**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA**

Programa de Pós-Graduação em Ciências Biológicas (Bioquímica)

SÁVIO DE SIQUEIRA FERREIRA

**Estudo do transcriptoma de genótipos
ancestrais de cana-de-açúcar com enfoque
em genes do metabolismo de parede celular**

Versão Corrigida

São Paulo

Data do Depósito na SPG:

29/10/2013

SÁVIO DE SIQUEIRA FERREIRA

**Estudo do transcriptoma de genótipos
ancestrais de cana-de-açúcar com enfoque
em genes do metabolismo de parede celular**

Tese apresentada ao Instituto de Química da
Universidade de São Paulo para obtenção do
Título de Doutor em Ciências (Bioquímica)

Orientadora: Profa. Dra. Glaucia Mendes Souza

São Paulo

2013

Ficha Catalográfica
Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

Ferreira, Sávio de Siqueira
F383es Estudo do transcriptoma de genótipos ancestrais de cana-de-
açúcar com enfoque em genes do metabolismo de parede celular /
Sávio de Siqueira Ferreira. -- São Paulo, 2013.
253p.

Tese (doutorado) - Instituto de Química da Universidade de
São Paulo. Departamento de Bioquímica.
Orientador: Souza, Glaucia Mendes

1. Expressão gênica 2. Saccharum : Botânica 2. Paredes celulares :
Botânica I. T. II Souza, Glaucia Mendes, orientador.

574.88 CDD

SAVIO DE SIQUEIRA FERREIRA

Estudo do transcriptoma de genótipos ancestrais de cana-de-açúcar com enfoque
em genes do metabolismo de parede celular

Tese apresentada ao Instituto de Química da
Universidade de São Paulo para obtenção do
Título de Doutor em Ciências (Bioquímica)

Aprovado em: _____

Banca Examinadora

Prof(a).Dr(a). GlauCIA Mendes Souza (Orientadora e Presidente)

Instituição: IQ – USP

Assinatura: _____

Prof(a).Dr(a). Aline Maria da Silva

Instituição: IQ – USP

Assinatura: _____

Prof(a).Dr(a). João Carlos Setubal

Instituição: IQ – USP

Assinatura: _____

Prof(a).Dr(a). Antonio Vargas de Oliveira Figueira

Instituição: CENA – USP

Assinatura: _____

Prof(a).Dr(a). Marcelo Menossi Teixeira

Instituição: IB – UNICAMP

Assinatura: _____

Aos meus pais, Maria Augusta e Jaci,

Dedico

AGRADECIMENTOS

A Deus, por sempre me guiar e me inspirar nos momentos difíceis.

Aos meus pais, Maria Augusta e Jaci, meus irmãos, Sílvio e Daniel, e minha avó, Matilde Lima, pelo apoio e confiança no meu potencial.

À minha namorada, Vanessa, pela compreensão, ajuda no desenvolvimento deste trabalho e por estar ao meu lado em todos os momentos, sempre me apoiando.

À professora Glaucia Mendes Souza, pela oportunidade, orientação e apoio, principalmente nessa reta final.

Aos colaboradores do projeto, Dr. Marcos Silveira Buckeridge do IB-USP, Dr. Marcelo Ehlers Loureiro do DBV-UFV, Dra. Monalisa Sampaio Carneiro do CCA-UFSCAR e Dr. Márcio Henrique Pereira Barbosa do DFT-UFV.

Aos amigos Daniel Bonoto, Marcos de Oliveira, Maycon Campos, Guilherme Pinheiro e Rodrigo Siqueira, pela amizade de longa data e que mesmo de longe sempre tiveram participação fundamental na minha vida pessoal e profissional.

Aos amigos da velha guarda do laboratório, Carolina Lembke, Érica Bandeira, Alessandra Vanessa, Edwin Delgado; aos que já passaram pelo laboratório, Maximiller Costa, Paloma Sato, Rodrigo Fandiño, Roberta Alvares, Sideny Nunes, Fábio Vicente; a nova geração, Armando Casas, Danielle Silva, Pablo Andrade, Izabel Moraes, Caio Henrique; aos vizinhos de laboratório Luíza Dantas e Dr. Carlos Hotta; e em especial ao amigo Milton Yutaka, pela imensa ajuda no desenvolvimento deste trabalho, em especial pelas análises de sequenciamento; agradeço a todos pela amizade, pelas discussões científicas e papos nas horas de descontração. Obrigado por fazerem do trabalho uma tarefa mais leve e agradável.

Às amigas do Instituto de Química, Ale Vaz e Mariana Massafra, e do CATG, Cláudia Siqueira e Layla Martins.

Ao Sandro Ferrarez, do PMGCA, pela ajuda nas coletas de campo.

À Débora Leite e Viviane Poelking, pela ajuda nos experimentos de anatomia e na interpretação dos resultados.

Ao Germano Siqueira, pelos ensaios preliminares de sacarificação.

Ao CNPq pela bolsa concedida e à FAPESP pelo apoio financeiro nos projetos.

A todos os funcionários de Instituto de Química e todas as pessoas que de alguma forma contribuíram para esse trabalho.

RESUMO

Ferreira, S.S. **Estudo do transcriptoma de genótipos ancestrais de cana-de-açúcar com enfoque em genes do metabolismo de parede celular**. 2013. (253p). Tese - Programa de Pós-Graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

A cana-de-açúcar é uma gramínea C4 cuja principal característica é o acúmulo de altas concentrações de sacarose no colmo. As variedades comerciais são derivadas de cruzamentos de *S. officinarum*, que possui colmos grossos com alto teor de açúcar, e *S. spontaneum*, que apresenta colmos finos, fibrosos, com baixo teor de açúcar, mas com maior perfilhamento e resistência a estresses. O desenvolvimento da tecnologia do bioetanol celulósico, no qual o bagaço é utilizado para produção de bioetanol, tem gerado o interesse no desenvolvimento da cana energia, que possui mais fibra e biomassa, porém menos sacarose. Para isso, é importante o entendimento das vias regulatórias de biossíntese de parede celular, principal constituinte da biomassa vegetal. Ainda, as variedades comerciais possuem uma base genética estreita, que pode ser um dos fatores responsáveis pela diminuição dos ganhos de produtividade que vem sendo observada a cada variedade lançada. Isso mostra a importância de se voltar aos genótipos ancestrais para identificar genes que possam ser importantes para o melhoramento da cana-de-açúcar, incluindo o desenvolvimento da cana energia. Assim, o presente trabalho analisou o transcriptoma de três genótipos ancestrais e dois híbridos comerciais, juntamente com a caracterização de dados morfo-fisiológicos. Essa abordagem permitiu ligar padrões de expressão gênica aos fenótipos e identificar genes e vias regulatórias que estão possivelmente relacionadas aos mecanismos que determinam características como produtividade, acúmulo de sacarose e lignina, principalmente genes de regulação da estrutura da cromatina, metabolismo de parede celular e algumas famílias de fatores de transcrição. A expressão desses genes

apresentou maior correlação com acúmulo de sacarose do que a expressão dos próprios genes do metabolismo de sacarose. *S. spontaneum* foi o genótipo que apresentou perfil de expressão e fenótipos mais distintos, mostrando um potencial biotecnológico para produção de biomassa. A integração dos resultados sugere que o acúmulo de sacarose compete com o de lignina e que o controle do crescimento em entrenós mais jovens pode ser importante para a maior produtividade. Também foram construídas bibliotecas de cDNA *full length* e estas apresentaram 90% de insertos completos. O sequenciamento dessas bibliotecas revelou uma grande quantidade de transcritos novos e específicos do gênero *Saccharum*, juntamente com um número significativo de transcritos antissense e milhares de genes diferencialmente expressos que podem estar associados, principalmente, às diferenças de resistência a estresses de *S. officinarum* e *S. spontaneum*. Essas bibliotecas fornecem uma grande plataforma para estudos posteriores para descoberta de polimorfismos, *splicing* alternativo e alelos que possam ser importantes nas características dos genótipos. Ainda, foram identificadas sequências promotoras de genes do metabolismo de parede celular e de alguns candidatos a estarem envolvidos nesse processo; vários apresentaram sítios de ligação conservados responsivos a fatores de transcrição da biossíntese de parede celular, mas algumas diferenças importantes foram identificadas, evidenciando divergências evolutivas entre cana-de-açúcar e outras espécies. Nossos dados mostram-se importantes por avançar no entendimento das vias envolvidas no metabolismo de parede celular em cana-de-açúcar assim como por caracterizar parte da variabilidade dos genótipos ancestrais.

Palavras-chave: Cana-de-açúcar, *Saccharum officinarum*, *Saccharum spontaneum*, transcriptoma, parede celular, lignina.

ABSTRACT

Ferreira, S.S. **Transcriptomic study of sugarcane ancestor genotypes focusing on cell wall metabolism-related genes**. 2013. (253p). PhD Thesis - Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo

Sugarcane is C₄ grass whose main characteristic is the accumulation of high concentrations of sucrose in its stem. Commercial varieties are derived from crosses between *S. officinarum*, which has thick culms with high sugar content, and *S. spontaneum*, which has thin fibrous culms, with low sugar content, but presents high tillering and stress tolerance. The development of cellulosic bioethanol technology, on which the bagasse is used to produce bioethanol, has raised the interest to produce the energy cane, which has high fiber content and biomass yield instead of sucrose. For this purpose, it is important to obtain knowledge on the regulatory pathways of the biosynthesis of the cell wall, the main component of plant biomass. Still, commercial varieties have a narrow genetic basis and this could be one of the reasons for the reduction on yield gains that has been noticed for new varieties released. This shows the importance to return to ancestor genotypes to identify genes that might be important for sugarcane improvement, as well as for energy cane development. The present work analyzed the transcriptome of three ancestor genotypes and two commercial hybrids, together with the characterization of morphophysiological traits. This approach allowed us to link gene expression pattern and phenotypes to identify genes and regulatory pathways that might be related to the mechanisms determining traits, such as yield, sugar and lignin content, especially genes related to chromatin structure regulation, cell wall metabolism and some transcription factors families. The expression of those genes showed a greater correlation to sucrose content than expression of sucrose

metabolism genes itself. *S. spontaneum* is the genotype with the most distinct expression profile and phenotype, showing a biotechnological potential for biomass production. The integration of the results suggests that sucrose accumulation competes with lignin and the control of growth in younger internodes might be important in determining higher yield. In addition, full length cDNA library has been constructed and they showed 90% full length inserts confirmed. High-throughput sequencing revealed a large number of new and specific *Saccharum* transcripts, as well as a significant number of antisense transcripts and thousands of differentially expressed genes that might be related to, especially, the differences of stress tolerance of *S. officinarum* and *S. spontaneum*. These libraries provide a large platform for future studies to discover polymorphisms, splicing variants and alleles that might be important to genotypes' traits. We also identify promoter sequences from cell wall-related genes and some candidates to be involved in this process; several of them showed conserved binding sites responsive to cell wall-related transcription factors, however, some differences were also identified, showing an evolutionary divergence among sugarcane and other species. Our data are important for advancing knowledge on cell wall-related pathways in sugarcane as well as for the characterization of genetic variability of ancestor genotypes.

Keywords: Sugarcane, *Saccharum officinarum*, *Saccharum spontaneum*, transcriptome, cell wall, lignin.

Lista de Figuras

Figura 1. Esquema dos diferentes tipos de parede celular primária..	32
Figura 2. Via de fenilpropanoides.....	34
Figura 3. Modelo estrutural da parede celular de cana-de-açúcar.....	36
Figura 4. Modelo da regulação transcricional da biossíntese de parede celular secundária. .	38
Figura 5. Fotossíntese C4 (A) e C3 (B).....	44
Figura 6. Esquema de acúmulo de sacarose nas células do colmo de cana-de-açúcar..	47
Figura 7. Genótipos em estudo.	79
Figura 8. Dados agronômicos.	81
Figura 9. Dados fotossintéticos.	83
Figura 10. Quantificação de açúcares solúveis.....	85
Figura 11. Quantificação de lignina.	86
Figura 12. Histoquímica para detecção de lignina (em vermelho) em secções transversais do colmo, com reagente floroglucinol.....	89
Figura 13. Histoquímica para detecção de siringil (em vermelho) em secções transversais do colmo, com reagente de Maüle.....	92
Figura 14. Categorias funcionais dos 2003 SAS diferencialmente expressos (Apêndice 1) nas 9 hibridizações (folha+1, entrenó imaturo e intermediário) dos 3 genótipos ancestrais contra a variedade comercial RB867515.....	95
Figura 15. Validação por qPCR na terceira réplica biológica de folha dos genes diferencialmente expressos identificados por microarranjo..	97
Figura 16. Validação por qPCR na terceira réplica biológica de entrenó imaturo dos genes diferencialmente expressos identificados por microarranjo..	98

Figura 17. Validação por qPCR na terceira réplica biológica de entrenó intermediário dos genes diferencialmente expressos identificados por microarranjo.....	99
Figura 18. Diagrama de Venn dos 1018 SAS diferencialmente expressos (Apêndice 1) nas 9 hibridizações (folha+1, entrenó imaturo e entrenó intermediário) dos 3 genótipos ancestrais contra a variedade comercial RB867515.	100
Figura 19. Diagrama de Venn dos SAS diferencialmente expressos em (A) quatro experimentos (2 de folha e 2 de entrenó imaturo) de alto Brix contra baixo Brix (ABxBB) de genótipos da RIDESA e dos dois genótipos ancestrais de baixo Brix (<i>S. spontaneum</i> e <i>S. robustum</i>) contra a referência RB867515 (de alto Brix) e (B) apenas entre os dois genótipos ancestrais de baixo Brix (<i>S. spontaneum</i> e <i>S. robustum</i>) contra a referência de alto Brix RB867515..	102
Figura 20. Categorias funcionais dos 807 SAS diferencialmente expressos (Apêndice 2) nas 3 hibridizações de entrenó imaturo (referência) vs entrenó intermediário de RB867515, <i>S. officinarum</i> e <i>S. spontaneum</i>	110
Figura 21. Clusterização dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó imaturo.....	116
Figura 22. Clusterização dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó intermediário.	117
Figura 23. Perfil de expressão do transcrito senso pela análise de intensidade (gráfico de linhas) e por qPCR (gráfico de barras) em entrenó imaturo e intermediário.	118
Figura 24. Cluster hierárquico dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó imaturo.	120

Figura 25. Validação da expressão do antissenso dos genes COMT e kobito em entrenó imaturo, mostrando as análises de intensidade (gráfico de linhas) e qPCR (gráfico de barras).....	120
Figura 26. Cluster hierárquico dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó imaturo.	121
Figura 27. Cluster hierárquico dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó intermediário.....	122
Figura 28. Cluster hierárquico dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó intermediário.....	123
Figura 29. Análise da expressão gênica de <i>ScNAC86</i> (homólogo de <i>NST1</i>) com normalização para comparação entre os diferentes entrenós de <i>S. spontaneum</i>	131
Figura 30. Análise da expressão gênica das celulose sintase <i>ScCesA3</i> (barras brancas) e <i>ScCesA7</i> (barras pretas).	135
Figura 31. Rede de genes coexpressos com o gene 4CL de milho.	141
Figura 32. Cluster hierárquico dos dados de sinal de expressão de entrenó imaturo dos genes da lista de coexpressos.	144
Figura 33. Cluster hierárquico dos dados de sinal de expressão de entrenó intermediário dos genes da lista de coexpressos.....	145
Figura 34. Alinhamento entre sequência do clone da biblioteca FL e o transcrito homólogo de sorgo.	151
Figura 35. Porcentagem de genes de cada uma das 7 espécie de gramíneas que possui <i>hit</i> contra os <i>reads</i> (A) e <i>contigs</i> (B) do FL-RNAseq (<i>cutoff</i> e-20).	153
Figura 36. Porcentagem de contigs do FL-RNAseq que possui <i>hit</i> contra cada uma das 7 espécies de gramíneas (<i>cutoff</i> e-20).....	154

Figura 37. Diagrama de Venn comparando o número de sequências do SUCEST, chip Agilent e <i>contigs</i> do FL-RNAseq	154
Figura 38. Porcentagem de cobertura dos <i>contigs</i> nos genes de cada uma das 7 espécies de gramíneas (<i>cutoff</i> e-20).	155
Figura 39. Número de <i>contigs</i> que alinham em cada gene das 7 espécies de gramíneas (<i>cutoff</i> e-20), considerando todos os <i>hits</i>	156
Figura 40. Principais funções e categorias presentes nos <i>contigs</i> das cinco amostras.	160
Figura 41. Genes diferencialmente expressos pela análise de FL-RNAseq.	163
Figura 42. Número de <i>reads</i> antissenso (NAT) em cada amostra por espécie.	167
Figura 43. Resumo das principais características apresentadas pelos genótipos analisados, mostrando relação com a expressão gênica e o fenótipo.....	227

Lista de Tabelas

Tabela 1. Exemplificação da nobilitação. O processo inicia com o cruzamento entre <i>S. officinarum</i> (SO) e <i>S. spontaneum</i> (SS), com número de cromossomos 80 e 64, respectivamente, dando origem à F ₁ com 112 cromossomos.	25
Tabela 2. Catálogo de genes de parede celular de cana-de-açúcar.	77
Tabela 3. Diâmetro (mm) e conteúdo de água (%) do entrenó maduro.	81
Tabela 4. Esquema das hibridizações.	93
Tabela 5. Categorias funcionais enriquecidas dos genes diferencialmente expressos nos experimentos dos ancestrais, considerando os 3 tecidos analisados.	96
Tabela 6. Comparação entre genes diferencialmente expressos em folha e entrenó imaturo em 4 experimentos (2 de folha e 2 de entrenó) de genótipos da RIDESA de alto Brix x baixo Brix (Sato 2012) e os diferencialmente expressos observados em 4 experimentos (2 de folha e 2 de entrenó imaturo) dos genótipos ancestrais de baixo Brix (<i>S. robustum</i> e <i>S. spontaneum</i> , Figura 8A) contra a referência de alto Brix, RB867515.	103
Tabela 7. Categorias funcionais enriquecidas dos genes diferencialmente expressos nos experimentos da RIDESA (Sato 2012) e dos ancestrais.	106
Tabela 8. Principais genes diferencialmente expressos em experimentos de folha, entrenó imaturo, intermediário e maduro de plantas de alto e baixo Brix derivadas de cruzamentos entre <i>S. officinarum</i> e <i>S. spontaneum</i> (Papini-Terzi <i>et al.</i> 2009) realizados no chip SUCAST.	107
Tabela 9. Esquema das hibridizações de entrenó imaturo vs entrenó intermediário.	109

Tabela 10. Categorias funcionais enriquecidas dos genes diferencialmente expressos nos experimentos de entrenó imaturo vs entrenó intermediário do mesmo genótipo (tabela 9 e figura 20).....	111
Tabela 11. Sondas com sinal acima do <i>background</i> em todos os experimentos.....	113
Tabela 12. Sondas do catálogo de genes de parede celular com sinal acima do <i>background</i> em todos os experimentos.	115
Tabela 13. Expressão gênica em entrenó imaturo.....	127
Tabela 14. Expressão gênica em entrenó intermediário.....	128
Tabela 15. Expressão gênica em entrenó maduro.	129
Tabela 16. Ortologia deduzida pelo programa Inparanoid.	130
Tabela 17. Expressão gênica em folha.	138
Tabela 18. Lista de genes de milho coexpressos com o gene 4CL e seus respectivos homólogos de cana-de-açúcar.	142
Tabela 19. Número de cis-elementos de parede celular nos promotores dos genes escolhidos pelos clusters hierárquicos (Figuras 24, 26-28).....	148
Tabela 20. Número de cis-elementos de parede celular nos promotores dos genes da lista de coexpressos (Tabela 18).	149
Tabela 21. Validação das bibliotecas de cDNA full length (FL).....	151
Tabela 22. Dados do sequenciamento do Ion PGM e 454 e montagem dos <i>contigs</i> com o software Trinity.....	153
Tabela 23. Número de contigs agrupado por amostra.	158
Tabela 24. Predição de proteínas nos contigs do RNAseq utilizando o programa ESTscan. .	161

Tabela 25. Enriquecimento de Categorias funcionais do Gene Ontology dos transcritos com expressão diferencial nos dados de FL-RNAseq, separadas pelas três principais categorias: Biological Process, Cellular Component e Molecular Function.	164
Tabela 26. Porcentagem de <i>reads</i> senso e NAT alinhados contra cada espécie e suas respectivas proporções em relação ao número total de <i>reads</i> sequenciados no PGM (29.260.184 <i>reads</i>).	167

Lista de Abreviaturas

4CL - 4-coumarato CoA ligase

C3H - p-coumarato 3-hidroxilase

C4H - cinamato-4-hidroxilase

CA - anidrase carbônica

CAD - álcool cinamil desidrogenase

CCR - cinamoil-CoA redutase

CesA - celulose sintase

COMT - ácido cafeico O-metiltransferase

Csl - celulose sintase-like

ESTs - *Expression sequence tags*

F5H - ferulato 5-hidroxilase

GAX – glucoronoarabinoxilanas

GO – *Gene Ontology*

Gs – condutância estomática

KNAT7 - KNOTTED ARABIDOPSIS THALIANA7

M46RE - MYB46-responsive cis-regulatory element

NAT - natural antisense transcripts

NST1/2 - NAC SECONDARY WALL THICKENING PROMOTING FACTOR1/2

ORF – open reading frame

PAL - fenilalanina amônia liase

PEP – fosfoenolpiruvato

PEPC - fosfoenolpiruvato carboxilase

PPDK – piruvato ortofosfato dicinase

RIDESA - Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético

Rubisco - ribulose-1,5-bisphosphate carboxylase oxygenase

SAS - *Sugarcane assembled sequence*

SMRE - *secondary wall MYB responsive element*

SNBE - *secondary wall NAC binding element*

SND1 /2/3- SECONDARY WALL ASSOCIATED NAC DOMAIN PROTEIN1/2/3

SPS - sacarose fosfato sintase

SS – sacarose sintase

SUCAMET - *Sugarcane Metabolism Catalogue*

SUCAST - *Sugarcane Signal Transduction Catalogue*

SUCEST - *SugarCane EST Genome Project*

SWNs - *secondary wall NACs*

TF – fator de transcrição

VND6/7 - VASCULAR-RELATED NAC-DOMAIN6 /7

WUE – eficiência do uso de água

XTH - xiloglucano endo-transglicosilase/hidrolase

SUMÁRIO

1. INTRODUÇÃO	23
1.1. Cana-de-açúcar: histórico e importância econômica	23
1.2. O etanol celulósico e a cana energia	26
1.3. A parede celular	29
1.4. Regulação transcricional da biossíntese de parede celular secundária	36
1.5. Fotossíntese, acúmulo de sacarose e biomassa	43
1.6. Perfil de expressão gênica em larga escala	48
1.6.1. Estudos de transcriptoma com enfoque em parede celular	51
2. OBJETIVOS	55
3. METODOLOGIA.....	57
3.1. Construção de um catálogo de genes relacionados a parede celular em cana-de-açúcar.....	57
3.2. Material vegetal	57
3.2.1. Plantas utilizadas nas análises de qPCR, Microarranjo, fisiologia e anatomia....	57
3.2.2. Plantas utilizadas para FL-RNAseq.....	59
3.3. Medida de Brix e análise de crescimento	60
3.4. Taxa de fotossíntese e transpiração	61
3.5. Conteúdo de água.....	61
3.6. Análise de açúcares solúveis.....	61
3.7. Quantificação de lignina	62

3.8.	Histoquímica	63
3.9.	Gráficos e análises estatísticas	63
3.10.	Extração de RNA.....	64
3.11.	Hibridização em chip de oligonucleotídeos da plataforma Agilent	65
3.12.	PCR em tempo real.....	66
3.13.	<i>Clusterização</i> dos dados de expressão.....	67
3.14.	Redes de coexpressão de genes.....	68
3.15.	Identificação da região promotora e cis-elementos regulatórios associados ao metabolismo de parede celular	69
3.16.	Construção de bibliotecas de cDNA e seqüenciamento (FL-RNAseq)	70
4.	RESULTADOS	75
4.1.	Construção de um catálogo de genes de parede celular para cana-de-açúcar	75
4.2.	Experimento de campo.....	78
4.3.	Dados Agronômicos	79
4.4.	Análise de parâmetros fotossintéticos	82
4.5.	Análise de carboidratos solúveis	83
4.6.	Quantificação de lignina	85
4.7.	Histoquímica	87
4.8.	Identificação de genes diferencialmente por microarranjo de oligonucleotídeos ...	93
4.8.1.	Análise do perfil de expressão diferencial entre os genótipos	93
4.8.2.	Genes diferencialmente expressos entre entrenó imaturo e intermediário ...	108

4.8.3.	Identificação de padrões de expressão dos genes com sinal de expressão acima do <i>background</i>	111
4.9.	Análise da expressão por qPCR.....	125
4.9.1.	Expressão de genes da via de biossíntese de fenilpropanoides	125
4.9.2.	Expressão de fatores de transcrição	130
4.9.3.	Genes relacionados ao metabolismo de parede celular	132
4.9.4.	Expressão de genes do metabolismo de sacarose e de geração de energia ...	135
4.9.5.	Expressão gênica em folha	136
4.10.	Construção de uma rede de coexpressão de genes de metabolismo de lignina	139
4.11.	Identificação de promotores de genes associados ao metabolismo de parede celular em cana-de-açúcar.....	146
4.12.	Construção de bibliotecas de cDNA <i>full length</i> e análise do transcriptoma de <i>S. officinarum</i> , <i>S. spontaneum</i> e do híbrido SP803280 por sequenciamento de RNA em larga escala (FL-RNAseq).....	149
5.	DISCUSSÃO	168
5.1.	Partição de carbono, acúmulo de sacarose e lignina são diferencialmente regulados entre os genótipos	168
5.2.	Análise do transcriptoma.....	173
5.2.1.	Comparação da expressão gênica entre genótipos por oligoarranjos e qPCR.	173
5.2.1.1.	Metabolismo energético e de sacarose	174
5.2.1.2.	Transportadores	176

5.2.1.3.	Parede celular.....	180
5.2.1.4.	Biossíntese de Hormônios.....	184
5.2.1.5.	Transdução de sinal.....	185
5.2.1.5.1.	Regulação transcricional	186
5.2.1.5.2.	Proteínas quinases e fosfatases	192
5.2.1.6.	Metabolismo de DNA e RNA	194
5.2.1.7.	Genes de captação de luz e fixação de carbono	196
5.2.2.	Comparação da expressão gênica entre entrenó imaturo e intermediário	198
5.2.3.	Comparação da expressão gênica entre os genótipos ancestrais e progênes segregantes para Brix.....	203
5.2.4.	Bibliotecas de cDNA <i>full length</i>	206
5.2.5.	Expressão de NATs.....	211
5.3.	Redes de coexpressão de genes associados a biossíntese de lignina	215
5.4.	Identificação de promotores associados a genes de parede celular.....	219
5.5.	Cana energia e potencial biotecnológico de <i>S. spontaneum</i>	222
6.	CONCLUSÃO	226
7.	REFERÊNCIAS	230
	LISTA DE APÊNDICES.....	248
	LISTA DE ANEXOS.....	249
	A - SÚMULA CURRICULAR	249

1. INTRODUÇÃO

1.1. Cana-de-açúcar: histórico e importância econômica

A cana-de-açúcar é um dos principais produtos do agronegócio brasileiro. As receitas geradas pela indústria sucroalcooleira ultrapassam 50 bilhões de dólares, sendo mais de 13 bilhões em exportações (Unica 2012). O Brasil é o líder mundial na produção de cana-de-açúcar e açúcar, sendo responsável por cerca de 22% da produção mundial de açúcar bruto (Faostat 2011). A safra 2012/2013 atingiu a marca de 588,9 milhões de toneladas de cana-de-açúcar, com média de 69,4 ton/ha (Conab 2013). Deste total, 292,9 milhões de toneladas foram destinados à produção de açúcar e 295,9 milhões de toneladas para a produção de etanol. O estado de São Paulo é o maior produtor do país, responsável por mais de 50% da produção nacional (330,7 milhões de toneladas na safra 2012/2013), muito à frente do segundo colocado, Goiás, com 52,7 milhões de toneladas (Conab 2013). Além da importância econômica, a cana-de-açúcar tem impacto significativo na matriz energética nacional, uma vez que seus derivados compreendem cerca de 18% da energia consumida no país (Unica 2012).

Pertencente ao gênero *Saccharum* da família *Poaceae*, a cana-de-açúcar é uma gramínea tropical que se destaca pela sua capacidade de acumular altas quantidades de sacarose nos colmos, podendo chegar a 42% de seu peso seco (Moore 1995). A cana-de-açúcar tem sua origem no sudeste Asiático e evidências indicam que a domesticação da cultura ocorreu na Nova Guiné com a seleção de *Saccharum officinarum* a partir da espécie selvagem *Saccharum robustum* (Paterson *et al.* 2013). As variedades modernas são derivadas de cruzamentos interespecíficos entre *S. officinarum* e a espécie selvagem *S. spontaneum*, com menores contribuições de outras espécies do mesmo gênero,

principalmente *S. robustum*, *S. sinense* e *S. barberi*, e também de outros gêneros, especialmente *Erianthus* e *Miscanthus* (Paterson *et al.* 2013). Conhecida como cana nobre, *S. officinarum* tem como principais características os colmos grossos ricos em sacarose e menor teor de fibras. Já *S. spontaneum* possui alta capacidade de perfilhamento, colmos mais finos, fibrosos, com pouco ou nenhum conteúdo de sacarose, porém, apresenta maior resistência a estresses bióticos e abióticos. *S. robustum* possui características semelhantes à *S. spontaneum*, mas, geralmente, com colmos mais grossos e conteúdo de sacarose ligeiramente superior.

Os primeiros programas de melhoramento genético datam de 1888 em Java e Barbados (Ming *et al.* 2006). Inicialmente, utilizavam-se apenas canas nobres (*S. officinarum*), que eram selecionadas pelo seu conteúdo de sacarose. Entretanto, os melhoristas identificaram que essas plantas eram muito suscetíveis a doenças, ataque de insetos e restritas a poucos habitats. Com isso, a cana nobre foi cruzada com a espécie selvagem *S. spontaneum*, visando à introdução de características de resistência e adaptabilidade, seguindo-se diversos retrocruzamentos com canas nobres, num processo denominado “nobilização” (Ming *et al.* 2006). A primeira etapa desse processo envolve a não redução do número de cromossomos do óvulo de *S. officinarum* e polinização interespecífica com *S. spontaneum* gerando uma progênie “ $2n+n$ ”. Considerando o número mais frequente de cromossomos entre os acessos de *S. officinarum* ($2n=80$) e *S. spontaneum* ($2n=64$) (Irvine 1999), a nobilização pode ser exemplificada como descrito na tabela 1. As progênies F_1 e BC_1 possuem o número somático não reduzido do parental feminino ($2n$) e o número gamético do parental masculino (n). Assim, a nobilização foi um processo que ajudou a aumentar a poliploidia e aneuploidia do genoma da cana-de-açúcar e, atualmente, as variedades comerciais possuem 100-130 cromossomos (Ming *et al.* 2006). Após o processo de

nobilização, os programas passaram a conduzir o melhoramento das canas nobilizadas produzindo as cultivares híbridas modernas. Devido à nobilitação, o genoma das variedades modernas, em sua maior parte, é constituído de cromossomos de *S. officinarum*. A análise da estrutura do genoma de variedades comerciais demonstrou que 70-80% dos cromossomos são derivados de *S. officinarum*, 10-20% de *S. spontaneum* e 10% de cromossomos recombinantes dessas duas espécies (D'Hont *et al.* 1996; D'Hont 2005; D'Hont *et al.* 2008).

Tabela 1. Exemplificação da nobilitação. O processo inicia com o cruzamento entre *S. officinarum* (SO) e *S. spontaneum* (SS), com número de cromossomos 80 e 64, respectivamente, dando origem à F_1 com 112 cromossomos. Seguem-se os retrocruzamentos (BC) das progênes com *S. officinarum*. O número de cromossomos está indicado entre parênteses. A coluna da direita indica a % do genoma da progênie derivado de *S. spontaneum* (S%). Note que nas progênes F_1 e BC_1 não ocorre a redução do número de cromossomos do parental feminino. A % de *S. spontaneum* diminui à medida que são feitos novos cruzamentos entre a progênie e *S. officinarum*.

Parental feminino		Parental masculino	Progênie (cromossomos)	% <i>S. spontaneum</i>
SO(80)	X	SS(64)	$F_1(80+32=112)$	S%=29
SO(80)	X	$F_1(112)$	$BC_1(80+56=136)$	S%=12
SO(80)	X	$BC_1(136)$	$BC_2(40+68=108)$	S%=7

Adaptado de (Paterson *et al.* 2013)

As variedades modernas possuem uma base genética muito similar, sendo essencialmente derivadas de 15-20 genótipos de cultivares “nobilizadas” (Roach 1989; Ming *et al.* 2006). Com isso, em 1965, iniciou-se uma tentativa de aumentar essa base genética estreita, utilizando-se clones diferentes daqueles usados inicialmente em Java e na Índia nas etapas de nobilitação ou cruzando-se canas selvagens com cultivares comerciais. Os programas de ampliação da base genética produziram diversos clones semicomerciais que foram introduzidos nos programas de melhoramento desde o final dos anos 80 (Paterson *et al.* 2013).

A inabilidade de rastrear, por seleção visual, um germoplasma incorporado em uma população de um programa de melhoramento foi, possivelmente, uma das principais causas do fracasso dos programas de ampliação da base genética (Paterson *et al.* 2013). Existe uma grande variabilidade genética entre os clones do gênero *Saccharum* (Tai & Miller 2002), mas faltam ferramentas para ajudar os melhoristas a incorporar genes de interesse de qualquer fonte em uma população de cultivares modernas. Os avanços recentes em biotecnologia estão começando a gerar informação e tecnologia que, certamente, irão auxiliar os programas de melhoramento a ampliar a base genética em suas populações para produzir novas cultivares no futuro (Paterson *et al.* 2013).

1.2. O etanol celulósico e a cana energia

Nas últimas décadas, a busca por fontes renováveis de energia tem despertado muito interesse. Com o crescimento da demanda de energia, a instabilidade dos preços de combustíveis fósseis e o aumento da preocupação com questões ambientais, principalmente a emissão de gases do efeito estufa e aumento de CO₂ atmosférico, os biocombustíveis apresentam-se como uma alternativa promissora (Hill 2006). No Brasil, a cana-de-açúcar tem sido utilizada com sucesso por décadas para a produção de bioetanol a partir da sacarose estocada no colmo, denominado etanol de primeira geração. Estudos sobre a cadeia produtiva demonstraram que o bioetanol de cana-de-açúcar gera uma diminuição significativa na emissão de gases do efeito estufa, chegando a 80% de redução em comparação à gasolina (Macedo *et al.* 2008; OECD 2008). O Brasil é o segundo maior produtor de bioetanol, com 23,6 bilhões de litros na safra 2012/2013 (Conab 2013), perdendo apenas para o bioetanol de milho dos EUA (Unica 2012), mas a eficiência energética da produção de bioetanol de cana-de-açúcar é de 5 a 8 vezes superior à de milho

(Goldemberg 2008). Além disso, o desenvolvimento de novas tecnologias está tornando possível a produção de etanol a partir de resíduos vegetais (por exemplo, bagaço da cana, palha de milho, madeira, etc), também conhecido como etanol celulósico, de segunda geração ou biocombustível avançado. Nessa tecnologia, a biomassa presente no bagaço é hidrolisada liberando carboidratos que serão fermentados para produzir bioetanol (Amorim *et al.* 2011). Potencialmente, a utilização do bagaço pode aumentar em 40-50% a produção atual de bioetanol de cana-de-açúcar (Waclawovsky *et al.* 2010).

Uma vez que a biomassa vegetal é constituída basicamente de parede celular, estratégias visando o aumento da deposição de parede celular poderão aumentar a densidade da biomassa produzida (Yang *et al.* 2013) reduzindo os custos de transporte, os quais contribuem significativamente para o preço da biomassa destinada às biorrefinarias (Kumar *et al.* 2005; Searcy *et al.* 2007). Nesse contexto, plantas C4, como cana-de-açúcar, milho, *Miscanthus* e sorgo, destacam-se por ter maior eficiência fotossintética e, consequentemente, pelo maior potencial para acúmulo de biomassa (Ming *et al.* 2006; de Siqueira Ferreira *et al.* 2013). Dentre essas, a cana-de-açúcar se destaca por ter menores requerimentos de adubos nitrogenados e por ser cultivada de forma semi-perene, podendo ser utilizada durante sete ou mais safras sem a necessidade de replantio, diminuindo, assim, os custos de produção (Buckeridge *et al.* 2010). Entretanto, algumas questões de sustentabilidade afetam diretamente o desenvolvimento de culturas dedicadas à produção de bioenergia, principalmente em relação à competição por áreas cultiváveis com a produção de alimentos, podendo afetar a segurança alimentar dos países produtores. Com isso, culturas dedicadas a biocombustíveis devem apresentar características que as permitam ser cultivadas em áreas marginais, diminuindo a competição por área para a produção de alimentos (Somerville *et al.* 2010; Byrt *et al.* 2011; Sang 2011).

A parede celular vegetal evoluiu para garantir rigidez e resistência ao ataque de patógenos, sendo, portanto, recalcitrante à hidrólise enzimática, o que acarreta um problema para a produção de etanol celulósico (Himmel *et al.* 2007). Os principais constituintes da parede celular vegetal são a celulose, hemicelulose, pectinas e lignina. Destes, a lignina, o único componente não formado por carboidratos, é um polímero hidrofóbico altamente ramificado que está entrelaçado com a hemicelulose conferindo força e rigidez à parede celular (Carpita 1996; Boerjan *et al.* 2003). Por esta razão, a lignina impede o acesso das enzimas hidrolíticas aos carboidratos de parede. Adicionalmente, seus produtos de degradação podem ser inibitórios para enzimas e processos de fermentação (Keating *et al.* 2006). Isso faz com que a lignina seja considerada um dos principais obstáculos para a produção de etanol celulósico (Vanholme *et al.* 2008). Por outro lado, a lignina tem alto valor calorífico (White 1987a; Raveendran & Ganesh 1996), possuindo em média 30% mais energia que carboidratos celulósicos (Shafizadeh & Chin Peter P 1977; White 1987b), e pode ser usada como matéria prima para diversos produtos químicos de alto valor agregado, como dimetilsulfóxido (DMSO) e vanilina (Calvo-Flores & Dobado 2010), sendo um interessante biopolímero com aplicações biotecnológicas e comerciais.

Recentemente, iniciou-se o desenvolvimento de um novo tipo de cana, denominada cana energia, com menor acúmulo de sacarose, mais fibra e melhores características adaptativas, voltada para produção de biomassa. Em 2030, espera-se que a composição da cana energia referência seja de 12% de sacarose e 23% de fibras, diferentemente da cana-de-açúcar convencional, com 16,5% de sacarose e 11% de fibras (Leal *et al.* 2013). Entretanto, além do melhoramento genético, de maneira geral, estar focado em acúmulo de sacarose, os ganhos de produtividade a cada nova variedade lançada pelos programas de melhoramento estão se tornando cada vez menores, aproximando-se de um estado

estacionário (Dal-Bianco *et al.* 2012). Provavelmente, a base genética estreita das variedades comerciais atuais pode ser uma das causas desse fenômeno. Por outro lado, os valores teóricos máximos de produtividade de cana-de-açúcar chegam a 381 ton/ha, longe da média mundial atual de 84 ton/ha (Waclawovsky *et al.* 2010). Nesse cenário, é importante buscar por novas características, genes ou promotores para ajudar a contornar este problema, introduzindo-os nas futuras cultivares comerciais, seja via transgenia ou por melhoramento genético assistido por marcadores moleculares, para o desenvolvimento de novas variedades com as características desejadas, incluindo a cana energia.

Visto que existe uma grande variabilidade genética nas espécies de *Saccharum* (Tai & Miller 2002) e as espécies ancestrais de cana-de-açúcar possuem diversas características de interesse, aumenta-se a relevância de estudá-las para identificar os alvos moleculares e introduzi-los no melhoramento e no desenvolvimento da cana energia de uma forma mais direcionada. Até o presente, estudos moleculares com espécies ancestrais de cana-de-açúcar estão mais focados na identificação de marcadores moleculares e polimorfismos (Silva *et al.* 1993; Ming *et al.* 2002; Aitken *et al.* 2007; Li *et al.* 2011; Bundock *et al.* 2012), genotipagem e filogenia (Pan *et al.* 2000; Takahashi *et al.* 2005; Khan *et al.* 2009; Chang *et al.* 2012), mapeamento de cromossomos (D'Hont 1996; Ha *et al.* 1999; Piperidis *et al.* 2010) e apenas um estudo sobre miRNAs (Zanca *et al.* 2010) e um sobre transposons (Rossi *et al.* 2004). Contudo, nenhum estudo foi realizado englobando a expressão gênica, fisiologia e características agrônômicas como descrito no presente trabalho.

1.3. A parede celular

A parede celular é uma matriz complexa formada por diferentes polímeros que, juntos, formam uma estrutura funcional. A parede primária é formada durante os estágios

finais da divisão celular, quando o fragmoplasto, formado por microtúbulos, microfilamentos e retículo endoplasmático, guia a deposição de vesículas contendo o material da parede celular para formar a placa celular rica em pectinas (Vermeris & Saballos 2013). Esta placa forma a lamela média na qual os demais componentes da parede celular são depositados por ambas as células recém formadas. A parede celular secundária é depositada em células especializadas, principalmente nos vasos condutores do xilema e nas células de sustentação do esclerênquima, sendo depositada entre a lamela média e a membrana celular. Células com parede celular secundária normalmente estão sujeitas à morte celular programada, na qual o conteúdo celular é removido formando um tubo oco. A parede secundária é então exposta ao lúmen no novo tubo formado (Vermeris & Saballos 2013).

A parede celular vegetal pode ser classificada com base na arquitetura da parede primária (Carpita & Gibeaut 1993). A parede tipo I (Figura 1), comumente encontrada em dicotiledôneas e monocotiledôneas não comelinoides, contém uma rede de microfibrilas de celulose estruturada por cadeias da hemicelulose xiloglucano, embebida em uma matriz gelatinosa de pectinas, com proteínas estruturais fornecendo suporte adicional. A parede tipo II (Figura 1) é encontrada em monocotiledôneas comelinoides, grupo que inclui as gramíneas (Carpita & Gibeaut 1993). A celulose é entrelaçada por cadeias de hemicelulose glucuronoarabinoxilanas (GAX), possui pouco conteúdo de pectinas e proteínas e contém ácidos hidroxicinâmicos e β -D-glicanos de ligação mista, sendo este último hidrolisado ao término do crescimento da parede primária (Carpita 1996). Em ambos os tipos, I e II, os grupos carboxilas das pectinas podem se ligar a íons cálcio formando zonas de junção entre diferentes cadeias de pectinas (Figura 1), dando solidez ao gel. A matriz péctica fornece ambiente para a deposição da celulose e hemicelulose, além de ser o principal material adesivo entre as células (Willats *et al.* 2001). Entretanto, as pectinas estão praticamente

ausentes na parede celular secundária, sendo mais restrita à parede primária e a lamela média (Willats *et al.* 2001).

A celulose consiste de uma cadeia linear de moléculas de D-glicose unidas por ligações β -1,4, o que gera uma rotação de 180° entre moléculas adjacentes. As microfibrilas de celulose, tipicamente, consistem de 36 destas cadeias mantidas unidas por ligações de hidrogênio e cada cadeia possui milhares de unidades de glicose (Vermerris & Saballos 2013). Em gramíneas, sete microfibrilas podem se unir formando uma macrofibrila (Ding & Himmel 2006). A celulose é sintetizada a partir de moléculas de UDP-glicose por um complexo localizado na membrana plasmática formado pelas proteínas celulose sintase (Somerville 2006). Cada roseta é formada por 6 grupos de 6 subunidades de celulose sintase (CesA), possuindo 36 proteínas CesA individuais no total (Mueller & Brown 1980). São necessárias 3 diferentes proteínas CesA para a formação do complexo, e a identidade dessas proteínas difere entre a parede primária e secundária (Somerville 2006). Por exemplo, em arroz (*Oryza sativa*), as proteínas OsCesA1, 3 e 8 são específicas de parede primária enquanto as OsCesA4, 7 e 9 estão envolvidas com a síntese de parede secundária apenas (Tanaka *et al.* 2003; Wang *et al.* 2010d).

Os diferentes tipos de hemicelulose são sintetizados no complexo de Golgi (Carpita 2012) e direcionados para o meio extracelular, no sítio de montagem das microfibrilas de celulose (Saxena & Brown-Jr 2005). As hemiceluloses são sintetizadas por proteínas celulose sintase-like (Csl), e alguns membros dessa família são específicos de gramíneas e dicotiledôneas, como o CslF e o CslB, respectivamente (Yin *et al.* 2009). A principal hemicelulose de gramíneas comelinoides, GAX, consiste de cadeias lineares de xilose com ramificações de resíduos de α -L-arabinose e β -D-ácido glicurônico (Vermerris & Saballos 2013). Os resíduos de arabinose são esterificados à xilose na posição O-3, enquanto o ácido

glicurônico na posição *O*-2. O ácido ferúlico, um ácido com anel aromático, é esterificado nos resíduos de arabinose, permitindo a formação de um sítio de ligação entre diferentes moléculas de GAX através de pontes diferulato (Marita *et al.* 2003).

Figura 1. Esquema dos diferentes tipos de parede celular primária. Note que no tipo II (gramíneas) a principal e mais abundante hemicelulose que faz ligações com as microfibrilas de celulose é a glucuronoarabinoxilana, enquanto que no tipo I é o xiloglicano. Compostos em vermelho/alaranjado na parede tipo II indicam a rede de fenilpropanóides. Adaptado de (Carpita & McCann 2008).

A lignina é um polímero fenólico formado pelo acoplamento oxidativo dos monolignóis álcool *p*-coumaril, álcool coniferil e álcool sinapil (Boerjan *et al.* 2003) e constitui aproximadamente 25% da parede celular secundária. Os monolignóis são produzidos pela via de fenilpropanóides (Figura 2) que se inicia com a desaminação da fenilalanina pela enzima fenilalanina amônia liase (PAL), seguido de hidroxilações em uma, duas ou três posições do anel aromático, metilação em um ou dois destes grupos hidroxila, e duas reduções sucessivas na cadeia lateral, passando de ácido carboxílico a aldeído e, então,

a álcool (Boerjan *et al.* 2003). Além da PAL, a via envolve três diferentes enzimas citocromo P450 monooxigenase (C4H, C3'H e F5H), duas metiltransferases (CCoAOMT e COMT), duas oxirredutases (CCR e CAD) e duas enzimas adicionais (4CL e HCT) que catalisam a formação de intermediários que servirão de substrato para reações seguintes (Boerjan *et al.* 2003; Bonawitz & Chapple 2010). Os monolignóis são transportados para o apoplasto onde sofrem acoplamento oxidativo, catalisado por peroxidases e lacases, para serem incorporados no polímero e os resíduos derivados dos monolignóis passam a ser denominados *p*-hidroxifenil (H), guaiacil (G) e siringil (S). A proporção entre cada resíduo varia de acordo com a espécie, tecido e resposta a estímulos ambientais (Bonawitz & Chapple 2010). Por também ter a capacidade de sofrer oxidação radicalar como os monolignóis, o ácido ferúlico esterificado a moléculas de GAX funciona como uma ponte de ligação com a lignina (Ralph *et al.* 2004a; Burr & Fry 2009). Até hoje, pouco se sabe sobre o transporte dos monolignóis para o apoplasto, mas acredita-se que sejam transportados na forma de glicosídeos e liberados para a polimerização pela ação de glicosidases ((Bonawitz & Chapple 2010) e referências citadas por ele). Os primeiros modelos propostos sugeriram que o transporte de monolignóis ocorria por meio de vesículas derivadas do Golgi (Pickett-Heaps 1968), mas evidências mais recentes sugerem que o transporte não estaria associado a essas vesículas (Kaneda *et al.* 2008). Alternativamente, transportadores ABC (*ATP-binding Cassette*) estariam envolvidos, mas não existem evidências concretas para confirmar essa hipótese (Kaneda *et al.* 2008; Bonawitz & Chapple 2010).

Figura 2. Via de fenilpropanoides. PAL, fenilalanina amônia-liase; C4H, cinamato-4-hidroxilase; 4CL, 4-coumarato:CoA ligase; HCT, hidroxicinamoil:CoA transferase; C3H, 5-O-(4-coumaroil)shikimate 3-hidroxilase; CCoAOMT, caffeoil-CoA O-metil transferase; CCR, cinamoil-CoA redutase; CAD, cinamil álcool desidrogenase, F5H, ferulato/coniferaldeído 5-hidroxilase; COMT, cafeato/5-hidroxi-coniferaldeído O-metiltransferase. As moléculas sombreadas em cinza são os monolignóis que sofrerão oxidação e serão incorporados na lignina. Asterisco (*) identifica as enzimas cuja expressão foi analisada por qPCR neste trabalho.

O modelo mais tradicional da polimerização da lignina propõe o acoplamento combinatório entre os monolignóis oxidados sem qualquer tipo de controle mediado por proteínas, gerando um polímero com número astronômico de isômeros, o que sugere uma baixa probabilidade de duas moléculas de lignina ser iguais (Ralph *et al.* 2004b). Nas últimas décadas, alguns autores tem sugerido que a polimerização da lignina ocorreria de forma direcionada por proteínas dirigentes que ditariam a ordem das ligações entre as subunidades (Gang *et al.* 1999; Davin & Lewis 2005). Entretanto, é importante ressaltar que

as evidências suportam claramente o modelo de combinação aleatória (Bonawitz & Chapple 2010).

Recentemente, a estrutura química da parede celular de folhas e colmo jovens de uma variedade comercial de cana-de-açúcar foi estudada (Souza *et al.* 2012), identificando-se uma grande semelhança de composição e estrutura da parede nos dois tecidos. A parede celular de cana-de-açúcar seria composta aproximadamente de 28% de celulose, 58% de hemiceluloses (arabinoxilanas, β -glucanos de ligação mista e xiloglucanos), 8% de pectinas e 6% de lignina. Os autores destacaram a extrema complexidade da parede celular de cana-de-açúcar e propuseram um modelo para hidrólise utilizando a ação sequencial de diferentes enzimas. Também destacaram que os compostos fenólicos presentes na parede (ácido ferúlico e lignina) formam uma barreira importante para a hidrólise, de forma que as primeiras enzimas a serem utilizadas para abrir caminho para as pectinases, celulasas e hemicelulasas seriam aquelas capazes de delignificar e/ou quebrar as ligações fenólicas da parede. O modelo de parede celular de cana-de-açúcar proposto (Figura 3) indica que as microfibrilas de celulose estão envolvidas por cadeias de xiloglucanos e recobertas por compostos fenólicos, como lignina e ácido ferúlico. Em torno dessa estrutura, estariam moléculas de GAX que poderiam estar ligadas aos xiloglucanos por pontes diferúlicas. Estes feixes de microfibrilas cercados por hemiceluloses estariam embebidos em uma matriz composta de pectinas e β -glucanos de ligação mista.

Figura 3. Modelo estrutural da parede celular de cana-de-açúcar. A parede celular de cana-de-açúcar seria formada por macrofibrilas de celulose ligadas à xiloglicanos e GAX envoltos por compostos fenólicos, embebidos numa matriz de pectinas e β -glicanos. Adaptado de (Souza et al. 2012).

1.4. Regulação transcricional da biossíntese de parede celular secundária

A regulação da biossíntese de parede celular secundária é um processo extremamente complexo que envolve uma série de fatores de transcrição tecido-específico atuando em diferentes níveis hierárquicos, que irão ativar ou reprimir a transcrição dos genes envolvidos na biossíntese dos constituintes da parede. O principal modelo para o estudo das vias regulatórias de parede celular secundária é *Arabidopsis thaliana* e diversos fatores de transcrição já foram caracterizados nesta espécie. Outras espécies como álamo (*Populus trichocarpa*), eucalipto (*Eucalyptus gunnii*), tabaco (*Nicotiana tabacum*), *Antirrhinum majus*, pinheiro (*Pinus taeda*), uva (*Vitis vinifera*) e *Medicago truncatula* (revisado por (Zhong & Ye 2009; Zhao & Dixon 2011; Handakumbura & Hazen 2012; Wang & Dixon 2012)) também já tiveram etapas elucidadas das vias regulatórias. Fatores de transcrição das famílias NAC e MYB são os principais reguladores da biossíntese de parede celular secundária (Zhong & Ye 2007). Em um primeiro nível de controle (Figura 4), os fatores

de transcrição *NAC SECONDARY WALL THICKENING PROMOTING FACTOR1* (*NST1*), *NST2*, *SECONDARY WALL ASSOCIATED NAC DOMAIN PROTEIN1* (*SND1*), *VASCULAR-RELATED NAC-DOMAIN6* (*VND6*) e *VND7* atuam ativando a biossíntese de celulose, xilanas, lignina e deposição de parede celular secundária nas anteras (*NST1/NST2*), fibras (*SND1*) e xilema (*VND6* e *VND7*) em *Arabidopsis* (Kubo *et al.* 2005; Mitsuda *et al.* 2005; Zhong *et al.* 2006; Zhong *et al.* 2007b). Esses fatores da família NAC, coletivamente conhecidos como *secondary wall NACs* (SWNs), ativam a transcrição de uma série de fatores de transcrição à jusante, incluindo *SND2*, *SND3*, *MYB103*, *MYB85*, *MYB52*, *MYB54*, *MYB46*, *MYB69*, *MYB63*, *MYB83*, *MYB20* e *KNOTTED ARABIDOPSIS THALIANA7* (*KNAT7*) (Zhong *et al.* 2008; Ko *et al.* 2009; McCarthy *et al.* 2009; Wang & Dixon 2012), bem como vários genes da biossíntese de elementos de parede celular e morte celular programada (Zhong *et al.* 2008; Ohashi-Ito *et al.* 2010; Zhong *et al.* 2010; Yamaguchi *et al.* 2011). Os fatores *MYB46* e *MYB83*, alvos diretos dos SWNs (Figura 4), atuam de forma redundante em um segundo nível de regulação, sendo capazes de ativar todo o processo de formação de parede celular secundária (Zhong *et al.* 2007a; McCarthy *et al.* 2009). Os demais fatores à jusante atuam em um terceiro nível de regulação, sendo alguns ativadores específicos da síntese de celulose (*SDN2*, *SND3*, *MYB103*) ou lignina (*MYB58* e *MYB63*) (Zhong *et al.* 2008; Zhou *et al.* 2009). O *MYB58* é capaz de ativar a maioria dos genes da via de fenilpropanoides, com exceção do gene *F5H*, gene chave na síntese de unidades de siringil (Zhou *et al.* 2009). Posteriormente, identificou-se que a expressão de *F5H*, e consequente ativação da biossíntese de siringil, é diretamente regulada por *SND1/NST1* (Zhao *et al.* 2010).

Figura 4. Modelo da regulação transcricional da biossíntese de parede celular secundária. Adaptado de (Handakumbura & Hazen 2012).

Alguns indícios mostram que a síntese de lignina e celulose pode ser regulada de forma coordenada. Álamo (*Populus tremuloides*) transgênico com silenciamento do gene *4CL* apresentou redução de 40% no conteúdo de lignina com um aumento simultâneo de 15% no conteúdo de celulose (Hu *et al.* 1999). Da mesma forma, alguns trabalhos com fatores de transcrição suportam a hipótese de regulação coordenada de celulose e lignina. Um fator de transcrição de eucalipto, *EgMYB1*, quando expresso em *Arabidopsis* ou álamo (*Populus trichocarpa*), foi capaz de reprimir tanto genes da via de lignina quanto genes relacionados à síntese de polissacarídeos (Legay *et al.* 2010). Já o fator de transcrição de *Arabidopsis*, *AtSHN2*, da sub-classe *SHINE/WAX INDUCER* da família *EP2/ERF*, quando superexpresso em arroz (Ambavaram *et al.* 2011), acarretou em diminuição de lignina (45%) e aumento de celulose (34%), com indução da expressão de genes da família *CesA* e concomitante repressão dos genes da via de fenilpropanídeos *CAD* e *4CL*, além da repressão de fatores de transcrição de arroz homólogos aos *SND1/NST1*, *VND6* e *MYB58/63* de *Arabidopsis*.

Entretanto, a análise do transcriptoma de *Arabidopsis* transgênica superexpressando *AtSHN1* indicou um perfil diferente do encontrado nas linhagens transgênicas de arroz. Alguns genes relacionados tanto à síntese de lignina quanto de polissacarídeos foram regulados negativamente, enquanto outros genes das mesmas vias e também o fator de transcrição *NST1* foram regulados positivamente. Além disso, análises de lignina e carboidratos não indicaram alterações significativas no conteúdo de lignina e celulose no transgênico em relação à planta não transformada (Kannangara *et al.* 2007; Ambavaram *et al.* 2011). Esses resultados mostram um padrão de resposta bem distinto entre arroz e *Arabidopsis* superexpressando *AtSHN*, indicando que este gene pode ter diferentes funções em mono e dicotiledôneas e que, talvez, existam pontos de regulação não conservados entre essas duas classes.

Além dos níveis de regulação descritos acima, outros pontos de controle transcricional permitem o ajuste temporal e espacial preciso da ativação ou repressão das vias, aumentando a complexidade da rede (Figura 4). Em células da medula e do córtex do caule de *Arabidopsis*, o fator de transcrição *WRKY12* atua reprimindo a síntese e deposição de parede celular secundária, mantendo o aspecto de células parenquimáticas, apenas com parede celular primária. Experimentos *in vitro* indicaram que *WRKY12* é capaz de se ligar ao promotor de *NST2*, sendo, portanto, um possível alvo de regulação (Wang *et al.* 2010a). Um dos fatores principais, *SND1*, também possui a capacidade de ligar-se ao seu próprio promotor, ativando sua transcrição (Wang *et al.* 2011). Além de estar submetido a esta retroalimentação, *SND1* também controla indiretamente sua retroinibição. Isso ocorre porque *MYB46*, alvo direto de *SND1*, ativa o *MYB32* que, por sua vez, atua como um repressor de *SND1* (Wang *et al.* 2011). Além disso, esse controle parece estar sob regulação hormonal, uma vez que *MYB32* tem sua expressão fortemente ativada por ácido

indolacético (Preston *et al.* 2004). Além de *MYB32*, Li e colaboradores (Li *et al.* 2012) demonstraram que *KNAT7*, alvo direto dos SWNs, também atua como repressor da síntese de parede celular secundária. *Arabidopsis* mutante com perda de função de *KNAT7* apresenta aumento da espessura da parede celular das fibras e regulação positiva de genes da via de fenilpropanoides, síntese de xilanas e celulose. Por outro lado, a superexpressão de *KNAT7* leva a uma diminuição da espessura das fibras interfasciculares. Os autores sugeriram que *KNAT7* poderia regular diferentes genes e aspectos da deposição de parede celular secundária em diferentes tipos de células.

