UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE BIOCÊNCIAS

FLAVIANE LOPES FERREIRA

Predição *in silico* do papel de *ScRAV1* sobre os genes de parede celular e fatores de transcrição durante a formação do aerênquima na raiz de cana-de-açúcar

In silico prediction of the role of ScRAV1 on cell wall genes and transcription factors during the formation of the aerenchyma in the sugarcane root

São Paulo

2019

FLAVIANE LOPES FERREIRA

Predição *in silico* do papel de ScRAV1 sobre os genes de parede celular e fatores de transcrição durante a formação do aerênquima na raiz de cana-de-açúcar

In silico prediction of the role of ScRAV1 on cell wall genes and transcription factors during the formation of the aerenchyma in the sugarcane root

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para obtenção de título de Mestre em Ciências, na Área de Concentração de Biologia de Sistemas.

Versão Corrigida (original se encontra disponível no IB-USP)

Orientador : Prof. Dr. Marcos S. Buckeridge

São Paulo

2019

FICHA CATALOGRÁFICA

Lopes Ferreira, Flaviane

Predição in silico do papel de ScRAV1 sobre os genes de parede celular e fatores de transcrição durante a formação do aerênquima na raiz de cana-de-açúcar / Flaviane Lopes Ferreira; orientador Marcos Silveira Buckeridge. -- São Paulo, 2019.

201 f.

Dissertação (Mestrado) - Instituto de Biociências da Universidade de São Paulo, Departamento de Botânica.

1. ScRAV1 . 2. Redes de Co-expressão . 3. Redes de Regulação Gênica . 4. Elementos cis-regulatórios. 5. Aerênquima. I. Buckeridge, Marcos Silveira, orient. II. Título.

Comissão Julgadora:

Prof (a). Dr (a).

Prof (a). Dr (a).

Prof. (a) Dr. (a). Orientador (a)

The important thing is not to stop questioning. Curiosity has its own reason for existing. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of this mystery every day.

Albert Einstein

AGRADECIMENTOS

Ao Prof. Dr. Marcos Silveira Buckeridge pela oportunidade de desenvolver este trabalho.

À Profa. Dra. Rosemeire Aparecida Pessoni da Universidade Metodista de São Paulo por ter sempre acreditado no meu potencial desde da graduação.

À Profa. Dra Marília Gaspar pelas sugestões de correção desta tese.

À Dra. Paula Elbl e Dra. Eveline Tavares pelos ensinamentos na parte de bioinformática que foram essenciais no desenvolvimento deste projeto.

À Dra. Adriana Grandis pelo incentivo e pela colaboração nesta dissertação.

Ao Ms. Vinícius Carvalho Jardim pela contribuição na parte de redes biológicas.

Aos colegas de laboratório – Bruno Flausino, Grayce Romim, Lauana, Eglee Igarashi e Viviane Lopes pela ajuda e pelas conversas.

E por último, ao CNPq pela bolsa concedida durante estes 24 meses de pesquisa.

ÍNDICE

RESUMO	1
ABSTRACT	2
I. INTRODUÇÃO	3
1. A parede celular da cana-de-açúcar	3
2. A cana-de-açúcar e a bioenergia	7
3. Formação do aerênquima nas plantas	10
3.1. Morte celular programada	13
3.2. Modificações na parede celular	14
4. Cascata de sinalização de etileno.....	16
4.1. Fatores de transcrição RAV (<i>Related to ABI3/VP1</i>).....	18
4.1.1. Papéis do fator de transcrição RAV nas plantas.....	21
II. OBJETIVOS	24
1. Objetivo geral	24
2. Objetivos específicos.....	24
III. MATERIAIS E MÉTODOS	25
1. Árvore filogenética do fator de transcrição RAV.....	25
2. Análise dos domínios proteicos de <i>ScRAV1</i>	26
3. Redes de co-expressão.....	26
3.1. Construção das redes de co-expressão	27
3.2. Filtro das redes de co-expressão através da centralidade de grau.....	29
3.3. Anotação dos transcritos de parede celular e fatores de transcrição de cana-de-açúcar em <i>S. bicolor</i> e <i>A. thaliana</i>	30
3.4. Enriquecimento funcional dos transcritos de parede celular e fatores de transcrição através dos termos do Gene Ontology (GO).....	31
3.5. Predição <i>in silico</i> de sítios de ligação de RAV em <i>S. bicolor</i>	32
3.6. Construção das redes de regulação gênica.....	32
3.7. Avaliação do perfil de expressão dos transcritos	33
3.8. Seleção dos transcritos de parede celular e fatores de transcrição	33
IV. RESULTADOS	36

1. Filogenia do fator de transcrição RAV	36
2. Caracterização dos domínios proteicos de <i>ScRAV1</i>	38
3. Redes de co-expressão de <i>ScRAV1</i> com os transcritos de parede celular	39
3.1. Categorização dos transcritos de parede celular presentes nas redes de co-expressão de <i>ScRAV1</i>	42
3.2. Seleção dos transcritos de parede celular pela centralidade de grau	45
3.3. Busca do homólogo mais próximo em <i>S. bicolor</i> e em <i>A. thaliana</i>	47
3.4. Enriquecimento funcional dos transcritos de parede celular através dos termos do Gene Ontology (GO)	52
3.5. Sítios de ligação de RAV ao DNA nas sequências homólogas de <i>S. bicolor</i>	54
3.6. Rede de regulação gênica para os genes de parede celular	61
3.7. Perfil de expressão dos transcritos de parede celular com base no RNA-Seq.....	63
4. Redes de co-expressão de <i>ScRAV1</i> com os fatores de transcrição	66
4.1. Avaliação das redes de co-expressão dos fatores de transcrição com <i>ScRAV1</i>	69
4.2. Seleção dos fatores de transcrição pela centralidade de grau	70
4.3. Busca do homólogo mais próximo em <i>S. bicolor</i> e em <i>A. thaliana</i>	72
4.4. Enriquecimento funcional dos fatores de transcrição através dos termos do Gene Ontology (GO)	76
4.5. Sítios de ligação ao DNA para RAV nas sequências homólogas de <i>S. bicolor</i>	78
4.6. Rede de regulação gênica para os fatores de transcrição	82
4.7. Perfil de expressão dos fatores de transcrição com base no RNA-Seq.....	84
V. DISCUSSÃO	86
1. <i>ScRAV1</i> apresenta três sequências homólogas em cana-de-açúcar possivelmente advindas de duplicação gênica e vários ortólogos em diferentes espécies	87
2. <i>ScRAV1</i> apresenta os domínios AP2 e B3 e o motivo BRD (<i>B3 repressor domain</i>) repressor conservados em diferentes espécies de plantas.....	90
3. Redes de co-expressão de <i>ScRAV1</i> como método de busca de transcritos de parede celular e fatores de transcrição que possivelmente agem durante a formação do aerênquima nos diferentes segmentos (S1 ao S4) da raiz em cana-de-açúcar	91
3.1. As redes de co-expressão de <i>ScRAV1</i> com os transcritos de parede celular e fatores de transcrição sugerem uma relação com as diferentes partes do desenvolvimento da raiz em cana-de-açúcar.....	93
3.2. <i>ScRAV1</i> co-expressa com os transcritos de parede celular associados à biossíntese quanto à hidrólise da parede celular na raiz de cana-de-açúcar	93
3.3. <i>ScRAV1</i> co-expressa com várias famílias de fatores de transcrição com diferentes papéis nas plantas	97

3.4. Identificação dos nós essenciais nas redes de co-expressão de <i>ScRAV1</i> através da centralidade de grau.....	102
3.5. Os homólogos em <i>S. bicolor</i> e em <i>A. thaliana</i> para os transcritos de parede celular e para os fatores de transcrição de cana-de-açúcar.....	103
3.6. Anotação funcional dos genes de parede celular e fatores de transcrição através dos termos do Gene Ontology (GO) de <i>A. thaliana</i>	104
3.7. Predição de sítios putativos de ligação de RAV ao DNA para os genes de parede celular e para os fatores de transcrição em <i>S. bicolor</i>	106
3.8. Redes regulatórias gênicas de genes de parede celular e de fatores de transcrição por <i>SbRAV1</i> .	111
3.9. Perfil de expressão dos transcritos de GPs/TFs com base no RNA-Seq.....	113
4. Seleção dos transcritos de parede celular e fatores de transcrição.....	116
4.1. <i>ScRAV1</i> possivelmente modula a expressão de glicosil hidrolases durante a formação do aerênquima	117
4.2. <i>ScRAV1</i> provavelmente modula a expressão dos fatores de transcrição que respondem à auxina	121
4.3. <i>ScRAV1</i> possivelmente coordena a expressão de fatores de transcrição relacionados à MCP no córtex da raiz de cana-de-açúcar.....	126
4.4. <i>ScRAV1</i> possivelmente coordena a expressão de genes relacionados à biossíntese de lignina no córtex durante a formação do aerênquima na raiz de cana-de-açúcar.....	132
VI. CONCLUSÕES	137
VII. REFERÊNCIAS BIBLIOGRÁFICAS.....	140
VIII. APÊNDICES	184

RESUMO

As modificações nas paredes celulares de células de raízes de cana-de-açúcar têm sido usadas como modelo para entender como as paredes poderiam ser alteradas *in vivo*. As alterações na parede de células corticais de raízes de cana-de-açúcar se iniciam com modificações nas pectinas que são responsáveis pelas junções intercelulares (lamela média). O fator de transcrição *ScRAV1* demonstrou controlar a expressão de uma endopoligalacturonase que degrada pectinas, iniciando o processo de formação do aerênquima que ocorre concomitantemente com a morte celular programada e forte alteração na parede, gerando um compósito que constitui a formação dos espaços de transporte de gases nas raízes. Paralelamente, os padrões de expressão de vários transcritos para genes de parede celular foram descritos recentemente por nosso grupo. Este trabalho teve por objetivo prever possíveis transcritos de parede celular e fatores de transcrição que podem ser modulados pelo fator de transcrição *ScRAV1* através das redes de co-expressão e de regulação gênica durante a formação do aerênquima em cana-de-açúcar. Nossas análises mostraram que *ScRAV1* apresenta co-expressão com vários transcritos de parede celular, bem como com outros fatores de transcrição ao longo dos primeiros 4 centímetros mais apicais da raiz de cana-de-açúcar. As redes de regulação gênica do ortólogo de *ScRAV1* em *S. bicolor* mostraram uma predominância de transcritos de parede celular associada à biossíntese de lignina, dois transcritos relacionados à hidrólise de hemicelulose/celulose (*ScGH3 2* e *ScEGase4*) e dois transcritos relacionados à hidrólise de pectina (*ScPAE2* e *ScPGase1*). Já para os fatores de transcrição foram identificados cinco transcritos que possivelmente regulam a biossíntese de lignina (*ScNAC3*, *ScNAC5*, *ScNAC7*, *ScHB6* e *ScMYB2*), dois associados à sinalização de auxina (*ScARF1* e *ScARF3*), e dois relacionados à morte celular programada (*ScNAC2* e *ScNAC8*) durante o desenvolvimento do aerênquima lisígeno na raiz de cana-de-açúcar. Nossos dados confirmam a hipótese prévia de que a formação do aerênquima nas raízes de cana-de-açúcar é feita através de módulos de ação de enzimas hidrolíticas, tendo o controle da expansão celular por auxinas e a morte celular programada como processo concorrente em paralelo às modificações na parede celular.

Palavras-chave: *ScRAV1*, aerênquima, hidrólise de pectina e hemicelulose, lignina, sinalização de auxina e MCP.

ABSTRACT

Modifications in the cell walls of roots of sugarcane have been used as a model to understand how its cell walls could be altered *in vivo*. The alteration in the walls of cortex cells initiate with modifications in the pectins, which are responsible for the cohesion in the middle lamella. The transcription factor *ScRAV1* has been demonstrated to control the expression of an endo-polygalacturonase that degrades pectins, initiating the process of formation of the aerenchyma, which occurs concomitantly with programmed cell death. At the same time, the expression patterns of cell wall-related genes have been studied by our group recently. Thus, this work aimed to predict possible cell wall transcripts and transcription factors that may be modulated by *ScRAV1* through coexpression and gene regulatory networks during the aerenchyma formation in sugarcane. The co-expression networks of *ScRAV1* showed that this transcription factor can co-express with types of cell wall transcripts and transcription factors in the different segments of the sugarcane roots. The gene regulatory networks of the ortholog of *ScRAV1* in *S. bicolor* showed a predominance of cell wall transcripts associated with lignin biosynthesis, two transcripts related to hemicellulose/cellulose hydrolysis (*ScGH3 2* and *ScEGase4*) and two transcripts related to pectin hydrolysis (*ScPAE2* and *ScPGase1*). As for the transcription factors were identified five transcripts that possibly regulate the lignin biosynthesis (*ScNAC3*, *ScNAC5*, *ScNAC7*, *ScHB6* and *ScMYB2*), two transcripts associated with auxin signaling (*ScARF1* and *ScARF3*), and two transcripts related to programmed cell death (*ScNAC2* and *ScNAC8*) during the development of the lysigenous aerenchyma in the sugarcane root. Our results confirm our previous hypothesis that aerenchyma formation occurs through modules that include the action of hydrolases and have cell expansion controlled by auxin and programmed cell death as a concurrent process with modifications in the cell wall.

Key words: *ScRAV1*, aerenchyma, hydrolysis of pectin and hemicellulose, lignin, auxin signaling and PCD.

I. INTRODUÇÃO

1. A parede celular da cana-de-açúcar

A cana-de-açúcar (*Saccharum* spp.) é uma gramínea que apresenta parede celular do tipo II devido as suas principais hemiceluloses consistirem de glucano de ligação mista (GLM) e o arabinoxilano (AX), e apresentar um domínio pécico reduzido (Carpita & Gibeaut, 1993; De Souza *et al.*, 2013). Além disso, o xiloglucano (XG) é encontrado em baixas concentrações, e uma alta quantidade de compostos fenólicos interligam os polissacarídeos na parede celular, principalmente com o ácido ferúlico (FA) (Silva, 2005; De Souza, 2011; Crivellari, 2012; De Souza *et al.*, 2013) (Figura 1).

Figura 1. Representação da parede celular do tipo II em gramíneas. As microfibrilas de celulose estão principalmente interligadas ao glucuroarabinoxilano (GAX) e ao glucano de ligação mista (GLM) ao invés do xiloglucano (XG). Os GAXs também podem se ligar a pectina na lamela média e também ao ácido ferúlico (FA), este último fazendo pontes entre os polissacarídeo e a lignina. O XG e a pectina estão em menores proporções neste tipo de parede celular. Retirado de Carpita *et al.* (2015).

As paredes celulares de cana-de-açúcar são compostas de cerca de 30% de celulose, 40% de AX; 10 % GLM; 8% de XG; 8% de pectinas incluindo o homogalacturonano (HG), o arabinogalacturonano e o arabinano, e 6% de lignina (De Souza *et al.*, 2013). A celulose está organizada em forma de microfibrilas, a qual contém várias cadeias lineares de $\beta(1,4)$ -D-glucano ligadas umas a outras por pontes de hidrogênio (Carpita *et al.*, 2015). Este polissacarídeo está ligado as hemiceluloses como o xiloglucano e o arabinoxilano (Buckeridge *et al.*, 2017), os quais estão envolvidos pelas matrizes de pectina e proteínas estruturais (Buckeridge *et al.*, 2015; 2017). A celulose liga-se mais fortemente ao xiloglucano (De Souza *et al.*, 2013), mas também se liga ao arabinoxilano que mantém a arquitetura da parede celular através das ligações cruzadas com a lignina (Buckeridge *et al.*, 2019).

Os arabinoxilanos (AXs) estão geralmente aderidos à superfície da celulose (Buanafina & Cosgrove, 2014) e nas células adjacentes na lamela média (Ishii, 1984; Tabuchi *et al.*, 2011). AXs são polissacarídeos heterogêneos que apresentam uma cadeia principal com ligações β -1,4 de resíduos de xilose, sendo substituída com arabinose (Ara), nas posições O-3 da cadeia principal do xilano. Em algumas espécies também podem ocorrer substituições nas posições O-2 da cadeia principal (Carpita, 1996). A cadeia principal do xilano ainda pode ser substituída com resíduos de ácido 4-O-metilglucurônico. Esta característica confere ao polímero o nome de glucuroarabinoxilano (GAX) (Carpita, 1996). Além disso, parte dos resíduos de arabinose podem ser substituídos por moléculas de ácido hidroxicinâmico (HCA), ácido ferúlico (4-hidroxi-3-metoxicinâmico) e substituída, em menor extensão, com o ácido *p*-cumárico (4-hidroxicinâmico) (Buanafina & Cosgrove, 2014). Nas gramíneas, FA faz ligação covalente com os resíduos de arabinose do AX ou GAX via ligação éster (Buanafina, 2009). A presença de FA permite os polissacarídeos-feruloil-éster reagirem com monolignóis via acoplamento oxidativo e gerar complexos polissacarídeos-ferulato-lignina através de ligações éster-éter

(Jacquet *et al.*, 1995). O acúmulo de ferulatos e de ligações cruzadas de AXs via diferulatos éster influenciam nas propriedades da parede celular, no crescimento foliar, na lignificação e na digestibilidade da parede celular (Buanafina & Cosgrove, 2014).

O GLM é um polímero não ramificado com ligações glicosídicas β (1,3) e β (1,4) dispersas ao longo da cadeia com intervalos de 3-6 resíduos de glicose (Buckeridge *et al.*, 2004). A sua estrutura consiste principalmente de subunidades de celotriose e celotetraose na razão de 2.2-2.6:1 (Gibeaut & Carpita, 1993; Fincher & Stone, 2004; Fincher, 2009). No glucano de ligação mista, as ligações β (1,3) são sempre flanqueadas por pelo menos duas ligações β (1,4) (Held *et al.*, 2015).

O XG é composto por uma cadeia principal de β (1,4)-D-glucano a qual é parcial e regularmente substituída com resíduos de α -D-xilose nas posições O-6 dos resíduos de glicose, enquanto as galactoses terminais se ligam na posição O-2 dos resíduos de xilose através de ligações β (1,2) (Hayashi, 1989). Este polissacarídeo ainda pode apresentar uma α (1,2) L-fucose terminal nos resíduos de galactose (Hayashi, 1989). Este polímero está ligado à celulose na parede celular (Park & Cosgrove, 2012).

As pectinas estão presentes na lamela média e são importantes na separação das microfibrilas de celulose e na adesão celular (Jarvis *et al.*, 2003; Caffall & Mohnen, 2009; Peaucelle *et al.*, 2012). Os polissacarídeos pécticos englobam quatro classes de polímeros: homogalacturonano (HG), ramnogalacturonano I (RGI), ramnogalacturonano II (RGII), e as pectinas neutras como arabinanos/galactanos (Buanafina & Cosgrove, 2014). Dentre as pectinas, o HG é o mais abundante, e apresenta ligações α (1,4)-ácido galacturônico com cadeias não-ramificadas (Buanafina & Cosgrove, 2014). Os HGs não-esterificados fazem pontes de cálcio formando um gel rígido, uma característica importante na adesão celular e na extensibilidade da parede celular (Buanafina & Cosgrove, 2014). A segunda pectina mais

abundante é o RGI, o qual consiste em dissacarídeos repetidos ligados (α -(1,4)-D-ácido galacturônico 1,2- α -L-ramnose), apresentando cadeias laterais neutras altamente variáveis (galactanos, arabinanos e/ou arabinogalactanos (Talbot & Ray, 1992), ligadas à cadeia principal do RG através dos resíduos de ramnose (Caffall & Mohnen, 2009). O RGII é o polímero mais complexo da parede celular vegetal, apresentando uma cadeia principal de HG ramificado com 12 açúcares diferentes ligados através de mais 20 ligações químicas distintas (Buanaína & Cosgrove, 2014). Este polissacarídeo pode formar ligações cruzadas com éster borato, permitindo fazer ligações covalentes com os HGs e formar uma grande rede de polissacarídeos (Buanaína & Cosgrove, 2014).

Já a lignina é um polímero aromático complexo que envolve os polissacarídeos da parede celular (principalmente a celulose e as hemiceluloses) e as glicoproteínas. Este polímero é sintetizado a partir do acoplamento oxidativo de monômeros de álcool *p*-hidroxicinâmico e compostos relacionados (Vanholme *et al.*, 2010; 2012). As principais unidades deste polímero são os álcoois *p*-hidroxifenil (H), guaiacil (G) e siringil (S), os quais são derivados dos monolignóis álcoois *p*-cumaril, coniferil e sinapsil. Os monômeros H, G e S diferem apenas pelo número de substituintes metoxil no anel aromático (Vanholme *et al.*, 2012; Calvo-Flores *et al.*, 2015). Em cana-de-açúcar, a lignina está presente principalmente nas paredes secundárias, mas também está em menores proporções nas paredes primárias das células parenquimáticas (Buckeridge *et al.*, 2017). Em ambos os casos, a lignina está associada ao AX ou GAX e pectinas (galactano) via ligações ésters que servem como pontos de nucleação para a síntese deste polímero na parede celular (Buanaína, 2009; Buckeridge *et al.*, 2019).

Como descrito acima, as paredes celulares são constituídas de vários polímeros com estruturas variadas que estão ligadas por ligações covalentes e não-covalentes (Buckeridge &

De Souza, 2014; Buckeridge *et al.*, 2015; Tavares & Buckeridge, 2015). A organização destes polissacarídeos na parede celular e as ligações cruzadas entre eles definem a arquitetura da parede celular (Tavares *et al.*, 2015). Acredita-se que o acesso destes polissacarídeos é impedido pela organização e pela localização destes polímeros dentro da parede celular, conhecido como o código glicômico, levando à resistência na quebra destes polissacarídeos por ação mecânica ou microbial (Buckeridge & De Souza, 2014), isto é a recalcitrância à hidrólise destes polímeros (Himmel *et al.*, 2007).

Entre as gramíneas mais estudadas para o uso na bioenergia está a cana-de-açúcar (De Souza *et al.*, 2013), pois esta planta produz altas concentrações de sacarose no colmo (De Souza *et al.*, 2014), que é a principal matéria-prima usada pelo Brasil - um dos grandes produtores mundiais de bioetanol (RFA, 2018).

2. A cana-de-açúcar e a bioenergia

A cana-de-açúcar tornou-se uma das principais fontes de bioenergia no planeta, sendo esta biomassa usada para a produção de etanol e açúcar. As plantações de cana-de-açúcar no Brasil estão concentradas principalmente na região Centro-Sul e no Nordeste. De acordo com dados da UNICA (2017), a região Centro-Sul produziu na safra de 2017/2018, cerca de 596 milhões de toneladas de cana-de-açúcar. Somente o estado de São Paulo produziu no mesmo período cerca de 366 milhões de toneladas de cana-de-açúcar (UNICA, 2017), sendo o responsável por mais da metade da produção de cana-de-açúcar nesta região. Consequentemente, a região Centro-Sul produziu cerca de 36 milhões de toneladas de açúcar e 26 milhões de litros de etanol (UNICA, 2017).

A produção de bioetanol de primeira geração (1G) a partir de sacarose acumulada nos colmos da cana-de-açúcar se tornou extremamente eficiente e com custo efetivo no sudeste

do Brasil a partir da terceira metade do século XX (De Souza *et al.*, 2014). Entretanto, a produção de bioetanol ainda poderia ser significativamente melhor aproveitada se a parede celular presente no bagaço e nas folhas da cana-de-açúcar pudessem ser usadas para a produção de bioetanol através do processo de segunda geração (2G) (Buckeridge *et al.*, 2010; Dos Santos *et al.*, 2011; Buckeridge *et al.*, 2012).

De acordo com Goldemberg (2008), apenas um terço da biomassa de cana-de-açúcar é convertida em etanol, enquanto os outros dois terços são deixados como bagaço e palha. Cerca de 92% do bagaço de cana-de-açúcar é queimado para produzir eletricidade. Se apenas os 8% restantes deste material fossem convertidos em etanol, poderiam ser produzidos cerca de 2200 L de etanol/hectare e isto reduziria o uso da terra em aproximadamente 29% (Macedo, 2005). E ainda, se a palha fosse usada junto com 50% do bagaço, seriam produzidos entre 3700 a 4000 L/hectare de etanol, reduzindo cerca de 33 a 38% do uso da terra (Macedo, 2005). A utilização desta biomassa pode levar ao aumento da demanda global por etanol e também atender aos critérios de proteção ambiental (Buckeridge *et al.*, 2012; Jaiswal *et al.*, 2017).

Porém, a recalcitrância da parede celular tem sido uma barreira no desenvolvimento dos processos do bioetanol de (2G) devido a falta de conhecimento sobre a complexidade desta estrutura (Buckeridge *et al.*, 2015). A fim de diminuir a recalcitrância da parede celular e expor os polissacarídeos à hidrólise, têm sido propostos vários tipos de pré-tratamentos para a produção de bioetanol. Entre eles estão: os pré-tratamentos físicos (explosão à vapor e hidrotérmico), químicos (ácido, álcali e organosolve) e os pré-tratamentos biológicos (uso de enzimas de micro-organismos e a fermentação biológica) (Cardona *et al.*, 2010).

Entretanto, sabe-se que os pré-tratamentos físicos e químicos consomem muita energia, geram inibidores que afetam a hidrólise enzimática e a fermentação (Mosier *et al.*,

2005). Já os pré-tratamentos biológicos se mostram uma tecnologia promissora, por apresentar várias vantagens, inclusive sendo uma estratégia economicamente viável e amigavelmente correta para aumentar a taxa de sacarificação, pois não há a necessidade de reciclagem de compostos químicos, e não há liberação de compostos tóxicos ao meio ambiente (Sindhu *et al.*, 2016).

Neste sentido, várias pesquisas têm sido feitas para o uso de glicosil hidrolases (GHs) produzidas por micro-organismos, a fim de usá-las para modificar arquitetura e a organização da parede celular para a produção de bioetanol de 2G (Soccol *et al.*, 2010). Várias enzimas têm sido clonadas e expressas em diferentes micro-organismos como na bactéria *Escherichia coli* (Ebanks *et al.*, 2000) e nas leveduras *Saccharomyces cerevisiae* (Das & Schultz, 1987) e *Pichia pastoris* (Cregg, 1999). Os fungos também têm sido estudados devido a sua capacidade natural de secretar grandes quantidades de proteínas, entre os fungos filamentosos mais estudados estão os gêneros *Aspergillus* e *Trichoderma* (De Vries *et al.*, 2011). Algumas desvantagens deste tipo de pré-tratamento é que pouco se têm explorado quanto as características de pH, temperatura e a adaptabilidade destes micro-organismos em substratos de baixo custo (Polizeli *et al.*, 2011). Além disso, as enzimas para hidrolisar os polissacarídeos da parede celular representam cerca de 9% do custo total para a conversão da biomassa (Aden *et al.*, 2002), o que dificulta o uso destas enzimas para a produção de bioetanol de 2G (Banerjee *et al.*, 2010).

Recentemente, foi proposto um novo tipo de pré-tratamento biológico que permite usar mecanismos que consistem no ataque a polissacarídeos por meio do emprego de processos de degradação endógena de parede celular de cana-de-açúcar (Tavares *et al.*, 2015). Este tipo de processo seguiria um modelo hierárquico de hidrólise enzimática seguindo a arquitetura da parede celular de cada tecido, como proposto por De Souza *et al.* (2013). Os

processos endógenos de degradação da parede celular ocorrem naturalmente nas plantas, e alguns exemplos são: a mobilização de reservas, o amadurecimento de frutos, a abscisão foliar e a formação do aerênquima (Grandis *et al.*, 2014). Estes sistemas usam uma série sequencial de módulos que levam à modificação da parede celular. O início destes módulos se dá com 1) a percepção do hormonal, 2) a separação das células, 3) a expansão celular, 4) a morte celular programada (MCP), 5) o afrouxamento e a hidrólise do domínio hemicelulose-celulose e 6) a hidrólise de celulose (Grandis *et al.*, 2014). A aquisição do controle sobre o mecanismo que a planta possui de degradar suas próprias paredes celulares poderia diminuir a demanda de energia para o pré-tratamento e facilitar a hidrólise da biomassa (Grandis *et al.*, 2014).

De acordo com Tavares (2015), dentre os mecanismos endógenos de degradação da parede celular que poderiam ser viáveis na produção de bioetanol 2G a partir da cana-de-açúcar, a formação do aerênquima pode ser um dos processos mais promissores para esta finalidade. De fato, a formação deste tecido já foi identificada na raiz e no colmo desta planta (Gilbert *et al.*, 2007; Tetsushi & Karim, 2007; Begum *et al.*, 2008). O nosso grupo realizou a caracterização do desenvolvimento deste tecido e identificou vários transcritos/ proteínas de degradação de parede celular que podem estar envolvidas neste processo (Grandis, 2015; Tavares, 2015; Leite *et al.*, 2017).

3. Formação do aerênquima nas plantas

Os aerênquimas são espaços formados durante o desenvolvimento de tecidos vegetais que permitem o transporte de gases pelos tecidos, principalmente o oxigênio (Armstrong, 1979), o dióxido de carbono, o metano e o etileno (Colmer, 2003). A condução destes gases pode ocorrer tanto por difusão como por fluxo de pressão (Raven, 1996; Jackson & Armstrong, 1999). Essas lacunas interconectadas formam um sistema de aeração, permitindo que partes

da planta sobrevivam e cresçam em ambientes deficientes ou completamente desprovidos de O_2 (Drew *et al.*, 2000). A formação deste tecido leva à diminuição na demanda de oxigênio pela raiz devido a MCP de algumas células do córtex (Armstrong, 1979).

Há dois tipos de aerênquima: o lisígeno e o esquizógeno (Sachs, 1882; De Bary, 1884; Arber, 1920) (Figura 2). O aerênquima esquizógeno é formado a partir espaços já existentes através da separação e da expansão celular (Arber, 1920; Justin & Armstrong, 1987; Jackson & Armstrong, 1999). Alguns exemplos do aerênquima esquizógeno são o do gênero *Rumex* que forma este tipo de aerênquima na sua raiz principal e nas suas raízes adventícias (Justin & Armstrong, 1987; Colmer *et al.* 2004), e na espécie *Sagittaria lancifolia* a qual forma aerênquima esquizógeno em seus pecíolos, mas também pode formar aerênquima lisígeno em suas raízes (Schussler & Longstreth, 1996). Já o aerênquima lisígeno é o resultado da morte e lise celular de células específicas no córtex da raiz (Arber, 1920; Justin & Armstrong, 1987; Jackson & Armstrong, 1999; Evans, 2003). Foi visto que este tipo de aerênquima se forma nas raízes, no caule, na bainha da folha, nos entrenós e também nas nervuras das folhas de *Oryza sativa* (Matsukura *et al.*, 2000; Colmer & Pedersen, 2008; Parlanti *et al.*, 2011; Steffens *et al.*, 2011).

Figura 2. Tipos de aerênquima nas plantas. Aerênquima esquizógeno na raiz de *Acorus calamus* (A), e o aerênquima lisígeno na raiz de *Oryza sativa* (B). Retirado de Takahashi *et al.* (2014).

