação com outras enzimas, como a cafeoil CoA 3-*o*-metiltransferase (CCOMT), regula o conteúdo e a composição da lignina nas células. Uma correlação entre os níveis de lignina de entrenós de *Medicago sativa* de maturidade progressiva e a atividade das enzimas COMT e CCOMT foi demonstrado (Inoue *et al.*, 1998). Um gene codificador para uma COMT de cana-de-açúcar foi clonado e exibiu um pico de expressão nos colmos (Selman-Housein *et al.*, 1999). O banco SUCEST indica a presença de cinco sequências completas para essa enzima, as quais podem indicar alvos promissores para a engenharia genética de cana-de-açúcar com objetivo de se modificar o conteúdo e/ou a composição do bagaço da cana-de-açúcar, permitindo o seu uso como uma matéria-prima barata para a produção de papel.

Dentre os SAS presentes no banco de dados do Projeto SUCEST, aproximadamente 30% não possui similaridade a proteínas conhecidas e constituem, então, genes novos de função desconhecida (Vettore *et al.*, 2003). Sete genes dessa categoria mostraram-se expressos predominantemente nos entrenós, sete em folhas e um em raízes (**Tabela 8**). Adicionalmente, dois *no matches* aparentam ser mais expressos nas amostras de flores, podendo representar transcritos não-codificadores, uma vez que uma fase aberta de leitura para esses SAS não é detectada pelo programa ESTScan (Iseli *et al.*, 1999). O envolvimento de transcritos não-codificadores no desenvolvimento floral já foi descrito (Aukerman & Sakai, 2003, Schmid *et al.*, 2003, Chen, 2004), o que aponta para um provável papel desses genes da cana-de-açúcar nesse processo fisiológico. Adicionalmente, a matriz de razões de expressão virtuais mostra oito *no matches* mais expressos nas amostras de flores (**Figura 9**),

Além da identificação de genes diferencialmente expressos, os dados de microarranjos gerados também permitiram a identificação de genes de expressão ubíqua e homogênea dentre os seis órgãos em estudo. Genes de expressão ubíqua são altamente interessantes, uma vez que podem representar controles experimentais para ensaios de avaliação de expressão gênica, bem

como para a clonagem de promotores úteis para a biotecnologia de plantas. Dentre os genes analisados, 153 apresentaram pelo menos 55% dos seus pontos réplicas dentro dos limites dependentes de intensidade determinados em todas as amostras utilizadas (**Tabela 9**). Dentre esses, encontram-se 35 genes sem similaridades nos bancos de dados públicos, indicando que esses novos genes podem desempenhar funções centrais na fisiologia de cana-de-açúcar. Além disso, os três SAS que codificam para proteínas quinases contendo domínios *U-box*, descritos acima, demonstraram um perfil de expressão ubíquo e homogêneo entre todos os órgãos analisados.

Como já foi dito anteriormente, todas as hibridizações foram realizadas contra uma referência comum (*Ref Pool*), constituída por uma mistura de amostras de RNA dos seis órgãos em estudo. Apesar dessa abordagem permitir a identificação de genes diferencialmente expressos e ubíquos entre os órgãos de cana-de-açúcar, os dados de expressão gerados são apenas relativos, não absolutos, ou seja, não é possível fazer comparações entre os órgãos par a par. Para ter acesso a esse tipo de informação, as razões de expressão virtuais entre pares de órgãos, usando-se os valores de *Ref Pool* como denominador comum, foram calculadas. Os valores de razões virtuais calculados para todas as comparações par a par foram utilizados para gerar um agrupamento hierárquico contendo os perfis de expressão para cada um dos genes representados nos microarranjos SUCAST (**Figura 9**). Essa matriz de perfis de expressão permite uma visão comparativa da abundância dos transcritos, que pode complementar a informação obtida com a análise de genes diferencialmente expressos pelo método HT-self. Como exemplo da utilidade desse tipo de análise, destacamos os quatro fatores de transcrição MADS, selecionados como expressos em menor grau nas raízes ou outros órgãos em relação à referência, mas na matriz indicados como expressos preferencialmente nas flores (marcados com asterisco na **Figura 9**), apesar de apenas um desses SAS ter sido selecionado como tal na detecção de genes diferencialmente expressos.

Para as validações por PCR em tempo real, inicialmente foi necessário definir qual gene seria adotado como normalizador para a avaliação dos níveis de expressão dos genes de interesse nos seis órgãos em estudo. A necessidade da utilização de mais do que um único normalizador

para ensaios de PCR em tempo real foi descrita, indicando-se as deficiências que cada normalizador individualmente apresenta (Vandesompele *et al.*, 2002). Levando-se isso em consideração, empregaram-se dois normalizadores concomitantemente para a análise da expressão gênica em diferentes órgãos de cana-de-açúcar. Um gene para proteína 14-3-3 e um para poliubiquitina foram utilizados como referência, sendo que as normalizações dos dados para raiz usando o gene de 14-3-3 como normalizador e os dados de flor tendo o gene de poliubiquitina como normalizador foram desconsiderados, já que para esses órgãos os níveis de expressão se destacavam dos demais (**Figura 10**). Os gráficos da **Figura 10A** mostram o aumento da fluorescência (diretamente proporcional à concentração de dsDNA) no decorrer dos ciclos da PCR. Quanto mais expresso for o gene de interesse em determinada amostra, mais as suas curvas são deslocadas para a esquerda em relação às demais curvas. Dessa forma, pode-se concluir que o gene de poliubiquitina é mais expresso nas flores em relação aos demais órgãos, e o gene de 14-3-3, menos expresso nas raízes.

Para a validação dos dados de microarranjos, foram selecionados 18 genes diferencialmente expressos e sete genes com perfil de expressão ubíquo. O desafio encontrado na análise dos dados de PCR em tempo real foi definir qual critério seria adotado para validar um gene de expressão ubíqua e como diferenciá-lo de um gene considerado órgão-enriquecido. Para isso, foi calculada a entropia (H) da distribuição dos níveis de expressão entre os órgãos para cada um dos genes. A entropia é largamente empregada para a avaliação de quão distante a distribuição observada está de uma distribuição uniforme (MacKay, 2003). Idealmente, os genes ubíquos devem apresentar uma distribuição uniforme dos níveis de expressão em todos os órgãos considerados. De acordo com a definição de entropia, essa propriedade traduz-se como H próximo de um. De forma oposta, genes órgão-enriquecidos devem apresentar uma expressão pronunciada em apenas um dos órgãos considerados e a entropia da distribuição dos níveis de expressão traduz-se como H próximo de zero (deve-se ressaltar que essa não é uma escala linear). A análise dos dados de PCR em tempo real mostrou que 6 de 7 genes classificados como tendo um perfil de expressão ubíquo de acordo com os dados de microarranjos apresentaram valor H igual ou maior do que 0,9 (**Tabela 10**), indicando que esses genes podem representar verdadeiros genes

housekeeping. Dentre os genes diferencialmente expressos, quatro mostraram-se altamente específicos de um órgão, com valores H abaixo de 0,6. Oito genes diferencialmente expressos mostraram-se enriquecidos em determinado órgão, como já apontado pelos dados de microarranjos, mas também foram expressos em altos níveis em outros órgãos. Nesses casos, os valores H foram maiores do que os obtidos para genes expressos em um único órgão, mas sempre abaixo de 0,9.

Em resumo, dos 25 genes selecionados, 18 (72%) mostraram um perfil de expressão nos ensaios de PCR em tempo real consistente com o obtido nos experimentos de microarranjos. Deve-se ressaltar que as amostras de RNA utilizadas para os ensaios de PCR em tempo real são provenientes de uma terceira amostra biológica, diferente, portanto, das amostras utilizadas para os microarranjos de cDNA, o que sugere que os dados de expressão gerados pelos microarranjos são robustos para indicar genes ubíquos e diferencialmente expressos. O critério utilizado para selecionar os genes como diferencialmente expressos nos experimentos de microarranjos é arbitrário (mínimo de 70% dos pontos-réplicas acima ou abaixo dos intervalos de razão determinados para genes considerados diferencialmente expressos e 55% dos pontos-réplicas dentro dos intervalos de razão para genes ubíquos), mas provou-se eficiente, tendo em vista a alta porcentagem de validação obtida.

7.4. Análise da expressão gênica em resposta a estímulos ambientais e a fitormônios

7.4.1. Visão geral dos genes responsivos a estímulos ambientais e a fitormônios em cana-de-açúcar

Hibridizações contra microarranjos de cDNA contendo 1.545 elementos válidos foram realizadas para a identificação de alterações no transcriptoma da cana-de-açúcar em resposta a fatores ambientais e aos fitormônios MeJA e ABA. Os estímulos investigados incluíram a seca, deficiência em fosfato, herbivoria e a interação com bactérias diazotróficas endofíticas.

A busca por valores extremos nos dados de expressão revelou 179 genes diferencialmente expressos (**Tabela 11**). Seca foi a condição responsável pelo maior número de alterações nos níveis de expressão, com 52% dos 179 genes que se mostraram alterados.

Os experimentos de estresse hídrico foram conduzidos usando a variedade sensível à seca SP90-1638. A análise dos dados indicou que 22 dos genes responsivos a esse estresse (23%) mostraram o mesmo perfil de expressão para os pontos 72 h e 120 h, revelando dois grupos de genes regulados: aqueles que respondem precocemente ao estímulo de seca (24 h) e os responsivos mais tardiamente (72 h e 120 h).

Devido ao seu alto impacto na agricultura, muitos estudos se dedicaram a identificar genes regulados por estresse hídrico, em particular estudos de regulação focados em fatores de transcrição (Yamaguchi-Shinozaki & Shinozaki, 2006). Sequências específicas de reconhecimento para determinados fatores de transcrição já foram detectadas nas sequências promotoras de muitos genes responsivos à seca, como é o caso das sequências de reconhecimento por fatores MYC, MYB e W-box, no caso dos fatores de transcrição do tipo WRKY (Abe *et al.*, 1997, Narusaka *et al.*, 2004, Riera *et al.*, 2005). Foram obtidos onze fatores de transcrição regulados por seca. Dentre esses, estão incluídos um fator do tipo MYB (SCRURT2007E11.g) e dois da família WRKY (SCCCLR1C04F11.g e SCEQRT1027D04.g), identificados como induzidos tardiamente pelo estresse hídrico. A análise de agrupamento por SOM com relação aos resultados de seca mostra também que esses SAS foram incluídos no grupo D4, com um perfil de indução ao longo do tempo após a supressão de rega.

Também foi encontrado um fator de transcrição do tipo MYB (SCCCLR1C05B03.g) de expressão diminuída após 6 h de deficiência em fosfato. Fatores de transcrição da família MYB já foram relatados como sendo relacionados com a resposta à deficiência em fosfato e seqüências de ligação a fatores MYB já foram encontradas nos promotores de diversos genes responsivos à deficiência nutricional (Rubio *et al.*, 2001, Hammond *et al.*, 2004, Todd *et al.*, 2004).

Diversos genes responsivos a frio em outros organismos foram regulados pela seca nos dados aqui apresentados. Quatro SAS homólogos a genes de proteínas induzidas pelo frio (LTI)

foram identificados como responsivos ao estímulo de seca, sendo um deles também reprimido após 6 h de exposição ao ABA. Dois desses SAS (SCUTST3084F06.g e SCACCL6008H06.g) pertencem ao mesmo grupo do SOM (D1), com um perfil de indução pela seca. As vias de seca e de frio apresentam grande sobreposição quanto a genes responsivos e muitas das respostas em comum a essas vias são desencadeadas pelo hormônio ABA, como é o caso das proteínas do tipo LEA, induzidas por frio, ABA e desidratação (Chinnusamy *et al.*, 2004). Genes regulados de forma comum em resposta à seca e ao frio foram identificados em *Arabidopsis thaliana* utilizando-se microarranjos de cDNA e foi visto que a maioria dos genes induzidos por frio e por seca possui em comum o motivo DRE ou motivos relacionados a DRE (elemento responsivo à desidratação, de *Dehydration-Responsive Element*) em seus promotores (Seki *et al.*, 2001). É importante notar que um fator de transcrição (SCBGLR1002A09.g) da família AP2 e homólogo a DREB2 (*DRE binding factor 2*) de arroz mostrou-se induzido após 72 h da supressão de rega e pode, portanto, representar um fator de transcrição importante para a regulação de genes responsivos à seca em cana-de-açúcar. Já se mostrou que a superexpressão do fator de transcrição DREB2A de *Arabidopsis thaliana* levou à geração de plantas transgênicas mais tolerantes à seca (Sakuma *et al.*, 2006). Portanto, a manipulação dos níveis do fator de transcrição DREB2 em cana-de-açúcar pode ser um caminho para o desenvolvimento de plantas mais resistentes à seca.

Como as vias de sinalização de seca são parcialmente dependentes de ABA, genes com perfil de expressão comum a esses dois estímulos foram procurados. Quatro genes foram identificados como sendo induzidos por esses dois estímulos: duas delta-12 oleato dessaturases (SCCCLR1C03G01.g e SCVPST1061G05.g), uma S-adenosilmetionina descarboxilase (SCCCLR1C05G07.g) e uma proteína fosfatase da família PP2C (SCEPRZ1010E06.g) homóloga às proteínas ABI1 e ABI2 de *Arabidopsis thaliana*. Fosfatases PP2C fazem parte da via de sinalização de ABA, como é o caso das fosfatases ABI1, ABI2, AtPP2CA e AtP2CHA, que têm seus transcritos induzidos pela ação desse fitormônio (Merlot *et al.*, 2001, Tahtiharju & Palva, 2001). ABI1 e ABI2 regulam diversas funções, como, por exemplo, o fechamento dos estômatos, o que ilustra a estreita relação entre as vias de sinalização de ABA e respostas à seca (Merlot *et al.*, 2001), corroborando os dados obtidos para cana-de-açúcar.

Uma outra fosfatase da família PP2C (SCJLRZ3077G10.g) mostrou-se induzida em plantas inoculadas com *Gluconacetobacter*. O envolvimento de fosfatases do tipo PP2C em respostas de nodulação foi mostrado em um trabalho com *Lotus japonica* (Kapranov *et al.*, 1999). Além disso, cinco fatores de transcrição mostraram-se regulados pela interação das plantas com as bactérias diazotróficas. Dentre esses, encontram-se dois fatores de transcrição com dedos de zinco (SCEQRT1033F01.g e SCEZST3147A10.g). A análise das bibliotecas do Projeto SUCEST mostrou que fatores de transcrição com dedos de zinco e proteínas fosfatases situam-se dentre as categorias enriquecidas nas bibliotecas AD e HR, referentes a plantas infectadas com *Gluconacetobacter diazotrophicus* e *Herbaspirillum rubrisubalbicans*, respectivamente (Vargas *et al.*, 2003), o que está de acordo com os dados obtidos por microarranjos.

Dos genes regulados pelo tratamento de plantas com ABA, 58% se mostraram regulados após 12 h de tratamento com esse hormônio. A comparação dos resultados de expressão aqui obtidos para o tratamento com ABA com os descritos em arroz mostra resultados coincidentes para um gene de frutose 1,6-bisfosfato aldolase, uma glicoxalase e para Rubisco, sendo todos esses genes induzidos por esse fitormônio (Lin *et al.*, 2003). O SAS de frutose 1,6-bisfosfato aldolase detectado como diferencialmente expresso nas nossas análises (SCRULR1020D11.g) é homólogo ao gene *NpAldP1*, que foi detectado a partir de folhas de *Nicotiana paniculata* e se mostrou reprimido frente a estresse salino nesse organismo (Yamada *et al.*, 2000). Além da indução pelo tratamento com ABA, o SAS SCRULR1020D11.g foi reprimido no tratamento com MeJA, pelo estímulo de seca e também está incluído nos grupos da análise por SOM A3 e D6, ressaltando a sua indução após tratamento com ABA e repressão após a supressão de rega. Adicionalmente, esse SAS apresentou, nos experimentos de microarranjos para amostras de órgãos, seus transcritos enriquecidos nas folhas. A análise da expressão diferencial de genes relacionados a enzimas da via glicolítica e fermentação frente a diferentes estresses em caule e raiz de arroz indicou a indução de um gene de aldolase em resposta ao estresse salino (Minhas & Grover, 1999). Em outro trabalho, um gene de frutose-bisfosfato aldolase se mostrou reprimido no tratamento com ABA em *Arabidopsis thaliana* (Seki *et al.*, 2002). Com isso, a regulação da aldolase de cana-de-açúcar em questão vem se somar a dados interessantes que mostram que as enzimas

da via glicolítica e podem estar envolvidas em vias relacionadas à adaptação às condições do ambiente, e que resultados diferentes podem ser obtidos dependendo do órgão, do organismo e do estímulo analisados.