A grande maioria dos estudos sobre parede celular secundária, incluindo os fatores de transcrição descritos acima, foram realizados em dicotiledôneas, principalmente em *Arabidopsis*, e pouco se sabe sobre monocotiledôneas. Mesmo os estudos de fatores de transcrição de monocotiledôneas são, geralmente, realizados em dicotiledôneas como *Arabidopsis* ou *Nicotiana sp.* devido a uma série de razões técnicas, como a dificuldade de se obter plantas transgênicas para gramíneas e a indisponibilidade de mutantes (Gray *et al.* 2012). Recentemente, foi demonstrado que fatores NAC de arroz (*Oryza sativa*) e milho (*Zea mays*), filogeneticamente próximos dos SWNs de *Arabidopsis*, são capazes de ativar a biossíntese de parede celular secundária quando superexpressos em *Arabidopsis* e também de complementar um mutante de *Arabidopsis* defectivo para *SND1* e *NST1* (Zhong *et al.* 2011). Os autores também demonstraram que *OsMYB46* e *ZmMYB46* são ortólogos funcionais de *AtMYB46/83*, sendo capazes de ativar todo o programa de biossíntese de parede celular secundária quando superexpresso em *Arabidopsis*. Ainda, ensaios em protoplastos de *Arabidopsis* mostraram que os SWNs de arroz e milho eram capazes de se ligarem aos promotores de *OsMYB46* e *ZmMYB46* e transativá-los. Entretanto, vale ressaltar que estes fatores de transcrição de gramíneas foram testados funcionalmente em uma

planta dicotiledônea e suas reais funções na espécie de origem ainda não foram demonstradas. Em cevada (*Hordeum vulgare*), apesar de não terem sido caracterizados funcionalmente, alguns fatores da família NAC tiveram sua expressão analisada em diferentes tecidos, e para três deles, *HvNAC033*, *HvNAC034* e *HvNAC039*, sendo o primeiro um homólogo putativo de *AtNST1*, observou-se uma correlação positiva com tecidos com maior biossíntese de parede celular secundária (Christiansen *et al.* 2011). Até hoje, provavelmente apenas um SWN de gramínea foi caracterizado em sua espécie de origem. O fator *BdSWN5*, da gramínea modelo *Brachypodium distachyon*, quando expresso sob controle de promotor induzível, demonstrou a capacidade de ativar a expressão de uma celulose sintase (*BdCesA4*), uma protease específica de xilema (*BdXCP1*), envolvida na diferenciação deste tecido, e *BdMYB1*, um homólogo putativo de *AtMYB46* (Valdivia *et al.* 2013). É importante destacar que *BdCesA4* é homólogo de *OsCesA4*, específico de parede celular secundária (Tanaka *et al.* 2003), mas *BdSWN5* não induziu a expressão de *BdCesA1*, homólogo de *OsCesA1*, específico de parede primária, indicando a função de *BdSWN5* na ativação de parede celular secundária (Valdivia *et al.* 2013). Os autores também testaram todos os 8 fatores NAC presentes no genoma de *B. distachyon* em ensaios de expressão transiente em tabaco, mostrando que todos eram capazes de ativar a síntese de parede celular secundária e/ou morte celular. Surpreendentemente, ensaios de transativação mostraram que *BdSWN5* não é capaz de ativar diretamente a expressão do promotor de *BdMYB1*, mas, sim, do promotor de *AtMYB46* e *BdXCP1*. Os autores sugerem que esses resultados indicam que muito do que foi estudado em *Arabidopsis* pode ser relevante para gramíneas, mas também indicam a possibilidade da existência de algumas diferenças importantes.

Além destes fatores de transcrição de gramíneas envolvidos com a regulação da biossíntese de parede celular secundária descritos acima, apenas 4 outros já foram caracterizados, e todos pertencem à família MYB, ao subgrupo relacionado a repressores transcricionais. Quando superexpressos em *Arabidopsis*, os genes de milho *ZmMYB42* e *ZmMYB31* foram capazes de reprimir diversos genes da via de fenilpropanoides, bem como diminuir o conteúdo de lignina (Fornale *et al.* 2006; Sonbol *et al.* 2009; Fornale *et al.* 2010). Estes fatores foram os primeiros a serem caracterizados em gramínea, onde experimentos de imunoprecipitação de cromatina e expressão transiente em protoplastos de milho demonstraram a ligação de *ZmMYB31* aos promotores dos genes da via de fenilpropanoides *ZmF5H* e *ZmCOMT* (Fornale *et al.* 2010), e a expressão simultânea em *Arabidopsis* de *ZmMYB31* ou *ZmMYB42* e GFP sob controle do promotor de *ZmCOMT* levou a repressão da expressão de GFP (Fornale *et al.* 2006). Outro fator de gramínea, *TaMYB4* de trigo (*Triticum aestivum*), quando superexpresso em tabaco leva a uma repressão da expressão de *CCR* e *CAD* e diminuição do conteúdo de lignina (Ma *et al.* 2011). O fator MYB de gramínea mais recentemente caracterizado foi o *PvMYB4* de *switchgrass* (*Panicum virgatum*) (Shen *et al.* 2012). A superexpressão de *PvMYB4* tanto em tabaco quanto em *switchgrass* causou a redução da expressão de quase todos os genes da via de fenilpropanoides e também a redução do conteúdo de lignina.

Os sítios de ligação dos SWNs e dos TFs MYB nos promotores alvo já foram identificados. Os SWNs se ligam a uma sequência palindrômica imperfeita de 19 pb (Zhong *et al.* 2010) (T/A)NN(C/T)(T/C/G)TNNNNNNNA(A/C)GN(A/C/T)(A/T), conhecidos como *secondary wall NAC binding element* (SNBE). Diversos alvos de *AtSND1* possuem esse cis-elemento, incluindo *MYB83*, *MYB103*, *SND3* e *KNAT7* (Zhong *et al.* 2010), assim como *OsMYB46* e *ZmMYB46* (Zhong *et al.* 2011). Em *B. distachyon*, *BdSWN5* também é capaz de se

ligar ao SNBE presente no promotor de *BdXCP1* (Valdivia *et al.* 2013). Já os fatores MYB se ligam aos cis-elementos AC (Lois *et al.* 1989; Hatton *et al.* 1995; Raes *et al.* 2003): AC-I (ACCTACC), AC-II (ACCAACC) e AC-III (ACCTAAC). Recentemente, a análise do sítio de ligação de AtMYB46 e AtMYB83 permitiu a identificação da sequência consenso ACC(A/T)A(A/C)(T/C), que engloba os três elementos AC e expande o número de elementos, denominada *secondary wall MYB responsive element* (SMRE) (Zhong & Ye 2012). Outro grupo identificou uma sequência consenso diferente com o sítio de ligação de MYB46, (A/G)(G/T)T(A/T)GGT(A/G) (Kim *et al.* 2012), denominado M46RE. Diversos genes da via de fenilpropanoides em *Arabidopsis* contêm elementos AC, ou SMREs, em seus promotores (Raes *et al.* 2003). Os fatores de gramíneas ZmMYB31 e PvMYB4 também são capazes de se ligar a SMREs presente no promotor dos genes da via de fenilpropanoides (Fornale *et al.* 2010; Shen *et al.* 2012). Entretanto, apesar de reprimir a expressão do gene *COMT* em *Arabidopsis* (Fornale *et al.* 2006), ZmMYB31 não é capaz de se ligar ao promotor de *AtCOMT*, sugerindo uma repressão indireta. Por outro lado, foi demonstrado a ligação tanto *in vitro* como *in vivo* de ZmMYB31 ao promotor de *ZmCOMT* (Fornale *et al.* 2010), sugerindo uma divergência evolutiva na regulação do gene *COMT* por fatores MYB entre *Arabidopsis* e milho.

1.5. Fotossíntese, acúmulo de sacarose e biomassa

Plantas como cana-de-açúcar, milho, sorgo, *Miscanthus*, entre outras, pertencem à tribo Andropogoneae, cujos membros utilizam a fotossíntese C4. Esse tipo de fotossíntese permite uma eficiência de 6% na conversão da energia solar, superior aos 4,6% obtidos pela fotossíntese C3 (Zhu *et al.* 2008), utilizada por plantas como *Arabidopsis*, tabaco e outras gramíneas como arroz e trigo. Além da maior eficiência energética, plantas C4 apresentam

maior eficiência de uso de água e nitrogênio (Ming *et al.* 2006; de Siqueira Ferreira *et al.* 2013). A alta capacidade fotossintética das plantas C4 é devida a integração metabólica de dois tipos diferentes de célula: as células do mesófilo e da bainha do feixe vascular (Figura 5). A assimilação de CO₂ ocorre nas células do mesófilo e é catalisada pela enzima fosfoenolpiruvato carboxilase (PEPC), juntamente com a anidrase carbônica (CA) que é fundamental para o rápido equilíbrio entre o CO₂ e HCO₃⁻. Os ácidos com quatro carbonos resultantes (C4) difundem para as células da bainha do feixe vascular, onde o CO₂ é liberado por uma decarboxilase (enzima málica em cana-de-açúcar). Assim, a concentração de CO₂ é aumentada em cerca de 10 vezes no sítio de carboxilação pela ribulose-1,5-bisfosfato carboxilase oxigenase (Rubisco) em relação ao valor atmosférico (de Siqueira Ferreira *et al.* 2013). A Rubisco utiliza tanto o CO₂ quanto o O₂ como substrato e a fixação deste último leva a um processo de perda de eficiência energética denominado fotorrespiração. Dessa forma, a competição entre o CO₂ e o O₂ pelo sítio catalítico da Rubisco nas plantas C4 é menor em comparação às plantas C3, aumentando a assimilação de carbono em altas temperaturas por reduzir a fotorrespiração (Sheen 1999).

Figura 5. Fotossíntese C4 (A) e C3 (B). Números em vermelho indicam as enzimas envolvidas na reação sendo: 1, Anidrase carbônica; 2, Fosfoenolpiruvato carboxilase; 3, NADP malato desidrogenase; 4, Enzima málica; 5, Rubisco; 6, piruvato ortofosfato diquinase. Modificado de (de Siqueira Ferreira *et al.* 2013).

Uma vez fixado nas folhas (fonte), o carbono é utilizado para sintetizar sacarose, a principal molécula transportada pelo floema (Komor 2000), sendo, então, distribuídos para os demais tecidos (drenos), como raiz, colmo e frutos. A síntese de sacarose inicia-se no citosol a partir das trioses fosfatos produzidas no cloroplasto pelo ciclo de Calvin, que serão transformadas em frutose-1,6-fosfato e esta será posteriormente defosforilada pela frutose-1,6-bifosfatase citosólica (Serrato *et al.* 2009). A frutose-6-fosfato resultante pode ser então convertida em glicose-6-fosfato, ou ser combinada com UDP-glicose para formar sacarose-6-fosfato pela ação da enzima sacarose fosfato sintase (Stitt & Quick 1989), sendo subsequentemente defosforilada por uma fosfatase para ser direcionada ao floema e demais tecidos (drenos).

Tanto a fonte quanto o dreno podem ser limitantes para o acúmulo de açúcares. Limitações da fonte implicam na falta de eficiência fotossintética para atingir o potencial do dreno em importar o carbono fixado, enquanto as limitações do dreno ocorrem quando a capacidade fotossintética excede a capacidade do dreno em importar o carbono fixado (Patrick *et al.* 2013). Experimentos de manipulação da relação fonte/dreno geralmente indicam que drenos meristemáticos são limitados pela fonte, enquanto a expansão celular e os tecidos de reserva são limitados pelo dreno (Smith & Stitt 2007).

A taxa fotossintética é regulada por diversos mecanismos. Em cana-de-açúcar, a regulação da fotossíntese está relacionada à demanda de carboidratos do dreno (McCormick *et al.* 2006) e à maturidade do colmo (McCormick *et al.* 2008b; McCormick *et al.* 2008a). Além disso, a fotossíntese sofre inibição com o aumento do acúmulo de açúcares nas folhas, fato já demonstrado em cana-de-açúcar (McCormick *et al.* 2008b) e diversas outras plantas (Krapp *et al.* 1993; Krapp & Stitt 1995), sendo, portanto, um mecanismo de regulação conservado entre as espécies. Os mecanismos desta retroinibição ainda não são bem

entendidos (Patrick *et al.* 2013), mas podem envolver enzimas como a hexoquinase (Dai *et al.* 1999) e estarem relacionados a limitações no dreno, envolvendo a atividade de transportadores de sacarose (revisado por (Patrick *et al.* 2013)).

Em cana-de-açúcar, o descarregamento da sacarose do floema até o parênquima de reserva (Figura 6) ocorre pela rota simplástica (por dentro das células), via plasmodesmas (Patrick *et al.* 2013). Chegando às células do parênquima, a sacarose é transportada para os vacúolos por um transportador antiporte de sacarose e prótons. Cria-se, assim, um potencial osmótico negativo intracelular (π_i) fazendo com que a água mova-se para dentro dos vacúolos através de proteínas de membrana chamadas aquaporinas, as quais formam canais hidrofílicos que permitem a rápida difusão da água (Agre *et al.* 1998). A sacarose pode “vazar” para o apoplasto, sendo recaptada por um transportador simporte de sacarose e prótons. A barreira apoplástica formada pela suberização (Jacobsen *et al.* 1992) e lignificação da parede das células que separam o feixe vascular do parênquima de reserva permite que os potenciais osmóticos de seus apoplastos sejam regulados independentemente (Patrick *et al.* 2013). Como consequência, as diferenças na pressão hidrostática que direcionam o descarregamento da sacarose pela rota simplástica são homeostaticamente mantidas, independente da concentração de sacarose no vacúolo (Patrick 1997; Patrick *et al.* 2013).

Figura 6. Esquema de acúmulo de sacarose nas células do colmo de cana-de-açúcar. Setas verdes, cinzas e azuis indicam o fluxo de sacarose (Sac), prótons (H^+) e água (H_2O), respectivamente. F, elementos de tubos crivados do floema; PF, parênquima do floema; PR, parênquima do tecido de reserva; Vac, vacúolo; π_i potencial osmótico intracelular; π_e potencial osmótico extracelular (apoplasto); π_{ef} potencial osmótico extracelular (apoplasto) do floema. As barras em vermelho indicam a barreira hidrofóbica entre os dois tecidos. Modificado de (Patrick et al. 2013).

Ao chegar aos tecidos dreno, a sacarose pode ser armazenada nas células do parênquima do colmo ou ser hidrolisada pela invertase, produzindo glicose e frutose (Zeng *et al.* 1999), ou pela sacarose sintase, produzindo frutose e UDP-glicose (Geigenberger & Stitt 1993) que serão utilizadas para síntese de carboidratos complexos (Rae *et al.* 2005) como os de parede celular, além de serem utilizadas na respiração e em outros processos metabólicos. Além disso, a sacarose e as hexoses derivadas de sua clivagem funcionam como moléculas sinalizadoras primárias, promovendo a proliferação celular, a diferenciação e a maturação (Roitsch & Gonzalez 2004).

Pode-se definir, de maneira geral, que os dois grandes drenos de carboidratos em cana-de-açúcar são o acúmulo de sacarose e a síntese de biomassa na forma de material lignocelulósico, ou seja, parede celular secundária. De toda biomassa disponível nas partes

aéreas da planta, 1/3 está na forma de sacarose estocada nos colmos e os outros 2/3 na forma de biomassa lignocelulósica nas folhas e colmos (Goldemberg 2008). Assim, a síntese de parede celular e o acúmulo de sacarose competem pelo carbono fixado na fotossíntese. Entretanto, essa correlação pode ser mais complicada, uma vez que o acúmulo de sacarose nos vacúolos necessita da expansão celular que, por sua vez, depende do metabolismo de parede celular.

1.6. Perfil de expressão gênica em larga escala

A identificação das bases moleculares associadas a características agronômicas de interesse é importante para a biotecnologia de plantas cultivadas. Entretanto, uma determinada característica pode ser controlada por múltiplos genes que sofrem interferência do ambiente, formando uma complexa rede de regulação, dificultando o estudo (Benfey & Mitchell-Olds 2008). Com o aperfeiçoamento das tecnologias, tornou-se possível a criação de plataformas para análise de expressão gênica em larga escala, cobrindo virtualmente todo o transcriptoma de uma determinada espécie. Tecnologias de análise do perfil de expressão gênica em larga escala fornecem uma fonte rica de informação biológica que permite aos pesquisadores integrar e entender interações entre múltiplos componentes nas células e organismos (Baginsky *et al.* 2010) e, portanto, possibilita a geração de dados capazes de ajudar na identificação de vias e genes associados às redes regulatórias das características de interesse.

A análise do perfil de expressão gênica em larga escala em cana-de-açúcar iniciou-se com a produção de catálogos de ESTs (*expression sequence tags*) por diferentes grupos de pesquisa em vários países (revisado por (Manners & Casu 2011; de Siqueira Ferreira *et al.* 2013; Nishiyama-Jr *et al.* 2013)). O mais importante destes, o projeto SUCEST (*SugarCane*

EST Genome Project) (Vettore *et al.* 2001; Vettore *et al.* 2003) financiado em parceria pela FAPESP e COPERSUCAR, gerou 237.954 ESTs a partir de 26 bibliotecas diferentes de cDNA as quais foram agrupadas em 43.141 transcritos (ou SAS, de *sugarcane assembled sequence*). A partir do SUCEST, foram gerados catálogos de genes para a utilização em experimentos de microarranjo de cDNA na análise da expressão gênica de cana-de-açúcar em diferentes tecidos, variedades ou submetidas a diferentes tratamentos. A técnica de microarranjo permite a comparação de duas populações de mRNAs, as quais são marcadas na forma de cDNA com diferentes fluoróforos e hibridizadas em uma lâmina contendo sequências alvo. A intensidade da fluorescência de cada uma das populações de cDNA permite a identificação da expressão relativa entre elas (Liang *et al.* 2004). Experimentos de microarranjo permitem gerar uma grande quantidade de informação sobre a genômica funcional, mesmo sem ter conhecimento prévio da função ou envolvimento do gene, produzindo uma visão global dos processos biológicos e sendo uma fonte rica para formulação de hipóteses (Schena *et al.* 1998; Liang *et al.* 2004).

Os catálogos SUCAST e SUCAMET, que compreendem 1830 SAS relacionados à transdução de sinal e 4594 SAS relacionados a vias metabólicas, respectivamente, permitiram a identificação de genes diferencialmente expressos em diferentes tecidos, associados a características como acúmulo de açúcar, resposta a seca, interação com patógenos, resposta a hormônios, elevação de CO₂, entre outros (Souza *et al.* 2001; Papini-Terzi *et al.* 2005; Rocha *et al.* 2007; De Souza *et al.* 2008; Felix *et al.* 2009; Papini-Terzi *et al.* 2009). Recentemente, uma nova plataforma foi desenvolvida para análise de expressão gênica de cana-de-açúcar utilizando uma lâmina customizada de oligonucleotídeos (Agilent Technologies). Essa lâmina possui 45220 elementos, com 1417 controles Agilent e 43803 oligonucleotídeos representando os SAS, sendo 21901 em duplicatas e 1 oligonucleotídeo

sem replicata. Destes, 14522 detectam o transcrito na orientação senso e 7380 detectam na orientação antissenso. Utilizando essa plataforma, Lembke e colaboradores (Lembke *et al.* 2012) relataram genes diferencialmente expressos em cana-de-açúcar em resposta a estresse por seca, incluindo transcritos antissenso corregulados com seus respectivos transcritos senso. Além do grande número de genes presentes, esta lâmina possui alta taxa de validação por qPCR, sendo uma plataforma muito confiável para análise de expressão gênica em larga escala, permitindo a identificação de genes associados a características de interesse e a formulação de hipóteses sobre as redes regulatórias envolvidas. A validade de experimentos de microarranjo na identificação de genes associados a características de interesse pode ser exemplificada pela identificação de um gene de função desconhecida diferencialmente expresso em plantas submetidas a estresse por seca (Rocha *et al.* 2007), o qual quando superexpresso em plantas de tabaco confere tolerância a diversos tipos de estresse, incluindo seca e estresse oxidativo (Begcy *et al.* 2012).

Uma das limitações dos estudos de expressão gênica por microarranjos é a necessidade de ter um conhecimento prévio da sequência a ser depositada na lâmina e o próprio número de sondas presentes na lâmina. Com o aperfeiçoamento das plataformas de sequenciamento de nova geração, tornou-se possível a análise de expressão gênica por uma técnica conhecida como RNA-Seq. Essa técnica permite a identificação e sequenciamento dos transcritos de uma determinada amostra, obtendo um perfil muito mais completo e possibilitando a detecção de RNAs não codificadores, descoberta de novos transcritos e variantes de *splicing*, quantificação da expressão de genes, entre várias outras aplicações, tornando-se uma técnica cada vez mais comum (Mutz *et al.* 2013). Existem várias plataformas de sequenciamento de nova geração, cada uma com suas vantagens e desvantagens; a escolha depende de vários fatores como aplicação, objetivo da análise,

custo-benefício, entre outras (Mutz *et al.* 2013). Muitos dos transcritos já sequenciados em cana-de-açúcar não possuem a região codificadora completa e o RNA-seq pode fornecer sequências adicionais para, juntamente com as sequências dos bancos de ESTs, ser possível a montagem de transcritos completos (Casu *et al.* 2010). Em cana-de-açúcar, alguns estudos já demonstraram a validade do uso de RNA-Seq na identificação de pequenos RNAs (Ferreira *et al.* 2012; Thiebaut *et al.* 2012; Carnavale Bottino *et al.* 2013; Gentile *et al.* 2013) e análise de retrotransposons (Domingues *et al.* 2012). Porém, a análise global do transcriptoma de cana-de-açúcar por RNA-Seq ainda está apenas começando.

1.6.1. Estudos de transcriptoma com enfoque em parede celular

Por se tratar de uma estrutura que é a fonte principal de biomassa, o completo entendimento das redes moleculares envolvidas na montagem e remodelamento da parede celular é crucial para a conversão eficiente da biomassa em biocombustíveis (Carpita & McCann 2008). Com isso, o perfil de expressão dos genes de parede celular de várias espécies, tanto comerciais quanto plantas modelo, é comumente alvo de estudo visando à identificação de genes e vias que possam ajudar a avançar no conhecimento dessa área. Mesmo com inúmeros estudos já conduzidos, estima-se que existam por volta de 1.000 genes com função ainda desconhecida que exerçam algum papel no metabolismo de parede celular (Yong *et al.* 2005), o que indica a necessidade de continuar e expandir tais estudos, incluindo o transcriptoma, para identificar esses genes ainda não caracterizados.

Apesar do sequenciamento completo do genoma de *Arabidopsis thaliana* ter ajudado a identificar vários genes envolvidos no metabolismo de parede celular, incluindo os respectivos homólogos em outras espécies como gramíneas e espécies madeireiras, a utilização de gramíneas como fonte de biomassa para biocombustíveis requer a elucidação

das funções de seus genes específicos a fim de se obter “controle” da qualidade e quantidade de biomassa produzida (Carpita & McCann 2008). Assim, para gramíneas como milho e arroz que já possuem mais recursos biotecnológicos (principalmente a disponibilidade de um genoma referência), já foi possível identificar diversos membros de famílias gênicas associadas à parede celular, como genes da via de lignina, glicosiltransferases e hidrolases, expansinas, entre outras, incluindo genes específicos de gramíneas como as *CsIF* (Penning *et al.* 2009; Wang *et al.* 2010d). Recentemente, Sekhon e colaboradores (Sekhon *et al.* 2011) relataram uma análise global do transcriptoma de milho utilizando microarranjos de oligonucleotídeos para estudar a expressão gênica em 60 tecidos diferentes e constataram que tecidos biologicamente relacionados possuem perfis de expressão similares. A análise de genes da via de biossíntese de lignina mostrou que sua expressão varia com a idade e a relação entre os tecidos. Por exemplo, partes aéreas apresentam padrão bem distinto de raiz; entrenós mais jovens também apresentam padrão diferente de entrenós mais velhos (Sekhon *et al.* 2011). No mesmo ano, outro grupo reportou a análise de expressão por microarranjo em entrenós de milho em fases de alongação e não-elongação (Bosch *et al.* 2011), numa tentativa mais direcionada de identificar os principais genes de biossíntese de parede celular envolvidos na maturação do colmo. Os autores conseguiram identificar novos genes candidatos à biossíntese de parede celular primária e secundária, uma vez que apresentavam padrão de expressão bem distinto em tecidos com deposição desses dois tipos de parede celular (Bosch *et al.* 2011). Dentre esses, vários fatores de transcrição, das famílias NAC, MYB, AP2-EREB, WRKY, Platz, bHLH, entre outros, demonstraram perfil de expressão que os tornam candidatos à reguladores da biogênese de parede celular. Entretanto, dos 10 fatores NAC e 5 MYB, apenas um MYB é homólogo de algum gene de *Arabidopsis*, *AtMYB85*, já caracterizado como regulador desse

processo (Bosch *et al.* 2011). Por outro lado, em estudos de RNAseq em diferentes regiões da folha de milho, foi demonstrado que expressão gênica da base da folha, onde a elongação está mais ativa e ocorre e a transição de dreno para fonte, está enriquecida em genes de metabolismo de parede celular, podendo-se destacar vários fatores de transcrição das famílias NAC e MYB, homólogos de *AtSND1*, *AtSND2*, *AtKNAT7*, *AtMYB85*, *AtMYB46*, entre outros (Li *et al.* 2010).

Diferentemente de milho e *Arabidopsis*, poucos estudos voltados para a expressão de genes do metabolismo de parede celular em cana-de-açúcar foram conduzidos. Existem vários estudos de transcriptoma de cana-de-açúcar que identificaram expressão diferencial de genes do metabolismo de parede celular em, por exemplo, plantas submetidas à seca (Lembke *et al.* 2012) e em variedades contrastantes para teor de sacarose (Papini-Terzi *et al.* 2009), onde genes como expansinas e xiloglucano transglicosilase/hidrolase (XTH) foram identificados e sugerido uma possível correlação da expressão desses genes com o teor de sacarose. Por outro lado, até onde vai nosso conhecimento, existem pouquíssimos estudos cujo objetivo é estudar o perfil de expressão de genes de parede celular de cana-de-açúcar. Em um dos primeiros trabalhos (se não o primeiro) abordando a expressão de genes em diferentes tecidos em uma escala relativamente alta, Lima e colaboradores (Lima *et al.* 2001) utilizaram BlastX e 25 palavras-chave (celulose sintase, expansina, xilanase, entre outras) para identificar 3.283 clones, dentre as 200 mil ESTs do SUCEST, relacionados ao metabolismo de parede celular, posteriormente agrupados em 459 genes possíveis. A partir do número de clones por gene, os autores estimaram o nível de expressão em diferentes tecidos e conseguiram identificar que as celulose sintases são mais expressas em folha e raiz e concluíram que o perfil de expressão encontrado refletiu bem as características fisiológicas esperadas para cada tecido (Lima *et al.* 2001). Anos depois, análises de oligoarranjos de três

diferentes partes do colmo permitiu identificar 119 genes diferencialmente expressos de diversas categorias, incluindo parede celular (Casu *et al.* 2007). Em uma segunda abordagem, os mesmos autores identificaram 17 sondas anotadas como membros das famílias CesA e Csl e estas podiam ser divididas em 5 *clusters* de acordo com sua expressão nos diferentes entrenós, mostrando que essas famílias são diferencialmente reguladas no colmo (Casu *et al.* 2007). Em seguida, utilizando-se critérios bem estridentes (correlação $>0,95$), foram buscados genes que apresentassem o mesmo perfil de um dos 5 *clusters* das celulose sintases, identificando-se 45 genes os quais são todos de metabolismo de parede celular, como expansinas, CAD, glucanases, etc, mostrando que, de fato, diferentes genes do metabolismo de parede celular possuem padrão de expressão correlacionado. Entretanto, os autores não conseguiram identificar um padrão característico de expressão das CesA em tecidos sob deposição de parede primária e secundária, não sendo possível determinar CesA específicas de parede primária ou secundária (Casu *et al.* 2007).

2. OBJETIVOS

Espécies ancestrais de cana-de-açúcar possuem fenótipos bem distintos e podem ser fontes de características agronômicas de interesse como perfilhamento, teor de fibra e resistência a estresses de *S. robustum* e, principalmente, *S. spontaneum* e teor de sacarose de *S. officinarum*. A introgressão destes genótipos nos programas de melhoramento já é realizada, mas pouco se sabe sobre a biologia molecular e transcriptoma destas espécies. Visto o grande interesse no uso de cana-de-açúcar como fonte de biomassa para produção de bioenergia, o presente trabalho utilizou-se de diferentes abordagens para caracterizar alguns parâmetros agronômicos e fisiológicos de interesse e analisar o perfil de expressão gênica de espécies ancestrais de cana-de-açúcar com enfoque nos genes de metabolismo de parede celular para a identificação de genes, vias metabólicas e redes regulatórias que possam ter envolvimento nas respostas fisiológicas analisadas e que possam ser relevantes para o entendimento da biologia molecular e o melhoramento genético da cana-de-açúcar e da cana energia.

Esse trabalho tem por objetivo caracterizar alguns parâmetros fisiológicos de interesse de genótipos ancestrais e um híbrido comercial de cana-de-açúcar e também analisar o transcriptoma dessas plantas, correlacionando-os aos parâmetros fisiológicos caracterizados para identificar as vias metabólicas e os genes que possam estar relacionados às diferenças fenotípicas apresentadas, principalmente em relação à biossíntese de parede celular secundária, acúmulo de biomassa e sacarose, bem como outras características agrotecnológicas de interesse como produtividade, fotossíntese e resistência a estresses.

Fazem parte deste trabalho os seguintes objetivos específicos:

- Identificação dos genes de cana-de-açúcar do catálogo SUCEST que sejam homólogos aos genes de outras espécies já caracterizados como envolvidos metabolismo de parede celular;
- Caracterização de parâmetros fisiológicos e morfológicos de três espécies ancestrais (*S. officinarum*, *S. robustum* e *S. spontaneum*) e um híbrido comercial (RB867515) para identificar os principais fenótipos contrastantes que possam ser relevantes para produção de sacarose e biomassa;
- Análise do perfil de expressão gênica desses genótipos via microarranjo de oligonucleotídeos e PCR em tempo real, com enfoque em genes de metabolismo de parede celular, correlacionando com os fenótipos de interesse;
- Identificação de genes com potencial para o melhoramento genético;
- Identificação de sequências promotoras de genes relacionados ao metabolismo de parede celular;
- Construção de uma biblioteca de cDNA *full length* de *S. officinarum*, *S. spontaneum* e da variedade comercial SP803280 e análise de expressão em larga escala por FL-RNAseq;

3. METODOLOGIA

3.1. Construção de um catálogo de genes relacionados a parede celular em cana-de-açúcar

A partir do banco de ESTs do SUCEST (Vettore *et al.* 2003), foi realizada a busca por genes relacionados ao metabolismo de parede celular de cana-de-açúcar. Utilizando como referência os bancos de dados de genes de parede celular Maizewall (<http://www.polebio.scsv.ups-tlse.fr/MAIZEWALL/>) e *Cell Wall Genomics* (<http://cellwall.genomics.purdue.edu/>), foi realizada uma análise de BlastX (Altschul *et al.* 1990) e identificou-se mais de 6 mil sequências prováveis. Todas essas 6 mil sequências foram individualmente checadas por BlastX contra os bancos de dados NR (<http://blast.ncbi.nlm.nih.gov/>) e Uniprot (<http://www.uniprot.org/>) e as sequências também foram e checadas pela presença de domínios conservados no Pfam (<http://pfam.sanger.ac.uk/>).

3.2. Material vegetal

3.2.1. Plantas utilizadas nas análises de qPCR, Microarranjo, fisiologia e anatomia

Os genótipos ancestrais de cana-de-açúcar *S. officinarum* (caiana listrada), *S. robustum* (IM76229), *S. spontaneum* (IN8458) e a variedade comercial RB867515 foram utilizados para os experimentos de microarranjo, qPCR, análise de carboidratos, lignina, umidade, características agronômicas e fotossintéticas. As plantas foram mantidas em uma linha de 5 metros no campo experimental do Centro de Ciências Agrárias da Universidade

Federal de São Carlos (CCA-UFSCAR), campus Araras, e as amostras foram cedidas pela professora Dra. Monalisa Sampaio. As amostras utilizadas para análises de oligoarranjos e validação por qPCR foram coletadas em julho de 2011 (9 meses de idade), sendo que duas réplicas biológicas foram utilizadas para os oligoarranjos (Tabela 4) e os resultados foram validados por qPCR numa terceira réplica biológica (Figuras 15-17). As amostras utilizadas para análise de expressão gênica apenas por qPCR (Tabelas 13-15 e 17), análises agronômicas, fotossíntese de carboidratos, lignina e anatomia foram coletadas em março de 2012 (7 meses de idade). Diferentemente das análises de expressão gênica das planta de 9 meses de idade (coleta de julho de 2011), as análises de qPCR com as amostras de março de 2012 foram realizadas com 3 réplicas biológicas, uma vez que não foi realizado nenhum experimento de oligoarranjo. As amostras coletadas foram de folha+1, entrenó imaturo, intermediário e maduro. Folha+1 é a primeira folha com bainha visível (*dewlap*), enquanto o entrenó imaturo é o primeiro entrenó, no ápice do colmo. Os entrenós intermediário e maduro são o 5º e 9º entrenós, respectivamente, contando-se do ápice para a base, a partir do entrenó 1 (imaturo).

Os experimentos com genótipos da RIDESA (Sato 2012), utilizados para comparar com os resultados dos experimentos dos genótipos ancestrais, são derivados de dois cruzamentos biparentais, SP832874 x TUC717 e RB855002 x RB855035. As progênies de cada cruzamento foram avaliadas para o conteúdo de Brix e classificadas em plantas de alto e baixo Brix. Para as hibridizações, os tecidos de 10 genótipos alto Brix e 10 baixo Brix das plantas com 7 meses de idade foram selecionados dentro de cada cruzamento biparental e, para diminuir a variação da expressão gênica de cada genótipo, para cada tecidos (folha +1 ou entrenó imaturo) as amostras foram agrupadas em *pools* e submetidas à extração do RNA total. Nos experimentos descritos por Papini-Terzi e colaboradores (Papini-Terzi *et al.* 2009),

os genótipos foram gerados a partir de dois cruzamentos sucessivos. Primeiramente, os genótipos foram gerados por policruzamentos intra-específicos, um de 21 genótipos de *S. officinarum* e o outro de 13 genótipos de *S. spontaneum*. Os indivíduos desses policruzamentos foram cruzados entre si, gerando híbridos de *S. officinarum* x *S. spontaneum*, que foram divididos em grupos de alto e baixo Brix. Para as hibridizações, plantas de 10 meses de idade foram coletadas e um pool de 8 indivíduos foi utilizado para as extrações de RNA total.

3.2.2. Plantas utilizadas para FL-RNAseq

Para as análises de sequenciamento em larga escala (FL-RNAseq), mini-toletes de cada um dos genótipos *S. officinarum* (caiana listrada) e *S. spontaneum* (IN8458) foram enviados pelo professor Dr. Márcio Barbosa (Departamento de Fitotecnia, Universidade Federal de Viçosa), plantados em vasos e mantidos em sala de cultura no Instituto de Química – USP, até a data da coleta das amostras (5 meses de idade). Amostras de folhas de 5 plantas de cada genótipo foram utilizadas para a construção de uma primeira rodada de bibliotecas, as quais foram sequenciadas apenas no 454 (Roche). Posteriormente, essas plantas foram levadas para o CCA-UFSCAR e mantidas em casa de vegetação. Para uma segunda rodada de construção de bibliotecas, 6 meses após a primeira, amostras de folhas das mesmas plantas, agora mantidas em casa de vegetação e com 11 meses de idade, foram coletadas da mesma maneira anterior. Ainda, também foram coletadas amostras da variedade comercial SP803280 mantidas em campo, cedidas pela professora Dra. Monalisa Sampaio (CCA-UFSCAR). Dessa maneira, a primeira construção das bibliotecas de cDNA *full length* foi realizada utilizando amostras de um *pool* de folhas de 5 plantas por genótipo de *S.*

officinarum e *S. spontaneum* com 5 meses de idade mantidas em sala de cultura. A segunda rodada de construção de bibliotecas foi realizada a partir um *pool* de 5 indivíduos por genótipo de plantas mantidas em casa de vegetação com 11 meses de idade, com exceção das bibliotecas de SP803280, que foram produzidas também a partir de amostras de 5 indivíduos porém mantidos em campo e com 9 meses de idade. Para as espécies *S. officinarum* e *S. spontaneum*, foram construídas apenas bibliotecas a partir de folha, enquanto para a SP803280 foram construídas bibliotecas de tecidos de folha, entrenó imaturo e intermediário.

3.3. Medida de Brix e análise de crescimento

O conteúdo de Brix, altura da planta e massa do colmo foram determinados para 3 réplicas biológicas de cada genótipo. O caldo foi extraído com auxílio de um perfurador manual, no terço médio do colmo, e o Brix foi medido através da leitura em refratômetro portátil (N1 model, ATAGO, Japan). A altura da planta foi determinada utilizando-se uma trena da base do colmo rente ao solo até a última bainha foliar visível. A massa do colmo foi determinada em balança após a remoção de todas as folhas. O número de colmos foi contado na linha de 5 metros onde cada genótipo estava plantado. Como havia apenas uma linha plantada para cada genótipo, não foi possível realizar análise estatística para o número de colmos. O diâmetro dos entrenós foi medido no entrenó maduro com auxílio de um paquímetro.

3.4. Taxa de fotossíntese e transpiração

As medidas de fotossíntese, condutância estomática e transpiração foram realizadas na porção mediana da folha+1 utilizando-se o sistema de fotossíntese portátil *LCi* (ADC Bioscientific) que mede a assimilação de CO₂ pelo sistema de infravermelho (*InfraRed Gas Analyser*, IRGA). As medidas foram realizadas em campo entre 11h da manhã e 13h da tarde. Luz solar foi utilizada como fonte de luz. Um total de 10 medidas por genótipo foi realizado, sendo duas réplicas técnicas para 5 réplicas biológicas.

3.5. Conteúdo de água

Cerca de 200-300 mg de amostra congelada já macerada foram pesados em balança analítica e o valor foi anotado com precisão de quatro casas decimais. As amostras foram liofilizadas e pesadas periodicamente até atingir peso constante. O conteúdo de água da amostra foi determinado calculando a diferença entre o valor de massa inicial (massa fresca) e final (massa seca).

3.6. Análise de açúcares solúveis

Cerca de 10 mg de amostra seca já pulverizada foram pesados em um tubo de microcentrífuga de 2 ml e os açúcares solúveis foram extraídos por incubação das amostras a 80°C com 1,5 ml de etanol 80% por 20 minutos. Após a incubação, as amostras foram centrifugadas por 10 minutos a 12000 x g e o sobrenadante foi coletado e armazenado. Este procedimento foi repetido até um total de 5 lavagens, sempre coletando o sobrenadante. Seguiu-se a secagem do sobrenadante sob vácuo em concentrador de amostras *SpeedVac*

Concentrator System (Thermo Scientific). As amostras foram ressuspensas em 1 mL água deionizada e, para a retirada de pigmentos lipossolúveis, foi adicionado 0,5 mL de clorofórmio às amostras de folha apenas. Após 5 minutos, estas amostras foram centrifugadas ($12.000 \times g$ por 5 minutos) e a fase aquosa (superior) separada para análise. Alíquotas de cada uma das amostras foram analisadas por cromatografia de troca iônica de alta performance (HPAEC/PAD), utilizando-se uma coluna Carbowax PA1 eluída com hidróxido de sódio 100 mM com fluxo constante de 1 mL/min em sistema *Dionex-ICS3000*[®] (de Souza *et al.* 2013). Padrões de glicose, frutose e sacarose em concentração de 50, 100 e 200 μ M foram utilizados para cálculo das quantidades equivalentes nas amostras. O valor foi então corrigido pela a quantidade inicial exata de matéria seca.

3.7. Quantificação de lignina

A quantificação de lignina foi realizada como prestação de serviço no exterior, conduzida no *Complex Carbohydrate Research Center* na Universidade da Georgia (Athens, Georgia, EUA). As amostras liofilizadas foram pesadas (4-5 mg) e pirolizadas utilizando a 500°C e os resíduos foram analisados com um espectrômetro de massa por feixe molecular (*Molecular Beam Mass Spectrometer*), onde os resíduos de lignina foram identificados por seu perfil de massa. Cada amostra foi analisada em duplicata. Os valores brutos de quantificação de lignina foram calculados utilizando análise multivariada (análise de componentes principais, PCA) e corrigidas com uma amostra padrão de bagaço de cana-de-açúcar (lignina=24,4% da massa seca).

3.8. Histoquímica

Para análises de anatomia vegetal, amostras de 3-4 cm da parte mediana dos entrenós intermediário e maduro foram cortadas longitudinalmente em quatro pedaços e fixados em solução de FAA50 (formaldeído PA, ácido acético glacial e etanol 50%; 5:5:90, v/v) sob vácuo por 48 horas, seguindo-se a lavagem com água destilada para remoção do excesso de fixador. As amostras foram então emblocadas em resina hidrossolúvel de PEG (polietilenoglicol) e secções transversais de 40-60 µm foram cortadas em micrótomo de deslize *Leica SM2000R* (Leica Biosystems). Para detecção de lignina, os cortes foram tratados com solução de floroglucinol 1% (m/v) em etanol 95% (v/v) e seguido de solução de HCl 25% (v/v). Para detecção de unidades de siringil (reação de Maüle), os cortes foram corados com KMnO_4 1% (m/v) por 4 minutos, lavados com água destilada, tratados com HCl 37% (v/v) por 1 minuto, novamente lavados em água destilada e a lâmina montada em NH_4OH concentrado. As fotos foram tiradas em microscópio de luz *Olympus BX51* e câmera digital *Olympus Evolt E-330* em até 10 minutos após a coloração. Também foram utilizadas três réplicas biológicas, sendo que as imagens mostradas nas Figuras 12 e 13 são fotos representativas das três réplicas.

3.9. Gráficos e análises estatísticas

Todos os gráficos foram gerados pelo programa *GraphPad Prism 5* onde também foi realizada a estatística dos dados. Todos os resultados foram submetidos à análise de variância simples (*one way ANOVA*) seguido de teste Tukey ($p < 0,05$), para os dados de qPCR, carboidratos e fisiologia, e teste Student-Newman-Keuls ($p < 0,05$) para os dados de lignina.

Para a confirmação dos dados de microarranjo por qPCR, a expressão do gene foi avaliada para o par RB867515 (referência) e genótipo ancestral por teste-t com $p < 0,05$.

3.10. Extração de RNA

Três réplicas biológicas foram utilizadas para os estudos de expressão gênica. Amostras de tecido de folha+1, entrenó imaturo, intermediário e maduro foram coletadas e imediatamente congeladas em nitrogênio líquido e mantidas em gelo seco até a transferência para freezer -80°C . As amostras congeladas foram maceradas com auxílio de almofariz e pistilo ou utilizando o moinho de bolas TE-350 (Tecnal). A extração de RNA total foi realizada utilizando o reagente *Trizol* (Invitrogen) de acordo com informações do fabricante. Para cada 100 mg de amostra foi adicionado 1 ml de *Trizol* e homogeneizado em vórtex por 5 minutos a temperatura ambiente. Adicionou-se 200 μl de clorofórmio em cada amostra, homogeneizando-as por inversão, seguido de incubação à temperatura ambiente por 3 minutos. Após esse passo, as amostras foram centrifugadas por 12000 x g por 15 minutos a 4°C . A fase aquosa (superior) foi transferida para um novo tubo contendo 500 μl de isopropanol. Após 10 minutos à temperatura ambiente, as amostras foram centrifugadas a 10000 x g por 10 minutos a 4°C . O sobrenadante foi descartado e o RNA precipitado foi lavado com etanol 75%, seguido de centrifugação a 7500 x g por 5 minutos a 4°C . O sobrenadante foi novamente descartado e o RNA precipitado foi mantido à temperatura ambiente para secagem. O RNA foi ressuscitado em 20 μl de água tratada com DEPC (dietilpirocarbonato) e a concentração foi determinada em *NanoDrop* (Thermo Scientific). Para eliminar contaminações com DNA genômico, as amostras foram então tratadas com DNase I (Invitrogen) por 15 minutos a temperatura ambiente, utilizando 1 U de enzima para

cada 1 µg de RNA total. A reação foi parada adicionando-se EDTA para concentração final de 2,25 mM e a enzima inativada por incubação a 65°C por 10 minutos. O RNA então foi purificado utilizando-se o *RNeasy mini kit* (Qiagen) de acordo com instruções do fabricante. A integridade do RNA foi checada por eletroforese no *Agilent 2100 Bioanalyzer* (Agilent Technologies) com o *Agilent RNA 6000 Pico Kit* (Agilent Technologies) checando o *RNA Integrity Number* (RIN). Para a construção das bibliotecas de cDNA, devido a necessidade de uma maior quantidade de RNA total (1 mg), partiu-se de 5-10 g de tecido e a extração foi realizada da mesma forma descrita acima, porém o RNA precipitado foi mantido em etanol 75% para envio via fedex para a Life Technologies (Carlsbad, CA, EUA) onde o RNA foi ressuspendido, checado quanto a sua integridade e utilizado para a construção das bibliotecas.

3.11. Hibridização em chip de oligonucleotídeos da plataforma Agilent

A marcação do cRNA com Cy3 e Cy5 para as hibridizações no *chip* de oligos customizados (Lembke *et al.* 2012) foi realizada com *Agilent Low RNA Input Fluorescent Linear Amplification kit* (Agilent Technologies), seguindo-se as instruções do fabricante, partindo de 200 ng de RNA total. Para as reações de amplificação e marcação do cRNA, foi utilizado um controle *spike* de mRNA sintetizado *in vitro* (*Agilent RNA Spike-In kit*). As hibridizações foram conduzidas com duas réplicas biológicas no *chip* de oligos customizados 44k (Agilent Technologies). Após as hibridizações, os *chips* foram lavados de acordo com o protocolo de hibridização desenvolvido pelo fabricante, escaneadas no *GenePix 4000B scanner* (Molecular Devices) e os dados das imagens extraídos com auxílio do programa *Feature Extraction 9.5.3* (Agilent Technologies) utilizando o referencial para microarranjo de

duas cores da plataforma Agilent (Kerr 2007). Os dados da intensidade derivada de Cy3 e Cy5 de uma mesma amostra foram corrigidos e normalizados utilizando-se a função Lowess (Yang *et al.* 2002) através de uma ferramenta do ambiente R do nosso servidor e levando-se em consideração os sinais dos controles *spikes*. Os genes diferencialmente expressos foram definidos através do método do HT-self (Vencio & Koide 2005), modificado de (Rocha *et al.* 2007) e adaptado à plataforma do microarranjo de oligonucleotídeos Agilent e ao *Feature Extraction*.

3.12. PCR em tempo real

A síntese de cDNA foi realizada a partir de 1 µg de RNA total utilizando o *SuperScript III First-Strand Synthesys SuperMix* (Invitrogen) com oligo(dT) de acordo com instruções do fabricante. Para as validações do nível de expressão de genes identificados como tendo transcritos nas orientações senso e antissenso, a síntese de cDNA foi realizada de forma fita específica, utilizando o mesmo kit citado acima. Para síntese da fita de cDNA do transcrito senso, foi utilizado 2 µM do *primer* com sequência antissenso e o *primer* senso para o transcrito antissenso, como descrito por Lembke *et al* (2012). As reações de PCR em tempo real (qPCR) foram realizadas com o *Fast SYBR Green Master Mix* (Applied Biosystems) no equipamento *7500 Fast Real-Time PCR System* (Applied Biosystems) de acordo com protocolo padrão. Todos os *primers* foram desenhados com auxílio do programa *Primer Express 2.0* (Applied Biosystems) e analisados quanto a eficiência de amplificação, sendo que apenas os *primers* com eficiência entre 90 e 110% foram utilizados. A especificidade da reação foi verificada pela curva de dissociação dos produtos de amplificação gerada pelo equipamento. Os resultados foram analisados como descrito por Hellemans *et al* (2007) com

auxílio do programa *qBase 2.0* (Biogazelle) utilizando como fator de normalização a média geométrica de, no mínimo, dois genes endógenos mais estáveis. Os genes actina (ACT, SCCCLR1069D05.g), subunidade delta da ATP sintase mitocondrial (SCCCLR1072A03.g), subunidade 60S do ribossomo (60S, SCJFRZ2009G01.g), gliceraldeído-3-fosfato desidrogenase (GAPDH, (Iskandar *et al.* 2004)), tubulina (TUB, SCCRZ1002H03.g), enzima E2 de ubiquitinação (UBE2, SCBGLR1002D06.g), poliubiquitina (PUB, SCCST2001G02.g) e ubiquitina (UB, SCCCLR1048F12.g) foram testados para identificação de genes normalizadores endógenos com expressão homogênea entre as amostras. Para maior acurácia dos dados, em cada análise foram selecionados os genes endógenos mais apropriados, identificados pelo programa *geNorm* (que compõe o programa *qBase*), tendo, portanto, um fator de normalização diferente para cada caso específico. Para as validações de dados de microarranjo (plantas com 9 meses, coleta julho de 2011), as reações de qPCR foram realizadas numa terceira réplica biológica. Já para as análises de expressão apenas por qPCR (plantas com 7 meses, coleta março 2012) foi utilizada três réplicas biológicas e, após a normalização no programa *qBase*, seguiu-se a normalização para réplicas biológicas para cada gene alvo em cada tecido, utilizando a média das amostras (*mean centering*) como descrito por Willems *et al* (2008). Devido a isso, o nível de expressão de genes diferentes não pode ser comparado.

3.13. Clusterização dos dados de expressão

Foram utilizados os valores de log2 da intensidade do sinal de microarranjo para as análises de *clusterização* hierárquica. Todas as análises foram geradas pelo programa MEV (*Multi Experiment Viewer*, versão 4.8.1 (Saeed *et al.* 2003)). Primeiramente, foi definido o

número ideal de *clusters* com auxílio da ferramenta *Figure of Merit*, que compõe o programa MEV, com configurações *default*. O gráfico gerado indica o número ideal de *clusters* pela identificação do ponto onde a inclinação da reta se aproxima de zero. Em seguida, os *clusters* foram gerados com a ferramenta *k-means* do programa MEV, com configuração *default*, com exceção da opção do número de *clusters*, definido pela *Figure of Merit*. Posteriormente, os sinais de intensidade de cada gene para os quatro genótipos foram normalizados com a opção *normalize gene/row*.

3.14. Redes de coexpressão de genes

Das diversas ferramentas disponíveis para esse tipo de análise, nenhuma utiliza dados de expressão de cana-de-açúcar ou permite ao usuário utilizar seu próprio conjunto de dados. Assim, foi escolhida a ferramenta *Genevestigator* (Hruz *et al.* 2008) utilizando dados de expressão de milho (*Zea mays*), por ser a espécie mais próxima de cana-de-açúcar com esse tipo de ferramenta disponível. A primeira etapa para essa análise foi a identificação do homólogo de milho do SAS de interesse, SCCCCL3002A03.b (4-coumarato:CoA ligase, 4CL). Com a ferramenta “*Search gene VS Homologs*” disponível em sucest-fun.org/wsapp, que utiliza o programa Inparanoid (Ostlund *et al.* 2010) para identificar homologia entre sequências, foi identificado o transcrito de milho GRMZM2G075333, o qual foi utilizado como “isca” para buscar os genes coexpressos na base de dados de milho. O *Genevestigator* utiliza dados públicos de expressão de 562 experimentos realizados no chip *Maize Genome 15K* da plataforma *Affymetrix* e calcula a correlação de Pearson entre os valores de sinal de expressão de cada par de genes. Todos os experimentos presentes no banco de dados do *Genevestigator* foram curados e

normalizados para permitir a realização de uma metanálise com diversos experimentos. Para a escolha da base de dados a ser utilizada, foi selecionada a opção “*Anatomy*” do *Genevestigator*, que analisa todos os genes que apresentam sinal de expressão, identificando, para cada par de genes, quais possuem um mesmo perfil de expressão em cada tecido da planta. Essa opção exclui experimentos de resposta a estresses e de plantas mutantes e transgênicas. Quanto mais semelhante for o perfil de expressão de dois genes nos diferentes tecidos, maior será a correlação entre eles, o que os credencia a estarem envolvidos num mesmo processo biológico. Assim, essa análise identificará todos os genes cuja expressão varia de forma semelhante em vários tecidos, calculando uma correlação de Pearson para o sinal de expressão de cada par de genes. Em nossa análise, após a geração da rede de genes, foram considerados apenas os genes que possuem correlação de Pearson, em relação à “isca” GRMZM2G075333 e mútua entre todos os genes da lista, acima de 0,8.

3.15. Identificação da região promotora e cis-elementos regulatórios associados ao metabolismo de parede celular

Inicialmente, os SAS de interesse foram alinhados por BlastN (Altschul *et al.* 1990) contra o genoma de sorgo (v2.1, www.phytozome.net) e, nos casos onde o SAS não continha a sequência completa do transcrito, o transcrito desta espécie com maior similaridade ao SAS foi utilizado como referência para mapear a localização e estimar o tamanho das UTRs e da ORF do SAS correspondente. Em seguida, os SAS foram alinhados por BlastN contra os *contigs* e *singletons* obtidos pelo sequenciamento de BACs e *whole genome shotgun* da variedade de cana-de-açúcar SP8032080, armazenados em sucest-fun.org/wsapp. O banco

de dados possui 32 Mb sequenciados a partir de 280 BACs e 9,6 Gb de sequência do *whole genome shotgun* (aproximadamente 1 milhão de *contigs* e 3,4 milhões de *singletons*). Com o alinhamento múltiplo do SAS, transcrito de sorgo e *contig* do genoma, foi possível identificar nos *contigs* a sequência imediatamente à montante da 5'-UTR do transcrito de cana-de-açúcar, que compreende o possível promotor. Para transcritos que apresentaram expressão do antissenso, também foi pesquisada a sequência imediatamente à jusante da 3'-UTR do transcrito senso, a qual seria o 5'-UTR do antissenso. O tamanho das sequências promotoras foi limitado em 1,5 kb e em alguns casos não foi possível identificá-las.

Os promotores foram então submetidos à análise para identificação dos cis-elementos específicos de parede celulares já caracterizados em outras espécies: SNBEs (Zhong *et al.* 2010), SMRE (Zhong & Ye 2012) e M46RE (Kim *et al.* 2012). Inicialmente, a matriz para busca dos cis-elementos foi gerada pelo programa TOMTOM (<http://meme.nbcr.net/meme/cgi-bin/tomtom.cgi>) que utiliza como *input* a sequência consenso do DNA e gera como resultado a matriz referente à sequência do *input*. As matrizes de cada cis-elemento geradas pelo TOMTOM foram utilizadas na ferramenta FIMO (Grant *et al.* 2011), disponível online em <http://meme.nbcr.net/meme/cgi-bin/fimo.cgi>, para analisar as sequências promotoras em busca dos SNBEs, SMREs e M46RE, utilizando $p < 0,01$.

3.16. Construção de bibliotecas de cDNA e seqüenciamento (FL-RNAseq)

A construção e as análises das bibliotecas de cDNA foram conduzidas em colaboração com a Life Technologies (Carlsbad, CA, EUA). O RNA total extraído foi enviado para a Life Technologies, onde foram realizadas as etapas de isolamento de mRNA, síntese de cDNA, produção das bibliotecas, clonagem e sequenciamento na plataforma *Ion Personal Genome*

Machine (PGM). O sequenciamento na plataforma *Genome Sequencer FLX System* (454 – Roche) foi conduzido no IQ-USP.

A partir de 1-2 mg de RNA total, os mRNAs com cauda poli-A foram isolados utilizando-se o *FastTrack MG mRNA Isolation kit* (Life Technologies). As bibliotecas *full length* foram construídas com o *SuperScript Full-Lenght cDNA Library Construction Kit* (Life Technologies) seguindo-se as instruções do fabricante com pequenas modificações. Resumidamente, a primeira fita de cDNA foi sintetizada utilizando-se o *primer* attB2-dT19 contendo uma sequência específica (*barcode*) para cada uma das cinco bibliotecas ao invés do *primer* padrão. Após a síntese de cDNA, um adaptador contendo o sítio de recombinação attB1 foi ligado à extremidade 3' do cDNA para permitir a clonagem por recombinação (Gateway, Invitrogen). Quantidades iguais de cada biblioteca foram reunidas em uma única amostra para os procedimentos posteriores. A biblioteca física, clonada em *E. coli*, foi construída no vetor de entrada pENTR222 utilizando o sistema Gateway (Invitrogen), de acordo com protocolo do fabricante, que utiliza os sítios de recombinação attB1 e attB2, e posteriormente transferidas para vetor de destino pCMVSPORT6.1, também utilizando o sistema Gateway.

De cada amostra, foi utilizada uma alíquota de 10 µg de mRNA com cauda poli-A, isolado com o kit descrito acima, para a síntese da primeira fita de cDNA com o kit *SuperScript III* (Life Technologies) e o *primer* oligo-dT21VN. Para enriquecer o cDNA *full length*, o híbrido mRNA:cDNA foi tratado com RNase I, que degrada especificamente RNA fita simples, seguido de seleção por anticorpo específico para a estrutura Cap-5' do mRNA (*Cap antibody selection*, Life Technologies) de acordo com as instruções do fabricante.