O aerênquima lisígeno pode ser constitutivo ou induzido (Fagerstedt, 2010). Alguns exemplos de aerênquima constitutivo em raízes ocorrem nas espécies *Juncus effusus* (Smirnoff & Crawford, 1983), *Zea nicaraguensis* (Mano & Omori, 2013), *Oryza sativa* (Jackson *et al.*, 1985) e *Saccharum* spp. (Leite *et al.*, 2017). O segundo tipo de aerênquima é formado nas raízes de *Zea mays* (Drew *et al.*, 1981; Gunawardena *et al.*, 2001a). Há outras espécies de gramíneas em que ainda não foi definido se o aerênquima é constitutivo ou induzido. Entre elas estão *Triticum aestivum* (Trought & Drew, 1980), *Hordeum vulgare* (Arikado & Adachi, 1955) e *Sorghum bicolor* (Promkhambut *et al.*, 2011).

A formação do aerênquima lisígeno pode ser causada por estímulos externos e internos (Evans, 2003). Dentre os estímulos externos, estão a impedância mecânica, a deficiência de nitrato, fosfato (Drew *et al.*, 1989) e sulfato (Bouranis *et al.*, 2003), bem como a resposta à hipóxia, à seca e a alta temperatura (Evans, 2003). De acordo com Evans (2003), a resposta a estes diferentes estresses podem levar ao mesmo tipo de MCP durante a formação do aerênquima. Em relação ao estímulo interno, o hormônio etileno desempenha um papel importante no desenvolvimento do aerênquima (Jackson & Armstrong, 1999; Drew *et al.*, 2000; Evans, 2003). Já foi observado que quando a taxa de etileno diminui na raiz de *Z. mays*, a mesma se torna mais responsiva a este hormônio, levando à formação do aerênquima (He *et al.*, 1992). Mais recentemente, sugeriu-se que a sensibilidade ao etileno e o balanço entre etileno-auxina podem ter um papel importante durante na formação do aerênquima constitutivo na raiz de cana-de-açúcar (Tavares *et al.*, 2018). No caso da auxina, acredita-se que esta pode estar envolvida na expressão de genes relacionados à expansão da parede celular, evento este que antecede a formação do aerênquima na raiz (Tavares *et al.*, 2018).

3.1. Morte celular programada

Um dos primeiros eventos da formação do aerênquima é a morte celular programada (MCP) no córtex da raiz (Evans, 2003). O termo MCP refere-se à um processo ativo e geneticamente controlado que leva à eliminação seletiva de células indesejadas no organismo (Gadjev *et al.*, 2008). Nos vegetais, este processo ocorre no desenvolvimento do endosperma (Young & Gallie, 2000), na formação dos elementos de vasos (Fukuda, 1996), na senescência foliar (Gunawardena *et al.*, 2004), nas respostas do vegetal aos estresses ambientais, e no ataque à patógeno (Yan *et al.*, 2018) e na formação do aerênquima (Gunawardena *et al.*, 2001a; Leite *et al.*, 2017).

A MCP durante a formação do aerênquima lisígeno caracteriza-se pelos seguintes processos celulares: a fragmentação do DNA, a condensação da cromatina, a invaginação da membrana plasmática e a formação de vesículas (Gunawardena *et al.*, 2001a). Leite *et al.* (2017) verificaram resultados similares na MCP durante a formação do aerênquima nas raízes de cana-de-açúcar.

Foi observado que a sinalização de etileno tem uma importância na regulação da MCP durante a formação do aerênquima lisígeno em gramíneas (He *et al.*, 1996; Drew *et al.*, 2000). Este hormônio induz a produção de espécies reativas de oxigênio (ROS, *Reactive Oxygen Species*), que levam ao início do desenvolvimento do aerênquima na raiz de várias espécies como *O. sativa* (Parlanti *et al.*, 2011; Steffens *et al.*, 2011; Yamauchi *et al.*, 2017), *Z. mays* (Yamauchi *et al.*, 2011) e *T. aestivum* (Yamauchi *et al.*, 2014). Entre as moléculas que fazem parte do grupo de ROS estão o radical superóxido ($O_2^{\cdot-}$), o radical hidroxila ($OH^{\cdot}$), o peróxido de hidrogênio (H_2O_2) e o oxigênio singlete (1O_2) (Gechev *et al.*, 2006). Estas moléculas têm a capacidade de causar dano oxidativo nas moléculas de DNA, proteínas e lipídeos, e também levar à morte celular (Apel & Hirt, 2004; Mittler *et al.*, 2004).

A produção de ROS tem como enzima-chave a RBOH (*Respiratory burst oxidase homolog H*), a qual é uma homóloga do domínio catalítico da subunidade de gp91^{phox} da NADPH oxidase em mamíferos (Keller *et al.*, 1998). Ela tem como função converter O₂ a O₂⁻ e mediar a sinalização de ROS (Torres & Dangl, 2005). Estas enzimas são superexpressas durante a formação do aerênquima lisígeno em *Z. mays* (Rajhi *et al.*, 2011, Yamauchi *et al.*, 2011), *T. aestivum* (Yamauchi *et al.*, 2014) e em cana-de-açúcar (Tavares, 2015). Isto sugere que a produção de ROS mediado por RBOH pode ser essencial para a formação do aerênquima em gramíneas (Yamauchi *et al.*, 2017).

Segundo Gunawardena *et al.* (2001a), as paredes celulares permaneceram intactas durante os eventos de MCP durante a formação do aerênquima em *Z. mays*, e somente após o término deste estágio, a degradação da parede celular e a formação dos espaços de ar ocorreram. Diferentemente, em cana-de-açúcar a MCP e as modificações nas paredes celulares ocorrem quase que concomitantemente no córtex da raiz durante a formação do aerênquima (Leite *et al.*, 2017).

3.2. Modificações na parede celular

Outro aspecto durante o desenvolvimento deste tecido são as modificações nas paredes celulares (Yamauchi *et al.*, 2013). Vários autores já evidenciaram a presença de várias glicosil hidrolases incluindo pectinases, celulasas e xiloglucano transglicosilases que agem na parede celular das células que irão formar o aerênquima (He *et al.*, 1994; Saab & Sachs, 1996; Jackson & Armstrong, 1999; Rajhi *et al.*, 2011).

Foram identificadas modificações nas pectinas desesterificadas e esterificadas usando os anticorpos monoclonais JIM5 e JIM7, respectivamente, nas células corticais das raízes de *Z. mays* que apresentaram aerênquima (Gunawardena *et al.*, 2001b). Em cana-de-açúcar na

variedade SP80-3280, Leite *et al.* (2017) analisando os 5 segmentos (de um centímetro cada) mais apicais da raiz, observou que nos dois segmentos iniciais não há formação do aerênquima. O fenômeno é iniciado a partir do segmento 3, quando o GLM e uma pectina (homogalacturonano) presente nas junções triangulares na lamela média, são os primeiros polissacarídeos a sofrer modificações, e desaparecem ao longo da formação do aerênquima. Além disso, também foram identificadas várias proteínas como arabinofuranosidases, β -1,3-glucanases, celulasas, liquenases, pectinases e xilosidases durante este processo (Grandis, 2015), assim dando suporte a hidrólise do homogalacturonano na lamela média e o GLM durante o desenvolvimento do aerênquima em cana-de-açúcar.

Dados anteriores do nosso grupo mostraram em função dos níveis crescentes da expressão do transcrito da endopoligalacturonase I (*ScEPG1*) em direção ao segmento 5 na raiz de cana-de-açúcar, esta enzima está envolvida na hidrólise do homogalacturonano, e pode promover a perda de adesão celular (Tavares, 2015). No contexto das modificações observadas no córtex das raízes de cana-de-açúcar, foi sugerido que a hidrólise da pectina presente na lamela média no terceiro segmento da raiz levaria à diminuição da adesão celular, e ao aumento da porosidade do tecido e à expansão celular (Tavares, 2015). Assim, a atuação de pectinases e de outras proteínas na parede celular poderia ser a chave para as etapas iniciais de degradação da parede e, no contexto biotecnológico, teria o potencial de predispor a parede celular à hidrólise (Tavares *et al.*, 2015).

4. Cascata de sinalização de etileno

O hormônio etileno está envolvido na formação do aerênquima (Jackson & Armstrong, 1999; Drew *et al.*, 2000). Ele é sintetizado a partir da metionina, a qual é convertida a S-adenosil-metionina (SAM) pela S-AdoMet sintetase (SAM sintetase). A conversão de SAM a ácido 1-carboxílico-1-aminociclopropano (ACC) é feita pela enzima ACC sintase (ACS). Posteriormente, ACC é oxidado pela ACC oxidase (ACO) para formar o etileno (Wang *et al.*, 2002).

Uma vez sintetizado, o etileno se liga aos receptores com a ajuda de íons de cobre fornecido pela RAN1 (*RESPONSE TO ANTAGONIST1*) (Figura 3) (Hirayama *et al.*, 1999; Hall *et al.*, 2007). A ligação do hormônio aos receptores, leva à inativação do complexo receptor-CTR1 (*CONSTITUTIVE TRIPLE RESPONSE 1*), e o acúmulo de fatores de transcrição como EIN3 (*ETHYLENE INSENSITIVE3*) e EIN3-like (EIL) no núcleo (Guo & Ecker, 2003), que são dependentes da atividade de EIN2 (*ETHYLENE INSENSITIVE2*), um regulador positivo na sinalização de etileno (Alonso *et al.*, 1999; Qiao *et al.*, 2009). Na ausência de etileno, a interação entre os receptores e CTR1, reguladores negativos da via de etileno (Stepanova & Alonso, 2009), leva à inativação dos componentes de sinalização de EIN2 e EIN3 (Chao *et al.*, 1997; Alonso *et al.*, 1999). A transitoriedade entre os estados ativos e inativos de um dos receptores, ETR1 (*ETHYLENE RECEPTOR1*), é facilitada pela RTE1 (*REVERSION-TO-ETHYLENE SENSITIVITY1*) (Barry *et al.*, 2005).

Os receptores ligados a etileno são degradados via proteossomo (Stepanova & Alonso, 2009), enquanto os níveis das proteínas EIN2 e EIN3 são controlados pelas proteínas F-box específicas. As proteínas F-box ETP1 (*EIN2 TARGETING PROTEIN1*) e ETP2 (*EIN2 TARGETING PROTEIN2*) regulam a produção de EIN2, enquanto as proteínas F-box EBF1 (*EIN3-BINDING F-BOX PROTEIN1*) e EBF2 (*EIN3-BINDING F-BOX PROTEIN2*) regulam a produção de EIN3. As

proteínas F-box específicas são componentes do complexo SCF ubiquitina E3 ligase que degradam EIN2 e EIN3 na ausência do gás etileno (Guo & Ecker, 2003; Potuschak *et al.*, 2003; Qiao *et al.*, 2009).

A transdução do sinal da via de etileno do complexo receptor-CTR1 para os fatores de transcrição EIN3 e EIL1 (*ETHYLENE INSENSITIVE3- like 1*) é feita pela proteína transmembrana EIN2 (Alonso *et al.*, 1999; Qiao *et al.*, 2009). Já a regulação das proteínas F-box – EBF1 e EBF2 é feita pelo componente da via EIN5 (*ETHYLENE INSENSITIVE5*), é uma 5'-3' exoribonuclease, que promove a diminuição de mRNA destes genes alvos (Olmedo *et al.*, 2006; Potuschak *et al.* 2006), e permite o acúmulo de EIN3. A estabilidade desta última proteína é regulada através da fosforilação de resíduos específicos de Thr de EIN3 pela cascata MKK9-MPK3/6 (Yoo *et al.*, 2008).

Assim, na presença de etileno, as proteínas F-box EBF1 e EBF2 não conseguem degradar EIN3 e EIL, que passam a se acumular no núcleo, e ativam a expressão de genes alvos como fatores de transcrição ERF1 (*ETHYLENE RESPONSE FACTOR1*) (Solano *et al.*, 1998), EBF2 (Konishi & Yanagisawa, 2008) e EDF1-4 (*ETHYLENE RESPONSE DNA BINDING FACTOR1-4*). Estes últimos são também conhecidos como RAVs (*RELATED TO ABI3/VP1*) (Alonso *et al.*, 1999; Riechmann *et al.*, 2000). A expressão destes fatores de transcrição leva a uma cascata transcricional que resulta na ativação ou repressão de vários genes de resposta ao etileno (Stepanova & Alonso, 2009).

Figura 3. Via de sinalização de etileno. O hormônio liga-se aos receptores com a ajuda dos íons de cobre que são fornecidos pela RAN1. Os receptores e CTR1 agem como reguladores negativos da via na ausência de etileno, e reprimem EIN2 e 3. Os estados ativos e inativos dos receptores são facilitados pela RTE1. Na presença de etileno, EIN2 e 3 acumulam-se, mas são reguladas negativamente pelas proteínas F-box ETP1 e 2 (no caso EIN2) e EBF1 e 2 (no caso de EIN3), na ausência de etileno. Já a regulação negativa de EBF1 e 2 é feita pela exoribonuclease EIN5. EIN3 e EIL1 ativam a expressão de genes alvo como fatores de transcrição ERF1, EBF2 e EDF1-4. Retirado de Stepanova & Alonso (2009).

4.1. Fatores de transcrição RAV (*Related to ABI3/VP1*)

A família de fatores de transcrição RAV está envolvida na resposta ao etileno (Alonso *et al.*, 2003) e possivelmente na formação do aerênquima (Rajhi *et al.*, 2011). Essas proteínas fazem parte da família AP2/ERF ou AP2/EREBP (*APETALA2/ETHYLENE RESPONSIVE FACTOR* ou *ETHYLENE RESPONSIVE ELEMENT BINDING PROTEIN*) de fatores de transcrição (Mizoi *et al.*, 2011). Essa família é caracterizada pela presença de ao menos um domínio AP2/ERF (Jofuku *et al.*, 1994), e compreende três subfamílias. A subfamília AP2 (*APETALA2*) tem dois domínios AP2/ERF, a subfamília ERF ou EREBP tem um domínio AP2/ERF, e a subfamília RAV (*RELATED*

TO ABI3/VP1) tem um domínio AP2/ERF e um domínio B3 (Riechmann *et al.*, 2000; Sakuma *et al.*, 2002; Nakano *et al.*, 2006).

Em *Arabidopsis thaliana*, a subfamília RAV apresenta seis membros (Riechmann *et al.*, 2000). Quatro destes fatores de transcrição são conhecidos como EDF1 (AT1G25560), EDF2 (AT1G68840), EDF3 (AT3G25730) e EDF4 (AT1G13260) (*ETHYLENE RESPONSE DNA BINDING FACTORS*) e são proteínas que respondem ao hormônio etileno de acordo com Alonso *et al.* (2003), como pode ser visto na Figura 3. As proteínas EDF1 e EDF2/RAV2 são também conhecidas como TEM1 (*TEMPRANILLO1*) e TEM2 (*TEMPRANILLO2*), respectivamente, devido as suas funções na floração (Castillejo & Pelaz, 2008). Mas somente RAV1 (AT1G13260) e RAV2 (AT1G68840) foram funcionalmente caracterizados (Kagaya *et al.*, 1999).

As proteínas RAV têm a capacidade de se ligarem ao DNA e regularem a expressão de genes-alvo, através do reconhecimento de sequências específicas na região promotora destes genes (Kagaya *et al.*, 1999). O domínio AP2 se liga a um motivo CAACA e o domínio B3 se liga a um motivo CACCTG dos promotores dos seus genes regulados (Kagaya *et al.*, 1999). Os domínios AP2 e B3 têm atividade de ligação ao DNA de forma independente, mas a afinidade pela sequência alvo aumenta quando ambos os domínios se ligam ao DNA (Kagaya *et al.*, 1999).

A subfamília RAV pode ser tanto classificada na família AP2/ERF como na família B3 de fatores de transcrição (Matías-Hernández *et al.*, 2014). O domínio AP2 tem cerca 60 aminoácidos (Okamuro *et al.*, 1997; Riechmann & Meyerowitz, 1998; Riechmann *et al.*, 2000; Sakuma *et al.*, 2002; Magnani *et al.*, 2004), e foi primeiramente identificado em tabaco (*Nicotiana tabacum*) como domínio conservado de ligação ao DNA na família EREBPs (*Ethylene Response Element Binding Proteins*) (Ohme-Takagi & Shinshi, 1995), e posteriormente em *A. thaliana* na família AP2 (*APETALA2*) (Jofuku *et al.*, 1994). O domínio AP2/ERF contém uma α

hélice C-terminal e 3 folhas β N-terminais que são responsáveis por reconhecer a sequência alvo no DNA (Allen *et al.*, 1998) (Figura 4A). Há um grande número de proteínas que tem o domínio AP2 nas plantas (Weigel, 1995; Sakuma *et al.*, 2002). Em *A. thaliana*, ANT (*AINTEGUMENTA*), TINY e CBF1 (*C-REPEAT/DRE BINDING FACTOR 1*) estão envolvidos principalmente no desenvolvimento floral, na proliferação celular, na resposta a hormônios e a diferentes estresses nas plantas (Jofuku *et al.*, 1994; Wilson *et al.*, 1996; Jaglo-Ottosen *et al.*, 1998).

O domínio B3 consiste em 7 folhas β dispostas em *open barrel* e duas α hélices curtas, que geralmente compartilham uma estrutura comum de reconhecimento do DNA (Yamasaki *et al.*, 2004; Waltner *et al.*, 2005) (Figura 4B). Este domínio apresenta cerca de 110 aminoácidos (Yamasaki *et al.*, 2013), e foi inicialmente identificado na proteína VP1 (*VIVIPAROUS1*) de *Z. mays* (McCarty *et al.*, 1991), e no seu ortólogo de VP1 em *A. thaliana* ABI3 (*ABSCISIC ACID INSENSITIVE3*) (Giraudat *et al.*, 1992). Há três grandes classes de fatores de transcrição que contêm o domínio B3, que são ABI3, RAV e ARFs (*AUXIN RESPONSE FACTORS*) (Riechmann *et al.*, 2000). As proteínas que contêm o domínio B3 estão associadas à resposta ao ácido abscísico e a auxina, e desempenham um papel importante no crescimento vegetal e na maturação da semente (McCarty *et al.*, 1989; Ulmasov *et al.*, 1997).

Dentre os fatores de transcrição da superfamília B3, encontraram-se os membros que contêm atividade repressiva devido ao fato de apresentar as sequências motivo RLFGV/KLFGV, o qual é nomeada como domínio repressor B3 (BRD) (Ikeda & Ohme-Takagi, 2009). Em *A. thaliana*, há 42 genes que apresentam a sequência motivo para RLFGV ou KLFGV, destes 29 são genes que codificam para fatores de transcrição, entre eles, estão os membros das famílias ARF, RAV, Hsf, MYB, JUMONJI e GARP (Ikeda & Ohme-Takagi, 2009). Na família RAV há 4 membros que compartilham o domínio BRD - TEM1, TEM2, RAV1 e RAV1-like (Ikeda

& Ohme-Takagi, 2009), uma sequência similar, MLFGV, está presente em RAV3 e RAV3-like (Causier *et al.*, 2012). Isso sugere que os fatores de transcrição RAV podem atuar como repressores transcripcionais dos seus genes alvos (Matías-Hernández *et al.*, 2014).

Figura 4. Modelo da estrutura do domínio AP2 de AtERF1 (AT4G17500) (A) mostrando uma α hélice C-terminal (em vermelho) e 3 folhas β N-terminais (em amarelo) (Allen *et al.*, 1998; código PDB 1gcc), e a estrutura do domínio B3 de RAV1 (AT1G13260) (B) com as 7 folhas β arranjadas em *open barrel* (em amarelo) e as duas α hélices (em vermelho) (Yamasaki *et al.*, 2013; código PDB 1wid), ambas as estruturas estão em complexo com o DNA (em cinza). Retirado de Yamasaki (2015).

4.1.1. Papéis do fator de transcrição RAV nas plantas

Dentre os fatores de transcrição da família RAV, RAV1 é o mais descrito na literatura (Alonso *et al.*, 1999). Estes fatores de transcrição podem responder a diferentes hormônios. De acordo com Alonso *et al.* (2003) essas proteínas são importantes na resposta ao etileno. O experimento usando tratamento com ACC (um precursor na biossíntese de etileno) e também com metil jasmonato (MJ) observou um aumento da expressão dos genes de RAVs em *A. thaliana* (Woo *et al.*, 2010). Por outro lado, a expressão de RAV1 diminui em resposta a um hormônio esteroide epibrassinolídeo (classe dos brassinoesteróides) (Hu *et al.*, 2004). Mais recentemente, observou-se que RAV1 tem um papel importante na sinalização de ácido

abscísico (ABA) durante a germinação de sementes e no desenvolvimento precoce de plântulas (Feng *et al.*, 2014).

Em experimentos relacionados a diferentes níveis de expressão deste fator de transcrição, a superexpressão de RAV1 em *A. thaliana* retardou a produção de raízes laterais e o desenvolvimento foliar em roseta, enquanto o silenciamento deste fator de transcrição causou a floração adiantada, levando a sugerir que RAV1 funciona como um regulador negativo no crescimento e no desenvolvimento vegetal (Hu *et al.*, 2004). Já a superexpressão de GmRAV levou ao nanismo, redução no alongamento da raiz, atraso no tempo de floração, baixa taxa fotossintética e no conteúdo de clorofila nas folhas, sugerindo a participação deste fator de transcrição na fotossíntese e no crescimento do vegetal (Zhao *et al.*, 2008).

Os fatores de transcrição RAV1 e RAV2 também participam em vários tipos de estresses bióticos e abióticos como, por exemplo, na resposta à baixa temperatura (Fowler *et al.*, 2005), à defesa a patógenos, à seca, ao estresse salino (Sohn *et al.*, 2006) e à falta de luz (Lee *et al.*, 2005).

Além disso, RAV1 pode estar envolvido em processos de MCP nas plantas. De acordo com Woo *et al.* (2010), RAV1 pode agir como regulador positivo da senescência foliar. Similarmente, em *Z. mays*, RAV1L (RAV1-Like) homólogo de RAV1 pode estar envolvido na formação do aerênquima, pois o seu transcrito é mais abundante em células corticais da raiz sob condições que favorecem a formação de aerênquima (Rajhi *et al.*, 2011). Já em cana-de-açúcar na variedade SP80-3280, observou-se que os transcritos referentes à *ScRAV1* são diferencialmente expressos em amostras com formação de aerênquima contrastante, atingindo menores níveis quando do favorecimento do desenvolvimento deste tecido (Tavares, 2015).

Em relação a *ScRAV1*, Tavares *et al.* (2019) demonstrou em experimentos de ensaio de transativação realizados com *Nicotiana benthamiana*, que este fator de transcrição pode apresentar efeito repressor sobre o promotor de *ScEPG1*, sugerindo que o fator de transcrição RAV pode modular a expressão do gene que codifica para uma endopoligacturonase. De fato, ambos os genes parecem ter um papel importante durante a formação do aerênquima, uma vez que um dos primeiros eventos relacionados à parede celular durante este processo é o ataque às pectinas da lamela média no córtex das raízes de cana-de-açúcar (Leite *et al.*, 2017). Além disso, também foi observado que conforme os níveis do transcrito de *ScRAV1* decaem, o efeito repressor sobre *ScEPG1* possivelmente torna-se mais brando e, consequentemente, a digestão da pectina presente na lamela média pode ocorrer (Tavares *et al.*, 2019). De acordo com os autores, *ScRAV1* atuaria como regulador negativo dos primeiros passos da formação do aerênquima.

Devido a alguns estudos mostrarem que membros da família dos fatores de transcrição RAV podem estar envolvidos na formação do aerênquima em algumas gramíneas, especialmente a cana-de-açúcar, planta com importante papel em bioenergia, este trabalho analisou dados do transcriptoma (RNA-Seq) dos quatro segmentos da raiz de cana-de-açúcar durante a formação do aerênquima visando a identificação de transcritos alvos da regulação por *ScRAV1* dentre genes que codificam para enzimas de biossíntese ou de degradação de parede celular, bem como para fatores de transcrição importantes neste processo. Os pontos de regulação de *ScRAV1* preditos neste trabalho abrem caminho para análises funcionais que ampliarão o conhecimento sobre a regulação da expressão gênica subjacente à degradação da parede celular. A compreensão destes mecanismos pode ser de grande valia para o entendimento de como *ScRAV1* modula as modificações da parede celular, o que consiste em uma importante ferramenta para a tecnologia do bioetanol 2G.

II. OBJETIVOS

1. Objetivo geral

- Identificar o papel de *ScRAV1* na regulação transcricional relacionado à formação do aerênquima lisígeno em raízes de cana-de-açúcar na variedade SP80-3280.

2. Objetivos específicos

- Construir a árvore filogenética do fator de transcrição RAV em plantas;
- Identificar os domínios conservados da proteína de *ScRAV1*;
- Construir redes de co-expressão e de regulação gênica de *ScRAV1* com os transcritos de parede celular e fatores de transcrição nos quatro segmentos da raiz de cana-de-açúcar;
- Identificar os possíveis homólogos dos transcritos de parede celular e fatores de transcrição em *S. bicolor* e *A. thaliana*;
- Identificar elementos *cis* nas regiões promotoras de genes de parede celular e fatores de transcrição com base nas sequências homólogas em *S. bicolor*;
- Avaliar o perfil de expressão dos transcritos de parede celular e fatores de transcrição por RNA-Seq obtidas nas redes de regulação gênica;
- Apontar potenciais transcritos relacionados com a formação do aerênquima.

III. MATERIAIS E MÉTODOS

1. Árvore filogenética do fator de transcrição RAV

Para construir a árvore filogenética, a sequência completa de *ScRAV1* presente no BAC (cromossomo artificial bacteriano) SHCRBa_021_A10 (Tavares *et al.*, 2019), foi utilizada como isca para buscar outros transcritos no programa BLASTN no banco de ESTs (*Expressed Sequence Tags*) de cana-de-açúcar no Sucest (<http://www.sucest-fun.org/>; Vettore *et al.*, 2003), para aquelas sequências que tiveram $\geq 80\%$ de similaridade. As ESTs obtidas neste programa foram avaliadas quanto a sua presença dentro do SAS (*Sugarcane Assembled Sequences*), e somente as sequências que tiveram os maiores SAS (sequência de proteína com o maior número de aminoácidos) foram usadas para as análises posteriores.

Em seguida, a sequência de *ScRAV1* foi utilizada como isca para buscar os seus possíveis homólogos em várias espécies de plantas dos bancos de dados do Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html/>; Goodstein *et al.*, 2012) e do PlantTFDB 4.0 (<http://planttfdb.cbi.pku.edu.cn/>; Jin *et al.*, 2017), utilizando os parâmetros padrões do BLASTP. Nestes bancos de dados, foram selecionadas somente as sequências que apresentaram os domínios AP2 e B3 característicos do fator de transcrição RAV. A localização dos domínios AP2 e B3 característicos do fator de transcrição RAV foi identificada utilizando a ferramenta Pfam (<http://pfam.xfam.org/>; Finn *et al.*, 2016) (Apêndice A). Além disso, observou-se a presença de sequências duplicadas ou incompletas, e somente aquelas que apresentaram o maior número de aminoácidos foram utilizadas para construir a árvore filogenética.

Por último, as sequências de proteína foram alinhadas na ferramenta MAFFT (<http://mafft.cbrc.jp/alignment/server/>; Katoh & Standley, 2013). O alinhamento gerado

neste programa foi usado na ferramenta IQ-TREE (<http://iqtree.cibiv.univie.ac.at/>; Trifinopoulos *et al.*, 2016), o qual usou os valores de *ultrasfast bootstrap* de 1000 replicatas (Minh *et al.*, 2013) e o melhor modelo de máxima verossimilhança (JTT+I+G4) para construir a árvore filogenética para a família de fator de transcrição RAV. Para a visualização da árvore filogenética foi usada a ferramenta FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>; Rambaut, 2016).

2. Análise dos domínios proteicos de *ScRAV1*

Para localizar os domínios conservados e de sinal de localização nuclear de *ScRAV1* (SHCRBa_021_A10) (Tavares *et al.*, 2019), foram utilizadas as ferramentas Pfam (<http://pfam.xfam.org/>; Finn *et al.*, 2016) e NSL Mapper (Kosugi *et al.*, 2009), respectivamente. Essas informações foram usadas após gerar o alinhamento múltiplo com as sequências ortólogas nas espécies de *M. sinensis* (*MisinRAV1*, Misin06G229100.1; *MisinRAV1.1*, MisinT241800.1), *S. bicolor* (*SbRAV1*, SOBIC.003G265200.1), *Z. mays* (*ZmRAV1L*, GRMZM2G169654_T01), *O. sativa* (*OsRAV1*, LOC_OS01G49830.1), *A. thaliana* (EDF1/TEM1, AT1G25560.1; EDF2/TEM2, AT1G68840; EDF3, AT3G25730; EDF4/RAV1, AT1G13260), e que foram obtidas conforme o parágrafo 2 do item 1. O alinhamento múltiplo foi gerado na ferramenta MAFFT (<https://mafft.cbrc.jp/alignment/server/>; Katoh & Standley, 2013), e visualizado no programa BioEdit (Hall, 1999).

3. Redes de co-expressão

O Laboratório de Fisiologia Ecológica de Plantas (LAFIECO) conta com um banco de transcritos proveniente do sequenciamento do transcriptoma por RNA-Seq de três réplicas biológicas para cada um dos quatro segmentos de 1 cm cada (S1 a S4) da porção mais apical da raiz, na plataforma ABI SOLiD. A montagem deste banco de dados, bem como as análises

gerais do RNA-Seq das raízes de cana-de-açúcar durante a formação do aerênquima encontram-se descritas na tese de doutorado de Amanda Piovezani (Piovezani, 2017). Esses dados foram utilizados para a construção de redes de co-expressão descritas abaixo:

3.1. Construção das redes de co-expressão

As redes de co-expressão são redes aonde as arestas representam uma relação estatística (correlação de Pearson) entre os valores de expressão dos transcritos. Neste trabalho, foi utilizado o pacote básico do software R para realizar os cálculos de correlações. Nas redes, os nós representam os transcritos de parede celular ou fatores de transcrição anotados no RNA-Seq que tiveram expressão em todas as 3 réplicas biológicas em cada segmento da raiz de cana-de-açúcar, aqueles que não apresentaram expressão em alguma das réplicas biológicas foram excluídos da análise (Figura 5). Os níveis de expressão dos transcritos no RNA-Seq foram dados em TPM (*Transcript per Million*), e foram normalizados em $\log_2$ para cada uma das réplicas biológicas de cada segmento da raiz da cana-de-açúcar antes de fazer as redes de co-expressão. A Figura 5 ainda mostra a quantidade de transcritos de parede celular e fatores de transcrição que são comuns aos quatro segmentos da raiz de cana-de-açúcar presentes no RNA-Seq.

Figura 5. Diagrama de Venn dos transcritos de parede celular (A) e fatores de transcrição (B) que são comuns aos S1 ao S4, e que apresentaram todas as 3 réplicas biológicas no RNA-Seq em cada segmento da raiz de cana-de-açúcar.