Ainda com relação aos genes induzidos após 12 h de tratamento com ABA, observou-se a presença de um gene (SCCCLR1C07B07.g) para uma proteína rica em glicina que se liga em RNA. A função desses tipos de proteínas ainda não é bem definida (Sachetto-Martins *et al.*, 2000, Mousavi & Hotta, 2005), mas sabe-se que algumas delas estão envolvidas com respostas a estresses e que desempenham um papel importante no processo de regulação pós-transcricional, controlando o *turnover* das moléculas de RNA (van Nocker & Vierstra, 1993). Uma proteína dessa categoria foi relatada como induzida por ABA, luz e salinidade em *Sorghum bicolor* (Annetta *et al.*, 2002). A proteína MA16 de milho também é um exemplo de proteína que se liga com RNA e que é induzida por ABA (Freire & Pages, 1995, Himmelbach *et al.*, 2003). Um SAS que codifica para uma proteína rica em glicina e com motivo de ligação ao RNA também foi encontrado regulado pela deficiência em fosfato (SCRUFL1112F04.b). Como essas proteínas podem estar relacionadas com a estabilidade do RNA, seu papel na deficiência em fosfato pode estar relacionado à mobilização desse nutriente a partir de estoques de ácidos nucleicos (Hammond *et al.*, 2004).

Diversos reguladores das vias de sinalização por ABA já foram descritos e caracterizados. Dentre esses, uma pequena GTPase (ROP10) foi caracterizada como regulador negativo da via de transdução de sinal de ABA (Zheng *et al.*, 2002). No presente trabalho, um gene codificador de GTPase do tipo Rab (semelhante a Ras) (SCACCL6006D08.g) foi induzido após 1 h de tratamento com esse hormônio. Adicionalmente, um outro SAS codificando para GTPase da família Rab encontra-se no grupo A6, de perfil de indução por ABA (SCJFRT1059D05.g). As GTPases monoméricas constituem moléculas que alternam entre a forma GTP-ligada (estado ativo) e a forma GDP-ligada (estado inativo) e desempenham diversos papéis regulatórios relacionados à proliferação celular, reorganização da actina, controle do tráfego de vesículas, participação em vias de MAPKs e transporte através da membrana nuclear (Vernoud *et al.*, 2003). As GTPases do tipo Rab, especificamente, participam no trânsito intracelular de vesículas e podem também participar de eventos de fusão de membranas (Vernoud *et al.*, 2003). Os dois SAS em questão codificam

para GTPases semelhantes às do tipo Rab11. Os genes codificadores para pequenas GTPases aqui identificados como regulados por ABA somam-se a alguns estudos que já indicaram a regulação de GTPases do tipo Rab por esse fitormônio (O'Mahony & Oliver, 1999, Nahm *et al.*, 2003,).

Dentre os genes induzidos pelo tratamento com MeJA, há três genes que codificam para histonas no grupo B1 de agrupamento por SOM, sendo dois do tipo H4 e um do tipo H2B. Somente um desses SAS (SCJLHR1028C12.g) encontra-se também dentre os detectados como diferencialmente expressos por esse tratamento. Essa observação aponta para o valor da análise de SOM como complementar à análise de detecção de *outliers*. No exemplo aqui dado, o agrupamento por SOM foi capaz de mostrar mais dois SAS referentes a histonas que apresentam um perfil de expressão induzido em decorrência do tratamento com MeJA, com pico em 12h. Provavelmente, os valores de M para esses SAS não foram suficientes para extrapolar os limites de razão calculados para considerar um gene como diferencialmente expresso, mas a análise de SOM revela que esses SAS podem estar sendo responsivos ao tratamento com MeJA. A regulação da expressão gênica de histonas aponta para um provável mecanismo de remodelação da cromatina disparada pela ação dos jasmonatos, podendo ser esse um importante mecanismo para o controle da expressão gênica mediada por MeJA. Já é sabido que os jasmonatos podem regular a expressão gênica interferindo nos mecanismos de acetilação/desacetilação de histonas, já que COI1, uma proteína do tipo *F-box* necessária para as respostas aos jasmonatos, é capaz de marcar uma histona desacetilase de *Arabidopsis* para degradação via proteólise (Devoto *et al.*, 2002). Adicionalmente, Kim e colaboradores (1998) observaram a indução das histonas CaH2B e CaH4 de *Capsicum annuum* por metiljasmonato.

Dentre os estímulos analisados pelo presente projeto, a inoculação de cana-de-açúcar com bactérias diazotróficas endofíticas é interessante por representar um estímulo biótico que interfere com o desenvolvimento da planta positivamente. *Herbaspirillum* spp. e *Gluconacetobacter diazotrophicus* são bactérias fixadoras de nitrogênio que colonizam os espaços intercelulares e os órgãos vasculares de cana-de-açúcar sem causar sintomas de doença. Esses microorganismos são os responsáveis pelos elevados níveis de BNF que alguns cultivares de cana-de-açúcar

apresentam e pela reduzida necessidade de uso de fertilizantes para esses cultivares. Apesar da interação planta-endofíticos ser uma associação benéfica para ambos os organismos, acredita-se que a cana-de-açúcar seja capaz de reconhecer essas bactérias e ativar mecanismos de defesa até o estabelecimento de uma associação eficiente (Vinagre *et al.*, 2006). Dentre os genes responsivos à interação com endofitas e que possam estar relacionados a mecanismos de percepção desses microorganismos pela planta, encontram-se quatro genes-R, sendo que dois desses foram induzidos tanto na interação com bactérias do gênero *Herbaspirillum* quanto com *Gluconacetobacter diazotrophicus*. A interação com *Gluconacetobacter* também levou à indução de um gene relacionado à biossíntese de ácido salicílico. O ácido salicílico acumula-se nos órgãos das plantas em resposta à infecção por patógenos e é essencial para a indução da resistência sistêmica adquirida e também para algumas respostas mediadas por genes de resistência (Gaffney *et al.*, 1993, Delaney *et al.*, 1994; Mur *et al.*, 1997). Dos 34 genes responsivos à interação planta-endofita, 23 foram detectados somente na inoculação com *Herbaspirillum* spp. e sete somente na inoculação com *Gluconacetobacter diazotrophicus*. Apenas quatro genes foram regulados por ambas as interações e da mesma forma, mostrando que o reconhecimento e a resposta a essas bactérias possuem mecanismos em comum, mas também variáveis, dependendo do microorganismo em questão.

Os dados de expressão para herbivoria mostram a indução de um gene de proteína semelhante à taumatina no ponto 24 h. Essas proteínas apresentam similaridade com proteínas relacionadas à patogênese do grupo 5 (PR-5). Algumas proteínas dessa categoria já foram estudadas e podem apresentar atividades antifúngica, de endo- β 1,3-glucanase ou de inibição de tripsina e de α -amilase, importantes para as defesas das plantas (Grenier *et al.*, 1999, Franco *et al.*, 2002).

A expressão diferencial de genes sem homologia nos bancos de dados (*no matches*) bem como de SAS que codificam para proteínas hipotéticas frente às condições aqui estudadas é de extrema valia para a caracterização funcional desses genes. A função para esses genes, bem como o seu potencial biotecnológico, poderão ser avaliados com o uso de técnicas de superexpressão e de silenciamento gênico e análise dos fenótipos resultantes. Foram encontrados

28 SAS sem homólogos nos bancos de dados públicos dentre os genes diferencialmente expressos em pelo menos um dos estímulos analisados. Desses, 13 são regulados por seca, 7 por inoculação com *Herbaspirillum*, 5 por tratamento com ABA, 4 por tratamento com MeJA, 3 por herbivoria e 3 por deficiência em fosfato. Interessantemente, quatro desses genes sem similares nos bancos de dados públicos são mais expressos em folhas, de acordo com os experimentos realizados de comparação da expressão gênica em diferentes órgãos da cana-de-açúcar. Um desses genes folha-enriquecidos (SCCCCL1001H07.b) mostrou-se induzido tanto pela inoculação com bactérias diazotróficas endofíticas quanto pelo ataque por *Diatraea saccharalis*. Esse dado pode apontar para um provável papel desse gene em vias gerais de defesa contra estímulos bióticos e a ativação de respostas de defesa, até o estabelecimento de uma associação eficiente, no caso de interação com bactérias fixadoras de nitrogênio.

É também interessante notar que cinco dos genes *no matches* encontrados como diferencialmente expressos não apresentam uma fase aberta de leitura pela predição do ESTScan, podendo representar transcritos não-codificadores. Desses, três foram regulados por seca, um por MeJA e um por inoculação com *Herbaspirillum* spp. Estudos recentes têm atribuído importantes funções a alguns microRNAs de plantas na regulação de processos como desenvolvimento, respostas a patógenos e sinalização hormonal (Zhang *et al.*, 2006). Contudo, é necessário que esses genes sejam funcionalmente analisados para a confirmação de que eles realmente representam transcritos não-codificadores, já que predições como as feitas pelo ESTScan sempre são sujeitas a erros.

As vias ativadas pelos fitormônios apresentam um elevado grau de sinalização cruzada, sendo essas interações altamente complexas e tempo e espaço-dependentes (Moller & Chua, 1999, Gazzarrini & McCourt, 2003). Genes relacionados a vias de sinalização e de biossíntese de hormônios foram regulados pelos tratamentos com ABA, MeJA e por condições de estresses. Apesar de não ser possível, mediante a análise dos dados de expressão aqui relatados, inferir conclusões definitivas a respeito das interações entre as vias de sinalização por fitormônios em cana-de-açúcar, as análises de genes diferencialmente expressos selecionados, bem como os grupos obtidos pelas análises de SOM, podem revelar alguns aspectos dessas interações.

Dos 29 genes identificados como diferencialmente expressos em resposta a mais de um estímulo, 18 foram responsivos à seca e aos fitormônios. Durante estresses abióticos, as plantas aumentam a produção de hormônios, como ABA, etileno e, provavelmente, ácido salicílico e ácido jasmônico (Xiong *et al.*, 2002). Sendo assim, muitos dos genes responsivos ao estresse hídrico são regulados por fitormônios induzidos pelo estímulo de seca. Interações entre as respostas à salinidade, que são parcialmente similares às respostas disparadas pela seca, e vias de ABA e jasmonatos já foram descritas em cevada (Walia *et al.*, 2006), mostrando-se a indução de genes responsivos a ABA, responsivos a jasmonatos e de biossíntese de jasmonatos quando essa planta foi submetida à condição de elevada salinidade. Os nossos dados mostram a indução de duas delta-12 oleato dessaturases após 24 h de supressão de rega. Também foram detectadas duas ômega-3 dessaturases menos expressas após 72 h e 120 h de secas (pertencentes aos grupos D5 e D6). Esses dados sugerem uma modulação na síntese de jasmonatos disparada como resposta à seca em cana-de-açúcar e essa indução parece ocorrer no início da resposta à supressão da rega, até 24 h após o início do estímulo.

Muitos dos resultados obtidos apontam para um envolvimento de resposta a auxinas nos tratamentos em estudos. A análise dos grupos gerados por SOM para tratamento com ABA evidencia algumas respostas antagônicas existentes entre a via de ABA e a de auxina: enquanto no grupo A4 encontra-se um gene para proteína responsiva à auxina (com perfil reprimido ao longo do tempo), um gene para proteína reprimida por auxina encontra-se induzido após 1 h de tratamento com ABA (grupo A6). As características das interações entre ABA e auxinas são dependentes do órgão em análise. No caso de folhas de *Arabidopsis thaliana*, ABA e auxinas interagem de forma antagônica para regular a abertura dos estômatos (Eckert & Kaldenhoff, 2000).

O estresse hídrico levou à indução de genes codificando para enzimas de biossíntese de auxinas (nitrilases) após 72 h de supressão de rega e a repressão de fatores de transcrição da categoria Aux/IAA após 120 h de seca. Esses fatores de transcrição agem inibindo a sinalização de auxinas e são rapidamente induzidos após exposição às auxinas (Abel & Theologis, 1996, Woodward & Bartel, 2005). O acúmulo de fatores Aux/IAA, porém, é sujeito a um processo de retroalimentação negativa, uma vez que as auxinas agem marcando esses fatores para a

degradação pelo proteossomo 26S (Woodward & Bartel, 2005). Os grupos D1 e D2 contêm dois genes codificando para nitrilases e um gene codificando para uma proteína ligadora de auxina (ABP, *auxin binding protein*). Além disso, um fator responsivo à auxina está presente no grupo D4, também com perfil induzido pela seca. Fatores de transcrição AUX/IAA podem ser encontrados entre os grupos D2, D3 e D6. Foi mostrado que o tratamento com IAA, o membro mais importante da família das auxinas, em plântulas de trigo submetidas à seca, levou à diminuição do acúmulo de prolina induzido por esse estresse, indicando o envolvimento de auxinas nas respostas à seca (Sadiqov *et al.*, 2002).

A deficiência em fósforo leva a alterações da arquitetura da raiz, de forma a aumentar a exploração do solo e a aquisição de fósforo. Nesse processo de adaptação fisiológica, muitos fitormônios desempenham importantes papéis, como é o caso da auxina, que está envolvida em processos como alongamento das raízes e desenvolvimento de raízes laterais (Hammond *et al.*, 2004). Em *Arabidopsis*, a deficiência em fósforo aumenta a sensibilidade à auxina, o que poderia explicar o aumento no número de raízes laterais observadas quando a planta é submetida a esse estresse nutricional (López-Bucio *et al.*, 2002). De modo coerente com essa observação, foi encontrado um SAS (SCCCLR2002F08.g) homólogo a genes de outras plantas que são reprimidos por auxina sendo reprimido após 6 h de deficiência em fósforo. Portanto, esse gene pode estar sendo responsivo a modificações na sensibilidade às auxinas que ocorrem na cana-de-açúcar quando essa se encontra em carência de fósforo.

7.4.2. Componentes do quinoma de cana-de-açúcar responsivos a fitormônios e a estímulos ambientais

A fosforilação por proteínas quinases constitui um dos principais mecanismos pelos quais diversos processos, incluindo respostas a estímulos ambientais, controle do ciclo celular e defesa contra patógenos, são controlados. Além disso, um elevado número de proteínas quinases de plantas são receptores, desempenhando importantes papéis na detecção de hormônios, fatores de

virulência de patógenos e sinais relacionados a crescimento e desenvolvimento (Shiu & Bleecker, 2001a).

Sete proteínas quinases mostraram-se diferencialmente expressas em decorrência da interação da planta com bactérias fixadoras de nitrogênio. Dessas, três são envolvidas com vias de cálcio, sendo uma CDPK (caneCDPK-9) e duas CIPKs (caneCIPK-18 e caneCIPK-22). Como já dito anteriormente, proteínas do tipo CIPK interagem com proteínas semelhantes à calcineurina, proteínas essas que agem como sensores de Ca^{2+} . Essas quinases fazem parte do subgrupo SnRK3 das quinases de plantas semelhantes à SNF1 de levedura (Shi *et al.*, 1999, Hrabak *et al.*, 2003). Alguns trabalhos já mostraram o envolvimento de vias dependentes de cálcio nos processos de simbiose e de nodulação (Weaver & Roberts, 1992, Webb *et al.*, 2000; White, 2001, Levy *et al.*, 2004). Adicionalmente, é sabido que algumas CDPKs participam em vias de sinalização de defesa a patógenos, como é o caso das respostas de defesa do tomate contra o fungo *Cladosporium fulvum* (Romeis *et al.*, 2000).