A primeira fita de cDNA foi utilizada para o sequenciamento no Ion PGM. A biblioteca foi preparada utilizando-se o kit Ion-Total RNA-seq (Life Technologies) de acordo com o

protocolo versão 2 para preparo de bibliotecas para *reads* de 300 bases na plataforma PGM. As partículas monoclonais *ion sphere particles* (ISPs) foram produzidas e enriquecidas usando o sistema *One Touch 2* seguindo-se o protocolo base *Ion PGM Template OT2 400*. As ISPs fita molde positiva enriquecidas de cada uma das amostras foram corridas no *chip* Ion 318 de acordo com instruções do fabricante, totalizando cinco corridas de sequenciamento.

Para o sequenciamento na plataforma *Genome Sequencer FLX System* (454), a segunda fita de cDNA foi sintetizada com o kit *cDNA Synthesis* (Life Technologies). O cDNA fita dupla enriquecido em *full length* foi sonicado utilizando-se o sonificador *Covaris 220* para obter fragmentos de cDNA com tamanho entre 1 a 2 kb, os quais foram seqüenciados no 454 de acordo com as recomendações do fabricante. Foram utilizadas MIDtags para cada biblioteca e o sequenciamento realizado em três corridas; os *reads* de cada amostras foram separados posteriormente pelas sequências das MIDtags.

O sequenciamento dos insertos dos clones descrito na Tabela 21 foi realizado com o kit *Big Dye Terminator 3.1* (Applied Biosystems) e o sequenciador automático *ABI Prism 3100 Genetic Analyser* (Applied Biosystems), de acordo com as instruções do fabricante. Os *primers* utilizados nas reações de sequenciamento foram o M13F e M13R. Os 30 clones selecionados aleatoriamente e sequenciados foram analisados no programa *BioEdit Sequence Alignment Editor* (Hall 1999) e posteriormente alinhados por BlastX contra os bancos de dados de proteínas *full length* de trigo (Mochida *et al.* 2009) e sorgo (V1.4, <http://www.phytozome.net/>) e também contra banco de dados do NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Todas as análises dos dados de sequenciamento no 454 e PGM foram conduzidas pelo bioinformata Milton Yutaka. A primeira etapa foi a *trimagem* dos dados de sequenciamento, utilizando os parâmetros de qualidade $Q < 20$ e tamanho da sequência $L < 70$

para os *reads* do 454 e $Q < 15$ com qualidade média mínima de $Q < 20$, tamanho de sequência $L < 50$ e poli-A/T para os *reads* do PGM, e para ambas as plataformas o filtro por sequências de baixa complexidade com método *dust* e valor de *threshold* < 7 e *trimagem* das pontas com qualidade $Q < 20$, Poly-N's ≥ 5 . A montagem dos *contigs* foi realizada com o programa Trinity (Grabherr *et al.* 2011) utilizando todos os *reads* do PGM e 454 e com tamanho mínimo do *contig* ≥ 200 pb e parâmetros *default* de sobreposição e identidade. Após a montagem, os *contigs* foram mapeados por BlastN contra 6 espécies de gramíneas (*Zea mays*, *Sorghum bicolor*, *Oryza sativa*, *Setaria italica*, *Panicum virgatum* e *Brachypodium distachyon*) além do próprio SUCEST, utilizando critério de *cutoff* de *e-value* $\leq e^{-20}$. Nos casos onde mais de um *contig* alinhava contra um mesmo gene de uma das gramíneas, a porcentagem de cobertura foi determinada somando-se a cobertura final com algoritmo para concatenar regiões separadas em uma região única.

Para identificação da origem de cada *read*, a ferramenta *bowtie2* foi utilizada para mapear os *reads* do PGM nos *contigs* do Trinity, utilizando os parâmetros de alinhamento *sensitive-local* e *max penalty for mismatch* = 6.

A análise de expressão utilizou RPKM (Mortazavi *et al.* 2008), mapeando os *reads* do PGM contra os *contigs* para gerar um valor de expressão para cada *contig* em cada amostra. Os dados então foram analisados com o pacote RSEM (Li & Dewey 2011) para estimar os níveis de expressão dos genes e isoformas para os dados de RNA-Seq e os mais abundantes em cada amostra. Para identificar os genes diferencialmente expressos em pelo menos um par de amostras com intensidade da expressão pelo menos 4 vezes ($\log_2(4)=2$) diferencialmente expresso com significância estatística corrigida para falsa descoberta de pelo menos 0,001. Os dados submetidos à análise de *clusterização* hierárquica dado o conjunto de transcritos diferencialmente expressos, extraídos os valores de expressão

normalizado e agrupados por padrões de expressão entre as amostras nos parâmetros definidos acima.

A estratégia utilizada para a análise de antissenso foi mapear os *reads* de cada biblioteca com o *software bowtie2* com parâmetro *very-sensitive* contra os genes das 6 espécies de gramíneas. Baseado no arquivo de mapeamento dos genes nos respectivos genomas (gff3), é extraída a informação da orientação do gene. Assim, se o hit de um *read* no gene for nas orientações *forward/forward* ou *reverse/reverse* este é considerado senso; se o hit for *reverse/forward* ou *forward/reverse* o *read* é considerado antissenso. Para o SUCEST, primeiramente foi criada uma matriz de *codon-usage* de cana-de-açúcar com auxílio do programa ESTScan para a predição das proteínas, gerando 42524 SAS únicos com predição, 66 predições duplicadas e 617 SAS sem predição. A orientação da ORF foi definida e seguiu-se análise descrita acima.

A anotação baseada no Gene Ontology (Ashburner *et al.* 2000) foi conduzida mapeando-se os *reads* do FL-RNAseq contra o banco de dados do Gene Ontology (GO) com o *software blastX*, critério de *cutoff* de *e-value* $\leq 1e-08$ para identificar as categorias associadas.

4. RESULTADOS

4.1. Construção de um catálogo de genes de parede celular para cana-de-açúcar

Para a identificação de genes do metabolismo de parede celular de cana-de-açúcar, o banco de ESTs do SUCEST (Vettore *et al.* 2001; Vettore *et al.* 2003) foi utilizado como referência para as sequências expressas de cana-de-açúcar. Como descrito na introdução, o SUCEST possui 43.141 sequências (SAS) geradas a partir de 26 bibliotecas de cDNA que compreendiam diferentes tecidos como folhas, flores, gema lateral, colmo, raiz, entre outros, de 13 variedades diferentes, sendo a maioria das bibliotecas produzida a partir da variedade SP803280 (Vettore *et al.* 2001). Apesar de ter sido realizada a anotação automática dos SAS, a identificação e categorização dos genes relacionados ao metabolismo de parede celular ainda se encontravam pouco organizadas. Então, o SUCEST foi cruzado por BlastX com os bancos de dados de genes de parede celular de milho, *Maizewall* (Guillaumie *et al.* 2007), e *Arabidopsis*, *Cell Wall Genomics* (Yong *et al.* 2005), para identificar os genes relacionados ao metabolismo de parede celular em cana-de-açúcar. O *Maizewall* foi gerado a partir de buscas com aproximadamente 100 palavras-chave relacionadas à parede celular nos bancos de dados do NCBI, incluindo outras espécies além de milho, e também a partir de genes diferencialmente expressos em *Zinnia elegans* sob deposição de parede celular secundária (Pesquet *et al.* 2005). Os genes identificados foram alinhados por Blast contra banco de dados de milho identificando-se 735 genes, classificados em 19 categorias funcionais (Guillaumie *et al.* 2007). O *Cell Wall Genomics* reúne informação de diversos estudos que já identificaram genes de parede celular de *Arabidopsis*, principalmente pela caracterização de mutantes, e apresenta mais de 1.000 genes divididos em 6 categorias

principais: produção de substratos; síntese de polissacarídeos; secreção; montagem, arquitetura e crescimento; diferenciação e formação de parede secundária; sinalização e mecanismos de resposta (<http://cellwall.genomics.purdue.edu/families/index.html>). A comparação do SUCEST com o *Maizewall* e *Cell Wall Genomics* identificou mais de 6 mil sequências prováveis e todas foram individualmente checadas por BlastX contra os bancos de dados NR (<http://blast.ncbi.nlm.nih.gov/>), Uniprot (<http://www.uniprot.org/>) e também analisadas para a presença de domínios conservados usando o Pfam (<http://pfam.sanger.ac.uk/>). Após essa filtragem, foi possível identificar 1606 sequências únicas relacionadas ao metabolismo de parede celular em cana-de-açúcar.

As 1606 sequências foram separadas em dois grupos (Tabela 2) de acordo com as informações obtidas sobre as sequências bem como a anotação do gene: (i) cell-wall related, composto por 732 sequências e inclui genes envolvidos na biossíntese, montagem e degradação da parede celular, como, por exemplo, celulose sintase, expansinas, XTHs e genes da biossíntese de lignina; (ii) probable, que inclui 874 SAS que provavelmente estão envolvidos no metabolismo de parede celular, mas também podem fazer parte de outras vias/rotas, como algumas glicosiltransferases ou sequências com alta similaridade com genes de outras espécies nas quais estes já foram caracterizados experimentalmente e evidenciado suas funções no metabolismo de parede celular, como os fatores de transcrição da família MYB e NAC, mas sua função em gramíneas ainda não foi demonstrada. Em ambos os grupos, os genes identificados podem ser divididos em categorias de acordo com a função. Genes de metabolismo de carboidrato representam cerca de metade do catálogo, incluindo hidrolases, liases, esterases, glicosiltransferases processivas, como Cesa e Csl, glicosiltransferases não processivas e enzimas que atuam na conversão de nucleotídeos de açúcar. Também estão presentes as categorias de genes de biossíntese de lignina, proteínas

envolvidas no crescimento celular, transportadores, fatores de transcrição e quinases, proteínas estruturais de parede, entre outros.

Tabela 2. Catálogo de genes de parede celular de cana-de-açúcar.

Grupo	Função	Número de SAS
Cell Wall Related (732)	Cell wall differentiation	351
	Cell wall growth/extension	61
	Polysaccharide biosynthesis	112
	Structure proteins of the cell wall	25
	Lignin metabolism	176
	Unknown function	5
Probable (874)	Fructose/Mannose/Galactose	37
	Nucleotide sugar synthesis, conversion and transport	58
	Sucrose/trehalose/sorbitol and others mono-oligosaccharides	47
	Several pathways	274
	Lignin metabolism	152
	Multiple functions	10
	Transcription Factor	114
	Kinase	39
	Membrane transporters	105
Total		1606

Uma vez que o número estimado de genes em cana-de-açúcar está em torno de 35.000 e especula-se que as plantas dedicam 10% de seu genoma para genes envolvidos no metabolismo de parede celular (Yong *et al.* 2005), o número de genes identificado nesse catálogo está abaixo do esperado. Isso pode ser devido ao número de ESTs presentes no SUCEST não abranger todos os genes de cana-de-açúcar. Além disso, vale lembrar que também foi utilizado para essa análise o banco de genes de Arabidopsis, e acredita-se que gramíneas possam ter um grande número de genes de parede celular sem ortólogos em Arabidopsis (Carpita & McCann 2008). Assim, é muito provável que diversos genes não foram identificados por essa análise. Por outro lado, o catálogo de cana-de-açúcar possui mais do dobro de sequências do *Maizewall* (1606 contra 735) e número similar ao do *Cell*

wall Genomics, mostrando que, mesmo não compreendendo o número ideal de genes, a abordagem foi eficaz na identificação dos alvos.

Definido o catálogo, este servirá de base para os estudos de transcriptoma para direcionar as análises de genes de parede celular, visando entender sua biossíntese e regulação. Esse conhecimento gerado pode representar um grande avanço para a produção de novas variedades de cana-de-açúcar com maior teor de biomassa (cana energia), ou biomassa menos recalcitrante, favorecendo sua utilização em processos de produção de biocombustíveis.

4.2. Experimento de campo

As plantas utilizadas nesse trabalho foram mantidas no campo experimental da Universidade Federal de São Carlos, campus Araras. Os alvos do estudo foram genótipos de *S. officinarum*, *S. spontaneum*, *S. robustum* e RB867515 (Figura 7) os quais eram mantidos no campo plantados em uma linha de 5 metros. A RB867515 foi escolhida pelo fato de já ter sido identificada uma diferença significativa no teor de lignina em relação à *S. spontaneum* em plantas mantidas em casa de vegetação (Poelking 2012), além de ser uma das variedades mais plantadas no Brasil, correspondendo a cerca de 20-25% da área plantada com cana-de-açúcar no país (Dal-Bianco *et al.* 2012). As plantas foram coletadas com 9 meses de idade, em julho de 2011, e posteriormente com 7 meses de idade, em março de 2012. O plano inicial previa três coletas ao longo do ano de 2011, mas o campo experimental foi cortado antes do previsto, possibilitando apenas uma coleta em 2011, o que nos levou a realizar uma segunda coleta apenas no ano seguinte. As coletas incluíram, além de amostras para transcriptoma, dados fisiológicos e morfológicos para a caracterização de alguns parâmetros

agronômicos de interesse, como por exemplo, altura, Brix, fotossíntese e número de plantas na linha central. Entretanto, devido a problemas com o armazenamento das amostras e durante a coleta, como o número de réplicas nas análises fotossintéticas, algumas análises da coleta de julho de 2011, plantas com 9 meses de idade, ficaram prejudicadas e, por isso, apenas os dados morfo-fisiológicos da coleta de março de 2012, das plantas com 7 meses de idade, serão apresentados.

Figura 7. Genótipos em estudo.

4.3. Dados Agronômicos

Os quatro genótipos analisados, os ancestrais *S. officinarum*, *S. robustum* e *S. spontaneum* e o híbrido comercial RB867515, apresentam características bem distintas. Historicamente já se sabe que os genótipos considerados selvagens, *S. robustum* e *S. spontaneum*, apresentam maior resistência a estresses bióticos e abióticos, maior teor de

fibra e pouca sacarose, enquanto que *S. officinarum* possui baixa resistência a estresses, mas tem alto teor de sacarose, apresentando plantas geralmente mais baixas e restritas a poucos habitats (Paterson *et al.* 2013). Mesmo possuindo algumas características similares a *S. robustum*, *S. spontaneum* apresenta características tão distintas que já foi proposto que este genótipo seria uma espécie em separado, enquanto *S. robustum*, *S. officinarum*, *S. edule*, *S. barberi* e *S. sinense* seriam todas representantes de uma única espécie (Irvine 1999).

Para análises mais criteriosas das diferenças entre os genótipos, foram medidos os principais parâmetros agronômicos de interesse para posteriormente poder relacioná-los com os dados moleculares. Assim, foram medidos parâmetros de crescimento, como altura da planta, diâmetro e massa do colmo, teor de sacarose e açúcares, número de plantas na linha central e também alguns parâmetros fotossintéticos. *S. officinarum* e RB867515 possuem colmos grossos (Tabela 3) e alto Brix (Figura 8A), enquanto *S. spontaneum* e *S. robustum* possuem colmos finos (Tabela 3), baixo Brix (Figura 8A) e praticamente não produzem caldo após a moagem do colmo (dados não mostrados), principalmente *S. spontaneum*. Brix é uma medida de sólidos solúveis, sendo uma quantificação indireta de sacarose. O conteúdo de água do colmo também foi analisado e *S. spontaneum* se destaca das demais com o menor valor (Tabela 3). Em relação à produtividade, RB867515 e *S. spontaneum* possuem plantas altas (Figura 8B), mas diferem entre si pelo fato de que a RB867515 acumula muito mais biomassa fresca por colmo (Figura 8C). Por outro lado, *S. spontaneum* perfilha muito mais, quase 4 vezes mais que a variedade comercial numa mesma área plantada (Figura 8D).

Tabela 3. Diâmetro (mm) e conteúdo de água (%) do entrenó maduro.

Genótipo	Diâmetro (mm)	% de água
RB867515	20,15 ± 0,15 (a)	83,7 ± 0,8 (ab)
<i>S. officinarum</i>	25,94 ± 1,34 (b)	87,0 ± 0,8 (a)
<i>S. robustum</i>	15,22 ± 0,10 (c)	78,8 ± 0,8 (b)
<i>S. spontaneum</i>	8,75 ± 0,16 (d)	63,1 ± 2,5 (c)

Letras diferentes representam médias diferentes pela análise com OneWay ANOVA seguido de teste Tukey. N=3, Erro=Erro padrão, $p < 0,05$.

Figura 8. Dados agrônômicos. A, Brix; B, altura da planta; C, massa fresca do colmo e D, número de colmos (perfilamento) na linha central de 5 metros. Letras diferentes no topo da barra representam médias diferentes por One-way ANOVA seguido de teste Tukey. Barra de erro=erro padrão; N=3; $p < 0,05$.

4.4. Análise de parâmetros fotossintéticos

Décadas atrás, já foi reportado que *S. spontaneum* tem maior taxa de fotossíntese que *S. officinarum*, mostrando que fotossíntese e acúmulo de sacarose não estão sempre diretamente correlacionados (Irvine 1975). Por outro lado, foi demonstrado que *S. spontaneum* possui menor valor de condutância estomática que *S. officinarum* mas consegue ser mais resistente à seca por manter parte da condutividade hidráulica em situações de estresse severo (Neufeld *et al.* 1992). Assim, para verificar se as plantas analisadas neste trabalho apresentavam características semelhantes às descritas anteriormente, foram avaliados alguns parâmetros de trocas gasosas nas folhas, entre eles, assimilação de carbono (A), transpiração (E), condutância estomática (Gs) e eficiência do uso de água (WUE) (Figura 9). Diferentemente dos dados de Irvine (Irvine 1975), a assimilação de carbono foi menor em *S. spontaneum* (cerca de $15 \mu\text{mol. CO}_2.\text{m}^2.\text{s}^{-1}$) quando comparada aos outros genótipos estudados, os quais apresentaram valores similares, acima de $20 \mu\text{mol. CO}_2.\text{m}^2.\text{s}^{-1}$ (Figura 9A). Da mesma forma, a transpiração (Figura 9B) e a condutância estomática (Figura 9C) foram menores em *S. spontaneum* e similares nos outros genótipos, concordando com as observações de Neufeld (Neufeld *et al.* 1992) sobre condutância estomática. A menor taxa de transpiração resulta em uma melhor eficiência do uso de água em *S. spontaneum* (Figura 9D). Todos esses resultados (Figuras 8 e 9) mostram *S. spontaneum* que possui uma fisiologia diferente das demais.

Figura 9. Dados fotossintéticos. A, taxa fotossintética; B, transpiração; C, condutância estomática e D, eficiência no uso de água. Letras diferentes no topo da barra representam médias diferentes por One-way ANOVA seguido de teste Tukey. Barra de erro=erro padrão; N=10; $p < 0,05$.

4.5. Análise de carboidratos solúveis

Para aprofundar as análises de Brix, mostrando um panorama mais preciso do conteúdo de carboidratos das plantas, foi realizada uma análise de açúcares redutores (glicose e frutose), açúcares totais (glicose, frutose e sacarose) e sacarose (Figura 10). RB867515 e *S. spontaneum* apresentaram os menores valores para açúcares redutores nas folhas (Figura 10A). Em entrenó intermediário, houve um aumento dos açúcares redutores em relação aos da folha (Figura 10A). Em entrenó maduro, RB867515 e *S. officinarum*

apresentaram valores similares àqueles observados em entrenó intermediário, enquanto que *S. robustum* e *S. spontaneum* possuem um menor conteúdo de açúcares redutores (Figura 10A). Em entrenó maduro de RB867515 e *S. officinarum*, os níveis de açúcares redutores foram duas vezes maiores que em *S. robustum* e cinco vezes o observado em *S. spontaneum*. Os açúcares totais (Figura 10B) mostraram valores similares nas folhas para todos os genótipos e, em entrenó intermediário, *S. spontaneum* apresentou o menor valor. Em entrenó maduro (Figura 10B), o perfil geral é semelhante ao observado na medida de Brix (Figura 8A), confirmando este resultado. Para o teor de sacarose (Figura 10C), não foram identificadas diferenças significativas nas folhas entre os genótipos. RB867515 apresentou o maior valor de sacarose em entrenó maduro, seguido de *S. officinarum* e *S. robustum*, com valores próximos, e, por último, *S. spontaneum*. Vale ressaltar, também, que estas plantas ainda não estão na maturidade do ciclo e, portanto, essas diferenças se acentuarão à medida que a planta atinja o fim do ciclo. Curiosamente, a RB867515 mostrou um aumento significativamente maior no teor de sacarose em entrenó maduro em relação ao intermediário (Figura 10D), aumento superior ao observado para os mesmos entrenós de *S. officinarum*.

Figura 10. Quantificação de açúcares solúveis. A, açúcares redutores (glucose e fructose); B, açúcares redutores + sacarose; C, sacarose; D, sacarose, comparando-se tecidos diferentes do mesmo genótipo. Devido à pequena quantidade de tecido, não foi possível quantificar açúcares no entrenó imaturo. Barras brancas, cinzas e pretas indicam folha, entrenó intermediário e maduro, respectivamente. Letras diferentes no topo da barra representam médias diferentes por One-way ANOVA seguido de teste Tukey. Barra de erro=erro padrão; N=3; $p<0,05$.

4.6. Quantificação de lignina

Recentemente, foi demonstrado que *S. spontaneum* possui mais lignina que a RB867515 em diferentes entrenós, incluindo intermediário e maduro, em plantas cultivadas em casa de vegetação (Poelking 2012). Assim, foi verificado se esse resultado também é válido para plantas cultivadas em campo e se essa mesma diferença pode ser observada em relação aos outros genótipos ancestrais. Quatro tecidos diferentes foram analisados para o

conteúdo de lignina. Exceto por *S. officinarum*, o teor de lignina é maior em entrenó maduro em relação aos outros entrenós analisados, variando de 12,9 – 14,5% em entrenó imaturo a 14,8 – 19,2% em entrenó maduro (Figura 11A). Comparando-se os genótipos, não há diferenças significativas em folha, entretanto, em entrenó intermediário e maduro, *S. spontaneum* apresentou uma quantidade maior que as demais plantas, concordando com os resultados de plantas de casa de vegetação (Poelking 2012). Além disso, a composição da lignina difere entre os genótipos, sendo que *S. spontaneum* possui maior quantidade de unidades de siringil em ambos os entrenós, intermediário e maduro (Figura 11B). Ademais, o aumento na quantidade de unidades de siringil do entrenó intermediário para o entrenó maduro em *S. spontaneum* (cerca de 20%) é maior do que o observado para as demais plantas analisadas (cerca de 10%).

Figura 11. Quantificação de lignina. A, lignina total; B (próxima página), lignina S (siringil). Letras diferentes no topo da barra representam médias diferentes por One-way ANOVA seguido de teste Newman-Keuls. Barra de erro=erro padrão; N=3; $p<0,05$. Análise estatística foi realizada apenas para diferenças entre os genótipos para um determinado tecido e não para tecidos diferentes.

Figura 11. Continuação. Quantificação de lignina. A, lignina total; B, lignina S (siringil). Letras diferentes no topo da barra representam médias diferentes por One-way ANOVA seguido de teste Newman-Keuls. Barra de erro=erro padrão; N=3; $p < 0,05$. Análise estatística foi realizada apenas para diferenças entre os genótipos para um determinado tecido e não para tecidos diferentes.

4.7. Histoquímica

Com o objetivo de visualizar o padrão de deposição de lignina nos tecidos das espécies analisadas, foram feitos cortes histológicos corados com floroglucinol (Patten *et al.* 2005) e reagente de Maüle (Chapple *et al.* 1992). Ambos reagentes detectam fenilpropanoides na parede celular, tanto o ácido ferúlico, o ácido cinâmico e a lignina são detectáveis, com reação positiva visualizada pela coloração vermelho/púrpura, mas o reagente de Maüle é mais específico para siringil. Os padrões de deposição variam entre os genótipos. Em entrenó intermediário corados com floroglucinol (Figura 12A-D), *S. officinarum* mostrou deposição restrita aos elementos traqueais (meta e protoxilema) e pouca coloração nas fibras dos feixes vasculares (Figura 12B), enquanto que os demais

apresentaram deposição de lignina no xilema e em volta de todo o feixe vascular, sendo mais intenso em *S. robustum* e *S. spontaneum* (Figura 12C e D). Adicionalmente, é possível observar uma coloração mais intensa no parênquima medular de *S. spontaneum*, (Figura 12D, centro do colmo) e a quantidade de fibras associadas ao feixe vascular de *S. officinarum* é bem menor em comparação aos demais genótipos analisados (compare periferia do colmo na Figura 12B e na Figura 12 A, C e D). Além disso, *S. robustum* e *S. spontaneum* apresentaram lignificação mais intensa na epiderme e no parênquima subepidérmico (Figura 12C e D, periferia do colmo), mais destacada nesta última. Apesar de não ter coloração positiva, também é possível notar um espessamento da parede celular do parênquima cortical e subepidérmico em *S. spontaneum* (Figura 12D, periferia do colmo).

Figura 12. Histoquímica para detecção de lignina (em vermelho) em secções transversais do colmo, com reagente floroglucinol. A e E, RB867515; B e F, *S. officinarum*; C e G, *S. robustum*; e D e H, *S. spontaneum*. MX=metaxilema; PX=protoxilema; F=floema; Fi=fibras; E=epiderme; PA=parênquima.

Em entrenó maduro (Figura 12E-H), o padrão de deposição é mais similar entre as espécies. Entretanto, observa-se ainda a lignificação do parênquima medular em *S. spontaneum* (Figura 12H, centro do colmo), o que não ocorre com as demais. Em *S. spontaneum*, *S. robustum* e RB867515 há lignificação no parênquima subepidérmico e cortical, padrão não observado em *S. officinarum* (Figura 12E-H, periferia do colmo), correlacionando com seu menor conteúdo de lignina (Figura 11A). É importante ressaltar que, em *S. spontaneum*, a coloração é mais intensa e a quantidade de fibras associadas ao feixe vascular é maior do que nos outros genótipos (Figura 12H, periferia do colmo), correlacionando, também, com a maior quantidade de lignina (Figura 11A).

A coloração com reagente de Maüle (Figura 13) demonstrou um padrão mais similar entre os genótipos em entrenó intermediário, ficando restrita ao protoxilema (Figura 13A-D). Da mesma forma, o padrão de coloração no centro do entrenó maduro é semelhante entre as amostras (Figura 13E-H, centro do colmo), com reação positiva no proto e metaxilema e também nas fibras associadas ao feixe vascular, sendo menos intensa em *S. officinarum* (Figura 13F). Diferentemente da reação com floroglucinol, não foi possível observar reação positiva com reagente de Maüle no parênquima medular de *S. spontaneum* (Figura 13D e H, centro do colmo), indicando pouca ou nenhuma deposição de siringil neste tecido deste genótipo. Próximo à epiderme, é possível notar uma coloração mais intensa em relação ao entrenó intermediário (compare Figura 13A-D e E-H, periferia do colmo). Ainda, a intensidade da coloração e o número de células com reação positiva é maior em *S. spontaneum* em relação às demais (Figura 13H, periferia do colmo), confirmando os valores de quantificação de unidades de siringil (Figura 11B), especialmente nas fibras. Por outro lado, *S. officinarum* e *S. robustum* apresentaram um padrão peculiar, com reação positiva apenas no xilema e em volta do feixe vascular, mas não nas fibras (Figura 13 F e G, periferia

do colmo). O padrão para a variedade comercial foi semelhante ao observado em *S. spontaneum*, porém com uma menor intensidade de cor na epiderme e nas fibras (Figura 13E, periferia do colmo).

Figura 13. Histoquímica para detecção de siringil (em vermelho) em secções transversais do colmo, com reagente de Maüle. A e E, RB867515; B e F, *S. officinarum*; C e G, *S. robustum*; e D e H, *S. spontaneum*. MX=metaxilema; PX=protoxilema; F=floema; Fi=fibras; E=epiderme; PA=parênquima.

4.8. Identificação de genes diferencialmente por microarranjo de oligonucleotídeos

4.8.1. Análise do perfil de expressão diferencial entre os genótipos

Foi realizada uma análise em larga escala do transcriptoma com o uso de microarranjo de oligonucleotídeos na plataforma Agilent para identificar os genes diferencialmente expressos entre os genótipos. A identificação desses genes pode ajudar no entendimento das vias regulatórias relacionadas às principais diferenças fenotípicas entre os genótipos analisados, como o acúmulo de sacarose (Figura 10), biomassa (Figura 8C) e lignina (Figura 11). O esquema das hibridizações, e a lista de genes diferencialmente expressos estão indicados na Tabela 4 e Apêndice 3.

Tabela 4. Esquema das hibridizações. Foram utilizadas duas réplicas biológicas (R1 e R2) para as hibridizações. Foram coletados tecidos de folha+1, entrenó imaturo e entrenó intermediário para cada genótipo. Ao lado, estão indicados os números de SAS diferencialmente expressos em cada experimento identificados pelo método HTself.

Experimento realizado para análise de expressão					SAS
Tecido	Nº	Hibridização			diferencialmente expressos
		Cy3		Cy5	
Folha+1	1	<i>S. officinarum</i> R1	vs	RB867515 (referência) R1	223
		RB867515 (referência) R2	vs	<i>S. officinarum</i> R2	
	2	<i>S. robustum</i> R1	vs	RB867515 (referência) R1	217
		RB867515 (referência) R2	vs	<i>S. robustum</i> R2	
	3	<i>S. spontaneum</i> R1	vs	RB867515 (referência) R1	252
		RB867515 (referência) R2	vs	<i>S. spontaneum</i> R2	
Entrenó imaturo	4	<i>S. officinarum</i> R1	vs	RB867515 (referência) R1	227
		RB867515 (referência) R2	vs	<i>S. officinarum</i> R2	
	5	<i>S. robustum</i> R1	vs	RB867515 (referência) R1	272
		RB867515 (referência) R2	vs	<i>S. robustum</i> R2	
	6	<i>S. spontaneum</i> R1	vs	RB867515 (referência) R1	261
		RB867515 (referência) R2	vs	<i>S. spontaneum</i> R2	
Entrenó intermediário	7	<i>S. officinarum</i> R1	vs	RB867515 (referência) R1	231
		RB867515 (referência) R2	vs	<i>S. officinarum</i> R2	
	8	<i>S. robustum</i> R1	vs	RB867515 (referência) R1	185
		RB867515 (referência) R2	vs	<i>S. robustum</i> R2	
	9	<i>S. spontaneum</i> R1	vs	RB867515 (referência) R1	135
		RB867515 (referência) R2	vs	<i>S. spontaneum</i> R2	

Amostras de folha+1, entrenó imaturo e entrenó intermediário de plantas com nove meses de idade (coleta de julho de 2011) foram hibridizadas em *chips* de oligonucleotídeos cruzando-se cada genótipo ancestral separadamente contra a RB867515 (referência), totalizando nove hibridizações (Tabela 4). Essas análises identificaram 2003 SAS (1018 SAS únicos) diferencialmente expressos (Apêndice 1), sendo 1945 na orientação senso e 58 na orientação antissenso. Curiosamente, apenas o SAS SCEQRT2025G01.g foi diferencialmente expresso (induzido) nas duas orientações, somente no experimento de entrenó intermediário de *S. spontaneum* em relação à RB867515. Todos os 1018 SAS foram buscados nos bancos de dados NCBI, UniprotKB (Magrane & Consortium 2011), KEGG (Kanehisa *et al.* 2011), GO (Ashburner *et al.* 2000) e Pfam (Punta *et al.* 2011) e manualmente categorizados e anotados de acordo com sua função deduzida pelo resultado do alinhamento. As categorias funcionais dos genes diferencialmente expressos são apresentadas na Figura 14. Praticamente metade dos genes diferencialmente expressos não possui função conhecida e 54 destes não possuem homologia de sequência com nenhuma outra espécie, sendo provavelmente transcritos específicos de cana-de-açúcar. Os demais genes pertencem principalmente às categorias *signal transduction*, *others*, *protein metabolism*, *transport* e *carbohydrate metabolism*. Curiosamente, apesar de esses genótipos possuírem grandes diferenças relativas à parede celular, principalmente no padrão de deposição e composição da lignina (Figuras 11, 12 e 13), a categoria *cell wall metabolism* apresentou apenas 32 SAS (3% em relação aos 1018 SAS) diferencialmente expressos. A análise apresentada na Tabela 5 mostrou que os genótipos ancestrais possuem diferentes categorias funcionais enriquecidas, sugerindo, provavelmente, uma divergência de padrão de expressão gênica. O enriquecimento de categorias utiliza a ferramenta *GeneMerge* (Castillo-Davis & Hartl 2003) que realiza uma análise estatística para identificar categorias funcionais que estejam mais

representadas entre os genes diferencialmente expressos em relação à população de sondas presentes na lâmina. Essa ferramenta gera um *e-score* para cada categoria funcional que varia de zero a um, onde valor mais próximo de zero indica o maior enriquecimento.

O microarranjo de oligonucleotídeos da plataforma Agilent para cana-de-açúcar já se mostrou anteriormente uma ferramenta robusta para análise de expressão gênica em larga escala, com taxa de validação entre 80 a 90% (Costa 2012; Lembke *et al.* 2012; Sato 2012). Da mesma forma, foram selecionados 40 pontos experimentais para validação dos experimentos de folha+1, entrenó imaturo e intermediário (Figuras 15-17), alcançando-se uma taxa validação de 82%.

Figura 14. Categorias funcionais dos 2003 SAS diferencialmente expressos (Apêndice 1) nas 9 hibridizações (folha+1, entrenó imaturo e intermediário) dos 3 genótipos ancestrais contra a variedade comercial RB867515.

Tabela 5. Categorias funcionais enriquecidas dos genes diferencialmente expressos nos experimentos dos ancestrais, considerando os 3 tecidos analisados.

<i>S. officinarum</i> vs RB867515		<i>S. robustum</i> vs RB867515		<i>S. spontaneum</i> vs RB867515	
e-score	Descrição	e-score	Descrição	e-score	Descrição
6.63E-63	Unknown	2.79E-43	Unknown	4.07E-50	Unknown
3.36E-07	Hormone biosynthesis	7.78E-05	Others	2.23E-05	Others
4.30E-06	Pathogen Resistance	6.90E-02	Carbohydrate metabolism	1.43E-02	RNA metabolism
2.10E-05	Others	0.12116	Redox metabolism	1.63E-02	Protein metabolism
7.47E-05	Protein metabolism	0.14272	Protein metabolism	5.68E-02	Lipid metabolism
2.10E-04	Redox metabolism	0.17438	Transport	0.20129	Vitamin and Cofactor metabolism
2.25E-04	Transposable element-related	0.17700	RNA metabolism		
1.09E-03	Signal Transduction	0.18299	Hormone biosynthesis		
9.12E-03	Flavonoid and anthocyanin metabolism	0.20249	AminoAcid and nitrogen metabolism		
1.32E-02	Transport				
7.02E-02	DNA metabolism				
0.19963	AminoAcid and nitrogen metabolism

Figura 15. Validação por qPCR na terceira réplica biológica de folha dos genes diferencialmente expressos identificados por microarranjo. O sinal de referência (RB867515) foi definido como 1 em todos os casos. Os quadros verdes e vermelhos junto aos gráficos indicam o resultado do microarranjo, sendo: SO, *S. officinarum*; SR, *S. robustum*; SS, *S. spontaneum*; up, induzido; down, reprimido. Símbolos indicam o ponto experimental cujo resultado do qPCR confirma (✓) ou não (X) o resultado do microarranjo após análise com teste-t ($p < 0,05$) para cada par hibridizado (genótipo ancestral vs RB867515). Barras sem símbolos indicam genes que não foram identificados por microarranjo como diferencialmente expressos no genótipo ancestral em relação à RB867515 pelo método HTself. Gráficos mostram erro padrão com $n=3$.

Figura 16. Validação por qPCR na terceira réplica biológica de entrenó imaturo dos genes diferencialmente expressos identificados por microarranjo. O sinal de referência (RB867515) foi definido como 1 em todos os casos. Os quadros verdes e vermelhos junto aos gráficos indicam o resultado do microarranjo, sendo: SO, *S. officinarum*; SR, *S. robustum*; SS, *S. spontaneum*; up, induzido; down, reprimido. Símbolos indicam o ponto experimental cujo resultado do qPCR confirma (✓) ou não (X) o resultado do microarranjo após análise com teste-t ($p < 0,05$) para cada par hibridizado (genótipo ancestral vs RB867515). Barras sem símbolos indicam genes que não foram identificados por microarranjo como diferencialmente expressos no genótipo ancestral em relação à RB867515 pelo método HTself. Gráficos mostram erro padrão com $n=3$.

Figura 17. Validação por qPCR na terceira réplica biológica de entrenó intermediário dos genes diferencialmente expressos identificados por microarranjo. O sinal de referência (RB867515) foi definido como 1 em todos os casos. Os quadros verdes e vermelhos junto aos gráficos indicam o resultado do microarranjo, sendo: SO, *S. officinarum*; SR, *S. robustum*; SS, *S. spontaneum*; up, induzido; down, reprimido. Símbolos indicam o ponto experimental cujo resultado do qPCR confirma (✓) ou não (X) o resultado do microarranjo após análise com teste-t ($p < 0,05$) para cada par hibridizado (genótipo ancestral vs RB867515). Barras sem símbolos indicam genes que não foram identificados por microarranjo como diferencialmente expressos no genótipo ancestral em relação à RB867515 pelo método HTself. Gráficos mostram erro padrão com $n=3$.

O diagrama de Venn apresentado na Figura 18 mostra que, entre os ancestrais, *S. robustum* e *S. spontaneum* possuem perfis de expressão gênica mais similares e *S. officinarum* forma um grupo à parte. Além de *S. officinarum* apresentar mais genes diferencialmente expressos exclusivos, os dois primeiros compartilham 191 (60+131) genes diferencialmente expressos enquanto que *S. officinarum* compartilha apenas 106 (60+46) e 138 (60+78) genes induzidos ou reprimidos com *S. spontaneum* e *S. robustum*, respectivamente. A diferença de perfil de expressão entre os ancestrais é refletida em alguns parâmetros fenotípicos, como acúmulo de açúcares (Figura 10) e resistência a estresses. Esse fato reforça a idéia de que os genes identificados pelos experimentos de microarranjo possam ser alvos interessantes para estudos posteriores visando o aperfeiçoamento de características agronômicas e também para o entendimento dos mecanismos moleculares de regulação dessas características. As funções dos genes identificados, e possivelmente correlacionados com as diferenças fenotípicas entre os genótipos, serão abordadas na discussão.

Figura 18. Diagrama de Venn dos 1018 SAS diferencialmente expressos (Apêndice 1) nas 9 hibridizações (folha+1, entrenó imaturo e entrenó intermediário) dos 3 genótipos ancestrais contra a variedade comercial RB867515. Note que as áreas são proporcionais ao número de genes.

Por se tratar de genótipos com um contraste para Brix com uma variabilidade genética maior que variedades comerciais, é interessante comparar os dados de expressão dos genótipos ancestrais com outros experimentos de variedades contrastantes para Brix, a fim de identificar possíveis padrões de expressão, similares ou divergentes, que possam ajudar a identificar redes regulatórias ou alvos de interesse. Assim, foi conduzida uma análise comparativa com dados de expressão de quatro experimentos (2 de folha e 2 de entrenó imaturo) de genótipos contrastantes para teor de sacarose provenientes de cruzamentos entre variedades comerciais (Sato 2012) realizados pelo programa de melhoramento da RIDESA (Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético). *S. officinarum*, por também possuir alto Brix, não foi utilizada nessas comparações, uma vez que a hibridização *S. officinarum* vs RB867515 não seria uma comparação alto e baixo Brix e, sim, ancestral contra híbrido, o que poderia incluir um viés na comparação alto e baixo Brix. A comparação dos resultados para entrenó imaturo entre esses dois grupos (Figura 19A) mostrou que apenas 91 SAS diferencialmente expressos aparecem em ambas as listas, tendo de 390 a 500 SAS exclusivos em cada grupo de experimento. Por outro lado, há quase o dobro (177) de genes diferencialmente expressos em comum entre os experimentos *S. robustum* x RB867515 e *S. spontaneum* x RB867515 (Figura 19B). Os genes diferencialmente expressos em comum aos experimentos com os genótipos contrastantes para Brix da RIDESA e os ancestrais de baixo Brix (*S. spontaneum* e *S. robustum*, contra a referência de alto Brix RB867515) estão listados na Tabela 6. A análise de categorias funcionais enriquecidas (Tabela 7) também mostrou um padrão distinto entre os genótipos da RIDESA e os ancestrais, onde o enriquecimento de categorias é bem diferente entre aquelas que figuram em ambos os grupos e as que são exclusivas dos genótipos da RIDESA e dos ancestrais. Por exemplo, com exceção da categoria *unknown*, a

categoria mais enriquecidas dos SAS que aparecem em comum nos genótipos da RIDESA e nos ancestrais é *protein metabolism*, a qual não está enriquecida nem para os SAS exclusivos dos genótipos da RIDESA nem para os ancestrais.

Figura 19. Diagrama de Venn dos SAS diferencialmente expressos em (A) quatro experimentos (2 de folha e 2 de entrenó imaturo) de alto Brix contra baixo Brix (ABxBB) de genótipos da RIDESA e dos dois genótipos ancestrais de baixo Brix (*S. spontaneum* e *S. robustum*) contra a referência RB867515 (de alto Brix) e (B) apenas entre os dois genótipos ancestrais de baixo Brix (*S. spontaneum* e *S. robustum*) contra a referência de alto Brix RB867515. Note que as áreas sobrepostas são proporcionais ao número de genes de cada conjunto.

Tabela 6. Comparação entre genes diferencialmente expressos em folha e entrenó imaturo em 4 experimentos (2 de folha e 2 de entrenó) de genótipos da RIDEA de alto Brix x baixo Brix (Sato 2012) e os diferencialmente expressos observados em 4 experimentos (2 de folha e 2 de entrenó imaturo) dos genótipos ancestrais de baixo Brix (*S. robustum* e *S. spontaneum*, Figura 8A) contra a referência de alto Brix, RB867515. Números em parênteses mostram a quantidade de genes de cada coluna que estão induzidos (up) ou reprimidos (down). SAS sombreado em cinza indica o único SAS antissenso da lista.

SAS	Categoria funcional	Anotação	Ridesa		Ancestrais	
			up (46)	down (62)	up (62)	down (90)
SCSBSB1057D05.g	Carbohydrate metabolism	Alpha-glucan phosphorylase	1			1
SCUTAM2089E05.g	Carbohydrate metabolism	beta-amylase		1	2	
SCMCCL6052H01.g	Carbohydrate metabolism	Fructose and Mannose Metabolism		1		2
SCCCCL4007E06.g	Carbohydrate metabolism	glucosyltransferase		1		1
SCEQHR1082B01.g	Carbohydrate metabolism	Glycosyl hydrolase family 1		1		2
SCEQRT1026E05.g	Cell wall metabolism	beta-1,3-glucanase GH17	1	1	1	
SCCCLR2C02C12.g	DNA metabolism	histone		1		1
SCCCLR2C02B06.g	DNA metabolism	histone		3		1
SCRUSB1078A08.g	DNA metabolism	nuclease	1			1
SCRURT3061C01.g	DNA metabolism	nuclease	2		1	
SCCCLR2002F04.g	DNA metabolism	regulation of transcription		1	1	
SCSGHR1066C03.b	Flavonoid and anthocyanin metabolism	chalcone synthase		1		2
SCRLLR1038D05.g	Hormone biosynthesis	jasmonic acid		2	2	
SCCCRT1001E01.g	Hormone biosynthesis	jasmonic acid		1		2
SCUTST3086G11.g	Light harvesting	chlorophyll a/b binding protein (CAB)	1			1
SCJLLR1106F01.g	Lipid metabolism	Acyltransferases	1		1	
SCCCLR2C03F01.g	Lipid metabolism	Non-specific lipid-transfer protein		1		1
SCBFRZ2045A12.g	Lipid metabolism	Phosphoethanolamine N-methyltransferase		1	1	
SCEQLB1068C05.g	Nucleotide metabolism	Dihydrofolate reductase		1		2
SCEPAM2015E05.g	Others	basic blue protein	1			2
SCSGHR1069F04.b	Others	cytochrome P450		1		1
SCRFLR1034A11.g	Others	desidrogenase		1	1	
SCJLAM1063F08.g	Others	hydrolase		1	1	

Tabela 6. Continuação

SAS	Categoria funcional	Anotação	Ridesa		Ancestrais	
			up (46)	down (62)	up (62)	down (90)
SCCCLR1070B06.g	Others	synthase	1		1	
SCVPRZ2035C02.g	Others	uncharacterized	1		1	
SCMCLR1122E10.g	Others	uncharacterized		1		1
SCAGCL6013C08.g	Pathogen Resistance	Barwin		1	2	
SCRLFL1014B12.g	Pathogen Resistance	R-gene		1	2	
SCCCLR2C01F06.g	Pathogen Resistance	wound-induced		1	1	
SCACLR2022B02.g	Protein metabolism	Eukaryotic peptide chain release factor	1	2		4
SCCCLR2001C04.g	Protein metabolism	heat-shock protein		1		1
SCCCLRT2004D09.g	Protein metabolism	peptidase		1	1	
SCBGLR1097G03.g	Protein metabolism	protease		1		2
SCCCLR1C01H05.g	Protein metabolism	protease		1	1	
SCSBAM1087G05.g	Protein metabolism	ubiquitination	1	1	2	
SCJFLR1013A04.g	Protein metabolism	uncharacterized		1		2
SCBFRT1071H10.g	Redox metabolism	Reductase		1	1	
SCCCLR1001F09.g	Signal Transduction	calmodulin-binding		1	1	
SCRUFL1115A08.b	Signal Transduction	Serine/threonine protein phosphatase 7	1		1	
SCVPRT2076A04.g	Transport	ABC transporter	1		4	
SCJFRZ2025E05.g	Transport	Major intrinsic protein	1			2
SCQSHR1022H06.g	Transport	Probable sulfate transporter		1	1	
SCBGAM1090F08.g	Unknown	no matches	1		1	
SCRUFL3069A10.g	Unknown	no matches		3	1	
SCVPCL6038G07.g	Unknown	no matches	1			3
SCACSD2014F07.g	Unknown	no matches		2		4
SCCCCL4004E05.g	Unknown	no matches	1		4	
SCACFL5031C04.g	Unknown	no matches	2			2
SCEQLR1092A03.g	Unknown	uncharacterized		1	1	
SCAGLR2018G01.g	Unknown	uncharacterized	1		1	
SCQGRT1039C10.g	Unknown	uncharacterized		1	1	
SCBGLR1117C12.g	Unknown	uncharacterized		1		2
SCEPFL3081E03.g	Unknown	uncharacterized	1			2
SCUTRZ2022D02.g	Unknown	uncharacterized	1			1
SCUTRZ2022E05.g	Unknown	uncharacterized	1	1		2
SCJLLR1107G06.g	Unknown	uncharacterized	1			2
SCEZRZ3015D02.g	Unknown	uncharacterized		2		2
SCVPLB1020B12.g	Unknown	uncharacterized		1		1
SCJLLR1033B04.g	Unknown	uncharacterized		1	4	
SCBFAD1067F08.g	Unknown	uncharacterized	2		1	

Tabela 6. Continuação

SAS	Categoria funcional	Anotação	Ridesa		Ancestrais	
			up (46)	down (62)	up (62)	down (90)
SCCCLB1C05B09.g	Unknown	uncharacterized		1	1	
SCJFRT1061G09.g	Unknown	uncharacterized	2			1
SCUTRZ3106F01.g	Unknown	uncharacterized		1	1	
SCEPAM2053C07.g	Unknown	uncharacterized		1	1	1
SCEPRZ1008A06.g	Unknown	uncharacterized	1			1
SCAGRT3047G04.g	Unknown	uncharacterized		1	2	
SCSGLR1025D03.g	Unknown	uncharacterized		1		3
SCSGAM2102C11.g	Unknown	uncharacterized	1		1	
SCEZRZ1016C10.g	Unknown	uncharacterized	1			3
SCCCAM1073A05.g	Unknown	uncharacterized	1		1	
SCRUFL3066B09.g	Unknown	uncharacterized		1		3
SCCCLR2C03C07.g	Unknown	uncharacterized	1			4
SCJLLR1108B01.g	Unknown	uncharacterized		1		1
SCEQRT1025E08.g	Unknown	uncharacterized	1		2	
SCRFAM1029A07.g	Unknown	uncharacterized	1		2	
SCEQRT2098E08.g	Unknown	uncharacterized		1	1	
SCCCRZ3004F06.g	Unknown	uncharacterized		1		1
SCUTLR1058H04.g	Unknown	uncharacterized	1			1
SCJFST1014E10.b	Unknown	uncharacterized	1			2
SCUTAM2005B05.g	Unknown	uncharacterized	1	1		2
SCUTFL1055C05.g	Unknown	uncharacterized	1			3
SCRUHR1073A05.g	Unknown	uncharacterized		1	1	
SCCCLR1C03A08.g	Unknown	uncharacterized	1			2
SCEQLB1067F01.g	Unknown	uncharacterized		1	2	1
SCVPFL3046G12.b	Unknown	uncharacterized	1		2	
SCCCRZ1002G05.g	Unknown	uncharacterized	1			2
SCACLR1128A09.g	Unknown	uncharacterized	1			1
SCCART1002B11.g	Unknown	uncharacterized	1			2
SCCCLR1001H03.g	Unknown	uncharacterized	1			2
SCEQRZ3090H08.g	Unknown	uncharacterized		1	1	
SCJLRZ1024B01.g	Unknown	uncharacterized		1		1

Tabela 7. Categorias funcionais enriquecidas dos genes diferencialmente expressos nos experimentos da RIDESA (Sato 2012) e dos ancestrais.

Comuns RIDESA e ancestrais		Exclusivos RIDESA		Exclusivos Ancestrais	
e-score	Descrição	e-score	Descrição	e-score	Descrição
8.64E-12	Unknown	2.04E-31	Unknown	2.68E-19	Unknown
6.58E-02	Protein metabolism	1.39E-14	Others	0.61399	Others
0.18996	Carbohydrate metabolism	5.35E-06	Signal Transduction		
0.28383	Transport	4.73E-05	Redox metabolism		
		8.12E-05	Transport		
		1.99E-04	Lipid metabolism		
		3.26E-02	Hormone biosynthesis		
		0.11991	Cell wall metabolism		
		0.14190	DNA metabolism		
		0.24382	Carbohydrate metabolism		

Da mesma forma, os dados de expressão também foram comparados com experimentos de microarranjo de quatro tecidos (folha, entrenó imaturo, intermediário e maduro) de plantas de alto e baixo Brix derivadas de cruzamentos entre genótipos de *S. officinarum* e *S. spontaneum* (Papini-Terzi *et al.* 2009). Diferentemente dos genótipos da RIDESA, esses experimentos foram conduzidos no chip SUCAST, que contém um número bem menor de sondas em relação ao chip Agilent. Nenhum SAS em comum foi identificado entre esses experimentos dessas progênies (Papini-Terzi *et al.* 2009) e os deste presente trabalho, mas isso é devido, principalmente, à grande diferença das sondas presentes no chip Agilent e SUCAST, os quais compartilham poucos SAS. Ainda assim, na Tabela 8 estão listados os principais genes diferencialmente expressos nas progênies do cruzamento *S. officinarum* e *S. spontaneum* (Papini-Terzi *et al.* 2009). Pode-se observar que genes da via de biossíntese de lignina, três PAL e uma C4H, aparecem reprimidos, assim como genes de biossíntese de hormônios e duas aquaporinas, sendo uma destas reprimida em três tecidos. De forma geral, os genes da categoria de transdução de sinal estão induzidos, incluindo um

fator da família MYB, mas o único gene responsivo a hormônios (SCJFLR1035E04.g) está reprimido. A comparação com os dados dos genótipos ancestrais será abordada na discussão.

Tabela 8. Principais genes diferencialmente expressos em experimentos de folha, entrenó imaturo, intermediário e maduro de plantas de alto e baixo Brix derivadas de cruzamentos entre *S. officinarum* e *S. spontaneum* (Papini-Terzi *et al.* 2009) realizados no chip SUCAST. Up e down indicam indução ou repressão, respectivamente, nas plantas de alto Brix em relação às de baixo Brix.

SAS	Categoria funcional	anotação	up	down
SCEZRZ1017E02.g	Carbohydrate metabolism	trehalase	1	
SCRULR1020D11.g	Carbohydrate metabolism	aldolase		1
SCCCLR1001E04.g	Carbon fixation	Rubisco small subunit		1
SCBFFL5074C09.g	Cell wall metabolism	reversibly glycosylated polypeptide		1
SCEQRT1024E12.g	Cell wall metabolism	phenylalanine ammonia-lyase		1
SCSGAM1094D05.g	Cell wall metabolism	phenylalanine ammonia-lyase		1
SCSGFL4193B05.g	Cell wall metabolism	cinnamate 4-hydroxylase (C4H)		1
SCCCLR1048D07.g	Cell wall metabolism	phenylalanine ammonia-lyase		1
SCAGLR1043E04.g	Hormone biosynthesis	Allene oxide synthase		2
SCRFS1020D03.g	Hormone biosynthesis	jasmonic acid		1
SCJLLR1011C06.g	Hormone biosynthesis	zeaxanthin epoxidase		1
SCCCRZ1001H05.g	Signal Transduction	Transcription Factor HLH	1	
SCCCCL3120C09.g	Signal Transduction	receptor Ser/Thr kinase		1
SCCCAM2004F11.g	Signal Transduction	Protein kinase	1	
SCMCCL6049C11.g	Signal Transduction	Calmodulin	1	
SCSGLR1084B06.g	Signal Transduction	kinase/phosphatase	1	
SCCCRZ1002F10.g	Signal Transduction	receptor Ser/Thr kinase	1	
SCJFRZ2012F04.g	Signal Transduction	receptor Ser/Thr kinase	1	
SCACLR1036B06.g	Signal Transduction	CBL-interacting protein kinase		1
SCCCCL5001D11.g	Signal Transduction	SNF-like kinase	1	
SCEQRT2099H01.g	Signal Transduction	Calcium-dependent protein kinase		1
SCJFLR1035E04.g	Signal Transduction	gibberellin-responsive protein		1
SCBFSB1046D04.g	Signal Transduction	CBL-interacting protein kinase	1	
SCJLLR1054C03.g	Signal Transduction	kinase/phosphatase	1	
SCVPFL3046C06.b	Signal Transduction	kinase/phosphatase	1	
SCCCLR2C01F06.g	Signal Transduction	others		1
SCRURT2007E11.g	Signal Transduction	MYB-like protein	1	
SCCCAM1001A03.g	Transport	ABC transporter	1	
SCCCRZ1001F02.g	Transport	aquaporin		1
SCCCRZ1002E08.g	Transport	aquaporin		3

4.8.2. Genes diferencialmente expressos entre entrenó imaturo e intermediário

Com o desenvolvimento de novas ferramentas de análise, foi possível realizar análises “híbridas”, ou seja, cruzando-se amostras de diferentes experimentos para gerar novos resultados com outros objetivos, sem a necessidade de realizar novos experimentos na bancada, apenas rodando novas análises estatísticas. Em outras palavras, essa ferramenta permite cruzar dados de expressão de uma amostra com dados de qualquer outra amostra presente no banco de dados do sucest-fun (sucest-fun.org), mesmo sendo de experimentos diferentes e não relacionados. Por exemplo, as hibridizações foram realizadas de forma a comparar um mesmo tecido de dois genótipos diferentes, como no caso entrenó imaturo de *S. spontaneum* contra entrenó imaturo de RB867515 e, da mesma forma, no caso de entrenó intermediário. Com a nova ferramenta, é possível cruzar os dados de expressão de entrenó imaturo de *S. spontaneum* com os dados de entrenó intermediário desse mesmo genótipo e rodar a estatística do método HTself para gerar os genes diferencialmente expressos entre esses dois tecidos, mesmo que essa hibridização nunca tenha sido feito na prática, num experimento de bancada. Assim, com o intuito de identificar os genes induzidos e reprimidos ao longo do desenvolvimento do colmo, amostras de entrenó imaturo (utilizado como referência) e entrenó intermediário de *S. officinarum* foram cruzadas e identificou-se 297 genes diferencialmente expressos (Tabela 9 e Apêndice 2). O mesmo foi feito para RB867515 e *S. spontaneum*, gerando 284 e 226 genes diferencialmente expressos, respectivamente (Tabela 9 e Apêndice 2). Do total de 807 SAS identificados (507 SAS únicos), apenas 6 são na orientação antissenso. A Figura 20 indica as principais categorias funcionais alteradas, sendo primeiramente *unknown*, seguido de *others*, *signal transduction*, *cell wall metabolism*, *DNA metabolism* e *cell cycle*. A análise de categorias funcionais enriquecidas

(Tabela 10) mostrou que as categorias *cell cycle* e *cell wall metabolism* estão entre as mais enriquecidas nos três experimentos. Ao contrário do que ocorre para os experimentos anteriores (Tabela 5), a categoria *unknown* não é a mais enriquecida pra RB867515 e *S. officinarum*. A análise dos genes diferencialmente expressos será abordada na discussão.

Curiosamente, nenhum gene relacionado diretamente ao metabolismo de sacarose está diferencialmente expresso nos genótipos analisados, mesmo para *S. officinarum* que apresenta um aumento de sacarose já no entrenó intermediário (Figura 10D).

Tabela 9. Esquema das hibridizações de entrenó imaturo vs entrenó intermediário. Foram utilizados dados de expressão de duas réplicas biológicas (R1 e R2) entrenó imaturo (E1) e entrenó intermediário (E5) para cada genótipo. Ao lado, estão indicados os números de SAS diferencialmente expressos em cada experimento identificados pelo método HTself. *Up* e *down* indicam o nível de expressão em entrenó intermediário em relação ao imaturo.

Genótipo	Cruzamento			SAS diferencialmente expressos		
	Cy3		cy5	up	down	Total
<i>S. officinarum</i>	E1 (referência) R1	vs	E5 R1	152	145	297
	E5 R2	vs	E1 (referência) R2			
RB867515	E1 (referência) R1	vs	E5 R1	150	134	284
	E5 R2	vs	E1 (referência) R2			
<i>S. spontaneum</i>	E1 (referência) R1	vs	E5 R1	125	101	226
	E5 R2	vs	E1 (referência) R2

Figura 20. Categorias funcionais dos 807 SAS diferencialmente expressos (Apêndice 2) nas 3 hibridizações de entrenó imaturo (referência) vs entrenó intermediário de RB867515, *S. officinarum* e *S. spontaneum*.