Para selecionar os nós das redes de *ScRAV1* utilizou-se como isca o SAS referente como SCAGLR2011B02.g (*ScRAV1*) no Sucest (<http://www.sucest-fun.org/>; Vettore *et al.*, 2003) e no RNA-Seq da raiz de cana-de-açúcar. As arestas das redes de co-expressão foram feitas somente com os transcritos de parede celular e fatores de transcrição que estavam significativamente correlacionados com *ScRAV1* (SCAGLR2011B02.g) e que apresentaram um $p\text{-valor} \leq 0,05$ (Apêndice B e C). Os nós que não possuíram nenhuma conexão com o transcrito *ScRAV1* foram excluídos das redes.

Ao final, foram geradas oito redes – sendo quatro redes de correlação de *ScRAV1* com os transcritos de parede celular e quatro com os fatores de transcrição, uma para cada segmento da raiz. Para a visualização das redes de co-expressão foi utilizado o programa Cytoscape (<http://www.cytoscape.org/>, Shannon *et al.*, 2003). Os transcritos de parede celular foram classificados manualmente em categorias de acordo com o seu papel funcional na parede celular, utilizando o procedimento descrito por Piovezani (2017), com modificações.

3.2. Filtro das redes de co-expressão através da centralidade de grau

Para filtrar os transcritos de parede celular e fatores de transcrição presentes nas redes de *ScRAV1*, analisaram-se as 10 maiores centralidades de grau de cada uma das redes utilizando o pacote Network Analyzer (Assenov *et al.*, 2008). A centralidade de grau é definida como o número de conexões que um nó forma dentro da rede (Pavlopoulos *et al.*, 2011). A partir da centralidade de grau é possível destacar os nós que estão envolvidos em um grande número de interações. Assim, é possível identificar os nós que podem ser funcionalmente importantes dentro de uma rede (Kranthi *et al.*, 2016). Neste trabalho, a centralidade de *ScRAV1* não foi calculada, pois ele já foi usado como isca para gerar as redes de co-expressão.

Assim, os transcritos de parede celular e fatores de transcrição subsequentes foram analisados quanto à sua centralidade de grau em cada uma das redes de cada segmento da raiz de cana-de-açúcar.

3.3. Anotação dos transcritos de parede celular e fatores de transcrição de cana-de-açúcar em *S. bicolor* e *A. thaliana*

As sequências de proteínas dos fatores de transcrição e os transcritos relacionados ao metabolismo de parede celular selecionados como as 10 maiores centralidades de grau em cada uma das redes de *ScRAV1* foram inicialmente anotadas em *S. bicolor*, pois é uma das espécies mais próximas da cana-de-açúcar e com o genoma completo disponível em bancos de dados online (Paterson *et al.*, 2009). Para esta anotação buscou-se o melhor *hit* desta espécie para cada sequência alvo de cana-de-açúcar (SAS – *Sugarcane Assembled Sequences*), utilizando a plataforma do Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html/>; Goodstein *et al.*, 2012).

Logo após, verificou-se a porcentagem de identidade e o E-value das sequências de proteínas de cana-de-açúcar com os melhores *hits* de *S. bicolor*. Para isto, foi feito um alinhamento local utilizando o BLASTP (<http://blast.ncbi.nlm.nih.gov/>; Altschul *et al.*, 1990) das sequências de cana-de-açúcar contra as sequências de *S. bicolor*.

Para análises posteriores foram consideradas somente as sequências de proteína de cana-de-açúcar que apresentaram valores $\geq 90\%$ de identidade com as sequências de *S. bicolor*. Foram excluídas da análise as sequências de cana-de-açúcar que não apresentaram nenhum *hit* ou que apresentaram duas, ou mais sequências para o mesmo *hit* da sequência de *S. bicolor*, neste último caso foi mantido o maior SAS.

Depois, prosseguiu-se para a anotação dos transcritos de parede celular e fatores de transcrição em *A. thaliana*, pois esta espécie é uma planta modelo (Meinke *et al.*, 1998) e tem uma grande quantidade de dados disponíveis sobre a função das proteínas (Rhee & Mutwil, 2014), principalmente quando comparado com cana-de-açúcar e *S. bicolor*. Para a anotação, foram utilizadas as sequências de proteínas de *S. bicolor* como iscas para buscar os melhores *hits* em *A. thaliana*, utilizando a plataforma do Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html/>; Goodstein *et al.*, 2012). Para obter as porcentagens de identidade e E-value dos melhores *hits* de *A. thaliana* com base nas sequências de *S. bicolor*, seguiu-se o procedimento como foi descrito no segundo parágrafo desta seção.

3.4. Enriquecimento funcional dos transcritos de parede celular e fatores de transcrição através dos termos do Gene Ontology (GO)

A partir da anotação dos transcritos de parede celular e fatores de transcrição em *A. thaliana* no item 3.3, buscaram-se manualmente os termos GOs pela plataforma UniProt (<http://www.uniprot.org/>; The UniProt Consortium, 2017), para verificar os possíveis papéis ou funções já descritas destas proteínas nesta planta. Os termos GOs foram carregados no programa Blast2GO (Götz *et al.*, 2008), o qual permitiu a validação (retirada dos termos redundantes) e a contagem dos termos para cada uma das categorias: componente celular, processo biológico e função molecular. Neste trabalho, foram avaliados apenas os 20 primeiros termos GOs para cada uma das categorias tanto para os transcritos de parede celular como para os fatores de transcrição que foram anotados em *A. thaliana*. Todos os termos GOs encontrados para os genes de parede celular e fatores de transcrição em *A. thaliana* estão nos Apêndices D e E, respectivamente.

3.5. Predição *in silico* de sítios de ligação de RAV em *S. bicolor*

Para avaliar uma possível regulação transcricional de *ScRAV1* com os transcritos de parede celular e fatores de transcrição de cana-de-açúcar, utilizou-se a sequência ortóloga deste fator de transcrição em *S. bicolor* - *SbRAV1* (parágrafo 2, item 1). Depois, buscaram-se possíveis elementos *cis* para RAV nas regiões promotoras das sequências preditas em *S. bicolor* (item 3.3 do parágrafo 3), devido a falta de genoma completo em cana-de-açúcar. A região promotora foi definida como uma sequência 2 Kb à montante do ATG. Tais sequências foram utilizadas na plataforma PLACE (<https://www.dna.affrc.go.jp/htdocs/PLACE/>; Higo *et al.*, 1999), para buscar os motivos reconhecidos pelos domínios AP2 e B3 do fator de transcrição RAV (Figura 6).

Figura 6. Representação esquemática da região promotora com 2 Kb à montante do ATG. O sítio de início da transcrição está representado como +1. O quadrado verde, amarelo e laranja referem-se a possíveis elementos *cis* aos quais RAV pode se ligar na região promotora dos genes anotados em *S. bicolor*.

3.6. Construção das redes de regulação gênica

Nas redes de regulação gênica (GRN, *Gene Regulatory Network*), os nós representam os genes de parede celular ou fatores de transcrição e as arestas representam uma possível regulação de *SbRAV1*. As redes de regulação gênica foram feitas com base na análise do domínio proteico de *SbRAV1* e para aqueles genes que apresentaram elementos *cis* para possível ligação de RAV nas regiões promotoras das sequências de *S. bicolor*, conforme descrito nos itens 2 e 3.5, respectivamente. Para identificar os homólogos em *S. bicolor* para os transcritos de parede celular e para os fatores de transcrição de cana-de-açúcar, trocou-se

a nomenclatura utilizada em cana-de-açúcar por apenas de *Sc* por *Sb* para gerar as redes regulatórias gênicas. A visualização das redes de regulação gênica foi feita no Cytoscape (<http://www.cytoscape.org/>, Shannon *et al.*, 2003). Além disso, os genes de parede celular e os fatores de transcrição foram classificados conforme Piovezani (2017) (com modificações) e manualmente de acordo com os termos GOs para *A. thaliana* (item 3.4), respectivamente.

3.7. Avaliação do perfil de expressão dos transcritos

Foram avaliados os perfis de expressão dos transcritos de parede celular e dos fatores de transcrição em cana-de-açúcar, para os seus genes homólogos em *S. bicolor* que estão presentes nas redes regulatórias gênicas de *SbRAV1* (item 3.6), e também para o transcrito de *ScRAV1*. A expressão destes transcritos foi visualizada através de heatmaps, os quais usaram os valores de expressão do RNA-Seq em TPM (*Transcript per Million*) em $\log_2$ da média de três réplicas biológicas de cada segmento da raiz da cana-de-açúcar, utilizando a função `heatmap.2` do pacote `gplots` na linguagem R (Warnes *et al.*, 2009). Para a clusterização hierárquica do perfil de expressão dos transcritos foi usado o método Ward.D2 (Murtagh & Legendre, 2014) e a distância Euclidiana. A expressão foi dada em Row Z-score, o qual é o número de desvios-padrão de um dado valor em relação à média.

3.8. Seleção dos transcritos de parede celular e fatores de transcrição

Foram selecionados quatro transcritos de parede celular que codificam glicosil hidrolases - duas que atacam a pectina (*ScPAE2*, *ScPGase1*) e duas que atacam a celulose ou a hemicelulose (*ScGH3 2* e *ScEGase4*), e para os fatores de transcrição foram selecionados dois transcritos relacionados à sinalização de auxina (*ScARF1*, *ScARF3*) e dois associados à MCP (*ScNAC2* e *ScNAC8*).

Estes transcritos de parede celular e fatores de transcrição foram escolhidos com base nos trabalhos do nosso grupo, que mostraram que no aerênquima de cana-de-açúcar ocorre separação celular na lamela média, degradação do GLM e MCP (Leite *et al.*, 2017). Recentemente, demonstrou-se por ensaio de transativação em *N. benthamiana* que *ScRAV1* pode regular negativamente *ScEPG1*, uma endopoligalacturonase, que age possivelmente na lamela média no córtex da raiz durante a formação do aerênquima (Tavares *et al.*, 2019). Além disso, um outro trabalho do grupo aponta que a auxina pode ter um papel importante durante a formação do aerênquima constitutivo na raiz de cana-de-açúcar (Tavares *et al.*, 2018). Também foi levado em consideração os transcritos de cana-de-açúcar que apresentaram homólogos em *A. thaliana* com funções em processos similares descritos na literatura referente ao que já foi observado no aerênquima de cana-de-açúcar em trabalhos anteriores. Foram excluídos os transcritos de parede celular (*ScIFR/PrR2*, *ScF5H/F3H1*) e o fator de transcrição (*ScABI2*), pois apresentaram uma anotação em cana-de-açúcar, e apresentou outra em seus homólogos de *A. thaliana*.

Além disso, os transcritos de parede celular relacionados à biossíntese de lignina (*ScCAD1*, *ScCCoAOMT4*, *ScCCR2*, *ScCOMT2*, *ScC3H1*, *ScC4H1* e *ScPAL2*) e os fatores de transcrição associados à parede secundária (*ScNAC3*, *ScNAC5*, *ScNAC7*, *ScHB6* e *ScMYB2*) foram selecionados devido ao padrão consistente destes transcritos ao longo da metodologia deste trabalho, além da lignina ter um papel fundamental para o desenvolvimento do bioetanol de 2G.

A estratégia utilizada para a construção de redes de correlação e as análises posteriores com os transcritos de parede celular e fatores de transcrição encontra-se descrita na Figura 7.

Figura 7. Fluxograma para a busca de transcritos candidatos na formação do aerênquima em cana-de-açúcar. A metodologia iniciou-se com a construção das redes de co-expressão de *ScRAV1* com os transcritos de parede celular e fatores de transcrição para cada um dos segmentos (S1 a S4) da raiz a partir do banco de dados do RNA-Seq. Depois, foram escolhidos apenas as 10 maiores centralidades de grau de cada uma das redes (S1 ao S4) tanto para os transcritos de parede celular quanto para os fatores de transcrição. Logo depois, somente as sequências de cana-de-açúcar que obtiveram $\geq 90\%$ de identidade (ID) com as sequências de *S. bicolor* foram anotadas nesta espécie. As sequências de *S. bicolor* foram anotadas em *A. thaliana* para permitir o acesso aos termos do Gene Ontology (GO) e a investigação na literatura. Posteriormente, foi feita a predição *in silico* dos sítios de ligação de RAV nos genes homólogos de *S. bicolor*. Em paralelo, foi construída a árvore filogenética do fator de transcrição RAV em plantas e analisado os domínios conservados de *ScRAV1* junto as espécies *M. sinensis*, *S. bicolor*, *Z. mays*, *O. sativa* e *A. thaliana*. Para construir as redes de regulação gênica dos genes de parede celular e fatores de transcrição utilizou-se das informações da predição *in silico* dos elementos *cis* para RAV nos genes homólogos de *S. bicolor* e da caracterização dos domínios conservados para RAV. A análise do perfil de expressão dos transcritos de parede celular e para fatores de transcrição foi avaliada pelo heatmap através dos valores obtidos no RNA-Seq da raiz de cana-de-açúcar. Com base na literatura para os genes anotados em *A. thaliana*, foram escolhidos os transcritos em cana-de-açúcar que possivelmente podem estar envolvidos na formação do aerênquima.

IV. RESULTADOS

1. Filogenia do fator de transcrição RAV

A árvore filogenética do fator de transcrição RAV foi feita com a finalidade de encontrar os principais ortólogos de *ScRAV1* nas plantas. A Figura 8 mostra que há dois clados para o grupo de monocotiledôneas e dois para as eudicotiledôneas. Já para o grupo das algas (em retângulo amarelo) foi apenas encontrada uma única sequência para o fator de transcrição RAV, e foi utilizada como *outgroup* na filogenia (Figura 8).

Figura 8. Filogenia do fator de transcrição RAV em monotilédôneas e eudicotilédôneas. Para *outgroup* da árvore foi utilizada a sequência (KFL00094) referente à espécie *Klebsormidium flaccidum*. A árvore filogenética foi construída com o programa IQ-Tree com *ultrabostrap* de 1000 replicatas, valores ≥ 95 de *ultrabostrap* são confiáveis. Os asteriscos (*) correspondem aos possíveis ortólogos de *ScRAV1* nestas espécies. Os códigos das sequências são referentes a diferentes espécies usadas nesta filogenia, como pode ser visto no Apêndice A.

Em monocotilédôneas, principalmente no que se refere às gramíneas, podemos verificar que foram encontradas 4 proteínas características do fator de transcrição RAV em cana-de-açúcar e seus ortólogos em várias espécies que estão representados nos clados I, II, III e IV da árvore filogenética (Figura 8).

No clado III e IV que pertence as eudicotiledôneas, podem ser vistos os possíveis ortólogos de *ScRAVs* em diferentes espécies para aqueles observados em cana-de-açúcar. Dentro do clado III, podemos observar que *A. thaliana* obteve quatro sequências para o fator de transcrição RAV (Figura 8).

Além disso, em asterisco estão os possíveis ortólogos de *ScRAV1* em gramíneas e em *A. thaliana* (Figura 8). Esta informação será usada no item 2.

2. Caracterização dos domínios proteicos de *ScRAV1*

Os domínios conservados de *ScRAV1* foram analisados através do alinhamento múltiplo das possíveis sequências ortólogas de *M. sinensis*, *S. bicolor*, *Z. mays*, *O. sativa* e *A. thaliana* (em asterisco na Figura 8) (Figura 9).

Pode-se verificar que os domínios AP2 e B3 e o motivo BRD são altamente conservados nestas espécies. Parte da informação deste dado será posteriormente usada nos itens 3.6 e 4.6.

3. Redes de co-expressão de *ScRAV1* com os transcritos de parede celular

Foram feitas 4 redes de co-expressão do fator de transcrição *ScRAV1* com os transcritos de parede celular para cada segmento (S1 a S4) da raiz de cana-de-açúcar (Figura 10), os valores de correlação podem ser vistos no Apêndice B. As redes foram feitas visando entender se *ScRAV1* tem relação com a modificação da parede celular durante a formação do aerênquima na raiz de cana-de-açúcar.

Na Figura 10, pode-se observar que *ScRAV1* está correlacionado com um número maior de transcritos de parede celular nas redes dos S1 e S4 (Figura 10A e D), enquanto nas redes dos S2 e S3 (Figura 10B e C) apresentam um número menor de correlações. A quantidade e a porcentagem de correlações positivas e negativas em cada uma das redes de co-expressão dos transcritos de parede com *ScRAV1* também podem ser vistas na Tabela 1.

Tabela 1. Quantidades e porcentagens de correlações positivas e negativas de *ScRAV1* com os transcritos de parede celular nas redes de cada segmento (S1 ao S4) da raiz de cana-de-açúcar.

Redes dos Segmentos	Nº de Correlações Positivas	Porcentagem (%)	Nº de Correlações Negativas	Porcentagem (%)	Total
S1	46	85,19	8	14,81	54
S2	2	9,52	19	90,48	21
S3	12	57,14	9	42,86	21
S4	21	65,63	11	34,37	32
Total	81		47		128

Em S1 e S4, as redes relacionadas ao *ScRAV1* apresentam um maior número de correlações positivas com os transcritos de parede celular. Em S2, a rede relacionada à *ScRAV1* apresenta uma maior parte de correlações negativas. A rede do S3 mostrou-se com maior

equilíbrio entre as correlações positivas e negativas (Figura 10 e Tabela 1). Isto indica que há uma variabilidade do padrão do tipo de correlação de *ScRAV1* com os transcritos de parede celular nas redes de cada segmento da raiz de cana-de-açúcar.

No que concerne à formação do aerênquima, processo que ocorre principalmente nos S2 e S3 (Figura 10B e C), as principais características são redes menos densas quando comparadas com as redes dos S1 e S4 (Figura 10A e D).

Figura 10. Redes de co-expressão de *ScRAV1* (círculo em cor de vermelho escuro) com os transcritos de parede celular nos diferentes segmentos da raiz da cana-de-açúcar, feitas com a correlação de Pearson e com o corte de p-valor $\leq 0,05$. Rede do S1 (A), Rede do S2 (B), Rede do S3 (C) e Rede do S4 (D). A legenda à direita mostra as cores para as diferentes categorias (círculos menores) e para os tipos de correlação (linhas - vermelha escuro representa correlações negativas e rosa claro as correlações positivas) observadas para os transcritos presentes nas diferentes redes de *ScRAV1*. A identificação dos transcritos de parede celular correlacionados com *ScRAV1* em cada segmento da raiz, e seus respectivos valores de correlação e p-valores encontram-se no Apêndice B.

3.1. Categorização dos transcritos de parede celular presentes nas redes de co-expressão de *ScRAV1*

A fim de entender o possível significado de cada uma das redes de co-expressão de *ScRAV1* com os transcritos de parede celular (Figura 10), os transcritos de parede presentes em cada uma das redes de co-expressão com *ScRAV1* foram categorizados de acordo com suas funções na parede celular, ou seja biossíntese, hidrólise e outros, conforme a Figura 11.

Na análise que se segue, as correlações são vistas como relações entre *ScRAV1* e grupos de genes relacionados à parede celular. Nesta análise não se consideram correlações positivas ou negativas, mas somente a existência de correlações significativas.

Figura 11. Distribuição das categorias dos transcritos de parede celular em biossíntese, hidrólise e outros, presentes nas redes de *ScRAV1* nos S1 (A), S2 (B), S3 (C) e S4 (D) na raiz de cana-de-açúcar.

Na rede do S1, a maioria dos transcritos de parede celular co-expressos com *ScRAV1* se destacam principalmente os transcritos que pertencem a biossíntese de lignina (20,37%) e ao metabolismo de carboidratos (14,81%) (Figura 11A). No total, os transcritos de biossíntese de polissacarídeos representam 27,78% versus 24,07% para os transcritos que pertencem à hidrólise de polissacarídeos presentes na parede celular. Dentre os membros da categoria de biossíntese de polissacarídeos observa-se maior representatividade para os transcritos de biossíntese de hemicelulose e hemicelulose/pectina/outros. Já aqueles referentes à hidrólise verifica-se uma maior proporção de transcritos de degradação de hemicelulose e pectina.

Já na rede do S2, *ScRAV1* é co-expresso predominantemente com transcritos de parede celular relacionados à biossíntese de lignina (19,05%) e à hidrólise de hemicelulose (19,05%) e celulose/hemicelulose (14,29%) (Figura 11B). Comparando a porcentagem dos transcritos associados à biossíntese e à hidrólise de polissacarídeos, pode-se verificar que estes transcritos apresentam apenas 14,29% em relação aos transcritos de hidrólise que apresentam cerca de 47,62%. Dos transcritos referentes à biossíntese está a celulose, hemicelulose e hemicelulose/pectina com porcentagens iguais, enquanto para os transcritos de degradação de polissacarídeos de parede celular, destacam-se os transcritos que hidrolisam celulose e hemicelulose, como visto anteriormente.

Na rede do S3, destaca-se a co-expressão de *ScRAV1* com transcritos de biossíntese de lignina (19,05%) (Figura 11C). Além disso, pode-se verificar que a porcentagem total de transcritos que representam a biossíntese é um pouco menor em relação aos transcritos de hidrólise de polissacarídeos, os quais obtiveram-se 28,57% e 33,33%, respectivamente. Dentre os transcritos de biossíntese que apresentam as maiores porcentagens estão os de hemicelulose e pectina, enquanto para os transcritos relacionados à hidrólise da parede celular estão os de hemicelulose, pectina e quitina.

Por último, na rede do S4 de *ScRAV1* estão presentes em maior proporção os transcritos relacionados à biossíntese de hemicelulose (18,75%), o metabolismo de carboidratos (18,75%) e biossíntese de lignina (12,50%) (Figura 11D). No que se refere à porcentagem de transcritos relacionados à biossíntese e à hidrólise, observa-se que os transcritos associados à degradação de polissacarídeos sobrepõem aos transcritos de biossíntese de parede celular, (31,25% e 28,13%, respectivamente). Dos transcritos de hidrólise se destaca aqueles referentes à hidrólise de celulose, hemicelulose e pectina, enquanto aqueles referentes à biossíntese de polissacarídeos são maiores para biossíntese de hemicelulose, como descrito anteriormente.

De forma geral, verificamos que os transcritos de biossíntese de lignina apresentaram sempre expressão distintiva nas redes dos S1, S2 e S3, estando em menor proporção na rede S4 comparando com as outras redes. Os transcritos do metabolismo de carboidratos apresentam maiores porcentagens particularmente nas redes dos S1 e S4. Já no que se refere à relação hidrólise versus à biossíntese de polissacarídeos da parede celular, observamos que na rede do S1 a biossíntese de parede celular é maior, enquanto nas redes dos S2, S3 e S4 a hidrólise é maior, principalmente em S2, onde a diferença na porcentagem da hidrólise em relação biossíntese de polissacarídeos é mais significativa.

3.2. Seleção dos transcritos de parede celular pela centralidade de grau

Com o intuito de buscar os transcritos de parede celular mais influentes de cada uma das redes que co-expressão com *ScRAV1* e que também podem estar relacionados com a formação do aerênquima na raiz de cana-de-açúcar (Figura 10), foi avaliado o aspecto topológico das redes de co-expressão de *ScRAV1* com os transcritos de parede celular. Esta abordagem levou em consideração as 10 principais centralidades de grau em cada uma das redes de co-expressão dos transcritos de parede celular com *ScRAV1* dos segmentos da raiz de cana-de-açúcar (Tabela 2).

Tabela 2. As dez maiores centralidades de grau em cada rede de *ScRAV1* com os transcritos de parede celular nos diferentes segmentos (S1 a S4) da raiz de cana-de-açúcar, e a categorização deles conforme a Figura 11.

Redes dos Segmentos	SAS	Identificação	Categorização	Correlação	Centralidade de Grau
S1	SCBFLB2091E01.g	ScCaIS2	Biossíntese de Hemicelulose	Negativa	54
	SCCCCL3001G09.b	ScCesA1	Biossíntese de Celulose	Positiva	54
	SCJFRT1060D04.g	ScGT8 3	Biossíntese de Hemicelulose/Pectina/Outros	Positiva	54
	SCCCLB1024D03.g	ScPAE2	Hidrólise de Pectina	Positiva	54
	SCBFAD1089C07.g	ScGT8 1	Biossíntese de Hemicelulose/Pectina/Outros	Positiva	53
	SCEPLR1051B04.g	ScFLA1	Adesão Celular	Positiva	53
	SCCCRZ1001H03.g	ScGaIT1	Biossíntese de Hemicelulose/Pectina	Positiva	53
	SCQSRT1035C08.g	ScEGase3	Hidrólise de Hemicelulose/Celulose	Positiva	51
	SCQSRT1036E09.g	ScC3H1	Biossíntese de Lignina	Negativa	51
	SCEZLB1007A09.g	ScGH3 2	Hidrólise de Hemicelulose/Celulose	Positiva	50
S2	SCCCHR1004C07.g	ScCCoAOMT1	Biossíntese de Lignina	Negativa	20
	SCRLFL4004D05.g	ScUGD3	Metabolismo de Carboidratos	Negativa	19
	SCJFAD1010G05.b	ScABF1	Hidrólise de Hemicelulose/Pectina	Negativa	19
	SCVPRT2076C12.g	ScXYN3	Hidrólise de Hemicelulose	Negativa	19
	SCSFAD1125C08.g	ScPGI1	Outros	Negativa	18
	SCUTFL3079A10.g	ScCsI2	Biossíntese de Hemicelulose	Negativa	18
	SCCCCL4009H01.g	ScC4H1	Biossíntese de Lignina	Negativa	18
	SCJLRT1023B09.g	ScCOMT2	Biossíntese de Lignina	Positiva	17
	SCEQSB1017F08.g	ScPGase2	Hidrólise de Pectina	Positiva	17
	SCCCLR1C01A03.g	ScSUS2	Metabolismo de Carboidratos	Negativa	17
S3	SCACLR1014A07.g	ScCaIS1	Biossíntese de Hemicelulose	Positiva	20
	SCVPLB1020E05.g	ScIFR/PrR2	Outros	Positiva	20
	SCSBAD1054A07.g	ScCAD1	Biossíntese de Lignina	Negativa	20
	SCQSST1037F05.g	ScUGPase1	Metabolismo de Carboidratos	Positiva	20
	SCRUSB1061D02.g	ScGaIT2	Biossíntese de Hemicelulose/Pectina	Positiva	19
	SCEZRZ3015E11.g	ScCHI3	Hidrólise de Quitina	Negativa	19
	SCEQRT2029E12.g	ScCHI2	Hidrólise de Quitina	Negativa	18
	SCSBSD1057C12.g	ScCCR2	Biossíntese de Lignina	Negativa	16
	SCRUAD1064C10.g	ScF5H/F3H1	Biossíntese de Lignina	Positiva	16
	SCCCCL3080E10.b	ScPGase1	Hidrólise de Pectina	Positiva	15
S4	SCQSRT1034C09.g	ScCCoAOMT4	Biossíntese de Lignina	Negativa	32
	SCCCST1006H12.g	ScGH1 1	Hidrólise de Celulose	Negativa	31
	SCCCLR1066B10.g	ScCesA2	Biossíntese de Celulose	Positiva	30
	SCQSRT1034C08.g	ScFUT1	Biossíntese de Hemicelulose	Negativa	30
	SCCCLB1C06E01.g	ScCsI1	Biossíntese de Hemicelulose	Positiva	30
	SCMCSB1114E08.g	ScFLA2	Adesão Celular	Negativa	29
	SCVPRZ3027D06.g	ScEGase4	Hidrólise de Hemicelulose/Celulose	Positiva	29
	SCCCLB1025F10.g	ScGT47 1	Biossíntese de Hemicelulose/Pectina	Positiva	29
	SCJFLR1017B11.g	ScPAL2	Biossíntese de Lignina	Negativa	28
	SCCCCL3140B03.g	ScXI1	Metabolismo de Carboidratos	Negativa	27

A visualização das principais centralidades dos transcritos de parede celular nas redes de co-expressão de *ScRAV1*, pode ser observada na Figura 10. Na rede do S1, os transcritos mais representados como centralidades de grau estão relacionados à biossíntese de hemicelulose e hemicelulose/pectina/outros, e à hidrólise de celulose e hemicelulose (Tabela 2). Já na rede do S2, a maioria das centralidades de grau são referentes à biossíntese de lignina e à hidrólise de hemicelulose e pectina. Na rede do S3, é prevalente as centralidades de grau que estão associadas à biossíntese de lignina. E por último, na rede do S4 pode-se verificar a predominância das centralidades de grau de biossíntese de hemicelulose e lignina, e apenas uma relacionada ao metabolismo de carboidratos. Assim, vemos que as maiores centralidades têm representantes de transcritos de parede celular que estão em maior proporção em cada uma das redes de *ScRAV1*, conforme o item 3.1 (Figura 11).

3.3. Busca do homólogo mais próximo em *S. bicolor* e em *A. thaliana*

A seleção das 40 maiores centralidades de grau encontradas nas redes de co-expressão com *ScRAV1* nos segmentos S1 ao S4 (Tabela 2), permitiu a busca dos melhores *hits* em homólogos de *S. bicolor* (Tabela 3). Pode-se observar que 32 (80%) das sequências em cana-de-açúcar apresentaram $\geq 90\%$ de identidade com as sequências de *S. bicolor* (Tabela 3).

Posteriormente, essas sequências de *S. bicolor* foram utilizadas para buscar os melhores *hits* de possíveis homólogos em *A. thaliana* (Tabela 4). Com a anotação em *A. thaliana*, pode-se verificar que 11 (34,38 %) das sequências de *S. bicolor* apresentaram $\geq 70\%$ de identidade com as sequências de *A. thaliana* (Tabela 4).

Tabela 3. Os melhores *hits* preditos em *S. bicolor* que apresentaram $\geq 90\%$ de identidade com os transcritos de parede celular da raiz de cana-de-açúcar para as maiores centralidades de grau (Tabela 2).