Outra proteína quinase regulada transcricionalmente pela interação planta-endofita é uma quinase da categoria similar a GSK3/shaggy (caneGSK3-5). Essas quinases, em plantas, são divididas em quatro grupos e estão associadas com diferentes funções: desenvolvimento floral, sinalização hormonal da via de brassinosteróides e resposta a estresses como ferimento e salinidade (Jonak & Hirt, 2002). Foi também induzida, pela interação com *Herbaspirillum* spp., a expressão de um SAS (canePBS1-4) similar ao gene *PBS1* de *Arabidopsis*. A proteína PBS1 tem a função de reconhecer fatores de avirulência de *Pseudomonas syringae* (Swiderski & Innes, 2001), o que indica uma função potencial em cana-de-açúcar dessa proteína na participação em vias de reconhecimento de *Herbaspirillum* spp.

Alguns receptores envolvidos com a regulação da simbiose entre bactérias e plantas já foram caracterizados (Endre *et al.*, 2002, Madsen *et al.*, 2003, Capoen *et al.*, 2005). Recentemente, um RLK de cana-de-açúcar (SHR5) foi descrito como sendo reprimido em plantas associadas com bactérias diazotróficas endofíticas (Vinagre *et al.*, 2006). Os autores também demonstraram que os níveis dessa proteína são inversamente relacionados à eficiência da associação planta-bactérias.

Dentre os dados de expressão obtidos para a interação de cana-de-açúcar com microorganismos diazotróficos endofíticos, deve-se ressaltar um RLK indefinido (caneURLK-13) que se mostrou induzido em decorrência da interação com *Herbaspirillum* e, portanto, pode representar um receptor importante para mediar tal interação.

Dez proteínas quinases mostraram-se diferencialmente expressas em resposta ao estresse hídrico. Sete dessas quinases pertencem à categoria de SnRKs e 4 dessas SnRKs são induzidas pela seca (*cane osmotic stress-activated protein kinase-2*, caneCIPK-8, caneCIPK-13 e caneCIPK-14). SRK2C, uma SnRK ativada pelo estresse osmótico, leva, quando superexpressa, ao melhor desempenho de *Arabidopsis thaliana* frente a condições de seca (Umezawa *et al.*, 2004). Adicionalmente, a SnRK OST1, de *Arabidopsis thaliana*, já foi associada à transdução de sinais relacionadas à seca (Mustilli *et al.*, 2002). Estudos de mutagênese nesse gene mostram que seus efeitos estão confinados às células-guarda, levando à insensibilidade ao ABA, e que a sua atividade catalítica é acentuada pela presença desse fitormônio no meio ou por condições de baixa umidade (Mustilli *et al.*, 2002, Riera *et al.*, 2005). Outros trabalhos descrevem funções dessa família de quinases em respostas à seca, incluindo um papel no fechamento dos estômatos (Li & Assmann, 1996, Li *et al.*, 2000, Mustilli *et al.* 2002, Yoshida *et al.*, 2002). Dentre as vias de sinalização ao estímulo de seca, a via de SOS dependente de cálcio (a qual envolve SOS2, uma SnRK semelhante às CIPKs) desempenha um importante papel na regulação da homeostasia de íons (Xiong *et al.*, 2002), sendo os mutantes *sos2* hipersensíveis ao estresse salino (Liu *et al.*, 2000).

Ainda com relação às quinases responsivas aos estímulos em estudo, convém destacar um SAS relacionado à proteína de arroz G11A (com uma inserção em seu domínio catalítico, como descrito anteriormente) no grupo A2 (*caneG11A kinase-2*) e a presença de duas outras quinases semelhantes à G11A no grupo D2 (*caneG11A kinase-3* e *caneG11A kinase-5*), com um perfil aparente de indução por ABA e por seca, respectivamente. Dentre os genes diferencialmente expressos frente a fitormônios e a estímulos ambientais, encontram-se seis genes de quinases indefinidas classificadas pela abordagem filogenética (*canePK-BIII3*, *caneRLCK-AVI2*, *caneRLCK-DII3*, *caneRLK-AX1*, *caneRLK-AX2* e *caneRLK-C5*). A descoberta de que esses genes são

regulados por seca, inoculação com *Herbaspirillum* spp. e/ou tratamentos com fitormônios adicionados dados funcionais inéditos na caracterização dessas proteínas.

7.5. Análise da expressão gênica em genótipos contrastantes para acúmulo de sacarose

A cana-de-açúcar é uma gramínea C4 capaz de acumular sacarose nos seus colmos em níveis que excedem 50% do seu peso seco. O teor de sacarose, avaliado através de medições de brix, é uma característica fenotípica usada para seleções em programas de melhoramento genético. Apesar das vias metabólicas de síntese de sacarose já estarem bem definidas, o controle do seu acúmulo em cana-de-açúcar ainda não é compreendido. Os cultivares comerciais modernos derivam principalmente de cruzamentos e retrocruzamentos de duas espécies, *Saccharum officinarum* e *Saccharum spontaneum*. Como parte do presente trabalho, a progênie F3 de cruzamentos desse tipo foi selecionada para identificar genótipos altamente contrastantes para níveis de brix. Amostra de folhas +1 e de entrenós 1, 5 e 9 de genótipos de AB (alto brix) e BB (baixo brix) foram comparadas entre si através de hibridizações contra microarranjos de cDNA. A folha +1 é a folha mais jovem com inserção da bainha com o limbo foliar visível e caracteriza-se por ser totalmente expandida e com maior fotossíntese líquida, sendo um órgão fonte de sacarose. Já os entrenós correspondem a diferentes porções do colmo de cana-de-açúcar, um órgão-dreno, e diferem entre si pelo acúmulo de sacarose, sendo o entrenó 1 o mais apical, o mais imaturo e o de menor acúmulo de sacarose. Os dados de expressão obtidos foram validados por PCR em tempo real para alguns dos indivíduos extremos.

Pelo método de HT-self (Vencio & Koide, 2005), 51 genes foram identificados como diferencialmente expressos entre populações de AB e BB, em pelo menos um dos órgãos analisados (**Tabela 20**). Cinco desses genes apareceram como diferencialmente expressos em duas das amostras analisadas e um deles, em três amostras. Os genes diferencialmente expressos pertencem a diversas categorias funcionais, incluindo sinalização de cálcio, respostas a estresses, transcrição e ubiquitinação.

Cinco dos genes enriquecidos nas amostras de folhas de indivíduos de BB (SCCCLR1001E04.g, SCBFST3136A06.g, SCACLR1036B06.g, SCEQRT2099H01.g e SCEQRT1026H08.g) são expressos preferencialmente nesse órgão, de acordo com dados de microarranjos obtidos anteriormente. Entre esses genes, há dois que codificam para proteínas quinases envolvidas com vias de sinalização de cálcio. Uma dessas proteínas é similar a membros da família de CDPKs (caneCDPK-27) e a outra, à família de CIPKs (caneCIPK-18). Além disso, há uma terceira proteína quinase, também similar às CIPKs (caneCIPK-16), mais expressa nos entrenós maduros (9º entrenó) de plantas de AB. A sinalização por cálcio é efetuada através de alterações nas concentrações intracelulares de cálcio e via proteínas sensoras de cálcio, como calmodulina e calcineurina (Sanders *et al.*, 2002). Essas proteínas sensoras permitem a ativação de uma cascata de sinalização, levando à fosforilação de proteínas e a alterações na expressão gênica. O gene *CIPK14* de *Arabidopsis thaliana* mostrou-se induzido por sacarose e elementos responsivos à sacarose foram identificados no seu promotor (Lee *et al.*, 2005). Um estudo envolvendo experimentos de microarranjos com material de indivíduos de cana-de-açúcar de uma população F1 segregante para teor de sacarose, proveniente de um cruzamento entre as variedades comerciais SP80-180 e SP80-4966 (CTC, Brasil), mostrou que a *caneCIPK-8* é mais expressa nas amostras de folha+1 da população de BB em relação à população de AB (Felix, 2006). Apesar de essa não ser exatamente a mesma CIPK encontrada no presente trabalho, esses dados em conjunto apontam para uma relação entre acúmulo de sacarose e regulação das CIPKs.

O controle da síntese de sacarose depende da atividade da enzima sacarose fosfato sintase, que catalisa a síntese de sacarose 6-fosfato a partir de UDP-glicose e frutose 6-fosfato. Já a clivagem da sacarose depende da atividade da invertase, que cliva a sacarose em glicose e frutose, e da atividade da enzima sacarose sintase, que converte a sacarose em frutose e UDP-glicose na presença de UDP (Moore, 1995). Diversos estudos mostraram que algumas CDPKs fosforilam e regulam a enzima sacarose sintase (Huber *et al.*, 1996, Zhang *et al.*, 1999, Hardin *et al.*, 2003, Hardin *et al.*, 2004). Estudos com a enzima sacarose sintase de milho mostraram que a fosforilação dessa enzima no resíduo Ser-15 por CDPKs estimula a atividade de quebra de sacarose dessa enzima (Huber *et al.*, 1996, Hardin *et al.*, 2003). Além disso, CDPKs podem

fosforilar o resíduo Ser-170 dessa enzima e direcioná-la para ser degradada pela via dependente do proteossomo 26S (Hardin & Huber, 1999, Hardin *et al.*, 2003). A enzima sacarose sintase está relacionada a diversos processos fisiológicos, incluindo relações entre órgãos fonte e dreno (Zrenner *et al.*, 1995, Hanggi & Fleming, 2001) que podem contribuir para o acúmulo de sacarose em cana-de-açúcar. Adicionalmente, já foi visto que algumas CDPKs podem fosforilar e inativar a enzima sacarose-fosfato sintase (McMichael *et al.*, 1995, Pagnussat *et al.*, 2002). Como a biossíntese de sacarose parece ser, ao menos parcialmente, um processo regulado por cálcio, as CDPKs e CIPKs de cana-de-açúcar diferencialmente expressas detectadas podem representar pontos críticos no controle do acúmulo de sacarose nessa gramínea. Alternativamente, como já discutido acima, algumas CDPKs participam do controle da abertura dos estômatos em folhas (Li *et al.*, 1998, Assmann & Shimazaki, 1999). Sendo assim, será interessante averiguar se essa CDPK realmente possui essa atividade em cana-de-açúcar e, caso isso seja confirmado, se a regulação da expressão dessa quinase é um ponto-chave de controle de captação de CO₂ por essas plantas, o que também explicaria a sua regulação diferencial entre genótipos de AB e de BB.

Dois genes (SCCCRZ1002E08.g e SCCCRZ1001F02.g) codificando para aquaporinas foram identificados dentre os genes diferencialmente expressos. As aquaporinas constituem uma grande e diversa família de proteínas de membrana envolvidas com a regulação do movimento de água entre as células e compartimentos celulares, apesar de muitas dessas proteínas também facilitarem a passagem de pequenos solutos (Maurel & Chrispeels, 2001, Chaumont *et al.*, 2005). De acordo com a localização subcelular, as aquaporinas dividem-se entre PIPs (proteínas intrínsecas da membrana plasmática) ou TIPs (proteínas intrínsecas do tonoplasto). Cada uma das duas aquaporinas identificadas como diferencialmente expressas classifica-se em uma dessas categorias. Uma das aquaporinas identificadas como diferencialmente expressa é mais expressa nas folhas e entrenós de plantas de baixo brix e é similar às aquaporinas do tipo PIP, mais exatamente, às do tipo PIP1 de milho. Ao passo que as aquaporinas do tipo PIP2 são bons canais de água, as do tipo PIP1 não o parecem ser (Chaumont *et al.*, 2000). O acúmulo de sacarose em elevadas concentrações, como ocorre em cana-de-açúcar, representa um estresse osmótico para as células e leva à necessidade de um controle eficiente da compartimentalização de solutos.

Como as aquaporinas são importantes para regular a permeabilidade de membranas, essas proteínas devem desempenhar um importante papel no processo de armazenamento de açúcar nos vacúolos das células de colmo de cana-de-açúcar. Em *Arabidopsis thaliana*, observou-se que a perda da aquaporina TIP1;1 afeta severamente o metabolismo e o transporte de carboidratos (Ma *et al.*, 2004) e os autores postulam que essa aquaporina esteja envolvida no direcionamento de metabólitos via vesículas para o vacúolo central. Devido à diversidade de papéis descritos para membros dessa família, experimentos adicionais são necessários para que os papéis desempenhados pelas aquaporinas em cana-de-açúcar sejam realmente definidos.

Foram também encontrados seis genes relacionados a citocromos P450 monooxigenases (CYPs) diferencialmente expressos entre genótipos de alto e de baixo brix. Proteínas do tipo citocromo P450 monooxigenases são altamente empregadas em vias biossintéticas e de detoxificação de plantas, incluindo síntese de ligninas, protetores de raios UV, pigmentos, compostos de defesa, ácidos graxos, hormônios, moléculas de sinalização e também na degradação de compostos endógenos e tóxicos (Schuler & Werck-Reichhart, 2003). O SAS SCSGHR1069F04.b codifica para uma proteína citocromo P450 monooxigenase que participa da via de biossíntese dos ácidos hidroxâmicos cíclicos DIBOA (2,4-diidróxi-1,4-benzoxazin-3-ona) e DIMBOA (2,4-diidróxi-7-metóxi-1,4-benzoxazin-3-ona). Esses compostos são metabólitos secundários tóxicos que agem como pesticidas naturais contra insetos e contra fungos e bactérias patogênicos (Frey *et al.*, 1997). Além do papel de ácidos hidroxâmicos cíclicos contra estresses bióticos, o trabalho recente de Poschenrieder *et al.* (2005) mostrou que esses compostos parecem estar também envolvidos na detoxificação de alumínio em milho, indicando que esses compostos podem desempenhar outras funções além do papel de pesticidas naturais. É interessante notar que esse SAS mostrou-se induzido precocemente pela seca, além de ser mais expresso nas folhas de genótipos de AB em relação aos de BB.

Já o SAS SCAGLR1043E04.g codifica para uma AOS (também conhecida como CYP74A), uma enzima da biossíntese de jasmonatos. Esse gene mostrou-se mais expresso em folhas e entrenós 9 de plantas de BB, além de ter sido reprimido por ABA e pelo estímulo de seca. Similarmente a esse gene, outros três genes codificando para as citocromo P450 monooxigenases

CYP73 e CYP75 e para a enzima fenilalanina amônia-liase, mostraram-se mais expressos em folhas da população de BB. Esses três genes codificam para enzimas envolvidas na síntese de fenilpropanóides. Os fenilpropanóides são sintetizados a partir de fenilalanina e incluem compostos com função na defesa contra patógenos, herbívoros e contra ferimentos, além de componentes da parede celular, protetores contra a radiação UV, pigmentos e moléculas sinalizadoras, como é o caso do ácido salicílico (Weisshaar & Jenkins, 1998). Com isso, os dados aqui apresentados apontam para uma maior síntese de fenilpropanóides e oxilipinas nas folhas dos genótipos de baixo brix. Um dos trabalhos de Casu e colaboradores (2003) relatou a maior expressão de genes relacionados à biossíntese de fenilpropanóides em entrenós maduros em relação aos entrenós imaturos, o que é proposto como sendo relacionado a uma maior atividade dos processos de lignificação e suberização nos entrenós maduros. Entretanto, esse dado refere-se a colmos e, dentre os dados aqui relatados, a expressão diferencial de genes relacionados a vias de fenilpropanóides ocorreu nos tecidos de folha, o que pode explicar o fato desses resultados serem aparentemente discrepantes entre si.