Tabela 10. Categorias funcionais enriquecidas dos genes diferencialmente expressos nos experimentos de entrenó imaturo vs entrenó intermediário do mesmo genótipo (tabela 9 e figura 20).

RB867515		<i>S. officinarum</i>		<i>S. spontaneum</i>	
e-score	Descrição	e-score	Descrição	e-score	Descrição
2.3E-14	Cell cycle	2.5E-18	Cell cycle	3.3E-08	Unknown
2.2E-09	Others	3.1E-10	Cell wall metabolism	2.3E-05	Cell cycle
2.6E-08	Cell wall metabolism	5.1E-09	Unknown	2.8E-04	Others
5.6E-05	DNA metabolism	3.6E-07	Others	4.0E-04	Cell wall metabolism
3.5E-04	Unknown	9.3E-05	DNA metabolism	2.6E-03	Biosynthesis of other Secondary Metabolites
2.2E-03	Redox metabolism	4.8E-04	Signal Transduction	4.6E-03	Signal Transduction
4.7E-03	Signal Transduction			1.6E-02	Redox metabolism
				0.5E-02	DNA metabolism

4.8.3. Identificação de padrões de expressão dos genes com sinal de expressão acima do *background*

Além dos genes identificados pelo método HTself (ver Metodologia), a partir dos dados de expressão dos oligoarranjos, também é possível analisar os genes pela sua intensidade do sinal de expressão (Lembke *et al.* 2012; Sato 2012). Assim, pode-se identificar um grande número de genes expressos para os quais o sinal de expressão pode não ser alto o suficiente para que o gene seja classificado como diferencialmente expressos pelo método HTself, permitindo visualizar o transcriptoma de um maneira mais global. A Tabela 11 mostra os números gerais dos SAS com sinal de expressão acima do *background*, divididos por experimentos e discriminando-se o número de sondas que apresentaram exclusivamente sinal do transcrito senso, do antissenso ou ambos. Cada experimento apresenta cerca de 10

a 12 mil sondas com sinal de expressão, sendo 9.600 a 11.400 sondas senso e 300 a 500 sondas antissenso. Das 14.522 sondas senso e 7.380 antissenso presentes no *chip*, somando-se todos os experimentos foram observadas 12.621 (86,9%) sondas senso e 995 (13,5%) sondas antissenso com sinal de expressão, números similares aos obtidos anteriormente em outros experimentos de nosso laboratório (Costa 2012; Lembke *et al.* 2012; Sato 2012). Curiosamente, é possível notar que existe um número significativo de genes expressos apenas na orientação antissenso (110 a 197 genes) e também de genes coexpressos nas duas orientações (181 a 338 genes).

A Tabela do Apêndice 4 mostra que as categorias funcionais enriquecidas dos genes com sinal de expressão para cada genótipo possuem algumas semelhanças. Entre as categorias com enriquecimento mais significativo (*e-score*=0), *oxidative phosphorylation*, *light harvesting* e *Cytoskeleton and vesicle trafficking* aparecem em todos os genótipos. *S. officinarum* e RB867515 formam um grupo à parte de das demais, pois essas plantas de alto Brix ainda compartilham as categorias *RNA metabolism*, *Protein metabolism* e *Circadian Clock* com *e-score* igual a zero, as quais não aparecem na lista para *S. spontaneum* e *S. robustum*. Essa diferença pode ser devido à composição genética de *S. officinarum* nos genomas dos híbridos comerciais ser maior (D'Hont *et al.* 1996; D'Hont 2005; D'Hont *et al.* 2008) e reflete nas características fenotípicas, principalmente no acúmulo de sacarose (Figura 10) e também em características visuais (Figura 7). Apenas em RB867515 nota-se que a categoria *DNA metabolism* aparece entre as categorias com *e-score* igual a zero, a qual inclui genes relacionados ao controle da estrutura da cromatina, como histonas e DNA metilases, além de genes de replicação e reparo. Esse fato levanta a hipótese de que a regulação epigenética parece estar mais presente no genótipo híbrido e pode ser devida ao “estresse genômico” causado pela hibridização de dois genomas de espécies diferentes,

como sugerido para híbridos de *Arabidopsis thaliana* e *Arabidopsis arenosa* (Chen & Tian 2007), podendo estar associada à melhor performance da RB867515 em comparação às demais plantas analisadas.

Tabela 11. Sondas com sinal acima do *background* em todos os experimentos. SS, senso; AS, antissenso.

		Sondas com sinal	SS	AS	Apenas SS	Apenas AS	SS e AS
Folha	<i>S. officinarum</i>	10616	10215	401	9976	162	239
	<i>S. robustum</i>	10278	9947	331	9733	117	214
	<i>S. spontaneum</i>	10318	9970	348	9732	110	238
	RB867515	11418	10913	505	10583	175	330
Entrenó imaturo	<i>S. officinarum</i>	11687	11162	525	10827	190	335
	<i>S. robustum</i>	10820	10472	348	10271	147	201
	<i>S. spontaneum</i>	11274	10829	445	10535	151	294
	RB867515	12023	11488	535	11150	197	338
Entrenó intermediário	<i>S. officinarum</i>	11135	10745	390	10522	167	223
	<i>S. robustum</i>	10681	10317	364	10103	150	214
	<i>S. spontaneum</i>	9976	9660	316	9479	135	181
	RB867515	11402	10939	463	10651	175	288
Total		13616	12621 (86,9%)	995 (13,5%)	-	-	-

Existem muitas diferenças entre os genótipos ancestrais já descritas na literatura. Algumas dessas diferenças foram confirmadas nas análises agronômicas e fisiológicas, como o maior Brix de *S. officinarum* (Figura 8A), o maior perfilhamento (Figura 8D) e fibra (lignina, Figura 11A) de *S. spontaneum*, maior biomassa por colmo da variedade comercial (Figura 8C). Outras características já conhecidas, como a maior resistência à estresses de *S. spontaneum* e *S. robustum* não puderam ser constatadas uma vez que o experimento não

propunha analisar plantas sob estresse comparando planta tratada e controle, mas foi possível identificar um indício que *S. spontaneum* é mais resistente à seca por apresentar maior eficiência no uso de água (Figura 9D). Dentre todas essas características analisadas, Brix (Figura 8A) e lignina (Figura 11A) foram as que apresentaram as maiores diferenças entre os genótipos. Como o acúmulo de sacarose já foi alvo de outros estudos anteriores em nosso laboratório (Papini-Terzi *et al.* 2009; Sato 2012) e a identificação de genes relacionados ao metabolismo de parede celular, incluindo lignina, são de extrema importância para o desenvolvimento da cana energia e da melhoria da biomassa para bioetanol celulósico, escolhemos focar as análises de expressão em genes de parede celular e lignina. Como descrito anteriormente (Seção 4.8.1), poucos genes de metabolismo de parede celular foram identificados como diferencialmente expressos pelo método HTself, mesmo os genótipos apresentando diferenças importantes quanto à composição de parede celular. Assim, uma análise direcionada aos genes do catálogo dessa categoria funcional (Seção 4.1) foi conduzida utilizando-se o sinal de expressão acima do *background* (Tabela 12). Essa análise visa identificar padrões de expressão que possam explicar os fenótipos analisados, principalmente o padrão diferenciado de *S. spontaneum* em relação à quantidade e padrão de deposição de lignina (Figuras 11-13), avançando no conhecimento das redes regulatórias da biossíntese de parede celular. Dos 1606 SAS do catálogo, 541 estão presentes no *chip* Agilent, sendo 332 presentes em ambas orientações e 209 apenas na orientação senso. Desse total, 513 sondas apresentaram sinal de expressão acima do *background* em pelo menos um experimento. A maioria dos SAS apresentaram expressão apenas na orientação senso e, apesar de ter mais de 300 sondas para detecção de antissenso, apenas 40 (10 a 17 em cada experimento) apresentaram sinal.

Tabela 12. Sondas do catálogo de genes de parede celular com sinal acima do *background* em todos os experimentos. SS, senso; AS, antissenso.

		Sondas com sinal	SS	AS	Apenas SS	Apenas AS	SS e AS
Folha	<i>S. officinarum</i>	327	315	12	307	4	8
	<i>S. robustum</i>	320	309	11	302	4	7
	<i>S. spontaneum</i>	342	331	11	323	3	8
	RB867515	372	355	17	342	4	13
Entrenó imaturo	<i>S. officinarum</i>	388	374	14	364	4	10
	<i>S. robustum</i>	373	363	10	355	2	8
	<i>S. spontaneum</i>	411	395	16	381	2	14
	RB867515	407	392	15	381	4	11
Entrenó intermediário	<i>S. officinarum</i>	417	404	13	395	4	9
	<i>S. robustum</i>	402	387	15	376	4	11
	<i>S. spontaneum</i>	384	371	13	364	6	7
	RB867515	427	412	15	403	6	9
Total		513	473	40	-	-	-

Com os sinais de intensidade, foi possível gerar uma *clusterização* hierárquica para identificar genes do catálogo com padrão de expressão diferencial entre os genótipos e, assim, identificar genes relacionados com os fenótipos descritos. Os *clusters* foram gerados pelo programa MEV (*Multi Experiment Viewer*, versão 4.8.1) para os dados de expressão de entrenó imaturo e intermediário após uma análise para identificar o número ideal de *clusters*. Foram identificados 12 *clusters* de genes que apresentaram um padrão de expressão bem característico (Figuras 21 e 22), os quais podem ser consultados nos Apêndices 6 e 7 com a anotação de cada gene. Assim como para o método HTself, também foi avaliado o perfil de expressão em uma terceira réplica biológica para validar por qPCR a análise baseada na intensidade do sinal. A validação foi realizada apenas para alguns poucos

genes, mas mostrou um perfil similar entre o qPCR e a intensidade do sinal na lâmina (Figura 23).

Figura 21. Clusterização dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó imaturo. Os clusters mostram a variação de expressão dos genes em cada genótipo, separando-os em grupos de genes com um mesmo perfil. A linha rosa indica o perfil de expressão de cada cluster e as linhas cinzas indicam a expressão de cada gene do cluster.

Figura 22. Clusterização dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó intermediário. Os clusters mostram a variação de expressão dos genes em cada genótipo, separando-os em grupos de genes com um mesmo perfil. A linha rosa indica o perfil de expressão de cada cluster e as linhas cinzas indicam a expressão de cada gene do cluster.

Figura 23. Perfil de expressão do transcrito senso pela análise de intensidade (gráfico de linhas) e por qPCR (gráfico de barras) em entrenó imaturo e intermediário.

Dos 12 *clusters* de cada tecido, é interessante analisar aqueles que apresentem expressão diferencial, principalmente, em *S. spontaneum*, visto que este genótipo possui a maior diferença no teor de lignina (Figura 11A), pois, assim, é possível identificar os genes que possam ter relação com a maior quantidade de lignina nesse genótipo. Para entrenó imaturo, nos *clusters* onde *S. spontaneum* possui os maiores valores de expressão, podemos identificar alguns genes iniciais da via de lignina (*4CL* e *PAL*), como exemplificado na Figura 24. Também é possível identificar o antissenso da *COMT*, sugerindo algum tipo de regulação por esse mecanismo nos genes finais da via. A validação por qPCR do antissenso do gene *COMT* (Figura 25) confirmou o perfil de expressão para três dos quatro genótipos (exceto para RB867515), mostrando que a análise de sinal de expressão acima do background pode

ser útil para identificação de diferenças não identificadas pelo HTself, com boa taxa de validação. É importante ressaltar que os genes *PAL* e *4CL* aparecem quase que exclusivamente nos clusters onde *S. spontaneum* tem maior expressão (Apêndice 5). Ainda, pode-se observar alguns fatores das famílias NAC e MYB mais expressos nesse genótipo em entrenó imaturo (Apêndice 5) correlacionado com o fato desse genótipo iniciar a lignificação mais precocemente, tornando-os candidatos interessantes para serem ativadores da biossíntese de parede celular secundária. Por outro lado, nos *clusters* onde *S. spontaneum* apresenta menor expressão (Figura 26), podemos identificar várias glicosil hidrolases e transferases, incluindo dois genes *kobito*, o qual possui função ainda desconhecida, mas possui domínios glicosiltransferase-like e está envolvido com a regulação da permeabilidade dos plasmodesmas (Kong *et al.* 2012) e também com a síntese de celulose em *Arabidopsis* (Pagant *et al.* 2002). Esse gene também apresenta expressão do antissenso, o qual, de forma análoga ao antissenso da *COMT*, validou para os três genótipos ancestrais e não para a RB867515 (Figura 25).

Figura 24. Cluster hierárquico dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó imaturo. O cluster acima mostra os genes com maior expressão em *S. spontaneum*. Setas vermelhas indicam os genes *4CL* e *PAL*, e seta preta indica o antissenso da *COMT*.

Figura 25. Validação da expressão do antissenso dos genes *COMT* e *kobito* em entrenó imaturo, mostrando as análises de intensidade (gráfico de linhas) e qPCR (gráfico de barras).

Figura 26. Cluster hierárquico dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó imaturo. O *cluster* acima mostra os genes com menor expressão em *S. spontaneum*. Setas vermelhas indicam os genes kobito.

Da mesma forma que no entrenó imaturo, pode-se identificar genes da biossíntese de lignina nos clusters de entrenó intermediário onde *S. spontaneum* apresenta maior expressão (Figura 27 e Apêndice 6), principalmente o gene *4CL*. O mesmo é válido para os clusters onde *S. spontaneum* tem menor expressão, onde podemos identificar o gene *kobito* menos expresso nesse genótipo também em entrenó intermediário (Figura 28).

Figura 27. Cluster hierárquico dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó intermediário. O cluster acima mostra os genes com maior expressão em *S. spontaneum*. Setas vermelhas indicam os genes *4CL*.

Figura 28. Cluster hierárquico dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó intermediário. Setas vermelhas indicam os genes *kobito*.

Embora não possua a mesma robustez e confiabilidade do método HTself, essa análise de *clusterização* (Figuras 21, 22, 24, 26-28 e Apêndices 6 e 7) permite a identificação de genes com padrão de expressão similar e será útil para identificar genes que participem de um mesmo processo biológico, como por exemplo a identificação de fatores MYB e NAC que tenham o mesmo perfil de expressão de genes da via de lignina, tornando-os bons candidatos para serem reguladores desse processo. Ainda assim, a *clusterização* dos sinais de expressão apresentou uma boa concordância com a validação feita por qPCR numa terceira réplica biológica (Figuras 25 e 27) indicando ser possível utilizar esse método para identificar padrões de expressão de interesse não identificados pelo HTself. Por exemplo, consistente com maior quantidade de lignina *S. spontaneum* (Figura 11A), a indução de genes relacionados à lignificação pode ser visualizado pelos genes *PAL* e *4CL* mais expressos nesse genótipo em entrenó imaturo (Figura 26), despertando maior interesse para esses genes para a manipulação temporal e espacial da lignificação. Uma vez identificados os principais genes da via responsáveis pelo fenótipo, a manipulação genética com o uso de promotores induzíveis ou tecido-específico pode ajudar a alterar a intensidade e o padrão da lignificação em determinados tecidos, dependendo do objetivo. Esse tipo de abordagem já se mostrou ser promissora resultando no aumento da sacarificação em *Arabidopsis* transgênica pela manipulação das relações entre promotor e sequência codificadora utilizando-se promotores tecido-específico e um gene da via de lignina de forma a restringir a lignificação aos feixes vasculares e aumentar da deposição de polissacarídeos de parede celular secundária nos outros tecidos, sem prejuízo para o crescimento da planta (Yang *et al.* 2013). A identificação dos principais genes associados a essas características em cana-de-açúcar, seguida da caracterização dos promotores, é o primeiro passo para possibilitar essa manipulação.

4.9. Análise da expressão por qPCR

Em complementação às análises de oligoarranjos, as plantas com 7 meses de idade (coleta de março de 2012) tiveram a expressão gênica de alguns alvos específicos avaliada por qPCR, principalmente genes da via de lignina. Como já mencionado anteriormente, as análises foram focadas em lignina por ser a característica com um grande contraste entre os genótipos e ser menos estudada em cana-de-açúcar do que o acúmulo de sacarose.

4.9.1. Expressão de genes da via de biossíntese de fenilpropanoides

Para obter um panorama geral da expressão de genes da via de biossíntese de lignina (Figura 2) e posteriormente correlacioná-los com os fenótipos observados (Figuras 11-13) foram selecionados os seguintes genes para análise por qPCR: *fenilalanina amônia liase* (*PAL*), *4-coumarato CoA ligase* (*4CL*), *cinamato-4-hidroxilase* (*C4H*), *p-coumarato 3-hidroxilase* (*C3H*), *cinamoil-CoA redutase* (*CCR*), *álcool cinamil desidrogenase* (*CAD*), *ácido cafeico O-metiltransferase* (*COMT*) e *ferulato 5-hidroxilase* (*F5H*). Em todos os tecidos, a expressão de *C3H* e *CAD* foi muito baixa para permitir a identificação de diferenças entre os genótipos. Em entrenó imaturo (Tabela 13), é possível notar uma maior expressão de vários destes genes em *S. robustum*, especialmente *C4H*. Em entrenó intermediário (Tabela 14), dois genes do início da via, *PAL* e *4CL*, e dois genes chave para a síntese de siringil, *COMT* e *F5H*, estão regulados positivamente em *S. spontaneum*. Esse fato concorda com a observação de *S. spontaneum* ter mais lignina total (Figura 11A) e unidades de siringil em entrenó intermediário (Figura 11B). Note também que os genes *PAL* e, principalmente, *4CL*

já foram identificados nos *clusters* hierárquicos com maior expressão em *S. spontaneum* (Figuras 24 e 27).

Em entrenó maduro (Tabela 15), o perfil de expressão dos genes da via de biossíntese de lignina é mais similar entre os genótipos, com maior expressão de *COMT* em RB867515 e *S. officinarum* e *C4H* em *S. robustum*.

Tabela 13. Expressão gênica em entrenó imaturo. O sinal de referência é indicado por ++. Cada + significa uma diferença no valor de expressão de 0,5 em relação à referência.

Categoria	SAS	Gene	RB867515	<i>S. officinarum</i>	<i>S. robustum</i>	<i>S. spontaneum</i>
Biossíntese de lignina	SCCCLR1048D07.g	PAL	++	+++	++	+++
	SCCCCL3002A03.b	4CL	++	++++	++++	++
	SCCCRZ2C01A04.g	CCR	++	+++	++++	+++
	SCRFLR1012F12.g	COMT	+++	+++	++++	++
	SCEZHR1087F06.g	F5H	++	++	++++	+++
	SCSGFL4193B05.g	C4H	+	+	+++++	++
Fatores de transcrição	SCEQRT1031E06.g	ScNAC86 (NST1)	+	+++	+	+++++
	SCJLRZ1019A02.g	ScMYB45 (MYB58/63)	nd	nd	nd	nd
	SCEZLB1006E02.g	ScHB26 (KNAT7)	nd	nd	nd	nd
	SCCCLR1C05B03.g	ScMYB22 (ZmMYB31)	++	++	++	++
Metabolismo de Sacarose	SCCCLR1C06G07.g	Sucrose-phosphate synthase (SPS)	++	+++	+++	+++
	SCEPCL6023F02.g	Sucrose synthase (SS)	+++	+++	+++	++
	SCJLLR2020E06.g	Cytosolic fructose-1,6-bisphosphatase	+	++	+	+++++
	SCVPRT3086D06.g	Invertase	+	+++++	+	+++++
Geração de energia	SCCCLR1070D04.g	Pyruvate kinase	++++	+++	++	++
	SCMCLR1122H07.g	NADH dehydrogenase (Complex I)	++	++++	++	++++
Metabolismo de parede celular	SCVPLB1020G05.g	α -L-Arabinofuranosidase	++	++	++++	++++
	SCJLST1020H07.g	XTH	++	+	+++++	+++++
	SCCCCL4006H09.g	Expansin	++	+	+	+++++
	SCRLFL1004C08.g	Csl C	++	++	++	++
	SCAGLR1064F05.g	ScCesA3	++	++	+++	++
	SCJLLR1106H06.g	ScCesA7	+	+	++	++++
Transportador	SCAGLR2033E03.g	Aquaporin	++	++	++	++++

Nd, não determinado devido à vários picos na curva de dissociação

Tabela 14. Expressão gênica em entrenó intermediário. O sinal de referência é indicado por ++. Cada + significa uma diferença no valor de expressão de 0,5 em relação à referência.

Categoria	SAS	Gene	RB867515	<i>S. officinarum</i>	<i>S. robustum</i>	<i>S. spontaneum</i>
Biossíntese de lignina	SCCCLR1048D07.g	PAL	++	++	++	++++
	SCCCCL3002A03.b	4CL	++	++	++	++++
	SCCCRZ2C01A04.g	CCR	++	+++	+++	++
	SCRFLR1012F12.g	COMT	++	+	++	+++++
	SCEZHR1087F06.g	F5H	++	++	++	+++++
	SCSGFL4193B05.g	C4H	++	++	++++	++
Fatores de transcrição	SCEQRT1031E06.g	ScNAC86 (NST1)	++	+++++	++	+
	SCJLRZ1019A02.g	ScMYB45 (MYB58/63)	+++	++++	+	++
	SCEZLB1006E02.g	ScHB26 (KNAT7)	++	++	++	++
	SCCCLR1C05B03.g	ScMYB22 (ZmMYB31)	++	++	++	+++++
Metabolismo de Sacarose	SCCCLR1C06G07.g	Sucrose-phosphate synthase (SPS)	++	++++	++++	++++
	SCEPCL6023F02.g	Sucrose synthase (SS)	++	++++	++++	++++
	SCJLLR2020E06.g	Cytosolic fructose-1,6-bisphosphatase	++	++	++	+++++
	SCVPRT3086D06.g	Invertase	++	+++	++	+++
Geração de energia	SCCCLR1070D04.g	Pyruvate kinase	++	++	+	+++++
	SCMCLR1122H07.g	NADH dehydrogenase (Complex I)	++	+++++	++	+++++
Metabolismo de parede celular	SCVPLB1020G05.g	α -L-Arabinofuranosidase	++	++	++	++
	SCJLST1020H07.g	XTH	+++	++	+++	++
	SCCCCL4006H09.g	Expansin	+++++	++	++	++
	SCRLFL1004C08.g	Csl C	++	++	++	+
	SCAGLR1064F05.g	ScCesA3	++	++++	+++	+++
	SCJLLR1106H06.g	ScCesA7	++	++	++	++
Transportador	SCAGLR2033E03.g	Aquaporin	++	++	++	+++++

Tabela 15. Expressão gênica em entrenó maduro. O sinal de referência é indicado por ++. Cada + significa uma diferença no valor de expressão de 0,5 em relação à referência.

Categoria	SAS	Gene	RB867515	<i>S. officinarum</i>	<i>S. robustum</i>	<i>S. spontaneum</i>
Biossíntese de lignina	SCCCLR1048D07.g	PAL	++	++	+++	++
	SCCCCL3002A03.b	4CL	++	++	+++	++
	SCCCRZ2C01A04.g	CCR	++	++	+++	++
	SCRFLR1012F12.g	COMT	++++	+++++	+++	++
	SCEZHR1087F06.g	F5H	++	++	+++	+++
	SCSGFL4193B05.g	C4H	+	++	+++++++	++
Fatores de transcrição	SCEQRT1031E06.g	ScNAC86 (NST1)	++	+++++	+++	++++
	SCJLRZ1019A02.g	ScMYB45 (MYB58/63)	++	++	++	++
	SCEZLB1006E02.g	ScHB26 (KNAT7)	++	+++	+++++	+
	SCCCLR1C05B03.g	ScMYB22 (ZmMYB31)	+	+++	++	+++++
Metabolismo de Sacarose	SCCCLR1C06G07.g	Sucrose-phosphate synthase (SPS)	++	++	++++	++
	SCEPCL6023F02.g	Sucrose synthase (SS)	++	++++	++++	++
	SCJLLR2020E06.g	Cytosolic fructose-1,6-bisphosphatase	++	+++++	+++++	+++
	SCVPRT3086D06.g	Invertase	++	+++++	++++	+++++
Geração de energia	SCCCLR1070D04.g	Pyruvate kinase	+++	++	++	+++
	SCMCLR1122H07.g	NADH dehydrogenase (Complex I)	++	+++	+++	+++++
Metabolismo de parede celular	SCVPLB1020G05.g	α -L-Arabinofuranosidase	++	++	++++	++++
	SCJLST1020H07.g	XTH	++	+++++	+++	+
	SCCCCL4006H09.g	Expansin	+	++++	++++	++
	SCRLFL1004C08.g	Csl C	++	++	++++	++++
	SCAGLR1064F05.g	ScCesA3	++++	++++	++	++
	SCJLLR1106H06.g	ScCesA7	++	++++	+	+++
Transportador	SCAGLR2033E03.g	Aquaporin	++	++	+++	+++++

4.9.2. Expressão de fatores de transcrição

Vários fatores de transcrição (TFs) associados à regulação do metabolismo de parede celular já foram caracterizados em *Arabidopsis* e outras dicotiledôneas. Entretanto, como descrito na introdução, poucos TFs envolvidos no metabolismo de parede celular de gramíneas foram caracterizados. Na tentativa de identificar os TFs com correlação positiva com a lignificação, foi avaliada a expressão de alguns TFs de cana-de-açúcar homólogos putativos de *Arabidopsis* e milho identificados pelo programa Inparanoid (Tabela 16).

Tabela 16. Ortologia deduzida pelo programa Inparanoid.

SAS	Ortólogo putativo	Anotação	Organismo	Confiança*
SCEQRT1031E06.g	AT2G46770.1	NST1	<i>Arabidopsis thaliana</i>	1
SCEZLB1006E02.g	AT1G62990.1	KNAT7	<i>Arabidopsis thaliana</i>	1
SCJLRZ1019A02.g	AT1G16490.1/AT1G79180.1	MYB58/63	<i>Arabidopsis thaliana</i>	1
SCCCLR1C05B03.g	GRMZM2G050305_T01	MYB31	<i>Zea mays</i>	1

*Valor de confiança varia de 0 a 1, sendo que quanto mais próximo de 1, maior será a significância estatística da ortologia definida pelo programa.

O *NST1* atua como um regulador principal e também ativa a síntese de siringil em *M. truncatula* regulando a expressão de *F5H* (Zhao *et al.* 2010). Portanto, seria esperada uma maior expressão do homólogo de cana-de-açúcar de *NST1*, *ScNAC86* (Yilmaz *et al.* 2009), em *S. robustum* e *S. spontaneum*, uma vez que esses dois genótipos apresentaram maior expressão de *F5H* (Tabelas 13-15) e *S. spontaneum* possui mais unidades de siringil (Figura 11B). Porém, nota-se que *S. officinarum* foi o genótipo que apresentou a maior expressão de *ScNAC86* em tecidos em lignificação (Tabelas 14 e 15). Apenas em entrenó imaturo, *S. spontaneum* apresentou maior expressão deste gene (Tabela 13). Além disso, *S. spontaneum* também apresenta um aumento de siringil ao longo do desenvolvimento do colmo (Figuras

11B e 12D e H), não acompanhado do aumento de expressão de *ScNAC86*, a qual se mantém relativamente constante em todos os entrenós analisados, o que pode ser observado ao se fazer uma normalização exclusiva para os entrenós de *S. spontaneum* (Figura 29).

Figura 29. Análise da expressão gênica de *ScNAC86* (homólogo de *AtNST1*) com normalização para comparação entre os diferentes entrenós de *S. spontaneum*. Letras diferentes no topo da barra representam médias diferentes por One-way ANOVA seguido de teste Tukey. Barra de erro=erro padrão; N=3; $p < 0,05$.

O *MYB58/63* de *Arabidopsis* ativa genes de biossíntese de lignina ao se ligar nos elementos AC, ou SMREs, dos promotores alvos. Novamente, não é observada uma correlação entre o homólogo de cana-de-açúcar de *AtMYB58/63*, *ScMYB45* (Yilmaz *et al.* 2009), e a expressão de genes de biossíntese de lignina. Em entrenó intermediário, *S. officinarum* e RB867515 apresentaram os maiores valores de expressão, não acompanhada da regulação positiva de genes da biossíntese de lignina (Tabela 14).

Outro TF, *KNAT7*, repressor da síntese de parede celular secundária em *Arabidopsis* e *Populus* (Li *et al.* 2012) também foi analisado. Em entrenó intermediário (Tabela 14), não há diferenças nos níveis de expressão do homólogo de cana-de-açúcar de *KNAT7*, *ScHB26* (Yilmaz *et al.* 2009), entre os quatro genótipos analisados. Para entrenó maduro (Tabela 15),

é possível notar uma regulação positiva de *ScHB26* em *S. robustum* não acompanhada da repressão dos genes de biossíntese de lignina. Ao contrário, ocorre uma regulação positiva de *C4H* (Tabelas 13-15) que também é regulado positivamente em *Arabidopsis* mutante com perda de função de *KNAT7* (Li *et al.* 2012).

Para o homólogo putativo de cana-de-açúcar do repressor de milho *ZmMYB31*, *ScMYB22* (Yilmaz *et al.* 2009), há uma expressão relativamente similar em todos os entrenós de *S. officinarum*, *S. robustum* e RB867515 (Tabelas 13-15). *S. spontaneum* apresentou uma expressão cinco vezes maior em entrenó intermediário em comparação com os outros genótipos (Tabela 14), mantendo o nível de expressão em entrenó maduro (Tabela 15). Conforme supracitado, o entrenó intermediário de *S. spontaneum* possui a maior expressão dos genes de biossíntese de lignina o que indica uma falta de repressão desses genes por *ScMYB22*.

4.9.3. Genes relacionados ao metabolismo de parede celular

Vários experimentos de variedades contrastantes para Brix (Waclawovsky *et al.*, dados não publicados) mostraram uma *alpha-L-arabinofuranosidase* diferencialmente expressa em entrenós imaturos. Isso levantou a hipótese de que esta mesma enzima seria diferencialmente regulada nas espécies ancestrais. De fato, em entrenó imaturo e maduro (Tabelas 13 e 15) nos genótipos de baixo Brix essa enzima está regulada negativamente, sugerindo uma possível correlação inversa com o acúmulo de sacarose no colmo, mesmo não estando diretamente envolvida no metabolismo de sacarose. Os resíduos de arabinose presentes na hemicelulose podem ser esterificados a resíduos de ácido ferúlico, formando uma ponte de ligação entre a hemicelulose e a lignina (Ralph *et al.* 2004a; Burr & Fry 2009),

aumentando a recalcitrância da biomassa. Nesse contexto, é possível que o aumento da expressão desta enzima esteja relacionado à diminuição da recalcitrância da biomassa, uma vez que a remoção dos resíduos de arabinose pela *alpha-L-arabinofuranosidase* diminui as possibilidades de ligações entre a hemicelulose e a lignina. Isto pode indicar que esta enzima esteja envolvida no rearranjo da parede celular e, possivelmente, na alteração das propriedades da parede celular.

A enzima xiloglucano endo-transglicosilase/hidrolase (XTH) atua no remodelamento das cadeias de xiloglucano na parede celular e está envolvida na expansão celular, mobilização de carboidratos de reserva da parede celular, germinação da semente, amadurecimento dos frutos, entre outros processos (Buckeridge 2010; Eklof & Brumer 2010). Em cana-de-açúcar submetida à elevação do CO₂ atmosférico, uma XTH e uma *alpha-L-arabinofuranosidase* foram observadas como diferencialmente expressas (De Souza *et al.* 2008). As expansinas são proteínas não enzimáticas relacionadas à expansão celular e acredita-se que elas atuem rompendo as ligações de hidrogênio entre os polissacarídeos de parede celular promovendo o relaxamento da parede (McQueen-Mason & Cosgrove 1994; Cosgrove & Moore 2002). Uma vez que a expansão celular pode estar relacionada ao acúmulo de sacarose, a expressão destes genes foi analisada. Em entrenó intermediário (Tabela 14), não houve diferenças na expressão de *XTH* entre as amostras. Em entrenó imaturo (Tabela 13), genótipos de alto e baixo Brix mostraram um perfil oposto na expressão destes genes. Contudo, em entrenó maduro (Tabela 15), apenas *S. officinarum* apresentou maior expressão de *XTH*. *S. spontaneum* apresentou maior expressão de uma expansina em entrenó imaturo (Tabela 13), enquanto RB867515 apresentou uma expansina regulada positivamente em entrenó intermediário (Tabela 14). Esses resultados sugerem que *S. spontaneum* pode ter expansão celular mais ativa em entrenós mais jovens e RB867515 em

entrenó intermediário, enquanto *S. robustum* e *S. officinarum* não mostraram um padrão definido. Isto pode explicar a baixa altura das plantas destes dois últimos genótipos (Figura 8B). Ainda, o gene *ClsC*, pertencente à subfamília envolvida na síntese de xiloglucanos (Cocuron *et al.* 2007), está mais expresso em plantas de baixo Brix (Tabela 15) e isso sugere uma possível correlação negativa com o acúmulo de sacarose. Os resultados mostram um padrão de expressão oposto ao de *XTH*, uma vez que *CslC* está regulada positivamente em entrenó maduro nas plantas de baixo Brix (Tabela 15).

O perfil de expressão de duas celulose sintases (*CesA*) também foi avaliado. Os complexos de síntese de celulose (rosetas) atuam com três diferentes subunidades *CesA*, sendo algumas específicas de parede primária e outras de parede secundária. Foram avaliados dois genes *CesA* de cana-de-açúcar sendo, após uma análise filogenética (Apêndice 3), um homólogo de *OsCesA3* e outro de *OsCesA7*, então denominados *ScCesA3* e *ScCesA7*, possivelmente específicos de parede primária e secundária, respectivamente, de acordo com a especificidade dos homólogos de arroz. Foi observado que *ScCesA3* apresenta um perfil de expressão contrastante entre plantas de alto e baixo Brix em entrenó maduro (Tabela 15). Para a *ScCesA7*, em entrenó imaturo (Tabela 13), *S. spontaneum* apresentou maior expressão que as demais, sugerindo que este genótipo iniciaria o programa de síntese de parede celular secundária em tecidos mais jovens. Para confirmar o padrão de expressão entre *CesA* de parede primária e secundária, foi realizada uma nova normalização dos resultados de qPCR para identificar a diferença de expressão entre os entrenós imaturo e o intermediário (Figura 30) e foi observado que *ScCesA7* tem sua expressão fortemente induzida em entrenó intermediário em relação ao imaturo.

Figura 30. Análise da expressão gênica das celulose sintase *ScCesA3* (barras brancas) e *ScCesA7* (barras pretas). Note que o eixo X está em escala logarítmica. Barra de erros=erro padrão; N=3.

4.9.4. Expressão de genes do metabolismo de sacarose e de geração de energia

Várias enzimas estão envolvidas na síntese e degradação de sacarose. A sacarose fosfato sintase (SPS) é uma das enzimas mais importantes na síntese da sacarose catalisando, a partir de UDP-glicose e frutose-6-fosfato, a formação de sacarose-fosfato que será defosforilada posteriormente. A degradação da sacarose produz UDP-glicose e frutose quando catalisada pela sacarose sintase (SS) e glicose e frutose pela ação da invertase. Outra enzima importante é a frutose-1,6-bisfosfatase (FBPase) citosólica, que é responsável pela

primeira reação irreversível na conversão de triose-fosfato em sacarose. Curiosamente, não foi observado nenhum padrão ligando o acúmulo de sacarose e a expressão destes genes, de forma que os genótipos de alto e baixo Brix não apresentaram perfil de expressão contrastante nos entrenós (Tabelas 13-15) e em folha (Tabela 17). De um modo geral, RB867515 apresentou os menores valores de expressão para estes genes em todos os tecidos analisados, independente se eram genes de síntese ou degradação de sacarose (Tabelas 13-15 e 17).

Também foram avaliados dois genes relacionados à geração de energia da célula: piruvato quinase (SCCCLR1070D04.g) que catalisa a última etapa da glicólise, e a NADH desidrogenase (SCMCLR1122H07.g) que faz parte do complexo I da cadeia de transporte de elétrons mitocondrial, responsável pela oxidação do NADH e redução da ubiquinona. Tanto em entrenó intermediário quanto em entrenó maduro (Tabelas 14 e 15), *S. spontaneum* apresentou a tendência de maior expressão desses genes que as demais plantas, principalmente para a NADH desidrogenase, sugerindo um possível maior gasto energético desse genótipo.

4.9.5. Expressão gênica em folha

A análise da expressão de genes relacionados à captação de luz e fixação de carbono foi realizada para identificar diferenças moleculares associadas às taxas fotossintéticas. Entre todos os genes analisados, apenas uma proteína do centro de reação do fotossistema II (pspB), envolvida com a evolução de oxigênio (Bricker *et al.* 2012), e uma subunidade da Rubisco (Rbcs) foram menos expressas em *S. spontaneum* (Tabela 17), correlacionando com

a menor taxa fotossintética deste genótipo (Figura 9A). Entretanto, outros genes analisados não seguiram este padrão.

As enzimas piruvato ortofofato diquinase (PPDK), anidrase carbônica (CA) e fosfoenol piruvato carboxilase (PEPC) catalisam reações sequenciais no início do ciclo C4. PPDK, utilizando piruvato e ATP como substrato, catalisa a formação de fosfoenolpiruvato (PEP), enquanto CA catalisa a hidratação de CO_2 a HCO_3^- que, juntamente com o PEP, será substrato para a PEPC produzir oxalacetato. O perfil destes três genes foi similar nos genótipos, com *S. officinarum* e *S. spontaneum* apresentando os menores níveis de expressão (Tabela 17).

Tabela 17. Expressão gênica em folha. O sinal de referência é indicado por ++. Cada + significa uma diferença no valor de expressão de 0,5 em relação à referência.

Categoria	SAS	Gene	RB867515	<i>S. officinarum</i>	<i>S. robustum</i>	<i>S. spontaneum</i>
Captação de luz	SCJLLR1102C07.g	Photosystem II reaction center PsbP family protein	++	++	+++	+
	SCJFLR1073E07.g	Photosystem II 10 kDa polypeptide	++	++	+++	+++
Fixação de carbono	SCCCLR1001E04.g	Rubisco small subunit (RbcS)	+++	++	++	+
	SCEQLB1063G04.g	Carbonic anhydrase (AC)	++++	++	+++++	++
	SCJFRT1060F11.g	Pyruvate phosphate dikinase (PPDK)	+++	++	++	++
	SCRULR1020D09.g	Phosphoenolpyruvate carboxylase (PEPC)	++++	++	++++	++
Metabolismo de Sacarose	SCCCLR1C06G07.g	Sucrose-phosphate synthase (SPS)	++	++++	++++	++
	SCEPCL6023F02.g	Sucrose synthase (SS)	++	+++++	++++	+
	SCJLLR2020E06.g	Cytosolic fructose-1,6-bisphosphatase	++	+++	++++	+++
	SCVPRT3086D06.g	Invertase	++	+++++	+	+++++
Transportador	SCEQRT2028C04.g	Aquaporin	+++++	+	+++++	+++++
	SCAGLR2033E03.g	Aquaporin	++	+	++++	++++

O perfil de expressão de aquaporinas, proteínas envolvidas no movimento de água e solutos entre células e compartimentos celulares, também foi analisado (Tabela 17). Claramente, há um padrão distinto entre as plantas de alto e baixo Brix, sendo que *S. spontaneum* e *S. robustum* apresentaram uma regulação positiva de duas aquaporinas em folha. O mesmo foi observado para uma aquaporina em entrenó imaturo e maduro (Tabelas 13 e 15).

4.10. Construção de uma rede de coexpressão de genes de metabolismo de lignina

Redes de coexpressão são ferramentas importantes para identificação de genes associados a determinados processos celulares. Essa análise se vale do conceito *guilt-by-association* (Quackenbush 2003), no qual genes que desempenham funções em um mesmo processo biológico tendem a ter padrão de expressão similar. Assim, é possível inferir uma função ou envolvimento num processo biológico a um gene não-caracterizado conforme sua expressão siga um padrão exibido por genes conhecidos. Essa análise já é muito utilizada para identificar envolvimento de genes em processos biológicos diversos em diferentes organismos, como mamíferos, insetos, levedura e plantas (Stuart *et al.* 2003; Lee *et al.* 2004; Ruprecht *et al.* 2011; Movahedi *et al.* 2012). Em plantas, já foi demonstrada a utilidade das redes de coexpressão para análise comparativa entre espécies através da integração da homologia e da expressão dos genes (Movahedi *et al.* 2012) e também para a identificação de genes envolvidos em diversos processos celulares, incluindo o metabolismo de parede celular (Ruprecht *et al.* 2011; Ruprecht & Persson 2012; Wang *et al.* 2012).

Como abordado na introdução, o conhecimento do metabolismo de parede celular de gramíneas ainda está muito aquém do que se sabe para outras plantas e a identificação de novos genes de gramíneas envolvidos nesse processo irá ajudar no avanço desse conhecimento. Assim, com base no conceito *guilt-by-association*, para identificar genes desconhecidos e fatores de transcrição candidatos à participantes do metabolismo de parede celular foi gerada uma rede de coexpressão de genes utilizando como “isca” o gene da via de fenilpropanoides *4CL* (SCCCCL3002A03.b), o qual foi identificado com padrão de expressão diferencial nos clusters de genes de parede celular (Figuras 24 e 27) e nos experimentos de qPCR (Tabela 14). Entretanto, pela falta de ferramentas disponíveis que utilizem dados de expressão de cana-de-açúcar, as análises foram realizadas com dados de expressão de milho, a gramínea mais próxima de cana-de-açúcar que possuía a ferramenta disponível. Assim, a estratégia adotada foi utilizar o homólogo do SAS SCCCCL3002A03.b (*4CL*) de milho como isca com auxílio da ferramenta *Genevestigator* (Hruz *et al.* 2008) e identificar posteriormente os homólogos correspondentes de cana-de-açúcar da lista de genes coexpressos de milho (Figura 31 e Tabela 18). Essa ferramenta utiliza dados públicos de centenas de experimentos de microarranjo, em diversas condições diferentes, para analisar o padrão de expressão dos genes e calcular a correlação entre a expressão dos mesmos. Após a análise da lista de genes identificados, a validade dessa rede pode ser verificada pela identificação de vários genes de biossíntese de parede celular (*PAL*, *COMT*, *CAD*, *CesA*, glicosiltransferases, entre outros, Tabela 18), concordando com o a hipótese *guilt-by-association*. Além dos genes já esperados, podemos destacar alguns genes desconhecidos (*Unknown*) e fatores de transcrição (*Platz*, *knotted-1-like* e *MYB*) que também aparecem na rede, os quais podem ser genes de gramíneas envolvidos na biossíntese e regulação da parede celular. Vale ressaltar que o TF da família *knotted-1-like*

(SCEZLB1006E02.g) é o homólogo putativo do regulador da biossíntese de parede celular de *Arabidopsis*, *KNAT7*. TFs da família Platz são pouco conhecidos, mas sabe-se que são capazes de se ligarem em regiões ricas em AT (Nagano *et al.* 2001). Entre os SAS *unknown*, SCJLRT1015F06.g e SCCCL4017H11.g possuem os domínios de função desconhecida DUF563 e DUF231, respectivamente. Ambos apresentam indícios de estar envolvido no metabolismo de parede celular pois DUF563 é encontrado em xilosiltransferases de *Trichomonas vaginalis* (Hwa *et al.* 2012) e arroz (Chiniquy *et al.* 2012) enquanto DUF231 é importante para a acetilação de xilanas (Yuan *et al.* 2013). Os demais SAS *unknown* (SCCCAM2001D03.g e SCSGST3118D04.g) não possuem nenhum domínio conhecido.

Figura 31. Rede de genes coexpressos com o gene 4CL de milho. Ponto central preto indica o gene 4CL e círculos concêntricos mostram os limites do coeficiente de correlação de Pearson até o *cutoff* de 0.8. Genes com correlação mútua acima de 0.8 estão conectados por linhas azuis.

Tabela 18. Lista de genes de milho coexpressos com o gene 4CL e seus respectivos homólogos de cana-de-açúcar. Os genes estão dispostos de forma decrescente pelo valor de coeficiente de correlação. Os genes podem ser identificados na Figura 31 pela numeração da primeira coluna.

Nº	Transcrito de milho	Anotação	Homólogo de cana
1	GRMZM2G000812	Sugar transporter	SCJLRZ1021E01.g*
2	GRMZM2G074604	Phenylalanine ammonia-lyase (PAL)	SCJFLR1017B11.g*
3	GRMZM2G150302	Glycosyltransferase GT43	SCCCRZ2C01D11.g*
4	GRMZM2G035584	Hydroxyanthranilate hydroxycinnamoyltransferase 2	SCCCCL4009E02.g
5	GRMZM5G844562	Cinnamyl alcohol dehydrogenase (CAD)	SCEPRZ1011A02.g
6	GRMZM2G029048	Phenylalanine ammonia-lyase (PAL)	SCEQRT1024E12.g
7	Zm.1502.1.A1_at	Caffeic acid 3-O-methyltransferase (COMT)	SCRFLR1012F12.g*
8	GRMZM2G028104	3-N-debenzoyl-2-deoxytaxol N-benzoyltransferase	SCJFRT2054E01.g
9	GRMZM2G019404	Plasma-membrane H ⁺ ATPase2	SCCCRT1002F07.g*
10	GRMZM2G094168	Platz transcription factor	SCCCLR2C03B01.g
11	GRMZM2G127948	Caffeoyl-CoA O-methyltransferase (CCoAOMT)	SCCCLR1069B09.g
12	Zm.8633.1.A1_x_at	Glycosyltransferase GT43	SCCCRZ2C01D11.g*
13	GRMZM2G159759	GDSL-like Lipase/Acylhydrolase	SCEPLB1044A08.g
14	GRMZM2G151519	ZIM motif family protein	SCSFRT2072D04.g
15	GRMZM2G006894	Proton-exporting ATPase Fragment	SCCCRT1002F07.g
16	GRMZM2G441347	Phenylalanine ammonia-lyase (PAL)	SCJLRT1013C01.g
17	GRMZM2G159724	Malic enzyme	SCCCCL2001A05.b*
18	GRMZM2G059845	Cell wall-related glycosyltransferase family 47 (GT47)	SCSBHR1053F06.g
19	GRMZM2G122937	Remorin	SCJFLR1013D11.g
20	Zm.118693	Probable peptide/nitrate transporter	SCACLR1057B09.g*
21	GRMZM5G860590	Development and cell death domain	SCEQLR1091B08.g
22	GRMZM2G132866	Beta-1,3-galactosyltransferase	SCSGLR1045E04.g
23	Zm.38134	Unknown	SCJLRT1015F06.g*
24	GRMZM2G125923	Prephenate dehydratase (phenylalanine biosynthesis)	SCEQRT2093C02.g*
25	GRMZM5G862540	UDP-glucose dehydrogenase	SCQGLR1019G02.g
26	GRMZM2G066165	Unknown	SCCCCL4017H11.g
27	GRMZM2G068158	Putative methyltransferase	SCRFLB1053H12.g*
28	GRMZM2G042818	3-hydroxyisobutyryl-CoA hydrolase/ catalytic	SCBFLR1083F12.g
29	Zm.13519	Unknown	nd
30	GRMZM2G007466	Protein tyrosine kinase	SCRUSB1078B08.g
31	GRMZM2G444845	BAG domain (Chaperone activity)	nd
32	GRMZM2G165357	TDP-glucose dehydratase	SCCCRZ2001D09.g*
33	GRMZM2G141473	Aldehyde oxidase-2	SCCCCL4011A06.g
34	GRMZM2G424832	Cellulose synthase-4	SCCCLR1066B10.g*
35	GRMZM2G131205	Cinnamoyl CoA reductase (CCR)	SCBFRT1064A01.g
36	GRMZM2G159431	Homeobox protein knotted-1-like (KNAT7)	SCEZLB1006E02.g
37	GRMZM2G151516	Unknown	SCSGST3118D04.g
38	GRMZM2G018241	Cellulose synthase-9	SCCCLR1065F07.g*
39	Zm.92707	Platz transcription factor	SCCCLR1078B06.g*
40	GRMZM2G142898	Cellulose synthase catalytic subunit 12	SCCCLR2C02H01.g
41	GRMZM2G150187	GDSL-like Lipase/Acylhydrolase	SCVPST1059E11.g
42	GRMZM2G017520	Myb-like DNA-binding domain	SCJFRZ2015F02.g

Tabela 18. Continuação.

Nº	Transcrito de milho	Anotação	Homólogo de cana
43	GRMZM5G803318	Saposin-like type B domain	SCCCLR1070A11.g
44	Zm.3374.1.S1_at	Protease inhibitor/seed storage/LTP family protein	SCCCRZ1001H02.g*
45	GRMZM2G144155	Ankyrin protein kinase-like	SCEZLB1013G06.g
46	GRMZM2G329181	Fasciclin-like arabinogalactan protein	SCEPRZ1010A01.g
47	Zm.19108	Flavonoid 3',5'-hydroxylase	SCEQRT1026H08.g*
48	GRMZM2G104449	Unknown	SCCCAM2001D03.g

*Homologia identificada por Blast. Os demais foram identificados pelo programa Inparanoid nd, não determinado

Um total de 46 genes dos 48 presentes na lista possuem algum homólogo ou gene com alta identidade em cana-de-açúcar sendo que 20 estão presentes no chip Agilent. Assim, da mesma forma conduzida na seção 4.8.3, foram gerados *clusters* hierárquicos com os dados de sinal de expressão desses genes para entrenó imaturo e intermediário para identificar os genes que possuem expressão diferencial entre nos genótipos ancestrais (Figuras 32 e 33). A maioria desses SAS (14/20 e 16/20 para entrenó imaturo e intermediário, respectivamente) apresenta maior expressão em *S. spontaneum*. Por outro lado, apenas 2 ou 3 genes apresentam maior expressão em *S. officinarum*.

Figura 32. Cluster hierárquico dos dados de sinal de expressão de entrenó imaturo dos genes da lista de coexpressos.

Figura 33. Cluster hierárquico dos dados de sinal de expressão de entrenó intermediário dos genes da lista de coexpressos.

4.11. Identificação de promotores de genes associados ao metabolismo de parede celular em cana-de-açúcar

Além dos genes, a identificação de promotores associados a características de interesse também é importante para se alcançar o fenótipo pretendido através da manipulação genética. Desse modo, foram escolhidos alguns SAS com padrão de expressão bem característico nos *clusters* hierárquicos (Figuras 24, 26-28), com expressão distinta em *S. spontaneum* e *S. officinarum*, e seus promotores foram buscados no genoma de cana-de-açúcar nos banco de dados de sequenciamento do *whole genome shotgun* e dos BACs, como descrito na metodologia (Seção 3.15). Os genes foram selecionados principalmente nos *clusters* onde *S. spontaneum* apresentava maior expressão, uma vez que esse genótipo parece ter o desenvolvimento precoce de parede celular secundária em relação às demais plantas analisadas. Assim, fatores da família NAC e MYB com esse perfil de expressão podem ser bons candidatos à reguladores desse processo, como ocorre em outras espécies. Além desses TFs, também foram selecionados genes da via de fenilpropanoides (*CAD*, *COMT*, *CCoACOMT*, *C4H*, *F5H*, *PAL*, *4CL* e *CCR*) e o gene de função desconhecida *kobito*. Para os genes com expressão detectada do transcrito antissenso, foram buscadas ambas as sequências à jusante do 3' quanto à montante do 5' do transcrito. Das 36 sequências escolhidas, para 25 não foi possível identificar nenhuma sequência promotora. Nas demais, conseguiu-se identificar 9 SAS com sequência promotora do transcrito senso, 1 SAS apenas com sequência promotora do antissenso e 1 SAS com ambas. Dentre esses SAS, temos apenas os genes *COMT*, *kobito* e TFs NAC e MYB. De forma semelhante, foram buscados os promotores (apenas senso) dos SAS homólogos aos de milho identificados pela análise de coexpressão encontrando-se apenas para 10 SAS dos 46 presentes na lista (Tabela 18),

sendo de 6 categorias diferentes. Em todos os casos, a sequência da região promotora foi limitada em 1,5kb. Entretanto, devido ao tamanho relativamente curto dos *contigs* do *shotgun*, a sequência promotora identificada na maior parte dos casos possuía tamanho entre 200 e 400 nt (Tabela 19 e 20).

Para verificar a presença nos promotores de sítios de ligação de fatores de transcrição relacionados ao metabolismo de parede celular secundária, as sequências foram analisadas utilizando-se a ferramenta FIMO (Grant *et al.* 2011) buscando-se pelos elementos regulatórios SNBE (Zhong *et al.* 2010), SMRE (Zhong & Ye 2012) e M46RE (Kim *et al.* 2012), os quais se ligam aos TFs da família NAC, MYB e ao *AtMYB46*, respectivamente (Tabela 19 e 20). Essa análise, além de caracterizar parcialmente esses promotores, acrescenta uma evidência importante da possível participação dos genes da análise de coexpressão (Figura 31) nesse processo biológico, dando uma maior confiabilidade a esses dados. Tanto os sítios de ligação que possuem a sequência consenso perfeita quanto aqueles com 1 *mismatch* foram contabilizados (Tabela 19 e 20).

Nos genes escolhidos nos *clusters* hierárquicos (Tabela 19), o sítio de ligação M46RE aparece apenas em poucos genes, como o MYB SCCCCL3004A05.b e o NAC SCJFRZ2014D06.g. Já o SMRE aparece em mais ocasiões, estando presente nos promotores da *COMT*, 3 MYBs (SCCCCL3004A05.b, SCEQRT2027F03.g e SCMCRT2105F02.g) e 1 NAC (SCJFRZ2014D06.g). O gene *COMT* de milho também possui SMREs em seu promotor (Fornale *et al.* 2010) diferentemente da *COMT* de *Arabidopsis*, a qual não possui SMREs (Raes *et al.* 2003; Fornale *et al.* 2010), sugerindo uma divergência entre gramíneas e *Arabidopsis*. Os SNBEs estão presentes em promotores de 1 MYBs (SCCCCL3004A05.b) e 1 NAC (SCACLR2007C06.g), sugerindo que estes possam ser constituintes do segundo nível de regulação, alvos diretos dos SWNs. A presença desses sítios de ligação ajudam a dar suporte

à hipótese na qual esses TFs possam ser reguladores biossintese da parede celular secundária. O promotor de transcrito MYB antissenso (SCACST3159E04.g) apresenta 4 SNBEs, mesmo sendo uma sequência curta (300 nt). Isso levanta a hipótese que a regulação da biossintese de parede celular secundária possa ter influencia de mecanismos via antissenso. Para o MYB SCJLRZ1019A02.g, homólogo do *AtMYB58/63*, nenhum sítio de ligação foi identificado nos 1500 nt do promotor. O mesmo acontece com kobito, mas, diferentemente de SCJLRZ1019A02.g, possui apenas 200 nt de sequência promotora identificada.

Tabela 19. Número de cis-elementos de parede celular nos promotores dos genes escolhidos pelos clusters hierárquicos (Figuras 24, 26-28). AS, antissenso.

SAS	Categoria	Anotação	Tamanho	M46RE*	SMRE*	SNBE*
SCEQLB2018G06.g	Cell wall Metabolism	AS - COMT	745 nt	0 (6)	0 (11)	0 (5)
SCEQLB2018G06.g	Cell wall Metabolism	COMT	1500 nt	0 (21)	3 (34)	0 (33)
SCACST3159E04.g	Signal transduction	AS - MYB	300 nt	0 (3)	0 (10)	4 (0)
SCCCCL3004A05.b	Signal transduction	MYB	900 nt	1 (3)	1 (10)	2 (14)
SCEQRT1030H12.g	Signal transduction	MYB	600 nt	0 (5)	0 (7)	0 (11)
SCEQRT2027F03.g	Signal transduction	MYB	300 nt	0 (1)	1 (0)	0 (3)
SCJLRZ1019A02.g	Signal transduction	ScMYB45	1500 nt	0 (14)	0 (21)	0 (33)
SCMCRT2105F02.g	Signal transduction	MYB	214 nt	0 (1)	1 (5)	0 (2)
SCACLR2007C06.g	Signal transduction	NAC	450 nt	0 (0)	0 (1)	5 (0)
SCCCRZ1004G12.g	Signal transduction	NAC	650 nt	0 (2)	0 (5)	0 (9)
SCJFRZ2014D06.g	Signal transduction	NAC	1500 nt	3 (13)	6 (22)	0 (21)
SCMCRT2087B09.g	Unknown function	Kobito	200 nt	0 (2)	0 (2)	0 (4)

*Número de cis-elementos identificados com sequência que se encaixa no consenso. Números entre parênteses indicam o número de cis-elementos identificados com 1 *mismatch* em relação ao consenso.

Nos promotores dos genes coexpressos com *4CL* (Tabela 20), apenas os genes com sequências promotoras curtas (<350 nt) não apresentaram algum dos 3 sítios de ligação analisados. Genes da via de lignina (*CCR* e *PAL*) e um gene envolvido na formação de precursores de carboidratos de hemicelulose, *UDP-glicose desidrogenase*, apresentaram SNBEs e, principalmente, sítios de ligação de fatores MYB (Tabela 20). Uma *Akyrin-kinase like*

apresentou sítios M46RE e SMRE; o MYB SCJFRZ2015F02.g apresenta 3 sítios SNBEs, sugerindo que seja alvo dos SWNs (Tabela 20), assim como alguns MYBs da Tabela 19. Curiosamente, uma proteína com função desconhecida apresenta os três sítios analisados (Tabela 20). Assim, essa proteína se torna um alvo interessante para análises posteriores para comprovar um envolvimento na biossíntese de parede celular secundária.

Tabela 20. Número de cis-elementos de parede celular nos promotores dos genes da rede de coexpressos (Tabela 18).