SAS	Sub_category1	Identificação	Correlação de ScRAV1		Hit	<i>S. bicolor</i>	Identidade (%)	E-value
			Segmento	Tipo				
SCJFRT1060D04.g	Glycosyl transferase family 8 (cell wall metabolism, starch metabolism or lipopolysaccharide biosynthesis)	ScGT8 3	S1	Positiva	Sobic.002G274700.1	similar to Glycosyl transferase family 8 protein-like	98	0
SCCCLB1024D03.g	Pectinacetyltransferase family protein	ScPAE2	S1	Positiva	Sobic.006G204500.1	similar to B0403H10-OSIGBa0105A11.19 protein	96	3,00E-123
SCBFAD1089C07.g	Glycosyl transferase family 8 (cell wall metabolism, starch metabolism or lipopolysaccharide biosynthesis)	ScGT8 1	S1	Positiva	Sobic.001G338400.1	similar to Putative uncharacterized protein	98	0
SCEPLR1051B04.g	Fasciclin-like arabinogalactan protein	ScFLA1	S1	Positiva	Sobic.009G055900.1	similar to Putative uncharacterized protein	92	0
SCCCRZ1001H03.g	Galactosyltransferase	ScGalT1	S1	Positiva	Sobic.010G059700.2	similar to Beta 1,3-glycosyltransferase-like protein II	99	1,00E-178
SCQSRT1035C08.g	endo-1,4-beta-glucanase Glycosyl hydrolase family 9	ScEGase3	S1	Positiva	Sobic.006G282300.1	PTHR22298:SF39 - ENDOGLUCANASE 7 (1 of 2)	98	0
SCQSRT1036E09.g	putative p-Coumarate 3-hydroxylase (C3H)	ScC3H1	S1	Negativa	Sobic.003G327800.1	similar to Cytochrome P450	96	0
SCEZLB1007A09.g	glycosyl hydrolase family 3 protein	ScGH3 2	S1/S2	Positiva/Negativa	Sobic.001G089000.1	similar to Exhydrolase II	98	0
SCRLFL4004D05.g	UDP-D-glucuronate decarboxylase	ScUGD3	S2	Negativa	Sobic.009G112600.1	similar to UDP-221xylase	91	2,00E-108
SCJFAD1010G05.b	Alpha-L-arabinofuranosidase	ScABF1	S2	Negativa	Sobic.008G017800.1	similar to Alpha-L-arabinofuranosidase C-terminus family protein, expressed	93	2,00E-84
SCVPRT2076C12.g	Glycoside hydrolase family 10 protein (Xylanase)	ScXYN3	S2	Negativa	Sobic.002G127400.2	similar to Putative 1,4-beta-D xylan xylanohydrolase	98	0
SCSFAD1125C08.g	Polygalacturonase inhibitor	ScPGI1	S2	Negativa	Sobic.002G343600.1	similar to Polygalacturonase inhibitor 1 precursor	97	1,00E-170
SCUTFL3079A10.g	cellulose synthase like (CSL)	ScCsl2	S2	Negativa	Sobic.003G308100.1	2.4.1.168//2.4.2.39 - Xyloglucan 6-xylosyltransferase / Xyloglucan 6-alpha-D-xylosyltransferase (1 of 4)	97	3,00E-140
SCCCCL4009H01.g	cinnamate 4-hydroxylase (C4H)	ScC4H1	S2	Negativa	Sobic.002G126600.1	similar to Cytochrome P450	97	0
SCJLRT1023B09.g	caffeic acid 3-O-methyltransferase	ScCOMT2	S2	Positiva	Sobic.007G047300.1	similar to Caffeic acid 3-O-methyltransferase	96	4,00E-133
SCVPLB1020E05.g	isoflavone reductase/pinoresinol reductase (lignan biosynthesis)	ScIFR/PrR2	S3	Positiva	Sobic.003G104600.1	similar to Isoflavone reductase homolog IRL	94	0
SCSBAD1054A07.g	Cinnamyl alcohol dehydrogenase (CAD)	ScCAD1	S3	Negativa	Sobic.002G195600.1	similar to Putative uncharacterized protein	96	1,00E-132
SCQSST1037F05.g	UDP-glucose pyrophosphorylase	ScUGPase1	S3	Positiva	Sobic.002G291200.2	similar to UTP--glucose-1-phosphate uridylyltransferase	98	2,00E-38
SCRUSB1061D02.g	Galactosyltransferase	ScGalT2	S2/S3	Negativa/Positiva	Sobic.007G118500.1	similar to Putative Avr9 elicitor response protein	97	1,00E-79
SCEZRZ3015E11.g	chitinase	ScCHI3	S3	Negativa	Sobic.001G516000.1	similar to Basic endochitinase, putative, expressed	94	1,00E-175
SCEQRT2029E12.g	chitinase	ScCHI2	S3	Negativa	Sobic.009G130100.1	similar to Chitinase precursor	94	0
SCSBSD1057C12.g	Cinnamoyl-coA reductase (CCR)	ScCCR2	S1/S3	Positiva/Negativa	Sobic.002G249900.1	weakly similar to Putative cinnamoyl-CoA reductase	96	0
SCRUAD1064C10.g	putative ferulate 5 hydroxylase (F5H); flavonoid 3 hydroxylase (F3H)	ScF5H/F3H1	S3	Positiva	Sobic.001G360800.2	similar to Flavonoid 3'5' hydroxylase	94	1,00E-80
SCCCCL3080E10.b	Glycoside hydrolase family 28 protein (polygalacturonase)	ScPGase1	S3	Positiva	Sobic.002G090400.2	similar to Putative uncharacterized protein	94	5,00E-176
SCQSRT1034C09.g	caffeoyl-CoA 3-O-methyltransferase	ScCCoAOMT4	S1/S4	Positiva/Negativa	Sobic.007G218800.1	similar to Caffeoyl-CoA O-methyltransferase	97	0
SCCCLR1066B10.g	Cellulose synthase (CesA)	ScCesA2	S4	Positiva	Sobic.002G075500.1	similar to Cellulose synthase-9	99	0
SCQSRT1034C08.g	Fucosyltransferase	ScFUT1	S1/S4	Negativa/Negativa	Sobic.008G053600.1	similar to 4-alpha-L-fucosyltransferase	99	0

Tabela 3. Conclusão.

SCMCSB1114E08.g	Fasciclin-like arabinogalactan protein	ScFLA2	S4	Negativa	Sobic.001G521850.1	PF02469 - Fasciclin domain (Fasciclin) (1 of 27)	93	3,00E-123
SCVPRZ3027D06.g	endo-1,4-beta-glucanase Glycosyl hydrolase family 9	ScEGase4	S4	Positiva	Sobic.004G244600.1	PTHR22298//PTHR22298:SF52 - ENDO-1,4-BETA-GLUCANASE // SUBFAMILY NOT NAMED (1 of 2)	97	3,00E-108
SCCCLB1025F10.g	glycosyltransferase family 47	ScGT47 1	S4	Positiva	Sobic.009G220200.1	similar to P0482D04.16 protein	97	0
SCJFLR1017B11.g	phenylalanine ammonia-lyase (PAL)	ScPAL2	S1/S4	Positiva/Negativa	Sobic.004G220300.1	similar to Phenylalanine ammonia-lyase	96	0
SCCCCL3140B03.g	Xylose isomerase	ScXI1	S4	Negativa	Sobic.002G409800.1	similar to Xylose isomerase	92	0

Tabela 4. Os melhores *hits* em *A. thaliana* com base nas sequências de *S. bicolor* preditas anteriormente para os transcritos de parede celular na raiz de cana-de-açúcar.

SAS	Sub_category1	Identificação	Correlação de ScRAV1		S. bicolor		A. thaliana		Símbolo	Identidade (%)	E-value
			Segmento	Tipo	Hit	Descrição	Hit	Descrição			
SCJFRT1060D04.g	Glycosyl transferase family 8 (cell wall metabolism, starch metabolism or lipopolysaccharide biosynthesis)	ScGT8 3	S1	Positiva	Sobic.002G274700.1	similar to Glycosyl transferase family 8 protein-like	AT3G61130.1	galacturonosyltransferase 1	GAUT1, LGT1	71,01	0
SCCCLB1024D03.g	Pectinacetyltransferase family protein	ScPAE2	S1	Positiva	Sobic.006G204500.1	similar to B0403H10-OSIGBa0105A11.19 protein	AT4G19410.1	Pectinacetyltransferase family protein	PAE7, ATPAE7	61,87	9,00E-175
SCBFAD1089C07.g	Glycosyl transferase family 8 (cell wall metabolism, starch metabolism or lipopolysaccharide biosynthesis)	ScGT8 1	S1	Positiva	Sobic.001G338400.1	similar to Putative uncharacterized protein	AT1G18580.1	galacturonosyltransferase 11	GAUT11	64,56	0
SCEPLR1051B04.g	Fasciclin-like arabinogalactan protein	ScFLA1	S1	Positiva	Sobic.009G055900.1	similar to Putative uncharacterized protein	AT4G12730.1	FASCICLIN-like arabinogalactan 2	FLA2	41,35	5e-97
SCCCRZ1001H03.g	Galactosyltransferase	ScGalT1	S1	Positiva	Sobic.010G059700.2	similar to Beta 1,3-glycosyltransferase-like protein II	AT4G32120.1	Galactosyltransferase family protein	HPGT2	72,27	4,00E-143
SCQSRT1035C08.g	endo-1,4-beta-glucanase Glycosyl hydrolase family 9	ScEGase3	S1	Positiva	Sobic.006G282300.1	PTHR22298:SF39 - ENDOGLUCANASE 7 (1 of 2)	AT1G65610.1	Six-hairpin glycosidases superfamily protein	ATGH9A2, KOR2, ATKOR2	57,98	0
SCQSRT1036E09.g	putative p-Coumarate 3-hydroxylase (C3H)	ScC3H1	S1	Negativa	Sobic.003G327800.1	similar to Cytochrome P450	AT2G40890.1	cytochrome P450, family 98, subfamily A, polypeptide 3	CYP98A3, REF8	69,47	0
SCEZLB1007A09.g	glycosyl hydrolase family 3 protein	ScGH3 2	S1/S2	Positiva/Negativa	Sobic.001G089000.1	similar to Exhydrolase II	AT5G20950.2	Glycosyl hydrolase family protein	BGLC1	72,09	0
SCRFL4004D05.g	UDP-D-glucuronate decarboxylase	ScUGD3	S2	Negativa	Sobic.009G112600.1	similar to UDP-glucuronic acid decarboxylase	AT2G47650.1	UDP-xylose synthase 4	UXS4	73,86	0
SCJFAD1010G05.b	Alpha-L-arabinofuranosidase	ScABF1	S2	Negativa	Sobic.008G017800.1	similar to Alpha-L-arabinofuranosidase C-terminus family protein, expressed	AT3G10740.1	alpha-L-arabinofuranosidase 1	ARAF, ARAF1, ASD1	63,50	0
SCVPRT2076C12.g	Glycoside hydrolase family 10 protein (Xylanase)	ScXYN3	S2	Negativa	Sobic.002G127400.2	similar to Putative 1,4-beta-D xylan xylanohydrolase	AT1G10050.1	glycosyl hydrolase family 10 protein / carbohydrate-binding domain-containing protein	ATXYN2	51,22	0
SCSFAD1125C08.g	Polygalacturonase inhibitor	ScPGI1	S2	Negativa	Sobic.002G343600.1	similar to Polygalacturonase inhibitor 1 precursor	AT5G06860.1	polygalacturonase inhibiting protein 1	PGIP1, ATPGIP1	39,81	8,00E-72
SCUTFL3079A10.g	cellulose synthase like (CSL)	ScCsl2	S2	Negativa	Sobic.003G308100.1	2.4.1.168//2.4.2.39 - Xyloglucan 6-xylosyltransferase / Xyloglucan 6-alpha-D-xylosyltransferase (1 of 4)	AT4G07960.1	Cellulose-synthase-like C12	ATCSLC12, CSLC12	70,74	0
SCCCCL4009H01.g	cinnamate 4-hydroxylase (C4H)	ScC4H1	S2	Negativa	Sobic.002G126600.1	similar to Cytochrome P450	AT2G30490.1	cinnamate-4-hydroxylase	C4H	76,29	0
SCJLRT1023B09.g	caffeic acid 3-O-methyltransferase	ScCOMT2	S2	Positiva	Sobic.007G047300.1	similar to Caffeic acid 3-O-methyltransferase	AT5G54160.1	O-methyltransferase 1	ATOMT1, OMT1, COMT1, ATCOMT, OMT3	60,93	4,00E-163
SCVPLB1020E05.g	isoflavone reductase/pinoresinol reductase (lignan biosynthesis)	ScIFR/PrR2	S3	Positiva	Sobic.003G104600.1	similar to Isoflavone reductase homolog IRL	AT1G75280.1	NmrA-like Negative transcriptional regulator family protein	P3	60,77	6,00E-139
SCSBAD1054A07.g	Cinnamyl alcohol dehydrogenase (CAD)	ScCAD1	S3	Negativa	Sobic.002G195600.1	similar to Putative uncharacterized protein	AT4G37970.1	cinnamyl alcohol dehydrogenase 6	ATCAD, CAD6	55,59	2,00E-146

Tabela 4. Conclusão.

SCQSST1037F05.g	UDP-glucose pyrophosphorylase	ScUGPase1	S3	Positiva	Sobic.002G291200.2	similar to UTP--glucose-1-phosphate uridylyltransferase	AT5G17310.2	UDP-glucose pyrophosphorylase 2	UGP2	82,40	0
SCRUSB1061D02.g	Galactosyltransferase	ScGalT2	S2/S3	Negativa/Positiva	Sobic.007G118500.1	similar to Putative Avr9 elicitor response protein	AT5G53340.1	Galactosyltransferase family protein	HPGT1	63,66	1,00E-163
SCEZRZ3015E11.g	chitinase	ScCHI3	S3	Negativa	Sobic.001G516000.1	similar to Basic endochitinase, putative, expressed	AT3G12500.1	basic chitinase	ATHCHIB, PR3, PR-3, CHI-B, B-CHI, HCHIB	57,74	2,00E-98
SCEQRT2029E12.g	chitinase	ScCHI2	S3	Negativa	Sobic.009G130100.1	similar to Chitinase precursor	AT3G12500.1	basic chitinase	ATHCHIB, PR3, PR-3, CHI-B, B-CHI, HCHIB	67,58	9,00E-145
SCS8SD1057C12.g	Cinnamoyl-coA reductase (CCR)	ScCCR2	S1/S3	Positiva/Negativa	Sobic.002G249900.1	weakly similar to Putative cinnamoyl-CoA reductase	AT1G76470.1	NAD(P)-binding Rossmann-fold superfamily protein		49,85	5,00E-102
SCRUAD1064C10.g	putative ferulate 5 hydroxylase (F5H); flavonoid 3 hydroxylase (F3H)	ScF5H/F3H1	S3	Positiva	Sobic.001G360800.2	similar to Flavonoid 3'S' hydroxylase	AT5G07990.1	Cytochrome P450 superfamily protein	CYP75B1,TT7, DS01	44,53	6,00E-158
SCCCCL3080E10.b	Glycoside hydrolase family 28 protein (polygalacturonase)	ScPGase1	S3	Positiva	Sobic.002G090400.2	similar to Putative uncharacterized protein	AT3G62110.1	Pectin lyase-like superfamily protein		65,34	0
SCQSRT1034C09.g	caffeoyl-CoA 3-O-methyltransferase	ScCCoAOMT4	S1/S4	Positiva/Negativa	Sobic.007G218800.1	similar to Caffeoyl-CoA O-methyltransferase	AT4G34050.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	CCoAOMT1	59,03	3,00E-105
SCCCLR1066B10.g	Cellulose synthase (CesA)	ScCesA2	S4	Positiva	Sobic.002G075500.1	similar to Cellulose synthase-9	AT5G05170.1	Cellulose synthase family protein	ELI1, MRE1, CESA3, IXR1	80,37	0
SCQSRT1034C08.g	Fucosyltransferase	ScFUT1	S1/S4	Negativa/Negativa	Sobic.008G053600.1	similar to 4-alpha-L-fucosyltransferase	AT1G71990.1	fucosyltransferase 13	FUT13, ATFUT13,FUCTC, FT4-M,ATFT4	55,99	8,00E-178
SCMCSB1114E08.g	Fasciclin-like arabinogalactan protein	ScFLA2	S4	Negativa	Sobic.001G521850.1	PF02469 - Fasciclin domain (Fasciclin) (1 of 27)	AT2G45470.1	FASCICLIN-like arabinogalactan protein 8	AGP8, FLA8	41,67	2,00E-80
SCVPRZ3027D06.g	endo-1,4-beta-glucanase Glycosyl hydrolase family 9	ScEGase4	S4	Positiva	Sobic.004G244600.1	PTHR22298//PTHR22298:SF52 - ENDO-1,4-BETA-GLUCANASE // SUBFAMILY NOT NAMED (1 of 2)	AT2G32990.1	glycosyl hydrolase 9B8	GH9B8	66,00	0
SCCCLB1025F10.g	glycosyltransferase family 47	ScGT47 1	S4	Positiva	Sobic.009G220200.1	similar to P0482D04.16 protein	AT1G27440.1	Exostosin family protein	GUT2, IRX10, ATGUT1	78,73	0
SCJFLR1017B11.g	phenylalanine ammonia-lyase (PAL)	ScPAL2	S1/S4	Positiva/Negativa	Sobic.004G220300.1	similar to Phenylalanine ammonia-lyase	AT3G53260.1	phenylalanine ammonia-lyase 2	PAL2	70,89	0
SCCCCL3140B03.g	Xylose isomerase	ScXI1	S4	Negativa	Sobic.002G409800.1	similar to Xylose isomerase	AT5G57655.2	xylose isomerase family protein		76,95	0

3.4. Enriquecimento funcional dos transcritos de parede celular através dos termos do Gene Ontology (GO)

Com a anotação manual em *A. thaliana* com base nos homólogos em *S. bicolor* (Tabela 4), foram preditas as possíveis funções ou papéis biológicos dos transcritos de parede celular em cana-de-açúcar utilizando os termos do Gene Ontology (GO) para componente celular, processo biológico e função molecular (Figura 12). Para verificar todos os termos GOs encontrados em *A. thaliana*, pode ser visto no Apêndice D.

Figura 12. Distribuição dos primeiros 20 termos do Gene Ontology (GO) para componente celular (A), processo biológico (B) e função molecular (C) para os transcritos de parede celular de cana-de-açúcar anotados em *A. thaliana* com base nos homólogos em *S. bicolor*.

Pode-se verificar no termo para componente celular (Figura 12A), que a maioria destes transcritos está associada principalmente ao termo *integral component of membrane* (GO:0016021), mas também pode-se observar termos para outros compartimentos celulares.

Com relação aos termos para processos biológicos (Figura 12B), pode-se verificar que para os transcritos de parede celular é predominante o termo *cell wall organization* (GO:0071555). Ainda foram encontrados vários transcritos de parede celular associados aos termos de biossíntese de lignina e ao metabolismo secundário como: *oxidation-reduction process* (GO:0055114), *lignin biosynthetic process* (GO:0009809), *plant-type secondary cell wall biogenesis* (GO:0009834), *coumarin biosynthetic process* (GO:0009805) e *flavonoid biosynthetic process* (GO:0009813).

No que se refere aos termos para função molecular para os transcritos de parede celular (Figura 12C), pode-se observar que estes podem apresentar vários tipos de atividades enzimáticas ou de ligação à substratos. Dentre eles estão principalmente os termos: *metal ion*

binding (GO:0046872), *iron ion binding* (GO:0005506) e *heme binding* (GO:0020037) (Figura 12C).

Pode-se verificar ainda que os termos GOs para componente celular, processo biológico e função molecular encontrados para os transcritos de parede celular anotados em *A. thaliana* estão de acordo com o que foi observado anteriormente nos itens 3 e 3.1.

3.5. Sítios de ligação de RAV ao DNA nas sequências homólogas de *S. bicolor*

Com a anotação dos transcritos de parede celular de cana-de-açúcar em *S. bicolor* (Tabela 3), foi possível utilizar as regiões promotoras dos genes desta espécie para buscar sítios de ligação de RAV ao DNA. Pode-se verificar na Tabela 5 a quantidade de sítios putativos de ligação de RAV para os motivos AP2 (CAACA) e B3 (CACCTG) encontrados nas regiões promotoras dos genes de parede celular de *S. bicolor*.

No total, foram encontrados 190 sítios para RAV, 162 (85,26%) destes sítios referem-se para ao motivo AP2 (CAACA) e apenas 28 (14,74%) referem-se para o motivo B3 (CACCTG) (Tabela 5). Pode-se observar que 18 (56,25%) destas sequências apresentam ambos os motivos AP2 e B3 (CAACA/CACCTG) para RAV, enquanto 14 (43,75%) das sequências apresentam somente o motivo AP2 (CAACA) para RAV (Tabela 5). Isto evidencia que há uma diferença na proporção entre os motivos AP2 e B3 para os sítios de ligação de RAV ao DNA nas regiões promotoras dos genes de parede celular em *S. bicolor*.

Tabela 5. Sítios putativos de ligação de RAV (AP2/B3) na região promotora de genes de parede celular dentre os possíveis ortólogos previamente selecionados de *S. bicolor*.

SAS	Sub_category1	Identificação	Segmento	Correlação	<i>S. bicolor</i>		Localização	Fita	Motivo	Quantidade de Sítios		
					Hit	Descrição				AP2	B3	Total
SCJFRT1060D04.g	Glycosyl transferase family 8 (cell wall metabolism, starch metabolism or lipopolysaccharide biosynthesis)	ScGT8 3	S1	Positiva	Sobic.002G274700.1	similar to Glycosyl transferase family 8 protein-like	91	+	CAACA	4	1	5
							186	-	CAACA			
							231	-	CACCTG			
							249	+	CAACA			
							448	-	CAACA			
SCCCLB1024D03.g	Pectinacetyltransferase family protein	ScPAE2	S1	Positiva	Sobic.006G204500.1	similar to B0403H10-OSIGBa0105A11.19 protein	182	-	CAACA	6	0	6
							434	-	CAACA			
							561	+	CAACA			
							663	+	CAACA			
							700	-	CAACA			
SCBFAD1089C07.g	Glycosyl transferase family 8 (cell wall metabolism, starch metabolism or lipopolysaccharide biosynthesis)	ScGT8 1	S1	Positiva	Sobic.001G338400.1	similar to Putative uncharacterized protein	773	+	CAACA	10	0	10
							74	+	CAACA			
							148	+	CAACA			
							151	+	CAACA			
							343	-	CAACA			
							408	-	CAACA			
							1239	+	CAACA			
							1355	-	CAACA			
							1372	-	CAACA			
							1375	-	CAACA			
SCEPLR1051B04.g	Fasciclin-like arabinogalactan protein	ScFLA1	S1	Positiva	Sobic.009G055900.1	similar to Putative uncharacterized protein	1461	-	CAACA	6	1	7
							431	+	CAACA			
							450	-	CAACA			
							1485	+	CAACA			
							1598	-	CAACA			
							1601	-	CAACA			
SCCCRZ1001H03.g	Galactosyltransferase	ScGalT1	S1	Positiva	Sobic.010G059700.2	similar to Beta 1,3-glycosyltransferase-like protein II	1801	+	CAACA	3	0	3
							1863	+	CACCTG			
							113	-	CAACA			
SCQSRT1035C08.g	endo-1,4-beta-glucanase Glycosyl hydrolase family 9	ScEGase3	S1	Positiva	Sobic.006G282300.1	PTHR22298:SF39 - ENDOGLUCANASE 7 (1 of 2)	121	-	CAACA	5	0	5
							144	+	CAACA			

Tabela 5. Continuação.

							145	-	CAACA			
							312	+	CAACA			
							1142	-	CAACA			
							1286	-	CAACA			
SCQSRT1036E09.g	putative p-Coumarate 3-hydroxylase (C3H)	ScC3H1	S1	Negativa	Sobic.003G327800.1	similar to Cytochrome P450	210	+	CACCTG	7	1	8
							395	+	CAACA			
							612	-	CAACA			
							1013	-	CAACA			
							1068	-	CAACA			
							1124	-	CAACA			
							1390	+	CAACA			
							1672	-	CAACA			
SCEZLB1007A09.g	glycosyl hydrolase family 3 protein	ScGH3 2	S1/S2	Positiva/Negativa	Sobic.001G089000.1	similar to Exhydrolase II	18	+	CAACA	5	1	6
							229	-	CAACA			
							1200	+	CAACA			
							1380	-	CAACA			
							1497	+	CACCTG			
SCRLFL4004D05.g	UDP-D-glucuronate decarboxylase	ScUGD3	S2	Negativa	Sobic.009G112600.1	similar to UDP-glucuronic acid decarboxylase	1547	-	CAACA	2	2	4
							375	-	CAACA			
							1321	+	CAACA			
							1583	+	CACCTG			
SCJFAD1010G05.b	Alpha-L-arabinofuranosidase	ScABF1	S2	Negativa	Sobic.008G017800.1	similar to Alpha-L-arabinofuranosidase C-terminus family protein, expressed	1682	+	CACCTG	9	0	9
							37	-	CAACA			
							105	-	CAACA			
							535	+	CAACA			
							548	+	CAACA			
							917	-	CAACA			
							1012	+	CAACA			
							1243	+	CAACA			
SCVPRT2076C12.g	Glycoside hydrolase family 10 protein (Xylanase)	ScXYN3	S2	Negativa	Sobic.002G127400.2	similar to Putative 1,4-beta-D xylan xylanohydrolase	1643	+	CAACA	2	0	2
							1955	+	CAACA			
SCSFAD1125C08.g	Polygalacturonase inhibitor	ScPGI1	S2	Negativa	Sobic.002G343600.1	similar to Polygalacturonase inhibitor 1 precursor	372	-	CAACA	4	0	4
							768	-	CAACA			

Tabela 5. Continuação.

SCUTFL3079A10.g	cellulose synthase like (CSL)	ScCsl2	S2	Negativa	Sobic.003G308100.1	2.4.1.168//2.4.2.39 - Xyloglucan 6-xylosyltransferase / Xyloglucan 6-alpha-D-xylosyltransferase (1 of 4)	735	-	CAACA	7	2	9
							886	-	CAACA			
							889	-	CAACA			
							154	-	CAACA			
							317	+	CAACA			
							374	-	CAACA			
							426	+	CAACA			
							525	+	CAACA			
							712	-	CAACA			
							1240	+	CACCTG			
SCCCCL4009H01.g	cinnamate 4-hydroxylase (C4H)	ScC4H1	S2	Negativa	Sobic.002G126600.1	similar to Cytochrome P450	1244	-	CAACA	4	1	5
							1872	-	CACCTG			
							444	+	CACCTG			
							722	-	CAACA			
							884	-	CAACA			
SCJLRT1023B09.g	caffeic acid 3-O-methyltransferase	ScCOMT2	S2	Positiva	Sobic.007G047300.1	similar to Caffeic acid 3-O-methyltransferase	1924	+	CAACA	7	0	7
							1927	+	CAACA			
							431	+	CAACA			
							495	+	CAACA			
							631	-	CAACA			
							672	+	CAACA			
							691	-	CAACA			
SCVPLB1020E05.g	isoflavone reductase/pinoresinol reductase (lignan biosynthesis)	ScIFR/PrR2	S3	Positiva	Sobic.003G104600.1	similar to Isoflavone reductase homolog IRL	1560	+	CAACA	9	1	10
							1899	+	CAACA			
							74	+	CAACA			
							227	-	CAACA			
							384	-	CAACA			
							1158	+	CACCTG			
							1223	+	CAACA			
							1719	+	CAACA			
							1724	-	CAACA			
							1823	+	CAACA			
SCSBAD1054A07.g	Cinnamyl alcohol dehydrogenase (CAD)	ScCAD1	S3	Negativa	Sobic.002G195600.1	similar to Putative uncharacterized protein	1922	+	CAACA	3	0	3
							1932	+	CAACA			
							60	-	CAACA			

Tabela 5. Continuação.

SCQSST1037F05.g	UDP-glucose pyrophosphorylase	ScUGPase1	S3	Positiva	Sobic.002G291200.2	similar to UTP--glucose-1-phosphate uridylyltransferase	635	-	CAACA	2	1	3
							1962	+	CAACA			
							439	+	CACCTG			
							1147	+	CAACA			
							1413	-	CAACA			
SCRUSB1061D02.g	Galactosyltransferase	ScGalT2	S2/S3	Negativa/Positiva	Sobic.007G118500.1	similar to Putative Avr9 elicitor response protein	107	+	CAACA	5	2	7
							203	+	CAACA			
							899	-	CAACA			
							1425	-	CAACA			
							1454	+	CACCTG			
							1491	+	CAACA			
SCEZRZ3015E11.g	chitinase	ScCHI3	S3	Negativa	Sobic.001G516000.1	similar to Basic endochitinase, putative, expressed	1971	-	CACCTG	5	0	5
							221	+	CAACA			
							530	-	CAACA			
							1112	+	CAACA			
							1412	-	CAACA			
SCEQRT2029E12.g	chitinase	ScCHI2	S3	Negativa	Sobic.009G130100.1	similar to Chitinase precursor	1836	+	CAACA	5	1	6
							61	-	CAACA			
							491	-	CAACA			
							1083	+	CAACA			
							1258	-	CACCTG			
							1282	+	CAACA			
SCSBSD1057C12.g	Cinnamoyl-coA reductase (CCR)	ScCCR2	S1/S3	Positiva/Negativa	Sobic.002G249900.1	weakly similar to Putative cinnamoyl-CoA reductase	1583	+	CAACA	6	0	6
							79	-	CAACA			
							115	+	CAACA			
							193	+	CAACA			
							954	+	CAACA			
							957	+	CAACA			
SCRUAD1064C10.g	putative ferulate 5 hydroxylase (F5H); flavonoid 3 hydroxylase (F3H)	ScF5H/F3H1	S3	Positiva	Sobic.001G360800.2	similar to Flavonoid 3'5' hydroxylase	1865	-	CAACA	1	0	1
							681	-	CAACA			
SCCCCL3080E10.b	Glycoside hydrolase family 28 protein (polygalacturonase)	ScPGase1	S3	Positiva	Sobic.002G090400.2	similar to Putative uncharacterized protein	222	+	CAACA	4	4	8
							422	+	CACCTG			

Tabela 5. Continuação.

							481	+	CACCTG			
							754	+	CACCTG			
							1471	-	CAACA			
							1911	-	CACCTG			
							1915	-	CAACA			
							1953	-	CAACA			
							113	-	CAACA			
							381	+	CAACA			
							765	+	CAACA			
							1112	+	CAACA			
SCQSRT1034C09.g	caffeoyl-CoA 3-O-methyltransferase	ScCCoAOMT4	S1/S4	Positiva/Negativa	Sobic.007G218800.1	similar to Caffeoyl-CoA O-methyltransferase	1115	+	CAACA	8	2	10
							1337	+	CAACA			
							1438	-	CACCTG			
							1583	+	CAACA			
							1787	+	CAACA			
							1952	+	CACCTG			
							55	+	CAACA			
							689	+	CAACA			
SCCCLR1066B10.g	Cellulose synthase (CesA)	ScCesA2	S4	Positiva	Sobic.002G075500.1	similar to Cellulose synthase-9	713	+	CAACA	5	0	5
							1109	+	CAACA			
							1112	+	CAACA			
SCQSRT1034C08.g	Fucosyltransferase	ScFUT1	S1/S4	Negativa/Negativa	Sobic.008G053600.1	similar to 4-alpha-L-fucosyltransferase	694	+	CAACA	2	0	2
							1603	+	CAACA			
SCMCSB1114E08.g	Fasciclin-like arabinogalactan protein	ScFLA2	S4	Negativa	Sobic.001G521850.1	PF02469 - Fasciclin domain (Fasciclin) (1 of 27)	437	-	CAACA	4	1	5
							735	-	CAACA			
							744	-	CACCTG			
							943	-	CAACA			
							998	-	CAACA			
SCVPRZ3027D06.g	endo-1,4-beta-glucanase Glycosyl hydrolase family 9	ScEGase4	S4	Positiva	Sobic.004G244600.1	PTHR22298//PTHR22298:SF52 - ENDO-1,4-BETA-GLUCANASE // SUBFAMILY NOT NAMED (1 of 2)	280	+	CAACA	5	2	7
							370	+	CAACA			
							677	-	CAACA			
							937	+	CAACA			
							965	+	CACCTG			
							1368	+	CAACA			
SCCCLB1025F10.g	glycosyltransferase family 47	ScGT47 1	S4	Positiva	Sobic.009G220200.1	similar to P0482D04.16 protein	1958	-	CACCTG	7	1	8
							419	+	CAACA			

Tabela 5. Conclusão.