Uma proteína relacionada ao processo de senescência também foi detectada como sendo mais expressa nos genótipos de AB ao se comparar amostras de entrenó 5. Já foi relatado que os entrenós maduros de cana-de-açúcar acumulam diferentes proteínas relacionadas à senescência e outros transcritos relacionados a vias de estresses (Casu *et al.*, 2004). Esse mesmo gene também foi induzido pelos estímulos bióticos analisados anteriormente (herbivoria e inoculação com *Gluconacetobacter diazotrophicus*) e também foi reprimido após 48 h de deficiência nutricional. Em Casu *et al.* (2004) também é observado um enriquecimento de transcritos relacionados a genes de defesa e de respostas a estresses em colmo maduro em relação ao colmo jovem. Durante a maturação dos entrenós de cana-de-açúcar, as células do parênquima se diferenciam em compartimentos altamente especializados no armazenamento de sacarose. Esse processo impõe uma reorganização celular de forma a permitir a adaptação ao estresse osmótico que irá ocorrer nessas células, e envolve também a lignificação e suberização progressiva das paredes celulares para prevenir a invasão por patógenos e a perda de água (Kolattukudy, 1984, Jacobsen *et al.*, 1992). Apesar dos estudos de Casu *et al.* (2004) terem sido realizados comparando-se colmos em

diferentes estágios de desenvolvimento, diferente da abordagem aqui relatada, em ambos os casos amostras de órgãos que diferem entre si pelo acúmulo de sacarose foram comparadas. Essas observações, em conjunto, apontam para a regulação de genes de cana-de-açúcar relacionados a estresses dependente dos níveis celulares de açúcares.

Como os açúcares são relacionados aos processos de fotossíntese, respiração mitocondrial e regulação do metabolismo de ácidos graxos, variações nos seus níveis resultam em alterações no balanço celular de ROS, levando a uma relação direta entre níveis de açúcares nas células e estresse oxidativo (Couée *et al.*, 2006). Com isso, os açúcares são importantes nas células vegetais não apenas como sensores da atividade fotossintética, mas também como sensores dos níveis de ROS intracelular, para que esses sejam controlados. Dentre os açúcares solúveis, a sacarose é postulada como sendo o açúcar mais importante para esses controles (Couée *et al.*, 2006). Também é sabido que diversos estresses, como patogenicidade, frio, seca e estresse salino, levam ao acúmulo de ROS e também estão associados a acúmulo de açúcares solúveis. Vias de indução de proteínas relacionadas a patogenicidade (proteínas PR) envolvem ROS e são aumentadas pelo tratamento com açúcares exógenos ou pelo aumento endógeno nos níveis de açúcares (Roitsch, 1999, Couée *et al.*, 2006), evidenciando uma relação entre genes envolvidos em patogenicidade e níveis intercelulares de açúcares. Baseado no fato de que alguns genes responsivos à patogenicidade são induzidos por açúcares, é proposto que a regulação de enzimas relacionadas a respostas de defesa à infecção por patógenos é mediada indiretamente pela indução inicial de um gene de invertase extracelular, o que resulta em aumento da concentração de açúcar (Jang & Sheen, 1994). Dentre os dados obtidos no presente trabalho, uma proteína PR do tipo taumatina está mais expressa em entrenós 9 dos genótipos de AB. Por outro lado, outros três genes para proteínas relacionadas a patogenicidade foram detectados mais expressos nas folhas dos genótipos de BB, o que pode indicar uma regulação diferenciada para essas proteínas entre órgãos fonte e dreno.

Diversos estudos relacionados ao papel de açúcares como reguladores de vias de transdução de sinais já foram conduzidos (Rolland *et al.*, 2002), mas é sempre importante ressaltar que, como já discutido em Couée *et al.* (2006), esses estudos se baseiam em experimentos em

que é feito um tratamento exógeno das plantas com açúcares ou derivados de açúcares. Com isso, o próprio açúcar adicionado ao meio pode ser um estresse e levar à indução de genes de defesa.

Foi também detectada a regulação de uma das subunidades da enzima ribulose-1,5-bisfosfato carboxilase (Rubisco) (SCCCLR1001E04.g), com um papel central na fixação de carbono. Esse gene mostrou-se mais expresso nas folhas da população de baixo brix. Já é sabido que essa enzima é reprimida quando os níveis de açúcares na célula são elevados, levando à diminuição nos níveis de fotossíntese (Koch, 1996). O fato desse gene ser menos expresso nos genótipos de alto brix em relação aos de baixo brix pode ser um indicativo de que as células das plantas de alto brix realmente possuem níveis elevados de açúcares em relação às plantas de baixo brix. A regulação gênica de Rubisco aponta para uma regulação da fotossíntese de forma a manter a homeostasia celular.

Dentre os genes encontrados como sendo diferencialmente expressos, é notável a detecção de um gene sem similaridade nos bancos de dados públicos (SCCCCL3001F04.g) e que apresentou, consistentemente, maior expressão na população de alto brix nas amostras de folhas, primeiro e quinto entrenós. É interessante notar que esse gene não possui ORF predita e, portanto, pode representar um transcrito não codificador, com algum papel regulatório que diferencia as duas populações em estudo entre si. Outros dois genes sem homologias nos bancos de dados mostraram-se altamente regulados pelas condições aqui analisadas. Um deles (SCAGLR2026G12.g) é induzido tardiamente pela seca, reprimido por ABA e induzido por MeJA. Além disso, esse mesmo gene, apesar de ter se mostrado mais expresso em raiz, é expresso diferencialmente ao se comparar folhas de genótipos de AB e de BB, sendo mais expresso em genótipos de BB. Já um outro gene sem homologias em bancos de dados (SCBFST3136A06.g) mostrou-se reprimido pelo estímulo de seca, induzido por ABA e reprimido por MeJA. Esse gene é expresso preferencialmente em folhas e é mais expresso em indivíduos de BB em relação aos de AB. A caracterização funcional desses possíveis “novos genes” de cana-de-açúcar poderá ser de grande valia para estudos futuros. Contudo, é importante lembrar que os experimentos de respostas a fatores bióticos e abióticos foram conduzidos com genótipos diferentes dos utilizados ao se comparar populações de AB x BB, o que limita a comparação direta desses resultados.

É também interessante notar a regulação de um provável RLK (cane RLK with L-lectin domain-33) com domínio extracelular de L-lectina. O gene para esse RLK se mostrou reprimido ao longo do tempo após a supressão de rega e é também mais expresso em indivíduos de BB em relação aos de AB, quando amostras de entrenós 9 foram analisadas. RLKs com domínio de L-lectina ainda são muito pouco caracterizados. Sabe-se que esses RLKs possuem um domínio extracelular com capacidade de se ligar a carboidratos ou a peptídeos e postula-se que esses RLKs possam se ligar a quitobiase, um componente de parede celular de fungos, insetos e nematóides (Shiu & Bleeker, 2001a, Gouget *et al.*, 2006). Alguns receptores com domínios extracelulares de L-lectina em *Arabidopsis* participam de vias de sinalização envolvendo adesão da membrana plasmática com a parede celular e são capazes de reconhecer proteínas com o motivo RGD (motivo de arginina-glicina-ácido aspártico), um motivo de adesão celular comum em mamíferos (Gouget *et al.*, 2006). Um RLK de *Populus nigra* com domínio de L-lectina foi visto como induzido por ferimento em folhas jovens (Nishiguchi *et al.*, 2002). No caso do RLK de cana-de-açúcar aqui relatado, sua regulação pela seca e também em entrenós maduros, com alta capacidade de acúmulo de sacarose, pode apontar para uma função desse RLK na sinalização de vias que levam a ajustes osmóticos nas células.

Dentre os fatores de transcrição regulados diferencialmente entre tecidos de genótipos de AB e de BB, destacam-se dois fatores de transcrição com dedos de zinco (SCEZST3147A10.g e SCEQRT1033F01.g). Esses fatores mostraram-se mais expressos na população de AB, quando amostras de entrenó 5 foram comparadas entre si. Além disso, como já discutido anteriormente, esses dois genes foram também induzidos pela interação planta-bactérias diazotróficas endofíticas, o que é um indício de que os fatores codificados por esses genes podem ter um papel em reconhecimento de invasão por bactérias. Fatores de transcrição com dedos de zinco desempenham diversos papéis em plantas, incluindo desenvolvimento floral, morfogênese regulada pela luz e respostas a patógenos (Takatsuji, 1998). Sendo assim, esses genes podem representar mais um caso de genes relacionados à patogênese com expressão diferencial entre os genótipos de AB e de BB.

Ao contrário dos experimentos de microarranjos, onde as amostras de RNA utilizadas foram extraídas a partir de material macerado de oito indivíduos da população de AB e oito da população de BB, optou-se por fazer os experimentos de PCR em tempo real a partir de amostras extraídas separadamente de seis indivíduos (três da população de AB e três da população de BB) para folhas ou de quatro indivíduos para amostras de entrenó 9, com o propósito de não apenas avaliar os resultados obtidos a partir de hibridização contra microarranjos de cDNA, mas também de verificar se os perfis de expressão detectados também se confirmam em cada indivíduo separadamente. Os resultados de PCR em tempo real mostraram que os níveis de expressão podem variar muito de indivíduo para indivíduo. Os resultados obtidos por PCR em tempo real foram comparados entre indivíduos de alto e de baixo brix, par a par, calculando-se, para isso, o valor P, como feito para os experimentos de respostas a hormônios e a fatores ambientais. Adotando-se como critério um valor P igual ou superior a 0,97, obtêm-se as porcentagens de validação mostradas, para cada um dos genes selecionados, na **Tabela 21**. Computando-se todos os dados conjuntamente, um índice de validação de 57% é obtido. Apesar desse índice parecer baixo, é importante notar que essa comparação envolve diferentes genótipos de cana-de-açúcar, e não réplicas biológicas, como feito nos experimentos de respostas a fitormônios e a estímulos bióticos e abióticos. No caso dos experimentos de microarranjos, os resultados de expressão obtidos correspondem a uma média de expressão entre a população AB *versus* BB. Já os ensaios de PCR em tempo real foram realizados com amostras dos genótipos isoladamente. Apesar do índice de validação variar muito de gene para gene, alguns chamam a atenção, como, por exemplo, o gene que codifica para uma proteína LTI (**Figura 15**), que se mostrou mais expresso em todas as amostras de folhas provenientes de genótipos de BB analisadas. Com isso, conclui-se que, a partir dos dados gerados por microarranjos, genes com perfis de expressão interessantes podem ser analisados em amostras de genótipos individualmente quanto ao potencial de serem bons marcadores moleculares para acúmulo de sacarose.

É também importante ressaltar que os experimentos de microarranjos conduzidos de forma a comparar órgãos de genótipos de AB e de BB entre si não é suficiente para distinguir o que é causa (ou seja, quais são os genes cuja regulação leva a um maior ou menor acúmulo de

sacarose) do que é consequência (ou seja, regulações gênicas que permitem a adaptação celular a uma condição de elevado acúmulo de sacarose). Para isso, estudos de superexpressão ou de silenciamento gênico serão importantes para complementar os resultados aqui discutidos.

É possível que alguns dos genes detectados como diferencialmente expressos não estejam diretamente relacionados a eventos do tipo causa/efeito associados aos níveis de sacarose, mas que estejam intimamente ligados a regiões cromossômicas selecionadas durante os cruzamentos. O uso do melhoramento genético clássico, através do cruzamento com espécies ancestrais e como forma de se incorporar características de interesse na progênie, usa parentais que possuem uma elevada proporção de variação genética inferior à dos cultivares comerciais modernos. Com isso, cruzamentos e seleções podem ser feitos com determinado objetivo, como, por exemplo, gerar cultivares com maior teor de sacarose, mas os cultivares gerados podem, também, ser menos resistentes a pragas ou trazer alterações nos níveis de expressão gênica não relacionadas à sacarose.

Os cruzamentos e as seleções estudadas no presente projeto objetivaram gerar genótipos de cana-de-açúcar contrastantes para teor de brix via a introdução de genes de espécies ancestrais que ainda não haviam sido explorados no desenvolvimento dos cultivares híbridos comerciais. O uso de germoplasma de 21 genótipos de *S. officinarum* e 13 de *S. spontaneum* permitiu a seleção de materiais mais divergentes do que o cruzamento entre cultivares comerciais. Os níveis de brix observados variaram de 4,8 a 19,6° Bx, uma diferença que nunca havia sido observada anteriormente nos cruzamentos tradicionais. Os genótipos gerados constituem, então, uma população altamente interessante para estudos de acúmulo de sacarose, ainda que com as considerações indicadas acima. Os resultados de expressão obtidos provavelmente são diferentes dos obtidos comparando-se populações derivadas de cruzamentos entre cultivares comerciais, com fenótipos não tão contrastantes quanto os dos genótipos aqui analisados. A abordagem utilizada é valiosa para detectar genes com potencial de serem adotados como marcadores moleculares em programas de melhoramento genético ou de serem manipulados por engenharia genética de forma a gerar plantas transgênicas com maior capacidade de síntese e/ou acúmulo de sacarose.

8. Conclusões & Perspectivas

O presente projeto teve como foco a análise de componentes de transdução de sinais em cana-de-açúcar. Em um primeiro momento, foi construído o Catálogo SUCAST, atualmente composto por mais de 3.500 SAS e a categoria referente ao quinoma de cana-de-açúcar foi analisada através de uma abordagem filogenética que permitiu a classificação de 527 proteínas quinases e RLKs em grupos e famílias. Posteriormente, a expressão gênica de componentes de transdução de sinais selecionados foi avaliada pela técnica de microarranjos de cDNA. Para isso, amostras de diferentes órgãos de cana-de-açúcar e de duas populações contrastantes para teor de brix foram comparadas. Além disso, a expressão gênica desses componentes foi avaliada frente a uma diversidade de estímulos ambientais e a tratamento com fitormônios. Aqui deve ser destacada a importância da tecnologia de microarranjos de DNA, que viabilizou a análise funcional, inferida a partir da expressão gênica, de mais de 1.000 genes concomitantemente, bem como permitiu esse tipo de análise sob as mais diversas condições experimentais. No entanto, para que essa tecnologia pudesse ser aplicada de forma confiável e reprodutível, cada uma de suas etapas, incluindo o método de normalização e de análise dos dados gerados, foi cuidadosamente padronizada.

Para os experimentos de microarranjos para órgãos, todas as hibridizações foram feitas contra uma referência comum, constituída de uma mistura equimolar de RNA de diferentes órgãos (raiz, folha, flor, entrenó e gema lateral). Com isso, foram obtidos 217 SAS diferencialmente expressos e 153 genes homogeneamente expressos em todos os órgãos estudados. Esses dados são particularmente valiosos para biotecnologia de cana-de-açúcar, já que os promotores desses genes podem apresentar atividade órgão-dirigida, ou permitir a expressão do transgene em toda a planta, no caso dos promotores dos genes ubíquos detectados. Poucos promotores estão disponíveis para serem utilizados para a construção de plantas transgênicas de cana-de-açúcar. Atualmente, os promotores mais utilizados em plantas geneticamente modificadas são o promotor 35S do vírus do mosaico da couve flor (CaMV) e os promotores dos genes da nopalina e octopina de *Agrobacterium tumefaciens*. Apesar dos grandes avanços que têm sido obtidos com esses

promotores, os padrões de expressão dos transgenes são variados e baixos em alguns casos (Zheng & Murai, 1997, Green *et al.*, 2002), não havendo garantia de expressão no órgão adequado (Neuteboom *et al.*, 2002). Além disso, o promotor CaMV 35S, freqüentemente empregado para dicotiledôneas por ser um promotor forte e ubíquo, apresenta baixa atividade em monocotiledôneas, como é o caso da cana-de-açúcar. Dentre os poucos promotores utilizados para a expressão de transgenes em cana-de-açúcar, o de maior uso é o promotor da ubiquitina de milho, Ubi1 (Lakshmanan *et al.*, 2005). Além da baixa disponibilidade de bons promotores para o desenvolvimento de plantas transgênicas de cana-de-açúcar, convém apontar o fato de que o uso de promotores constitutivos é uma das causas da grande preocupação a respeito dos efeitos dos transgênicos no meio ambiente. Promotores como o CaMV 35S quando ligados a genes utilizados para seleção de transformantes (genes de resistência a antibióticos) determinam, em geral, a expressão do produto gênico em todos os órgãos da planta. Tal característica, desnecessária e indesejável em vegetais geneticamente modificados, tende a diminuir a aceitação dos produtos derivados pelos consumidores. Vale também ressaltar que a expressão constitutiva de um gene-alvo em plantas pode acarretar problemas fenotípicos tais como nanismo e deficiência no desenvolvimento e crescimento. O uso de promotores órgão-dirigidos pode, então, minimizar esses efeitos negativos e incrementar as características fenotípicas desejadas.