SAS	Categoria	Anotação	Tamanho	M46RE*	SMRE*	SNBE*
SCSGLR1045E04.g	Carbohydrate Metabolism	Beta-1,3-galactosyltransferase	265	0 (2)	0 (2)	0 (1)
SCQGLR1019G02.g	Carbohydrate Metabolism	UDP-glucose dehydrogenase	1014	6 (5)	4 (18)	2 (16)
SCBFRT1064A01.g	Cell wall Metabolism	CCR	1157	7 (8)	6 (21)	1 (20)
SCJLRT1013C01.g	Cell wall Metabolism	PAL	1523	2 (21)	4 (43)	3 (32)
SCBFLR1083F12.g	Lipid metabolism	3-hydroxyisobutyryl-CoA hydrolase	350	0 (1)	0 (7)	0 (2)
SCEZLB1013G06.g	Others	Ankyrin kinase-like	264	1 (1)	1 (2)	0 (2)
SCSFRT2072D04.g	Others	Tify motif family protein	310	0 (0)	0 (1)	0 (0)
SCJFRZ2015F02.g	Signal transduction	Myb-like domain	1256	1 (12)	1 (20)	3 (18)
SCSGST3118D04.g	Unknown	Hypothetical protein	1539	4 (13)	4 (31)	3 (32)

*Número de cis-elementos identificados com sequência que se encaixa no consenso. Números entre parênteses indicam o número de cis-elementos identificados com 1 *mismatch* em relação ao consenso.

4.12. Construção de bibliotecas de cDNA *full length* e análise do transcriptoma de *S. officinarum*, *S. spontaneum* e do híbrido SP803280 por sequenciamento de RNA em larga escala (FL-RNAseq)

Com o objetivo de aprofundar as análises de transcriptoma e genômica funcional de cana-de-açúcar e dos genótipos ancestrais e, principalmente, produzir um catálogo referência de genes para definir o ORFeoma de cana-de-açúcar, foram produzidas

bibliotecas de cDNA *full length* a partir de tecido foliar das duas principais espécies ancestrais, *S. officinarum* e *S. spontaneum*, e também de folha, entrenó imaturo e intermediário do híbrido comercial SP803280. Esse genótipo comercial foi escolhido uma vez que aproximadamente metade das bibliotecas que compõem o SUCEST (Vettore *et al.* 2001) são dessa variedade, além do projeto de sequenciamento do genoma da cana-de-açúcar (<http://sugarcanegenome.org>) também utilizar a SP803280. Isso irá facilitar análises posteriores, como a identificação de promotores, mapeamento de éxons, separação de alelos e na própria montagem do genoma. Além disso, as análises das bibliotecas de cDNA *full length* irão ajudar futuros estudos de identificação dos alelos mais expressos, SNPs, variantes de *splicing*, além de ser um catálogo de proteínas completas que podem ser facilmente transferidas para vetores de expressão, visando tanto a expressão heteróloga quanto a transformação de plantas.

As bibliotecas foram produzidas em parceria com a empresa *Life Technologies*, pelos pesquisadores Pei-Zhong Tang e Antje Taliana, de forma a selecionar apenas RNAs que continham cauda poliA e estrutura cap-5' (ver Metodologia), eliminando qualquer tipo de RNA ribossomal e mRNAs truncados, produzindo, assim, bibliotecas enriquecidas com cDNAs *full length*. Após a síntese do cDNA, as cinco amostras foram utilizadas para (i) sequenciamento na plataforma 454 da Roche (IQ-USP), (ii) sequenciamento na plataforma *Ion Personal Genome Machine* (PGM, *Life Technologies*) e (iii) clonagem em *E. coli*. Para esta última etapa, o cDNA foi ligado à adaptadores contendo a sequência de recombinação *attB1* (Sistema Gateway, Invitrogen) e uma sequência específica para cada biblioteca (*barcode*) e então clonado no vetor de entrada pENTR 222. Os clones em *E. coli* das cinco bibliotecas foram reunidos em um único *pool* e transferidos para o vetor pCMVSPORT6.1 e posteriormente divididos em duas frações, S e L (*small* e *large*), de acordo com o tamanho

médio do inserto (Tabela 21). Para validar se os clones apresentavam cDNA realmente *full length*, foram feitas estrias das duas frações das bibliotecas e 30 clones foram selecionados aleatoriamente, 15 de cada fração, e sequenciados pelo método de Sanger a partir das duas extremidades do inserto (Tabela 21). Excluindo-se as sequências repetidas e com baixa qualidade, 21 sequências restantes foram alinhadas por BlastX contra bancos de proteínas *full length* de sorgo e trigo (TRIFLDB) e contra o banco de dados do NCBI alcançando-se ótima validação, com 90% de sequências possivelmente *full length* (Tabela 21). Sequências que apresentaram tamanho suficiente para cobrir toda a ORF e parte das UTRs da proteína homóloga identificada por BlastX foram consideradas *full length*, como exemplificado na Figura 34, uma vez que o objetivo principal é obter as sequências das ORFs.

Tabela 21. Validação das bibliotecas de cDNA full length (FL).

Fração	Número de Clones	Tamanho médio do inserto	Clones analisados	Transcritos específicos de cana	FL putativos	Não FL	% de FL putativos
L	3×10^8	1.4 kb	8	0	8	0	90%
S	7.5×10^8	1.1 kb	13	1	11	1

Figura 34. Alinhamento entre sequência do clone da biblioteca FL e o transcrito homólogo de sorgo. A linha preta na parte superior da figura indica a posição no cromossomo 7 de sorgo. Setas em azul indicam a sequência do clone de cana-de-açúcar. Setas em cinza e amarelo indicam a UTR e a ORF, respectivamente, do transcrito de sorgo.

Foram realizadas três corridas na plataforma 454 e cinco na plataforma Ion PGM e todos os *reads* foram utilizados para uma montagem *de novo* utilizando o software Trinity (Grabherr *et al.* 2011) gerando 195.765 *contigs* consenso (Tabela 22). Esses *contigs* foram mapeados contra bancos de genes preditos das gramíneas milho (*Zea mays*), sorgo (*Sorghum bicolor*), arroz (*Oryza sativa*), *Setaria italica*, *Panicum virgatum* e *Brachypodium distachyon*, além do próprio SUCEST. A Figura 35 mostra o número de genes das gramíneas que possuem *hit* contra os *reads* e *contigs* do FL-RNAseq. A diferença de porcentagem entre *reads* (Figura 35A) e *contigs* (Figura 35B) é principalmente devida ao tamanho curto dos *reads*, que impede um alinhamento maior com os genes em questão, principalmente pelo fato de que apenas os alinhamentos acima de 100 nt foram considerados. Mais de 90% dos genes de sorgo (cerca de 27 mil dos 29 mil genes preditos), espécie filogeneticamente mais próxima de cana-de-açúcar, apresenta *hit* contra os *contigs* do FL-RNAseq (Figura 35B), número semelhante do alinhamento contra o SUCEST. Os 10% do SUCEST que não possui *hit* contra o FL-RNAseq podem ser transcritos específicos de alguns tecidos, como raiz e flores, que foram usados na construção do SUCEST mas não nas bibliotecas de cDNA. A análise recíproca, gramíneas mapeadas contra o FL-RNAseq (Figura 36), mostrou que apenas 60% do FL-RNAseq está representado em todo o SUCEST, o que significa aproximadamente 78 mil novas sequências de cana-de-açúcar, ausentes no SUCEST (Figura 37). A comparação dos dados do FL-RNAseq com as sequências presentes no chip Agilent (Figura 37) mostrou que apenas 435 sondas do chip não são representadas pelo FL-RNAseq. Além disso, 25,7% dos *contigs* (50.346) não possuem homologia de sequência com nenhum gene das 7 gramíneas analisadas. Estes foram então alinhados contra os bancos de dados Uniprot chegando a um total de 49.110 *contigs* (25,0%) sem homologia em nenhum dos banco de dados testados, sendo possíveis transcritos específicos do gênero *Saccharum*.

Tabela 22. Dados do sequenciamento do Ion PGM e 454 e montagem dos *contigs* com o software Trinity.

Descrição	Ion PGM	454
Corridas	5	3
<i>Reads</i>	29.260.184	2.403.750
Tamanho total dos <i>reads</i>	5,69 Gb	0,97 Gb
Média	194 bases	404 bases
Mediana	197 bases	436 bases
<i>Read</i> mais longo	645 bases	861 bases
<i>Read</i> mais curto	8 bases	50 bases
Montagem com Trinity		
Número de <i>contigs</i>	195.765	
Tamanho total dos <i>contigs</i>	133.946.191 bases	
<i>Contig</i> mais longo	8.854 bases	
<i>Contig</i> mais curto	201 bases	
Número de <i>contigs</i> > 1Kb	39.266 (20,1%)	
Média	684 bases	
Mediana	462 bases

Figura 35. Porcentagem de genes de cada uma das 7 espécie de gramíneas que possui *hit* contra os *reads* (A) e *contigs* (B) do FL-RNAseq (*cutoff* e-20). Na figura A, apenas os alinhamentos >100nt foram considerados.

Figura 36. Porcentagem de contigs do FL-RNAseq que possui *hit* contra cada uma das 7 espécies de gramíneas (*cutoff* e-20).

Figura 37. Diagrama de Venn comparando o número de seqüências do SUCEST, chip Agilent e *contigs* do FL-RNAseq. Note que o número total de *contigs* do FL-RNAseq mostrado na figura é menor que o número real devido ao fato que alguns SAS batem com mais de um *contig*.

A porcentagem de cobertura dos contigs nos genes das 7 espécies de gramíneas (Figura 38) foi avaliada para confirmar se o protocolo de construção das bibliotecas de fato enriqueceu para transcritos *full length*. Uma vez que a região UTR possui tamanho mais variável, foi adotado como critério para considerar um transcrito como possivelmente *full length* se este apresentar acima de 80% de cobertura do gene homólogo, pois essa porcentagem é suficiente para cobrir toda a ORF e ainda parte das UTRs, similar ao exemplo da Figura 34 onde a cobertura é de 79,8%. Essa análise foi realizada de forma a considerar todos os *hits* entre os genes e os *contigs*, somando-se as coberturas não sobrepostas dos *hits* em um mesmo gene para chegar à cobertura total para este gene. Muitos genes são cobertos por mais de um *contig* (Figura 39), mostrando que possivelmente o cDNA está completo mas não foi sequenciado ou montado completamente. Utilizando-se os genes de milho (*Z. mays*) como exemplo, aproximadamente 30 mil genes (11 mil, 16 mil e 3 mil genes com cobertura de 80%, 90% e 99%, respectivamente) enquadram-se nesse critério (Figura 38).

Figura 38. Porcentagem de cobertura dos contigs nos genes de cada uma das 7 espécies de gramíneas (*cutoff* e-20).

Figura 39. Número de *contigs* que alinham em cada gene das 7 espécies de gramíneas (*cutoff* e-20), considerando todos os *hits*.

Para avaliar a presença e a ausência dos transcritos em cada amostra, os *reads* de cada biblioteca foram mapeados com auxílio do software *bowtie* contra os 195.765 *contigs* permitindo visualizar a especificidade de expressão entre as amostras. Primeiramente, foi possível identificar que todas as 5 bibliotecas possuem *reads* na maioria dos *contigs* (152.016 ou 77,6%), mostrando que são transcritos comuns entre todas as amostras (Tabela 23). Por outro lado, alguns transcritos (4.020 ou 2%) são expressos especificamente em determinados genótipos (Tabela 23, linhas sombreadas). Um número significativo de transcritos (13.273 ou 6,8%) é expresso apenas em *S. spontaneum* e SP803280, mas não em *S. officinarum*, enquanto que uma quantidade bem menor de transcritos é expressa exclusivamente em *S. officinarum* e SP803280 (7.380 ou 3,8%), mas não em *S. spontaneum*. Esse resultado deve ser analisado com cuidado, uma vez que essa diferença pode ser devida

à diferença no número de *reads* de cada amostra, pois *S. spontaneum* tem 500 mil *reads* a mais que *S. officinarum* (6.172.884 e 5.647.332 *reads* de *S. spontaneum* e *S. officinarum*, respectivamente), o que pode justificar essa diferença de *contigs*. Poucos *contigs* apresentaram expressão somente nos genótipos ancestrais (602 ou 0,3%) e 5 vezes mais *contigs* apresentam expressão apenas em SP803280 (3.418 ou 1,7%), provavelmente por ter mais tecidos analisados que os ancestrais, sendo 120 e 1162 com expressão exclusiva em folha e entrenós, respectivamente, e os demais com expressão em folha e entrenó ao mesmo tempo. Dos 602 transcritos somente dos ancestrais, 86 são expressos em *S. officinarum*, 189 em *S. spontaneum* e 327 em ambos (Tabela 23, linhas sombreadas em verde, vermelho e azul, respectivamente).

Tabela 23. Número de contigs agrupado por amostra. Os *reads* de cada biblioteca foram mapeados nos contigs para identificar quais contigs são expressos em cada uma das amostras. Linhas sombreadas indicam contigs específicos de cada amostra/grupo, da seguinte maneira: Cinza, SP803280 (um ou mais tecidos); Azul, genótipos ancestrais; Vermelho, *S. spontaneum*; Verde, *S. officinarum*. E1, entrenó imaturo; E5, entrenó intermediário.

Nº de amostras	Folha			E1	E5	Nº Contigs
	<i>S. officinarum</i>	<i>S. spontaneum</i>	SP803280	SP803280	SP803280	
5	x	x	x	x	x	152016
4	x	x	x		x	8260
4		x	x	x	x	7206
4	x	x		x	x	5964
4	x		x	x	x	4073
3		x	x		x	2125
3		x		x	x	2007
3	x	x			x	1918
4	x	x	x	x		1661
3			x	x	x	1418
3	x		x		x	1157
3	x			x	x	1124
2		x			x	983
3	x	x	x			810
2				x	x	688
2			x		x	585
2	x				x	494
3		x	x	x		459
3	x	x		x		445
1					x	383
2	x	x				327
2		x	x			289
3	x		x	x		249
2		x		x		204
1		x				189
2	x		x			174
2			x	x		133
1			x			120
2	x			x		109
1				x		91
1	x					86

Todos os *contigs* foram alinhados contra o banco de dados do *Gene Ontology* (Ashburner *et al.* 2000) para identificar as principais categorias e funções dos transcritos sequenciados. Dos 195 mil *contigs*, foram identificadas 5.876 funções diferentes, sendo a grande maioria (5.465) presente em todas as cinco amostras. A Figura 40 mostra as funções que aparecem em mais de 1 mil *contigs*. Nota-se um grande número de *contigs* da categoria *chloroplast*, concordando com o fato de que 3 das 5 amostras são de folha. Também é possível identificar várias categorias relacionadas a respostas a estresses (*response to wounding*, *response to cold*, *defense response to bacteria*, *defense response to fungus*, *response to water deprivation*, *response to salt stress*) e categorias ligadas à transdução de sinal (*kinase activity*, *regulation of transcription*, *positive regulation of transcription*, *DNA binding*, *sequence specific DNA-binding*).

Todos os *contigs* foram posteriormente submetidos à predição de proteínas com o programa ESTscan, utilizando a matriz de *codon-usage* de cana-de-açúcar. Foram preditas 187.935 proteínas e apenas 7.830 *contigs* não possuem proteína predita (Tabela 24). A maior proteína apresenta 2.952 aminoácidos; o tamanho médio das proteínas e mediana foram de 188 e 116 aminoácidos, respectivamente (Tabela 24). As proteínas preditas foram analisadas pela presença de domínios conservados e 57.526 apresentaram algum domínio do banco de dados do Pfam (Punta *et al.* 2011), sendo que os mais frequentes estão listados na Tabela 24. De acordo com a descrição do próprio Pfam, os domínios PPR, WD40 e LRR, são de função desconhecida ou presentes em uma grande gama de proteínas não relacionadas. Por outro lado, em concordância com a categorização do Gene Ontology (Figura 40), é possível observar domínios relacionados às categorias de transdução de sinal, principalmente quinases (Pkinase e Pkinase_Tyr), e de resposta a estresses, como o NB-ARC, domínio o qual está relacionado a respostas a doenças (van Ooijen *et al.* 2008). Por último,

observamos o domínio citocromo p450 (p450) pertencente a uma superfamília de proteínas e responsável por catalisar a oxigenação de uma grande variedade de compostos, estando presente em enzimas como C4H e F5H da via de biossíntese de lignina, bem como em diversas outras enzimas de inúmeras vias, como, por exemplo, em enzimas do metabolismo de lipídeos.

Figura 40. Principais funções e categorias presentes nos *contigs* das cinco amostras. As barras indicam o número de contigs com a função especificada em todas as cinco amostras, considerando apenas funções representadas por mais de 1 mil *contigs*.

Tabela 24. Predição de proteínas nos contigs do FL-RNAseq utilizando o programa ESTscan.

Descrição	Número
Contigs	195.765
Proteínas preditas	187.935
Contigs sem proteínas preditas	7.830
Tamanho médio	188 aminoácidos
Mediana	116 aminoácidos
Maior proteína predita	2.952 aminoácidos
Menor proteína predita	15 aminoácidos
Proteínas com <i>hit</i> no HMMPFAM	57.526
Domínios mais frequentes	
Pkinase	2.381
WD40	1.670
PPR	1.595
LRR_1	1.440
PPR_2	1.433
Pkinase_Tyr	1.193
NB-ARC	992
LRR_8	835
RRM_1	689
p450	646

O cDNA enviado ao IQ-USP para sequenciamento no 454 havia sido convertido em dupla fita antes da fragmentação, o que implica na perda da orientação do transcrito original. Assim, para a análise de expressão, apenas os *reads* do *Ion* PGM foram considerados. Como a cana-de-açúcar ainda não apresenta um genoma referência disponível nem um banco de genes anotados, os *contigs* foram utilizados como referência para a análise de expressão. Os *reads* foram então analisados utilizando RPKM (*Reads Per Kilobase of transcript per Million mapped reads*), que leva em consideração o tamanho da referência e o número de *reads* mapeados, para gerar o nível de expressão dos transcritos, possibilitando identificar genes diferencialmente expressos entre as amostras. Ainda, por se tratar de amostras muito heterogêneas (pool de diversas folhas diferentes de plantas diferentes), foi estipulado um critério bem estrigente para definir se o gene estava

diferencialmente expresso, sendo considerado apenas aqueles com diferença de expressão maior que 4 vezes e $p=0,001$ entre cada par de amostra, resultando em 17.990 transcritos diferencialmente expressos. A *clusterização* hierárquica dos dados (Figura 41) mostra um padrão bem distinto de expressão dos entrenós em relação às folhas, separando-os em *clusters* diferentes. Ainda, é possível notar que os genes mais expressos nos entrenós estão reprimidos nas folhas e vice versa. Mesmo entre as amostras de folha, fica evidente uma distinção entre as os genótipos ancestrais e a SP803280, separadas em *sub-clusters* diferentes, onde é possível detectar diferenças importantes, como por exemplo um grupo de genes na parte superior do *heatmap* que estão mais expressos apenas em SP803280 e um grupo na parte inferior do *heatmap* menos expressos apenas em *S. officinarum* (Figura 41).

Os transcritos diferencialmente expressos foram categorizados no Gene Ontology e posteriormente submetidos à análise de enriquecimento de categorias funcionais (Tabela 25), da mesma forma conduzida para os dados de oligoarranjos (Tabelas 5 e 7). As funções estão separadas pelas três categorias principais do Gene Ontology, *Biological Process*, *Cellular Component* e *Molecular Function*, e as 30 funções mais enriquecidas de cada categoria estão listadas (Tabela 25). Coerente com o fato de 3 das 5 bibliotecas serem de folha, pode-se observar várias funções relacionadas à fotossíntese, como *photosynthesis* e *photosynthetic electron transport* na categoria *Biological process*; *chloroplast*, *oxygen evolving complex*, *chloroplast stroma* e *chloroplast thylakoid membrane* na categoria *Cellular Component*; *oxygen binding* e *chlorophyll binding* na categoria *Molecular Function*. Assim como na Figura 40, funções de regulação transcricional e resposta a estresses também estão enriquecidas entre os genes diferencialmente expressos, como *regulation of transcript*, *sequence-specific DNA binding TF activity*, *response to salt stress* e *response to wounding*. Funções relacionados à biossíntese de parede celular também estão enriquecidas,

principalmente relacionadas à lignina, mostrando um concordância com todos os resultados anteriores mostrados de qPCR e oligoarranjos dos genótipos ancestrais e variedade comercial (Tabelas 13-15 e Figuras 24 e 27). Dentre essas funções, destacam-se *secondary cell wall biogenesis*, *lignin biosynthesis*, *cell wall*, *phenylalanine ammonia-lyase activity* e *cinnamyl-alcohol dehydrogenase activity*. Ainda, algumas funções podem conter genes recorrentes nas listas de genes diferencialmente expressos entre variedades contrastantes para Brix (Papini-Terzi *et al.* 2009; Sato 2012), como aquaporinas (*water channel activity*) e genes do metabolismo de hormônios, incluindo biossíntese e resposta (*ethylene biosynthesis*, *response to abscisic acid stimulus* e *response to ethylene stimulus*).

Figura 41. Genes diferencialmente expressos pela análise de FL-RNAseq. Os genes foram divididos em 39 clusters hierárquicos indicados por cada uma das cores na coluna ao lado do *heatmap*. In1, entrenó imaturo; In5, entrenó intermediário; L, folha.

Tabela 25. Enriquecimento de Categorias funcionais do Gene Ontology dos transcritos com expressão diferencial nos dados de FL-RNAseq, separadas pelas três principais categorias: Biological Process, Cellular Component e Molecular Function. Foram selecionadas apenas as 30 primeiras funções com menor *e-score*, ou seja, mais representadas entre os transcritos diferencialmente expressos.

Biological process		Cellular component		Molecular Function	
<i>e-score</i>	Descrição	<i>e-score</i>	Descrição	<i>e-score</i>	Descrição
2.74E-118	Biological process	0	chloroplast	8.08E-169	Molecular function
4.24E-77	regulation of transcription	2.85E-260	nucleus	1.85E-106	sequence-specific DNA binding TF activity
7.36E-70	isopentenyl diphosphate biosynthesis	1.30E-228	plasma membrane	3.36E-61	protein binding
3.56E-64	starch biosynthetic process	1.55E-144	cytosol	1.26E-50	kinase activity
1.11E-40	response to cold	1.06E-89	extracellular region	1.78E-27	ATP binding
4.97E-39	photosynthesis	3.15E-72	cytoplasm	8.56E-22	oxygen binding
3.26E-27	response to fructose stimulus	9.31E-37	mitochondrion	4.70E-21	chlorophyll binding
2.75E-25	photosynthetic electron transport	1.72E-20	plastid	4.02E-18	calcium ion binding
1.28E-20	plant-type cell wall organization	1.64E-17	chloroplast thylakoid membrane	8.69E-18	hydrolyzing O-glycosyl compounds
1.32E-18	response to salt stress	1.34E-11	chloroplast stroma	5.52E-17	copper ion binding
4.61E-17	response to ethylene stimulus	2.14E-10	membrane	1.24E-15	DNA binding
4.66E-16	toxin catabolic process	3.49E-09	plastoglobule	3.76E-15	hydrolase activity
2.04E-14	response to sucrose stimulus	5.59E-09	chloroplast inner membrane	1.22E-14	protein serine/threonine phosphatase activity
2.66E-14	response to wounding	1.08E-08	plasmodesma	1.51E-14	oxidoreductase activity
1.09E-12	response to abscisic acid stimulus	2.30E-08	cellular_component	1.66E-14	RNA binding
1.24E-12	divalent metal ion transport	6.54E-07	Golgi apparatus	6.65E-13	structural constituent of ribosome
2.69E-12	coumarin biosynthesis	1.04E-05	oxygen evolving complex	1.34E-12	transferring glycosyl groups
4.55E-12	cell tip growth	1.63E-05	integral to plasma membrane	3.60E-11	transporter activity
1.72E-11	lignin biosynthesis	4.67E-05	vacuolar membrane	1.14E-10	nucleoside transmembrane transporter activity
5.89E-11	response to far red light	9.35E-05	chloroplast thylakoid	1.48E-10	sucrose:hydrogen symporter activity

Tabela 25. Continuação.

Biological process		Cellular component		Molecular Function	
<i>e-score</i>	Descrição	<i>e-score</i>	Descrição	<i>e-score</i>	Descrição
7.37E-11	anthocyanin biosynthesis	0.000108	endosome	1.67E-10	calmodulin binding
4.11E-10	regulation of proton transport	0.00025	vacuole	6.10E-10	zinc ion binding
8.17E-10	microtubule nucleation	0.000413	plant-type vacuole	9.47E-10	cinnamyl-alcohol dehydrogenase activity
1.80E-09	ethylene biosynthesis	0.001257	integral to membrane	1.71E-09	poly(U) RNA binding
1.88E-09	response to water deprivation	0.003056	cell wall	1.33E-08	linoleate 13S-lipoxygenase activity
1.89E-09	photosynthesis, light reaction	0.00885	chloroplatic endopeptidase Clp complex	3.31E-08	structural molecule activity
5.74E-09	response to chitin	0.025526	transcription export complex	2.34E-07	glutathione transferase activity
6.35E-09	secondary cell wall biogenesis	0.026106	endoplasmic reticulum	2.46E-07	serine-type carboxypeptidase activity
1.24E-08	rRNA processing	0.046009	SCF ubiquitin ligase complex	9.30E-07	water channel activity
7.70E-08	proteolysis	0.095959	plant-type vacuole membrane	1.41E-06	phenylalanine ammonia-lyase activity

Com a possibilidade de manter a orientação da fita no sequenciamento do PGM, foi possível identificar a expressão de transcritos antissenso. Vale lembrar que o protocolo de construção das bibliotecas leva à seleção apenas de RNAs com cauda poliA e cap 5', o que indica que apenas NATs (*natural antisense transcripts*) com poliA foram detectados. Assim, expandindo as análises de microarranjo nas quais foi detectado um número razoável de NATs expressos (Tabela 11), os *reads* do FL-RNAseq foram mapeados com auxílio do software bowtie nas 6 espécies de gramíneas e no SUCEST detectando um número bem maior de NATs em relação ao microarranjo (Figura 42). Entretanto, esse número pode ser ainda maior visto que cerca de 10 milhões de *reads* não mapearam em nenhuma das

espécies analisadas (dados não mostrados), provavelmente em parte devida ao fato de que apenas alinhamentos acima de 100 pb foram computados e também devido à alta estringência do programa bowtie. De forma geral, as amostras de folha parecem ter mais NATs que as amostras de entrenós e parece não ter uma grande diferença em números de NATs entre os ancestrais e o híbrido (Figura 42). Dentre os *reads* alinhados em cada espécie, a grande maioria são transcritos senso (Tabela 26) e em algumas espécies apenas uma pequena porção, menos que 1%, são NATs (Tabela 26). Sorgo, espécie mais próxima de cana-de-açúcar, foi a que apresentou a menor porcentagem de *reads* mapeados como antissenso (0,09%), mas por outro lado foi a espécie na qual mais *reads* alinharam, mais até que no próprio SUCEST (40 e 27% do total de *reads* alinhados em sorgo e no SUCEST, respectivamente, Tabela 26), evidenciando a importância de sorgo para análises em cana-de-açúcar.

Figura 42. Número de *reads* antissenso (NAT) em cada amostra por espécie. Todos os *reads* do PGM foram mapeados contra o banco de genes das espécies de gramíneas. Nos casos onde o *read* alinha com o gene na orientação reversa à do gene (*plus/minus* ou *minus/plus*), o *read* foi considerado um NAT. In1, entrenó imaturo; In5, entrenó intermediário; L, folha. Note que a escala do eixo Y está dividida em três partes.

Tabela 26. Porcentagem de *reads* senso e NAT alinhados contra cada espécie e suas respectivas proporções em relação ao número total de *reads* sequenciados no PGM (29.260.184 *reads*).

	Porcentagem dos <i>reads</i> alinhados		Porcentagem de todos os <i>reads</i> do PGM (29.260.184)	
	Senso (%)	NAT (%)	Senso (%)	NAT (%)
<i>B. distachyon</i>	92,34	7,66	4,25	0,35
<i>O. sativa</i>	99,88	0,12	24,51	0,03
<i>P. virgatum</i>	99,73	0,27	11,53	0,03
<i>S. italica</i>	93,48	6,52	12,22	0,85
<i>S. bicolor</i>	99,91	0,09	40,69	0,04
SUCEST	82,82	17,18	27,95	5,80
<i>Z. mays</i>	99,53	0,47	17,95	0,09

5. DISCUSSÃO

5.1. Partição de carbono, acúmulo de sacarose e lignina são diferencialmente regulados entre os genótipos

No presente trabalho, foram avaliadas as diferenças na composição da parede celular (lignina), acúmulo de sacarose, taxa fotossintética, e outros parâmetros fisiológicos em genótipos de cana-de-açúcar. Esses dados permitiram realizar uma caracterização fisiologia dos genótipos em estudo, o que irá direcionar as análises moleculares na identificação de genes ou vias que possam explicar ou estar relacionadas às diferenças fenotípicas encontradas. Tratando-se de cana-de-açúcar, o acúmulo de sacarose ainda é a característica mais importante. Com o aumento do interesse do uso da biomassa como matéria-prima para etanol celulósico, a síntese de parede celular também tem atraído um grande interesse.

A partição de carbono entre acúmulo de sacarose e síntese de parede celular em cana-de-açúcar parece ser muito complexa. Os dados fisiológicos permitem chegar a algumas conclusões que ajudam a avançar nos conhecimentos dessa área. Os genótipos RB867515 e *S. officinarum* possuem valores similares de taxa fotossintética (Figura 9A) e açúcares solúveis totais em entrenó maduro (Figura 10B) e Brix (Figura 8A), mas apresentam valores bem diferentes de biomassa e perfilhamento (Figura 8C e 8D). Ao analisar apenas o acúmulo de sacarose (Figura 10C e D), a RB867515 apresenta níveis semelhantes de sacarose na folha (fonte, 33,1 mg/g) e no entrenó intermediário (dreno, 34,5 mg/g), enquanto que *S. officinarum* possui mais sacarose no entrenó intermediário (46,8 mg/g) do que em folha (31,5 mg/g). Isso pode sugerir que o entrenó intermediário de RB867515 ainda não é um tecido de reserva e em *S. officinarum* a estocagem de sacarose já se iniciou. Apenas em

entrenó maduro é possível observar um aumento considerável no teor de sacarose em RB867515 (75,3 mg/g) em relação ao tecido fonte (folha). Vale ressaltar que as plantas analisadas estão com 7 meses de idade e ainda não atingiram à maturidade e, portanto, continuarão a crescer e aumentar os níveis de sacarose. É possível que o entrenó intermediário de RB867515 comece a acumular sacarose no fim do ciclo produtivo. Ainda, *S. officinarum* possui o mesmo número de entrenós que RB867515 (dados não mostrados) com menos da metade da altura da planta (Figura 8B), resultando em entrenós mais curtos. Esses dados em conjunto sugerem que *S. officinarum* tende a iniciar o acúmulo de sacarose em entrenós mais jovens, enquanto que RB867515 ainda está nas etapas de expansão celular e iniciará a estocagem de sacarose em entrenós mais expandidos, resultando em plantas maiores com maior biomassa. Concordando com essa hipótese, dois genes relacionados à expansão celular, uma expansina e uma XTH, estão mais expressos em entrenó intermediário de RB867515 do que em *S. officinarum* (Tabela 14), sugerindo que em RB867515 esse tecido está realmente em expansão.

Por outro lado, ao analisar-se os dados de *S. spontaneum*, pode-se observar que, como esperado, sacarose e parede celular parecem competir pelo carbono assimilado. Essa conclusão vem da análise dos resultados de sacarose e lignina e tem suporte nos dados de expressão gênica, especialmente considerando a via de biossíntese de lignina em *S. spontaneum*. Enquanto *S. officinarum* parece começar o acúmulo de sacarose em tecidos mais jovens (como descrito acima), *S. spontaneum* parece iniciar a lignificação em tecidos mais jovens, uma vez que esse genótipo possui maior lignificação em entrenó intermediário, fato detectado por ambas análises, quantitativa (Figura 11) e qualitativa (Figura 12 e 13). As figuras 10B e 11A mostram uma correlação inversa entre acúmulo de açúcares e a lignificação em *S. spontaneum*, visto que em entrenó intermediário desse genótipo o total

de açúcares (103 mg/g) é maior do que em entrenó maduro (68 mg/g), enquanto o teor de lignina em entrenó intermediário (16,8%) é menor do que em entrenó maduro (19,2%). Comparado às demais plantas, esses dados sugerem uma lignificação precoce em *S. spontaneum*, iniciando-se em entrenós mais jovens. Em *S. officinarum*, a consequência do fluxo de carbono direcionado para sacarose reflete no menor nível de lignina em entrenó maduro (Figura 11A), enquanto *S. spontaneum* possui menor nível de açúcares (Figura 10B e C) devido ao fluxo de carbono direcionado para lignina. Como uma consequência indireta do baixo grau de lignificação de *S. officinarum*, o desenvolvimento de fibras em entrenó intermediário e maduro pode ser afetado (compare Figura 12B e F com Figura 12D e H) reduzindo, assim, a rigidez do colmo e, por conseguinte, dificultando o crescimento da planta. Assim, estratégias de melhoramento, incluindo transgenia, com o objetivo de aumentar a sacarose ou diminuir a recalcitrância da biomassa pela redução da lignificação, devem tomar cuidado uma vez que já foram reportados experimentos com plantas transgênicas nos quais a simples diminuição da lignificação pode, de fato, aumentar a disponibilidade de carboidratos para produção de etanol celulósico, mas também pode levar a fenótipos indesejáveis, como a redução do crescimento da planta (Besseau *et al.* 2007; Coleman *et al.* 2008b; Song & Wang 2011).

O fato de *S. spontaneum* apresentar uma regulação positiva de expansinas e XTH em entrenó imaturo (Tabela 13) concorda com a hipótese na qual a lignificação inicia-se em tecidos mais jovens nesse genótipo. Em outras palavras, os processos de expansão celular e crescimento tendem a ser iniciados e concluídos em porções mais jovens do colmo, seguindo-se à deposição de parede celular secundária ocorrendo mais cedo do que nos outros genótipos. Isso pode ser evidenciado pelo fato dos entrenós intermediários de *S. spontaneum* apresentam características visuais semelhantes às encontradas no entrenó

maduro, incluindo comprimento (dados não mostrados). A deposição de siringil, a qual está relacionada a estágios tardios da lignificação (Donaldson 2001; Rencoret *et al.* 2011), é maior em porções mais jovens do colmo de *S. spontaneum* (Figura 11B) do que nos outros genótipos, sugerindo, novamente, uma mudança temporal na síntese de lignina em comparação com as demais plantas analisadas. Esses resultados indicam que uma das razões para o baixo conteúdo de sacarose (Figura 8A e 10 C) e a menor espessura do colmo de *S. spontaneum* (Tabela 3) é a lignificação precoce, que impede a expansão celular necessária para acomodar grandes vacúolos e permitir o espessamento do colmo, direcionando a partição de carbono para a lignificação. De fato, em experimentos de casa de vegetação foi demonstrado que as células do parênquima de entrenó intermediário de *S. spontaneum* são menores do que as de RB867515 (Poelking 2012). Aliás, genes envolvidos na produção de energia (piruvato quinase e NADH desidrogenase) são mais expressos em entrenó intermediário de *S. spontaneum* (Tabela 14), sugerindo uma demanda maior de energia nesse tecido que seja devida ao aumento da lignificação e de formação de fibras, visto que a produção de lignina é um processo que exige muita energia, no qual cada grama de lignina sintetizada requer 2,6 a 3 gramas de glicose (Amthor 2003).

A lignificação prematura pode ser explicada, pelo menos em parte, pela ausência de repressão da expressão dos genes da via de biossíntese de lignina pelo *ScMYB22*, homólogo putativo do repressor da biossíntese de lignina de milho *ZmMYB31* (Fornale *et al.* 2010), ou até mesmo por uma inesperada função de ativação dos genes dessa via pelo *ScMYB22*. Em suporte a essa hipótese, pode-se observar que em entrenó intermediário de *S. spontaneum* ocorre uma regulação positiva de *ScMYB22* (Tabela 14) com concomitante regulação positiva de *COMT* e *F5H* (Tabela 14), lembrando que já foi demonstrado que *ZmMYB31* reprime ambos os genes em milho (Fornale *et al.* 2010). Aliás, *S. spontaneum* apresentou maior

conteúdo de lignina em ambos os entrenós, intermediário e maduro (Figura 11A), sugerindo que *ScMYB22* não atue como repressor da biossíntese de lignina neste genótipo.

Por ser a fonte principal de assimilação de carbono na planta, é intuitivo pensar que a fotossíntese correlacione diretamente com acúmulo de sacarose e biomassa. Entretanto, ao menos em cana-de-açúcar, esse assunto é muito mais complexo. Há várias décadas já se sabe que taxa fotossintética não está diretamente correlacionada com acúmulo de sacarose e produtividade (Irvine 1975). Outros fatores também influenciam no acúmulo de sacarose e biomassa, como as relações entre fonte e dreno, carregamento e descarregamento no floema, partição de carbono, mecanismos de transporte através de membranas e atividades de enzimas específicas (Souza *et al.* 2013). A inibição da fotossíntese desencadeada pelo acúmulo de carboidratos solúveis nas folhas já foi relatado em diversas espécies [(van Heerden *et al.* 2010; Wang *et al.* 2010b) e referências citadas por eles], incluindo cana-de-açúcar (McCormick *et al.* 2008b), levantando a hipótese de que a menor taxa fotossintética de *S. spontaneum* possa ter relação com essa via de inibição. Porém, a quantificação de açúcares (Figura 10A-C) não demonstrou nenhuma diferença significativa entre os genótipos que pudesse explicar as diferenças de taxa fotossintética, mas pode sugerir que o limiar de sinalização por açúcares para ativar a inibição da fotossíntese seja diferente entre os genótipos. É possível especular que a menor taxa fotossintética de *S. spontaneum* esteja relacionada ao fato dessa planta não acumular sacarose. Entretanto, em experimentos com duração de um ano com clones de cana-de-açúcar com alto teor de sacarose ou alto teor de fibras (Inman-Bamber *et al.* 2011), foi observado que a taxa fotossintética reduz-se com a maturidade da planta, mas o acúmulo de sacarose no colmo não é a base para essa redução. No mesmo estudo, os autores concluíram que a manutenção da fotossíntese não está associada ao baixo conteúdo de açúcares nas folhas, resultado que, de certa forma, é similar

ao encontrado nos genótipos ancestrais, onde o menor acúmulo de açúcares na folha não se correlaciona com uma menor fotossíntese (Figuras 9A e 10A-C). Inman-Bamber e colaboradores também concluíram que a inibição da fotossíntese pelo acúmulo de açúcares em folha deve acontecer em porções localizadas nas folhas jovens, mas essa inibição seria compensada pela fotossíntese nas folhas mais velhas, mantendo constante a fotossíntese da planta inteira durante todo o dia (Inman-Bamber *et al.* 2011). Assim, para obter-se um melhor entendimento da partição de carbono e sua relação com fotossíntese nos genótipos ancestrais, seriam necessários mais experimentos considerando um número maior de pontos experimentais, incluindo outras folhas além da folha+1.

5.2. Análise do transcriptoma

5.2.1. Comparação da expressão gênica entre genótipos por oligoarranjos e qPCR

O estudo dos genes e das vias que regulam o acúmulo de sacarose e biomassa em cana-de-açúcar sempre atraiu grande interesse. Experimentos anteriores conduziram a prospecção de genes diferencialmente expressos em genótipos contrastantes para acúmulo de sacarose e biomassa (Papini-Terzi *et al.* 2009; Sato 2012), o que permitiu a obtenção de resultados importantes com a identificação de genes e possíveis mecanismos regulatórios envolvidos nessas características. No entanto, devido à base genética estreita das variedades comerciais e à diminuição dos ganhos de produtividade a cada nova variedade comercial lançada, torna-se importante explorar a grande diversidade genética do gênero *Saccharum* para identificar alvos de interesse que possam ser introduzidos nos programas de melhoramento, bem como obter um melhor entendimento das vias regulatórias para

direcionar o melhoramento genético. Nesse contexto, o perfil de expressão gênica de três espécies ancestrais e uma variedade comercial importante, RB867515, foi analisado utilizando-se microarranjo de oligonucleotídeos e qPCR. Nos oligoarranjos, a RB867515 foi usada como referência em todos os experimentos, desta forma, possibilitando comparações entre os genótipos. Os genótipos ancestrais podem ser fontes importantes de variabilidade genética com potencial para contribuir no aumento dos ganhos de produtividade e desenvolvimento da cana energia.

De forma geral, o padrão de expressão de *S. officinarum* é mais similar ao de RB867515. É muito provável que isso seja devido à composição genética mais similar entre as duas, consequência da nobilitação. Isso é semelhante ao observado em sorgo, onde foi reportado que o padrão de expressão de duas variedades é determinado pela composição genética (Jiang *et al.* 2013).

5.2.1.1. Metabolismo energético e de sacarose

De forma geral, os genes relacionados à geração de energia, tanto nos oligoarranjos (Seção 4.8.1) como nos experimentos de qPCR (seção 4.9.4), estão mais expressos em *S. spontaneum* e menos expressos em *S. officinarum*. Genes como a fosfofrutoquinase (PFK, SCSBFL1105F08.g) e uma subunidade da ATP-sintase (SCJLFL3014G01.g) estão induzidos em mais de um tecido em *S. spontaneum* (Apêndice 1). Ao contrário, a piruvato quinase (PK, SCCCLR1070D04.g) e a fosfoglicerato quinase (PGK, SCCCLR1079B06.g) estão reprimidas em entrenó imaturo e intermediário de *S. officinarum*. A indução de PK em *S. spontaneum* também é vista em entrenó intermediário nos experimentos de qPCR (Tabela 14). Vale ressaltar que estas quinases participam de etapas que envolvem gasto ou produção de ATP na glicólise, sendo os principais pontos de controle da via (Lehninger *et al.* 2005). O

complexo I da cadeia de transporte de elétrons (NADH desidrogenase, SCMCLR1122H07.g) também está induzido em *S. spontaneum* nos três entrenós testados por qPCR (Tabelas 13-15). Ainda, podemos observar a indução da expressão do antissenso desse mesmo SAS em entrenó de *S. officinarum* em relação à RB867515 nos oligoarranjos (Apêndice 1), o que pode indicar um silenciamento pós-transcricional desse gene no genótipo ancestral, sugerindo novamente um menor gasto energético de *S. officinarum*. Isso pode ser explicado pela mesma hipótese de acúmulo precoce de sacarose em entrenós mais jovens, onde o fluxo de carbono é direcionado para estocagem em *S. officinarum*, enquanto que em RB867515 está direcionado para crescimento. Por outro lado, o maior gasto energético de *S. spontaneum* pode ser devido à produção de fibras lignificadas, uma vez que esse genótipo possui uma maior porcentagem de fibras lignificadas (Poelking 2012), e a lignina requer um maior *input* de energia para sua produção em comparação a carboidratos (Amthor 2003).

Da mesma forma observada nos experimentos de qPCR (Seção 4.9.4), genes do metabolismo de sacarose quase não apresentam um padrão entre plantas de alto e baixo Brix, nem qualquer outro padrão de expressão aparente. Em todos os experimentos de microarranjo, apenas um gene relacionado ao metabolismo de sacarose foi identificado, em *S. officinarum*, onde uma sacarose sintase (SCEQLB1065B01.g) está induzida em entrenó intermediário.

Essa fraca correlação entre a expressão de genes do metabolismo de sacarose e seu acúmulo em cana-de-açúcar já foi relatada anteriormente, onde colmos em desenvolvimento com acúmulo de sacarose ativo não apresentam enriquecimento na expressão de genes de metabolismo de sacarose (Casu *et al.* 2003). Provavelmente, a falta de correlação entre o padrão de expressão de genes e o acúmulo de sacarose pode ser explicada pelo fato do metabolismo de sacarose em geral ser fortemente influenciado por

modificações pós-traducionais, mediadas por proteínas regulatórias, como as quinases. Dentre elas, pode-se destacar as proteínas SnRKs, as quais controlam o metabolismo de carbono em plantas pela modulação da atividade enzimática por fosforilação direta e pela regulação da expressão gênica (Halford & Hey 2009). Enzimas como a sacarose sintase e sacarose-fosfato sintase são controladas direta ou indiretamente por SnRKs (Purcell *et al.* 1998; Sugden *et al.* 1999). Em cana-de-açúcar, já foi demonstrado uma associação entre a expressão de SnRKs e o acúmulo de sacarose, onde plantas com alto Brix apresentavam maior expressão dessas quinases (Papini-Terzi *et al.* 2009).

O único gene que pode estar relacionado ao metabolismo de sacarose que apresenta uma correlação direta com o conteúdo de Brix das plantas analisadas é uma *SnRK interacting prontein* (SCACRZ3111E02.g) reprimida em entrenó imaturo e intermediário de *S. spontaneum* e *S. robustum*. Acredita-se que essa proteína atue inibindo a ação das SnRKs (Thelander *et al.* 2007). Esse resultado eleva a complexidade da atuação das SnRKs no metabolismo de sacarose em cana-de-açúcar, uma vez que a observação de Papini-Terzi (Papini-Terzi *et al.* 2009) indica que a maior expressão de SnRKs correlaciona-se diretamente com o maior acúmulo de sacarose e, assim, seria esperado que uma proteína que inibe a ação de SnRKs estivesse induzida em variedades de baixo Brix, e não reprimida, como observado nos genótipos ancestrais.

5.2.1.2. Transportadores

As aquaporinas, como descrito anteriormente, são proteínas que formam canais de membrana que facilitam o movimento de água e solutos entre células ou compartimentos celulares (Heinen *et al.* 2009). Em cana-de-açúcar, aquaporinas já foram identificadas com expressão diferencial em experimentos de variedades contrastantes para Brix (Papini-Terzi

et al. 2009; Sato 2012) e os autores sugeriram uma possível associação entre a expressão dessa proteína e o acúmulo de sacarose. Em experimentos de seca (Costa 2012; Lembke *et al.* 2012), as aquaporinas também aparecem diferencialmente expressas. Recentemente, foi reportado que bananeiras transgênicas superexpressando aquaporinas apresentam um melhor desempenho contra diversos tipos de estresses abióticos (Sreedharan *et al.* 2013), demonstrando que essas proteínas são, de fato, interessantes para o melhoramento genético, principalmente, no caso da produção de cana energia, onde o cultivo em regiões menos adequadas é uma característica relevante.

Além disso, aquaporinas estão envolvidas no processo de expansão celular (Chen *et al.* 2013), podendo ser interessante no aumento da biomassa vegetal e, conseqüentemente, melhorando a produtividade, agregando mais uma característica significativa. Nos 3 genótipos ancestrais, pelo menos uma aquaporina apresentou expressão diferencial em todos os tecidos analisados por oligoarranjos (Apêndice 1), sugerindo que, também em cana-de-açúcar, essas proteínas podem estar associadas à expansão do colmo, além do acúmulo de sacarose. Nos experimentos de qPCR, é possível observar uma correlação mais clara, onde as plantas de baixo Brix, principalmente *S. spontaneum*, apresentam maior expressão de aquaporinas nos entrenós (Tabelas 13-15) e folha (Tabela 17).

Já foi demonstrado que as aquaporinas localizam-se nas células guarda de diversas espécies, incluindo gramíneas (Heinen *et al.* 2009) e, uma vez que a abertura e fechamento dos estômatos são controlados pela pressão osmótica nas células guarda (Heinen *et al.* 2009), é provável que as aquaporinas influenciem diretamente o turgor celular e, conseqüentemente, a abertura dos estômatos. Assim, é possível que a expressão de aquaporinas possa ter alguma relação direta ou indireta nas diferenças apresentadas por *S. spontaneum* para os parâmetros de trocas gasosas em folha (Figura 9), como transpiração e

condutância estomática (Gs). De fato, em experimentos com videiras submetidas a estresse hídrico moderado mostraram uma correlação entre a expressão de aquaporinas, Gs e condutância hidráulica da folha (Pou *et al.* 2012). Em outro estudo, plantas de arroz submetidas a estresse hídrico severo mostraram um aumento considerável de aquaporinas no proteoma da planta (Mirzaei *et al.* 2012). Esses resultados sugerem um papel importante das aquaporinas nas respostas à seca e relações hídricas na folha. Nos resultados do presente trabalho, nota-se uma correlação negativa entre a expressão de aquaporinas e Gs em *S. spontaneum* e *S. robustum* e também com transpiração em *S. spontaneum* (Tabela 17 e Figura 9B e C). Como é provável que as aquaporinas possam influenciar de algum modo o funcionamento dos estômatos, consequentemente também irão influenciar os valores de Gs, transpiração e WUE. Além disso, Neufeld e colaboradores (Neufeld *et al.* 1992) relataram que *S. spontaneum* possui uma resposta melhor à seca do que *S. officinarum* e uma variedade comercial, visto que, em baixos valores de potencial hídrico na folha, *S. spontaneum* foi capaz de manter 20% da condutividade hidráulica foliar máxima enquanto as demais plantas a perdiam completamente, conferindo à *S. spontaneum* uma resistência maior à perda de condutividade do xilema, protegendo-a contra a cavitação e o embolismo. As aquaporinas poderiam não somente evitar o embolismo como também poderiam reparar esse problema durante o estresse por seca (Heinen *et al.* 2009). Entretanto, a maior expressão de aquaporinas em *S. spontaneum* pode ser uma consequência indireta da maior lignificação desse genótipo. A lignificação pode acarretar inibição do transporte de água pela rota apoplástica, devido ao caráter hidrofóbico da mesma, o que pode ser compensado pelo aumento do fluxo por outras rotas (North & Peterson 2005). Nesse contexto, uma vez que *S. spontaneum* parece ter lignificação em um número muito maior de células em comparação com às outras plantas, incluindo células do parênquima e do feixe vascular (compare Figura

12D com 12A-C e 12H com 12E-F), o transporte de água pela rota apoplástica ficaria prejudicado, aumentando a importância da rota simplástica. A consequência seria o aumento da expressão de aquaporinas para auxiliar no transporte de água entre as células.

Os resultados do presente trabalho corroboram observações anteriores (Papini-Terzi *et al.* 2009; Sato 2012) que indicam uma possível correlação das aquaporinas com acúmulo de sacarose, mostrando que essas proteínas apresentam padrões de expressão diferencial desde as espécies ancestrais até os híbridos comerciais modernos, e avançam no conhecimento por mostrar uma relação da expressão das aquaporinas com a transpiração, WUE e Gs. Isso faz das aquaporinas alvos interessantes para análises futuras e manipulações genéticas.

Como descrito na introdução, o carbono fixado nas folhas é transportado pelo floema na forma de sacarose até os tecidos dreno, como os entrenós, onde a sacarose será estocada. Para isso, é fundamental a participação de transportadores de açúcares os quais farão o carregamento da sacarose da folha para o floema e deste para o parênquima de reserva (Ainsworth & Bush 2011). Esses transportadores regulam o fluxo de sacarose pelo floema e a regulação da atividade e da expressão desses genes controla as relações entre fonte e dreno (Ainsworth & Bush 2011). Já foi reportado que a repressão da expressão desses genes pode causar acúmulo de carboidratos nas folhas e inibição da fotossíntese (Zhang & Turgeon 2009). A manipulação da expressão de transportadores de sacarose já demonstrou ser capaz de aumentar o influxo de sacarose em tecidos dreno, levando ao aumento do crescimento dos cotilédones em ervilha (Rosche *et al.* 2002) e dos níveis de proteína em grãos de trigo (Weichert *et al.* 2010), provando ter aplicações biotecnológicas interessantes. Um total de 4 transportadores de açúcares estão diferencialmente expressos (SCCCT2002G08.g, SCEQRT3C03F06.g, SCCCRZ2C01E03.g e SCSFRT2067F07.g) e

correlacionam-se exatamente com o acúmulo de sacarose nos genótipos ancestrais. Nos genótipos de baixo Brix esses transportadores estão reprimidos em folha, e também em entrenó intermediário de *S. spontaneum*, e induzidos, com exceção de SCCCRT2002G08.g, em folha e entrenó intermediário de *S. officinarum*.

5.2.1.3. Parede celular

Em um dos poucos trabalhos focados na análise de genes de cana-de-açúcar relacionados à parede celular, Casu e colaboradores (Casu *et al.* 2007) conduziram uma análise transcricional em entrenós e relataram cinco grupos de genes com um perfil de expressão similar com cada grupo sendo diferencialmente regulado em diferentes partes do colmo. Os autores sugeriram que, em cana-de-açúcar, não haveria especificidade das proteínas Cesa na síntese de parede primária e secundária, ou que os dois tipos de parede poderiam ser sintetizados ao mesmo tempo em um mesmo entrenó, porém em diferentes tecidos. Entretanto, nossos resultados mostram claramente que há uma diferença de expressão de duas Cesa entre entrenó jovem e entrenó em lignificação (ou seja, com síntese de parede celular secundária), indicando uma expressão dez vezes maior de ScCesA7 em entrenó intermediário em comparação ao entrenó imaturo e vinte vezes maior em comparação à expressão de ScCesA3 em entrenó intermediário (Figura 30). Esses resultados sugerem que ScCesA3 e ScCesA7 são específicas de parede primária e secundária, respectivamente.

Curiosamente, apenas em folha de *S. officinarum* foi possível detectar genes da família Cesa com expressão diferencial nos oligoarranjos, sendo uma Cesa (SCCCLR2C02H01.g) e uma CsIF (SCEQLB1066G04.g) reprimidas, e uma Cesa induzida (SCJLLR1106H06.g). Também foi possível detectar beta-1,3 e beta-1,4 glucanases

diferencialmente expressas (famílias GH17 e GH9, respectivamente), principalmente em entrenó intermediário dos três genótipos, com 4 SAS induzidos e apenas 1 reprimido. Apesar da função de degradar a celulose (quebra de ligações beta-1,4), algumas evidências indicam que enzimas da família GH9 possam ter um papel importante na síntese de celulose e no desenvolvimento. Em aspen híbrido, um membro de cada uma das famílias GH9 e GH17 estão induzidos em tecidos com deposição de parede celular secundária (Aspeborg *et al.* 2005), o que também ocorre para um GH9 de *Arabidopsis* (Shani *et al.* 2006), sugerindo um papel dessas enzimas em processos de crescimento e deposição de parede celular secundária. Mutantes defectivos de arroz para uma GH9 apresentam problemas de crescimento, afetando o alongamento do entrenó e a composição da parede celular (Zhou *et al.* 2006). Além disso, análises de coexpressão de genes e de plantas mutantes de arroz indicaram um possível papel de membros da família GH9 tanto na síntese de celulose quanto na sua cristalinidade (Xie *et al.* 2013). Da mesma forma, as enzimas beta-1,3 glucanase (GH17) também estão envolvidas em diversos processos fisiológicos como desenvolvimento, crescimento e reprodução, além de resposta a patógenos (Balasubramanian *et al.* 2012). A superexpressão de uma beta-1,3 glucanase em linho causou aumento do conteúdo de celulose, hemicelulose e pectinas e, concomitante, diminuição de lignina (Wojtasik *et al.* 2013), mostrando que essa enzima pode estar envolvida nos processos de deposição de parede celular secundária, sendo interessante para manipulações genéticas em cana-de-açúcar, visando à produção de etanol celulósico.

Enzimas como a XTH e a (1-4)- β -manan endohidrolase atuam hidrolisando polissacarídeos de parede celular primária, promovendo o afrouxamento da parede e, conseqüentemente, permitindo a expansão da célula. Assim, essas enzimas são importantes para diversos processos fisiológicos como amadurecimento de frutos, germinação, entre

outros (Eklof & Brumer 2010). A pectinametilesterase, apesar de não ser uma glicosilhidrolase, também é importante para a expansão celular por ser capaz de alterar as propriedades biofísicas da parede primária pela remoção de grupos metil das pectinas e aumentar a susceptibilidade das pectinas às poligalacturonases. Além disso, são importantes em outros processos como resistência a patógenos, captação de água, modulação indireta da ação das expansinas e podendo também alterar o pH da parede celular (Lionetti *et al.* 2012; Muller *et al.* 2013). Foi possível observar genes de todas essas famílias diferencialmente expressos em pelo menos um experimento. Em conformidade com a hipótese de que *S. officinarum* está em menor expansão celular em entrenó intermediário, observou-se uma XTH (SCJLST1020H07.g) e uma mannan endohidrolase (SCQSRT2035G06.g) reprimidas mas, por outro lado, uma pectinametilesterase (SCEZAM2035C11.g) e uma expansina (SCCCLR1048B09.g) estão induzidas. Essa mesma XTH também está reprimida em entrenó imaturo de *S. officinarum* em relação à de *S. robustum* e *S. spontaneum* (Tabela 13). Entretanto, em entrenó maduro, essa enzima está induzida em *S. officinarum* (Tabela 15). Uma expansina (SCCCCL4006H09.g) apresenta um perfil semelhante ao da XTH, com indução em entrenó imaturo de *S. spontaneum* (Tabela 13). Já em entrenó maduro, a indução pode ser observada em RB867515 ao invés de *S. officinarum* (Tabela 15). Vale lembrar que Sato (Sato 2012) e Papini-Terzi (Papini-Terzi *et al.* 2009) também identificaram expansinas e XTHs diferencialmente expressas em experimentos com variedades contrastantes para Brix, mostrando que esses são genes recorrentes nesse tipo de abordagem, ajudando a reforçar a hipótese de que essas proteínas estejam envolvidas no acúmulo de sacarose. Genes do metabolismo de parede celular podem estar relacionados ao acúmulo de sacarose, além da própria competição pelo carbono assimilado, pelo rearranjo da parede celular necessário para a expansão da célula, que é importante para acomodar grandes vacúolos de estocagem

de sacarose e para conter a pressão osmótica crescente advinda do aumento na concentração de sacarose. Assim, genes como XTHs e expansinas podem ser ótimos candidatos para análises posteriores.