							581	+	CAACA				
							989	-	CAACA				
							1010	-	CAACA				
							1119	+	CAACA				
							1209	-	CAACA				
							1325	-	CACCTG				
SCJFLR1017B11.g							1383	-	CAACA	4			
							303	-	CAACA				
							523	-	CAACA				
							925	+	CAACA				1
							1606	+	CAACA				
							1637	-	CACCTG				
SCCCCL3140B03.g							152	-	CAACA	6			
							252	+	CAACA				
							473	-	CAACA				
							696	-	CACCTG				
							927	-	CAACA				3
							1204	+	CAACA				
							1207	+	CAACA				
							1617	-	CACCTG				
							1968	-	CACCTG				
							Total						

3.6. Rede de regulação gênica para os genes de parede celular

A predição *in silico* para sítios putativos de RAV nos genes homólogos em *S. bicolor* para os transcritos de cana-de-açúcar (Tabela 5) e o motivo repressor para *SbRAV1* (Figura 9) foram usados para construir a rede de regulação gênica (GRN, *Gene Regulatory Network*), como mostra a Figura 13.

Figura 13. Representação da rede de regulação gênica dos potenciais alvos dos genes de parede celular que possivelmente podem ser coordenados por *SbRAV1* (círculo em vermelho no centro) com base na predição *in silico* dos sítios putativos para RAV em genes de parede celular em *S. bicolor* (Tabela 5). A identificação dos homólogos em *S. bicolor* para os transcritos de parede celular de cana-de-açúcar foram nomeados apenas trocando *Sc* por *Sb* devido a busca de elementos *cis* nesta primeira espécie. As cores dos círculos na rede de regulação gênica referem-se à categorização em biossíntese, hidrólise ou outros para estes transcritos de parede celular em cana-de-açúcar, conforme Piovezani (2017) com modificações. À direita mostra a legenda com a atividade de *SbRAV1* baseada na conservação do seu motivo repressor B3 (BRD, *B3 repressor domain*) (Ikeda & Ohme-Takagi, 2009), como mostra a Figura 9.

A partir da rede de regulação gênica de *SbRAV1* podemos verificar que este fator de transcrição tem alvos para praticamente para todas as categorias apresentadas na Figura 13. A maior parte destes genes está relacionada à biossíntese da parede celular. Dentre estes, é evidente a predominância de genes alvos relacionados à categoria de biossíntese de lignina (Figura 13).

Contudo, ainda foram encontrados poucos genes de parede celular associados à hidrólise de polissacarídeos que podem ser alvos de *SbRAV1* (Figura 13). Por exemplo, podem-se verificar genes relacionados à hidrólise de pectina como *SbPAE2* e *SbPGase1*, de arabinoxilano e de celulose como *SbABF1*, *SbEGase3* e *SbEGase4*, respectivamente, bem como os genes relacionados à hidrólise do xiloglucano como *SbGH3 2* e *SbXYN3*.

De modo geral, pode-se concluir que *SbRAV1* apresenta genes alvos principalmente tanto na regulação de genes de biossíntese como na hidrólise da parede celular na raiz de cana-de-açúcar.

3.7. Perfil de expressão dos transcritos de parede celular com base no RNA-Seq

O heatmap mostra o perfil de expressão dos transcritos de parede celular nos quatro segmentos (S1 a S4) da raiz de cana-de-açúcar (Figura 14) para os genes homólogos em *S. bicolor* que estão presentes na rede regulatória de *SbRAV1* (Figura 13).

Figura 14. Expressão de *ScRAV1* e dos transcritos de parede celular nos quatro segmentos (S1 a S4) da raiz de cana-de-açúcar que apresentam genes homólogos em *S. bicolor* com sítios putativos de ligação a RAV (AP2/B3). Os valores de expressão estão em TPM (*Transcript per Million*) em $\log_2$ da média de três réplicas biológicas de cada segmento oriundo do RNA-Seq.

Pode-se verificar no heatmap que há dois agrupamentos que são identificados por I e II, os quais podem ser subdividido em 4 subgrupos (A – D) (Figura 14). O agrupamento I apresenta maiores níveis transcricionais em S1, e o oposto pode ser visto para o agrupamento II. Pela clusterização ainda pode-se verificar que os S2 e S3 apresentam padrões de expressão

mais similares entre si, e que o S4 é mais parecido com os S2 e S3 que em relação ao S1 (Figura 14).

Os transcritos do subgrupo A consistem em um grupo altamente heterogêneo. Os membros deste grupo estão relacionados tanto à biossíntese quanto a hidrólise de polissacarídeos (Tabela 2). Estes transcritos apresentam principalmente um pico de expressão nos S2 e S3 (Figura 14).

Por outro lado, os transcritos do subgrupo B estão principalmente relacionados à biossíntese dos componentes da parede celular (Tabela 2), excluindo os transcritos *ScEGase3* e *ScCHI2* que se relacionam com a hidrólise de hemicelulose/celulose e quitina, respectivamente (Tabela 2). Esses transcritos apresentam uma maior expressão principalmente no S1 e têm uma tendência a diminuir em direção ao S4 da raiz de cana-de-açúcar (Figura 14). Além disso, pode-se observar que o fator de transcrição *ScRAV1* apresentou um nível de expressão maior nos S1 e S2 comparando com os S3 e S4.

Os transcritos do subgrupo C também formam um grupo diverso, o qual apresenta alguns elementos relacionados à hidrólise de hemicelulose e pectina (Tabela 2). Pode-se verificar que estes transcritos apresentam maior expressão particularmente nos S1 e S4, com exceção do transcrito *ScEGase4* que apresenta uma maior expressão nos S2 e S4 (Figura 14).

Já os transcritos do subgrupo D estão predominantemente relacionados à biossíntese de lignina (Tabela 2), e apresentam maior expressão preferencialmente em S3 e S4 (Figura 14), com exceções como *ScABF1*, *ScC3H1* e *ScCOMT2* que demonstram um maior nível transcricional no S2.

Entretanto, podemos observar que a maioria dos níveis de expressão dos transcritos de parede celular não é o inverso da expressão de *ScRAV1* no RNA-Seq em cana-de-açúcar. Isto seria esperado pela possível regulação dos níveis transcricionais dos genes de parede

celular por *ScRAV1*, baseando-se na predição de sítios putativos de RAV em genes homólogos de *S. bicolor* (Tabela 5), e na atividade repressiva de *ScRAV1* e seu ortólogo *SbRAV1* em *S. bicolor* (Figura 9), conforme predito na rede de regulação gênica (Figura 13).

4. Redes de co-expressão de *ScRAV1* com os fatores de transcrição

Foram feitas também 4 redes de co-expressão de *ScRAV1* com os fatores de transcrição presentes nos diferentes segmentos (S1 a S4) da raiz de cana-de-açúcar (Figura 15), os valores de correlação podem ser vistos no Apêndice C. Com o propósito de verificar transcritos que podem estar co-expressos com *ScRAV1* e que poderiam estar envolvidos nos diferentes processos que ocorrem para o desenvolvimento do aerênquima na raiz de cana-de-açúcar.

Verifica-se que *ScRAV1* apresenta uma maior quantidade de correlações com estes fatores de transcrição nos S1 e S4 (Figura 15A e D), comparando com as redes dos S2 e S3 (Figura 15B e C). Já no que se refere à quantidade e às porcentagens de correlações positivas e negativas obtidas em cada uma das redes de *ScRAV1*, estas podem ser vistas na Tabela 6.

Tabela 6. Quantidades e porcentagens de correlações positivas e negativas de *ScRAV1* com os fatores de transcrição nas redes de cada segmento da raiz de cana-de-açúcar.

Rede dos Segmentos	Nº de Correlações Positivas	Porcentagem (%)	Nº de Correlações Negativas	Porcentagem (%)	Total
S1	51	87,93	7	12,07	58
S2	6	25	18	75	24
S3	18	52,94	16	47,06	34
S4	32	71,11	13	28,89	45
Total	107		54		161

Pode-se observar na rede no S2 que *ScRAV1* possui um maior número de correlações negativas com os fatores de transcrição. Já na rede do S3 apresenta porcentagens próximas a 50% entre os dois tipos de correlações de *ScRAV1* com os fatores de transcrição (Tabela 6). Diferentemente de S2 e S3, nas redes dos S1 e S4 de *ScRAV1* pode-se verificar que há um maior número de correlações positivas que negativas com os fatores de transcrição (Tabela 6). Esses dados demonstram uma variação no padrão do tipo de correlação de *ScRAV1* com

os fatores de transcrição nas redes de cada segmento da raiz de cana-de-açúcar, assim como visto para os transcritos de parede celular (item 3, Tabela 1).

Figura 15. Redes de co-expressão de *ScRAV1* (círculo em cor vermelho escuro) com os fatores de transcrição nos diferentes segmentos da raiz da cana-de-açúcar, feitas com a correlação de Pearson e com o corte de p-valor $\leq 0,05$. (A) Rede do S1, (B) Rede do S2, (C) Rede do S3 e (D) Rede do S4. A legenda à direita mostra as cores dos tipos de correlação (linhas - vermelha escuro representa correlações negativas e rosa claro as correlações positivas) observadas para os fatores de transcrição nas diferentes redes de *ScRAV1*. A identificação dos fatores de transcrição correlacionados com *ScRAV1* em cada segmento da raiz, e seus respectivos valores de correlação e p-valores encontram-se no Apêndice C.

4.1. Avaliação das redes de co-expressão dos fatores de transcrição com *ScRAV1*

Com base nas redes de correlações de *ScRAV1* com os fatores de transcrição nos S1 ao S4, como mostra a Figura 15, avaliaram-se as porcentagens relativas das famílias de fatores de transcrição em cada uma das redes (Figura 16).

Figura 16. Distribuição das famílias de fatores de transcrição presentes nas redes de *ScRAV1* nos S1, S2, S3 e S4 na raiz de cana-de-açúcar.

De forma geral, as redes dos S1 e S4 apresentam um maior número de famílias de fatores de transcrição que se co-expressam com *ScRAV1*, com 27 e 24, respectivamente. Diferentemente das redes dos S2 e S3 que apresentam apenas 19 famílias de fatores de transcrição em cada uma (Figura 16).

É possível verificar que as famílias de fatores de transcrição que estão co-expressas com *ScRAV1* em todas as redes são: AP2/EREBP, bZIP, C2H2, C3H, HOMEBOX, NAC, PHD e SET (Figura 16). Entretanto, há famílias de fatores de transcrição como ARR-B e MYB que aparecem somente nas redes dos S1 ao S3 (Figura 16). Outros exemplos são as famílias de fatores de transcrição como ARF e GRAS que estão presentes somente nas redes dos S2 ao S4

(Figura 16). Já outras famílias de fatores de transcrição aparecem apenas em algumas redes, como é o caso de DBP e WRKY, entre outros (Figura 16).

Em relação às famílias de fatores de transcrição mais representativas em cada uma das redes de *ScRAV1*, pode-se averiguar na rede do S1 que as famílias de fatores de transcrição que estão em maiores proporções são as famílias bZIP, DBP e MYB (Figura 16). Já na rede de co-expressão do S2, as famílias de fatores de transcrição que predominam são as famílias C2H2, bHLH, bZIP e GRAS (Figura 16). Por outro lado, nas redes de co-expressão dos S3 e S4, as famílias que estão mais representadas são C2H2, DBP, MYB e WRKY e C2H2, NAC, PHD e SET, respectivamente (Figura 16).

Assim, pode-se concluir que dentre as famílias mais significativas nas redes de *ScRAV1*, a família bZIP aparece principalmente nas redes S1 e S2, enquanto as famílias DBP e MYB aparecem nas redes S1 e S3 (Figura 16). Já a família C2H2 aparece nas redes S2, S3 e S4. Porém, outras famílias estão bem representadas em apenas uma das redes de *ScRAV1*, como é o caso das famílias de fatores de transcrição bHLH, GRAS, WRKY, NAC, PHD e SET.

4.2. Seleção dos fatores de transcrição pela centralidade de grau

Dentre os fatores de transcrição co-expressos com *ScRAV1* (Figura 15), verificou-se o aspecto topológico das redes levando-se em consideração as 10 principais centralidades de grau em cada um dos segmentos. O propósito foi buscar os fatores de transcrição mais importantes em cada uma das redes, e que possivelmente podem estar relacionados com a formação do aerênquima em cana-de-açúcar (Tabela 7).

Tabela 7. As dez maiores centralidades de grau em cada rede de *ScRAV1* com os fatores de transcrição nos diferentes segmentos (S1 a S4) da raiz de cana-de-açúcar.

Redes dos Segmentos	SAS	Identificação	Correlação	Centralidade de Grau
S1	SCVPST1060E04.g	ScbHLH6	Positiva	58
	SCVPRZ2042G02.g	ScMYBR1	Positiva	58
	SCACLR1127A11.g	ScbHLH1	Positiva	58
	SCCCCL3001D02.g	ScNAC2	Positiva	58
	SCVPLB1018G11.g	ScbZIP10	Positiva	58
	SCCCLR1C05B03.g	ScMYB2	Positiva	57
	SCCCRT1002F09.g	ScbZIP4	Positiva	56
	SCVPRZ2039D09.g	ScDBP11	Positiva	55
	SCSGFL4C07B02.g	ScABI2	Positiva	55
	SCMCSB1109C11.g	ScNAC8	Positiva	55
S2	SCSFLR2009G10.g	ScSET4	Negativa	24
	SCCCRZ2001A12.g	ScbHLH4	Negativa	24
	SCJFRZ2010A05.g	ScHB6	Positiva	23
	SCSFLR2016F09.g	ScbZIP8	Positiva	23
	SCEZLR1009A01.g	ScARF3	Positiva	21
	SCEQRT2025D01.g	ScJMJ1	Negativa	21
	SCBFRZ2016F05.g	ScGRAS1	Negativa	19
	SCEPCL6029A11.g	ScC2H5	Positiva	18
	SCCCFL2002D10.g	ScNAC3	Negativa	18
	SCCCCL7038B02.g	ScPHD1	Negativa	17
S3	SCSFRT2072F01.g	ScMYB9	Negativa	33
	SCQGLR2025E03.g	ScWRKY3	Negativa	32
	SCCCCL4014B04.g	ScC2H2	Negativa	30
	SCUTFL1055E01.g	ScC3H7	Positiva	30
	SCCCCL2001C11.b	ScSET2	Negativa	30
	SCQGLR2017G12.g	ScDBP6	Negativa	30
	SCCCLB1024B02.g	ScZIM2	Negativa	30
	SCBGLR1003E02.g	ScDOF1	Negativa	30
	SCBGLR1113F05.g	ScMYB1	Negativa	30
	SCEZLR1009A01.g	ScARF3	Positiva	29
S4	SCCCST3005F09.g	ScNAC7	Positiva	45
	SCCCLR1022F12.g	ScARF1	Negativa	44
	SCEQRT2026B09.g	ScBZR1	Positiva	44
	SCCCLR1C05E08.g	ScCOL1	Positiva	44
	SCCCLR1048F08.g	ScJUN2	Positiva	44
	SCCCLR1C04H01.g	ScNAC5	Positiva	44
	SCCCSB1002C05.g	ScABI1	Negativa	43
	SCACLR1130E07.g	ScbHLH2	Negativa	43
	SCCCRZ1004G09.g	ScC2H4	Positiva	42
	SCCCLR1C04E10.g	ScPHD3	Positiva	41

As centralidades de grau para fatores de transcrição presentes nas redes de co-expressão de *ScRAV1* (Tabela 7), podem ser visualizadas na Figura 15. Na rede do S1, as centralidades de grau que contêm maior representatividade são as famílias de fatores de transcrição bHLH, MYB, NAC e bZIP. Na rede do S2, observa-se que há apenas um

representante de cada família de fatores de transcrição. Já na rede do S3, a família MYB está em maior proporção, enquanto na rede do S4 é a família NAC. Porém, também se observa que algumas famílias de fatores de transcrição como ARF, bZIP, bHLH, C2H2, DBP, GRAS, PHD e WRKY estão presentes em algumas redes, como também descrito no item 4.1 (Figura 16).

4.3. Busca do homólogo mais próximo em *S. bicolor* e em *A. thaliana*

As 40 maiores centralidades de grau selecionadas para fatores de transcrição de cana-de-açúcar encontradas nas redes de co-expressão com *ScRAV1* nos S1 ao S4 (Tabela 7), permitiu buscar os melhores *hits* em homólogos de *S. bicolor* (Tabela 8). Pode-se verificar que 31 (77,5%) sequências em cana-de-açúcar apresentaram $\geq 90\%$ de identidade com as sequências de *S. bicolor* (Tabela 8).

Em seguida, foi feita a anotação em *S. bicolor* para cada fator de transcrição em cana-de-açúcar que permitiu buscar os melhores *hits* em *A. thaliana* (Tabela 9). Apenas 4 (12,90%) das sequências de *S. bicolor* apresentaram $\geq 70\%$ de identidade com as sequências de *A. thaliana* (Tabela 9).

Tabela 8. Os melhores *hits* preditos em *S. bicolor* que apresentaram $\geq 90\%$ de identidade com os fatores de transcrição da raiz de cana-de-açúcar para as maiores centralidades de grau (Tabela 7).

SAS	Sub_category1	Identificação	Correlação de <i>ScRAV1</i>		<i>S. bicolor</i>		Identidade (%)	E-value
			Segmento	Tipo	Hit	Descrição		
SCVPRZ2042G02.g	MYB-RELATED	ScMYBR1	S1	Positiva	Sobic.004G222800.1	similar to Os02g0631200 protein	93	0
SCCCCL3001D02.g	NAC	ScNAC2	S1	Positiva	Sobic.001G386000.1	similar to Putative uncharacterized protein	96	0
SCVPLB1018G11.g	BZIP	ScbZIP10	S1	Positiva	Sobic.010G081800.2	similar to Putative uncharacterized protein	93	6,00E-144
SCCCLR1C05B03.g	MYB	ScMYB2	S1	Positiva	Sobic.002G279100.1	similar to Transcription factor MYB31	94	0
SCVPRZ2039D09.g	DBP	ScDBP11	S1	Positiva	Sobic.001G406300.1	similar to Protein phosphatase type 2-C	97	0
SCSGFL4C07B02.g	ABI3/VP1	ScABI2	S1	Positiva	Sobic.010G229000.1	similar to Auxin response factor 17	97	6,00E-107
SCMCSB1109C11.g	NAC	ScNAC8	S1	Positiva	Sobic.010G111800.1	similar to Putative uncharacterized protein	95	0
SCSFLR2009G10.g	SET	ScSET4	S2/S4	Negativa/Positiva	Sobic.003G241700.2	PTHR13793//PTHR13793:SF102 - PHD FINGER PROTEINS // SUBFAMILY NOT NAMED (1 of 1)	96	1,00E-156
SCCCRZ2001A12.g	BHLH	ScbHLH4	S2	Negativa	Sobic.010G145200.1	similar to Helix-loop-helix DNA-binding domain containing protein, expressed	93	6,00E-128
SCJFRZ2010A05.g	HOMEODOMAIN	ScHB6	S2	Positiva	Sobic.009G159900.1	similar to Putative homeodomain protein	96	0
SCSFLR2016F09.g	BZIP	ScbZIP8	S2	Positiva	Sobic.001G380700.3		97	0
SCZLR1009A01.g	ARF	ScARF3	S2/S3	Positiva/Positiva	Sobic.003G411900.1	similar to Auxin response factor 4	97	0
SEQRT2025D01.g	JUMONJI	ScJMJ1	S2	Negativa	Sobic.001G284400.1	similar to Transcription factor jumonji, putative, expressed	96	0
SCBFRZ2016F05.g	GRAS	ScGRAS1	S2	Negativa	Sobic.006G172900.1	similar to OSJNBa0010H02.25 protein	97	0
SCPECL6029A11.g	C2H2	ScC2H5	S2	Positiva	Sobic.007G225100.1	similar to Putative uncharacterized protein	93	0
SCCCFL2002D10.g	NAC	ScNAC3	S2/S4	Negativa/Negativa	Sobic.003G035100.1	PTHR31079:SF3 - NAC DOMAIN CONTAINING PROTEIN 75-RELATED (1 of 2)	100	4,00E-67
SCCCCL7038B02.g	PHD	ScPHD1	S2	Negativa	Sobic.005G034700.1	similar to Expressed protein	94	0
SCSFRT2072F01.g	MYB	ScMYB9	S3	Negativa	Sobic.004G303600.1	similar to ATMYB4-like	90	1,00E-103
SCQGLR2025E03.g	WRKY	ScWRKY3	S3	Negativa	Sobic.001G332500.1	similar to WRKY DNA binding domain containing protein, expressed	91	2,00E-120
SCCCCL4014B04.g	C2H2	ScC2H2	S3	Negativa	Sobic.003G370700.2	similar to Putative transparent testa 1	97	0
SCCCCL2001C11.b	SET	ScSET2	S3/S4	Negativa/Negativa	Sobic.006G080300.1	similar to SET domain-containing protein SET102	95	0
SCQGLR2017G12.g	DBP	ScDBP6	S3	Negativa	Sobic.001G309500.1	similar to Protein phosphatase 2C, putative, expressed	98	0
SCCCLB1024B02.g	ZIM	ScZIM2	S3	Negativa	Sobic.002G374100.1	similar to Putative uncharacterized protein	95	5,00E-138
SCCCST3005F09.g	NAC	ScNAC7	S4	Positiva	Sobic.003G251800.1	similar to No apical meristem (NAM) protein-like	91	4,00E-157
SCCCLR1022F12.g	ARF	ScARF1	S4	Negativa	Sobic.005G132000.1	similar to Auxin response factor 23	94	0
SCCCLR1C05E08.g	C2C2-CO-LIKE	ScCOL1	S4	Positiva	Sobic.004G208400.1	similar to Zinc finger protein	92	2,00E-169
SCCCLR1048F08.g	JUN ACTIVADOR	ScJUN2	S4	Positiva	Sobic.003G124800.1	similar to 26S proteasome non-ATPase regulatory subunit 14	100	0
SCCCLR1C04H01.g	NAC	ScNAC5	S4	Positiva	Sobic.008G094700.1	similar to No apical meristem protein, expressed	90	4,00E-154
SCACLR1130E07.g	BHLH	ScbHLH2	S4	Negativa	Sobic.002G054600.3	PTHR16223:SF15 - TRANSCRIPTION FACTOR UNE12-RELATED (1 of 2)	96	0
SCCCRZ1004G09.g	C2H2	ScC2H4	S2/S4	Negativa/Positiva	Sobic.008G083000.3	similar to Zinc finger family protein, putative, expressed	98	0
SCCCLR1C04E10.g	PHD	ScPHD3	S3/S4	Positiva/Positiva	Sobic.001G522600.3	similar to OSJNBa0035I04.11 protein	91	0

Tabela 9. Os melhores *hits* em *A. thaliana* com base nas sequências de *S. bicolor* preditas anteriormente para os fatores de transcrição na raiz de cana-de-açúcar.

SAS	Sub_category1	Identificação	Correlação de ScRAV1		<i>S. bicolor</i>		<i>A. thaliana</i>		Símbolo	Identidade (%)	E-value
			Segmento	Tipo	Hit	Descrição	Hit	Descrição			
SCVPRZ2042G02.g	MYB-RELATED	ScMYBR1	S1	Positiva	Sobic.004G222800.1	similar to Os02g0631200 protein	AT1G60670.2	Protein of unknown function (DUF3755)		37,74	2,00E-46
SCCCCL3001D02.g	NAC	ScNAC2	S1	Positiva	Sobic.001G386000.1	similar to Putative uncharacterized protein	AT5G18270.1	<i>Arabidopsis</i> NAC domain containing protein 87	ANAC087	46,30	4,00E-99
SCVPLB1018G11.g	BZIP	ScbZIP10	S1	Positiva	Sobic.010G081800.2	similar to Putative uncharacterized protein	AT1G45249.1	abscisic acid responsive elements-binding factor 2	ABF2, AREB1, ATAREB1	65,09	7,00E-40
SCCCLR1C05B03.g	MYB	ScMYB2	S1	Positiva	Sobic.002G279100.1	similar to Transcription factor MYB31	AT4G38620.1	myb domain protein 4	ATMYB4, MYB4	55,92	8,00E-103
SCVPRZ2039D09.g	DBP	ScDBP11	S1	Positiva	Sobic.001G406300.1	similar to Protein phosphatase type 2-C	AT1G07160.1	Protein phosphatase 2C family protein		49,24	6,00E-118
SCSGFL4C07B02.g	ABI3/VP1	ScABI2	S1	Positiva	Sobic.010G229000.1	similar to Auxin response factor 17	AT1G30330.2	auxin response factor 6	ARF6	79,64	0
SCMCSB1109C11.g	NAC	ScNAC8	S1	Positiva	Sobic.010G111800.1	similar to Putative uncharacterized protein	AT1G25580.1	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein	NAC008, SOG1	51,18	1,00E-153
SCSFLR2009G10.g	SET	ScSET4	S2/S4	Negativa/Positiva	Sobic.003G241700.2	PTHR13793//PTHR13793:SF102 - PHD FINGER PROTEINS // SUBFAMILY NOT NAMED (1 of 1)	AT4G27910.1	SET domain protein 16	ATX4, SDG16	55,01	0
SCCCRZ2001A12.g	BHLH	ScbHLH4	S2	Negativa	Sobic.010G145200.1	similar to Helix-loop-helix DNA-binding domain containing protein, expressed	AT1G29950.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	SACL3	35,71	1,00E-32
SCJFRZ2010A05.g	HOMEODOMAIN	ScHB6	S2	Positiva	Sobic.009G159900.1	similar to Putative homeodomain protein	AT5G02030.1	POX (plant homeobox) family protein	LSN, PNY, HB-6, BLR, RPL, BLH9, VAN	41,38	5e-82
SCSFLR2016F09.g	BZIP	ScbZIP8	S2	Positiva	Sobic.001G380700.3		AT2G40620.1	Basic-leucine zipper (bZIP) transcription factor family protein	BZIP18	48,06	1,00E-93
SCEZLR1009A01.g	ARF	ScARF3	S2/S3	Positiva/Positiva	Sobic.003G411900.1	similar to Auxin response factor 4	AT5G62000.1	auxin response factor 2	ARF1-BP, ARF2, HSS, ORE14	55,46	0
SCEQRT2025D01.g	JUMONJI	ScJMJ1	S2	Negativa	Sobic.001G284400.1	similar to Transcription factor jumonji, putative, expressed	AT5G46910.1	Transcription factor jumonji (jmi) family protein / zinc finger (CSHC2 type) family protein		52,37	0
SCBFRZ2016F05.g	GRAS	ScGRAS1	S2	Negativa	Sobic.006G172900.1	similar to OSJNBa0010H02.25 protein	AT4G00150.1	GRAS family transcription factor	ATHAM3, HAM3, LOM3, SCL6-IV	48,04	7,00E-93
SCEPCL6029A11.g	C2H2	ScC2H5	S2	Positiva	Sobic.007G225100.1	similar to Putative uncharacterized protein	AT1G22310.1	methyl-CPG-binding domain 8	MBD8, ATMBD8	37,44	2,00E-33
SCCCFL2002D10.g	NAC	ScNAC3	S2/S4	Negativa/Negativa	Sobic.003G035100.1	PTHR31079:SF3 - NAC DOMAIN CONTAINING PROTEIN 75-RELATED (1 of 2)	AT4G29230.1	NAC domain containing protein 75	anac075, NAC075	51,61	2,00E-150
SCCCCL7038B02.g	PHD	ScPHD1	S2	Negativa	Sobic.005G034700.1	similar to Expressed protein	AT5G35210.2	metalloendopeptidases; zinc ion binding; DNA binding	PTM	32,46	5,00E-156
SCSFRT2027F01.g	MYB	ScMYB9	S3	Negativa	Sobic.004G303600.1	similar to ATMYB4-like	AT5G52260.1	myb domain protein 19	AtMYB19, MYB19	48,50	1,00E-58
SCQGLR2025E03.g	WRKY	ScWRKY3	S3	Negativa	Sobic.001G332500.1	similar to WRKY DNA binding domain containing protein, expressed	AT2G03340.1	WRKY DNA-binding protein 3	WRKY3	43,59	7,00E-110

Tabela 9. Conclusão.

SCCCCL4014B04.g	C2H2	ScC2H2	S3	Negativa	Sobic.003G370700.2	similar to Putative transparent testa 1	AT1G34370.2	C2H2 and C2HC zinc fingers superfamily protein	AtSTOP1, STOP1	55,02	2,00E-158
SCCCCL2001C11.b	SET	ScSET2	S3/S4	Negativa/Negativa	Sobic.006G080300.1	similar to SET domain-containing protein SET102	AT1G76710.1	SET domain group 26	ASHH1, SDG26	55,14	0
SCQGLR2017G12.g	DBP	ScDBP6	S3	Negativa	Sobic.001G309500.1	similar to Protein phosphatase 2C, putative, expressed	AT4G38520.2	Protein phosphatase 2C family protein	APD6	68,89	0
SCCCLB1024B02.g	ZIM	ScZIM2	S3	Negativa	Sobic.002G374100.1	similar to Putative uncharacterized protein	AT1G74950.1	TIFY domain/Divergent CCT motif family protein	JAZ2, TIFY10B	37,55	1,00E-33
SCCCST3005F09.g	NAC	ScNAC7	S4	Positiva	Sobic.003G251800.1	similar to No apical meristem (NAM) protein-like	AT4G28500.1	NAC domain containing protein 73	ANAC073, NAC073, SND2	71,72	4,00E-128
SCCCLR1022F12.g	ARF	ScARF1	S4	Negativa	Sobic.005G132000.1	similar to Auxin response factor 23	AT5G62000.1	auxin response factor 2	ARF1-BP, ARF2, HSS, ORE14	53,74	0
SCCCLR1C05E08.g	C2C2-CO-LIKE	ScCOL1	S4	Positiva	Sobic.004G208400.1	similar to Zinc finger protein	AT1G06040.1	B-box zinc finger family protein	BBX24, STO	51,69	1,00E-79
SCCCLR1048F08.g	JUN ACTIVADOR	ScJUN2	S4	Positiva	Sobic.003G124800.1	similar to 26S proteasome non-ATPase regulatory subunit 14	AT5G23540.1	Mov34/MPN/PAD-1 family protein		93,20	0
SCCCLR1C04H01.g	NAC	ScNAC5	S4	Positiva	Sobic.008G094700.1	similar to No apical meristem protein, expressed	AT5G13180.1	NAC domain containing protein 83	ANAC083, NAC083, VNI2	42,92	6,00E-58
SCACLR1130E07.g	BHLH	ScbHLH2	S4	Negativa	Sobic.002G054600.3	PTHR16223:SF15 - TRANSCRIPTION FACTOR UNE12-RELATED (1 of 2)	AT4G02590.3	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	UNE12	77,46	3,00E-78
SCCCRZ1004G09.g	C2H2	ScC2H4	S2/S4	Negativa/Positiva	Sobic.008G083000.3	similar to Zinc finger family protein, putative, expressed	AT5G63280.1	C2H2-like zinc finger protein		62,90	8,00E-115
SCCCLR1C04E10.g	PHD	ScPHD3	S3/S4	Positiva/Positiva	Sobic.001G522600.3	similar to OSJNBa0035I04.11 protein	AT2G19260.1	RING/FYVE/PHD zinc finger superfamily protein		46,38	1,00E-55

4.4. Enriquecimento funcional dos fatores de transcrição através dos termos do Gene Ontology (GO)

A anotação manual em *A. thaliana* com base nos homólogos em *S. bicolor* (Tabela 9), permitiu prever as possíveis funções ou papéis biológicos dos fatores de transcrição em cana-de-açúcar através dos termos do Gene Ontology (GO) para componente celular, processo biológico e função molecular para os fatores de transcrição (Figura 17). Todos os termos GOs podem ser encontrados em *A. thaliana* (Apêndice E).