Para os experimentos de expressão gênica frente a estímulos ambientais e a fitormônios, 179 genes foram detectados como diferencialmente expressos e para os experimentos de tratamentos com fitormônios, deficiência nutricional e seca, os perfis de expressão para 695 SAS puderam ser avaliados ao longo da série temporal, o que tornou a análise desses dados bem mais abrangente. A análise dos dados de microarranjos aponta para muitas categorias gênicas já descritas na literatura como reguladas no nível da transcrição, mas mostra também muitos genes ainda não descritos e que representam alvos interessantes para serem estudados em maior profundidade em cana-de-açúcar. Dentro dessa categoria, estão incluídas diversas proteínas desconhecidas, genes sem homologias nos bancos públicos de dados e proteínas quinases indefinidas. A informação gerada pela análise filogenética, aliada a dados de expressão, constitui uma ferramenta importante para guiar a futura caracterização dessas quinases. A caracterização

da regulação gênica relacionada aos estresses em estudo propiciará a melhor compreensão dos mecanismos pelos quais a cana-de-açúcar responde a esses estímulos ambientais. Tal conhecimento, associado ao uso da engenharia genética, poderá levar ao desenvolvimento de cultivares mais adaptados a condições adversas, como seca, herbivoria e carência nutricional, o que é de grande valor para a agricultura.

Por último, genótipos contrastantes para o acúmulo de sacarose tiveram seus transcriptomas comparados entre si, permitindo a identificação de 51 genes possivelmente relacionados ao acúmulo de sacarose. Esses genes podem ter papéis importantes nos mecanismos de regulação da síntese ou do acúmulo de sacarose em cana-de-açúcar e, portanto, poderão ser úteis para a melhor compreensão de como esses processos são regulados nessa gramínea. Adicionalmente, os genes identificados no projeto possuem também o potencial de serem usados como marcadores moleculares na seleção de parentais e progênies de interesse no melhoramento genético tradicional.

Cada um dos “blocos” de experimentos de microarranjos conduzidos (comparação entre seis órgãos, respostas a diversos estímulos e comparação de genótipos de alto e de baixo brix) foi analisado separadamente, de modo a permitir que a análise dos dados fosse a mais adequada a cada caso. Todos os resultados provenientes de experimentos de microarranjos e de PCR em tempo real, bem como dados de buscas por domínios conservados e a análise filogenética conduzida para o quinoma de cana-de-açúcar, foram integrados no banco de dados SUCAST. O banco de dados SUCAST será sempre atualizado, assim que novos dados forem sendo gerados pelo grupo de pesquisa da Dra. Gláucia Mendes Souza e certamente representará uma grande ferramenta para a busca de alvos para serem estudados funcionalmente em cana-de-açúcar e em outras plantas e também para o desenvolvimento de plantas transgênicas que sejam mais produtivas.

9. Colaboradores

Flávia Stal Papini Terzi

Ricardo Zorzetto Nicolliello Vêncio

Milton Yutaka Nishiyama Junior

Os grupos liderados pelos pesquisadores abaixo relacionados foram colaboradores do presente projeto:

Dr. Marcelo Menossi, Centro de Biologia Molecular e Engenharia Genética (CBMEG) - UNICAMP - Campinas, SP,

Dr. Eugênio C. Ulian, Centro de Tecnologia Canavieira, Piracicaba, SP,

Dra. Aline Maria da Silva, Instituto de Química (IQ) – USP – São Paulo, SP,

Dra. Adriana Silva Hemerly, Instituto de Ciências Biomédicas (ICB) – UFRJ – Rio de Janeiro, RJ,

Dr. Márcio de Castro Silva Filho, Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ) - USP - Piracicaba, SP,

Dr. Antônio Vargas de Oliveira Figueira, Centro de Energia Nuclear na Agricultura (CENA) - USP - Piracicaba, SP,

Dra. Sônia Marli Zingaretti, Faculdade de Ciências Agrárias e Veterinárias - UNESP - Jaboticabal, SP.

10. Referências

- Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D., Shinozaki, K. (1997). Role of *Arabidopsis* MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell*. 9:1859-1868.
- Abel, S., Nurnberger, T., Ahnert, V., Krauss, G.J., Glund, K. (2000). Induction of an extracellular cyclic nucleotide phosphodiesterase as an accessory ribonucleolytic activity during phosphate starvation of cultured tomato cells. *Plant Physiol*. 122:543-552.
- Abel, S., Theologis, A. (1996). Early genes and auxin action. *Plant Physiol*. 111:9-17.
- Abel, S., Ticconi, C.A., Delatorre, C.A. (2002). Phosphate sensing in higher plants. *Physiol Plant*. 115:1-8.
- Ahmad, M. (1999). Seeing the world in red and blue: insight into plant vision and photoreceptors. *Curr. Opin. Plant Biol*. 2, 230-235.
- Albrecht, V., Ritz, O., Linder, S., Harter, K., Kudla, J. (2001). The NAF domain defines a novel protein-protein interaction module conserved in Ca²⁺-regulated kinases. *The EMBO Journal* 20: 1051-1063.
- Altschul, S.F., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25: 3389-3402.
- Alvarez-Buylla, E.R., Liljegren, S.J., Pelaz, S., Gold, S.E., Burgeff, C., Ditta, G.S., Vergara-Silva, F., Yanofsky, M.F. (2000). MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J*. 24:457-466.
- Aneeta Sanan-Mishra N., Tuteja, N., Kumar Sopory S. (2002). Salinity- and ABA-induced up-regulation and light-mediated modulation of mRNA encoding glycine-rich RNA-binding protein from *Sorghum bicolor*. *Biochem Biophys Res Commun*. 296:1063-1068.
- Annen, F., Stockhaus, J. (1999). SbRLK1, a receptor-like protein kinase of *Sorghum bicolor* (L.) Moench that is expressed in mesophyll cells. *Planta* 208:420-425.
- Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815.
- Aravind, L., Koonin, E. V. (2000). The U box is a modified RING finger- a common domain in ubiquitination. *Curr. Biol*. 10:R132-134.
- Arimura, G., Kost, C., Boland, W. (2005). Herbivore-induced, indirect plant defences. *Biochim Biophys Acta*. 1734:91-111.
- Assmann, S. M., Shimazaki, K. (1999). The Multisensory Guard Cell. Stomatal Responses to Blue Light and Absciscic Acid. *Plant Physiol*. 119: 809-816.
- Aukerman, M.J., Sakai, H. (2003). Regulation of Flowering Time and Floral Organ Identity by MicroRNA and Its APETALA2-Like Target Genes. *The Plant Cell* 15, 2730-2741.
- Azevedo, C., Santos-Rosa, M.J., Shirasu, K. (2001). The U-box protein family in plants. *Trends Plant Sci*. 6:354-358.
- Bai, C., Sen, P., Hofmann, K., Ma, L., Goebel, M., Harper, J.W., Elledge, S.J. (1996). SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* 86: 263-274.
- Baldani, J.I., Baldani, V.L. (2005). History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. *An Acad Bras Cienc*. 77:549-579.
- Baldani, J.I., Baldani, V.L.D., Seldin, L., Dobereiner, J. (1986). Characterization of *Herbaspirillum seropedicae* gen. nov., sp. nov., a root-associated nitrogen-fixing bacterium. *Int. J. Syst. Bacteriol*. 36: 86-93.
- Baldani, J.I., Pot, B., Kirchhof, G., Falsen, E., Baldani, V.L., Olivares, F.L., Hoste, B., Kersters, K., Hartmann, A., Gillis, M., Dobereiner, J. (1996). 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. *Int. J. Syst. Bacteriol*. 46:802-810.
- Baldwin, I.T., Halitschke, R., Paschold, A., von Dahl, C.C., Preston, C.A. (2006). Volatile signaling in plant-plant interactions: "talking trees" in the genomics era. *Science*. 311:812-815.
- Bariola, P.A., MacIntosh, G.C., Green, P.J. (1999). Regulation of S-Like Ribonuclease Levels in *Arabidopsis*. Antisense Inhibition of RNS1 or RNS2 Elevates Anthocyanin Accumulation. *Plant Physiology* 119: 331-342.
- Barrett, T., Suzek, T.O., Troup, D.B., Wilhite, S.E., Ngau, W.C., Ledoux, P., Rudnev, D., Lash, A.E., Fujibuchi, W., Edgar, R. (2005). NCBI GEO: mining millions of expression profiles--database and tools. *Nucleic Acids Res*. 33(Database issue):D562-566.
- Barsalobres, C.F. Análise da expressão gênica induzida por *Diatraea saccharalis* em cana-de-açúcar via macroarranjos de colônias bacterianas. 2004. 228 p. Dissertação (mestrado)- Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, SP, 2004.

- Bartel, B., Fink, G.R. (1994). Differential regulation of an auxin-producing nitrilase gene family in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 91:6649-6653.
- Bateman, A., Bycroft, M. (2000). The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *J. Mol. Biol.* 299:1113-1119.
- Beisson, F., Koo, A.J.K., Ruuska, S., Schwender, J., Pollard, M., Thelen, J.J., Paddock, T., Salas, J.J., Savage, L., Milcamps, A., Mhaske, V.B., Cho, Y., Ohlrogge, J.B. (2003). *Arabidopsis* Genes Involved in Acyl Lipid Metabolism. A 2003 Census of the Candidates, a Study of the Distribution of Expressed Sequence Tags in Organs, and a Web-Based Database. *Plant Physiol.* 132:681-697.
- Bennett, V., Chen, L. (2001). Ankyrins and cellular targeting of diverse membrane proteins to physiological sites. *Curr Opin Cell Biol.* 13:61-67.
- Blanco, Y., Arroyo, M., Legaz, M.E., Vicente, C. (2005). Isolation from *Gluconacetobacter diazotrophicus* cell walls of specific receptors for sugarcane glycoproteins, which act as recognition factors. *J Chromatogr A*. 1093:204-211.
- Bohnert, H.J., Nelson, D.E., Jensen, R.G. (1995). Adaptations to Environmental Stresses. *Plant Cell*. 7:1099-1111.
- Bosse, D., Köck, M. (1998). Influence of phosphate starvation on phosphohydrolases during development of tomato seedlings. *Plant Cell Environ.* 21: 325-332.
- Bray, E.A. (1993). Molecular Responses to Water Deficit. *Plant Physiol.* 103:1035-1040.
- Bray, E.A., Bailey-Serres, J., Weretilnyk, E. Responses to abiotic stresses. In: Buchanan, B., Gruissem, W., Jones, R. (eds) *Biochemistry & Molecular Biology of Plants*. Rockyville, Maryland: American Society of Plant Physiologists, 2000. p. 1158-1203.
- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., Vingron, M. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet.* 29:365-371.
- Broner, I., Law, R.A.P. (1991). Evaluation of a modified atmometer for estimating reference ET. *Irrigation Science*, 12:21-26.
- Brueggeman, R., Rostoks, N., Kudrna, D., Kilian, A., Han, F., Chen, J., Druka, A., Steffenson, B., Kleinhofs, A. (2002). The barley stem rust-resistance Rpg1 is a novel disease-resistance gene with homology to receptor kinases *Proc. Natl. Acad. Sci. U S A* 99, 9328-9333.
- Canales, C., Bhatt, A.M., Scott, R., Dickinson, H. (2002). EXS, a putative LRR receptor kinase, regulates male germline cell number and tapetal identity and promotes seed development in *Arabidopsis*. *Curr Biol.* 12:1718-1727.
- Capoen, W., Goormachtig, S., De Rycke, R., Schroeyers, K., Holsters, M. (2005). SrSymRK, a plant receptor essential for symbiosome formation. *Proc. Natl. Acad. Sci. U S A*. 102:10369-10374.
- Carson, D.L., Botha, F.C. (2002). Genes expressed in sugarcane maturing internodal tissue. *Plant Cell Reports* 20, 1075-1081.
- Carson, D.L., Hockett, B.I., Botha, F.C. (2002). Sugarcane ESTs differentially expressed in immature and maturing internodal tissue. *Plant Sci.* 162, 289-300.
- Casu, R., Dimmock, C., Chapman, S., Grof, C., McIntyre, C.L., Bonnett, G., Manners, J. (2004). Identification of Differentially Expressed Transcripts from Maturing Stem of Sugarcane by in silico Analysis of Stem Expressed Sequence Tags and Gene Expression Profiling. *Plant Mol. Biol.* 54, 503-517.
- Casu, R.E., Grof, C.P.L., Rae, A.L., McIntyre, C.L., Dimmock, C.M., Manners, J.M. (2003). Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. *Plant Mol. Biol.* 52, 371-386.
- Casu, R.E., Manners, J.M., Bonnett, G.D., Jackson, P.A., McIntyre, C.L., Dunne, R., Chapman, S.C., Rae, A.L., Grof, C.P.L. (2005). Genomics approaches for the identification of genes determining important traits in sugarcane. *Field Crops Research* 92:137-147.
- Cavalcante, V.A., Dobereiner, J. (1988). A new acid-tolerant nitrogen-fixing bacterium associated with sugarcane. *Pl Soil* 108: 23-31.
- Chaumont, F., Barrieu, F., Jung, R., Chrispeels, M.J. (2000). Plasma membrane intrinsic proteins from maize cluster in two sequence subgroups with differential aquaporin activity. *Plant Physiol.* 122:1025-1034.
- Chaumont, F., Moshelion, M., Daniels, M.J. (2005). Regulation of plant aquaporin activity. *Biol Cell*, 97:749-64.
- Chen, J.G., Pandey, S., Huang, J., Alonso, J.M., Ecker, J.R., Assmann, S.M., Jones, A.M. (2004). GCR1 can act independently of heterotrimeric G-protein in response to brassinosteroids and gibberellins in *Arabidopsis* seed germination. *Plant Physiol.* 135:907-915.
- Chen, W., Provart, N.J., Glazebrook, J., Katagiri, F., Chang, H.S., Eulgem, T., Mauch, F., Luan, S., Zou, G., Whitham, S.A., Budworth, P.R., Tao, Y., Xie, Z., Chen, X., Lam, S., Kreps, J.A., Harper, J.F., Si-Ammour, A., Mauch-Mani, B., Heinlein, M., Kobayashi, K., Hohn, T., Dangl, J.L., Wang, X., Zhu, T. (2002). Expression

profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell*. 14:559-74.

Chen, X. (2004). A MicroRNA as a Translational Repressor of APETALA2 in *Arabidopsis* Flower Development. *Science* 303:2022-2025.

Cheong, J.J., Choi, Y.D. (2003). Methyl jasmonate as a vital substance in plants. *Trends Genet.* 19:409-13.

Cheong, Y.H., Chang, H-S., Gupta, R., Wang, X., Zhu, T., Luan, S. (2002). Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress and hormonal responses in *Arabidopsis*. *Plant Physiol.* 129:661-677.

Chinnusamy, V., Schumaker, K., Zhu, J.K. (2004). Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. *J. Exp. Bot.* 55:225-236.

Cho, Y., Fernandes, J., Kim, S.H., Walbot, V. (2002). Gene-expression profile comparisons distinguish seven organs of maize. *Genome Biol.* 3, research0045.

Clark, G.B., Thompson, G. Jr., Roux, S.J. (2001). Signal transduction mechanisms in plants: an overview. *Curr. Sci.* 80:170-177.

Clark, S.E., Williams, R.W., Meyerowitz, E.M. (1997). The CLAVATA1 gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell* 89:575-585.

Cooper, B., Clarke, J.D., Budworth, P., Kreps, J., Hutchison, D., Park, S., Guimil, S., Dunn, M., Luginbuhl, P., Ellero, C., Goff, S.A., Glazebrook, J. (2003). A network of rice genes associated with stress response and seed development. *Proc. Natl. Acad. Sci. U S A.* 100:4945-4950.

Couee, I., Sulmon, C., Gouesbet, G., El Amrani, A. (2006). Involvement of soluble sugars in reactive oxygen species balance and responses to oxidative stress in plants. *J. Exp. Bot.* 57:449-459.

Creelman, R.A., Mullet, J.E. (1997). Biosynthesis and action of jasmonates in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48:355-381.