Como descrito na seção 5.2.1.1, poucos genes relativos ao metabolismo de sacarose foram identificados como diferencialmente expressos pelo oligoarranjos ou por qPCR entre os grupos de alto e baixo Brix. Sendo assim, visto que além das diferenças de composição de parede celular, principalmente lignina (Figuras 11 e 12), e a biossíntese de parede celular ser um grande dreno de carbono e, portanto, competir com o acúmulo de sacarose, é possível que as diferenças de expressão de genes do metabolismo de parede celular possam explicar melhor as diferenças entre os genótipos do que os próprios genes de metabolismo de sacarose. De fato, foi possível identificar 32 genes de metabolismo de parede celular diferencialmente expressos. Da biossíntese de lignina, uma CAD (SCACHR1038E08.g), que catalisa da última etapa da via (Figura 2), está mais expressa nos genótipos de baixo Brix em entrenó imaturo, enquanto que uma PAL e uma C3H, genes mais iniciais da via (Figura 2), estão reprimidos em entrenó imaturo e folha, respectivamente, em *S. officinarum*. Também é possível identificar vários genes PAL e 4CL mais expressos em *S. spontaneum* nas análises de *clusterização* hierárquica, tanto em entrenó imaturo (Figura 25) quanto intermediário (Figura 27). Tais fatos concordam com os experimentos de qPCR (Tabelas 13-15) e com a quantidade de células lignificadas nesses genótipos (Figura 12), além de levantar a ideia de que as diferenças de níveis de transcrição de genes de parede celular entre os genótipos, principalmente da via de biossíntese de lignina, ocorrem em entrenós mais jovens, antes de ficar evidente as diferenças na parede em si, uma vez que as diferenças de expressão são identificadas também em entrenó imaturo, onde não há grandes diferenças de lignina (Figura 11). Ainda, esses genes são alvos interessantes em situações onde se procura

diminuir a quantidade de lignina, ou mesmo alterar sua composição. Em tabaco transgênico silenciando a expressão da PAL (Sewalt *et al.* 1997) foi possível diminuir significativamente a quantidade de lignina. Da mesma forma, a supressão de C3H em álamo híbrido foi capaz de diminuir a quantidade de lignina, além de alterar as quantidades dos monolignóis guaiacil e hidroxifenil (Coleman *et al.* 2008a). Em experimentos visando à produção de etanol celulósico, milho transgênico com repressão de CAD apresentou apenas uma pequena diminuição na razão S/G (siringil/guaiacil) sem alteração no conteúdo de lignina total, juntamente com aumento de celulose e arabinoxilanas, melhorando a sacarificação e levando a um aumento da produção de etanol celulósico (Fornale *et al.* 2012).

5.2.1.4. Biossíntese de Hormônios

O crescimento e desenvolvimento da planta, bem como respostas a estresses, são regulados por diversos tipos de hormônios. Os hormônios atuam disparando uma cascata de sinalização, mediada por fatores de transcrição, proteínas quinases e fosfatases culminando em uma resposta fisiológica (Santner & Estelle 2009). O uso de hormônios sintéticos para estimular o crescimento, germinação e controle de pragas é comum em diferentes culturas (Raven *et al.* 1996). A alteração da expressão gênica de plântulas de cana-de-açúcar em resposta a tratamentos com diferentes hormônios já foi avaliada anteriormente (Rocha *et al.* 2007; Brandão 2009) e foram identificados diversos genes diferencialmente expressos de várias categorias, incluindo fatores de transcrição, metabolismo de carboidratos, biossíntese de hormônios, quinases, metabolismo de parede celular, fosfatases, metabolismo de DNA, entre outros, mostrando uma grande gama de funções alteradas por hormônios.

Genes da biossíntese de ácido abscísico (SCBFRZ3008D06.g), jasmonato (SCJFRZ2032B09.g e SCCCCL4013D09.g) e etileno (SCEZLR1009E06.g, SCCCCL2001D09.b,

SCJLLR1054A06.g e SCCCLR1022A02.g) apresentaram-se diferencialmente expressos nos genótipos ancestrais. O SAS SCBFRZ3008D06.g codifica uma 9-cis-epoxicarotenoide dioxigenase, enzima chave na biossíntese de ácido abscísico (ABA) (Iuchi *et al.* 2001), um hormônio normalmente relacionado à resposta a estresse hídrico (Riera *et al.* 2005), apresentou repressão somente em *S. officinarum*, o que se correlaciona com a maior suscetibilidade dessa planta a estresses. Da mesma forma, os genes de biossíntese de jasmonato, hormônio ligado ao desenvolvimento da planta e também a resposta a estresses (Kombrink 2012), estão reprimidos em *S. officinarum* apenas. Na biossíntese de etileno, todos os 4 SAS identificados codificam a enzima ACC oxidase, enzima chave na biossíntese de etileno, catalisando a última etapa da via (Grierson 2013); três deles estão reprimidos em *S. officinarum* nos três tecidos.

É importante ressaltar que esses 7 SAS relacionados à biossíntese de hormônios estão diferencialmente expressos em 11 ocasiões, sendo preferencialmente reprimidos (9 de 11) e, dentre estes, *S. officinarum* apresentou o maior número de genes reprimidos (7 de 9) em todos os tecidos, incluindo ambos os SAS relacionados à síntese de jasmonato, e apenas 1 gene induzido (SCEZLR1009E06.g, em folha). O outro gene induzido (ACC oxidase, SCEZLR1009E06.g) ocorre em entrenó imaturo de *S. spontaneum*. Os resultados mostram que *S. officinarum* tem menor expressão de genes de biossíntese de hormônios, o que pode acarretar em deficiência de hormônios, correlacionando com sua baixa resistência a estresses e produtividade. É possível que a aplicação de hormônios exógenos possam melhorar a performance desse genótipo.

5.2.1.5. Transdução de sinal

A categoria *Signal Transduction* foi uma das que apresentou maior número de genes diferencialmente expressos (Figura 14). Dentre eles, podemos destacar os fatores de transcrição (TFs), proteínas regulatórias induzidas por hormônios, quinases e fosfatases, os quais atuam em diferentes níveis na via de sinalização, desde a percepção do estímulo até o controle da atividade das enzimas efetuatoras da resposta.

5.2.1.5.1. Regulação transcricional

Os TFs atuam se ligando ao DNA e interagindo com a maquinaria de transcrição de forma a recrutar ou bloquear o acesso da RNA polimerase ao promotor do gene (Udvardi *et al.* 2007). TFs das famílias bZIP, AP2/ERF, DREB, entre outras, aparecem na lista de genes diferencialmente expressos. Em plantas, TFs da família bZIP estão envolvidos em diversos processos fisiológicos desde a resposta a patógenos (Alves *et al.* 2013) ao controle do metabolismo de amido (Wang *et al.* 2013a). Um membro da família bZIP (SCBFRZ2019D02.g) apresenta expressão diferencial entre planta de alto e baixo Brix, sendo induzido em folha e entrenó intermediário de *S. spontaneum* e reprimido em folha de *S. officinarum*.

Também foi possível identificar TFs e outras proteínas envolvidas na transdução de sinal em resposta aos hormônios ABA, etileno, auxina, giberelina e jasmonato, mostrando o papel importante da regulação hormonal para o desenvolvimento da planta e nas diferenças entre os genótipos. Os efeitos dos diferentes hormônios podem interagir entre si em diferentes níveis de complexidade, atuando em vias que podem se sobrepor de forma sinérgica ou antagônica, dependendo da via ou do tecido (Anderson *et al.* 2004; Muday *et al.* 2012; Vanstraelen & Benkova 2012; Richter *et al.* 2013). Todos esses hormônios atuam na sinalização de vários processos de crescimento e desenvolvimento, bem como em respostas

a estresses, formando uma intrincada e complexa rede regulatória (Shibasaki *et al.* 2009; Ben-Ari 2012; Erb *et al.* 2012; Vanstraelen & Benkova 2012; Daviere & Achard 2013).

Fatores de transcrição da família AP2/ERF estão envolvidos nas respostas ao hormônio etileno (Xu *et al.* 2011) e também ao jasmonato (McGrath *et al.* 2005). Seis SAS diferencialmente expressos da família AP2/ERF foram identificados (Apêndice 1), sendo apenas 1 induzido em folha de *S. officinarum*, 4 reprimidos em todos os tecidos do três genótipos (com exceção de entrenó imaturo de *S. robustum*) e 1 induzido tanto o senso quanto o antissenso em entrenó intermediário de *S. spontaneum*. A repressão preferencial desses genes nos genótipos ancestrais podem ter alguma relação com a melhor performance de RB867515 ou mesmo em função de ser um híbrido interespecífico, o que pode levar a mudanças epigenéticas na expressão (Chen & Tian 2007; Hu *et al.* 2013) ativando genes que normalmente estariam reprimidos. De fato, plantas transgênicas com superexpressão de genes dessa família já demonstraram ter uma melhora significativa em diferentes características, incluindo resistência a estresses bióticos e abióticos, parâmetros fotossintéticos e aumento da biomassa (revisado por (Xu *et al.* 2011)). Além disso, recentemente foi identificado um membro dessa família relacionado à indução de florescência em arroz (Ren *et al.* 2013) e o entendimento e a manipulação da indução da florescência é particularmente importante em cana-de-açúcar, uma vez que o florescimento está normalmente acompanhado da diminuição da produtividade (van Heerden *et al.* 2010). Esses dados mostram que genes AP2/ERF são candidatos interessantes para engenharia genética de cana-de-açúcar, além de evidenciar um possível papel da resposta aos hormônios para as diferenças fenotípicas entre os genótipos.

O modelo atual para atuação das giberelinas aponta que proteínas inibitórias DELLA interagem com fatores de transcrição relacionados a genes de crescimento, impedindo-os de

se ligarem aos promotores dos genes alvos. O receptor de giberelina, GID1, ao se ligar ao hormônio sofre uma mudança de conformação passando a interagir com as proteínas DELLA, formando um complexo giberelina-GID1-DELLA. Esse complexo estimula a degradação das proteínas DELLA via proteossomo, removendo a inibição do crescimento (Daviere & Achard 2013). Em entrenó intermediário de *S. officinarum*, foi possível observar 3 TFs e/ou proteínas responsivas a giberelina (SCJFLR1035B01.g, SCRFLR1012G02.g eSCCCCL3004D07.b) regulados positivamente. A indução desses genes pode sugerir que as vias de sinalização de crescimento e resposta a estresses nessa planta possuem alguma etapa comprometida anterior à percepção do estímulo, ou seja, após a degradação das proteínas DELLA, uma vez que esse genótipo apresenta o menor crescimento. A identificação dessa etapa ajudaria a detectar genes de interesse para manipulação genética e também de marcadores moleculares para auxiliar o melhoramento clássico. Contudo, é necessária uma análise integrada de redes regulatórias, já que a estratégia comum de identificação de genes de vias de resposta a hormônios é feita pela análise de mutantes (Kombrink 2012; Daviere & Achard 2013), o que não é possível para cana-de-açúcar pela ausência de uma coleção de mutantes.

Genes de resposta a jasmonato (SCEZLB1009F12.g, SCRLLR1038D05.g e SCBGLR1002E05.g) estão diferencialmente expressos em *S. officinarum* e *S. robustum*, mas não em *S. spontaneum*. O SAS SCRLLR1038D05.g está induzido nos 3 tecidos de *S. robustum* e também em entrenó imaturo e intermediário de *S. officinarum*. O fitormônio jasmonato atua em diversos processos, desde fertilidade da planta até resposta a ataque de insetos (revisado por (Santino *et al.* 2013)). A transdução do sinal disparada por jasmonato ocorre primariamente via TFs com domínio bHLH, sendo o MYC2 o membro melhor caracterizado. Na ausência do hormônio, repressores JASMONATE-ZIM-DOMAIN (JAZ) interagem com CORONATINE INSENSITIVE 1 (COI1), o qual é responsável pela percepção da molécula

biologicamente ativa de jasmonato. JAZ bloqueia a atividade de MYC2, sendo liberado pela degradação de JAZ induzida por jasmonato (Santino *et al.* 2013). Em cana-de-açúcar, a mudança de expressão gênica em resposta a jasmonato já foi estudada (Rocha *et al.* 2007) e os autores evidenciaram a indução da expressão de três histonas em resposta a esse hormônio. Isso pode sugerir que jasmonato também atue de forma a regular epigeneticamente a expressão de outros genes, ampliando o espectro de ação do hormônio.

A auxina é um dos principais hormônios relacionados ao crescimento e desenvolvimento dos tecidos da planta (Perrot-Rechenmann 2010) e o direcionamento do crescimento se dá pela formação de um gradiente de concentração de auxina (Benkova *et al.* 2003). Em mais de um tecido dos 3 genótipos, 8 genes responsivos a auxina estão diferencialmente expressos. Destes, o SAS SCBGLR1114G12.g está exclusivamente induzido nos 3 tecidos de *S. spontaneum* e em 2 de *S. robustum*, evidenciando um perfil de expressão antagônico entre plantas de alto e baixo Brix. Esse gene pertence à família AUX/IAA, que atuam reprimindo genes associados à resposta a auxina (Wang *et al.* 2010e). Ainda, genes dessa família estão constantemente associados a respostas a seca (Wang *et al.* 2010e; Costa 2012; Lembke *et al.* 2012; Horta 2013), um estresse que altera o acúmulo de sacarose em cana-de-açúcar (Costa 2012). O fato desse repressor AUX/IAA estar induzido somente em genótipos de baixo Brix sugere sua associação com a repressão de genes que regulam o acúmulo de sacarose.

Recentemente, foi sugerido que os fitormônios possam estar envolvidos na regulação do acúmulo de sacarose e produtividade em cana-de-açúcar (Sato 2012) e os resultados aqui apresentados reforçam e ampliam essa hipótese, demonstrando que essa regulação pode estar conservada desde os genótipos ancestrais. Assim, fazem-se necessárias análises

posteriores mais aprofundadas do papel desses hormônios no metabolismo de cana-de-açúcar.

Em relação aos possíveis TFs envolvidos nas diferenças de parede celular secundária, nos oligoarranjos nenhum gene da família MYB está diferencialmente expresso e apenas um TF da família NAC aparece reprimido em folha de *S. robustum* (SCCCLR2003C07.g). Contudo, genes da família NAC estão envolvidos em outras funções fisiológicas, como senescência e resposta a estresses (Pinheiro *et al.* 2009), sendo improvável a função desse SAS em biossíntese de parede celular devido à falta de correlação entre expressão e diferenças de parede celular entre os genótipos neste tecido. Com isso, foi conduzida uma análise por qPCR para avaliar a expressão dos homólogos dos TFs *NST1*, *MYB58/63*, *KNAT7* de *Arabidopsis* e do *ZmMYB31* de milho (Tabelas 13-15), os quais já foram caracterizados como reguladores do metabolismo de parede celular em suas espécies de origem. De forma geral, pode-se observar que nenhum dos TFs testados apresentou a correlação entre sua função esperada e, principalmente, o padrão de deposição de lignina (ver Seção 4.9.2). Isso pode sugerir que, além de possuírem diferentes tipos de parede celular, gramíneas e dicotiledôneas também podem possuir diferenças no nível de regulação transcricional. Esse fato foi recentemente observado em plantas de transgênicas de arroz superexpressando o fator de *Arabidopsis*, *AtSHINE2* (Ambavaram *et al.* 2011). Essas plantas transgênicas apresentaram uma regulação negativa dos genes de síntese de lignina e uma regulação positiva dos genes de síntese de celulose com redução no teor de lignina e aumento simultâneo no conteúdo de celulose. Contudo, nenhuma alteração significativa foi observada em *Arabidopsis* superexpressando o *AtSHINE1* (Kannangara *et al.* 2007; Ambavaram *et al.* 2011). Na gramínea *Brachypodium distachyon*, fatores de transcrição da família NAC foram caracterizados (Valdivia *et al.* 2013) e também observou-se diferenças

funcionais em relação aos homólogos de *Arabidopsis*, principalmente na ligação aos promotores dos genes alvo. Os autores sugeriram que muito do que foi estudado em dicotiledôneas sobre parede celular pode ser válido para gramíneas, mas que existem algumas diferenças importantes. Todos esses dados somados sugerem que gramíneas e dicotiledôneas possam ter evoluído diferentemente em algumas de suas estratégias de regulação do metabolismo parede celular, podendo apresentar algumas etapas diferentes. Assim, para a manipulação e melhoramento de gramíneas visando a produção de biomassa para etanol celulósico é necessário pesquisar mais sobre a parede celular deste grupo de plantas. Ao mesmo tempo, porém, também é necessário um modelo mais adequado para estudos moleculares de gramíneas.

Mesmo não tendo sido observado fatores das famílias NAC e MYB diferencialmente expressos pelo método HTself, devido sua importância para biossíntese de parede celular secundária mesmo em gramíneas, evidenciada pela demonstração da função de TFs dessas famílias no metabolismo de parede celular em milho (Fornale *et al.* 2010), *Brachypodium* (Valdivia *et al.* 2013) e trigo (Ma *et al.* 2011), a análise de intensidade do sinal de expressão (Seção 4.8.3) foi utilizada para identificar TFs dessa família que possam ser candidatos a reguladores da biossíntese de parede celular. Nenhum homólogo dos fatores já caracterizados em outras espécies foi encontrado, mas foi possível identificar um fator MYB (SCQGST1032D08.g) mais expresso em entrenó imaturo (Figuras 23 e 24) e dois NACs (SCCCCL4014A04.g e SCCCCL3001D02.g) mais expressos em entrenó intermediário de *S. spontaneum* (Figura 27), correlacionando com a maior síntese e deposição de lignina nesse genótipo. Assim, esses genes são bons candidatos para análises posteriores de caracterização funcional.

Apesar de não ser um TF, o SAS SCCCLR1067B01.g está ligado à regulação da transcrição por possuir um domínio de ligação proteínas da família LIM, as quais possuem papel multifuncional, regulando a expressão de histonas e também atuando na montagem dos filamentos de actina (Moes *et al.* 2013). Por controlar a expressão de histonas, é possível que desempenhe um papel importante na regulação da estrutura da cromatina e, conseqüentemente, na regulação epigenética. Esse SAS está reprimido exclusivamente nos genótipos de baixo Brix, em todos os tecidos.

5.2.1.5.2. Proteínas quinases e fosfatases

Inúmeras vias metabólicas e de resposta a estímulos são reguladas por cascatas de fosforilação e defosforilação. Essas cascatas, além de promoverem a transdução de um sinal interno ou externo, também atuam de forma a amplificá-lo, aumentando a resposta ao estímulo. As proteínas quinases e fosfatases catalisam reações opostas e atuam em diversas vias, desde resposta a estresses bióticos e abióticos (Arnaud *et al.* 2012; Fuchs *et al.* 2013) à percepção de hormônios (Zhu *et al.* 2013). Nos experimentos com os genótipos ancestrais, 41 quinases e 5 fosfatases mostraram-se reprimidas ou induzidas em pelo menos um ponto experimental (Apêndice 1).

Dentre as 41 quinases identificadas, não foi possível identificar a família a qual pertence a grande maioria, pois nem sempre o SAS possui a sequência completa da proteína, o que dificulta a análise de alinhamento e uma anotação mais acurada. Entretanto, identificou-se 4 LRR-RLKs (*leucine-rich repeat receptor-like kinases*) que pertencem a um grande grupo de receptores quinase de plantas (Shiu & Bleecker 2001; ten Hove *et al.* 2011). Apesar da função de muitas dessas quinases ainda não estarem elucidadas, alguns autores sugerem que as LRR-RLKs de plantas podem ser divididas em duas categorias funcionais,

sendo uma relacionada a processos de crescimento e desenvolvimento e a outra envolvida em resposta a estresses bióticos e abióticos (Shiu & Bleecker 2001; Afzal *et al.* 2008; Wang *et al.* 2013b). Recentemente, demonstrou-se que uma LRR-RLK de *Arabidopsis* está envolvida na formação de parede celular secundária nas fibras do xilema (Wang *et al.* 2013b), despertando o interesse por essa família de quinases para o desenvolvimento da cana energia. Contudo, nos experimentos de microarranjo, *S. officinarum* apresentou duas LRR-RLKs induzidas em entrenó intermediário, enquanto que uma LRR-RLK estava reprimida em entrenó imaturo de ambos os genótipos de baixo Brix e também em entrenó intermediário de *S. spontaneum*, mesmo estes apresentando mais fibras lignificadas (ou seja, com parede celular secundária) nos entrenós (Figura 12B e D) em comparação a *S. officinarum*.

Quinases da família STRUBBELIG-RECEPTOR (SRF) possuem papel importante na morfogênese dos tecidos em diferentes órgãos e membros dessa família já foram demonstrados estarem envolvidos no desenvolvimento do óvulo, flor, folha, caule e raiz (Schneitz *et al.* 1997; Chevalier *et al.* 2005; Eyuboglu *et al.* 2007; Fulton *et al.* 2009; Vaddepalli *et al.* 2011). Um SAS representante dessa família (SCCCLR1068G01.g) foi diferencialmente expresso em *S. spontaneum* e, interessante, apenas o antissenso foi identificado, sendo induzido na folha desse genótipo. É possível que esse antissenso exerça algum tipo de controle sobre o transcrito senso desse SAS, resultando na repressão de sua tradução. Essa repressão pode estar associada à formação de fita dupla de RNA, disparando o mecanismo de RNA de interferência ou mesmo pela competição entre as RNA polimerases transcrevendo as fitas em direções opostas, impedindo fisicamente a transcrição de uma das fitas (Lapidot & Pilpel 2006). Em *Arabidopsis*, já foi demonstrado que SRF4 atua como um regulador positivo da área foliar (Eyuboglu *et al.* 2007). A expressão do antissenso dessa SFR correlaciona diretamente com o menor tamanho das folhas de *S. spontaneum* (Figura 7),

sugerindo um papel similar ao de *Arabidopsis*. Em experimentos de campo, foi demonstrado que *Miscanthus* é 59% mais produtivo que milho e essa diferença é explicada pela maior captação de luz pelas folhas resultante da maior área foliar, indução precoce da expansão e senescência tardia das folhas de *Miscanthus* em comparação a milho (Dohleman & Long 2009). Assim, um regulador positivo da área foliar pode influenciar diretamente na produtividade da cultura, sendo de grande interesse para a biotecnologia de cana-de-açúcar. A evidência sugere que o SCCCLR1068G01.g pode estar envolvido nesse processo, sendo um alvo interessante para estudos posteriores.

Apenas para uma fosfatase identificada (SCVPFL3046C06.b) foi possível determinar a família. Trata-se de uma PP2C, cujos membros estão relacionados à resposta a estresse, sendo principalmente reguladores negativos da via de sinalização por ABA, hormônio responsável por respostas a estresse hídrico (Schweighofer *et al.* 2004; Saez *et al.* 2006). Curiosamente, essa PP2C está induzida apenas nos genótipos de baixo Brix os quais são conhecidamente mais resistentes a estresses (Paterson *et al.* 2013). Isso pode sugerir que essas plantas, justamente por serem mais resistentes, tenham as vias de sinalização de resposta a estresse mais facilmente “ligadas” e, por isso, necessitem de uma repressão mais intensa em condições normais, aumentando a expressão de PP2C.

5.2.1.6. Metabolismo de DNA e RNA

É essencial que as plantas desenvolvam rápidas respostas às condições ambientais para sua adaptação e sobrevivência. Além da regulação da expressão gênica pela ação de TFs, o controle preciso da modificação da cromatina pode ter um papel fundamental nessa regulação, uma vez que a estrutura da cromatina pode permitir ou reprimir a expressão de determinados genes em resposta a estímulos externos. Vários estudos já demonstraram a

capacidade das plantas de adaptar o crescimento e desenvolvimento através de mudanças pós-traducionais nas histonas em resposta a estímulos ambientais como luz, temperatura, estresses bióticos e abióticos (Chen & Tian 2007; Luo *et al.* 2012). Em organismos aloploplóides, as modificações em histonas também podem atuar de forma a “selecionar” quais homeólogos serão expressos (Chen & Tian 2007; Hu *et al.* 2013). Alguns autores postulam a existência de um “código de histonas” no qual, de forma análoga ao código genético, cada modificação nas histonas ou conjunto de modificações, “codificaria” uma função, como a ativação ou repressão de genes (Jenuwein & Allis 2001). Entretanto, esse código não seria tão simples como inicialmente presumido, sendo na realidade uma linguagem muito mais sofisticada onde nem sempre existe uma distinção rigorosa entre modificações ativadores e repressoras (Berger 2007). Em todos os experimentos, 10 histonas estão diferencialmente expressas em 26 ocasiões, sendo 20 reprimidas e apenas 6 induzidas. Nos genótipos de baixo Brix, todas as histonas estão reprimidas e apenas em entrenó imaturo e intermediário de *S. officinarum* ocorre indução. Somado a isso, uma *Polycomb group protein* (PcG, SCQSLB1049H12.g) está reprimida apenas em entrenó imaturo e intermediário de *S. officinarum* e induzida em folha de *S. officinarum* e todos os tecidos de *S. robustum*. As PcGs fazem parte dos *polycomb repressive complexes* (PRC) e já foi demonstrado que PRC1 promove a monoubiquitinação da histona H2A (Bratzel *et al.* 2010) e PRC2 induz a trimetilação da histona H3 (Makarevich *et al.* 2006) levando a repressão de genes relacionados a diferenciação e desenvolvimento em *Arabidopsis*. Somando-se aos resultados anteriores, que apontam para uma mudança do metabolismo do entrenó de *S. officinarum* de desenvolvimento para acúmulo de sacarose mais precocemente que as demais, é razoável supor a existência de algum tipo de regulação epigenética, direta ou indireta, do acúmulo de sacarose via um PRC em cana-de-açúcar. De toda forma, apesar de

ter sido identificado apenas um gene de modificação de histonas diferencialmente expresso, o número significativo de histonas nos genótipos de baixo Brix especificamente reprimidas sugere uma correlação entre o metabolismo de histonas e regulação epigenética e as diferenças de Brix ou mesmo de resistência a estresses. Além disso, como o controle epigenético via histonas pode ser resultante do “estresse” genômico causado pela hibridização interespecífica (Chen & Tian 2007; Hu *et al.* 2013), é provável que a maior expressão de histonas na RB867515 tenha alguma relação com o fato de ser um híbrido das espécies *S. officinarum* e *S. spontaneum* e a seleção dos homeólogos específicos que tornam o híbrido mais produtivo, tanto em sacarose quanto em biomassa por colmo, em relação aos genótipos ancestrais. É importante ressaltar que um gene com domínio de ligação às proteínas LIM, as quais estão envolvidas no controle da expressão de histonas (Moes *et al.* 2013), está reprimido em todos os tecidos de *S. spontaneum* e *S. robustum*, mas não em *S. officinarum* (Seção 5.2.1.5.1), assim como as histonas. Esses fatos ajudam a reforçar a hipótese de regulação epigenética do acúmulo de sacarose.

Curiosamente, além das histonas, a subunidade Rpb6 da RNA polimerase (SCCCRZ2004C12.g) está reprimida em todos os tecidos de *S. spontaneum* e *S. robustum*, mas não aparece nos experimentos de *S. officinarum*. Entretanto, por se tratar de uma subunidade comum às RNA polimerases I, II e III, seria inadequado atribuir uma relação a qualquer fenótipo analisado.

5.2.1.7. Genes de captação de luz e fixação de carbono

Neste trabalho, o perfil de expressão das três primeiras enzimas relacionadas ao ciclo C4, PEPC, CA e PPDK, mostraram-se similares entre si nos genótipos (Tabela 17), sendo *S. officinarum* e *S. spontaneum* com os menores valores de expressão. Uma alta expressão de

PPDK, PEPC e uma subunidade da Rubisco pode ser observada em RB867515, a planta com maior biomassa por colmo (Tabela 17). Isso pode sugerir que a alta produtividade pode em parte estar relacionada a esse perfil de expressão, sendo mais importante que a própria taxa fotossintética, visto que já foi relatado que a taxa fotossintética e produtividade não são diretamente correlacionadas (Irvine 1975). Apesar disso, essa baixa correlação encontrada entre produtividade e fotossíntese pode ser devida ao fato de que apenas uma medida de fotossíntese foi realizada. Em *S. spontaneum*, apenas a expressão de uma subunidade da Rubisco e de uma subunidade do fotossistema II (Tabela 17) correlacionaram com a menor taxa fotossintética, indicando uma relação dessas duas enzimas com a taxa fotossintética, o que não ocorre com as enzimas do ciclo C4 PEPC, CA e PPDK. Esses resultados são em parte semelhante aos de De Souza *et al.*, onde plantas submetidas à alta concentração de CO₂ apresentaram um aumento da taxa fotossintética e produtividade com uma concomitante indução de genes dos fotossistemas levando os autores a sugerirem que o acréscimo da fotossíntese é devido ao aumento da taxa de transporte de elétrons nos fotossistemas (De Souza *et al.* 2008). Essa hipótese é parcialmente suportada pela observação de que a essa subunidade da Rubisco pode ser responsável pela constante catalítica da enzima (taxa de *turnover*, k_{cat}) (Ishikawa *et al.* 2011). A expressão da subunidade da Rubisco de sorgo, que possui alto k_{cat} , em plantas transgênicas de arroz levaram a um aumento significativo de k_{cat} (Ishikawa *et al.* 2011). Contudo, não foi observada uma mudança significativa na assimilação de CO₂. De toda forma, essa proteína pode ser um alvo interessante para o melhoramento de cana-de-açúcar e outras plantas.

5.2.2. Comparação da expressão gênica entre entrenó imaturo e intermediário

Além de um grande número de genes diferencialmente expressos entre os genótipos, também foi identificado um número similar de genes induzidos reprimidos no entrenó intermediário em relação ao entrenó imaturo da mesma planta pelos “experimentos híbridos” e esses genes podem estar relacionados ao desenvolvimento do colmo (Apêndice 2). O entendimento dos mecanismos desse processo pode aprimorar e direcionar novas abordagens para manipulação genética ou melhoramento clássico.

Uma análise mais detalhada desses experimentos permitiu identificar padrões de expressão entre as amostras. Os genótipos ancestrais apresentaram a grande maioria dos genes da categoria *cell cycle* regulados negativamente (8 de 8 em *S. spontaneum* e 17 de 21 em *S. officinarum*), ao contrário de RB867515, na qual a grande maioria estão regulados positivamente (14 de 16). O mesmo acontece na categoria *DNA metabolism*. Esse padrão de expressão apresentado pela RB867515 em relação a *S. officinarum* e *S. spontaneum* sugere que a variedade comercial tenha uma maior replicação celular, o que acarreta em um colmo mais alongado e mais grosso que os genótipos ancestrais, explicando o fato daquela produzir mais biomassa por colmo. De fato, isso pode ser consequência direta das observações anteriores, nas quais *S. spontaneum* iniciaria a lignificação mais precocemente, o que poderia prejudicar a replicação e expansão celular, e que *S. officinarum* direcionaria a partição de carbono para acúmulo de sacarose em tecidos mais jovens do que RB867515, o que também prejudicaria a replicação e expansão celular.

Na categoria *cell wall metabolism*, os genes da via da biossíntese de lignina, de forma geral, estão regulados positivamente em todos os genótipos. Também podemos observar a expressão diferencial de algumas glucanases das famílias GH9 e GH17 (beta-1,4 e beta-1,3

hidrolases, respectivamente). Como descrito na seção 5.2.1.3, essas enzimas podem estar associadas à modificações na parede relacionadas a processos normais de desenvolvimento, não sendo inesperado a expressão diferencial identificada em entrenós em desenvolvimento. Quatro *CesA* também estão diferencialmente expressas em pelo menos um dos experimentos: SCCCLR1066B10.g, SCCCRZ2002G07.g, SCCCLR2C02H01.g e SCJLLR1106H06.g (*ScCesA7*). Todas as quatro estão induzidas em *S. officinarum*, enquanto que em *S. spontaneum* apenas a SCCCLR2C02H01.g está induzida. Em RB867515 SCCCLR2C02H01.g e *ScCesA7* estão reprimidas. Destas quatro *CesA*, duas estão possivelmente relacionadas à parede primária (SCCCLR1066B10.g e SCCCRZ2002G07.g) e duas à parede secundária (SCCCLR2C02H01.g e *ScCesA7*) de acordo com a análise filogenética (Apêndice 3). Esses experimentos mostraram que o desenvolvimento do colmo de cana-de-açúcar leva à expressão diferencial de uma série de genes de metabolismo de parede celular, assim como previamente relatado (Casu *et al.* 2007). O fato de haver duas *CesA* de parede secundária reguladas negativamente em entrenó intermediário de RB867515 em relação ao entrenó imaturo (SCCCLR2C02H01.g e *ScCesA7*) está de acordo com a hipótese de que as células deste tecido ainda estão em expansão, enquanto os demais genótipos já cessaram esse processo neste mesmo tecido. Isso porque a síntese de parede celular secundária está associada ao término do crescimento celular, sugerindo que em entrenó intermediário a síntese de parede secundária é menor. Uma explicação para a maior expressão dessas *CesA* em entrenó imaturo de RB867515 seria o desenvolvimento dos vasos do xilema que requer deposição de parede celular secundária. Outra evidência para sustentar essa hipótese é a presença de dois receptores de giberelinas regulados positivamente em RB867515 (SCCCCL3004D07.b e SCJFLR1035B01.g), hormônio relacionado ao alongamento dos entrenós (Dayan *et al.* 2012). Da mesma forma, também é possível

observar duas peroxidases (SCEPRZ3085A12.g e SCQGLR2032G02.g) e uma laccase (SCUTST3084C11.g) reguladas positivamente em *S. officinarum*, enzimas responsáveis pela oxidação radicalar necessária à polimerização da lignina, sugerindo novamente um aumento de deposição de parede celular secundária em relação ao entrenó imaturo.

Uma das categorias com maior número de genes diferencialmente expressos foi a categoria relacionada a transdução de sinais, que engloba principalmente fatores de transcrição, quinases e fosfatases. Dentre os fatores de transcrição, enquanto *S. spontaneum* não apresentou nenhum fator de transcrição da família MYB diferencialmente expresso em entrenó intermediário, *S. officinarum* apresentou 3 fatores da família MYB regulados positivamente (SCQGST1032D08.g, SCACST3159E04.g e SCMCRT2105A02.g) e RB867515 apresentou 2 fatores MYB (SCQGST1032D08.g e SCMCRT2105A02.g) regulados negativamente. Esses dois fatores MYB, cuja família possui diversos membros relacionados ao controle transcricional de genes de parede celular secundária (Wang & Dixon 2012), reprimidos em RB867515 estão induzidos em *S. officinarum* e coexpressos com genes de biossíntese de lignina e CesaA, são passíveis a estarem associados ao controle transcricional da biossíntese de parede celular, sendo ótimos candidatos para análises posteriores. Por outro lado, nenhum fator da família NAC, também associada a parede celular em outras espécies (Wang & Dixon 2012), foi identificado nesses experimentos.

Ao comparar *S. officinarum* e RB867515, observa-se que as duas acumulam sacarose, mas apresentam uma grande diferença de biomassa (Figuras 8A e C). Pela análise dos dados de microarranjo entre os dois genótipos (Apêndice 1) e também dos experimentos entre os entrenós de cada genótipo (Apêndice 2), é possível identificar que *S. officinarum* possui genes de biossíntese de hormônios e também de transdução de sinal hormonal regulados positivamente em relação à RB867515, principalmente de hormônios ligados ao crescimento

da planta, porém com uma regulação negativa e genes de ciclo celular e de metabolismo de DNA, ambos envolvidos na replicação celular. Novamente, isso sugere que a diferença de crescimento de *S. officinarum* e RB867515 não esteja relacionada nem à biossíntese, nem à percepção do hormônio, mas provavelmente esteja ligada a alguma etapa entre a percepção do sinal e o alvo final, como discutido na seção 5.2.1.5.1. De todo modo, assim como nos experimentos de Brix (Sato 2012), a regulação hormonal parece ser um fator muito importante para a diferença de biomassa entre os genótipos analisados, porém, essa diferença está provavelmente ligada a uma das etapas à jusante da percepção do sinal.

Em *S. spontaneum*, a hipótese de que esse genótipo tende a alongar suas células em tecidos mais jovens, anteriores ao entrenó intermediário, é parcialmente suportada pela identificação de um receptor de giberelina (SCCCCL3004D07.b, o mesmo identificado em RB867515) regulado negativamente. Em suporte a essa hipótese, Poelking (Poelking 2012), através de análise micromorfométrica de entrenó intermediário de RB867515 e *S. spontaneum*, conseguiu identificar que a RB867515 tinha menor porcentagem de fibras lignificadas, com menor diâmetro e comprimento, sugerindo que esse tecido da variedade comercial ainda está em expansão e formação de fibras, enquanto que o mesmo tecido de *S. spontaneum* já está em processo mais tardio de deposição de parede celular secundária e, por isso, no fim das etapas de crescimento e expansão. Isso somado ao fato da lignificação das células do parênquima de *S. spontaneum* (Figura 12D e H) com menor diâmetro em relação à RB867515 (Poelking 2012) em entrenó intermediário, reforçam a hipótese de diferença de crescimento/expansão entre a variedade comercial e os ancestrais nesse tecido.

Apenas 1 SAS antissenso (SCACFL5029C09.g) diferencialmente expresso foi identificado em *S. officinarum*, sendo reprimido. A sequência senso desse SAS não possui

nenhuma proteína predita, mas uma análise de blast indica que é similar a uma proteína predita em sorgo (Sobic.010G220800) com função desconhecida e que possui o domínio 'PLATZ' (Pfam:04640). Esse domínio está relacionado a repressão transcricional e é capaz de se ligar a sequências ricas em AT (Nagano *et al.* 2001). Pouco se sabe sobre esse domínio e nenhuma proteína com o mesmo já foi caracterizada funcionalmente. Já para *S. spontaneum*, temos 3 SAS antissenso diferencialmente expressos. Nenhum deles possui proteína predita no transcrito senso, mas foi possível identificar que esses SAS cobrem com alta identidade (>80%) uma parte de transcritos de sorgo. Dentre esses, o SCEQRT2025G01.g também possui expressão diferencial do transcrito senso, e o transcrito de sorgo correspondente contém o domínio AP2 (Pfam:00847). Já para o SAS SCCCST1008F01.g, que parece ser específico de gramíneas por não ter homologia de sequência com nenhuma dicotiledônea, o transcrito de sorgo identificado não possui nenhum domínio conhecido. O transcrito de sorgo correspondente ao SAS SCCCLR2C03E03.g possui o domínio EamA-like transporter family (Pfam:00892). Curiosamente, todos os antissensos, e também o senso de SCEQRT2025G01.g, estão regulados positivamente em *S. spontaneum*, bem como no experimento de entrenó intermediário de *S. spontaneum* contra RB867515. No experimento da RB867515, nenhum transcrito antissenso foi diferencialmente expresso. É possível que esses genes sejam importantes para o desenvolvimento do colmo e são candidatos interessantes a serem caracterizados funcionalmente, principalmente SCCCST1008F01.g por ser, a princípio, específico de gramíneas e não ter nenhum domínio conhecido. O SAS SCEQRT2025G01.g aparece regulado negativamente em experimentos com plântulas tratadas com giberelina, indicando que este gene poderia atuar de forma oposta em alguma etapa da via de transdução de sinal desse hormônio ou poderia atuar como um repressor de alguma função controlada por giberelina. De fato, membros dessa família (AP2) já foram

caracterizados como repressores transcricionais (Song *et al.* 2005). Além disso, a hipótese de que este gene tenha uma correlação negativa com o crescimento/desenvolvimento do colmo é parcialmente suportada pela observação de que este mesmo gene está regulado positivamente em entrenó intermediário de *S. spontaneum* em relação à RB867515 (Apêndice 1), concordando com a hipótese anterior que o crescimento do colmo é mais intenso em RB867515 nesse tecido em relação aos ancestrais. O domínio EamA presente no SAS SCCCLR2C03E03.g não tem função conhecida, mas Vastermark (Vastermark *et al.* 2011) reportou a origem das famílias dos transportadores de moléculas nucleotídeo-açúcar a partir da duplicação do *cluster* gênico *EamA*, sendo possível, então, que este domínio tenha alguma função de transporte ou mesmo no metabolismo de moléculas nucleotídeo-açúcar. Assim, pode-se especular um papel deste SAS no transporte de carboidratos e acúmulo de sacarose ou até mesmo na biossíntese de celulose e hemicelulose, uma vez que a CesA e Csl utilizam nucleotídeo-açúcar (UDP-glicose, por exemplo) como substrato. Evidentemente, experimentos posteriores são necessários para testar essas hipóteses.

5.2.3. Comparação da expressão gênica entre os genótipos ancestrais e progênes segregantes para Brix

Após a análise cuidadosa dos resultados dos genes diferencialmente expressos nos genótipos da RIDESA, Sato (Sato 2012) observou que diversas etapas do metabolismo de hormônios apresentavam genes diferencialmente expressos entre os genótipos de alto e baixo Brix, desde à biossíntese, passando por receptores e fatores de transcrição responsivos a hormônios, até enzimas das vias metabólicas. Assim, foi sugerido um modelo no qual a regulação hormonal influenciaria diretamente o acúmulo de sacarose e a produtividade.

Entretanto, o diagrama de Venn (Figura 19A) dos experimentos de entrenó imaturo e folha mostram uma quantidade pequena de genes (~10%) em comum entre os dois grupos (RIDESA e Ancestrais). Porém, ao analisar a Figura 19B, pode-se observar uma quantidade significativamente maior (~30%) de genes diferencialmente expressos em comum entre os dois genótipos ancestrais de baixo Brix. Ainda, dos 91 genes em comum entre os experimentos com genótipos da RIDESA e os ancestrais, apenas 5 genes são da categoria metabolismo de carboidratos, 2 de biossíntese de hormônios e 2 de transdução de sinal (Tabela 6), mostrando que os possíveis pontos relacionados ao acúmulo de sacarose entre esses dois grupos podem ser diferentes. Ao proceder a análise de categorias funcionais enriquecidas (Tabela 7), as categorias *protein metabolism* e *unknown* são as únicas com *e-score* significativo entre os genes diferencialmente expressos em comum nos dois grupos. Já para os genes diferencialmente expressos apenas nos genótipos da RIDESA (Tabela 7), temos mais categorias enriquecidas, das quais podemos destacar *signal transduction* e *hormone biosynthesis*, o que ajuda a corroborar a conclusão de Sato (Sato 2012). Nos genes diferencialmente expressos exclusivamente nos genótipos ancestrais (Tabela 7), com exceção da categoria *Unknown*, nenhuma outra categoria está enriquecida. Todos esses resultados sugerem que o programa genético que determina a característica de baixo Brix é diferente entre os híbridos e os ancestrais. Isso pode ser devido principalmente ao fato da intensa seleção dos híbridos para o acúmulo de sacarose e, talvez, não exista, verdadeiramente, um genótipo híbrido de baixo Brix que apresente valores similares aos encontrados em *S. spontaneum* e *S. robustum*, principalmente considerando todo o ciclo produtivo da planta. Por outro lado, essa diferença da programação genética também indica que ainda há muita diversidade genética a ser explorada nos genótipos ancestrais que pode

ser muito útil para os programas de melhoramento, bem como para o estudo visando avançar no entendimento das vias regulatórias de interesse.

Ao analisar as categorias funcionais enriquecidas dos genes diferencialmente expressos entre *S. officinarum* e RB867515 (Tabela 5), apesar de não ser um experimento de contraste para Brix e, sim, para biomassa e produtividade, observa-se que as categorias *signal transduction* e *hormone biosynthesis* figuram entre as enriquecidas, da mesma forma que nos genótipos da RIDESA (Tabela 7), seguindo, portanto, a mesma ideia de Sato (Sato 2012) sobre a importância da regulação hormonal no acúmulo de sacarose e produtividade.

Além disso, comparando os experimentos de entrenó imaturo e folha dos genótipos ancestrais e da RIDESA, é possível observar uma quantidade 3 vezes maior de histonas diferencialmente expressas em *S. spontaneum* e *S. robustum*, sendo 14 ocorrências em 4 experimentos dos ancestrais e apenas 9 ocorrências em 8 experimentos da RIDESA. Novamente, como discutido na seção 5.2.1.6, a grande distinção na expressão de histonas entre os genótipos ancestrais e os híbridos interespecíficos sugere que a regulação epigenética é um fator importante na definição das características de interesse que resultam na melhor performance de sacarose e biomassa por colmo dos híbridos comerciais e, talvez, essa regulação seja um dos grandes responsáveis para o salto de performance entre os ancestrais e os híbridos resultantes de seu cruzamento.

No outro grupo de experimentos com plantas contrastantes para Brix provenientes de cruzamentos entre *S. officinarum* e *S. spontaneum* (Papini-Terzi *et al.* 2009), nenhum SAS em comum com os experimentos dos genótipos ancestrais foi encontrado, mas isso é provavelmente devido às diferenças entre as duas plataformas de microarranjo utilizadas: SUCAST (Papini-Terzi *et al.* 2009) e chip Agilent (este trabalho). Entretanto, genes de biossíntese de lignina PAL (SCCCLR1048D07.g) e C4H (SCSGFL4193B05.g) estão reprimidos

nas plantas de alto Brix (Tabela 8), os quais também estão reprimidos em *S. officinarum* e RB867515 nos experimentos de qPCR (Tabela 14). Esses dados ajudam a corroborar a hipótese de que a biossíntese de lignina compete com o acúmulo de sacarose (Seção 5.1), mostrando que vários genótipos com contraste para sacarose apresentam expressão diferencial de genes da via de fenilpropanoides. Outro ponto importante, apesar de não ser o mesmo SAS, duas aquaporinas também estão reprimidas nas plantas de alto Brix (Tabela 8), o que também ocorre com as plantas de alto Brix testadas neste trabalho por qPCR (Tabelas 13-15 e 17), mostrando, novamente, a importância de se estudar com mais detalhes as aquaporinas, uma vez que elas são genes frequentes em quaisquer lista de expressão diferencial entre plantas com contraste para teor de sacarose.

5.2.4. Bibliotecas de cDNA *full length*

Estudos de genômica funcional de cana-de-açúcar são relativamente limitados em comparação com outras gramíneas economicamente importantes como milho e arroz. Entretanto, as novas tecnologias de sequenciamento em larga escala e estratégias de montagem *de novo* podem ajudar a equiparar o conhecimento nessas espécies. No presente estudo, 195 mil *contigs* foram montados a partir do sequenciamento de um total de 5 amostras de 3 genótipos diferentes. Normalmente, a grande maioria dos estudos com RNAseq não se preocupam em realizar uma pré-seleção de transcritos completos, sendo conduzido o sequenciamento e a identificação dos transcritos com os mais variados propósitos desde a identificação de polimorfismo (Zhai *et al.* 2013) e variantes de *splicing* (Loraine *et al.* 2013) até expressão durante o desenvolvimento de tecidos (Li *et al.* 2010) e expressão gênica de genótipos diferentes de uma mesma espécie (Jiao *et al.* 2013;

Wachholtz *et al.* 2013). Até onde vai nosso conhecimento, o presente trabalho é o primeiro a descrever um transcriptoma em larga escala pra cana-de-açúcar e genótipos ancestrais com objetivo principal de criar uma biblioteca de transcritos *full length*. Para isso, um novo protocolo de seleção de transcritos *full length* foi desenvolvido pela equipe da *Life Technologies* (Carlsbad, EUA) e os resultados preliminares mostram que o protocolo foi eficiente, uma vez que o *screening* das bibliotecas mostrou 90% de transcritos completos (Tabela 21). Com a análise dos dados de sequenciamento em larga escala e mapeamento dos *contigs* em outras espécies (Figuras 35 e 38), pôde-se identificar um grande número de transcritos provavelmente *full length* ou muito próximos disso, considerando uma cobertura de >80% dos genes das outras espécies de gramíneas. O exemplo apresentado na seção 4.12 considerando os genes de milho mostra aproximadamente 30 mil transcritos que se enquadram nesse critério, ou seja, *full length* putativos. Esse número é mais que o dobro do reportado anteriormente para a coleção de ESTs do SUCEST (Vettore *et al.* 2003), mostrando que a abordagem de FL-RNAseq foi eficaz em obter um grande número de transcritos completos de cana-de-açúcar, principalmente considerando que o número total de genes estimados para esta espécie está em torno de 35 mil (Vettore *et al.* 2003). Ainda, o tamanho médio dos transcritos de sorgo (1,5 kb), *Setaria* (1,4 kb) e *Brachypodium* (1,6 kb) (Meyer *et al.* 2012) é próximo ao do tamanho médio de 1,1 e 1,4 kb dos insertos das frações S e L, respectivamente (Tabela 21). Somado a isso, ainda existem 39 mil *contigs* montados com tamanho >1 kb (Tabela 22).

Diversos trabalhos já demonstraram a grande semelhança na estrutura do genoma de sorgo e cana-de-açúcar, principalmente em regiões gênicas (Wang *et al.* 2010c; Figueira *et al.* 2012; Kim *et al.* 2013), e os resultados apresentados aqui reforçam essa proximidade entre as espécies ao demonstrá-la em nível transcricional, uma vez que aproximadamente

80 e 90% dos *reads* e *contigs*, respectivamente, possuem algum *hit* contra os genes de sorgo (Figura 35). Esse dados confirmam que o genoma de sorgo será de grande relevância para diversos estudos de genômica e transcriptômica em cana-de-açúcar, incluindo o sequenciamento e a montagem do genoma. Por outro lado, a espécie que tem sido frequentemente considerada um modelo para estudo de gramíneas, *Brachypodium (B. distachyon)*, teve uma das menores porcentagens de mapeamento contra os dados do FL-RNAseq, principalmente em relação aos *reads* (Figura 35A).

Com o FL-RNAseq, foi possível identificar um grande número de transcritos que são provavelmente específicos do gênero *Saccharum*. A comparação com 6 espécies de gramíneas e com o banco de dados do Uniprot revelou 25% dos *contigs* (aproximadamente 49 mil sequências) sem similaridade significativa nesses bancos. Essa porcentagem é similar à encontrada para o transcriptoma da gramínea *Panicum hallii* (Meyer *et al.* 2012) mas muito menor em relação à reportada para *buffalograss* (Wachholtz *et al.* 2013), onde cerca de 75% dos *contigs* não apresentam similaridade de sequência contra banco de proteínas de gramíneas (Ensembl commelinid), mesmo com um tamanho médio de *contigs* relativamente grande (899 bp).

Para se ter uma ideia da cobertura do FL-RNAseq, sua comparação com o SUCEST e o oligoarranjo (Figura 37) mostrou que praticamente todos os genes presentes no oligoarranjo (14.087 de 14.522 sequências) e no SUCEST (40.203 de 43.141) foram detectados, além de acrescentar 78 mil novas sequências. Esse alto número de sequências novas não se deve ao fato de terem sido seqüenciadas as espécies ancestrais de cana-de-açúcar, pois dos 195 mil *contigs* apenas cerca de 600 aparecem exclusivamente nesses genótipos. Essa diferença entre os dois métodos é similar ao reportado para análises do transcriptoma de milho, onde

RNAseq conseguiu detectar 82% de todos os transcritos anotados de milho enquanto o oligoarranjo apenas 56% (Sekhon *et al.* 2013).

O grande número de genes diferencialmente expressos (quase 18 mil, Figura 41) impede uma análise mais aprofundada da função dos genes como foi conduzida para os dados de oligoarranjos e qPCR (Seções 5.2.1 a 5.2.3), o que nos levou a realizar a análise de enriquecimento de categorias para as funções do Gene Ontology (GO). Genes de resposta a estresses estão entre as funções mais frequentes na lista baseada no GO (Figura 40) e entre as categorias funcionais enriquecidas dos genes diferencialmente expressos (Tabela 25). Vale lembrar que *S. officinarum* e *S. spontaneum* diferem entre si pela resistência a estresses (Paterson *et al.* 2013) e a identificação dos genes envolvidos nessas diferenças é a etapa inicial para o entendimento desse mecanismo, mesmo que o transcriptoma tenha sido analisado na ausência de qualquer estresse controlado. Da mesma forma, o transcriptoma de duas variedades de *buffalograss* contrastantes para resistência a ataque de insetos mostrou muitos genes de resposta a estresses expressos em condições normais, sem a indução de estresse biótico (Wachholtz *et al.* 2013). Foram encontrados genes diferencialmente expressos entre as duas variedades contendo os domínios LRR e NB-ARC (Wachholtz *et al.* 2013), os quais, principalmente NB-ARC, estão relacionados a resposta a estresses. Muitas proteínas de resposta a estresses possuem domínios LRR e NB-ARC e estão envolvidas no reconhecimento inicial do ataque de patógenos (Takken *et al.* 2006). Dois genes LRR foram induzidos mais de 100 vezes em sorgo em resposta à infecção pelo fungo *Bipolaris sorghicola* e os autores acreditam que essas proteínas sejam importantes para a sinalização da resposta à infecção (Yazawa *et al.* 2013). Esses domínios estão entre os mais frequentes nas proteínas preditas nos *contigs* (Tabela 24), sugerindo uma importância desse domínios para a diferença entre os genótipos *S. officinarum* e *S. spontaneum*.

Entre as funções enriquecidas do GO nos genes diferencialmente expressos (Tabela 25), várias apresentam relação com o metabolismo de parede celular. Isso é em parte devido à análise de dois entrenós diferentes, de forma análoga à que ocorre na comparação entre entrenó imaturo e intermediário nos oligoarranjos (Tabelas 9 e 10 e Figura 20). Recentemente foi relatada a expressão diferencial de vários genes da via da lignina entre os tecidos de *Panicum hallii* (Meyer *et al.* 2012), o que é coerente com as funções enriquecidas de biossíntese de lignina (*lignin biosynthesis* da categoria *Biological Process*) entre os genes diferencialmente expressos dos genótipos ancestrais e a SP803280 (Tabela 25). Poucas funções da categoria *Molecular Function* relacionadas ao metabolismo de carboidratos estão enriquecidas (Tabela 25), assim como nos dados de qPCR e oligoarranjos (Tabelas 5, 13-15 e Figura 14), apenas funções relacionadas ao transporte de carboidratos aparecem na lista (Tabela 25). Esse resultado está de acordo com observações anteriores onde a expressão de genes do metabolismo de carboidratos são relativamente raros em entrenós em amadurecimento de cana-de-açúcar e que, dentre os que são expressos, transportadores de açúcares são os mais abundantes (Casu *et al.* 2003). Genes que são recorrentes entre os genes diferencialmente expressos em variedades de cana-de-açúcar contrastantes para Brix (Papini-Terzi *et al.* 2009; Sato 2012) envolvem as funções de transporte de água (que inclui proteínas como as aquaporinas) e metabolismo de hormônios, as quais também estão enriquecidas no FL-RNAseq do transcriptoma dos ancestrais (Tabela 25). Esses resultados sugerem que os mecanismos que determinam a característica de alto ou baixo Brix envolvem a conservação da expressão desses genes, tanto nos genótipos ancestrais (este trabalho) quanto nos genótipos híbridos (Papini-Terzi *et al.* 2009; Sato 2012). Além de reafirmar a importância de genes dessas categorias, as semelhanças entre RNAseq e oligoarranjos mostram uma coerência entre essas duas plataformas de análise do

transcriptoma. Apesar das limitações dos oligoarranjos em relação ao RNAseq, uma alta correlação de Pearson ($>0,7$) entre essas plataformas foi reportada recentemente em análises do transcriptoma de milho (Sekhon *et al.* 2013).

Um passo fundamental para a manipulação genética de qualquer organismo é o conhecimento completo da sequência de interesse a ser manipulada, e a disponibilidade da biblioteca física, ou seja, dos clones em *E. coli* dos cDNAs *full length* descritos aqui, será útil para o fácil acesso a uma gene de interesse para propósitos de clonagem e manipulação genética. A biblioteca também será uma boa ferramenta por ter sido construída em vetores do sistema Gateway (Invitrogen), que permite a troca do inserto entre vetores de forma relativamente simples. Além disso, as sequências produzidas aqui ainda serão alvos de estudos futuros para identificar diferenças genéticas como, por exemplo, SNPs, variantes de *splicing* e expressão alelo-específica, que possam ser relevantes para identificação de marcadores moleculares ou variações associadas às diferenças nas características agronômicas dos genótipos e podem ajudar a aprofundar nosso conhecimento da função dos genes de cana-de-açúcar. É possível que sejam identificadas, por exemplo, casos de expressão alelo-específica, os quais podem estar associados ao ganho de performance do híbrido em relação às espécies ancestrais, como ocorre em híbridos de laranja (Jiao *et al.* 2013), *Cirsium arvense* (Bell *et al.* 2013), arroz (Zhai *et al.* 2013), e várias outras espécies.

5.2.5. Expressão de NATs

Uma das vantagens do *chip* de oligonucleotídeos é a possibilidade de detecção de transcritos antissenso (Lembke *et al.* 2012). Esses transcritos, denominados *natural antisense transcripts* (NATs), podem ser classificados como *cis*-NATs, quando originados da

transcrição da fita oposta ao transcrito senso no mesmo locus genômico apresentando complementaridade perfeita, e *trans*-NATs, que são originados de outros loci e possuem complementaridade imperfeita ao transcrito senso (revisado por (Lapidot & Pilpel 2006)). A presença de NATs já foi reportada em diversos organismos (revisado por (Lapidot & Pilpel 2006)) e um banco de NATs para mais de 60 espécies de plantas foi recentemente desenvolvido (Chen *et al.* 2012). Os NATs podem atuar na regulação da expressão do seu par senso através de diferentes mecanismos, como a interferência espacial de duas maquinarias de transcrição no mesmo locus caminhando em direções opostas ou pelo pareamento entre as fitas senso e antissenso formando dsRNA que pode disparar o mecanismo de silenciamento gênico (RNAi), levando à degradação do mRNA ou impedir a ligação de fatores importantes para outros processos, como *splicing*, tradução e transporte do mRNA, por exemplo (Britto-Kido *et al.* 2013). Ainda, os NATs podem direcionar alterações na estrutura da cromatina levando à regulação epigenética da expressão de genes (Magistri *et al.* 2012). Quase 1.000 NATs foram detectados nos quatro genótipos pelos oligoarranjos (Tabela 11), sendo que em todos os tecidos a RB867515 apresentou o maior número de NATs expressos. É possível que haja uma diferença de controle epigenético do genoma híbrido. Para os genes do catálogo de parede celular, foi possível detectar 40 NATs expressos, os quais podem ser importantes reguladores da biossíntese de parede celular secundária. De fato, já foi demonstrada a existência de transcritos antissenso em cevada que modulam a expressão de genes da família Cesa/Csl (Held *et al.* 2008). Dentre os NATs que apresentam expressão em todos os genótipos, pode-se destacar o NAT do gene *COMT*, que possui expressão diferencial em *S. spontaneum* (Figuras 24 e 25). Pelo fato da *COMT* ser um das principais responsáveis pela síntese de siringil (Figura 2), o antissenso desse gene pode desempenhar um papel importante na regulação temporal precisa da síntese desse monolignol, uma vez que siringil

está relacionado a estágios mais tardios de lignificação (Donaldson 2001; Rencoret *et al.* 2011). Essa hipótese também é suportada pela menor expressão desse NAT em entrenó intermediário de *S. spontaneum* (Apêndice 6, cluster 5), tecido o qual já apresenta maior quantidade de siringil (Figura 11B) e maior expressão da *COMT* (Tabela 14). A regulação da expressão da *COMT* é particularmente importante para cana-de-açúcar, uma vez que a regulação negativa desse gene via RNAi leva a melhora da sacarificação (Jung *et al.* 2012). Assim, o entendimento da regulação desse gene pode ajudar no desenvolvimento de novas estratégias de manipulação para diferentes objetivos.