Figura 17. Distribuição dos primeiros 20 termos do Gene Ontology (GO) para componente celular (A), processo biológico (B) e função molecular (C) dos fatores de transcrição de cana-de-açúcar anotados em *A. thaliana* com base nos homólogos em *S. bicolor*.

Pode-se verificar para os termos do componente celular (Figura 17A) que estas proteínas estão predominantemente associadas ao termo *nucleus* (GO:0005634), e em menor proporção estão presentes em outros compartimentos celulares.

Para os termos de processos biológicos (Figura 17B), pode-se observar que o principal termo é *regulation of transcription, DNA-templated* (GO:0006355). Além disso, pode-se observar que estes fatores de transcrição podem participar em diferentes processos biológicos. Entre os termos encontrados estão: *response to salt stress* (GO:0009651), *flower development* (GO:0009908), *regulation of defense response* (GO:0031347), *leaf senescence* (GO:0010150), *response to wounding* (GO:0009611). Outros fatores de transcrição estão associados à formação da parede secundária, conforme os termos *xylem development* (GO:0010089) e *regulation of secondary cell wall biogenesis* (GO:2000652). Por último, podemos verificar que alguns fatores de transcrição estão relacionados a hormônios, como pode ser visto pelos termos: *auxin-activated signaling pathway* (GO:0009734) e *response to abscisic acid* (GO:0009737).

Com relação aos termos de função molecular associados aos fatores de transcrição anotados neste trabalho (Figura 17C), predomina-se o termo *DNA-binding transcription factor activity* (GO:0003700). Além disso, essas proteínas podem apresentar diferentes tipos de atividades enzimáticas ou de ligação à substratos.

Com base nestes resultados, pode-se inferir que estes fatores de transcrição estão associados a diferentes papéis biológicos em cana-de-açúcar.

4.5. Sítios de ligação ao DNA para RAV nas sequências homólogas de *S. bicolor*

A partir das sequências homólogas de *S. bicolor* para os fatores de transcrição em cana-de-açúcar (Tabela 8), procurou-se nas regiões promotoras dos genes de *S. bicolor* sítios de ligação de RAV ao DNA. Na Tabela 10, pode-se verificar os sítios putativos de ligação de RAV para os motivos AP2 (CAACA) e B3 (CACCTG) encontrados nas regiões promotoras dos genes de fatores de transcrição de *S. bicolor*. Dentre as 31 sequências de *S. bicolor* que apresentaram alta identidade com as sequências de cana-de-açúcar, apenas a sequência *ScDBP11* (Sobic.001G406300.1) não apresentou sítios de ligação de RAV ao DNA, sobrando apenas 30 genes com os sítios putativos para este fator de transcrição (Tabela 10).

Foram encontrados 145 sítios para RAV. Cento e vinte (82,76%) destes sítios referem-se ao motivo AP2 (CAACA) e apenas 25 (17,24%) referem-se ao motivo B3 (CACCTG) (Tabela 10). Ao contrário do observado para os transcritos de parede celular, cerca de 17 (56,67%) das sequências apresentam sítios para o motivo AP2 (CAACA) para RAV e apenas 13 (43,33%) das sequências apresentam sítios para ambos os motivos AP2 e B3 (CAACA/CACCTG) para RAV. Pode-se concluir que há uma divergência na quantidade de motivos AP2 e B3 para os sítios de ligação de RAV ao DNA nas regiões promotoras dos genes de fatores de transcrição em *S. bicolor*.

Tabela 10. Sítios putativos de ligação de RAV (AP2/B3) na região promotora de fatores de transcrição dentre os possíveis ortólogos previamente selecionados de *S. bicolor*.

SAS	Sub_category1	Identificação	Correlação de <i>ScRAV1</i>		<i>S. bicolor</i> Hit	Localização	Fita	Motivo	Quantidade de Sítios		
			Segmento	Tipo					AP2	B3	Total
SCVPRZ2042G02.g	MYB-RELATED	ScMYBR1	S1	Positiva	Sobic.004G222800.1	443	+	CAACA	2	1	3
						479	-	CAACA			
						524	-	CACCTG			
SCCCCL3001D02.g	NAC	ScNAC2	S1	Positiva	Sobic.001G386000.1	947	-	CAACA	2	0	2
						1119	+	CAACA			
SCVPLB1018G11.g	BZIP	ScbZIP10	S1	Positiva	Sobic.010G081800.2	489	+	CACCTG	1	3	4
						570	+	CAACA			
						703	+	CACCTG			
						1316	+	CACCTG			
SCCCLR1C05B03.g	MYB	ScMYB2	S1	Positiva	Sobic.002G279100.1	112	-	CAACA	6	0	6
						400	-	CAACA			
						622	+	CAACA			
						650	+	CAACA			
						847	-	CAACA			
						1988	+	CAACA			
SCSGFL4C07B02.g	ABI3/VP1	ScABI2	S1	Positiva	Sobic.010G229000.1	113	+	CAACA	4	1	5
						380	+	CAACA			
						660	+	CAACA			
						782	-	CACCTG			
SCMCSB1109C11.g	NAC	ScNAC8	S1	Positiva	Sobic.010G111800.1	1611	-	CAACA	5	1	6
						230	+	CAACA			
						648	-	CAACA			
						814	+	CAACA			
						1235	+	CACCTG			
						1703	-	CAACA			
SCSFLR2009G10.g	SET	ScSET4	S2/S4	Negativa/Positiva	Sobic.003G241700.2	1944	-	CAACA	3	0	3
						527	+	CAACA			
						614	+	CAACA			
SCCCRZ2001A12.g	BHLH	ScbHLH4	S2	Negativa	Sobic.010G145200.1	1326	+	CAACA	2	0	2
						151	+	CAACA			
SCJFRZ2010A05.g	HOMEBOX	ScHB6	S2	Positiva	Sobic.009G159900.1	347	+	CAACA	4	1	5
						56	-	CAACA			
						62	+	CAACA			
						346	+	CAACA			
SCSFLR2016F09.g	BZIP	ScbZIP8	S2	Positiva	Sobic.001G380700.3	594	+	CAACA	3	0	3
						1366	+	CACCTG			
						203	-	CAACA			
SCEZLR1009A01.g	ARF	ScARF3	S2/S3	Positiva/Positiva	Sobic.003G411900.1	295	+	CAACA	1	0	1
						893	+	CAACA			
SCEQRT2025D01.g	JUMONJI	ScJMJ1	S2	Negativa	Sobic.001G284400.1	13	-	CAACA	4	1	5
						157	+	CAACA			
						890	+	CAACA			
						1218	+	CACCTG			
SCBFRZ2016F05.g	GRAS	ScGRAS1	S2	Negativa	Sobic.006G172900.1	1931	-	CAACA	5	2	7
						1980	-	CAACA			
						361	+	CAACA			
						391	+	CACCTG			
						948	+	CAACA			
						1012	+	CAACA			
SCEPCL6029A11.g	C2H2	ScC2H5	S2	Positiva	Sobic.007G225100.1	1622	-	CAACA	11	1	12
						1842	+	CAACA			
						1853	-	CACCTG			

Tabela 10. Continuação.

						450	+	CAACA			
						502	-	CAACA			
						610	-	CAACA			
						683	+	CAACA			
						735	-	CAACA			
						1174	+	CAACA			
						1205	+	CAACA			
						1414	+	CACCTG			
						1613	+	CAACA			
						1638	+	CAACA			
SCCCFL2002D10.g	NAC	ScNAC3	S2/S4	Negativa/Negativa	Sobic.003G035100.1	1871	+	CAACA	5	0	5
						102	-	CAACA			
						122	-	CAACA			
						819	-	CAACA			
						836	-	CAACA			
SCCCCL7038B02.g	PHD	ScPHD1	S2	Negativa	Sobic.005G034700.1	1727	+	CAACA	7	2	9
						71	+	CAACA			
						79	+	CAACA			
						251	+	CAACA			
						254	+	CACCTG			
						333	+	CAACA			
						640	+	CAACA			
						648	+	CAACA			
						932	+	CAACA			
						1356	+	CACCTG			
SCSFRT2072F01.g	MYB	ScMYB9	S3	Negativa	Sobic.004G303600.1	254	+	CACCTG	3	3	6
						461	-	CACCTG			
						1150	-	CACCTG			
						1396	-	CAACA			
						1701	+	CAACA			
SCQGLR2025E03.g	WRKY	ScWRKY3	S3	Negativa	Sobic.001G332500.1	1843	-	CAACA	3	0	3
						200	+	CAACA			
						220	+	CAACA			
SCCCCL4014B04.g	C2H2	ScC2H2	S3	Negativa	Sobic.003G370700.2	1420	+	CAACA	3	2	5
						493	+	CACCTG			
						516	+	CACCTG			
						1017	-	CAACA			
						1066	-	CAACA			
SCCCCL2001C11.b	SET	ScSET2	S3/S4	Negativa/Negativa	Sobic.006G080300.1	1989	-	CAACA	4	6	10
						8	+	CACCTG			
						141	+	CAACA			
						191	-	CAACA			
						256	+	CACCTG			
						309	+	CACCTG			
						506	+	CAACA			
						1431	+	CACCTG			
						1703	-	CACCTG			
						1773	+	CACCTG			
SCQGLR2017G12.g	DBP	ScDBP6	S3	Negativa	Sobic.001G309500.1	1964	-	CAACA	3	0	3
						338	-	CAACA			
						432	-	CAACA			
SCCCLB1024B02.g	ZIM	ScZIM2	S3	Negativa	Sobic.002G374100.1	1446	-	CAACA	1	1	2
						1299	+	CAACA			
SCCCST3005F09.g	NAC	ScNAC7	S4	Positiva	Sobic.003G251800.1	1500	+	CACCTG	3	0	3
						1110	+	CAACA			

Tabela 10. Conclusão.

SCCCLR1022F12.g	ARF	ScARF1	S4	Negativa	Sobic.005G132000.1	1871	+	CAACA	5	0	5									
						1874	+	CAACA												
						254	+	CAACA												
						866	-	CAACA												
						1085	-	CAACA												
						1104	+	CAACA												
SCCCLR1C05E08.g	C2C2-CO-LIKE	ScCOL1	S4	Positiva	Sobic.004G208400.1	1202	+	CAACA	6	0	6									
						89	-	CAACA												
						300	-	CAACA												
						724	-	CAACA												
						750	-	CAACA												
						1247	+	CAACA												
SCCCLR1048F08.g	JUN ACTIVADOR	ScJUN2	S4	Positiva	Sobic.003G124800.1	1993	-	CAACA	3	0	3									
						1036	+	CAACA												
						1308	+	CAACA												
						1618	-	CAACA												
						SCCCLR1C04H01.g	NAC	ScNAC5				S4	Positiva	Sobic.008G094700.1	382	+	CAACA	4	0	4
															890	+	CAACA			
1803	+	CAACA																		
1938	-	CAACA																		
SCACLR1130E07.g	BHLH	ScbHLH2	S4	Negativa	Sobic.002G054600.3				704	+	CAACA				5	0	5			
									789	+	CAACA									
						1054	+	CAACA												
						1261	+	CAACA												
						SCCCRZ1004G09.g	C2H2	ScC2H4	S2/S4	Negativa/Positiva	Sobic.008G083000.3	1688	-	CAACA				3	0	3
												768	+	CAACA						
1002	-	CAACA																		
1555	+	CAACA																		
SCCCLR1C04E10.g	PHD	ScPHD3	S3/S4	Positiva/Positiva	Sobic.001G522600.3							108	-	CAACA	9	0	9			
												243	+	CAACA						
						772	-	CAACA												
						1103	+	CAACA												
						1160	-	CAACA												
						1293	-	CAACA												
						1793	+	CAACA												
						1827	+	CAACA												
Total						1971	-	CAACA	120	25	145									

4.6. Rede de regulação gênica para os fatores de transcrição

O motivo repressor para *SbRAV1* (Figura 9) e a predição *in silico* para sítios putativos de RAV nos genes homólogos em *S. bicolor* para os fatores de transcrição de cana-de-açúcar (Tabela 10) foram usados para formar a rede de regulação gênica (GRN, *Gene Regulatory Network*), como mostra a Figura 18.

Figura 18. Representação da rede de regulação gênica dos potenciais alvos dos genes de fatores de transcrição que possivelmente podem ser coordenados por *SbRAV1* (círculo em vermelho) com base na predição *in silico* dos sítios putativos para RAV nos genes de fatores de transcrição em *S. bicolor* (Tabela 10). A identificação dos homólogos em *S. bicolor* para os fatores de transcrição de cana-de-açúcar foram nomeados apenas trocando *Sc* por *Sb* devido a busca de elementos *cis* nesta primeira espécie. As cores dos círculos na rede de regulação gênica referem-se aos processos biológicos descritos para estes transcritos com base em *A. thaliana*, conforme o Apêndice E. Contudo, as cores dos círculos *SbNAC2* (ANAC087) e *SbNAC3* (NAC075) foram anotados manualmente, de acordo com Buchanan-Wollaston *et al.* (2005) e Nakano *et al.*, (2015), respectivamente. À direita mostra a legenda com a atividade de *SbRAV1* baseada na conservação do seu motivo repressor B3 (BRD, *B3 repressor domain*) (Ikeda & Ohme-Takagi, 2009), como mostra a Figura 9.

Pode-se verificar na rede de regulação gênica de *SbRAV1* que este apresenta diferentes genes alvo de diferentes processos biológicos (Figura 18). A maioria dos fatores transcrição pertencem a classificação de regulação da transcrição e parede secundária (*SbNAC3*, *SbNAC5*, *SbNAC7*, *SbHB6* e *SbMYB2*) (Figura 18). Entre as famílias de fatores de transcrição mais representativas que podem ser alvos de *SbRAV1* estão cinco NACs (*SbNAC3*, *SbNAC5*, *SbNAC7* e *SbNAC8*) e três ARFs (*SbABI2*, *SbARF1* e *SbARF3*) (Figura 18).

4.7. Perfil de expressão dos fatores de transcrição com base no RNA-Seq

O heatmap abaixo mostra o perfil de expressão dos fatores de transcrição em cana-de-açúcar nos quatro segmentos da raiz (Figura 19) para os genes homólogos em *S. bicolor* que estão presentes na rede regulatória de *SbRAV1* (Figura 18).

Figura 19. Expressão de *ScRAV1* e dos fatores de transcrição nos quatro segmentos (S1 a S4) da raiz de cana-de-açúcar que apresentam genes homólogos em *S. bicolor* com sítios putativos de ligação a RAV (AP2/B3). Os valores de expressão estão em TPM (*Transcript per Million*) em $\log_2$ da média de três réplicas biológicas de cada segmento oriundo do RNA-Seq.

Pode-se verificar no heatmap que há dois agrupamentos que são identificados por I e II, os quais podem ser subdivididos em 4 subgrupos (A – D) (Figura 19). No agrupamento I, pode-se observar que os transcritos apresentam aumento de expressão principalmente nos S3 e S4, enquanto no agrupamento II, os transcritos apresentam maior expressão principalmente no

S1. A clusterização dos segmentos mostra que os S2 e S3 apresentam padrões de expressão mais similares entre si, e que o S4 é mais parecido com os S2 e S3 que em relação ao S1 (Figura 19).

Em relação à clusterização dos fatores de transcrição pode-se verificar que o subgrupo A apresenta uma maior expressão, principalmente no S3 (Figura 19). Já os transcritos do subgrupo B apresentam uma menor expressão no S1 e aumentam a expressão, tendo um pico nos S3 e S4. No entanto, os transcritos do subgrupo C apresentam maior expressão principalmente no S4, com exceção de *ScPHD3* e *ScNAC8* que apresentam maior expressão nos segmentos iniciais (S1 e S2) e também no S4.

O subgrupo D que apresenta um conjunto de transcritos altamente heterogêneo e são vistos com maior expressão principalmente no S1 (Figura 19). Algumas exceções deste tipo de padrão de expressão são *ScSET2* e *ScGRAS1* os quais apresentam um aumento na expressão nos S1 e S3 e diminuem nos S2 e S4, enquanto *ScHB6* apresenta uma alta expressão nos S1 a S3, e diminui no S4. Além disso, pode-se observar que o fator de transcrição *ScRAV1* apresentou um nível de expressão maior nos S1 e S2 comparando com os S3 e S4.

No entanto, podemos observar que os níveis de expressão da maioria dos transcritos de fatores de transcrição não é o inverso da expressão de *ScRAV1* no RNA-Seq em cana-de-açúcar, com exceção dos fatores de transcrição *ScARF1* e *ScNAC2* (subgrupo A) e os fatores de transcrição *ScbHLH2* e *ScC2H2* (subgrupo C). A expressão inversa dos fatores de transcrição em relação à expressão de *ScRAV1* seria esperada pela sua possível regulação dos níveis transcricionais dos genes de fatores de transcrição, baseando-se na predição de sítios putativos de RAV em genes homólogos de *S. bicolor* (Tabela 10), e na atividade repressiva de *ScRAV1* e seu ortólogo *SbRAV1* em *S. bicolor* (Figura 9), conforme predito na rede de regulação gênica (Figura 18).

V. DISCUSSÃO

O fator de transcrição *ScRAV1* de cana-de-açúcar pertence à família de fatores de transcrição RAV cujas proteínas estão envolvidas no processo de sinalização de etileno nas plantas (Alonso *et al.*, 2003). De acordo com a árvore filogenética da Figura 8, *ScRAV1* é ortólogo de *ZmRAV1L* (GRMZM2G169654) em *Z. mays*. Foi observado que este último é superexpresso especificamente nas células corticais de *Z. mays* durante a formação do aerênquima (Rajhi *et al.*, 2011) (Figura 8). Utilizando como base a sequência apontada por estes autores, *ScRAV1* foi recentemente caracterizado a partir de raízes de cana-de-açúcar por Tavares *et al.* (2019). Seus 8 alelos foram caracterizados e sua função foi demonstrada, através de expressão transiente em *N. benthamiana*, como um repressor de uma endopoligalacturonase associada à formação do aerênquima em cana-de-açúcar (Tavares *et al.*, 2019).

Após uma análise filogenética e a caracterização comparativa dos domínios de *ScRAV1* a outros RAV de plantas, no presente trabalho o fator de transcrição *ScRAV1* foi estudado com o objetivo de descobrir outras possíveis correlações com a expressão de genes relacionados ao metabolismo de parede celular e de outros fatores de transcrição em raízes de cana-de-açúcar em desenvolvimento. Redes de co-expressão foram montadas e identificaram padrões de conexão entre *ScRAV1* aos processos de metabolismo de carboidratos e biossíntese de polímeros de parede celular (lignina e polissacarídeos) no ápice cortical, bem como à degradação de algumas hemiceluloses e pectinas na região de formação do aerênquima. Além da correlação com a expressão de genes de parede celular, fatores de transcrição das famílias ARF e NAC de cana-de-açúcar foram identificados como co-expressos no processo de desenvolvimento das raízes de cana-de-açúcar.

O foco principal deste trabalho foi expandir o contexto de compreensão do processo de formação do aerênquima nas raízes de cana-de-açúcar, processo que vem sendo estudado por vários ângulos por nosso grupo nos últimos anos (Grandis, 2015; Leite *et al.*, 2017; Piovezani, 2017; Tavares *et al.*, 2018; Tavares *et al.*, 2019). A busca de informações sobre o contexto mais amplo do controle da formação do aerênquima é de grande importância para compreender o processo de desenvolvimento da raiz como um todo. Em cana-de-açúcar, a formação do aerênquima ocorre em meio a outros dois eventos principais. Por um lado, as divisões celulares no ápice da raiz, características de atividade meristemática e por outro o início da formação do sistema vascular, com a deposição de lignina e celulose. Até o momento, nossas análises permitiram discernir os eventos em segmentos de 1 cm até o quinto centímetro da raiz.

Na discussão que se segue, o ponto focal será a formação do aerênquima, mas não se deixará de abordar um contexto um pouco mais amplo, quando for o caso, de informações que permitem conclusões parciais e até mesmo algumas especulações que podem ser úteis para compreender os processos concorrentes durante o desenvolvimento da raiz de cana-de-açúcar.

1. *ScRAV1* apresenta três sequências homólogas em cana-de-açúcar possivelmente advindas de duplicação gênica e vários ortólogos em diferentes espécies

Foram encontradas 4 sequências proteicas para o fator de transcrição RAV em cana-de-açúcar, *A. thaliana* e *O. sativa*, de acordo com a árvore filogenética do fator de transcrição RAV (Figura 8). Diferentemente, alguns autores apontaram que *A. thaliana* e *O. sativa* tem 6 e 5 genes para o fator de transcrição RAV (Riechmann *et al.*, 2000, Swaminathan *et al.*, 2008), respectivamente. Contudo, verificamos que as outras duas sequências (AT1G50680 e

AT1G51120) em *A. thaliana* e uma sequência (LOC_Os07G17230) em *O. sativa* não estão presentes na árvore filogenética do fator de transcrição de RAV. Isto é devido as sequências de *A. thaliana* possuir somente o domínio B3, e apesar de estar dentro da subfamília RAV (Swaminathan *et al.*, 2008), a presença deste domínio se caracteriza da família de fatores de transcrição NGHATA (Lee *et al.*, 2015). Já a sequência de *O. sativa* contém os domínios NB-ARC e B3 (dado não mostrado), e foi descrita como da família de fatores de transcrição WRKY, conforme o Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html/>; Goodstein *et al.*, 2012).

O clado I da árvore filogenética mostrou que *ScRAV1* e *ScRAV2* podem ser possíveis isoformas ou gerados por duplicação gênica (Figura 8), pois também se observou duas sequências ortólogas em *M. sinensis* para estas sequências de cana-de-açúcar. Os gêneros *Saccharum* e *Miscanthus* vieram de duplicações recentes no genoma (Daniels *et al.*, 1975). No caso da cana-de-açúcar, sabe-se que é uma espécie poliploide derivado de um genoma inteiro duplicado (Garsmeur *et al.*, 2011). De acordo com Paterson *et al.* (2004), a duplicação e a diploidização moldaram os genomas de todas as culturas de Poaceae, incluindo os principais cereais e as culturas para a biomassa como a cana-de-açúcar. Isto pode sugerir que ambos os transcritos *ScRAV1* e *ScRAV2* vieram de uma duplicação gênica. Contudo, a primeira hipótese que estes transcritos podem ser isoformas não deve ser descartada, pois precisa obter a sequência proteica completa de *ScRAV2*. Além disso, o clado I ainda mostrou que *ScRAV1* é ortólogo de várias sequências proteicas em gramíneas (Figura 8), entre elas *SbRAV1* de sorgo, e *ZmRAV1L* (GRMZM2G169654), de milho. Este último gene é superexpresso somente nas células corticais de raízes de *Z. mays* durante a formação do aerênquima (Rajhi *et al.*, 2011). Os genes ortólogos são genes homólogos de diferentes espécies derivadas de um evento de especiação no último ancestral comum (Van de Velde *et al.*, 2014; Glover *et al.*, 2016). É

amplamente assumido que genes ortólogos podem apresentar a mesma função biológica em diferentes espécies (Fang *et al.*, 2010). Isto aponta para o fato de que *ScRAV1* e *SbRAV1* estão provavelmente envolvidos na formação do aerênquima em cana-de-açúcar e em *S. bicolor*, respectivamente.

Em contraposição, a bifurcação dos clados I e II, a presença de sequências ortólogas em várias gramíneas e os altos valores de *ultrabostrap* na árvore filogenética dão suporte à ideia que *ScRAV1* é parálogo, ou seja, aqueles originados por um evento de duplicação gênica (Glover *et al.*, 2016), de *ScRAV3* e *ScRAV4* (Figura 8). Visto que a cana-de-açúcar é um organismo poliploide (Garsmeur *et al.*, 2018), este evento é importante para a origem de novidades evolutivas através de processos como subfuncionalização – onde os genes duplicados podem compartilhar os atributos do gene ancestral, como função ou padrão de expressão (McGrath & Lynch, 2012). Porém, ambos os parálogos podem acumular mutações que alteram a função em relação ao gene ancestral, cada um restando apenas uma parte da função, ou seja, produzindo subfunções do gene ancestral (Supratim, 2014). Assim, cada parálogo pode perder a sua habilidade para substituir a função do gene ancestral, mas os dois parálogos juntos ainda podem ser capazes de complementar um ao outro para produzir a função ancestral (Supratim, 2014). Outro evento que colabora para novidades evolutivas é a neofuncionalização que significa que os genes duplicados podem ganhar uma nova função no genoma da espécie (McGrath & Lynch, 2012). Isto sugere que os fatores de transcrição *ScRAV3* e *ScRAV4* podem ter sido alvos dos processos de subfuncionalização e neofuncionalização no genoma de cana-de-açúcar.

Já o clado III e IV apresentou vários ortólogos de *ScRAV1* em eudicotiledôneas, ressaltando 4 possíveis ortólogos encontrados em *A. thaliana* (Figura 8). Mühlenbock *et al.* (2007) mostrou que plantas adultas são negativamente reguladas pelo gene envolvido na

defesa vegetal LSD1 (*LESION SIMULATING DISEASE1*) na formação do aerênquima lisígeno nos hipocótilos em *A. thaliana*. Já foi reportado que o hormônio etileno e o fator de transcrição RAV1 estão envolvidos na resposta à defesa a patógenos (Sohn *et al.*, 2006; van Loon *et al.*, 2006; Adie *et al.*, 2007). Assim, é possível concluir que os genes ortólogos para *ScRAV1* em *A. thaliana* possam estar envolvidos na formação do aerênquima, mas essa informação ainda precisa ser melhor investigada nesta espécie.

2. *ScRAV1* apresenta os domínios AP2 e B3 e o motivo BRD (*B3 repressor domain*) repressor conservados em diferentes espécies de plantas

No que se refere à caracterização dos domínios AP2, B3 e do motivo BRD (*B3 repressor domain*) de *ScRAV1* com os seus possíveis ortólogos, observou-se que eles são bem conservados nas diferentes espécies mencionadas neste trabalho (Figura 9). Vários autores já relataram que os domínios AP2 e B3 são característicos para os fatores de transcrição RAV em *A. thaliana* e *O. sativa* (Kagaya *et al.*, 1999; Riechmann *et al.*, 2000; Sakuma *et al.*, 2002; Nakano *et al.*, 2006; Swaminathan *et al.*, 2008), e que o motivo BRD é conservado nestas espécies (Ikeda & Ohme-Takagi, 2009). A conservação dos domínios AP2, B3 e do motivo BRD presente em *ScRAV1* pode indicar que eles são essenciais para a função desta proteína na transcrição dos seus genes alvos em cana-de-açúcar.

Para a família de fatores de transcrição RAV constatou-se que os quatro genes – RAV1, RAV2 (TEM2), TEM1 e AT3G25730 em *A. thaliana* apresentam o motivo repressor BRD (RLFGV) na região C terminal, e através do ensaio de expressão transiente foi sugerido que os genes relacionados à família de fatores de transcrição RAV podem ser repressores (Ikeda & Ohme-Takagi, 2009). De acordo com os autores, a atividade repressora destes fatores de transcrição é devida à presença de aminoácidos hidrofóbicos como metionina (M), leucina (L), valina (V),

fenilalanina (F) ou isoleucina (I) que ficam antes da sequência consenso do motivo BRD. A hidrofobicidade nesta posição pode ser necessária para a atividade deste motivo repressor (Ikeda & Ohme-Takagi, 2009).

Neste trabalho, observamos a presença de uma unidade de valina (V) antes da sequência consenso do motivo BRD - RLFGV (Figura 9), o que sugere que *ScRAV1* possui atividade repressora na transcrição dos seus genes alvos. De fato, Tavares *et al.* (2019) demonstraram por expressão transiente a atividade repressora de *ScRAV1*, o qual se liga ao promotor do gene de *ScEPG1*, uma endopoligalacturonase de cana-de-açúcar. A presença dos aminoácidos metionina (M), leucina (L) ou valina (V) para os ortólogos de *ScRAV1* em gramíneas (Figura 9), também pode sugerir uma possível atividade repressora destes fatores de transcrição nestas espécies.

3. Redes de co-expressão de *ScRAV1* como método de busca de transcritos de parede celular e fatores de transcrição que possivelmente agem durante a formação do aerênquima nos diferentes segmentos (S1 ao S4) da raiz em cana-de-açúcar

Baseando-se na hipótese que *ScRAV1* pode estar envolvido na formação do aerênquima em cana-de-açúcar (item 1, parágrafo 2), foram construídas redes de co-expressão de *ScRAV1* com os transcritos de parede celular (Figura 10A a D) e com os fatores de transcrição (Figura 15A a D) em cada segmento (S1-S4) da raiz, os quais representam os diferentes estágios de desenvolvimento do aerênquima em cana-de-açúcar. Recentemente, as redes de co-expressão tornaram-se uma abordagem atraente para conjuntos de dados *high-throughput* para compreender ao nível sistêmico o *status* dos componentes das células e determinar se essas moléculas interagem umas com as outras em processos biológicos

chaves que ocorrem nos organismos (Barabási & Oltvai, 2004; Usadel *et al.*, 2009; Costa *et al.*, 2015; Silva *et al.*, 2016).

Os genes altamente co-expressos teriam maior probabilidade de estarem envolvidos nos mesmos processos biológicos ou em processos biológicos do gene usado como isca para a formação das redes de co-expressão (Usadel *et al.*, 2009). Pelo princípio de “*guilt by association*” (culpado por associação) é afirmado que genes que apresentam a mesma função ou a mesma via regulatória apresentarão o mesmo perfil de expressão, portanto, formarão agrupamentos na rede (Wolfe *et al.*, 2005). Assim, dentro do mesmo agrupamento - genes com uma função conhecida são usados para predizer a função de genes co-expressos com a função desconhecida (Rhee & Mutwil, 2014). As duas principais aplicações das redes de co-expressão são encontrar novos genes e sugerir processos biológicos o qual um determinado gene está envolvido (Serin *et al.*, 2016).