De Rosa, Jr. V.E., Nogueira, F.T.S., Menossi, M., Ulian, E.C., Arruda, P. (2005). Identification of methyl jasmonate-responsive genes in sugarcane using cDNA arrays. *Braz. J. Plant Physiol.* 17, 173-180.

Delaney, T.P., Uknes, S., Vernooij, B., Friedrich, L., Weymann, K., Negrotto, D., Gaffney, T., Gut-Rella, M., Kessmann, H., Ward, E., and Ryals, J. (1994). A central role of salicylic acid in plant disease resistance. *Science* 266:1247-1250.

Devoto, A., Nieto-Rostro, M., Xie, D., Ellis, C., Harmston, R., Patrick, E., Davis, J., Sherratt, L., Coleman, M., Turner, J.G. (2002). CO1 links jasmonate signalling and fertility to the SCF ubiquitin-ligase complex in *Arabidopsis*. *Plant J.* 32:457-466.

Devoto, A., Turner, J.G. (2003). Regulation of jasmonate-mediated plant responses in *Arabidopsis*. *Ann. Bot.* 92:329-337.

Dharmasiri, N., Dharmasiri, S., Estelle, M. (2005). The *F-box* protein TIR1 is an auxin receptor. *Nature* 435:441-445.

Dumont, F., Joris, B., Gumusboga, A., Bruyninx, M., Loppes, R. (1993). Isolation and characterization of cDNA sequences controlled by inorganic phosphate in *Chlamydomonas reinhardtii*. *Plant Sci.* 89:55-67.

Eckert, M., Kaldenhoff, R. (2000). Light-induced stomatal movement of selected *Arabidopsis thaliana* mutants. *J Exp Bot.* 51:1435-1442.

Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kalo, P., Kiss, G.B. (2002). A receptor kinase gene regulating symbiotic nodule development. *Nature* 417:962-966.

Falco, M.C., Marbach, P.A.S., Pompermayer, P., Lopes, F.C.C., Silva-Filho, M.C. (2001). Mechanisms of sugarcane response to herbivory. *Genet. Mol. Biol.* 24:113-122.

Farmer, E.E., Almeras, E., Krishnamurthy, V. (2003). Jasmonates and related oxylipins in plant responses to pathogenesis and herbivory. *Curr. Opin. Plant Biol.* 6:372-378.

Felix, J.M. Análise da expressão gênica envolvida no metabolismo de sacarose em cana-de-açúcar (*Saccharum* spp.). 2006. 174 p. Tese (doutorado) - Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, 2006.

Finkelstein, R.R., Gibson, S.I. (2002). ABA and sugar interactions regulating development: cross-talk or voices in a crowd? *Curr. Opin. Plant Biol.* 5:26-32.

Franco, O.L., Rigden, D.J., Melo, F.R., Grossi-De-Sa, M.F. (2002). Plant alpha-amylase inhibitors and their interaction with insect alpha-amylases. *Eur. J. Biochem.* 269:397-412.

Franco-Zorrilla, J.M., Gonzalez, E., Bustos, R., Linhares, F., Leyva, A., Paz-Ares, J. (2004). The transcriptional control of plant responses to phosphate limitation. *J. Exp. Bot.* 55:285-293.

Freire, M.A., Pages, M. (1995). Functional characteristics of the maize RNA-binding protein MA16. *Plant Mol. Biol.* 29:797-807.

Frey, M., Chomet, P., Glawischnig, E., Stettner, C., Grun, S., Winklmair, A., Eisenreich, W., Bacher, A., Meeley, R.B., Briggs, S.P., Simcox, K., Gierl, A. (1997). Analysis of a chemical plant defense mechanism in grasses. *Science* 277:696-9.

- Gaffney, T., Friedrich, L., Vernooij, B., Negrotto, D., Nye, G., Uknes, S., Ward, E., Kessmann, H., Ryals, J. (1993). Requirement of salicylic acid for the induction of systemic acquired resistance. *Science* 261, 754–756.
- Gazzarrini, S., McCourt, P. (2003). Cross-talk in plant hormone signalling: what *Arabidopsis* mutants are telling us. *Ann. Bot.* 91:605-612.
- Gene Ontology Consortium. (2006). The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.* 34:D322-326.
- Glasziou, K. T., Gayler, K. R. (1972). Storage of sugars in stalks of sugar cane. *The Botanical Review* 38(4): 471-488.
- Glickman, M.H., Ciechanover, A. (2002). The ubiquitin-proteasome proteolytic pathway: Destruction for the sake of construction. *Physiological Reviews* 82: 373–428.
- Goebel, M., Yanagida, M. (1991). The TPR snap helix: a novel protein repeat motif from mitosis to transcription. *Trends Biochem. Sci.* 16:173-177.
- Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531-537.
- Gouget, A., Senchou, V., Govers, F., Sanson, A., Barre, A., Rouge, P., Pont-Lezica, R., Canut, H. (2006). Lectin receptor kinases participate in protein-protein interactions to mediate plasma membrane-cell wall adhesions in *Arabidopsis*. *Plant Physiol.* 140:81-90.
- Green, J., Vain, P., Fearnough, M.T., Worland, B., Snape, J.W., Atkinson, H.J. (2002). Analysis of the expression patterns of the *Arabidopsis thaliana* tubulin-1 and *Zea mays* ubiquitin-1 promoters in rice plants in association with nematode infection. *Physiological and Molecular Plant Pathology*, 60:197-205.
- Grenier, J., Potvin, C., Trudel, J., Asselin, A. (1999). Some thaumatin-like proteins hydrolyse polymeric beta-1,3-glucans. *Plant J.* 19:473-480.
- Gribskov, M., Fana, F., Harper, J., Hope, D. A., Harmon, A. C., Smith, D. S, Tax, F. E., Zhang, G. (2001). PlantsP: a functional genomics database for plant phosphorylation. *Nucleic Acids Res.* 29:111-113.
- Grivet, L., Arruda, P. (2001). Sugarcane genomics: depicting the complex genome of an important tropical crop. *Curr. Opin. Plant Biol.* 5:122-127.
- Grivet, L., Daniels, C., Glaszmann, J. C., D'Hont, A. (2004). A review of recent molecular genetics evidence for sugarcane evolution and domestication. *Ethnobotany Research & Applications* 2:9-17.
- Halford, N.G., Paul, M.J. (2003). Carbon metabolite sensing and signalling. *Plant Biotechnology Journal* 1:381-398.
- Hammond, J.P., Broadley, M.R., White, P.J. (2004). Genetic responses to phosphorus deficiency. *Ann. Bot.* 94:323-332.
- Hanggi, E., Fleming, A.J. (2001). Sucrose synthase expression pattern in young maize leaves: implications for phloem transport. *Planta* 214:326-329.
- Hanks, S. K., Hunter, T. (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* 9:576-596.
- Hanks, S. K., Quinn, A. M. (1991). Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol.* 200:38-62.
- Hanks, S. K., Quinn, A. M., Hunter, T. (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic domain. *Science* 241:42-52.
- Hardie, D.G. (1999). Plant protein serine threonine kinases: Classification and functions. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50:97-131.
- Hardin, S.C., Huber, S.C. (1999). Proteasome activity and the post-translational control of sucrose synthase stability in maize leaves. *Plant Physiol Biochem.* 42:197-208.
- Hardin, S.C., Tang, G.Q., Scholz, A., Holtgraewe, D., Winter, H., Huber, S.C. (2003). Phosphorylation of sucrose synthase at serine 170: occurrence and possible role as a signal for proteolysis. *Plant J.* 35:588-603.
- Hardin, S.C., Winter, H., Huber, S.C. (2004). Phosphorylation of the amino terminus of maize sucrose synthase in relation to membrane association and enzyme activity. *Plant Physiol.* 134:1427-1438.
- Hazen, S.P., Wu, Y., Kreps, J.A. (2003). Gene expression profiling of plant responses to abiotic stress. *Funct. Integr. Genomics.* 3:105-111.
- Hendre, K.R., Iyer, R.S., Kotwain, M., Kluspe, S.S., Mascarenhas, A.F. (1983). Rapid multiplication of sugarcane by tissue culture. *Sugarcane* 1:5–8.
- Henning, L., Gruissem, W., Grossniklaus, U., Kohler, C. (2004). Transcriptional programs of early reproductive stages in *Arabidopsis*. *Plant Physiol.* 135:1765-1775.
- Himmelbach, A., Yang, Y., Grill, E. (2003). Relay and control of abscisic acid signaling. *Curr Opin Plant Biol.* 6:470-479.
- Hirayama, T., Oka, A. (1992). Novel protein kinase of *Arabidopsis thaliana* (APK1) that phosphorylates tyrosine, serine and threonine. *Plant Mol. Biol.* 20:653-662.
- Hoagland, D.R., Arnon, D. I. (1950). The water culture method for growing plants without soil. *California Agr. Expt. Sta. Cir.* 347, Berkely, CA.

- Holford, I.C.R. (1997). Soil phosphorus: its measurement, and its uptake by plants. *Aust. J. Soil Res.* 35:227-240.
- Hrabak, E.M., Chan, C.W., Gribskov, M., Harper, J.F., Choi, J.H., Halford, N., Kudla, J., Luan, S., Nimmo, H.G., Sussman, M.R., Thomas, M., Walker-Simmons, K., Zhu, J.K., Harmon, A.C. (2003). The *Arabidopsis* CDPK-SnRK superfamily of protein kinases. *Plant Physiol.* 132:666-680.
- Huber, S.C., Huber, J.L., Liao, P.C., Gage, D.A., McMichael, R.W. Jr., Chourey, P.S., Hannah, L.C., Koch, K. (1996). Phosphorylation of serine-15 of maize leaf sucrose synthase. Occurrence in vivo and possible regulatory significance. *Plant Physiol.* 112:793-802.
- Ingram, J., Bartels, D. (1996). The molecular basis of dehydration tolerance in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47:377-403.
- Inoue, K., Sewalt, V.J., Murray, G.B., Ni, W., Sturzer, C., Dixon, R.A. (1998). Developmental expression and substrate specificities of alfalfa caffeic acid 3-O-methyltransferase and caffeoyl coenzyme A 3-O-methyltransferase in relation to lignification. *Plant Physiol.* 117:761-770.
- Iseli, C., Jongeneel, C.V., Bucher, P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 138-48.
- Iskandar, H.M., Simpson, R.S., Casu, R.E., Bonnett, G.D., MacLean, D.J., Manners, J.M. (2004). Comparison of Reference Genes for Quantitative Real-Time Polymerase Chain Reaction Analysis of Gene Expression in Sugarcane. *Plant Molecular Biology Reporter* 22:325-337.
- Ito, Y., Banno, H., Moribe, T., Hinata, K., Machida, Y. (1994). NPK15, a tobacco protein-serine/threonine kinase with a single hydrophobic region near the amino-terminus. *Mol Gen Genet.* 245:1-10.
- Jacobsen, K.R., Fisher, D.G., Maretzki, A., Moore, P.H. (1992). Developmental-changes in the anatomy of the sugarcane stem in relation to phloem unloading and sucrose storage. *Botanica Acta* 105:70-80.
- James, E.K., Reis, V.M., Olivares, F.L., Baldani, J.I., Dobereiner, J. (1994). Infection of sugar cane by the nitrogen-fixing bacterium *Acetobacter diazotrophicus*. *J. Exp. Bot.* 45:757-766.
- Jang, J.C., Sheen, J. (1994). Sugar sensing in higher plants. *Plant Cell.* 6:1665-1679.
- Jansen, R.C., Nap, J.P. (2001). Genetical genomics: the added value from segregation. *Trends in Genetics* 17:388-391.
- Johnson, K.L., Ingram, G.C. (2005). Sending the right signals: regulating receptor kinase activity. *Curr. Opin. Plant Biol.* 8:648-656.
- Jonak, C., Hirt, H. (2002). Glycogen synthase kinase 3/ SHAGGY-like kinases in plants: an emerging family with novel functions. *Trends in Plant Sci.* 7:457-461.
- Josefsson, L.-G., Rask, L. (1997). Cloning of a putative G-protein coupled receptor from *Arabidopsis thaliana*. *Eur. J. Biochem.* 249:415-420.
- Kapranov, P., Jensen, T.J., Poulsen, C., de Bruijn, F.J., Szczyglowski, K. (1999). A protein phosphatase 2C gene, LjNPP2C1, from *Lotus japonicus* induced during root nodule development. *Proc. Natl. Acad. Sci. U S A.* 96:1738-1743.
- Kates, R.W., Parris, T.M. (2003). Long-term trends and a sustainability transition. *Proc. Natl. Acad. Sci. U S A* 100:8062-8067.
- Kawasaki, S., Borchert, C., Deyholos, M., Wang, H., Brazille, S., Kawai, K., Galbraith, D., Bohnert, H.J. (2001). Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell* 13:889-905.
- Kepinski, S., Leyser, O. (2005). The *Arabidopsis* *F-box* protein TIR1 is an auxin receptor. *Nature* 435:446-451.
- Kim, B., Nam, H.Y., Kim, S.U., Kim, S.I., Chang, Y.J. (2003). Normalization of reverse transcription quantitative-PCR with housekeeping genes in rice. *Biotechnology Letters* 25:1869-1872.
- Kim, H., Zhao, B., Snesrud, E.C., Haas, B.J., Town, C.D., Quackenbush, J. (2002). Use of RNA and genomic DNA references for inferred comparisons in DNA microarray analyses. *Biotechniques* 33:924-930.
- Kim, K.N., Cheong, Y.H., Gupta, R., Luan, S. (2000). Interaction specificity of *Arabidopsis* calcineurin B-like calcium sensors and their target kinases. *Plant Physiol.* 124:1844-1853.
- Kim, S.A., Kwak, H.J., Park, M.C., Kim, S.R. (1998). Induction of reproductive organ-preferential histone genes by wounding or methyl jasmonate. *Mol Cells.* 8:669-677.
- Klein, D. (2002). Quantification using real-time PCR technology: applications and limitations. *Trends in Molecular Medicine* 8:257-260.
- Koch, K. (2004). Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Curr. Opin. Plant Biol.* 7:235-246.
- Koch, K.E. (1996). Carbohydrate-Modulated Gene Expression in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47:509-540.
- Koide, T., Salem-Izaac S.M., Gomes S.L., Vencio R.Z.N. (2006). SpotWhatR: a user-friendly microarray data analysis system. *Genet. Mol. Res.* 5: 93-107.
- Kolattukudy, P.E. (1984). Biochemistry and function of cutin and suberin. *Can. J. Bot.* 62:2918-2933.
- Kolomiets, M.V., Hannape, D.J., Chen, H., Tymeson, M., Gladon, R.J. (2001). Lipxygenase is involved in the control of potato tuber development. *Plant Cell* 13:613-626.