Outro gene que se destaca por ter padrão de expressão diferencial é o gene *kobito*, ainda pouco caracterizado. Acredita-se que *kobito* seja importante para a síntese de celulose (Pagant *et al.* 2002), mas recentemente foi demonstrado que esse gene também possui um papel na regulação da permeabilidade dos plasmodesmas, uma vez que *Arabidopsis* mutante com perda de função de *kobito* apresenta aumento da permeabilidade (Kong *et al.* 2012). Esse gene possui expressão do NAT (Figura 25) e também do transcrito senso com repressão em entrenó imaturo de *S. spontaneum* em relação às outras plantas (Figura 26). Como discutido anteriormente (seção 5.1), a lignificação de um maior número de células no córtex do colmo de *S. spontaneum* (Figura 12D e H), incluindo células parenquimáticas, pode resultar na diminuição do transporte de água pela rota apoplástica o que resultaria no aumento da expressão de aquaporinas (Tabelas 14 e 15), compensando o transporte pela rota simplástica. Essa hipótese também abre espaço para sugerir que a menor expressão de *kobito* também seria uma consequência da maior lignificação, de forma a aumentar a permeabilidade dos plasmodesmas para ajudar a compensar a diminuição do transporte pela via apoplástica não só de água, mas também de outras moléculas.

Os dados de FL-RNAseq também mostraram um número significativo de NATs expressos (Figuras 42 e Tabela 26). Entretanto, considerando o número total de transcritos (ou *reads*), o FL-RNAseq (Tabela 26) identificou uma porcentagem muito menor em relação ao oligoarranjo (Tabela 11), sendo que o FL-RNAseq identificou menos de 1% nas espécies mais próximas (Tabela 26) como sorgo (*S. bicolor*) e milho (*Z. mays*). Por outro lado, quando mapeado contra o SUCEST, 17% dos *reads* alinhados foram identificados como antissenso (Tabela 26). Talvez essa diferença seja explicada pelos transcritos específicos de cana-de-açúcar, que possam conter um grande número de transcritos antissenso. Além disso, a detecção de NATs não é uma tarefa simples, pois alguns transcritos podem ser mapeados como senso em uma espécie e antissenso em outras (dados não mostrados), dificultando a análise. NATs podem mapear também em regiões intergênicas (Sherstnev *et al.* 2012) e mesmo em espécies próximas como cana-de-açúcar e sorgo, essas regiões não são tão conservadas quanto às regiões gênicas (Wang *et al.* 2010c). A detecção de NATs é ainda mais complicada em cana-de-açúcar pela ausência de um genoma referência e genes preditos. Esse viés pode ser minimizado com o uso de uma anotação mais acurada dos transcritos de cada espécie.

Uma observação importante é que ambas as plataformas utilizam *primers* oligodT para a síntese do cDNA e a diferença principal entre os métodos para a síntese de cDNA é que as bibliotecas de cDNA *full length* foram também selecionados pela presença do cap-5' (ver metodologia), e a menor detecção relativa de NATs no FL-RNAseq sugere que a grande maioria não possuem o cap-5'. Ainda, mesmo sendo um transcrito antissenso, os NATs podem apresentar ORFs; em arroz, já foi demonstrado que 86% dos transcritos que possuem um par antissenso possuem ORF nas duas orientações (Osato *et al.* 2003). Entretanto, ainda não se sabe a importância biológica das ORFs antissenso e se elas são realmente traduzidas.

Provavelmente, a ausência do cap-5' dos NATs pode alterar algumas propriedades, como o reconhecimento da maquinaria de tradução ou mesmo sua estabilidade, o que dificultaria a tradução dessa ORFs.

O aprofundamento das análises dos dados do FL-RNAseq poderão ajudar a entender os mecanismos regulatórios de características como o acúmulo de sacarose em cana-de-açúcar. Observações como a existência de uma expressão maior de NATs em folha (Figura 42) em relação aos entrenós ainda requer estudos mais detalhados para decifrar seu significado biológico. Outras análises ainda podem responder questões relacionadas, por exemplo, à associação da expressão de NATs com diferentes padrões de *splicing* ou regulação da estrutura da cromatina em cana-de-açúcar e quais as diferenças em relação aos genótipos híbridos e ancestrais.

5.3. Redes de coexpressão de genes associados a biossíntese de lignina

As análises de redes de genes coregulados são interessantes para complementar as análises de genes diferencialmente expressos no sentido de identificar genes que possam estar relacionados a determinadas vias e que possam estar diretamente ligados às características de interesse. Mesmo a rede (Figura 31) tendo sido construída a partir de dados de expressão de milho, as informações extraídas desse resultado podem ser extrapoladas para cana-de-açúcar, pois muitos módulos de genes coexpressos são conservados entre as espécies (Zarrineh *et al.* 2010; Ruprecht *et al.* 2011), mesmo entre espécies distantes como arroz e *Arabidopsis* (Movahedi *et al.* 2012). Portanto, análises de coexpressão poderão ser úteis para revelar novos aspectos da biossíntese de parede celular,

e análises comparativas entre espécies ajudarão a destacar elementos conservados e divergentes nesse processo.

Enquanto as análises comparativas de coexpressão para genes relacionados à celulose e hemicelulose já foram conduzidas em várias espécies, as relações conservadas de expressão entre os genes do metabolismo de lignina tem sido pouco estudadas (Ruprecht & Persson 2012). A análise de genes coregulados com *4CL* (Figura 31 e Tabela 18) mostra uma relação entre genes que podem estar envolvidos no metabolismo de lignina em gramíneas. Além da óbvia relação entre os genes da via de fenilpropanoides presentes na lista (*PAL*, *CAD*, *CCR*, *COMT* e *CCoAOMT*, Tabela 18), que corrobora a hipótese de coregulação da via e rede, também é possível identificar diversos genes com função relacionada à parede celular, mas não diretamente com lignina. Glicosiltransferases das famílias 43 e 47 (GT43 e GT47) estão inclusas na rede (Tabela 18). Estas enzimas estão envolvidas na biossíntese do esqueleto de xilanas de hemiceluloses em *Arabidopsis* (Wu *et al.* 2009; Lee *et al.* 2011) e são muito provavelmente responsáveis pela biossíntese de GAX em gramíneas (Mitchell *et al.* 2007; Lovegrove *et al.* 2013), a principal hemicelulose dessas plantas. Ainda é possível observar o gene UDP-glicose desidrogenase (Tabela 18), responsável pela síntese de UDP-glucuronato que será substrato para produção de GAX. Os resultados apontam para uma regulação coordenada da biossíntese de GAX e lignina, o que faz sentido pelo fato dos resíduos de arabinose presentes em GAX serem pontos de ligação com a lignina (Ralph *et al.* 2004a; Burr & Fry 2009). Por outro lado, as celulose sintases da rede (*ZmCesA9* e *ZmCesA4*, Tabela 18) são filogeneticamente mais próximas das CesAs de parede primária (Holland *et al.* 2000), o que é, aparentemente, inesperado já que a lignina é restrita à parede secundária.

Entre os fatores de transcrição, a análise do programa Inparanoid indica que SCEZLB1006E02.g e SCJFRZ2015F02.g são homólogos putativos de *KNAT7* e *MYB61* de

Arabidopsis (Tabela 16 e dados não mostrados). Recentemente foi demonstrado que *KNAT7* é um regulador negativo da biossíntese de parede celular em Arabidopsis (Li *et al.* 2012) e que *MYB61* é capaz de se ligar aos elementos AC (SMREs) do promotor de *KNAT7* e transativá-lo em ensaios *in vitro*, sugerindo que a expressão de *KNAT7* seja diretamente regulada por *MYB61* (Romano *et al.* 2012). O fato desses dois genes estarem corregulados (Tabela 18) indica que a mesma relação observada em Arabidopsis pode ser conservada em gramíneas. Porém, a corregulação de *KNAT7* com genes da via de biossíntese de lignina parece ser contrária hipótese de que *KNAT7* seja um repressor dessa via, da mesma forma como observada nos resultados de qPCR (Tabelas 13-15 e seções 4.9.2 e 5.2.1.5.1). Esses dois resultados sugerem que a função de *KNAT7* em gramíneas seja diferente da função em Arabidopsis, mostrando uma possível divergência evolutiva entre gramíneas e dicotiledôneas. Diferentemente da rede da Figura 31, *KNAT7* e *MYB61* não são observados em redes de coexpressão com um gene da via de fenilpropanoides (*PAL1*) de Arabidopsis (Ruprecht & Persson 2012). Outros TFs que aparecem na lista, *Platz* e *Tify*, são menos caracterizados em plantas e podem ser bons candidatos ao cargo de reguladores do metabolismo de parede celular de gramíneas, sendo interessantes alvos para análises posteriores.

Como um dos principais objetivos de redes de coexpressão é identificar novos genes relacionados a um determinado processo biológico, a identificação de genes com função desconhecida (*unknown*, Tabela 18) é um resultado bem interessante. Essa abordagem já foi responsável por identificar genes com o domínio de função desconhecida DUF579 coexpressos com outros genes marcadores para parede celular (Brown *et al.* 2011; Jensen *et al.* 2011), tanto em gramíneas como em dicotiledôneas (Ruprecht *et al.* 2011). A partir desses resultados, grupos independentes conseguiram encontrar evidências que o DUF579

possa estar envolvido na síntese de xilanas (Brown *et al.* 2011; Jensen *et al.* 2011), apesar da exata função desse domínio ainda não ser conhecida. Entre os resultados da Tabela 18, genes que possuem dois domínios de função desconhecida, DUF563 e DUF231, mostram-se bons candidatos para análises posteriores, incluindo para manipulação genética visando ao melhoramento da biomassa para biocombustíveis. Mesmo que sua função ainda permaneça desconhecida, já foi reportado o envolvimento de DUF231 com a esterificação e acetilação de carboidratos de parede celular e sua consequente importância para a biossíntese de celulose de parede secundária em *Arabidopsis* (Bischoff *et al.* 2010; Yuan *et al.* 2013), e também evidências que a expressão de uma proteína contendo o domínio DUF231 é regulada por SND1 (Yuan *et al.* 2013), um dos principais reguladores da biossíntese de parede celular secundária em *Arabidopsis*. Da mesma forma, DUF563 também apresenta evidências de ser importante para o metabolismo de parede celular, por estar presente em xilosiltransferases (Chiniquy *et al.* 2012; Hwa *et al.* 2012). Não por acaso, ambos DUFs são recorrentes em lista de genes coexpressos com genes do metabolismo de parede celular, incluindo com genes da biossíntese de lignina, em diferentes espécies (Ruprecht *et al.* 2011; Ruprecht & Persson 2012) e já existe uma patente protegendo o uso de DUF563 para o aumento da biomassa vegetal (Apuya 2012).

A validade da extrapolação dos dados da rede de milho (Figura 31 e Tabela 18) para cana-de-açúcar fica evidente nas Figuras 32 e 33. Coerente com a hipótese de que *S. spontaneum* inicia a lignificação precocemente, a grande maioria dos genes coexpressos com *4CL* estão nos clusters onde *S. spontaneum* apresenta maior expressão em entrenó imaturo, dentre eles *PAL*, *4CL*, *MYB* (homólogo de *AtMYB61*) e UDP-glicose desidrogenase (Figura 32). Em entrenó intermediário, já pode-se observar o mesmo MYB mais expresso em RB867515 assim como uma expressão mais similar de *KNAT7* (knotted-1-like) em *S.*

spontaneum e RB867515, o que pode indicar que na variedade comercial a programação da biossíntese de parede celular secundária está sendo ativada nesse tecido, coerente com a observação que nesse tecido a expressão gênica está direcionada para o crescimento (Seção 5.2.2), que inclui parede celular secundária.

5.4. Identificação de promotores associados a genes de parede celular

Em alguns casos, a alteração da expressão de genes de interesse com promotores constitutivos, como o 35S e UBI1, não é suficiente para alcançar o fenótipo desejado. Isso porque os promotores constitutivos vão promover a alteração da expressão do gene de interesse em todas as células da planta, até mesmo em células onde essa alteração seja deletéria para planta. Um trabalho que expressa muito bem o uso de promotores específicos é a reprogramação da deposição de parede celular secundária em *Arabidopsis* (Yang *et al.* 2013). O promotor específico de feixes vasculares do gene *VND6* foi usado para expressar *C4H* em um mutante defeutivo de *C4H*, de forma a recuperar a síntese de lignina apenas nos feixes vasculares. Depois, essa planta foi transformada com uma construção com o TF ativador da biossíntese de parede celular secundária, *NST1*, sob controle do promotor do gene *IRX8*, o qual é ativado pelo próprio *NST1*, formando o *loop* de *feedback* positivo. Assim, as plantas transgênicas contendo simultaneamente as duas construções restringiam a síntese de lignina aos feixes vasculares enquanto aumentava a deposição de parede celular secundária nas fibras, com um consequente aumento da eficiência de sacarificação sem qualquer prejuízo de performance (Yang *et al.* 2013).

Um exemplo em cana-de-açúcar é o silenciamento do gene *COMT*. Com o objetivo de diminuir a recalcitrância da biomassa e aumentar o teor de açúcar no colmo pela redução da

lignina, dois grupos independentes produziram plantas transgênicas silenciadas para o gene *COMT* (Barboza 2012; Jung *et al.* 2012; Jung *et al.* 2013). Apesar de ambos os grupos conseguirem resultados similares na questão da diminuição do teor de lignina e aumento da degradabilidade do bagaço, as plantas transformadas com a construção de silenciamento sob controle do promotor constitutivo UBI1 de milho (Barboza 2012) apresentaram colmos mais finos e sérios problemas de sustentação, acamando com muita facilidade. Já as plantas transformadas com a construção de silenciamento sob controle do promotor de um gene da via de fenilpropanoides (*C4H*) não apresentaram problemas de acamamento, possuindo performance similar às plantas selvagens, tanto em casa de vegetação (Jung *et al.* 2012) como em experimentos de campo (Jung *et al.* 2013).

Um dos primeiros passos para a caracterização de promotores de genes do metabolismo de parede celular em cana-de-açúcar é a identificação dos cis-elementos responsáveis pela ligação dos fatores de transcrição que irão controlar sua atividade. Os principais cis-elementos de resposta a fatores de transcrição de biossíntese de parede celular são os SNBEs e SMREs, os quais foram inicialmente caracterizados em *Arabidopsis* (Raes *et al.* 2003; Zhong *et al.* 2010; Zhong & Ye 2012). Posteriormente, descobriu-se que eles atuam de forma semelhante em gramíneas (Fornale *et al.* 2010; Valdivia *et al.* 2013). Assim, esses *binding sites* foram buscados nos promotores de alguns genes de cana-de-açúcar (Tabelas 19 e 20) e, de fato, eles estão presentes em vários promotores. Alguns fatores NAC e MYB apresentaram SNBEs (Tabela 19), de forma similar ao que ocorre em *Arabidopsis* (Zhong *et al.* 2010), e o mesmo ocorre com os SMREs. Entretanto, o SAS SCJLRZ1019A02.g, o homólogo de *AtMYB58/63* (*ScMYB45*, Tabela 16), não apresentou nenhum dos cis-elementos analisados, mesmo em 1.500 nt de sequência promotora (Tabela 19), mesmo que o homólogo de *Arabidopsis AtMYB58/63* possua SMREs em seu promotor

(Zhong & Ye 2012). Essa observação, somada à falta de correlação com a expressão de seus alvos putativos (Tabelas 13-15, Seções 4.9.2 e 5.2.6.1) mostram uma diferença importante em relação à função de *AtMYB58/63* e seu homólogo em cana-de-açúcar *ScMYB45*, sugerindo uma diferença nos mecanismos regulatórios entre essas espécies. Interessantemente, o transcrito antissenso do MYB SCACST3159E04.g também apresentou 4 SNBEs em seu promotor, mesmo sendo uma sequência de apenas 300 nt (Tabela 19) abrindo espaço para a hipótese de mecanismos de regulação transcricional do metabolismo de parede celular via antissenso. Para o gene *COMT*, seu antissenso não apresentou nenhum *binding site*, mas possui alguns similar à sequência consenso, com 1 *mismatch* (Tabela 19). O transcrito senso do gene *COMT* possui 3 SMREs, analogamente ao que ocorre em milho (Fornale *et al.* 2010), mas diferentemente de *Arabidopsis*, que não possui SMREs no promotor da *COMT* (Raes *et al.* 2003). Novamente, esses resultados sugerem uma divergência evolutiva importante entre essas espécies. Já pra os promotores escolhidos da rede de genes coexpressos com *4CL* (Tabela 20), vemos que vários possuem algum cis-elementos de resposta à TFs de parede celular, complementando os resultados da análise de co-expressão (Figura 31). Assim como *COMT*, os genes *CCR* e *PAL* apresentam 6 e 4 SMREs, respectivamente (Tabela 20). Em *Arabidopsis*, esses mesmos genes também possuem SMREs (Raes *et al.* 2003), mas ao contrário de cana-de-açúcar, possuem apenas um sítio em cada promotor. O MYB SCJFRZ2015F02.g, homólogo de *AtMYB61*, apresenta 3 SNBEs e 1 SMRE, aumentando a possibilidade desse MYB estar envolvido no metabolismo de parede celular, mas vale lembrar que sua função não parece ser de reprimir genes como ocorre em *Arabidopsis* (Romano *et al.* 2012). Entre os genes desconhecidos, apenas o SCSGST3118D04.g teve seu promotor identificado e este apresentou todos os *binding sites* analisados, o que o torna um bom candidato para uma caracterização posterior, lembrando

que esse gene não possui nenhum domínio conhecido, nem mesmo domínios de função desconhecida (DUFs).

5.5. Cana energia e potencial biotecnológico de *S. spontaneum*

S. spontaneum apresenta um grande potencial para produção de biomassa, sendo maior que uma variedade comercial importante como a RB867515, a qual representa 20-25% da área plantada no Brasil (Dal-Bianco *et al.* 2012). Mesmo que decorrentes de uma análise em pequena escala, os resultados são promissores. Considerando o perfilhamento (Figura 8D), massa do colmo (Figura 8C), e porcentagem de matéria seca (Tabela 3), a estimativa de produtividade de *S. spontaneum* é de 45,5 tons/ha de massa seca, maior que RB867515 (37,2 tons/ha) e que a média comercial atual de 39 tons/ha (Waclawovsky *et al.* 2010). Ainda, *S. spontaneum* apresenta outras características agronômicas importantes, como maior WUE (Figura 9D), resistência à estresse hídrico (Neufeld *et al.* 1992), e resistência a outros estresses bióticos e abióticos já conhecidos por muito tempo (Paterson *et al.* 2013). Além disso, devido ao menor conteúdo de água (Tabela 3), o volume total para transporte e a perda de caldo com carboidratos durante a colheita e estocagem serão menores, reduzindo os prejuízos e os custos de transporte.

A lignina, de forma geral, é um dos principais obstáculos para a hidrólise da parede celular das plantas para a produção de biocombustíveis e, logo, o alto conteúdo de lignina de *S. spontaneum* seria um problema caso fosse utilizada como matéria-prima para produção de etanol celulósico. Entretanto, como mencionado anteriormente, a lignina tem densidade energética maior que a celulose (White 1987a; Raveendran & Ganesh 1996) e poderia ser usada para produção de eletricidade pela queima do bagaço, podendo gerar mais energia

que outros genótipos com menos lignina. Ainda, a lignina poderia ser recuperada, após etapas de pré-tratamento, e ser utilizada como matéria-prima de diversos produtos químicos de alto valor agregado, tais como vanilina e DMSO (Calvo-Flores & Dobado 2010). Já existem protocolos com resultados promissores para o fracionamento dos componentes da biomassa, incluindo a recuperação da lignina com pureza superior a 90% (Bozell *et al.* 2011). Nesse contexto, o alto teor de lignina não deve ser visto como um problema em todos os casos, mas também como uma característica potencialmente interessante. Além disso, *S. spontaneum* pode ter uma lignina menos interligada e, conseqüentemente, não ser tão recalcitrante como o teor total sugere. A regulação positiva da enzima α -L-arabinofuranosidase em entrenó imaturo (Tabela 13) e o maior teor de siringil em entrenó intermediário e maduro (Figura 11B) suportam essa hipótese, visto que essa enzima reduz as possibilidades de ligação entre a lignina e a hemicelulose e uma lignina rica em unidades de siringil tende a ser mais linear e menos entrelaçada (Bonawitz & Chapple 2010). É possível que *S. spontaneum* tenha uma lignina que seja removida eficientemente com pré-tratamentos específicos, produzindo uma biomassa viável para ser substrato para a produção de etanol celulósico. De fato, alguns trabalhos mostraram que *S. spontaneum* possui um bom potencial para a produção de etanol de segunda geração (Chandel *et al.* 2009; Scordia *et al.* 2010; Chaudhary *et al.* 2012). Dois grupos independentes testaram diversos tipos de pré-tratamentos e conseguiram converter 85 a 89,9% dos glucanos iniciais em açúcares monoméricos solúveis (Scordia *et al.* 2010; Chaudhary *et al.* 2012). Resultados preliminares de sacarificação sem pré-tratamento dessas amostras demonstraram uma baixa eficácia (dados não mostrados), porém, em experimentos similares com *S. spontaneum* e RB867515, o pré-tratamento brando com ácido permitiu uma sacarificação similar do colmo dos dois genótipos (Poelking 2012), sugerindo que a maior quantidade de

lignina de *S. spontaneum* pode, realmente, não ser um problema para a degradação do bagaço. Além desses dados, o uso de *S. spontaneum* como uma cultura dedicada à bioenergia é particularmente interessante para países como o Brasil onde já existe um grande conhecimento no cultivo de cana, uma vez que a introdução de uma nova cultura não levará à grandes mudanças na cadeia produtiva (práticas de agricultura, logística e manejo, por exemplo) que poderiam ocorrer com a introdução de outras plantas como *Miscanthus* ou *Switchgrass*.

Para a indústria baseada em bioetanol de cana-de-açúcar atingir uma maior sustentabilidade, seria necessário que a biomassa estivesse disponível ao longo de todo o ano, o que não acontece atualmente pelo fato da colheita acontecer apenas uma vez ao ano por um período relativamente curto para coincidir com o época na qual a planta está com maior acúmulo de sacarose (Souza *et al.* 2013). Uma vez que o plantio de *S. spontaneum* não seria focado em sacarose, a safra poderia ser administrada de forma a manter um fluxo contínuo durante o ano.

Os principais fatores fisiológicos associados à maior produtividade levados em consideração para o melhoramento da cana-de-açúcar incluem tolerância a seca e frio, baixo florescimento, quantidade de fibra, capacidade de rebrotamento, uniformidade de diâmetro e altura dos colmos, perfilhamento, crescimento ereto, velocidade do brotamento e crescimento rápido para competir com plantas daninhas (Souza *et al.* 2013). Muitos desses fatores são característicos de *S. spontaneum*, o que ajuda a corroborar o potencial biotecnológico desse genótipo para produção de biomassa. Historicamente, com o objetivo de aumentar o acúmulo de sacarose por sucessivos retrocruzamentos com *S. officinarum* da progênie híbrida (Tabela 1), a nobilitação optou por diminuir a porcentagem do genoma de *S. spontaneum* nas variedades produzidas. Assim, os programas de melhoramento podem

utilizar *S. spontaneum* de forma a produzir variedades “*desnobilizadas*”, ou seja, cruzando variedades de cana-de-açúcar com *S. spontaneum* visando aumentar a biomassa produzida e também introduzir todas essas características citadas acima. Alguns resultados já mostram que *S. spontaneum* pode ser utilizada com sucesso para esse propósito, onde o cruzamento de variedades de cana-de-açúcar com *S. spontaneum* gerou progênies com maior quantidade biomassa por área e maior porcentagem de matéria seca por colmo (Wang *et al.* 2008). Assim, além do potencial para ser usada diretamente para a produção de biomassa, *S. spontaneum* também pode ser introduzida de forma mais intensa nos programas de melhoramento visando ao aumento de produtividade e ao desenvolvimento de cana energia. Mesmo apenas para estudos voltados para o entendimento de parede celular, nota-se que *S. spontaneum* apresenta um maior número de genes do catálogo de genes (Tabela 2) com maior expressão (Figuras 24 e 27 e Apêndices 6 e 7), mostrando o potencial desse genótipo como modelo para identificação de genes envolvidos no metabolismo de parede celular.

6. CONCLUSÃO

Neste trabalho, analisamos o perfil de expressão gênica de milhares de genes em três genótipos ancestrais, *S. officinarum*, *S. robustum* e *S. spontaneum*, e dois híbridos comerciais, RB867515 e SP803280, por técnicas de microarranjo, qPCR e seqüenciamento em larga escala. As análises dos genótipos ancestrais vêm para avançar nos conhecimentos sobre a variabilidade genética do gênero *Saccharum*, ajudando na questão da necessidade de ampliar a base genética estreita dos híbridos comerciais modernos (Roach 1989; Ming *et al.* 2006), o que pode estar por trás da redução dos ganhos de produtividade (Dal-Bianco *et al.* 2012).

A comparação dos resultados do presente trabalho com outros estudos indicou que genes como aquaporinas e expansinas tem o perfil de expressão frequentemente associado ao contraste para Brix em variedades de cana-de-açúcar, mostrando pontos conservados na determinação dessa característica e que estes podem estar relacionados à expansão celular e ao controle da osmolaridade das células. Por outro lado, também foi possível identificar indícios que o baixo Brix de *S. robustum* e *S. spontaneum* pode ser regulado por vias diferentes das que regulam essa mesma característica nos genótipos híbridos, o que mostra que há muito a ser estudado para entendermos as relações de partição de carbono na planta.

A análise comparativa dos dados de expressão de fatores de transcrição e de caracterização parcial de promotores de outras espécies sugere alguns pontos conservados, mas também algumas divergências evolutivas importantes em relação ao controle do metabolismo de parede celular entre gramíneas e dicotiledôneas. Apesar de muito do conhecimento sobre dicotiledôneas poder ser extrapolado para gramíneas, essas

divergências indicam que será necessário um melhor entendimento dessas vias regulatórias em gramíneas para possibilitar a elaboração de estratégias mais adequadas para a produção de variedades focadas na produção e biomassa (parede celular) para etanol celulósico.

Em termos fisiológicos, identificamos que o acúmulo de sacarose pode competir diretamente com a lignificação e crescimento, e o controle do crescimento e expansão celular em determinados entrenós, principalmente entrenós mais jovens, pode ser fundamental para a maior produtividade da planta. Genes associados a essas características foram identificados pelo perfil transcricional, podendo-se destacar genes do metabolismo de hormônios, incluindo síntese e transdução de sinal, genes do metabolismo de parede celular e, genes relacionados à estrutura da cromatina. Tais genes podem ser alvos interessantes para manipulações genéticas futuras. Um resumo das principais características fisiológicas e os genes relacionados a essas características é apresentado na Figura 43.

Figura 43 (próxima página). Resumo das principais características apresentadas pelos genótipos analisados, mostrando relação com a expressão gênica e o fenótipo. A figura integra a expressão de genes que podem estar relacionados às características apresentadas pelo genótipo e como essas características se relacionam no fenótipo da planta. Na parte superior é apresentada a legenda da figura. Palavras em vermelho e verde indicam padrão aumentado e diminuído, respectivamente. Não são mostrados dados para *S. robustum* pois essa planta não apresentou uma correlação clara entre a expressão gênica e o fenótipo, como a apresentada por *S. spontaneum*, *S. officinarum* e RB867515. Apesar de vários genes estarem relacionados com o acúmulo de sacarose, apresentando perfil antagônico de expressão em plantas de alto e baixo Brix (vide discussão), genes do metabolismo de sacarose não apresentaram tal perfil, o que está representado na figura como um “?” para a característica acúmulo de sacarose. Os resultados também indicam que a regulação epigenética pode influenciar as características da RB867515, mas a correlação clara de qual característica ou gene pode sofrer maior influência não foi determinada, o que está representado na figura por um “?”. Da mesma forma, o “?” no fator de transcrição ScMYB22, homólogo de ZmMYB31, representa a possível participação desse fator de transcrição na ativação da lignificação, mesmo sendo esperada uma repressão dos genes da via de biossíntese de lignina por ScMYB22. A figura também evidencia a grande diferença de fenótipo de *S. spontaneum* em relação às demais, mostrando mais características alteradas. WUE, uso eficiente de água; Gs, condutância estomática; E, transpiração.

Vale ressaltar também a grande diferença fenotípica e de perfil de expressão apresentado por *S. spontaneum*. Essas diferenças apontam para um potencial biotecnológico desse genótipo para uso como fonte de biomassa ou mesmo para estudos visando o entendimento de vias relacionadas à síntese de lignina ou mesmo de fibras.

Também demos um passo inicial para identificação de genes, incluindo alguns com função desconhecida, relacionados ao metabolismo de parede celular pelas análises de co-expressão e caracterização parcial de promotores de genes com perfil de expressão com correlação positiva com acúmulo de lignina. À medida que novas ferramentas estiverem disponíveis, poderemos utilizar dados de expressão de cana-de-açúcar, ao invés da análise comparativa com dados de expressão de milho utilizada no presente trabalho, para atender às particularidades dessa planta.

As informações geradas pelo FL-RNAseq fornecem um complemento importante para os dados de microarranjo apresentados aqui e ajudará no futuro entendimento da dinâmica envolvida nas diferenças entre os genótipos ancestrais de cana-de-açúcar, além de fornecer um catálogo com milhares de genes completos que será de grande importância para estudos de genômica funcional e manipulação genética em cana-de-açúcar. Os milhares de transcritos sequenciados serão alvo de estudos posteriores em nosso laboratório e, dentre eles, podemos destacar cerca de 30 mil cDNAs *full length* putativos, transcritos antissenso, 49 mil transcritos específicos de cana-de-açúcar e 78 mil novos transcritos. Muita informação sobre os polimorfismos e alelos poderá ser identificada, as quais são interessantes para acessar a diversidade genética dos genótipos. Os novos transcritos identificados também serão importantes para estudar a diferença entre espécies do gênero *Saccharum* e demais gramíneas, e podem ser a resposta da característica única de cana-de-açúcar poder acumular altas concentrações de sacarose no colmo.

7. REFERÊNCIAS

- Afzal AJ, Wood AJ, Lightfoot DA (2008). Plant Receptor-Like Serine Threonine Kinases: Roles in Signaling and Plant Defense. *Molecular Plant-Microbe Interactions* 21(5): 507-517.
- Agre P, Bonhivers M, Borgnia MJ (1998). The Aquaporins, Blueprints for Cellular Plumbing Systems. *Journal of Biological Chemistry* 273(24): 14659-14662.
- Ainsworth EA & Bush DR (2011). Carbohydrate export from the leaf: a highly regulated process and target to enhance photosynthesis and productivity. *Plant Physiol* 155(1): 64-69.
- Aitken KS, Jackson PA, McIntyre CL (2007). Construction of a genetic linkage map for *Saccharum officinarum* incorporating both simplex and duplex markers to increase genome coverage. *Genome* 50(8): 742-756.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* 215(3): 403-410.
- Alves MS, Dadalto SP, Goncalves AB, De Souza GB, Barros VA, Fietto LG (2013). Plant bZIP Transcription Factors Responsive to Pathogens: A Review. *Int J Mol Sci* 14(4): 7815-7828.
- Ambavaram MM, Krishnan A, Trijatmiko KR, Pereira A (2011). Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice. *Plant Physiol* 155(2): 916-931.
- Amorim HV, Lopes ML, de Castro Oliveira JV, Buckeridge MS, Goldman GH (2011). Scientific challenges of bioethanol production in Brazil. *Appl Microbiol Biotechnol* 91(5): 1267-1275.
- Amthor JS (2003). Efficiency of Lignin Biosynthesis: a Quantitative Analysis. *Annals of Botany* 91(6): 673-695.
- Anderson JP, Badruzsaufari E, Schenk PM, Manners JM, Desmond OJ, Ehlert C, Maclean DJ, Ebert PR, Kazan K (2004). Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in *Arabidopsis*. *Plant Cell* 16(12): 3460-3479.
- Apuya N (2012). Transgenic plants having enhanced biomass composition. United States, Ceres, Inc.
- Arnaud D, Desclos-Theveniau M, Zimmerli L (2012). Disease resistance to *Pectobacterium carotovorum* is negatively modulated by the *Arabidopsis* Lectin Receptor Kinase LecRK-V.5. *Plant Signal Behav* 7(9): 1070-1072.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25(1): 25-29.
- Aspeborg H, Schrader J, Coutinho PM, Stam M, Kallas A, Djerbi S, Nilsson P, Denman S, Amini B, Sterky F, Master E, Sandberg G, Mellerowicz E, Sundberg B, Henrissat B, Teeri TT (2005). Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen. *Plant Physiol* 137(3): 983-997.
- Baginsky S, Hennig L, Zimmermann P, Gruissem W (2010). Gene expression analysis, proteomics, and network discovery. *Plant Physiol* 152(2): 402-410.
- Balasubramanian V, Vashisht D, Cletus J, Sakthivel N (2012). Plant beta-1,3-glucanases: their biological functions and transgenic expression against phytopathogenic fungi. *Biotechnol Lett* 34(11): 1983-1990.
- Barboza AF, S; Souza, GM; Carrer, H (2012). Antisense of Caffeic Acid 3-O-Methyltransferase (COMT) Decreases Lignin Content in Transgenic Sugarcane (*Saccharum* spp.). International Symposium on C4 and CAM Plant Biology, Urbana-Champaign, Illinois, USA.
- Begcy K, Mariano ED, Gentile A, Lembke CG, Zingaretti SM, Souza GM, Menossi M (2012). A novel stress-induced sugarcane gene confers tolerance to drought, salt and oxidative stress in transgenic tobacco plants. *PLoS One* 7(9): e44697.
- Bell GDM, Kane NC, Rieseberg LH, Adams KL (2013). RNA-Seq Analysis of Allele-Specific Expression, Hybrid Effects, and Regulatory Divergence in Hybrids Compared with Their Parents from Natural Populations. *Genome Biology and Evolution* 5(7): 1309-1323.

- Ben-Ari G (2012). The ABA signal transduction mechanism in commercial crops: learning from Arabidopsis. *Plant Cell Rep* 31(8): 1357-1369.
- Benfey PN & Mitchell-Olds T (2008). From genotype to phenotype: systems biology meets natural variation. *Science* 320(5875): 495-497.
- Benkova E, Michniewicz M, Sauer M, Teichmann T, Seifertova D, Jurgens G, Friml J (2003). Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell* 115(5): 591-602.
- Berger SL (2007). The complex language of chromatin regulation during transcription. *Nature* 447(7143): 407-412.
- Besseau S, Hoffmann L, Geoffroy P, Lapiere C, Pollet B, Legrand M (2007). Flavonoid accumulation in Arabidopsis repressed in lignin synthesis affects auxin transport and plant growth. *Plant Cell* 19(1): 148-162.
- Bischoff V, Nita S, Neumetzler L, Schindelasch D, Urbain A, Eshed R, Persson S, Delmer D, Scheible WR (2010). TRICHOME BIREFRINGENCE and Its Homolog AT5G01360 Encode Plant-Specific DUF231 Proteins Required for Cellulose Biosynthesis in Arabidopsis. *Plant Physiology* 153(2): 590-602.
- Boerjan W, Ralph J, Baucher M (2003). Lignin biosynthesis. *Annu Rev Plant Biol* 54: 519-546.
- Bonawitz ND & Chapple C (2010). The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu Rev Genet* 44: 337-363.
- Bosch M, Mayer CD, Cookson A, Donnison IS (2011). Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and non-elongating maize internodes. *J Exp Bot* 62(10): 3545-3561.
- Bozell JJ, Black SK, Myers M, Cahill D, Miller WP, Park S (2011). Solvent fractionation of renewable woody feedstocks: Organosolv generation of biorefinery process streams for the production of biobased chemicals. *Biomass and Bioenergy* 35(10): 4197-4208.
- Brandão AD (2009). Efeito da Giberelina 'A3' e do paclobutrazol no metabolismo de carboidratos e expressão genica da cana-de-acúcar (*Saccharum* sp.). Instituto de Biologia. Campinas, Universidade Estadual de Campinas. Doutor em Biologia Vegetal.
- Bratzel F, Lopez-Torrejon G, Koch M, Del Pozo JC, Calonje M (2010). Keeping cell identity in Arabidopsis requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination. *Curr Biol* 20(20): 1853-1859.
- Bricker TM, Roose JL, Fagerlund RD, Frankel LK, Eaton-Rye JJ (2012). The extrinsic proteins of Photosystem II. *Biochim Biophys Acta* 1817(1): 121-142.
- Britto-Kido SdA, #xe3, Ferreira Neto J, #xe9, Costa R, Pandolfi V, Marcelino-Guimar, es FC, #xea, Nepomuceno AL, Vilela Abdelnoor R, Benko-Iseppon AM, Kido EA (2013). Natural Antisense Transcripts in Plants: A Review and Identification in Soybean Infected with *Phakopsora pachyrhizi* SuperSAGE Library. *The Scientific World Journal* 2013: 14.
- Brown D, Wightman R, Zhang Z, Gomez LD, Atanassov I, Bukowski J-P, Tryfona T, McQueen-Mason SJ, Dupree P, Turner S (2011). Arabidopsis genes IRREGULAR XYLEM (IRX15) and IRX15L encode DUF579-containing proteins that are essential for normal xylan deposition in the secondary cell wall. *The Plant Journal* 66(3): 401-413.
- Buckeridge MS (2010). Seed cell wall storage polysaccharides: models to understand cell wall biosynthesis and degradation. *Plant Physiol* 154(3): 1017-1023.
- Buckeridge MS, Dos Santos WD, De Souza AP (2010). Routes for cellulosic ethanol in Brazil In: sugarcane bioethanol: R&D for productivity and sustainability. Sugarcane bioethanol: R&D for productivity and sustainability. L. Cortez. São Paulo, Edgard Blucher: 365–380.
- Bundock PC, Casu RE, Henry RJ (2012). Enrichment of genomic DNA for polymorphism detection in a non-model highly polyploid crop plant. *Plant Biotechnol J* 10(6): 657-667.
- Burr SJ & Fry SC (2009). Extracellular cross-linking of maize arabinoxylans by oxidation of feruloyl esters to form oligoferuloyl esters and ether-like bonds. *Plant J* 58(4): 554-567.
- Byrt CS, Grof CP, Furbank RT (2011). C4 plants as biofuel feedstocks: optimising biomass production and feedstock quality from a lignocellulosic perspective. *J Integr Plant Biol* 53(2): 120-135.

- Calvo-Flores FG & Dobado JA (2010). Lignin as renewable raw material. *ChemSusChem* 3(11): 1227-1235.
- Carnavale Bottino M, Rosario S, Grativol C, Thiebaut F, Rojas CA, Farrineli L, Hemerly AS, Ferreira PCG (2013). High-Throughput Sequencing of Small RNA Transcriptome Reveals Salt Stress Regulated MicroRNAs in Sugarcane. *PLoS One* 8(3): e59423.
- Carpita NC (1996). Structure and Biogenesis of the Cell Walls of Grasses. *Annu Rev Plant Physiol Plant Mol Biol* 47: 445-476.
- Carpita NC (2012). Progress in the biological synthesis of the plant cell wall: new ideas for improving biomass for bioenergy. *Current Opinion in Biotechnology* 23(3): 330-337.
- Carpita NC & Gibeaut DM (1993). Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J* 3(1): 1-30.
- Carpita NC & McCann MC (2008). Maize and sorghum: genetic resources for bioenergy grasses. *Trends Plant Sci* 13(8): 415-420.
- Castillo-Davis CI & Hartl DL (2003). GeneMerge--post-genomic analysis, data mining, and hypothesis testing. *Bioinformatics* 19(7): 891-892.
- Casu R, Hotta CT, Souza GM (2010). Functional Genomics: Transcriptomics of Sugarcane - Current Status and Future Prospects. *Genetics, Genomics and Breeding of Sugarcane*. R. Henry & C. Cole, Science Publishers: 167-191.
- Casu RE, Grof CP, Rae AL, McIntyre CL, Dimmock CM, Manners JM (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(2): 371-386.
- Casu RE, Jarmey JM, Bonnett GD, Manners JM (2007). Identification of transcripts associated with cell wall metabolism and development in the stem of sugarcane by Affymetrix GeneChip Sugarcane Genome Array expression profiling. *Funct Integr Genomics* 7(2): 153-167.
- Chandel AK, Narasu ML, Chandrasekhar G, Manikyam A, Rao LV (2009). Use of *Saccharum spontaneum* (wild sugarcane) as biomaterial for cell immobilization and modulated ethanol production by thermotolerant *Saccharomyces cerevisiae* VS3. *Bioresour Technol* 100(8): 2404-2410.
- Chang D, Yang FY, Yan JJ, Wu YQ, Bai SQ, Liang XZ, Zhang YW, Gan YM (2012). SRAP analysis of genetic diversity of nine native populations of wild sugarcane, *Saccharum spontaneum*, from Sichuan, China. *Genet Mol Res* 11(2): 1245-1253.
- Chapple CC, Vogt T, Ellis BE, Somerville CR (1992). An *Arabidopsis* mutant defective in the general phenylpropanoid pathway. *Plant Cell* 4(11): 1413-1424.
- Chaudhary G, Singh LK, Ghosh S (2012). Alkaline pretreatment methods followed by acid hydrolysis of *Saccharum spontaneum* for bioethanol production. *Bioresour Technol* 124(0): 111-118.
- Chen D, Yuan C, Zhang J, Zhang Z, Bai L, Meng Y, Chen L-L, Chen M (2012). PlantNATsDB: a comprehensive database of plant natural antisense transcripts. *Nucleic Acids Res* 40(D1): D1187-D1193.
- Chen W, Yin X, Wang L, Tian J, Yang R, Liu D, Yu Z, Ma N, Gao J (2013). Involvement of rose aquaporin RhPIP1;1 in ethylene-regulated petal expansion through interaction with RhPIP2;1. *Plant Mol Biol*.
- Chen ZJ & Tian L (2007). Roles of dynamic and reversible histone acetylation in plant development and polyploidy. *Biochim Biophys Acta* 1769(5-6): 295-307.
- Chevalier D, Batoux M, Fulton L, Pfister K, Yadav RK, Schellenberg M, Schneitz K (2005). STRUBBELIG defines a receptor kinase-mediated signaling pathway regulating organ development in *Arabidopsis*. *Proc Natl Acad Sci U S A* 102(25): 9074-9079.
- Chiniquy D, Sharma V, Schultink A, Baidoo EE, Rautengarten C, Cheng K, Carroll A, Ulvskov P, Harholt J, Keasling JD, Pauly M, Scheller HV, Ronald PC (2012). XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. *Proceedings of the National Academy of Sciences* 109(42): 17117-17122.

- Christiansen MW, Holm PB, Gregersen PL (2011). Characterization of barley (*Hordeum vulgare* L.) NAC transcription factors suggests conserved functions compared to both monocots and dicots. *BMC Res Notes* 4: 302.
- Cocuron JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K, Raikhel N, Wilkerson CG (2007). A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. *Proc Natl Acad Sci U S A* 104(20): 8550-8555.
- Coleman HD, Park JY, Nair R, Chapple C, Mansfield SD (2008a). RNAi-mediated suppression of p-coumaroyl-CoA 3'-hydroxylase in hybrid poplar impacts lignin deposition and soluble secondary metabolism. *Proc Natl Acad Sci U S A* 105(11): 4501-4506.
- Coleman HD, Samuels AL, Guy RD, Mansfield SD (2008b). Perturbed lignification impacts tree growth in hybrid poplar--a function of sink strength, vascular integrity, and photosynthetic assimilation. *Plant Physiol* 148(3): 1229-1237.
- Conab (2013). Acompanhamento de safra brasileira: cana-de-açúcar, Companhia Nacional de Abastecimento.
- Cosgrove DJ, Li, L. C., Cho, H. T., Hoffmann-Benning, S., & Moore RC, et al. (2002). The growing world of expansins. *Plant Cell Physiol*. 43: 1436-1444.
- Costa MDL (2012). Estudo do transcriptoma associado ao déficit hídrico e desenvolvimento de imunoprecipitação de cromatina em cana de açúcar para estudos de redes regulatórias transcricionais. Instituto de Química. São Paulo, Universidade de São Paulo. Doutor.
- D'Hont A (2005). Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenet Genome Res* 109(1-3): 27-33.
- D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann JC (1996). Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Mol Gen Genet* 250(4): 405-413.
- D'Hont A, Souza GM, Menossi M, Vincentz M, Van Sluys MA, Glaszmann JC, Ulian EC (2008). Sugarcane: a Major Source of Sweetness, Alcohol, and Bio-energy. *Genomics of Tropical Crop Plants*. P. Moore, Ming, R. New York, Springer: 483-513.
- D'Hont A, Grivet, L., Feldmann, P., Rao, P. S., Berding, N. (1996). Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp) by molecular cytogenetics. *Mol Gen Genet* 250: 405-413.
- Dai N, Schaffer A, Petreikov M, Shahak Y, Giller Y, Ratner K, Levine A, Granot D (1999). Overexpression of *Arabidopsis* hexokinase in tomato plants inhibits growth, reduces photosynthesis, and induces rapid senescence. *Plant Cell* 11(7): 1253-1266.
- Dal-Bianco M, Carneiro MS, Hotta CT, Chapola RG, Hoffmann HP, Garcia AA, Souza GM (2012). Sugarcane improvement: how far can we go? *Curr Opin Biotechnol* 23(2): 265-270.
- Daviere JM & Achard P (2013). Gibberellin signaling in plants. *Development* 140(6): 1147-1151.
- Davin LB & Lewis NG (2005). Lignin primary structures and dirigent sites. *Curr Opin Biotechnol* 16(4): 407-415.
- Dayan J, Voronin N, Gong F, Sun TP, Hedden P, Fromm H, Aloni R (2012). Leaf-induced gibberellin signaling is essential for internode elongation, cambial activity, and fiber differentiation in tobacco stems. *Plant Cell* 24(1): 66-79.
- de Siqueira Ferreira S, Nishiyama M, Paterson A, Souza G (2013). Biofuel and energy crops: high-yield *Saccharinae* take center stage in the post-genomics era. *Genome Biology* 14(6): 210.
- de Souza AP, Arundale RA, Dohleman FG, Long SP, Buckeridge MS (2013). Will the exceptional productivity of *Miscanthus x giganteus* increase further under rising atmospheric CO₂? *Agricultural and Forest Meteorology* 171-172(0): 82-92.
- De Souza AP, Gaspar M, Da Silva EA, Ulian EC, Waclawovsky AJ, Nishiyama MY, Jr., Dos Santos RV, Teixeira MM, Souza GM, Buckeridge MS (2008). Elevated CO₂ increases photosynthesis, biomass and productivity, and modifies gene expression in sugarcane. *Plant Cell Environ* 31(8): 1116-1127.
- Ding SY & Himmel ME (2006). The maize primary cell wall microfibril: a new model derived from direct visualization. *J Agric Food Chem* 54(3): 597-606.

- Dohleman FG & Long SP (2009). More productive than maize in the Midwest: How does *Miscanthus* do it? *Plant Physiol* 150(4): 2104-2115.
- Domingues DS, Cruz GM, Metcalfe CJ, Nogueira FT, Vicentini R, Alves Cde S, Van Sluys MA (2012). Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. *BMC Genomics* 13: 137.
- Donaldson LA (2001). Lignification and lignin topochemistry - an ultrastructural view. *Phytochemistry* 57(6): 859-873.
- Eklof JM & Brumer H (2010). The XTH gene family: an update on enzyme structure, function, and phylogeny in xyloglucan remodeling. *Plant Physiol* 153(2): 456-466.
- Erb M, Meldau S, Howe GA (2012). Role of phytohormones in insect-specific plant reactions. *Trends Plant Sci* 17(5): 250-259.
- Eyuboglu B, Pfister K, Haberer G, Chevalier D, Fuchs A, Mayer K, Schneitz K (2007). Molecular characterisation of the STRUBBELIG-RECEPTOR FAMILY of genes encoding putative leucine-rich repeat receptor-like kinases in *Arabidopsis thaliana*. *BMC Plant Biol* 7(1): 16.
- Faostat (2011). Food and Agricultural commodities production, Food and Agriculture Organization of the United Nations.
- Felix J, Papini-Terzi F, Rocha F, Vêncio R, Vicentini R, Nishiyama M, Jr., César Ulian E, Souza G, Menossi M (2009). Expression Profile of Signal Transduction Components in a Sugarcane Population Segregating for Sugar Content. *Tropical Plant Biology* 2(2): 98-109.
- Ferreira TH, Gentile A, Vilela RD, Costa GG, Dias LI, Endres L, Menossi M (2012). microRNAs associated with drought response in the bioenergy crop sugarcane (*Saccharum* spp.). *PLoS One* 7(10): e46703.
- Figueira TR, Okura V, Rodrigues da Silva F, Jose da Silva M, Kudrna D, Ammiraju JS, Talag J, Wing R, Arruda P (2012). A BAC library of the SP80-3280 sugarcane variety (*saccharum* sp.) and its inferred microsynteny with the sorghum genome. *BMC Res Notes* 5: 185.
- Fornale S, Capellades M, Encina A, Wang K, Irar S, Lapierre C, Ruel K, Joseleau JP, Berenguer J, Puigdomenech P, Rigau J, Caparros-Ruiz D (2012). Altered lignin biosynthesis improves cellulosic bioethanol production in transgenic maize plants down-regulated for cinnamyl alcohol dehydrogenase. *Mol Plant* 5(4): 817-830.
- Fornale S, Shi X, Chai C, Encina A, Irar S, Capellades M, Fuguet E, Torres JL, Rovira P, Puigdomenech P, Rigau J, Grotewold E, Gray J, Caparros-Ruiz D (2010). ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. *Plant J* 64(4): 633-644.
- Fornale S, Sonbol FM, Maes T, Capellades M, Puigdomenech P, Rigau J, Caparros-Ruiz D (2006). Down-regulation of the maize and *Arabidopsis thaliana* caffeic acid O-methyl-transferase genes by two new maize R2R3-MYB transcription factors. *Plant Mol Biol* 62(6): 809-823.
- Fuchs S, Grill E, Meskiene I, Schweighofer A (2013). Type 2C protein phosphatases in plants. *Febs J* 280(2): 681-693.
- Fulton L, Batoux M, Vaddepalli P, Yadav RK, Busch W, Andersen SU, Jeong S, Lohmann JU, Schneitz K (2009). DETORQUEO, QUIRKY, and ZERZAUST Represent Novel Components Involved in Organ Development Mediated by the Receptor-Like Kinase STRUBBELIG in *Arabidopsis thaliana*. *PLoS Genet* 5(1): e1000355.
- Gang DR, Costa MA, Fujita M, Dinkova-Kostova AT, Wang HB, Burlat V, Martin W, Sarkanen S, Davin LB, Lewis NG (1999). Regiochemical control of monolignol radical coupling: a new paradigm for lignin and lignan biosynthesis. *Chem Biol* 6(3): 143-151.
- Geigenberger P & Stitt M (1993). Sucrose synthase catalyses a readily reversible reaction in vivo in developing potato tubers and other plant tissues. *Planta* 189(3): 329-339.
- Gentile A, Ferreira TH, Mattos RS, Dias LI, Hoshino AA, Carneiro MS, Souza GM, Calsa T, Jr., Nogueira RM, Endres L, Menossi M (2013). Effects of drought on the microtranscriptome of field-grown sugarcane plants. *Planta* 237(3): 783-798.
- Goldemberg J (2008). The Brazilian biofuels industry. *Biotechnol Biofuels* 1(1): 6.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C,

- Lindblad-Toh K, Friedman N, Regev A (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7): 644-652.
- Grant CE, Bailey TL, Noble WS (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27(7): 1017-1018.
- Gray J, Caparros-Ruiz D, Grotewold E (2012). Grass phenylpropanoids: regulate before using! *Plant Sci* 184: 112-120.
- Grierson D (2013). Ethylene and the Control of Fruit Ripening. *The Molecular Biology and Biochemistry of Fruit Ripening*, Blackwell Publishing Ltd.: 43-73.
- Guillaumie S, San-Clemente H, Deswarte C, Martinez Y, Lapierre C, Murigneux A, Barriere Y, Pichon M, Goffner D (2007). MAIZEWALL. Database and developmental gene expression profiling of cell wall biosynthesis and assembly in maize. *Plant Physiol* 143(1): 339-363.
- Ha S, Moore PH, Heinz D, Kato S, Ohmido N, Fukui K (1999). Quantitative chromosome map of the polyploid *Saccharum spontaneum* by multicolor fluorescence in situ hybridization and imaging methods. *Plant Mol Biol* 39(6): 1165-1173.
- Halford NG & Hey SJ (2009). Snf1-related protein kinases (SnRKs) act within an intricate network that links metabolic and stress signalling in plants. *Biochem J* 419(2): 247-259.
- Hall T (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/08/NT. *Nucleic Acids Symp Ser* 41: 95-98.
- Handakumbura PP & Hazen SP (2012). Transcriptional Regulation of Grass Secondary Cell Wall Biosynthesis: Playing Catch-Up with *Arabidopsis thaliana*. *Front Plant Sci* 3: 74.
- Hatton D, Sablowski R, Yung MH, Smith C, Schuch W, Bevan M (1995). Two classes of cis sequences contribute to tissue-specific expression of a PAL2 promoter in transgenic tobacco. *Plant J* 7(6): 859-876.
- Heinen RB, Ye Q, Chaumont F (2009). Role of aquaporins in leaf physiology. *J Exp Bot* 60(11): 2971-2985.
- Held MA, Penning B, Brandt AS, Kessans SA, Yong W, Scofield SR, Carpita NC (2008). Small-interfering RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell wall biosynthesis in barley. *Proceedings of the National Academy of Sciences* 105(51): 20534-20539.
- Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 8(2): R19.
- Hill J (2006). From the Cover: Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proceedings of the National Academy of Sciences* 103(30): 11206-11210.
- Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315(5813): 804-807.
- Holland N, Holland D, Helentjaris T, Dhugga KS, Xoconostle-Cazares B, Delmer DP (2000). A comparative analysis of the plant cellulose synthase (CesA) gene family. *Plant Physiol* 123(4): 1313-1324.
- Horta CGL (2013). Identificação de genes e promotores relacionados ao estresse de seca em cana-de-açúcar e obtenção de plantas transgênicas. Instituto de Química. São Paulo, Universidade de São Paulo. Doutor.
- Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* 2008: 420747.
- Hu WJ, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD, Tsai CJ, Chiang VL (1999). Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nat Biotechnol* 17(8): 808-812.

- Hu Z, Han Z, Song N, Chai L, Yao Y, Peng H, Ni Z, Sun Q (2013). Epigenetic modification contributes to the expression divergence of three TaEXPA1 homoeologs in hexaploid wheat (*Triticum aestivum*). *New Phytol* 197(4): 1344-1352.
- Hwa K-Y, Subramani B, Khoo K-H (2012). Putative xylosyltransferase genes in *Trichomonas vaginalis*. *Soft Computing* 16(3): 381-391.
- Inman-Bamber NG, Jackson PA, Hewitt M (2011). Sucrose accumulation in sugarcane stalks does not limit photosynthesis and biomass production. *Crop and Pasture Science* 62(10): 848.
- Irvine JE (1975). Relations of Photosynthetic Rates and Leaf and Canopy Characters to Sugarcane Yield. *Crop Sci.* 15(5): 671-676.
- Irvine JE (1999). *Saccharum* species as horticultural classes. *Theoretical and Applied Genetics* 98(2): 186-194.
- Ishikawa C, Hatanaka T, Misoo S, Miyake C, Fukayama H (2011). Functional incorporation of sorghum small subunit increases the catalytic turnover rate of Rubisco in transgenic rice. *Plant Physiol* 156(3): 1603-1611.
- Iskandar HM, Simpson RS, Casu RE, Bonnett GD, MacLean DJ, Manners JM (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.
- Iuchi S, Kobayashi M, Taji T, Naramoto M, Seki M, Kato T, Tabata S, Kakubari Y, Yamaguchi-Shinozaki K, Shinozaki K (2001). Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in *Arabidopsis*. *The Plant Journal* 27(4): 325-333.
- Jacobsen KR, Fisher DG, Maretzki A, Moore PH (1992). Developmental-changes in the anatomy of the sugarcane stem in relation to phloem unloading and sucrose storage. *Botanica Acta* 105(1): 70-80.
- Jensen JK, Kim H, Cocuron J-C, Orlor R, Ralph J, Wilkerson CG (2011). The DUF579 domain containing proteins IRX15 and IRX15-L affect xylan synthesis in *Arabidopsis*. *The Plant Journal* 66(3): 387-400.
- Jenuwein T & Allis CD (2001). Translating the histone code. *Science* 293(5532): 1074-1080.
- Jiang S-Y, Ma Z, Vanitha J, Ramachandran S (2013). Genetic variation and expression diversity between grain and sweet sorghum lines. *BMC Genomics* 14(1): 18.
- Jiao W-B, Huang D, Xing F, Hu Y, Deng X-X, Xu Q, Chen L-L (2013). Genome-wide characterization and expression analysis of genetic variants in sweet orange. *The Plant Journal* 75(6): 954-964.
- Jung JH, Fouad WM, Vermerris W, Gallo M, Altpeter F (2012). RNAi suppression of lignin biosynthesis in sugarcane reduces recalcitrance for biofuel production from lignocellulosic biomass. *Plant Biotechnol J* 10(9): 1067-1076.
- Jung JH, Vermerris W, Gallo M, Fedenko JR, Erickson JE, Altpeter F (2013). RNA interference suppression of lignin biosynthesis increases fermentable sugar yields for biofuel production from field-grown sugarcane. *Plant Biotechnol J* 11(6): 709-716.
- Kaneda M, Rensing KH, Wong JC, Banno B, Mansfield SD, Samuels AL (2008). Tracking monolignols during wood development in lodgepole pine. *Plant Physiol* 147(4): 1750-1760.
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2011). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40(D1): D109-D114.
- Kannangara R, Branigan C, Liu Y, Penfield T, Rao V, Mouille G, Hofte H, Pauly M, Riechmann JL, Broun P (2007). The transcription factor WIN1/SHN1 regulates Cutin biosynthesis in *Arabidopsis thaliana*. *Plant Cell* 19(4): 1278-1294.
- Keating JD, Panganiban C, Mansfield SD (2006). Tolerance and adaptation of ethanologenic yeasts to lignocellulosic inhibitory compounds. *Biotechnol Bioeng* 93(6): 1196-1206.
- Kerr KF (2007). Extended analysis of benchmark datasets for Agilent two-color microarrays. *BMC Bioinformatics* 8(1): 371.
- Khan FA, Khan A, Azhar FM, Rauf S (2009). Genetic diversity of *Saccharum officinarum* accessions in Pakistan as revealed by random amplified polymorphic DNA. *Genet Mol Res* 8(4): 1376-1382.