Fica cada vez mais claro que uma determinada função biológica pode ser raramente atribuída a uma única molécula, e sim na interação complexa de vários constituintes presentes nas células. Portanto, o desafio para a biologia no século XXI é entender a estrutura e a dinâmica das redes complexas de interações intercelulares que contribuem para a estrutura e a função de uma célula viva (Barabási & Oltvai, 2004).

3.1. As redes de co-expressão de *ScRAV1* com os transcritos de parede celular e fatores de transcrição sugerem uma relação com as diferentes partes do desenvolvimento da raiz em cana-de-açúcar

As redes de co-expressão de *ScRAV1* com os transcritos de parede celular (Figura 10A a D, Tabela 1) e com os fatores de transcrição (Figura 15A a D, Tabela 6) mostraram diferentes porcentagens para os tipos de correlação e de correlações totais nos diferentes segmentos da raiz de cana-de-açúcar. Os tipos de correlação em cada segmento da raiz podem ser explicados pelo coeficiente de correlação de Pearson levar em consideração no cálculo de co-expressão os perfis de expressão dos dois genes de interesse nas amostras (Ruprecht & Persson, 2012), neste caso a expressão de *ScRAV1* em relação aos transcritos de parede celular e fatores de transcrição no RNA-Seq em cada segmento da raiz de cana-de-açúcar. Já as diferentes porcentagens de correlações totais em cada segmento podem estar relacionadas a diferentes partes da raiz, sendo o S1 – parte meristema apical/ parte zona de alongamento, S2/S3 – zona de alongamento e S4 – zona de maturação da raiz, de acordo com os dados produzidos por Leite *et al.* (2017).

3.2. *ScRAV1* co-expressa com os transcritos de parede celular associados à biossíntese quanto à hidrólise da parede celular na raiz de cana-de-açúcar

Em relação à categorização por biossíntese, hidrólise e outros para os transcritos de parede celular co-expressos com *ScRAV1*, observou-se que a biossíntese e a hidrólise dos polissacarídeos na parede celular são constantes na raiz durante o desenvolvimento do aerênquima em cana-de-açúcar (Figura 11). Alguns autores têm mostrado que os genes de biossíntese e hidrólise de polissacarídeos se co-expressam em *A. thaliana* (Mutwil *et al.*, 2009; Ruprecht & Persson, 2012). Em outro estudo, a co-expressão dos genes de síntese de celulose

(CesAs) com genes de biossíntese e hidrólise de outros polissacarídeos, sugeriu que os genes co-expressos são conservados em monocotiledôneas (*H. vulgare*, *O. sativa*, *T. aestivum*) e eudicotiledôneas (*A. thaliana*, *P. trichocarpa*, *G. max*, *M. truncatula*), assim como em parede celular primária e secundária destas plantas (Ruprecht *et al.*, 2011). Assim, pode-se sugerir que a síntese e a hidrólise de polissacarídeos são importantes para as plantas no remodelamento de suas paredes celulares durante o desenvolvimento de seus tecidos como, por exemplo, a formação do aerênquima.

Neste sentido, verificamos na categorização dos transcritos de parede celular co-expressos com *ScRAV1* no S1 que a biossíntese é maior em relação à hidrólise de polissacarídeos neste segmento. Neste segmento, prevaleceu principalmente a biossíntese de hemicelulose e hemicelulose/pectina/outros, enquanto na degradação de polissacarídeos predominou a hidrólise de hemicelulose e pectina (Figura 11A). Conforme Leite *et al.* (2017), o S1 apresenta o meristema apical da raiz e o início da zona de alongamento, a primeira região está em constante divisão celular e na biossíntese de parede celular (Sebastian & Lee, 2013), enquanto a segunda região está em expansão celular (Bizet *et al.*, 2015). Isto faz sentido uma vez que se pode verificar na rede do S1 que *ScRAV1* se co-expressa com vários transcritos de parede celular de ambas as categorias (Figura 10A).

Entretanto, nas redes de co-expressão de *ScRAV1* nos S2, S3 e S4 observou-se uma maior porcentagem de transcritos de parede celular associados à hidrólise de polissacarídeos, sendo esta porcentagem maior no S2. Nas redes de co-expressão do S2 e S4 houve uma predominância de transcritos de parede celular relacionados à hidrólise de celulose e hemicelulose, e somente transcritos de parede celular associados à hidrólise de pectina na rede de co-expressão do S4 (Figura 11B e D). Já na rede de co-expressão do S3 houve prevalência de transcritos associados à hidrólise hemicelulose e pectina (Figura 11C). Leite *et*

al. (2017) não obtiveram evidências que sugerissem a degradação de celulose durante a formação do aerênquima, mas observaram a degradação do glucano de ligação mista (GLM) em direção ao S5 durante o desenvolvimento do aerênquima na raiz de cana-de-açúcar. Porém, os autores não descartam que haja degradação das microfibrilas de celulose no córtex da raiz durante o desenvolvimento deste tecido nesta espécie. Em contrapartida, outros estudos já verificaram a presença de celulasas e hemicelulasas durante o desenvolvimento do aerênquima em *Z. mays* (He *et al.*, 1994; Bragina *et al.*, 2001; Rajhi *et al.*, 2011).

No que diz respeito à hidrólise de pectina, verificou-se uma modificação da pectina chamada homogalacturonano, presente na lamela média com início no S3 e um aumento até o S5 durante a formação do aerênquima na raiz de cana-de-açúcar (Leite *et al.*, 2017), sendo condizente com os transcritos relacionados à degradação de pectina nas redes de co-expressão de *ScRAV1* no S3 e S4 (Figura 11C e D). Corroborando com isso, outro trabalho do grupo mostrou um aumento da expressão da endopoligalacturonase *ScEPG1* em direção ao S5 na raiz de cana-de-açúcar durante o desenvolvimento deste tecido (Tavares *et al.*, 2019), esta enzima age na hidrólise do homogalacturonano (Atkinson *et al.*, 2002; González-Carranza *et al.*, 2007). Em *Z. mays*, também foram observadas modificações nas pectinas nas células corticais das raízes que apresentavam aerênquima através do uso de anticorpos monoclonais JIM5 e JIM7 para a identificação de pectinas desesterificadas e esterificadas, respectivamente (Gunawardena *et al.*, 2001b). Assim, as evidências indicam que a degradação ou modificações na pectina nas células do córtex seja um passo importante para a formação do aerênquima.

Como visto anteriormente, estes três últimos segmentos apresentaram uma menor porcentagem de transcritos de parede celular associados à biossíntese em relação à hidrólise de polissacarídeos nas redes de co-expressão de *ScRAV1* (Figura 11B a D). Nas redes dos S2, S3 e S4 verificaram-se transcritos associados à biossíntese de hemicelulose (Figura 11B a D),

os S2 e S3 também apresentaram transcritos associados à biossíntese de pectina (Figura 11B a C), enquanto o S4 apresentou somente transcritos associados à biossíntese de hemicelulose (Figura 11D). É provável que a biossíntese de celulose, hemicelulose e pectina nos S2 e S3 possam estar relacionadas à eventos de alongamento da raiz (Figura 11B a C), onde ocorre a expansão e extensão celular, a deposição e o rearranjo de polissacarídeos na parede celular (Tsang *et al.*, 2011), assim como a biossíntese de celulose em orientação específica nas células (Crowell *et al.*, 2010; Li *et al.*, 2014b). Quanto aos transcritos referentes à biossíntese de hemicelulose no S4 (Figura 11D), estes podem estar relacionados à formação do compósito de hemiceluloses como o xiloglucano e arabinoxilano neste segmento durante o desenvolvimento do aerênquima na raiz da cana-de-açúcar (Leite *et al.*, 2017).

No entanto, a co-expressão de *ScRAV1* foi mais significativa com os transcritos de biossíntese de lignina em todos os segmentos da raiz (Figura 11A a D), principalmente nos três primeiros segmentos da raiz. Estes transcritos podem estar envolvidos no desenvolvimento do sistema vascular (Vanholme *et al.*, 2010; Barros *et al.*, 2015; Yoon *et al.*, 2015; Zhong & Ye, 2015) ou no metabolismo secundário (Vanholme *et al.*, 2012; Labeeuw *et al.*, 2015), já que a proporção de lignina nas paredes celulares nesta porção da raiz de cana-de-açúcar foi constatada ser muito pequena (Leite *et al.*, 2017).

Também se verificou uma alta co-expressão de *ScRAV1* com os transcritos do metabolismo de carboidratos nos S1 e S4 (Figura 11A e D). Pouco se sabe sobre a participação de genes do metabolismo de carboidratos durante a formação do aerênquima. É provável que estes transcritos estejam agindo na produção de açúcares e açúcares-nucleotídeos envolvidos na biossíntese da parede celular no S1, tecido que está em constante divisão e expansão celular devido a atividade do meristema apical da raiz (Sebastian & Lee, 2013; Bizet *et al.*, 2015). Em contrapartida, no S4 é possível que estes transcritos estejam gerando açúcares

(Verbančič *et al.*, 2018) para a síntese de polissacarídeos que formarão o sistema vascular da raiz (Leite *et al.*, 2017).

3.3. *ScRAV1* co-expressa com várias famílias de fatores de transcrição com diferentes papéis nas plantas

No que se refere as redes de co-expressão de *ScRAV1* com os fatores de transcrição, observou-se que *ScRAV1* co-expressa com diferentes quantidades de famílias de fatores de transcrição em cada uma das redes de co-expressão de *ScRAV1* na raiz de cana-de-açúcar (Figura 16). Segundo Riechmann *et al.* (2000), há cerca de 30 diferentes famílias de fatores de transcrição que foram identificados em *A. thaliana*. Estes dados são similares aos encontrados neste trabalho nas redes de *ScRAV1* nos S1 e S4, e representa metade do que foi encontrado nos S2 e S3.

Alguns bancos de dados apresentam uma variação na quantidade de família de fatores de transcrição devido a definição dada a fator de transcrição em cada um deles, por exemplo, RARTF (<http://rarge.gsc.riken.jp/rartf/>; Iida *et al.*, 2005), AGRIS (<http://arabidopsis.med.ohio-state.edu/AtTFDB/>; Davuluri *et al.*, 2003), PlantTFDB (<http://plantfdb.cbi.pku.edu.cn/>; Jin *et al.*, 2017) e PlnTFDB (<http://plntfdb.bio.uni-potsdam.de/>; Riaño-Pachón *et al.*, 2007) contêm 51, 50, 58 e 67 famílias de fatores de transcrição, respectivamente. Assim, a quantidade de famílias de transcrição possivelmente pode estar associada a diferentes estágios do desenvolvimento da raiz e do aerênquima em cana-de-açúcar (Leite *et al.*, 2017), como visto anteriormente no item 3.1.

As redes de co-expressão de *ScRAV1* com os fatores de transcrição mostraram que as famílias AP2/EREBP, bZIP, C2H2, C3H, HOMEODOMAIN, NAC, PHD e SET estão presentes nas quatro redes dos segmentos da raiz de cana-de-açúcar (Figura 16). As famílias de fatores de

transcrição que são específicas nos vegetais e que possuem domínio de ligação ao DNA somente em plantas, incluem as famílias AP2/EREBP e NAC (*NAM-ATAF-CUC*), ao contrário das famílias bZIP (*BASIC REGION/LEUCINE ZIPPER*), C2H2, C3H, HOMEODOMAIN, PHD e SET (Riechmann *et al.*, 2000; Marmorstein, 2003; Bienz, 2006). Estas famílias têm um importante papel no controle no crescimento e no desenvolvimento do vegetal em diversas condições (Hong, 2016). As famílias de fatores de transcrição como AP2/EREBP, MYB, SET e WRKY e AP2/EREBP, bZIP, HOMEODOMAIN, MYB e NAC já foram encontradas durante a formação do aerênquima em *O. sativa* (Yoo *et al.*, 2015) e em *Z. mays* (Rajhi *et al.*, 2011), respectivamente. No entanto, pouco se sabe da função destes fatores de transcrição na formação do aerênquima nas raízes de gramíneas.

Em relação àqueles fatores de transcrição que co-expressaram em maior proporção nas redes S1 e S2 com *ScRAV1* estão os fatores de transcrição do tipo bZIP (Figura 16). Estes fatores de transcrição regulam vários processos como na defesa contra patógenos, na sinalização da luz e estresse, na maturação das sementes e no desenvolvimento floral (Jakoby *et al.*, 2002), bem como em resposta à ABA (Choi *et al.*, 2000; Uno *et al.*, 2000). Estudos anteriores sugerem a participação de RAV1 na senescência foliar em resposta à ABA em *A. thaliana* (Zhao *et al.*, 2016), e na ativação de CaRAV1 na defesa a patógenos por ABA em *Capsicum annuum* (Sohn *et al.*, 2006). Os processos como a senescência foliar, a defesa a patógeno, e a formação do aerênquima levam à MCP nas plantas (Lam *et al.*, 2004; Gadjev *et al.*, 2008; Woo *et al.*, 2013). Assim, não se deve descartar a hipótese que *ScRAV1* aja na modulação de ABA através dos fatores de transcrição bZIPs durante a formação do aerênquima em cana-de-açúcar nestes dois primeiros segmentos. Contudo, não há estudos mostrando a relação ABA e a formação do aerênquima em gramíneas.

A respeito das famílias de fatores de transcrição bHLH (*BASIC HELIX-LOOP-HELIX*) e GRAS observou-se uma alta porcentagem de transcritos co-expressos com *ScRAV1* nos S2, enquanto para a família de fatores de transcrição MYB co-expressa com *ScRAV1* nos S1 e S3 da raiz (Figura 16). Os fatores de transcrição bHLH estão envolvidos em diversos processos como no crescimento da raiz, na diferenciação e na proliferação celular (Tsukagoshi *et al.*, 2010; Feller *et al.*, 2011), enquanto a família GRAS está associada à sinalização de giberelina e à padronização da raiz (Hofmann, 2016). Já a família MYB está envolvida no controle de diversos processos nos vegetais, por exemplo, na arquitetura da raiz (Dai *et al.*, 2012), na identidade, no destino celular (Dubos *et al.*, 2010), e na biossíntese de fenilpropanóides (Jin *et al.*, 2000). Isto pode sugerir que *ScRAV1* esteja envolvido na modulação de fatores de transcrição GRAS, MYB e bHLH nos S1 e S2 para o crescimento, o desenvolvimento e a diferenciação da raiz, bem como na determinação das células que irão formar o aerênquima, enquanto no S3 *ScRAV1* possivelmente coordena os fatores de transcrição MYB que podem estar envolvidos na formação da parede secundária na raiz de cana-de-açúcar.

Em contrapartida, verificou-se uma alta porcentagem de fatores de transcrição da família C2H2 que se co-expressam com *ScRAV1* nos S2, S3 e S4 (Figura 16). Estes fatores de transcrição estão envolvidos principalmente no crescimento e no desenvolvimento do vegetal (Huang *et al.*, 2004). Já os fatores de transcrição PHD e SET apresentaram uma alta porcentagem de co-expressão com *ScRAV1* no S4 da raiz de cana-de-açúcar (Figura 16). Essas famílias de fatores de transcrição participam na regulação da cromatina (Marmorstein, 2003; Bienz, 2006). Isto indica que provavelmente *ScRAV1* está relacionado tanto no programa epigenético celular quanto no desenvolvimento da raiz durante a formação do aerênquima em cana-de-açúcar através das famílias de fatores de transcrição PHD, SET e C2H2, respectivamente.

Também se verificou uma porcentagem relevante de fatores de transcrição DBP co-expressos com *ScRAV1* nos S1 e S3, enquanto os fatores de transcrição WRKY têm uma alta co-expressão com *ScRAV1* no S3 (Figura 16). Para os fatores de transcrição DBP, há apenas um relato da proteína designada como DBP1 em *N. tabacum*, a qual apresenta tanto ligação específica de sequência ao DNA quanto a atividade fosfatase. A desfosforilação por proteínas fosfatase é um mecanismo importante na regulação da atividade de fatores de transcrição devido facilitar a expressão gênica para mudanças apropriadas no comportamento celular. Este processo pode regular vários aspectos da função dos fatores de transcrição como a localização celular, a estabilidade da proteína, a interação proteína-proteína e a ligação ao DNA (Whitmarsh & Davis, 2000).

De acordo com alguns autores, os fatores de transcrição DBP1 e WRKY regulam principalmente a resposta à patógeno (Carrasco *et al.*, 2003; Rushton *et al.*, 2010). Isto é, mecanismo da defesa da planta contra o ataque ao patógeno, que leva a uma resposta chamada reação de hipersensibilidade (HR; *Hypersensitive Reaction*), a qual é um tipo de programa de morte celular para direcionar a invasão do patógeno e prevenir que a infecção se espalhe pela planta (Camagna & Takemoto, 2018). Este processo se assemelha ao fenômeno de morte celular programada (MCP) do aerênquima de cana-de-açúcar (Leite, 2013), sugerindo que possivelmente os fatores de transcrição DBP co-expressos com *ScRAV1* nestes segmentos podem interagir com este fator de transcrição para modular a desfosforilação proteica de outros fatores de transcrição no S1, enquanto a co-expressão com DBP e WRKY no S3 pode levar a MCP e o espalhamento progressivo da formação do aerênquima na raiz de cana-de-açúcar (Leite *et al.*, 2017).

Evidências na literatura e as predições *in silico* deste trabalho indicam a participação de fatores de transcrição das famílias ARFs e NACs na formação do aerênquima em cana-de-

açúcar. Os fatores de transcrição ARFs estão co-expressos com *ScRAV1* nas redes dos S2 ao S4 (Figura 16). Esta família de fatores de transcrição confere especificidade na resposta à auxina (Li *et al.*, 2016). A importância da auxina já foi reportada durante a formação do aerênquima em *Z. mays* (Justin & Armstrong, 1991) e em cana-de-açúcar (Tavares *et al.*, 2018). Assim, *ScRAV1* possivelmente modula a atuação de fatores de transcrição como ARFs durante o desenvolvimento do aerênquima em cana-de-açúcar.

No que diz respeito à família de fatores de transcrição NAC, verificou-se que os transcritos são co-expressos com *ScRAV1* significativamente em todas as redes de co-expressão, com uma alta porcentagem no S4 (Figura 16). Estes fatores de transcrição estão envolvidos em vários processos do desenvolvimento nas plantas (Puranik *et al.*, 2012), por exemplo, na formação da parede secundária (Nakano *et al.*, 2015) e na senescência foliar (Podzimska-Sroka *et al.*, 2015), processos que incluem MCP, assim como na formação do aerênquima. Já foi reportada a presença de vários fatores de transcrição NACs em *A. thaliana*, *T. aestivum* e *O. sativa* mostrando a importância deles na regulação na senescência foliar (Guo *et al.*, 2004; Buchanan-Wollaston *et al.*, 2005; Uauy *et al.*, 2006; Liang *et al.*, 2014). Estes fatores de transcrição induzem a morte celular por estresse no retículo endoplasmático ou estresse osmótico parcialmente através da enzima VPE (*Vacuolar Processing Enzyme*) ou pela atividade da proteína caspase-like (Faria *et al.*, 2011; Mendes *et al.*, 2013; Yang *et al.*, 2014). Em relação ao aerênquima, dois autores mostraram que a superexpressão de OsNAC5 e OsNAC9 aumentou o eustelo e o aerênquima em *O. sativa* cv *Nipponbare* (Redillas *et al.*, 2012; Jeong *et al.*, 2013). Isto pode indicar que os fatores de transcrição NAC co-expressos com *ScRAV1* podem estar associados à formação do aerênquima ou na parede secundária, e provavelmente podem ser candidatos em regular a MCP na raiz de cana-de-açúcar.

Nossos levantamentos bibliográficos indicam que não há muitos trabalhos sobre fatores de transcrição durante o desenvolvimento do aerênquima em gramíneas. Porém, os dados existentes apontam que *ScRAV1* possivelmente participa em diversos processos biológicos na raiz, entre eles a formação do aerênquima que provavelmente é coordenada por vários fatores de transcrição que determinam o desenvolvimento deste tecido na raiz de cana-de-açúcar.

3.4. Identificação dos nós essenciais nas redes de co-expressão de *ScRAV1* através da centralidade de grau

A seleção dos transcritos de parede celular (Tabela 2) e os fatores de transcrição (Tabela 7) pela centralidade grau mostrou grupos de transcritos que tiveram maior representatividade nas redes de co-expressão de *ScRAV1* nos diferentes segmentos da raiz de cana-de-açúcar, conforme os itens 3.2 e 4.2. A centralidade de grau é um dos índices topológicos da estrutura de redes mais usados (Jalili *et al.*, 2016), e refere-se ao número de conexões diretas que um nó possui dentro da rede (Albert & Barabási, 2002). Estes nós altamente conectados ou também conhecidos como “*hubs*” são provavelmente nós essenciais dentro das redes biológicas (Kranthi *et al.*, 2016).

É assumido que uma vez que um nó possui a maior centralidade de grau – este está conectado a mais vértices, e provavelmente tem mais opções para se comunicar com o resto da rede, pois vários dos vértices ou dos caminhos possíveis passam por este nó na rede (Ghasemi *et al.*, 2014). Assim, o índice de centralidade de grau da rede pode ser utilizado para entender a função biológica dos componentes dentro da rede, ajudando a desvendar mecanismos centrais em processos biológicos nos organismos (Jalili *et al.*, 2016). Isto contrasta com o fato de que a identificação de proteínas essenciais por métodos

experimentais é demorada, cara e nem sempre viável (Zhong *et al.*, 2013; Li *et al.*, 2014a; Ashtiani *et al.*, 2018). Portanto, pode-se sugerir que algumas das maiores centralidades de grau encontradas nas redes de *ScRAV1* para transcritos de parede celular e para fatores de transcrição podem ser importantes para a formação do aerênquima assim como para o desenvolvimento geral da raiz.

3.5. Os homólogos em *S. bicolor* e em *A. thaliana* para os transcritos de parede celular e para os fatores de transcrição de cana-de-açúcar

Verificamos que a maioria das sequências de cana-de-açúcar como as maiores centralidades de grau para os transcritos de parede celular e para os fatores de transcrição (itens 3.3 e 4.3) apresentaram uma alta identidade com as sequências homólogas de *S. bicolor* (Tabela 3 e 8). Provavelmente, isto deve-se ao fato de *S. bicolor* ser uma espécie filogeneticamente próxima de cana-de-açúcar (Dufour *et al.*, 1997; Guimaraes *et al.*, 1997; Ming *et al.*, 1998; Le Cunff *et al.*, 2008; Manners & Casu, 2011; Kellogg, 2013) e devido as ambas as espécies pertencerem a família Poaceae (gramíneas) (Paterson *et al.*, 2009; De Siqueira Ferreira *et al.*, 2013). Acredita-se que cerca de 82% dos genes preditos em cana-de-açúcar podem ser encontrados no genoma de *S. bicolor* (Garsmeur *et al.*, 2018), e que as sequências de genes de ambas as espécies apresentam um alto grau de conservação entre si (Jannoo *et al.*, 2007; Garsmeur *et al.*, 2011). Por este motivo, vários autores têm postulado que *S. bicolor* pode ser um modelo para estudar os genes de cana-de-açúcar (Dufour *et al.*, 1997; Jannoo *et al.*, 2007; Paterson *et al.*, 2009).

Em relação à baixa identidade das sequências de *S. bicolor* com as sequências de *A. thaliana* para os genes de parede celular (Tabela 4) e para os fatores de transcrição (Tabela 9), acredita-se que aquela se deve ao fato da divergência evolutiva entre as duas espécies

(Paterson *et al.*, 2004), refletindo na conservação das sequências ao longo do tempo evolutivo entre dicotiledôneas e monocotiledôneas.

3.6. Anotação funcional dos genes de parede celular e fatores de transcrição através dos termos do Gene Ontology (GO) de *A. thaliana*

A comparação de sequências proteicas dos genes de parede celular (Tabela 4) e dos fatores de transcrição (Tabela 9) de *S. bicolor* com *A. thaliana*, permitiu buscar a anotação funcional através dos termos GOs para os genes de parede celular (Figura 12) e para os fatores de transcrição em *A. thaliana* (Figura 17). Sabe-se que cerca de 94% dos genes de *S. bicolor* têm ortólogos em *A. thaliana* (Paterson *et al.*, 2009). A escassez e a disparidade na anotação funcional em diferentes espécies, particularmente para outras plantas que não sejam *A. thaliana*, é um gargalo para a compreensão de como os processos biológicos são organizados, como funcionam e como evoluíram nas plantas (Rhee & Mutwil, 2014).

Apenas cerca de 40% em *A. thaliana* e 1% em *O. sativa* têm genes que codificam para proteínas com funções anotadas com base em evidências experimentais (Rhee & Mutwil, 2014). É difícil determinar o número de genes experimentalmente caracterizados para qualquer espécie a não ser *A. thaliana* ou *O. sativa* (Rhee & Mutwil, 2014). Assim, a comparação de análise de sequências é uma ótima ferramenta para estudar família de genes homólogos, isto é, genes que compartilham um ancestral comum. Pois, a maioria das anotações de novos genomas sequenciados são baseados na similaridade com outras sequências - as quais as funções são conhecidas (Tirosh *et al.*, 2007), ajudando na predição da anotação funcional de novos genomas (Movahedi *et al.*, 2012) ou transcriptomas. A anotação funcional de genes é essencial para entender como os processos biológicos funcionam (Hansen *et al.*, 2014).

Assim, no que diz respeito à anotação funcional dos genes de parede celular com base em *A. thaliana*, verificou-se que o termo *integral component of membrane* (GO:0016021) predominou entre os termos GOs para componente celular (Figura 12A). Este termo se refere a proteínas que se inserem na membrana plasmática (Taiz *et al.*, 2017). Para processo biológico prevaleceu o termo *cell wall organization* (GO:0071555) (Figura 12B). De acordo com a plataforma Amigo (<http://amigo.geneontology.org/>, Carbon *et al.*, 2009), este termo refere-se à construção ou desconstrução da parede celular através da biossíntese ou hidrólise dos polissacarídeos presentes na parede das plantas, respectivamente. Isto coincide com o papel de várias proteínas anotadas em *A. thaliana* (Apêndice D).

Ainda foram observados vários termos associados à parede secundária e ao metabolismo secundário para processos biológicos (Figura 12B) e para função molecular (Figura 12C, Apêndice D) para os genes de parede celular. A síntese de lignina se inicia com a via geral dos fenilpropanóides gerando vários precursores para diferentes compostos como flavonoides, cumarinas, quinonas e monolignóis, bem como os precursores de lignina para formar a parede secundária nas plantas (Vanholme *et al.*, 2012; Labeeuw *et al.*, 2015).

Em contrapartida, para os termos que predominaram para componente celular, processo biológico e função molecular para os fatores de transcrição anotados em *A. thaliana* (Figura 17A, B e C), foram os termos *nucleus* (GO:0005634), *regulation of transcription, DNA-templated* (GO:0006355) e *DNA-binding transcription factor activity* (GO:0003700), respectivamente. Isto coincide com a descrição para fatores de transcrição nas células, que têm como característica conter domínios de ligação ao DNA ao reconhecerem sequências específicas dentro das regiões promotoras dos seus genes alvos (Kummerfeld & Teichmann, 2006), regulando o processo de transcrição que ocorre dentro do núcleo das células (Gonzalez, 2016).

Também observamos vários termos GOs para processo biológico para fatores de transcrição (Figura 17B), coincidindo com os papéis biológicos já vistos na literatura para as proteínas RAV como, por exemplo, na defesa à patógeno, no estresse salino, na injúria (Sohn *et al.*, 2006; Li *et al.*, 2011), na senescência foliar (Zhao *et al.*, 2008; Woo *et al.*, 2010; Zhao *et al.*, 2016), na floração (Hu *et al.*, 2004; Castillejo & Pelaz, 2008; Zhao *et al.*, 2008) e na resposta à ABA (Zhao *et al.*, 2008; Feng *et al.*, 2014; Zhao *et al.*, 2016). Isto possivelmente pode indicar que *ScRAV1* e estes fatores de transcrição apresentam papéis em comum nas plantas e possivelmente regulam juntos estes diferentes processos biológicos. A formação do aerênquima na raiz de cana-de-açúcar via *ScRAV1* pode ser um deles.

3.7. Predição de sítios putativos de ligação de RAV ao DNA para os genes de parede celular e para os fatores de transcrição em *S. bicolor*

No que concerne a predição *in silico* de elementos *cis* regulatórios para RAV, esta foi feita em sequências homólogas de *S. bicolor* para os genes de parede celular e para os fatores de transcrição (Tabela 5 e 10), pois o genoma de cana-de-açúcar tem uma estrutura complexa e é um obstáculo no sequenciamento, na montagem e na anotação do genoma nesta espécie (Riaño-Pachón & Mattiello, 2017). Há várias sequências parciais do genoma (de Setta *et al.*, 2014; Grativol *et al.*, 2014; Okura *et al.*, 2016; Miller *et al.*, 2017) e também do transcriptoma de cana-de-açúcar (Vettore *et al.*, 2003; Mattiello *et al.*, 2015; Belesini *et al.*, 2017; Hoang *et al.*, 2017). Recentemente, o genoma de cana-de-açúcar foi sequenciado por dois grupos, um deles ainda está incompleto (Riaño-Pachón & Mattiello, 2017) e o outro ainda não foi liberado para a comunidade científica (Garsmeur *et al.*, 2018).

Apesar de *S. bicolor* ser uma espécie diploide com o genoma pequeno, completo e com poucas duplicações gênicas (Paterson *et al.*, 2009), esta espécie é filogeneticamente próxima

da cana-de-açúcar (Manners & Casu, 2011; Kellogg, 2013), e tem sido a melhor referência na genômica, no transcriptoma e em outras aplicações na biologia de sistema para a cana-de-açúcar (Manners & Casu, 2011). Ambas as espécies pertencem a família Poaceae (gramíneas) e representam um grande potencial para a produção de bioenergia (Paterson *et al.*, 2009; De Siqueira Ferreira *et al.*, 2013). Ademais, *S. bicolor* pode ser um importante modelo para pesquisas referentes à bioenergia para as gramíneas (Carpita *et al.*, 2008; Saballos, 2008; Paterson *et al.*, 2009), e tem sido amplamente reconhecido como genoma referência para análise comparativa (de Setta *et al.*, 2014) de outras culturas em que falta sequência genômica (Aitken *et al.*, 2014).

Além disso, *ScRAV1* é ortólogo de *SbRAV1* (SOBIC.003G265200.1), ambos apresentando o mesmo ortólogo para *ZmRAV1L* (GRMZM2G169654) que é expresso somente nas células do córtex durante formação do aerênquima em *Z. mays* (Rajhi *et al.*, 2011) (Figura 8). Contribuindo com esta informação, já foi reportada a formação do aerênquima nas raízes em *S. bicolor* (Promkhambut *et al.*, 2011). Estes dados sugerem que *SbRAV1* assim como *ScRAV1* também pode apresentar um possível papel nas células corticais durante a formação do aerênquima em *S. bicolor*. Assim, a busca de elementos *cis* nesta espécie também pode agregar conhecimento sobre a formação do aerênquima em *S. bicolor* que ainda é pouco estudado e permitir inferências sob o desenvolvimento do aerênquima em cana-de-açúcar, assim tornando *S. bicolor* um ótimo candidato para avaliar os sítios de ligação para RAV. Pois, genes de diferentes espécies que geralmente mostram alto grau de similaridade entre as sequências, permitem estudar processos análogos em organismos modelos (Chikina & Troyanskaya, 2011), como visto para *S. bicolor* e cana-de-açúcar no item 3.5 desta seção.