- Lakshmanan, P., Geijskes, R.J., Aitken, K. S., Grof, C. L. P., Bonnett, G. D., Smitha, G. R. (2005). Sugarcane biotechnology: the challenges and opportunities. *In Vitro Cell. Dev. Biol.-Plant* 41:345-363.
- Lamb, J.R., Tugendreich S., Hieter, P. (1995). Tetra-tryc peptide repeat interactions: to TPR or not to TPR? *Trends Biochem. Sci.* 20:257-259.
- Lawton, M.A., Yamamoto, R.T., Hanks, S.K., Lamb, C.J. (1989). Molecular cloning of plant transcripts encoding protein kinase homologs. *Proc Natl Acad Sci U S A* 86:3140-4.
- Lee, E.J., lai, H., Sano, H. and Koizumi, N. (2005). Sugar responsible and tissue specific expression of a gene encoding AtCIPK14, an *Arabidopsis* CBL-interacting protein kinase. *Biosci. Biotechnol. Biochem.*, 69:242-245.
- Levy, J., Bres, C., Geurts, R., Chalhou, B., Kulikova, O., Duc, G., Journet, E.P., Ane, J.M., Lauber, E., Bisseling, T., Denarie, J., Rosenberg, C., Debelle, F. (2004). A putative Ca^{2+} and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science*. 303:1361-1364.
- Lewin, B. Genes VII. In capítulo 26: Transdução de sinais. Porto Alegre: Artmed editora, 2001. p. 765-798.
- Li, J., Chory, J. (1997). A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell*. 90:929-938.
- Li, J., Lee, Y-R., J., Assmann, S.M. (1998). Guard cells possess a calcium-dependent protein kinase that phosphorylates the KAT1 potassium channel. *Plant Physiol.* 116: 785-795.
- Li, J., Wang, X.Q., Watson, M.B., Assmann, S.M. (2000). Regulation of abscisic acid-induced stomatal closure and anion channels by guard cell AAPK kinase. *Science* 287:300-303.
- Li, J.X., Assmann, S.M. (1996). An abscisic acid activated and calcium-independent protein kinase from guard cells of fava bean. *Plant Cell* 8:2359-2368.
- Limpens, E., Franken, C., Smit, P., Willemse, J., Bisseling, T., Geurts, R. (2003). LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. *Science* 302:630-633.
- Lin, F., Xu, S.L., Ni, W.M., Chu, Z.Q., Xu, Z.H., Xue, H.W. (2003). Identification of ABA-responsive genes in rice shoots via cDNA macroarray *Cell Res.*13:59-68.
- Lingle, S. E. (1997). Seasonal internode development and sugar metabolism in sugarcane. *Crop Sci.* 37:1222-1227.
- Liu, J., Ishitani, M., Halfter, U., Kim, C.S., Zhu, J.K. (2000). The *Arabidopsis thaliana* SOS2 gene encodes a protein kinase that is required for salt tolerance. *Proc. Natl. Acad. Sci. U S A*. 97:3730-3744.
- Liu, W., Saint, D.A. (2002a). A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. *Analytical Biochemistry* 302:52-59.
- Liu, W., Saint, D.A. (2002b) Validation of a quantitative method for real time PCR kinetics. *Biochem. Biophys. Res. Comm.* 294:347-353.
- Livak, K.J., Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* 25:402-408.
- Ljung, K., Hull, A.K., Celenza, J., Yamada, M., Estelle, M., Normanly, J., Sandberg, G. (2005). Sites and regulation of auxin biosynthesis in *Arabidopsis* roots. *Plant Cell* 17:1090-1104.
- Ljung, K., Hull, A.K., Kowalczyk, M., Marchant, A., Celenza, J., Cohen, J.D., Sandberg, G. (2002). Biosynthesis, conjugation, catabolism and homeostasis of indole-3-acetic acid in *Arabidopsis thaliana*. *Plant Mol. Biol.* 49:249-272.
- Lopez-Bucio, J., Hernandez-Abreu, E., Sanchez-Calderon, L., Nieto-Jacobo, M.F., Simpson, J., Herrera-Estrella, L. (2002). Phosphate availability alters architecture and causes changes in hormone sensitivity in the *Arabidopsis* root system. *Plant Physiol.* 129:244-256.
- Lu, Y. H., D'Hont, A., Paulet, F., Grivet, L., Arnaud, M., Glaszmann, J. C. (1994). Molecular diversity and genome structure in modern sugarcane varieties. *Euphytica* 78: 217-226.
- Ma, L., Chen, C., Liu, X., Jiao, Y., Su, N., Li, L., Wang, X., Cao, M., Sun, N., Zhang, X., Bao, J., Li, J., Pedersen, S., Bolund, L., Zhao, H., Yuan, L., Wong, G.K., Wang, J., Deng, X.W., Wang, J. (2005). A microarray analysis of the rice transcriptome and its comparison to *Arabidopsis*. *Genome Res.* 15:1274-1283.
- Ma, S., Quist, T.M., Ulanov, A., Joly, R., Bohnert, H.J. (2004). Loss of TIP1;1 aquaporin in *Arabidopsis* leads to cell and plant death. *Plant J.*, 40:845-859.
- MacKay, D. (2003). Probability, entropy, and inference. In: *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge, UK.
- Madsen, E.B., Madsen, L.H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J. (2003). A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* 425:637-640.
- Manning, G., Plowman, G.D., Hunter, T., Sudarsanam, S. (2002a). Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.* 27:514-520.
- Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S. (2002b). The protein kinase complement of the human genome. *Science* 298:1912-1934.
- Marschner A (1995) Mineral nutrition of higher plants. Academic Press, San Diego.

- Matsubayashi, Y., Ogawa, M., Morita, A., Sakagami, Y. (2002) An LRR receptor kinase involved in perception of a peptide plant hormone, phytosulfokine. *Science*. 296:1470-1472.
- Maurel, C., Chrispeels, M.J. (2001) Aquaporins. A molecular entry into plant water relations. *Plant Physiol.* 125:135-138.
- McCarty, D.R., Chory, J. (2000). Conservation and innovation in plant signaling pathways. *Cell* 103:201-209.
- McConn, M., Creelman, R.A., Bell, E., Mullet, J.E., Browse, J. (1997). Jasmonate is essential for insect defense in *Arabidopsis*. *Proc. Natl. Acad. Sci. U S A*. 94:5473-5477.
- McMichael, R.W. Jr., Bachmann, M., Huber, S.C. (1995). Spinach Leaf Sucrose-Phosphate Synthase and Nitrate Reductase Are Phosphorylated/Inactivated by Multiple Protein Kinases in Vitro. *Plant Physiol.*, 108:1077-1082.
- Merlot, S., Gosti, F., Guerrier, D., Vavasseur, A., Giraudat, J. (2001). The ABI1 and ABI2 protein phosphatases 2C act in a negative feedback regulatory loop of the abscisic acid signalling pathway. *Plant J.* 25:295-303.
- Minhas, D., Grover, A. (1999). Transcript level of genes encoding various glycolytic and fermentation enzymes change in response to abiotic stresses. *Plant Sci.* 146:41-51.
- Moller, S.G., Chua, N.H. (1999). Interactions and intersections of plant signaling pathways. *J Mol. Biol.* 293:219-234.
- Moore, P.M. (1995). Temporal and Spatial Regulation of Sucrose Accumulation in the Sugarcane Stem. *Aust. J. Plant Physiol.* 22:661-679.
- Moran, T.V., Walker, J.C. (1993). Molecular cloning of two novel protein kinase genes from *Arabidopsis thaliana*. *Biochim. Biophys. Acta*, 1216:9-14.
- Mousavi, A., Hotta, Y. (2005). Glycine-rich proteins: a class of novel proteins. *Appl. Biochem. Biotechnol.* 120:169-174.
- Muchhal, U.S., Pardo, J.M. and Raghothama, K.G. (1996). Phosphate transporters from the higher plant *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 93: 10519-10523.
- Mur, L., Bi, Y.-M., Darby, R.M., Firek, S., Draper, J. (1997). Compromising early salicylic acid accumulation delays the hypersensitive response and increases viral dispersal during lesion establishment in TMV-infected tobacco. *Plant J.* 12:1113-1126.
- Mustilli, A.C., Merlot, S., Vavasseur, A., Fenzi, F., Giraudat, J. (2002). *Arabidopsis* OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. *Plant Cell* 14:3089-3099.
- Nahm, M.Y., Kim, S.W., Yun, D., Lee, S.Y., Cho, M.J., Bahk, J.D. (2003). Molecular and biochemical analyses of OsRab7, a rice Rab7 homolog. *Plant Cell Physiol.* 44:1341-1349.
- Nambara, E., Marion-Poll, A. (2005). Absciscic acid biosynthesis and catabolism. *Annu. Rev Plant Biol.* 56:165-185.
- Narusaka, Y., Narusaka, M., Seki, M., Umezawa, T., Ishida, J., Nakajima, M., Enju, A., Shinozaki, K. (2004). Crosstalk in the responses to abiotic and biotic stresses in *Arabidopsis*: analysis of gene expression in cytochrome P450 gene superfamily by cDNA microarray. *Plant Mol. Biol.* 55:327-342.
- Neuteboom, L.W., Kunimitsu, W.Y., Webb, D., Christopher, D.A. (2002). Characterization and tissue-regulated expression of genes involved in pineapple (*Ananas comosus* L.) root development. *Plant Sci.* 163:1021-1035.
- Newton, R.P., Smith, C.J. (2004). Cyclic nucleotides. *Phytochemistry* 65:2423-2437.
- Nishiguchi, M., Yoshida, K., Sumizono, T., Tazaki, K. (2002). A receptor-like protein kinase with a lectin-like domain from lombardy poplar: gene expression in response to wounding and characterization of phosphorylation activity. *Mol Genet Genomics*. 267:506-14.
- Nogueira, E.M., Vinagre, F., Masuda, H.P., Vargas, C., de Pádua, V.L.M., da Silva, F.R., dos Santos, R.V., Baldani, J.I., Ferreira, P.C.G., Hemerly, A.S. (2001). Expression of sugarcane genes induced by inoculation with *Gluconacetobacter diazotrophicus* and *Herbaspirillum rubrisubalbicans*. *Genet. Mol. Biol.* 24:199-206.
- Nogueira, F.T.S., De Rosa, V.E., Menossi, M., Ulian, E.C., Arruda, P. (2003). RNA expression profiles and data mining of sugarcane response to low temperature. *Plant Physiol.* 132:1811-1824.
- Nonomura, K., Miyoshi, K., Eiguchi, M., Suzuki, T., Miyao, A., Hirochika, H., Kurata, N. (2003). The MSP1 gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice. *Plant Cell*. 15:1728-1739.
- Obayashi, T., Okegawa, T., Sasaki-Sekimoto, Y., Shimada, H., Masuda, T., Asamizu, E., Nakamura, Y., Shibata, D., Tabata, S., Takamiya, K., Ohta, H. (2004). Distinctive features of plant organs characterized by global analysis of gene expression in *Arabidopsis*. *DNA Res.* 11:11-25.
- O'Mahony, P.J., Oliver, M.J. (1999). Characterization of a desiccation-responsive small GTP-binding protein (Rab2) from the desiccation-tolerant grass *Sporobolus stapfianus*. *Plant Mol Biol.* 39:809-821.
- Oono, Y., Seki, M., Nanjo, T., Narusaka, M., Fujita, M., Satoh, R., Satou, M., Sakurai, T., Ishida, J., Akiyama, K., Iida, K., Maruyama, K., Satoh, S., Yamaguchi-Shinozaki, K., Shinozaki, K. (2003). Monitoring

expression profiles of *Arabidopsis* gene expression during rehydration process after dehydration using ca 7,000 full-length cDNA microarray. *Plant J.* 34:868-887.

Oono, Y., Seki, M., Satou, M., Iida, K., Akiyama, K., Sakurai, T., Fujita, M., Yamaguchi-Shinozaki, K., Shinozaki, K. (2006). Monitoring expression profiles of *Arabidopsis* genes during cold acclimation and deacclimation using DNA microarrays. *Funct Integr Genomics.* 4:1-23.

Osakabe, Y., Maruyama, K., Seki, M., Satou, M., Shinozaki, K., Yamaguchi-Shinozaki, K. (2005). Leucine-rich repeat receptor-like kinase1 is a key membrane-bound regulator of abscisic acid early signaling in *Arabidopsis*. *Plant Cell* 17:1105-1119.

Pagnussat, G.C., Fiol, D.F., Salerno, G.L. (2002) A CDPK type protein kinase is involved in rice SPS light modulation. *Physiol Plant.*, 115:183-189.

Pandey, S., Chen, J.G., Jones, A.M., Assmann, S.M. (2006). G-protein complex mutants are hypersensitive to abscisic acid regulation of germination and postgermination development. *Plant Physiol.* 141:243-256.

Papini-Terzi, F.S., Rocha, F.R., Vencio, R.Z., Oliveira, K.C., Felix, J.M., Vicentini, R., Rocha, C.S., Simoes, A.C., Ulian, E.C., di Mauro, S.M., da Silva, A.M., Pereira, C.A., Menossi, M., Souza, G.M. (2005). Transcription profiling of signal transduction-related genes in sugarcane tissues. *DNA Res.* 12:27-38.

Park, T., Yi, S.G., Kang, S.H., Lee, S., Lee, Y.S., Simon, R. (2003). Evaluation of normalization methods for microarray data. *BMC Bioinformatics.* 4:33.

Peirson, S.N., Butler, J.N., Foster, R.G. (2003). Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Res.* 31:e73.

Perez, G., De Prata, F., Chinea, A., Bernal, N., O'Reilly, J.P. (1997). Recursos genéticos de la caña de azúcar. Instituto Nacional de Investigaciones de La Caña de Azúcar (INICA), 249p.

Perfus-Barbeoch, L., Jones, A.M., Assmann, S.M. (2004). Plant heterotrimeric G protein function: insights from *Arabidopsis* and rice mutants. *Curr. Opin. Plant Biol.* 7:719-731.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29:2002-2007.

Porta, H., Rocha-Sosa, M. (2002). Plant lipoxygenases. Physiological and molecular features. *Plant Physiol.* 130, 15-21.

Poschenrieder, C., Tolra, R.P., Barcelo, J. (2005). A role for cyclic hydroxamates in aluminium resistance in maize? *J Inorg Biochem.* 99:1830-1836.

Quackenbush, J. (2001). Computational analysis of microarray data. *Nat. Rev. Genet.* 2:418-427.

Quackenbush, J. (2002). Microarray data normalization and transformation. *Nat. Rev. Genet.* 3:496-501.

Radutoiu, S., Madsen, L.H., Madsen, E.B., Felle, H.H., Umehara, Y., Gronlund, M., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., Stougaard, J. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* 425:585-592.

Raghothama, K.G. (2000). Phosphate transport and signaling. *Curr. Opin. Plant Biol.* 3:182-187.

Rambaut, A. (1996). Se-Al: Sequence Alignment Editor. Available at <http://evolve.zoo.ox.ac.uk/>.

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

Reis, V.M., Olivares, F.L., Döbereiner, J. (1994). Improved methodology for isolation of *Acetobacter diazotrophicus* and confirmation of its endophytic habitat. *World Journal of Microbiology and Biotechnology* 10:101-104.

Riera, M., Valon, C., Fenzi, F., Giraudat, J., Leung, J. (2005). The genetics of adaptive responses to drought stress: abscisic acid-dependent and abscisic acid-independent signalling components. *Physiologia Plantarum* 123:111-119.

Roach, B.T., Daniels, J. (1987). A review of the origin and improvement of sugarcane. International Sugarcane Breeding Workshop, pp 1-33, COPERSUCAR, Sao Paulo, Brasil.

Rockett, J.C., Hellmann, G.M. (2004). Confirming microarray data - is it really necessary? *Genomics* 83:541-549.

Roitsch, T. (1999). Source-sink regulation by sugar and stress. *Curr Opin Plant Biol.* 2:198-206.

Rolland, F., Moore, B., Sheen, J. (2002). Sugar sensing and signaling in plants. *Plant Cell* 14:S185-S205.

Romeis, T., Piedras, P., Jones, J.D. (2000). Resistance gene-dependent activation of a calcium-dependent protein kinase in the plant defense response. *Plant Cell.* 12:803-816.

Ruan, Y., Gilmore, J., Conner, T. (1998). Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1,400 genes using cDNA microarrays. *Plant J.* 15:821-833.

Rubio, V., Linhares, F., Solano, R., Martin, A.C., Iglesias, J., Leyva, A., Paz-Ares, J. (2001). A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. *Genes Dev.* 15:2122-2133.

Ruegger, M., Dewey, E., Gray, W.M., Hobbie, L., Turner, J., Estelle, M. (1998). The TIR1 protein of *Arabidopsis* functions in auxin response and is related to human SKP2 and yeast Grr1p. *Genes and Development* 12:198-207.