- Kim C, Lee TH, Compton RO, Robertson JS, Pierce GJ, Paterson AH (2013). A genome-wide BAC end-sequence survey of sugarcane elucidates genome composition, and identifies BACs covering much of the euchromatin. *Plant Mol Biol* 81(1-2): 139-147.
- Kim WC, Ko JH, Han KH (2012). Identification of a cis-acting regulatory motif recognized by MYB46, a master transcriptional regulator of secondary wall biosynthesis. *Plant Mol Biol* 78(4-5): 489-501.
- Ko JH, Kim WC, Han KH (2009). Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in *Arabidopsis*. *Plant J* 60(4): 649-665.
- Kombrink E (2012). Chemical and genetic exploration of jasmonate biosynthesis and signaling paths. *Planta* 236(5): 1351-1366.
- Komor E (2000). The physiology of sucrose storage in sugarcane. *Developments in Crop Science*. G. Anil Kumar & K. Narinder, Elsevier. Volume 26: 35-53.
- Kong D, Karve R, Willet A, Chen MK, Oden J, Shpak ED (2012). Regulation of plasmodesmatal permeability and stomatal patterning by the glycosyltransferase-like protein KOBITO1. *Plant Physiol* 159(1): 156-168.
- Krapp A, Hofmann B, Schafer C, Stitt M (1993). Regulation of the expression of *rbcS* and other photosynthetic genes by carbohydrates - a mechanism for the sink regulation of photosynthesis. *Plant Journal* 3(6): 817-828.
- Krapp A & Stitt M (1995). An evaluation of direct and indirect mechanisms for the sink-regulation of photosynthesis in spinach - changes in gas-exchange, carbohydrates, metabolites, enzyme-activities and steady-state transcript levels after cold-girdling source leaves. *Planta* 195(3): 313-323.
- Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, Ito J, Mimura T, Fukuda H, Demura T (2005). Transcription switches for protoxylem and metaxylem vessel formation. *Genes & Development* 19(16): 1855-1860.
- Kumar A, Cameron JB, Flynn PC (2005). Pipeline transport and simultaneous saccharification of corn stover. *Bioresour Technol* 96(7): 819-829.
- Landell M & Bressiani J (2008). Melhoramento genético, caracterização e manejo varietal. Cana de açúcar. L. Dinardo-Miranda, A. Vasconcelos & M. Landell. Campinas, São Paulo, Instituto Agronômico.
- Lapidot M & Pilpel Y (2006). Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep* 7(12): 1216-1222.
- Leal MV, Walter A, Seabra JA (2013). Sugarcane as an energy source. *Biomass Conversion and Biorefinery* 3(1): 17-26.
- Lee C, Zhong R, Ye ZH (2011). *Arabidopsis* Family GT43 Members are Xylan Xylosyltransferases Required for the Elongation of the Xylan Backbone. *Plant and Cell Physiology* 53(1): 135-143.
- Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P (2004). Coexpression analysis of human genes across many microarray data sets. *Genome Res* 14(6): 1085-1094.
- Legay S, Sivadon P, Blervacq AS, Pavy N, Baghdady A, Tremblay L, Levasseur C, Ladouce N, Lapierre C, Seguin A, Hawkins S, Mackay J, Grima-Pettenati J (2010). EgMYB1, an R2R3 MYB transcription factor from eucalyptus negatively regulates secondary cell wall formation in *Arabidopsis* and poplar. *New Phytol* 188(3): 774-786.
- Lehninger AL, Nelson DL, Cox MM (2005). *Lehninger principles of biochemistry*. New York, W.H. Freeman.
- Lembke CG, Nishiyama MY, Jr., Sato PM, de Andrade RF, Souza GM (2012). Identification of sense and antisense transcripts regulated by drought in sugarcane. *Plant Mol Biol* 79(4-5): 461-477.
- Li B & Dewey C (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12(1): 323.
- Li E, Bhargava A, Qiang W, Friedmann MC, Forneris N, Savidge RA, Johnson LA, Mansfield SD, Ellis BE, Douglas CJ (2012). The Class II KNOX gene *KNAT7* negatively regulates secondary wall formation in *Arabidopsis* and is functionally conserved in *Populus*. *New Phytol* 194(1): 102-115.

- Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, Reidel EJ, Turgeon R, Liu P, Sun Q, Nelson T, Brutnell TP (2010). The developmental dynamics of the maize leaf transcriptome. *Nat Genet* 42(12): 1060-1067.
- Li Y, He Y, Zu Y, Zhan F (2011). Identification and cloning of molecular markers for UV-B tolerant gene in wild sugarcane (*Saccharum spontaneum* L.). *J Photochem Photobiol B* 105(2): 119-125.
- Liang M, Cowley AW, Greene AS (2004). High throughput gene expression profiling: a molecular approach to integrative physiology. *J Physiol* 554(Pt 1): 22-30.
- Lima DU, Santos HP, Tiné MA, Molle FD, Buckeridge MS (2001). Patterns of expression of cell wall related genes in sugarcane. *Genetics and Molecular Biology* 24(1-4): 191-198.
- Lionetti V, Cervone F, Bellincampi D (2012). Methyl esterification of pectin plays a role during plant-pathogen interactions and affects plant resistance to diseases. *J Plant Physiol* 169(16): 1623-1630.
- Lois R, Dietrich A, Hahlbrock K, Schulz W (1989). A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification of elicitor and light responsive cis-acting elements. *EMBO J* 8(6): 1641-1648.
- Loraine AE, McCormick S, Estrada A, Patel K, Qin P (2013). RNA-Seq of Arabidopsis Pollen Uncovers Novel Transcription and Alternative Splicing. *Plant Physiol* 162(2): 1092-1109.
- Lovegrove A, Wilkinson MD, Freeman J, Pellny TK, Tosi P, Saulnier L, Shewry PR, Mitchell RAC (2013). RNA Interference Suppression of Genes in Glycosyl Transferase Families 43 and 47 in Wheat Starchy Endosperm Causes Large Decreases in Arabinoxylan Content. *Plant Physiology* 163(1): 95-107.
- Luo M, Liu X, Singh P, Cui Y, Zimmerli L, Wu K (2012). Chromatin modifications and remodeling in plant abiotic stress responses. *Biochim Biophys Acta* 1819(2): 129-136.
- Ma QH, Wang C, Zhu HH (2011). TaMYB4 cloned from wheat regulates lignin biosynthesis through negatively controlling the transcripts of both cinnamyl alcohol dehydrogenase and cinnamoyl-CoA reductase genes. *Biochimie* 93(7): 1179-1186.
- Macedo IC, Seabra JEA, Silva JEAR (2008). Green house gases emissions in the production and use of ethanol from sugarcane in Brazil: The 2005/2006 averages and a prediction for 2020. *Biomass and Bioenergy* 32(7): 582-595.
- Magistri M, Faghihi MA, St Laurent Iii G, Wahlestedt C (2012). Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. *Trends in Genetics* 28(8): 389-396.
- Magrane M & Consortium U (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database* 2011(0): bar009-bar009.
- Makarevich G, Leroy O, Akinci U, Schubert D, Clarenz O, Goodrich J, Grossniklaus U, Kohler C (2006). Different Polycomb group complexes regulate common target genes in Arabidopsis. *EMBO Rep* 7(9): 947-952.
- Manners J & Casu R (2011). Transcriptome Analysis and Functional Genomics of Sugarcane. *Tropical Plant Biology* 4(1): 9-21.
- Marita JM, Vermerris W, Ralph J, Hatfield RD (2003). Variations in the cell wall composition of maize brown midrib mutants. *J Agric Food Chem* 51(5): 1313-1321.
- McCarthy RL, Zhong R, Ye ZH (2009). MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in Arabidopsis. *Plant Cell Physiol* 50(11): 1950-1964.
- McCormick AJ, Cramer MD, Watt DA (2006). Sink strength regulates photosynthesis in sugarcane. *New Phytol* 171(4): 759-770.
- McCormick AJ, Cramer MD, Watt DA (2008a). Changes in photosynthetic rates and gene expression of leaves during a source-sink perturbation in sugarcane. *Ann Bot* 101(1): 89-102.
- McCormick AJ, Cramer MD, Watt DA (2008b). Regulation of photosynthesis by sugars in sugarcane leaves. *J Plant Physiol* 165(17): 1817-1829.
- McGrath KC, Dombrecht B, Manners JM, Schenk PM, Edgar CI, Maclean DJ, Scheible WR, Udvardi MK, Kazan K (2005). Repressor- and activator-type ethylene response factors functioning in

- jasmonate signaling and disease resistance identified via a genome-wide screen of *Arabidopsis* transcription factor gene expression. *Plant Physiol* 139(2): 949-959.
- McQueen-Mason S & Cosgrove DJ (1994). Disruption of hydrogen bonding between plant cell wall polymers by proteins that induce wall extension. *Proc. Natl. Acad. Sci. USA* 91: 6574-6578.
- Meyer E, Logan TL, Juenger TE (2012). Transcriptome analysis and gene expression atlas for *Panicum hallii* var. *filipes*, a diploid model for biofuel research. *Plant J* 70(5): 879-890.
- Ming R, Moore PH, Wu K, D'Hont A, Glaszmann JC, Tew TL, Mirkov E, Silva J, Jifon J, Rai M, Schnell RJ, Brumbley SM, Lakshmanan P, Comstock JC, Paterson AH (2006). Sugarcane Improvement through Breeding and Biotechnology. *Plant Breeding Reviews*. J. Janick, 27.
- Ming R, Wang W, Draye X, Moore H, Irvine E, Paterson H (2002). Molecular dissection of complex traits in autopolyploids: mapping QTLs affecting sugar yield and related traits in sugarcane. *Theor Appl Genet* 105(2-3): 332-345.
- Mirzaei M, Pascovici D, Atwell BJ, Haynes PA (2012). Differential regulation of aquaporins, small GTPases and V-ATPases proteins in rice leaves subjected to drought stress and recovery. *Proteomics* 12(6): 864-877.
- Mitchell RA, Dupree P, Shewry PR (2007). A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. *Plant Physiol* 144(1): 43-53.
- Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M (2005). The NAC transcription factors NST1 and NST2 of *Arabidopsis* regulate secondary wall thickenings and are required for anther dehiscence. *Plant Cell* 17(11): 2993-3006.
- Mochida K, Yoshida T, Sakurai T, Ogihara Y, Shinozaki K (2009). TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. *Plant Physiol* 150(3): 1135-1146.
- Moes D, Gatti S, Hoffmann C, Dieterle M, Moreau F, Neumann K, Schumacher M, Diederich M, Grill E, Shen WH, Steinmetz A, Thomas C (2013). A LIM domain protein from tobacco involved in actin-bundling and histone gene transcription. *Mol Plant* 6(2): 483-502.
- Moore PH (1995). Temporal and spatial regulation of sucrose accumulation in the sugarcane stem. *Austr J Plant Physiol* 22: 661-679.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5(7): 621-628.
- Movahedi S, Van Bel M, Heyndrickx KS, Vandepoele K (2012). Comparative co-expression analysis in plant biology. *Plant, Cell & Environment* 35(10): 1787-1798.
- Muday GK, Rahman A, Binder BM (2012). Auxin and ethylene: collaborators or competitors? *Trends Plant Sci* 17(4): 181-195.
- Mueller SC & Brown RM, Jr. (1980). Evidence for an intramembrane component associated with a cellulose microfibril-synthesizing complex in higher plants. *J Cell Biol* 84(2): 315-326.
- Muller K, Levesque-Tremblay G, Bartels S, Weitbrecht K, Wormit A, Usadel B, Haughn G, Kermode AR (2013). Demethylesterification of cell wall pectins in *Arabidopsis* plays a role in seed germination. *Plant Physiol* 161(1): 305-316.
- Mutz KO, Heilkenbrinker A, Lonne M, Walter JG, Stahl F (2013). Transcriptome analysis using next-generation sequencing. *Curr Opin Biotechnol* 24(1): 22-30.
- Nagano Y, Furuhashi H, Inaba T, Sasaki Y (2001). A novel class of plant-specific zinc-dependent DNA-binding protein that binds to A/T-rich DNA sequences. *Nucleic Acids Res* 29(20): 4097-4105.
- Neufeld HS, Grantz DA, Meinzer FC, Goldstein G, Crisosto GM, Crisosto C (1992). Genotypic variability in vulnerability of leaf xylem to cavitation in water-stressed and well-irrigated sugarcane. *Plant Physiol* 100(2): 1020-1028.
- Nishiyama-Jr MY, Vicente F, Sato PM, Ferreira SS, Feltus FA, Souza GM (2013). Transcriptome Analysis in the Saccharinae. *Genomics of the Saccharinae*. A. Paterson, Springer. 11: 121-140.
- North GB & Peterson CA (2005). Water flow in roots: Structural and regulatory features. *Vascular Transport in Plants*. N. M. Holbrook & A. Z. Maciej. Burlington, Academic Press: 131-156.
- OECD (2008). Biofuel Support Policies: An Economic Assessment. Paris, OECD Publications.

- Ohashi-Ito K, Oda Y, Fukuda H (2010). Arabidopsis VASCULAR-RELATED NAC-DOMAIN6 directly regulates the genes that govern programmed cell death and secondary wall formation during xylem differentiation. *Plant Cell* 22(10): 3461-3473.
- Osato N, Yamada H, Satoh K, Ooka H, Yamamoto M, Suzuki K, Kawai J, Carninci P, Ohtomo Y, Murakami K, Matsubara K, Kikuchi S, Hayashizaki Y (2003). Antisense transcripts with rice full-length cDNAs. *Genome Biol* 5(1): R5.
- Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, Roopra S, Frings O, Sonnhammer EL (2010). InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res* 38(Database issue): D196-203.
- Pagant S, Bichet A, Sugimoto K, Lerouxel O, Desprez T, McCann M, Lerouge P, Vernhettes S, Hofte H (2002). KOBITO1 encodes a novel plasma membrane protein necessary for normal synthesis of cellulose during cell expansion in Arabidopsis. *Plant Cell* 14(9): 2001-2013.
- Pan YB, Burner DM, Legendre BL (2000). An assessment of the phylogenetic relationship among sugarcane and related taxa based on the nucleotide sequence of 5S rRNA intergenic spacers. *Genetica* 108(3): 285-295.
- Papini-Terzi FS, Rocha FR, Vencio RZ, Felix JM, Branco DS, Waclawovsky AJ, Del Bem LE, Lembke CG, Costa MD, Nishiyama MY, Jr., Vicentini R, Vincentz MG, Ulian EC, Menossi M, Souza GM (2009). Sugarcane genes associated with sucrose content. *BMC Genomics* 10: 120.
- Papini-Terzi FS, Rocha FR, Vencio RZ, Oliveira KC, Felix Jde M, Vicentini R, Rocha Cde S, Simoes AC, Ulian EC, di Mauro SM, da Silva AM, Pereira CA, Menossi M, Souza GM (2005). Transcription profiling of signal transduction-related genes in sugarcane tissues. *DNA Res* 12(1): 27-38.
- Paterson AH, Moore PH, Tew TL (2013). The Gene Pool of Saccharum species and their Improvement. *Genomics of the Saccharinae*. A. Paterson, Springer. 11: 43-72.
- Patrick JW (1997). Phloem unloading: sieve element unloading and post sieve element transport. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48: 191-222.
- Patrick WJ, Botha CF, Birch GR (2013). Metabolic engineering of sugars and simple sugar derivatives in plants. *Plant Biotechnol J* 11(2): 142-156.
- Patten AM, Cardenas CL, Cochrane FC, Laskar DD, Bedgar DL, Davin LB, Lewis NG (2005). Reassessment of effects on lignification and vascular development in the *irx4* Arabidopsis mutant. *Phytochemistry* 66(17): 2092-2107.
- Penning BW, Hunter CT, 3rd, Tayengwa R, Eveland AL, Dugard CK, Olek AT, Vermerris W, Koch KE, McCarty DR, Davis MF, Thomas SR, McCann MC, Carpita NC (2009). Genetic resources for maize cell wall biology. *Plant Physiol* 151(4): 1703-1728.
- Perrot-Rechenmann C (2010). Cellular responses to auxin: division versus expansion. *Cold Spring Harb Perspect Biol* 2(5): a001446.
- Pesquet E, Ranocha P, Legay S, Digonnet C, Barbier O, Pichon M, Goffner D (2005). Novel markers of xylogenesis in zinnia are differentially regulated by auxin and cytokinin. *Plant Physiol* 139(4): 1821-1839.
- Pickett-Heaps JD (1968). Xylem Wall Deposition: Radioautographic Investigations Using Lignin Precursors. *Protoplasma* 65: 181-205.
- Pinheiro GL, Marques CS, Costa MD, Reis PA, Alves MS, Carvalho CM, Fietto LG, Fontes EP (2009). Complete inventory of soybean NAC transcription factors: sequence conservation and expression analysis uncover their distinct roles in stress response. *Gene* 444(1-2): 10-23.
- Piperidis G, Piperidis N, D'Hont A (2010). Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Mol Genet Genomics* 284(1): 65-73.
- Poelking VGC (2012). Caracterização anatômica, química e molecular e potencial bioenergético de dois genótipos de cana-de-açúcar. Departamento de Biologia Vegetal. Viçosa, Universidade Federal de Viçosa. Doutor.
- Pou A, Medrano H, Flexas J, Tyerman SD (2012). A putative role for TIP and PIP aquaporins in dynamics of leaf hydraulic and stomatal conductances in grapevine under water stress and re-watering. *Plant Cell Environ.*

- Preston J, Wheeler J, Heazlewood J, Li SF, Parish RW (2004). AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. *Plant J* 40(6): 979-995.
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD (2011). The Pfam protein families database. *Nucleic Acids Res* 40(D1): D290-D301.
- Purcell PC, Smith AM, Halford NG (1998). Antisense expression of a sucrose non-fermenting-1-related protein kinase sequence in potato results in decreased expression of sucrose synthase in tubers and loss of sucrose-inducibility of sucrose synthase transcripts in leaves. *Plant Journal* 14(2): 195-202.
- Quackenbush J (2003). Genomics. Microarrays--guilt by association. *Science* 302(5643): 240-241.
- Rae AL, Grof CPL, Casu RE, Bonnett GD (2005). Sucrose accumulation in the sugarcane stem: pathways and control points for transport and compartmentation. *Field Crops Research* 92(2-3): 159-168.
- Raes J, Rohde A, Christensen JH, Van de Peer Y, Boerjan W (2003). Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiol* 133(3): 1051-1071.
- Ralph J, Bunzel M, Marita J, Hatfield R, Lu F, Kim H, Schatz P, Grabber J, Steinhart H (2004a). Peroxidase-dependent cross-linking reactions of p-hydroxycinnamates in plant cell walls. *Phytochemistry Reviews* 3(1-2): 79-96.
- Ralph J, Lundquist K, Brunow G, Lu F, Kim H, Schatz P, Marita J, Hatfield RD, Christensen JH, Boerjan W (2004b). Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenylpropanoids. *Phytochemistry Reviews* 3(1-2): 29-60.
- Raveendran K & Ganesh A (1996). Heating value of biomass and biomass pyrolysis products. *Fuel* 75: 1715-1720.
- Raven PH, Evert RF, Eichhorn (1996). *Biologia Vegetal*, Guanabara Koogan.
- Ren D, Li Y, Zhao F, Sang X, Shi J, Wang N, Guo S, Ling Y, Zhang C, Yang Z, He G (2013). MULTI-FLORET SPIKELET1, which encodes an AP2/ERF protein, determines spikelet meristem fate and sterile lemma identity in rice. *Plant Physiol* 162(2): 872-884.
- Rencoret J, Gutierrez A, Nieto L, Jimenez-Barbero J, Faulds CB, Kim H, Ralph J, Martinez AT, Del Rio JC (2011). Lignin composition and structure in young versus adult *Eucalyptus globulus* plants. *Plant Physiol* 155(2): 667-682.
- Richter R, Behringer C, Zourelidou M, Schwechheimer C (2013). Convergence of auxin and gibberellin signaling on the regulation of the GATA transcription factors GNC and GNL in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* 110(32): 13192-13197.
- 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(2): 111-119.
- Roach BT (1989). Origin and improvement of the genetic base of sugarcane. *Proc Aust Soc Sug Cane Technol*: 34-47.
- Rocha FR, Papini-Terzi FS, Nishiyama MY, Jr., Vencio RZ, Vicentini R, Duarte RD, de Rosa VE, Jr., Vinagre F, Barsalobres C, Medeiros AH, Rodrigues FA, Ulian EC, Zingaretti SM, Galbiatti JA, Almeida RS, Figueira AV, Hemerly AS, Silva-Filho MC, Menossi M, Souza GM (2007). Signal transduction-related responses to phytohormones and environmental challenges in sugarcane. *BMC Genomics* 8: 71.
- Roitsch T & Gonzalez MC (2004). Function and regulation of plant invertases: sweet sensations. *Trends Plant Sci* 9(12): 606-613.
- Romano JM, Dubos C, Prouse MB, Wilkins O, Hong H, Poole M, Kang K-Y, Li E, Douglas CJ, Western TL, Mansfield SD, Campbell MM (2012). AtMYB61, an R2R3-MYB transcription factor, functions as a pleiotropic regulator via a small gene network. *New Phytologist* 195(4): 774-786.
- Rosche E, Blackmore D, Tegeder M, Richardson T, Schroeder H, Higgins TJ, Frommer WB, Offler CE, Patrick JW (2002). Seed-specific overexpression of a potato sucrose transporter increases sucrose uptake and growth rates of developing pea cotyledons. *Plant J* 30(2): 165-175.

- Rossi M, Araujo PG, de Jesus EM, Varani AM, Van Sluys MA (2004). Comparative analysis of Mutator - like transposases in sugarcane. *Mol Genet Genomics* 272(2): 194-203.
- Ruprecht C, Mutwil M, Saxe F, Eder M, Nikoloski Z, Persson S (2011). Large-scale co-expression approach to dissect secondary cell wall formation across plant species. *Front Plant Sci* 2: 23.
- Ruprecht C & Persson S (2012). Co-expression of cell-wall related genes: new tools and insights. *Front Plant Sci* 3: 83.
- Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34(2): 374-378.
- Saez A, Robert N, Maktabi MH, Schroeder JI, Serrano R, Rodriguez PL (2006). Enhancement of abscisic acid sensitivity and reduction of water consumption in *Arabidopsis* by combined inactivation of the protein phosphatases type 2C ABI1 and HAB1. *Plant Physiol* 141(4): 1389-1399.
- Sang T (2011). Toward the domestication of lignocellulosic energy crops: learning from food crop domestication. *J Integr Plant Biol* 53(2): 96-104.
- Santino A, Taurino M, De Domenico S, Bonsegna S, Poltronieri P, Pastor V, Flors V (2013). Jasmonate signaling in plant development and defense response to multiple (a)biotic stresses. *Plant Cell Rep* 32(7): 1085-1098.
- Santner A & Estelle M (2009). Recent advances and emerging trends in plant hormone signalling. *Nature* 459(7250): 1071-1078.
- Sato PM (2012). Regulação do Acúmulo de Sacarose em Cana-de-Açúcar e Análise Funcional de Uma Proteína Quinase Relacionada com o Conteúdo de Sacarose. Instituto de Química. São Paulo, Universidade de São Paulo. Doutor.
- Saxena IM & Brown-Jr RM (2005). Cellulose biosynthesis: current views and evolving concepts. *Ann Bot* 96(1): 9-21.
- Schena M, Heller RA, Theriault TP, Konrad K, Lachenmeier E, Davis RW (1998). Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol* 16(7): 301-306.
- Schneitz K, Hulskamp M, Kopczak SD, Pruitt RE (1997). Dissection of sexual organ ontogenesis: a genetic analysis of ovule development in *Arabidopsis thaliana*. *Development* 124(7): 1367-1376.
- Schweighofer A, Hirt H, Meskiene I (2004). Plant PP2C phosphatases: emerging functions in stress signaling. *Trends Plant Sci* 9(5): 236-243.
- Scordia D, Cosentino SL, Jeffries TW (2010). Second generation bioethanol production from *Saccharum spontaneum* L. ssp. *aegyptiacum* (Willd.) Hack. *Bioresour Technol* 101(14): 5358-5365.
- Searcy E, Flynn P, Ghafoori E, Kumar A (2007). The relative cost of biomass energy transport. *Appl Biochem Biotechnol* 137-140(1-12): 639-652.
- Sekhon RS, Briskine R, Hirsch CN, Myers CL, Springer NM, Buell CR, de Leon N, Kaeppler SM (2013). Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. *PLoS One* 8(4): e61005.
- Sekhon RS, Lin H, Childs KL, Hansey CN, Buell CR, de Leon N, Kaeppler SM (2011). Genome-wide atlas of transcription during maize development. *Plant J* 66(4): 553-563.
- Serrato AJ, de Dios Barajas-Lopez J, Chueca A, Sahrawy M (2009). Changing sugar partitioning in FBPase-manipulated plants. *J Exp Bot* 60(10): 2923-2931.
- Sewalt V, Ni W, Blount JW, Jung HG, Masoud SA, Howles PA, Lamb C, Dixon RA (1997). Reduced Lignin Content and Altered Lignin Composition in Transgenic Tobacco Down-Regulated in Expression of L-Phenylalanine Ammonia-Lyase or Cinnamate 4-Hydroxylase. *Plant Physiol* 115(1): 41-50.
- Shafizadeh F & Chin Peter P S (1977). Thermal Deterioration of Wood. *Wood Technology: Chemical Aspects*, AMERICAN CHEMICAL SOCIETY. 43: 57-81.

- Shani Z, Dekel M, Roiz L, Horowitz M, Kolosovski N, Lapidot S, Alkan S, Koltai H, Tsabary G, Goren R, Shoseyov O (2006). Expression of endo-1,4-beta-glucanase (cel1) in *Arabidopsis thaliana* is associated with plant growth, xylem development and cell wall thickening. *Plant Cell Rep* 25(10): 1067-1074.
- Sheen J (1999). C4 Gene Expression. *Annu Rev Plant Physiol Plant Mol Biol* 50: 187-217.
- Shen H, He X, Poovaiah CR, Wuddineh WA, Ma J, Mann DG, Wang H, Jackson L, Tang Y, Stewart CN, Jr., Chen F, Dixon RA (2012). Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. *New Phytol* 193(1): 121-136.
- Sherstnev A, Duc C, Cole C, Zacharaki V, Hornyik C, Oszolak F, Milos PM, Barton GJ, Simpson GG (2012). Direct sequencing of *Arabidopsis thaliana* RNA reveals patterns of cleavage and polyadenylation. *Nature Structural & Molecular Biology* 19(8): 845-852.
- Shibasaki K, Uemura M, Tsurumi S, Rahman A (2009). Auxin response in *Arabidopsis* under cold stress: underlying molecular mechanisms. *Plant Cell* 21(12): 3823-3838.
- Shiu SH & Bleecker AB (2001). Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc Natl Acad Sci U S A* 98(19): 10763-10768.
- Silva JA, Sorrells ME, Burnquist WL, Tanksley SD (1993). RFLP linkage map and genome analysis of *Saccharum spontaneum*. *Genome* 36(4): 782-791.
- Smith AM & Stitt M (2007). Coordination of carbon supply and plant growth. *Plant Cell Environ* 30(9): 1126-1149.
- Somerville C (2006). Cellulose synthesis in higher plants. *Annu Rev Cell Dev Biol* 22: 53-78.
- Somerville C, Youngs H, Taylor C, Davis SC, Long SP (2010). Feedstocks for lignocellulosic biofuels. *Science* 329(5993): 790-792.
- Sonbol FM, Fornale S, Capellades M, Encina A, Tourino S, Torres JL, Rovira P, Ruel K, Puigdomenech P, Rigau J, Caparros-Ruiz D (2009). The maize ZmMYB42 represses the phenylpropanoid pathway and affects the cell wall structure, composition and degradability in *Arabidopsis thaliana*. *Plant Mol Biol* 70(3): 283-296.
- Song CP, Agarwal M, Ohta M, Guo Y, Halfter U, Wang P, Zhu JK (2005). Role of an *Arabidopsis* AP2/EREBP-type transcriptional repressor in abscisic acid and drought stress responses. *Plant Cell* 17(8): 2384-2396.
- Song J & Wang Z (2011). RNAi-mediated suppression of the phenylalanine ammonia-lyase gene in *Salvia miltiorrhiza* causes abnormal phenotypes and a reduction in rosmarinic acid biosynthesis. *J Plant Res* 124(1): 183-192.
- Souza A, Leite DC, Pattathil S, Hahn M, Buckeridge M (2012). Composition and Structure of Sugarcane Cell Wall Polysaccharides: Implications for Second-Generation Bioethanol Production. *Bioenergy Research*: 1-16.
- Souza AP, Grandis A, Leite DCC, Buckeridge MS (2013). Sugarcane as a Bioenergy Source: History, Performance, and Perspectives for Second-Generation Bioethanol. *Bioenergy Research*.
- Souza GM, Simoes ACQ, Oliveira KC, Garay HM, Fiorini LC, Gomes FdS, Nishiyama-Junior MY, Silva AMd (2001). The sugarcane signal transduction (SUCAST) catalogue: prospecting signal transduction in sugarcane. *Genetics and Molecular Biology* 24(1-4): 25-34.
- Sreedharan S, Shekhawat UKS, Ganapathi TR (2013). Transgenic banana plants overexpressing a native plasma membrane aquaporin MusaPIP1;2 display high tolerance levels to different abiotic stresses. *Plant Biotechnol J*: n/a-n/a.
- Stitt M & Quick WP (1989). Photosynthetic carbon partitioning: its regulation and possibilities for manipulation. *Physiologia Plantarum* 77(4): 633-641.
- Stuart JM, Segal E, Koller D, Kim SK (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302(5643): 249-255.
- Sugden C, Donaghy PG, Halford NG, Hardie DG (1999). Two SNF1-related protein kinases from spinach leaf phosphorylate and inactivate 3-hydroxy-3-methylglutaryl-coenzyme A reductase, nitrate reductase, and sucrose phosphate synthase in vitro. *Plant Physiol* 120(1): 257-274.

- Tai PYP & Miller JD (2002). Germplasm diversity among four sugarcane species for sugar composition. *Crop science* 42: 958-964.
- Takahashi S, Furukawa T, Asano T, Terajima Y, Shimada H, Sugimoto A, Kadowaki K (2005). Very close relationship of the chloroplast genomes among *Saccharum* species. *Theor Appl Genet* 110(8): 1523-1529.
- Takken FL, Albrecht M, Tameling WI (2006). Resistance proteins: molecular switches of plant defence. *Curr Opin Plant Biol* 9(4): 383-390.
- Tanaka K, Murata K, Yamazaki M, Onosato K, Miyao A, Hirochika H (2003). Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. *Plant Physiol* 133(1): 73-83.
- ten Hove CA, Bochdanovits Z, Jansweijer VM, Koning FG, Berke L, Sanchez-Perez GF, Scheres B, Heidstra R (2011). Probing the roles of LRR RLK genes in *Arabidopsis thaliana* roots using a custom T-DNA insertion set. *Plant Mol Biol* 76(1-2): 69-83.
- Thelander M, Nilsson A, Olsson T, Johansson M, Girod PA, Schaefer DG, Zryd JP, Ronne H (2007). The moss genes PpSKI1 and PpSKI2 encode nuclear SnRK1 interacting proteins with homologues in vascular plants. *Plant Mol Biol* 64(5): 559-573.
- Thiebaut F, Grativol C, Carnavale-Bottino M, Rojas CA, Tanurdzic M, Farinelli L, Martienssen RA, Hemerly AS, Ferreira PC (2012). Computational identification and analysis of novel sugarcane microRNAs. *BMC Genomics* 13: 290.
- Udvardi MK, Kakar K, Wandrey M, Montanari O, Murray J, Andriankaja A, Zhang JY, Benedito V, Hofer JM, Chueng F, Town CD (2007). Legume transcription factors: global regulators of plant development and response to the environment. *Plant Physiol* 144(2): 538-549.
- Unica (2012). Números do setor sucroenergético brasileiro na safra 2010/2011, União da Indústria de Cana-de-açúcar.
- Vaddepalli P, Fulton L, Batoux M, Yadav RK, Schneitz K (2011). Structure-Function Analysis of STRUBBELIG, an *Arabidopsis* Atypical Receptor-Like Kinase Involved in Tissue Morphogenesis. *PLoS One* 6(5): e19730.
- Valdivia ER, Herrera MT, Gianzo C, Fidalgo J, Revilla G, Zarra I, Sampedro J (2013). Regulation of secondary wall synthesis and cell death by NAC transcription factors in the monocot *Brachypodium distachyon*. *J Exp Bot* 64(5): 1333-1343.
- van Heerden PD, Donaldson RA, Watt DA, Singels A (2010). Biomass accumulation in sugarcane: unravelling the factors underpinning reduced growth phenomena. *J Exp Bot* 61(11): 2877-2887.
- van Ooijen G, Mayr G, Kasiem MMA, Albrecht M, Cornelissen BJC, Takken FLW (2008). Structure-function analysis of the NB-ARC domain of plant disease resistance proteins. *Journal of Experimental Botany* 59(6): 1383-1397.
- Vanholme R, Morreel K, Ralph J, Boerjan W (2008). Lignin engineering. *Curr Opin Plant Biol* 11(3): 278-285.
- Vanstraelen M & Benkova E (2012). Hormonal interactions in the regulation of plant development. *Annu Rev Cell Dev Biol* 28: 463-487.
- Vastermark A, Almen MS, Simmen MW, Fredriksson R, Schioth HB (2011). Functional specialization in nucleotide sugar transporters occurred through differentiation of the gene cluster EamA (DUF6) before the radiation of Viridiplantae. *BMC Evol Biol* 11: 123.
- Vencio RZ & Koide T (2005). HTself: self-self based statistical test for low replication microarray studies. *DNA Res* 12(3): 211-214.
- Vermerris W & Saballos A (2013). Genetic Enhancement of Sorghum for Biomass Utilization. *Genomics of the Saccharinae*. A. H. Paterson, Springer New York. 11: 391-425.
- Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MI, Henrique-Silva F, Giglioti EA, Lemos MV, Coutinho LL, Nobrega MP, Carrer H, Franca SC, Bacci Junior M, Goldman MH, Gomes SL, Nunes LR, Camargo LE, Siqueira WJ, Van Sluys MA, Thiemann OH, Kuramae EE, Santelli RV, Marino CL, Targon ML, Ferro JA, Silveira HC, Marini DC, Lemos EG, Monteiro-Vitorello CB, Tambor JH, Carraro DM, Roberto PG, Martins VG, Goldman GH, de Oliveira RC,

- Truffi D, Colombo CA, Rossi M, de Araujo PG, Sculaccio SA, Angella A, Lima MM, de Rosa Junior VE, Siviero F, Coscrato VE, Machado MA, Grivet L, Di Mauro SM, Nobrega FG, Menck CF, Braga MD, Telles GP, Cara FA, Pedrosa G, Meidanis J, Arruda P (2003). Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Res* 13(12): 2725-2735.
- Vettore AL, Silva FRd, Kemper EL, Arruda P (2001). The libraries that made SUCEST. *Genetics and Molecular Biology* 24: 1-7.
- Wachholtz M, Heng-Moss T, Twigg P, Baird L, Lu G, Amundsen K (2013). Transcriptome analysis of two buffalograss cultivars. *BMC Genomics* 14(1): 613.
- Waclawovsky AJ, Sato PM, Lembke CG, Moore PH, Souza GM (2010). Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. *Plant Biotechnol J* 8(3): 263-276.
- Wang H, Avci U, Nakashima J, Hahn MG, Chen F, Dixon RA (2010a). Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. *Proceedings of the National Academy of Sciences* 107(51): 22338-22343.
- Wang H, Ma F, Cheng L (2010b). Metabolism of organic acids, nitrogen and amino acids in chlorotic leaves of 'Honeycrisp' apple (*Malus domestica* Borkh) with excessive accumulation of carbohydrates. *Planta* 232(2): 511-522.
- Wang H, Zhao Q, Chen F, Wang M, Dixon RA (2011). NAC domain function and transcriptional control of a secondary cell wall master switch. *The Plant Journal* 68(6): 1104-1114.
- Wang HZ & Dixon RA (2012). On-off switches for secondary cell wall biosynthesis. *Mol Plant* 5(2): 297-303.
- Wang J-C, Xu H, Zhu Y, Liu Q-Q, Cai X-L (2013a). OsbZIP58, a basic leucine zipper transcription factor, regulates starch biosynthesis in rice endosperm. *Journal of Experimental Botany*.
- Wang J, Kucukoglu M, Zhang L, Chen P, Decker D, Nilsson O, Jones B, Sandberg G, Zheng B (2013b). The Arabidopsis LRR-RLK, PXC1, is a regulator of secondary wall formation correlated with the TDIF-PXY/TDR-WOX4 signaling pathway. *BMC Plant Biol* 13: 94.
- Wang J, Roe B, Macmil S, Yu Q, Murray JE, Tang H, Chen C, Najjar F, Wiley G, Bowers J, Van Sluys MA, Rokhsar DS, Hudson ME, Moose SP, Paterson AH, Ming R (2010c). Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. *BMC Genomics* 11: 261.
- Wang L-P, Jackson PA, Lu X, Fan Y-H, Foreman JW, Chen X-K, Deng H-H, Fu C, Ma L, Aitken KS (2008). Evaluation of Sugarcane × Progeny for Biomass Composition and Yield Components. *Crop Science* 48(3): 951.
- Wang L, Guo K, Li Y, Tu Y, Hu H, Wang B, Cui X, Peng L (2010d). Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. *BMC Plant Biol* 10: 282.
- Wang S, Bai Y, Shen C, Wu Y, Zhang S, Jiang D, Guilfoyle TJ, Chen M, Qi Y (2010e). Auxin-related gene families in abiotic stress response in *Sorghum bicolor*. *Funct Integr Genomics* 10(4): 533-546.
- Wang S, Yin Y, Ma Q, Tang X, Hao D, Xu Y (2012). Genome-scale identification of cell-wall related genes in Arabidopsis based on co-expression network analysis. *BMC Plant Biol* 12(1): 138.
- Wei X, Jackson P, Stringer J, Cox M (2008). Relative economic genetic value (rEGV) – an improved selection index to replace net merit grade (NMG) in the Australian sugarcane variety improvement program. *Proceedings of the Australian Society of Sugar Cane Technologists*.
- Weichert N, Saalbach I, Weichert H, Kohl S, Erban A, Kopka J, Hause B, Varshney A, Sreenivasulu N, Strickert M, Kumlehn J, Weschke W, Weber H (2010). Increasing sucrose uptake capacity of wheat grains stimulates storage protein synthesis. *Plant Physiol* 152(2): 698-710.
- White R (1987a). Effect of lignin content and extractives on the higher heating value of Wood. *Wood Fiber Sci* 19: 446-452.
- White RH (1987b). Effect of lignin content and extractives on the higher heating value of wood. *Wood and Fiber Science* 19(4): 446-452.
- Willats WT, McCartney L, Mackie W, Knox JP (2001). Pectin: cell biology and prospects for functional analysis. *Plant Mol Biol* 47(1-2): 9-27.

- Willems E, Leyns L, Vandesompele J (2008). Standardization of real-time PCR gene expression data from independent biological replicates. *Anal Biochem* 379(1): 127-129.
- Wojtasik W, Kulma A, Dyminska L, Hanuza J, Zebrowski J, Szopa J (2013). Fibres from flax overproducing beta-1,3-glucanase show increased accumulation of pectin and phenolics and thus higher antioxidant capacity. *BMC Biotechnology* 13(1): 10.
- Wu A-M, Rihoey C, Seveno M, Hörnblad E, Singh SK, Matsunaga T, Ishii T, Lerouge P, Marchant A (2009). The Arabidopsis IRX10 and IRX10-LIKE glycosyltransferases are critical for glucuronoxylan biosynthesis during secondary cell wall formation. *The Plant Journal* 57(4): 718-731.
- Xie G, Yang B, Xu Z, Li F, Guo K, Zhang M, Wang L, Zou W, Wang Y, Peng L (2013). Global Identification of Multiple OsGH9 Family Members and Their Involvement in Cellulose Crystallinity Modification in Rice. *PLoS One* 8(1): e50171.
- Xu ZS, Chen M, Li LC, Ma YZ (2011). Functions and application of the AP2/ERF transcription factor family in crop improvement. *J Integr Plant Biol* 53(7): 570-585.
- Yamaguchi M, Mitsuda N, Ohtani M, Ohme-Takagi M, Kato K, Demura T (2011). VASCULAR-RELATED NAC-DOMAIN7 directly regulates the expression of a broad range of genes for xylem vessel formation. *Plant J* 66(4): 579-590.
- Yang F, Mitra P, Zhang L, Prak L, Verhertbruggen Y, Kim JS, Sun L, Zheng K, Tang K, Auer M, Scheller HV, Loque D (2013). Engineering secondary cell wall deposition in plants. *Plant Biotechnol J* 11(3): 325-335.
- Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 30(4): e15.
- Yazawa T, Kawahigashi H, Matsumoto T, Mizuno H (2013). Simultaneous transcriptome analysis of Sorghum and Bipolaris sorghicola by using RNA-seq in combination with de novo transcriptome assembly. *PLoS One* 8(4): e62460.
- Yilmaz A, Nishiyama-Jr MY, Fuentes BG, Souza GM, Janies D, Gray J, Grotewold E (2009). GRASSIUS: a platform for comparative regulatory genomics across the grasses. *Plant Physiol* 149(1): 171-180.
- Yin Y, Huang J, Xu Y (2009). The cellulose synthase superfamily in fully sequenced plants and algae. *BMC Plant Biol* 9: 99.
- Yong W, Link B, O'Malley R, Tewari J, Hunter CT, Lu CA, Li X, Bleecker AB, Koch KE, McCann MC, McCarty DR, Patterson SE, Reiter WD, Staiger C, Thomas SR, Vermerris W, Carpita NC (2005). Genomics of plant cell wall biogenesis. *Planta* 221(6): 747-751.
- Yuan Y, Teng Q, Zhong R, Ye ZH (2013). The Arabidopsis DUF231 Domain-Containing Protein ESK1 Mediates 2-O- and 3-O-Acetylation of Xylosyl Residues in Xylan. *Plant and Cell Physiology* 54(7): 1186-1199.
- Zanca AS, Vicentini R, Ortiz-Moreira FA, Del Bem LE, da Silva MJ, Vincentz M, Nogueira FT (2010). Identification and expression analysis of microRNAs and targets in the biofuel crop sugarcane. *BMC Plant Biol* 10: 260.
- Zarrineh P, Fierro AC, Sanchez-Rodriguez A, De Moor B, Engelen K, Marchal K (2010). COMODO: an adaptive coclustering strategy to identify conserved coexpression modules between organisms. *Nucleic Acids Res* 39(7): e41-e41.
- Zeng Y, Wu Y, Avigne WT, Koch KE (1999). Rapid repression of maize invertases by low oxygen. Invertase/sucrose synthase balance, sugar signaling potential, and seedling survival. *Plant Physiol* 121(2): 599-608.
- Zhai R, Feng Y, Zhan X, Shen X, Wu W, Yu P, Zhang Y, Chen D, Wang H, Lin Z, Cao L, Cheng S (2013). Identification of transcriptome SNPs for assessing allele-specific gene expression in a super-hybrid rice Xieyou9308. *PLoS One* 8(4): e60668.
- Zhang C & Turgeon R (2009). Downregulating the sucrose transporter VpSUT1 in Verbascum phoeniceum does not inhibit phloem loading. *Proc Natl Acad Sci U S A* 106(44): 18849-18854.

- Zhao Q & Dixon RA (2011). Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends Plant Sci* 16(4): 227-233.
- Zhao Q, Wang H, Yin Y, Xu Y, Chen F, Dixon RA (2010). Syringyl lignin biosynthesis is directly regulated by a secondary cell wall master switch. *Proc Natl Acad Sci U S A* 107(32): 14496-14501.
- Zhong R, Demura T, Ye ZH (2006). SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis. *Plant Cell* 18(11): 3158-3170.
- Zhong R, Lee C, McCarthy RL, Reeves CK, Jones EG, Ye ZH (2011). Transcriptional activation of secondary wall biosynthesis by rice and maize NAC and MYB transcription factors. *Plant Cell Physiol* 52(10): 1856-1871.
- Zhong R, Lee C, Ye ZH (2010). Global analysis of direct targets of secondary wall NAC master switches in Arabidopsis. *Mol Plant* 3(6): 1087-1103.
- Zhong R, Lee C, Zhou J, McCarthy RL, Ye ZH (2008). A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. *Plant Cell* 20(10): 2763-2782.
- Zhong R, Richardson EA, Ye ZH (2007a). The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in Arabidopsis. *Plant Cell* 19(9): 2776-2792.
- Zhong R, Richardson EA, Ye ZH (2007b). Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of Arabidopsis. *Planta* 225(6): 1603-1611.
- Zhong R & Ye ZH (2007). Regulation of cell wall biosynthesis. *Curr Opin Plant Biol* 10(6): 564-572.
- Zhong R & Ye ZH (2009). Transcriptional regulation of lignin biosynthesis. *Plant Signal Behav* 4(11): 1028-1034.
- Zhong R & Ye ZH (2012). MYB46 and MYB83 bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes. *Plant Cell Physiol* 53(2): 368-380.
- Zhou HL, He SJ, Cao YR, Chen T, Du BX, Chu CC, Zhang JS, Chen SY (2006). OsGLU1, a putative membrane-bound endo-1,4-beta-D-glucanase from rice, affects plant internode elongation. *Plant Mol Biol* 60(1): 137-151.
- Zhou J, Lee C, Zhong R, Ye ZH (2009). MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in Arabidopsis. *Plant Cell* 21(1): 248-266.
- Zhu JY, Sae-Seaw J, Wang ZY (2013). Brassinosteroid signalling. *Development* 140(8): 1615-1620.
- Zhu XG, Long SP, Ort DR (2008). What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Curr Opin Biotechnol* 19(2): 153-159.

LISTA DE APÊNDICES

Todos os apêndices estão disponíveis no CD-ROM que acompanha a tese.

Apêndice 1. Genes diferencialmente expressos entre os genótipos.

Apêndice 2. Genes diferencialmente expressos entre entrenó imaturo e intermediário.

Apêndice 3. Filogenia CesA.

Apêndice 4. Enriquecimento de termos dos genes com sinal de expressão acima do *background*.

Apêndice 5. *Clusterização* hierárquica dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó imaturo.

Apêndice 6. *Clusterização* hierárquica dos dados de sinal de expressão dos genes do catálogo de parede celular em entrenó intermediário.

LISTA DE ANEXOS

A - SÚMULA CURRICULAR

DADOS PESSOAIS

Nome: Sávio de Siqueira Ferreira

Nascimento: Ouro preto, 17 de fevereiro de 1984.

EDUCAÇÃO

Universidade Federal de Viçosa - MG, 2008.

Mestrado em Bioquímica Agrícola

Universidade Federal de Viçosa, Viçosa - MG, 2006.

Graduação - Bacharelado em Bioquímica

Colégio Imaculada Conceição, Passos - MG, 2001.

FORMAÇÃO COMPLEMENTAR

Treinamento em Aplicações e Fundamentos da PCR Quantitativa em Tempo Real, Life Technologies, São Paulo, 12 a 15 de fevereiro de 2011.

Advanced School on Biochemistry of Biofuels, promovido pela SBBq, Ubatuba, 26 de setembro a 2 de outubro de 2010.

PUBLICAÇÕES

Artigos Publicados

DE SIQUEIRA FERREIRA, SAVIO ; NISHIYAMA, MILTON ; PATERSON, ANDREW H ; SOUZA, GLAUCIA . Biofuel and energy crops: high-yield Saccharinae take center stage in the post-genomics era. *GenomeBiology.com* (London. Print), v. 14, p. 210, 2013

FERREIRA, S S ; Barros, D. R. ; de Almeida, M. R. ; ZERBINI, F. M. . Characterization of Passionfruit severe leaf distortion virus, a novel begomovirus infecting passionfruit in Brazil, reveals a close relationship with tomato-infecting begomoviruses. *Plant Pathology* (Print), v. 59, p. 221-230, 2010.

CALEGARIO, R. F. ; FERREIRA, S S ; ANDRADE, E. C. ; ZERBINI, F. M. . Characterization of Tomato yellow spot virus, a novel tomato-infecting begomovirus in Brazil. *Pesquisa Agropecuária Brasileira* (1977. Impressa), v. 42, p. 1335-1343, 2007.

ZERBINI, F. M. ; FERREIRA, S S ; ANDRADE, E. C. ; BARROS, D R ; LIMA, A. T. M. ; ALFENAS, P. F ; Mello, R. N. . Traditional and novel strategies for geminivirus management in Brazil. *Australasian Plant Pathology*, v. 34, p. 475, 2005.

Artigos em preparação

Milton Yutaka Nishiyama Jr.*, Savio Siqueira Ferreira*, Pei-Zhong Tang*, Scott Becker, Antje Taliana and Glaucia Mendes Souza. A transcriptome analysis of sugarcane ancestor and hybrid genotypes: defining the sugarcane ORFeome.

Sávio de Siqueira Ferreira, Débora Chaves Coelho Leite, Viviane Guzzo de Carli Poelking, Marcos Silveira Buckeridge, Marcelo Ehlers Loureiro, Márcio Henrique Pereira Barbosa, Monalisa Sampaio Carneiro and Glaucia Mendes Souza. Physiological and molecular characterization of ancestral species of sugarcane.

Paloma Mieko Sato, Rodrigo Fandino Andrade, Milton Yutaka Nishiyama-Jr, Carolina Gimiliani Lembke, Sávio de Siqueira Ferreira, Ana Paula Joaquim Silveira, Carlos Alberto Labate, Hermann Hoffmann, Monalisa Carneiro Sampaio and Glaucia Mendes Souza. Sucrose Accumulation Regulation Revealed Through Studies of the Transcriptome and Physiology in Breeding Populations of Sugarcane.

Capítulo de Livro

Nishiyama, Milton Yutaka; Vicente, Fabio; Sato, Paloma Mieko; Ferreira, Savio Siqueira; Feltus, Frank Alex; Souza, Glaucia Mendes. Transcriptome Analysis in the Saccharinae. In: Andrew H Paterson. (Org.). *Genomics of the Saccharinae*. 1ed. New York: Springer New York, 2013, v. 11, p. 121-139.

Trabalhos Completos Publicados Em Anais De Congressos

Lembke, Carolina G; Nishiyama, Milton Y; FERREIRA, S. S.; Dal-Bianco, Maximiller; Sato, Paloma M.; Hotta, Carlos T.; Carneiro, Monalisa Sampaio; LOUREIRO, M. E. ; CARRER, HELAINE; ENDRES, L.; Hoffmann, Hermann Paulo; SANCHES, M.; Menossi, Marcelo ; Souza, Glaucia M.. Sugarcane genome sequencing and gene Discovery:

getting closer to sugar content, fiber and drought traits.. In: XXVIII Congress International Society of Sugarcane Technologists, 2013, São Paulo. Proc. Int. Soc. Sugarcane Technol., 2013. p. 28.

Resumos Publicados Em Anais De Congressos

Ferreira, S. S. ; Leite, D. C. C. ; Poelking, V. G. C. ; Buckeridge, M. S. ; Loureiro, M. E. ; SOUZA, G. M. . Molecular and Physiological characterization of ancestral species of sugarcane. In: IV Simpósio Brasileiro de Genética Molecular de Plantas, 2013, Bento Gonçalves - RS. Simpósio Brasileiro de Genética Molecular de Plantas, 2013.

Barboza, A. L. ; Ferreira, S. S. ; SOUZA, G. M. ; Carrer, H. . Antisense of Caffeic Acid 3-O-Methyltransferase (COMT) Decreases Lignin Content in Transgenic Sugarcane (*Saccharum* spp.). In: The 2013 International Symposium on C4 and CAM Plant Biology, 2013, Urbana-Champaign. The 2013 International Symposium on C4 and CAM Plant Biology, 2013.

Kitano, E. S. ; Zelanis, A. ; Ferreira, S. S. ; Nishiyama-Jr, MY ; SOUZA, G. M. ; Serrano, S. M. T. . Unraveling the proteome of sugarcane leaves by mass spectrometric analysis. In: 12th HUPO World Congress, 2013, Yokohama. 12th HUPO World Congress, 2013.

Ferreira, S. S. ; Leite, D. C. C. ; Buckeridge, M. S. ; SOUZA, G. M. . Histochemical analysis of lignin in sugarcane ancestral species. In: II Congresso do Instituto de Química, 2012, Guarujá - SP. Congresso do Instituto de Química, 2012.

Ferreira, S. S. ; Nishiyama-Jr, MY ; SOUZA, G. M. . A cell wall gene catalogue for sugarcane. In: III Simpósio Brasileiro de Genética Molecular de Plantas, 2011, Ilhéus - BA. Simpósio Brasileiro de Genética Molecular de Plantas, 2011.

Ferreira, S. S. ; SOUZA, G. M. . Gene expression profiling of sugarcane varieties contrasting for sucrose content. In: Brazilian Bioenergy Science and Technology (BBEST), 2011, Campos do Jordão - SP. Brazilian Bioenergy Science and Technology, 2011.

Ferreira, S. S. ; Dal-Bianco, M ; Silva, G. L. ; Nishiyama-Jr, MY ; Loureiro, M. E. ; Barbosa, M. ; SOUZA, G. M. . Differentially Expressed Genes Between Contrasting Lignin Content Sugarcane Cultivars. In: 56 Congresso Brasileiro de Genética, 2010, Guarujá. Congresso Brasileiro De Genética, 2010.

FERREIRA, S S ; LIMA, A. T. M. ; ALMEIDA, M. R. ; ALFENAS-ZERBINI, P. ; ZERBINI, F. M. . Characterization of a begomovirus obtained from yellow passionfruit (*Passiflora edulis*

f. flavicarpa) and production of infectious clones. In: 19o Encontro Nacional de Virologia, 2008, Caxambú, MG. Virus Reviews and Research. Rio de Janeiro, RJ: Sociedade Brasileira de Virologia, 2008. v. 13. p. 270-271.

CHIARELLI NETO, O. ; VIDIGAL, P. M. P. ; SILVA JUNIOR, A. ; Silva, F. M. F. ; Costa, D. A. ; FERREIRA, S S ; Salgado, R. L. ; FIETTO, J. L. R. ; ALMEIDA, M. R. . CAPSIDE PROTEIN EPITOPES ANALYSIS OF BRAZILIANS PCV2 ISOLATES. In: XVIII ENCONTRO NACIONAL DE VIROLOGIA, 2007, ARMAÇÃO DE BÚZIOS. XVIII ENCONTRO NACIONAL DE VIROLOGIA, 2007.

FERREIRA, S S ; FONTES, E. P. B. ; ZERBINI, F. M. . Ocorrência de Begomovírus e DNAs satélites em tomateiro e plantas daninhas na região sudeste do Brasil. In: Simpósio de Iniciação Científica, 2006, Viçosa, MG. Em CD Room, 2006.

Pinheiro, Guilherme Luiz ; FERREIRA, S S ; OLIVEIRA, M. C. ; Gonçalves, Daniel Bonoto ; REIS, R. S. ; GRAVINA, H. D. ; DAMSCENO, B. L. ; REZENDE, S. T. ; QUEIROZ, J. H. . Caracterização Bioquímica e cinética da invertase de *Saccharomyces cerevisiae*. In: XV Simpósio de Iniciação Científica, 2006, Viçosa. XV Simpósio de INiciação Cientifica, 2006.

FERREIRA, S S ; FONTES, E. P. B. ; ZERBINI, F. M. . OCCURRENCE OF BEGOMOVIRUSES AND SATELLITES DNAs IN WEEDS. In: XXXIV Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2005, Águas de Lindóia - SP. Livro de resumos da XXXIV Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2005.

FERREIRA, S S ; ALFENAS, P. F ; LIMA, A. T. M. ; BARROS, D R ; SENNA, L. ; ZERBINI, F. M. . Detecção de begomovírus em tomateiro no município de Paty de Alferes, Rio de Janeiro. In: 38o Congresso Brasileiro de Fitopatologia, 2005, Brasília, DF. Fitopatologia Brasileira, 2005. v. 30. p. S183.

CALEGARIO, R. F. ; ANDRADE, E. C. ; FERREIRA, S S ; MANHANI, G G ; ZERBINI, F. M. . Biological and molecular characterization of a tomato isolate of *Sida micrantha* mosaic virus (SimMV).. In: 38o Congresso Brasileiro de Fitopatologia, 2005, Brasília, DF. Fitopatologia Brasileira, 2005. v. 30. p. S180.

CALEGARIO, R. F. ; ANDRADE, E. C. ; FERREIRA, S S ; ZERBINI, F. M. . Caracterização de um isolado do begomovírus *Sida micrantha* mosaic virus (SimMV) obtido de tomateiro. In: 37o Congresso Brasileiro de Fitopatologia, 2004, Gramado, RS.. Fitopatologia Brasileira., 2004. v. v.29. p. S150-S150.

MOREIRA, A. G. ; PEREIRA, De O. C. ; ANDRADE, E. C. ; CALEGARIO, R. F. ; FERREIRA, S S ; ZERBINI, F. M. . MOLECULAR CLONING OF TWO DIFFERENT BEGOMOVIRUS INFECTING SOYBEAN AND ASSOCIATED WEEDS IN BRAZIL AND THEIR POTENTIAL USE AS A VECTOR FOR TRANSIENT GENE EXPRESSION IN PLANTS. In: 15o Encontro Nacional de Virologia, 2004, São Pedro, SP. Livro de Resumos do Congresso Nacional de Fitopatologia, 2004. v. 9. p. 243-243.