- Saal, L.H., Troein, C., Vallon-Christersson, Gruvberger, S., Borg, A., Peterson, C. (2002). BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data. *Genome Biology* 3:software0003.1-003.6.
- Sachetto-Martins, G., Franco, L.O., de Oliveira, D.E. (2000). Plant glycine-rich proteins: a family or just proteins with a common motif? *Biochim Biophys Acta*. 1492:1-14.
- Sadiqov, S.T., Akbulut, M., Ehmedov, V. (2002). Role of Ca^{2+} in drought stress signaling in wheat seedlings. *Biochemistry (Mosc)*. 67:491-497.
- Sakuma, Y., Maruyama, K., Osakabe, Y., Qin, F., Seki, M., Shinozaki, K., Yamaguchi-Shinozaki, K. (2006). Functional analysis of an *Arabidopsis* transcription factor, DREB2A, involved in drought-responsive gene expression. *Plant Cell*. 18:1292-1309.
- Sanders, D., Brownlee, C., Harper, J.F. (1999). Communicating with calcium. *Plant Cell*. 11:691-706.
- Sanders, D., Pelloux, J., Brownlee, C., Harper, J.F. (2002). Calcium at the crossroads of signaling. *Plant Cell*, 14:Suppl:S401-17.
- Sasaki, Y., Asamizu, E., Shibata, D., Nakamura, Y., Kaneko, T., Awai, K., Amagai, M., Kuwata, C., Tsugane, T., Masuda, T., Shimada, H., Takamiya, K., Ohta, H., Tabata, S. (2001). Monitoring of methyl jasmonate-responsive genes in *Arabidopsis* by cDNA macroarray: self-activation of jasmonic acid biosynthesis and crosstalk with other phytohormone signaling pathways. *DNA Res*. 8:153-161.
- Schenk, P.M., Kazan, K., Wilson, I., Anderson, J.P., Richmond, T., Somerville, S.C., Manners, J.M. (2000). Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl. Acad. Sci. USA* 97:11655-11660.
- Schmid, M., Uhlenhaut, N.H., Godard, F., Demar, M., Bressan, R., Weigel, D., Lohmann, J.U. (2003). Dissection of floral induction pathways using global expression analysis. *Development* 130:6001-6012.
- Schmittgen, T.D., Zakrajsek, B.A. (2000). Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. *J. Biochem. Biophys. Meth.* 46:69-81.
- Schuler, M.A., Werck-Reichhart, D. (2003). Functional genomics of P450s. *Annu Rev Plant Biol.* 54: 629-667.
- Schultz, J., Milpetz, F., Bork, P., Ponting, C.P. (1998) SMART, a simple modular architecture research tool: Identification of signaling domains *Proc. Natl. Acad. Sci. U.S.A.* 95 : 5857-5864.
- Seki, M., Ishida, J., Narusaka, M., Fujita, M., Nanjo, T., Umezawa, T., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y., Shinozaki, K. (2002). Monitoring the expression pattern of around 7,000 *Arabidopsis* genes under ABA treatments using a full-length cDNA microarray. *Funct. Integr. Genomics* 2:282-291.
- Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y., Shinozaki, K. (2001). Monitoring the expression pattern of 1,300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell*. 13:61-72.
- Selman-Housein, G., Lopez, M., Hernandez, D., Civardi, L., Miranda, F., Rigau, J., Puigdomenech, P. (1999). Molecular cloning of cDNAs coding for three sugarcane enzymes involved in lignification. *Plant Sci.* 143:163-172.
- Seo, H.S., Song, J.T., Cheong, J.J., Lee, Y.H., Lee, Y.W., Hwang, I., Lee, J.S., Choi, Y.D. (2001). Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proc Natl Acad Sci U S A*. 98:4788-4793.
- Shi, J., Kim, K.N., Ritz, O., Albrecht, V., Gupta, R., Harter, K., Luan, S., Kudla, J. (1999). Novel protein kinases associated with calcineurin B-like calcium sensors in *Arabidopsis*. *Plant Cell*. 11:2393-2405.
- Shiu, S.H., Bleecker, A.B. (2001a). Plant receptor-like kinase gene family: diversity, function, and signaling. *Sci STKE*. 113:RE22.
- Shiu, S.H., Bleecker, A.B. (2001b). Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. U.S.A.* 98:10763-10768.
- Shiu, S.H., Bleecker, A.B. (2003). Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. *Plant Physiol.* 132:530-543.
- Shiu, S.H., Karlowski, W.M., Pan, R., Tzeng, Y.H., Mayer, K.F., Li, W.H. (2004). Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell*. 16:1220-1234.
- Smeekens, S. (2000). Sugar-induced signal transduction in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 51:49-81.
- Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X., Zhu, L.H., Fauquet, C., Ronald, P. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270:1804-1806.
- Sonnhammer, E.L.L., Eddy, S.R., Birney, E., Bateman, A., Durbin, R. (1998). Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res.* 26:320-322.
- Souza, G.M., Simoes, A.C.Q., Oliveira, K.C., Garay, H.M., Fiorini, L.C., Gomes, F.S., Nishiyama-Junior, M.Y., da Silva, A.M. (2001). The sugarcane signal transduction (SUCAST) catalogue: prospecting signal transduction in sugarcane. *Genet. Mol. Biol.* 24:25-34.

- Speed, T.P., Yang, Y.H. (2002). Direct versus indirect designs for cDNA microarray experiments. *Sankhyā: The Indian Journal of Statistics* 64:707-721.
- Stintzi, A., Weber, H., Reymond, P., Browse, J., Farmer, E.E. (2001). Plant defense in the absence of jasmonic acid: the role of cyclopentenones. *Proc. Natl. Acad. Sci. USA* 98:12837-12842.
- Swiderski, M.R., Innes, R.W. (2001). The *Arabidopsis* PBS1 resistance gene encodes a member of a novel protein kinase subfamily. *Plant J.* 26:101-112.
- Swofford, D.L. (1999). PAUP*. Phylogenetic Analysis using Parsimony (* and other methods). Sinauer Associates, Sunderland, Massachusetts, USA.
- Tahtiharju, S., Palva, T. (2001). Antisense inhibition of protein phosphatase 2C accelerates cold acclimation in *Arabidopsis thaliana*. *Plant J.* 26:461-470.
- Takatsui, H. (1998). Zinc-finger transcription factors in plants. *Cell Mol Life Sci.* 54:582-596.
- Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S., Golub, T.R. (1999). Interpreting patterns of gene expression with self-organizing maps. *Proc. Natl. Acad. Sci. USA* 96:2907-2912.
- Thomas, C.M., Dixon, M.S., Parniske, M., Golstein, C., Jones, J.D. (1998). Genetic and molecular analysis of tomato Cf genes for resistance to *Cladosporium fulvum*. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 353:1413-1424.
- Thompson, J.D., Higgins, D.G., Gibson, T.G. (1994). 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 : 4673-4680.
- Todd, C.D., Zeng, P., Huete, A.M., Hoyos, M.E., Polacco, J.C. (2004). Transcripts of MYB-like genes respond to phosphorous and nitrogen deprivation in *Arabidopsis*. *Planta.* 219:1003-1009.
- Torii, K.U. (2000). Receptor kinase activation and signal transduction in plants: An emerging picture. *Curr. Opin. Plant Biol.* 3:361-367.
- Torii, K.U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R.F., Komeda, Y. (1996). The *Arabidopsis* ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *Plant Cell* 8:735-746.
- Trewavas, A. Signal perception and transduction. In: Buchanan, B., Gruissem, W. and Jones, R. Biochemistry & Molecular Biology of Plants. Rockyville, Maryland : American Society of Plant Physiologists, 2000. p.938.
- Turner, J.G., Ellis, C., Devoto, A. (2002). The jasmonate signal pathway. *Plant Cell* 15:53-64.
- Uchida, S., Nishida, Y., Satou, K., Muta, S., Tashiro, K., Kuhara, S. (2005). Detection and normalization of biases present in spotted cDNA microarray data: a composite method addressing dye, intensity-dependent, spatially-dependent, and print-order biases. *DNA Res.* 12:1-7.
- Umezawa, T., Yoshida, R., Maruyama, K., Yamaguchi-Shinozaki, K. and Shinozaki, K. (2004). SRK2C, a SNF1-related protein kinase 2, improves drought tolerance by controlling stress-responsive gene expression in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A.* 101:17306-17311.
- Urao, T., Yamaguchi-Shinozaki, K., Shinozaki, K. (2000). Two-component systems in plant signal transduction. *Trend in Plant Science,* 5:67-74.
- Urao, T., Yamaguchi-Shinozaki, K., Shinozaki, K. (2001). Plant histidine kinases: an emerging picture of two-component signal transduction in hormone and environmental responses. *Sci STKE.* ;2001:RE18.
- Urquiaga, S., Cruz, K.H.S., Boddey, R.M. (1992). Contribution of nitrogen fixation to sugar cane: Nitrogen-15 and nitrogen-balance estimates. *Soil Science Society of America Journal* 56:105-114.
- van Nocker, S., Vierstra, R.D. (1993). Two cDNAs from *Arabidopsis thaliana* encode putative RNA binding proteins containing glycine-rich domains. *Plant Mol. Biol.* 21:695-699.
- Vandesompele, J., Preter, K.D., Pattyn, F., Poppe, B., Roy, N.V., Paepe, A.D., Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3: research 0034.
- Vargas, C., de Pádua, V., Nogueira, E.M., Vinagre, F., Masuda, H.P., da Silva, F.R., Baldani, J.I., Ferreira, P.C.G., Hemerly, A.S. (2003). Signaling pathways mediating the association between sugarcane and endophytic diazotrophic bacteria: a genomic approach. *Symbiosis* 35:159-180.
- Vencio, R.Z.N., Koide, T. (2005). HTself: self-self based statistical test for low-replication microarray studies *DNA Res.* 12:211-214.
- Vernoud, V., Horton, A.C., Yang, Z., Nielsen, E. (2003). Analysis of the small GTPase gene superfamily of *Arabidopsis*. *Plant Physiol.* 131:1191-208.
- Vettore, A.L., da Silva, F.R., Kemper, E.L., Arruda, P. (2001). The libraries that made SUCEST. *Genet. Mol. Biol.* 24, 1-7.
- Vettore, A.L., da Silva, F.R., Kemper, E.L., Souza, G.M., da Silva, A.M., Ferro, M.I., Henrique-Silva, F., Gigliotti, E.A., Lemos, M.V., Coutinho, L.L., Nóbrega, M.P., Carrer, H., Franca, S.C., Bacci Junior 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 Araújo, P.G., Sculaccio, S.A., Angella, A., Lima, M.M., de Rosa Junior, V.E., Siviero, F., Coscrato, V.E., Machado, M.A., Grivet, L., Di Mauro, S.M., Nóbrega, F.G., Menck, C.F., Braga, M.D., Telles, G.P., Cara, F.A., Pedrosa, G., Meidanis, J., Arruda, P. (2003). Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Res.* 13:2725-2735.
- Vierstra, R.D. (2003). The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. *Trends Plant Sci.*, 8, 135-142.
- Vinagre, F., Vargas, C., Schwarcz, K., Cavalcante, J., Nogueira, E.M., Baldani, J.I., Ferreira, P.C., Hemery, A.S. (2006). SHR5: a novel plant receptor kinase involved in plant-N₂-fixing endophytic bacteria association. *J Exp Bot.* 57:559-569.
- Walia, H., Wilson, C., Wahid, A., Condamine, P., Cui, X., Close, T.J. (2006). Expression analysis of barley (*Hordeum vulgare* L.) during salinity stress. *Funct. Integr. Genomics.* 6:143-156.
- Walker, J. C. (1994). Structure and function of the receptor-like protein kinases of higher plants. *Plant Mol. Biol.* 26:1599-1609.
- Wang, R., Guegler, K., Labrie, S.T., Crawford, N.M. (2000). Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. *Plant Cell* 12:1491-1510.
- Wang, Y.H., Garvin, D.F., Kochian, L.V. (2002). Rapid induction of regulatory and transporter genes in response to phosphorus, potassium, and iron deficiencies in tomato roots. Evidence for cross talk and root/rhizosphere-mediated signals. *Plant Physiol.* 130:1361-1370.
- Wasternack, C., Parthier, B. (1997). Jasmonate-signalled plant gene expression. *Trends Plant Sci.* 2:302-307.
- Weaver, C.D., Roberts, D.M. (1992). Determination of the site of phosphorylation of nodulin 26 by the calcium-dependent protein kinase from soybean nodules. *Biochemistry.* 31:8954-8959.
- Webb, K.J., Skot, L., Nicholson, M.N., Jorgensen, B., Mizen, S. (2000). *Mesorhizobium loti* increases root-specific expression of a calcium-binding protein homologue identified by promoter tagging in *Lotus japonicus*. *Mol Plant Microbe Interact.* 13:606-616.
- Weisshaar, B., Jenkins, G.I. (1998). Phenylpropanoid biosynthesis and its regulation. *Curr Opin Plant Biol.* 1:251-257.
- White, P.J. (2001). A Nod and a wave: calcium signals during nodulation. *Trends Plant Sci.* 6:141.
- Woodward, A.W., Bartel, B. (2005). Auxin: regulation, action, and interaction. *Ann. Bot.* 95:707-735.
- Xie, D.X., Feys, B.F., James, S., Nieto-Rostro, M., Turner, J.G. (1998). COI1: an *Arabidopsis* gene required for jasmonate-regulated defense and fertility. *Science* 280: 1091-1094.
- Xiong, L., Schumaker, K.S., Zhu, J.K. (2002). Cell signaling during cold, drought, and salt stress. *Plant Cell.* 14(Supplement):s165-s183.
- Yamada, S., Komori, T., Hashimoto, A., Kuwata, S., Imaseki, H., Kubo, T. (2000). Differential expression of plastidic aldolase genes in *Nicotiana* plants under salt stress. *Plant Sci.* 154, 61-69.
- Yamaguchi-Shinozaki, K., Shinozaki, K. (2006). Transcriptional Regulatory Networks in Cellular Responses and Tolerance to Dehydration and Cold Stresses. *Annu. Rev. Plant Biol.* 57:781-803.
- Yang, I.V., Chen, E., Hasseman, J.P., Liang, W., Frank, B.C., Wang, S., Sharov V., Saeed A.I., White, J., Li, J., Lee, N.H., Yeatman, T.J., Quackenbush, J. (2002a). Within the fold: assessing differential expression measures and reproducibility in microarray assays. *Genome Biol.* 3:research0062.
- Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., Speed, T.P. (2002a). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.*, 30, e15.
- Yoshida, R., Hobo, T., Ichimura, K., Mizoguchi, T., Takahashi, F., Aronso, J., Ecker, J.R., Shinozaki, K. (2002). ABA-activated SnRK2 protein kinase is required for dehydration stress signaling in *Arabidopsis*. *Plant Cell Physiol.* 43:1473-1483.
- Zhang, B., Pan, X., Cobb, G.P., Anderson, T.A. (2006). Plant microRNA: a small regulatory molecule with big impact. *Dev. Biol.* 289:3-16.
- Zhang, X.Q., Lund, A.A., Sarath, G., Cerny, R.L., Roberts, D.M., Chollet, R. (1999). Soybean nodule sucrose synthase (nodulin-100): further analysis of its phosphorylation using recombinant and authentic root-nodule enzymes. *Arch. Biochem. Biophys.* 371:70-82.
- Zhao, D.Z., Wang, G.F., Speal, B., Ma, H. (2002). The excess microsporocytes1 gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the *Arabidopsis* anther. *Genes Dev.* 16:2021-2031.
- Zheng, Z., Murai, N. (1997). A distal promoter regions of the rice seed storage protein glutelin gene enhanced quantitative gene expression. *Plant Sci.* 128:59-65.
- Zheng, Z.L., Nafisi, M., Tam, A., Li, H., Crowell, D.N., Chary, S.N., Schroeder, J.I., Shen, J., Yang, Z. (2002). Plasma membrane-associated ROP10 small GTPase is a specific negative regulator of abscisic acid responses in *Arabidopsis*. *Plant Cell* 14:2787-2797.

- Zhong, H., Simons, J.W. (1999). Direct comparison of GAPDH, b-actin, cyclophilin, and 28S rRNA as internal standards for quantifying RNA levels under hypoxia. *Biochem. Biophys. Res. Commun.* 259:523-26.
- Zhou, J., Loh, Y.T., Bressan, R.A., Martin, G.B. (1995). The tomato gene Pti1 encodes a serine/threonine kinase that is phosphorylated by Pto and is involved in the hypersensitive response. *Cell* 83:925-935.
- Zrenner, R., Salanoubat, M., Willmitzer, L., Sonnewald, U. (1995). Evidence of the crucial role of sucrose synthase for sink strength using transgenic potato plants (*Solanum tuberosum* L.). *Plant J.* 7, 97-107.