A busca de sítios de ligação de RAV ao DNA nas sequências homólogas de *S. bicolor* para genes de parede celular (Tabela 5) e para fatores de transcrição (Tabela 10) mostraram

uma maior quantidade de motivos AP2 (CAACA) em relação ao motivo B3 (CACCTG). Contudo, quando se diz respeito à porcentagem de ambos, o motivo AP2 (CAACA) e o motivo B3 (CACCTG) nas regiões promotoras das sequências homólogas em *S. bicolor*, observou-se que os genes de parede celular apresentaram uma maior quantidade para ambos os motivos para sítios de ligação de RAV ao DNA (Tabela 5), enquanto o oposto foi observado para os genes de fatores de transcrição nas sequências homólogas de *S. bicolor* (Tabela 10). Essas duas situações também já foram verificadas para os motivos AP2 (CAACA) e B3 (CACCTG) para sítios de ligação de RAV ao DNA em *Arabidopsis thaliana* (Kagaya *et al.*, 1999) e *Solanum lycopersicum* (Li *et al.*, 2011). Além disso, sabe-se que os domínios AP2 e B3 de RAV podem reconhecer as suas sequências-alvos de forma independente (Kagaya *et al.*, 1999), ainda que a ligação de RAV seja fraca para o reconhecimento de um domínio só ao DNA (Yamasaki *et al.*, 2004). Contudo, a afinidade e a eficiência pela sequência alvo aumenta quando os dois domínios se ligam juntos ao DNA (Kagaya *et al.*, 1999).

Para alguns genes de parede celular tem sido reportado sítios de ligação de RAV ao DNA na região promotora em diferentes espécies. Em cana-de-açúcar, observou-se apenas o motivo AP2 (CAACA) para os sítios de ligação de RAV ao DNA no promotor da glicosil hidrolase *ScEPG1*. Ao mesmo tempo, verificou-se que *ScRAV1* reprimiu a transcrição do gene da *ScEPG1* no ensaio de transativação em *N. benthamiana* (Tavares *et al.*, 2019). Nas espécies *T. aestivum* e *O. sativa* (OSJNBa0084K01.10), foi reportado a presença do motivo B3 (CACCTG) para sítios de ligação a RAV nas regiões promotoras do gene F3H, uma enzima importante na biossíntese de flavonoides (Himi *et al.*, 2011). De acordo Chen *et al.* (2015), RAV1 poderia coordenar o metabolismo de fenilpropanóides controlando a expressão de gene de biossíntese como F3H ou interagindo com a via de sinalização de biossíntese de jasmonato na raiz de *Isatis indigotica*. Diferentemente do que foi visto anteriormente, o gene homólogo de *S. bicolor*

para *ScF5H/F3H1* apresentou apenas o motivo AP2 (CAACA) de ligação de RAV ao DNA (Tabela 5), sugerindo um possível papel de RAV no metabolismo secundário nas plantas e uma possível conservação dos motivos para RAV em espécies evolutivamente distante. Além disso, verificou-se que os genes homólogos de *S. bicolor* dos transcritos *ScCHI2* e *ScCHI3* apresentou sítios de ligação para RAV ao DNA (Tabela 5), e são sequências homólogas da sequência (AT3G12500) em *A. thaliana* (Tabela 4). Em *S. lycopersicum*, observou-se sítios de ligação de RAV ao DNA para o motivo AP2 (CAACA) na região promotora de uma endoquitinase (Li *et al.*, 2011). Este gene é homólogo da sequência de *A. thaliana* citada anteriormente neste parágrafo. Isto sugere uma relação das quitinases e do fator de transcrição RAV em processos de resposta imune nas plantas. Em relação aos fatores de transcrição, não foi encontrado nenhum trabalho na literatura que reportasse elementos *cis* para RAV na região promotora de genes para fatores de transcrição.

Com base na Figura 9, sugeriu-se que *SbRAV1* e *ScRAV1* atuam como repressores da transcrição de seus genes alvos devido a presença do motivo BRD (*B3 repressor domain*) (item 2, parágrafo 3). É sabido que os fatores de transcrição repressores como estes regulam a transcrição dos seus genes alvos em diferentes formas, uma delas é através da interação com outras proteínas chamadas co-repressores (Gaston & Jayaraman, 2003), as quais são incapazes de se ligar ao DNA independentemente, sendo recrutados diretamente ou indiretamente por fatores de transcrição para reprimir a expressão de genes alvos (Long *et al.*, 2006; Causier *et al.*, 2012). Foi observado que os co-repressores como TPL/TPR (*TOPELESS/TPL-related*) podem interagir com as proteínas RAV através do motivo RLFGV ou MLFGV destas proteínas (Causier *et al.*, 2012). Este motivo também foi identificado para *ScRAV1* e *SbRAV1*, em seus ortólogos em gramíneas e em *A. thaliana* (Figura 9). Durante a formação do aerênquima de cana-de-açúcar, observou no RNA-Seq vários transcritos

anotados como TPL/TPR (Piovezani, 2017), enquanto na proteômica verificou-se a presença de 3 proteínas desta família de co-repressores (Grandis, 2015). As proteínas TPL/TPR mediam a repressão da transcrição por agir com histona deacetilases (HDACs) ou metiltransferases (HMTs) como SDG19 que pertence à família de transcrição SET, para induzir a compactação da cromatina, gerando um estado de repressão transcrricional (Kieffer *et al.*, 2006; Long *et al.*, 2006; Zhu *et al.*, 2010; Causier *et al.*, 2012).

Foram identificados genes relacionados a modificações das histonas - duas histonas acetiltransferase e um fator de transcrição Jumonji em *Z. mays* (Rajhi *et al.*, 2011), este último fator de transcrição atua na desmetilação de histonas (Mosammaparast & Shi, 2010), e um fator de transcrição da família SET na espécie *O. sativa* (Yoo *et al.*, 2015) durante a formação do aerênquima lisígeno desta gramínea em experimento de microarranjo. Várias HMTs da família de fatores de transcrição SET e apenas um fator de transcrição Jumonji são co-expressos com *ScRAV1* nos diferentes segmentos da raiz em cana-de-açúcar (Figura 16, Apêndice C). Os homólogos destes genes em *S. bicolor* são alvos da regulação/interação do fator de transcrição *SbRAV1* (Figura 18). Segundo Rajhi *et al.* (2011), é possível que a dinâmica das modificações nas histonas ocorra na cromatina em certos genes no aerênquima de *Z. mays* em resposta ao etileno sob condições de alagamento. Isto pode sugerir a participação dos fatores de transcrição RAV em resposta ao etileno na regulação do estado da cromatina, devido à sua possível interação com fatores de transcrição que modificam as histonas presentes no DNA nas células corticais que formarão o aerênquima lisígeno em gramíneas.

Como visto anteriormente, *ScRAV1* e o seu ortólogo *SbRAV1* em *S. bicolor* apresentam o motivo RLFGV que está fora do domínio B3 que se liga ao DNA (Figura 9), este é parte essencial para que as proteínas RAVs possam interagir com outras proteínas (Causier *et al.*, 2012), bem como parte fundamental na atividade repressora desta família de fatores de

transcrição (Ikeda & Ohme-Takagi, 2009). As interações proteína-proteína podem ocorrer dentro ou fora do domínio de ligação ao DNA e são essenciais na ação dos fatores de transcrição (Gonzalez, 2016). Essas interações são importantes na determinação da estabilidade do complexo transcricional, e também na taxa de transcrição de um determinado gene (Gonzalez, 2016). Assim, a presença de apenas um motivo AP2 (CAACA) ou ambos os motivos AP2 (CAACA) e B3 (CACCTG) para os sítios de ligação de RAV ao DNA na região promotora dos genes alvos em *S. bicolor* (Tabela 5 e 10), sugere que dependendo como *SbRAV1* se liga ao DNA - com apenas um ou com os dois domínios, este pode atuar em diferentes complexos com fatores de transcrição/co-repressores para silenciar os seus genes alvos, levando a diferentes taxas de expressão gênica. Além disso, acredita-se que a descoberta de novos sítios de ligação para RAV torna-se um passo importante para melhorar nosso entendimento sobre a regulação gênica deste fator de transcrição sobre os genes alvos em diferentes processos biológicos que este possa estar envolvido.

3.8. Redes regulatórias gênicas de genes de parede celular e de fatores de transcrição por *SbRAV1*

Com a inferência de sítios de ligação de RAV ao DNA na região promotora de sequências para os genes de parede celular e para os fatores de transcrição em *S. bicolor* (Tabela 5 e 10), foi possível construir as redes regulatórias gênicas para *SbRAV1* que é ortólogo de *ScRAV1* (Figura 8) com os genes de parede celular (Figura 13) e com os fatores de transcrição (Figura 18). De acordo com Van de Velde *et al.* (2014), a identificação individual de sítios de ligação para fatores de transcrição nas sequências do genoma é o principal objetivo para inferir as redes regulatórias, pois os fatores de transcrição (TFs) desempenham um papel fundamental na regulação transcricional que é um processo complexo e dinâmico.

As inferências das redes regulatórias de *SbRAV1* mostraram que este fator de transcrição provavelmente pode coordenar negativamente a transcrição de vários genes de parede celular (Figura 13) e alguns fatores de transcrição (*SbNAC3*, *SbNAC5*, *SbNAC7*, *SbHB6* e *SbMYB2*) associados ao desenvolvimento da parede secundária (Figura 18, Apêndice E). Já foi observado que o ortólogo de *A. thaliana* RAV2 em *E. grandis* pode estar associado à regulação da formação de elementos traqueais, especialmente as células do xilema (Cassan-Wang *et al.*, 2013). A sequência de RAV2 *A. thaliana* e quatro sequências em *E. grandis* são ortólogos de *ScRAV1* e *SbRAV1* (Figura 8), sugerindo que os genes para fatores de transcrição associados à formação da parede secundária são possivelmente inibidos por *SbRAV1* durante o desenvolvimento do aerênquima no córtex da raiz. Contudo, ainda não há evidências empíricas que RAV reprime a formação da parede secundária nas plantas.

A rede de regulação gênica de *SbRAV1* com os genes de parede celular também mostrou que *SbRAV1* pode coordenar negativamente alguns genes de degradação de polissacarídeos (Figura 13). Entre estes que podem ter uma relação com as modificações da parede celular durante a formação do aerênquima nas raízes de cana-de-açúcar (Leite *et al.*, 2017), estão os genes *SbPAE2* e *SbPGase1* que atuam na hidrólise de pectinas, e os genes *SbGH3 2* e *SbEGase4* que podem degradar tanto os polissacarídeos de celulose como hemicelulose (Figura 13, Apêndice D).

Dentre os fatores de transcrição mais promissores na regulação ou interação com *SbRAV1* durante a formação do aerênquima estão os ARFs e NACs (Figura 18). Na primeira família, observou-se os fatores de transcrição *SbARF1* e *SbARF3* que respondem à auxina (Figura 18, Apêndice E), e que também podem ser alvos da atividade transcricional negativa de *SbRAV1* durante a formação do aerênquima talvez mediando a relação entre auxina e etileno na raiz. Essas observações corroboram a hipótese do papel importante da

sinalização da auxina junto ao etileno na formação do aerênquima em cana-de-açúcar (Tavares *et al.*, 2018).

Já na família NAC de fatores de transcrição verificou-se aqueles que podem estar envolvidos na senescência foliar (*SbNAC2*) e dano ao DNA (*SbNAC8*), conforme anotação funcional dos homólogos em *A. thaliana* (Figura 18, Apêndice E). Ambos a senescência foliar e o dano ao DNA podem levar à MCP das células (Gunawardena *et al.*, 2004; Yoshiyama *et al.*, 2015), processo este que também ocorre no córtex durante a formação do aerênquima de cana-de-açúcar (Leite, 2013). Sugerindo que *SbRAV1* possivelmente age negativamente na MCP na formação do aerênquima através da regulação da transcrição dos fatores de transcrição NACs. Além disso, já foi reportado anteriormente que o fator de transcrição RAV participa na senescência foliar em *A. thaliana* (Woo *et al.*, 2010) e na formação do aerênquima em *Z. mays* (Rajhi *et al.*, 2011).

As hipóteses inferidas a partir da atividade repressora de *SbRAV1* (Figura 9) devem ser investigadas em maior profundidade com ensaios que demonstrem a capacidade deste fator de transcrição se ligar a região promotora dos seus genes alvos, e se estas inferências em *S. bicolor* se aplicam aos genes homólogos para os genes de parede celular e para os fatores de transcrição em relação à atividade repressora do seu ortólogo *ScRAV1* (Figura 9) na formação do aerênquima na raiz em cana-de-açúcar.

3.9. Perfil de expressão dos transcritos de GPs/TFs com base no RNA-Seq

Os perfis de expressão dos transcritos de parede celular (Figura 14) e fatores de transcrição (Figura 19) em cana-de-açúcar mostram quatro subgrupos com diferentes padrões de expressão no RNA-Seq da raiz de cana-de-açúcar. O agrupamento I mostrou uma alta expressão no S1 para os transcritos de parede celular (Figura 14), enquanto os fatores de

transcrição apresentou uma alta expressão nos S3 e S4. Os transcritos de parede celular neste agrupamento podem estar relacionados à formação de novas células e divisão celular devido à presença do meristema apical da raiz (Sebastian & Lee, 2013; Bizet *et al.*, 2015). Isto é evidenciado nos subgrupos A, B e C para vários transcritos de parede celular na síntese e nas modificações de polissacarídeos (*ScPGase1*, *ScEGase3*, *ScEGase4* e *ScPAE2*) na parede celular (Tabela 2, Apêndice D).

Em relação ao agrupamento I para fatores de transcrição (Figura 19), verificou-se transcritos na sinalização de auxina *ScARF1* e *ScARF3* (Figura 18, Apêndice E). Recentemente, observou-se que a sinalização de auxina pode ter um importante papel na formação do aerênquima de cana-de-açúcar (Tavares *et al.*, 2018). Já para os fatores de transcrição, observou-se que o transcrito *ScNAC2* pode estar envolvido na MCP nas células do córtex da raiz, com base no seu homólogo em *A. thaliana* (Figura 18, Apêndice E). A expressão deste fator de transcrição é condizente com o início da MCP a partir do S2 na formação do aerênquima na raiz de cana-de-açúcar (Leite, 2013). Para alguns fatores de transcrição como *ScNAC3*, *ScNAC5*, *ScNAC7* e *ScMYB2* que apresentaram padrão de expressão nos subgrupos A e B (Figura 19), observou-se que estes também podem estar relacionados ao desenvolvimento da parede secundária, conforme a anotação funcional em *A. thaliana* (Figura 18, Apêndice E).

No agrupamento II, observou-se que os transcritos de parede celular mostrou uma maior expressão nos S2 ao S4, com a presença de vários transcritos associados à biossíntese de lignina (Figura 14). Estes transcritos possivelmente estão relacionados à formação do sistema vascular (Vanholme *et al.*, 2010; Barros *et al.*, 2015; Yoon *et al.*, 2015; Zhong & Ye, 2015) ou no metabolismo secundário (Vanholme *et al.*, 2012; Labeeuw *et al.*, 2015) na raiz em cana-de-açúcar, como visto anteriormente. Já os fatores de transcrição apresentou uma maior expressão no S1 no agrupamento II (Figura 19). Estes fatores de transcrição estão relacionados

à divisão celular (*ScGRAS1*) e ao meristema apical (*ScSET2* e *ScHB6*), conforme a anotação funcional em *A. thaliana* – (Apêndice E), respectivamente. Com exceção dos fatores de transcrição *ScABI2*, *ScHB6*, *ScNAC8* que também estão envolvidos na resposta à auxina, no desenvolvimento do xilema e ao dano ao DNA (Figura 18, Apêndice E), respectivamente.

A clusterização da expressão nos segmentos mostrou que os S2 e S3 apresentam padrões de expressão mais similares entre si, e que o S4 é mais parecido com os S2 e S3 que em relação ao S1 para os transcritos de parede celular (Figura 14) e fatores de transcrição (Figura 19). Isto pode sugerir que alguns transcritos expressos principalmente nos segmentos S2 até o S4 podem ser importantes na formação do aerênquima, uma vez que a MCP se inicia no S2 (Leite, 2013) e modificações nas paredes celulares do córtex são visíveis a partir do S3 da raiz de cana-de-açúcar (Leite *et al.*, 2017).

Além disso, pode-se verificar que o perfil de expressão da maioria dos transcritos de parede celular (Figura 14) e dos fatores de transcrição (Figura 19) não são condizentes com a expressão de *ScRAV1* nos diferentes segmentos da raiz de cana-de-açúcar (Figura 14 e 19), bem como com a possível regulação transcricional proposta com o fator de transcrição *SbRAV1* para os genes de parede celular (Figura 13) e para os fatores de transcrição (Figura 18). Como se sabe, a regulação transcricional está sujeita a muitos mecanismos reguladores potencialmente sobrepostos, como regulação de microRNA (miRNA) e a acessibilidade à cromatina coordenada por modificações de histonas e metilação do DNA, a ligação de TFs em locais genômicos específicos modulando os níveis de expressão gênica que é essencial para a regulação adequada de diferentes processos biológicos (Kaufmann *et al.*, 2010; Van de Velde *et al.*, 2014).

Os eventos de ligação de TFs podem ter um efeito direto ou indireto na ativação ou repressão da transcrição gênica (Van de Velde *et al.*, 2014). A regulação da expressão gênica

pode ser mais complexa através da ligação cooperativa de diferentes TFs, adicionando um nível extra de regulação combinatória (Riechmann & Ratcliffe, 2000). Esses mecanismos regulatórios permitem que os organismos processem diferentes sinais endógenos relacionados ao crescimento e ao desenvolvimento e respondam às mudanças nas condições ambientais, incluindo diferentes tipos de estresses abióticos e bióticos (Van de Velde *et al.*, 2014). Contudo, o conhecimento sobre os genes regulados por diferentes TFs durante a formação do aerênquima é ainda escasso na literatura para plantas.

4. Seleção dos transcritos de parede celular e fatores de transcrição

Foram selecionados oito candidatos que podem estar associados à formação do aerênquima lisígeno em cana-de-açúcar. Entre eles estão quatro transcritos de parede celular que participam da hidrólise dos polissacarídeos de pectina (*ScPAE2*, *ScPGase1*) e de celulose ou hemicelulose (*ScGH3 2* e *ScEGase4*), e quatro transcritos para os fatores de transcrição que possivelmente podem estar associados à sinalização de auxina (*ScARF1*, *ScARF3*) e dois à MCP (*ScNAC2* e *ScNAC8*).

Estes transcritos foram escolhidos com base nos trabalhos do nosso grupo que mostraram alguns dos eventos que ocorrem no aerênquima de cana-de-açúcar (Leite *et al.*, 2017; Tavares *et al.*, 2018; Tavares *et al.*, 2019), e nas evidências dos seus homólogos em *A. thaliana* que participam em processos biológicos que se assemelham a formação do aerênquima.

Além disso, foram selecionados sete transcritos de parede celular relacionados à biossíntese de lignina (*ScCAD1*, *ScCCoAOMT4*, *ScCCR2*, *ScCOMT2*, *ScC3H1*, *ScC4H1* e *ScPAL2*) e cinco fatores de transcrição associados à parede secundária (*ScNAC3*, *ScNAC5*, *ScNAC7*, *ScHB6* e *ScMYB2*) devido ao padrão consistente destes transcritos neste trabalho.

4.1. *ScRAV1* possivelmente modula a expressão de glicosil hidrolases durante a formação do aerênquima

Como visto anteriormente, a hidrólise dos polissacarídeos de hemicelulose, celulose e pectina predominam nos S2, S3 e S4 nas redes de co-expressão de *ScRAV1* com os transcritos de parede celular (Figura 11). Entre os transcritos que apresentaram um pico de expressão nestes segmentos (Figura 14), possivelmente podem ser coordenados por *SbRAV1* (Figura 13) e podem estar envolvidos com a formação do aerênquima de cana-de-açúcar - estão os transcritos *ScPAE2* (uma pectina acetil esterase), *ScPGase1* (uma endo-poligalacturonase), *ScGH3 2* (uma beta-glucanase) e *ScEGase4* (uma endo-glucanase).

Os transcritos *ScPAE2* e *ScPGase1* são homólogos de PAE7 e AT3G62110.1 em *A. thaliana* (Tabela 4), respectivamente. O primeiro transcrito codifica uma pectina acetil esterase (Apêndice B), que tem como função clivar as ligações acetil éster das pectinas como o homogalacturonano (HG) e ramnogalacturonano (RG-I), os quais são acetilados nas posições O-2 e O-3 nos resíduos de ácido galacturônico (Philippe *et al.*, 2017), enquanto o segundo transcrito codifica para uma poligalacturonase (Apêndice B), que catalisa a hidrólise das ligações α -1,4-D-galacturônico do HG (Martel & Giovannoni, 2007). O papel das pectinas acetil esterases no desenvolvimento em plantas até agora permanece elusivo (Philippe *et al.*, 2017). Já as poligalacturonases têm sido principalmente relacionadas a diferentes eventos que levam à separação celular (González-Carranza *et al.*, 2007), por exemplo, na maturação de frutos (Atkinson *et al.*, 2002, 2012), no desenvolvimento do vegetal (Xiao *et al.*, 2014), na deiscência e nas zonas de abscisão de frutos (Jenkins *et al.*, 1999; Sander *et al.*, 2001).

Sabe-se que as pectinas são polissacarídeos envolvidos na adesão celular e relacionados com a porosidade da parede celular (Baron-Epel *et al.*, 1988; Willats *et al.*, 2001). Esta última característica se deve ao fato de o polímero controlar a penetração de outras

enzimas dentro da parede celular, delimitando a dimensão e o tamanho das enzimas que podem penetrar a malha de polímeros da parede celular (Buckeridge *et al.*, 2015). Acredita-se que a diminuição da adesão celular e o aumento da porosidade celular possam ser aumentadas com a ação de pectinases (Pakarinen *et al.*, 2012; Latarullo *et al.*, 2016).

Os níveis de expressão no RNA-Seq mostram que o transcrito *ScPAE2* é preferencialmente expresso entre os S1 e S4, enquanto o transcrito da *ScPGase1* tem maior expressão no S3 (Figura 14). A formação do aerênquima lisígeno em cana-de-açúcar é evidente nos S3 e S4, regiões em que ocorre a separação celular no córtex da raiz durante o desenvolvimento deste tecido (Leite *et al.*, 2017). Estes autores demonstraram, através do ensaio de imunolocalização utilizando anticorpo monoclonal para o homogalacturonano (CCRC-M38), que a marcação do anticorpo diminui quando se inicia a abertura dos espaços gasosos no córtex da cana-de-açúcar. A hipótese é que haja a remoção de pectinas como o homogalacturonano presente na lamela média das células corticais (Knox, 1992). Os dados de imunolocalização de ouro coloidal reforçam esta hipótese devido à uma pequena quantidade ter sido encontrada deste polímero nas paredes celulares depois da formação do aerênquima em cana-de-açúcar (Leite *et al.*, 2017). Além disso, o padrão observado no córtex da raiz de cana-de-açúcar nos S2 e S3 é de uma ruptura das conexões feitas pela lamela média entre as ligações triangulares (junção pectínica entre as três células). Este processo parece ser o primeiro passo para a formação do aerênquima em raízes de cana-de-açúcar (Leite *et al.*, 2017).

De forma similar, no aerênquima de *Z. mays* também foram previamente observadas modificações nas pectinas desesterificadas e esterificadas usando os anticorpos monoclonais JIM5 e JIM7 (Gunawardena *et al.*, 2001b), respectivamente. A atividade enzimática e expressão gênica de pectinases também foi detectada durante a formação do aerênquima em

Z. mays (Bragina *et al.*, 2003; Rajhi *et al.*, 2011). Em cana-de-açúcar, foi reportado que *ScRAV1*, um fator de transcrição de resposta ao etileno, reprime a expressão da *ScEPG1* quando este se liga na região promotora desta endopoligalacturonase (Tavares *et al.*, 2019). Corroborando com essas observações, verificamos uma correlação negativa de *ScRAV1* com *ScEPG1* no S3 da raiz de cana-de-açúcar (Figura 10C, Apêndice B) quando do exame das redes de correlações de *ScRAV1* com os transcritos de parede celular.

No que se refere ao *ScGH3 2*, observou-se que este transcrito pertence à família GH3 das glicosil hidrolases (Apêndice B), e é homólogo de BGCL1 em *A. thaliana* (Tabela 4), uma enzima que codifica para uma β -glucosidase que se caracteriza por ser uma xiloglucano exoglucosidase (Sampedro *et al.*, 2017). Ela é capaz de remover os resíduos de glicose não substituídos na extremidade não redutora das moléculas do xiloglucano, sem afetar a ligação do polímero à parede celular (Sampedro *et al.*, 2017). O seu ortólogo (TmBGCL1), uma enzima caracterizada em *T. majus* tem um importante papel na remobilização da reserva do xiloglucano durante a germinação em plântulas (Crombie *et al.*, 1998), a qual também é membro da família GH3 (Lombard *et al.*, 2014). É possível que BGCL1 seja ortólogo de outras duas β -glucosidase chamadas ExoI e ExoII, que hidrolisam ambas as ligações $\beta(1,4)$ e $\beta(1,3)$, assumindo que elas podem agir no glucano de ligação mista (GLM) durante a germinação de plântulas em *H. vulgare* (Hrmova *et al.*, 1996; Varghese *et al.*, 1999).

Já a endo- β -1,4-glucanase *ScEGase4* pertence à família GH9 das glicosil hidrolases (Apêndice B) e é conhecida como uma celulase (Cantarel *et al.*, 2009). Esta enzima cliva as ligações β -1,4-glucosídicas da celulose, do xiloglucano, do glucomanano e do GLM (Cosgrove, 2003; Yoshida & Komae, 2006; Lopez-Casado *et al.*, 2008; Buchanan *et al.*, 2012). *ScEGase4* é possivelmente homólogo da GH9B8 ou endo- β -1,4-glucanase 11 de *A. thaliana* (Tabela 4). Este gene pertence à classe B da família GH9 (Urbanowicz *et al.*, 2007) e está associado a

diferentes estágios do desenvolvimento e do crescimento vegetal (Wilson & Urbanowicz, 2018). Essa classe de endo- β -1,4-glucanase foi reportada como sendo importante na expansão celular, nas zonas de abscisão, e durante o amadurecimento do fruto (Libertini *et al.*, 2004; Wilson & Urbanowicz, 2018). Além disso, um estudo demonstrou que a enzima GH9B8 em *A. thaliana* é superexpressa em resposta a aplicação de auxina (Goda *et al.*, 2004).

O perfil de expressão para o transcrito *ScGH3 2* mostra que este teve um pico de expressão a partir S3 e S4, enquanto para o transcrito *ScEGase4* observou-se uma maior expressão nos S2 e S4 no RNA-Seq da raiz de cana-de-açúcar (Figura 14). Apesar de não ter sido observada a hidrólise da celulose durante o aerênquima de cana-de-açúcar pela caracterização bioquímica da parede celular pelos métodos utilizados por Leite *et al.* (2017), mas na proteômica e no transcriptoma observou-se a expressão de várias celulasas durante o desenvolvimento deste tecido nesta espécie (Grandis, 2015; Tavares *et al.*, em preparação), sugerindo possivelmente a participação destas enzimas neste processo celular. No que se refere à formação do aerênquima de *Z. mays* verificou-se que a atividade enzimática celulasas aumentam (He *et al.*, 1992; Bragina *et al.*, 2003). Rajhi *et al.*, (2011) também verificou um alta expressão gênica para esta classe de enzimas. A expressão destes transcritos coincide parcialmente com a diminuição do glucano de ligação mista (GLM) no córtex da raiz durante o desenvolvimento do aerênquima (Leite *et al.*, 2017), o qual pode ser substrato de hidrólise destas celulasas, como descrito anteriormente.

O GLM no córtex da raiz pode estar atuando como um “andaime” (*scaffold*) para a montagem da parede celular. A sua remoção normalmente ocorre após completar-se o processo de montagem da parede celular. Em suporte a esta ideia, há observações prévias de que o GLM atua como um polissacarídeo de transição, sendo sintetizado em tecidos onde

ocorre o alongamento e degradado quando o alongamento celular termina (Carpita, 1996; Kim *et al.*, 2000).

Apesar destes transcritos para glicosil hidrolases não apresentarem um perfil de expressão inverso ao perfil de expressão de *ScRAV1* (Figura 14) e, portanto, não condizente com o possível papel regulatório predito para *ScRAV1* com base no seu ortólogo em *S. bicolor* (Figura 13), acredita-se que outros fatores podem estar afetando a expressão destes transcritos nos diferentes segmentos da raiz durante a formação do aerênquima, como descrito no parágrafo 5 do item 3.9. Além disso, reportou-se que vários genes de poligacturonases, pectina acetil esterases, endo- β -1,4-glucanase e beta-glucosidases são também responsivos ao hormônio auxina (Goda *et al.*, 2004).

4.2. *ScRAV1* provavelmente modula a expressão dos fatores de transcrição que respondem à auxina

Nas redes de co-expressão de *ScRAV1* com os fatores de transcrição, foi observado que os fatores de transcrição ARFs co-expressam com *ScRAV1* do S2 ao S4 (Figura 16). Os fatores de transcrição chamados ARFs (*Auxin Response Factors*) são componentes que conferem especificidade na resposta à auxina (Li *et al.*, 2016). Sabe-se que a auxina tem um importante papel no crescimento e no desenvolvimento dos vegetais. Entre eles estão o alongamento, a divisão e a diferenciação celular (Abel & Theologis, 1996; Li *et al.*, 2016). Apesar do hormônio etileno ser descrito como o principal responsável por promover a formação de aerênquima (Jackson & Armstrong, 1999; Drew *et al.*, 2000; Yamauchi *et al.*, 2013), a influência da auxina na formação do aerênquima também já foi reportada em *Z. mays*, onde verificou-se que baixos níveis de auxina (ácido 1-naftilacético – NAA, hormônio sintético de auxina) aumentaram a formação do aerênquima e reduziram-o quando houve altos

níveis deste hormônio. Acredita-se que esta resposta se deve ao fato de a auxina promover a biossíntese de etileno (Justin & Armstrong, 1991).

Recentemente, foi sugerido que não somente a sensibilidade ao etileno, mas o balanço entre etileno-auxina pode ter um papel importante durante na formação do aerênquima constitutivo na raiz de cana-de-açúcar (Tavares *et al.*, 2018). Os autores observaram que a formação do aerênquima em cana-de-açúcar ocorre quando os níveis de etileno e auxina diminuem na raiz, e que a auxina pode estar envolvida na expressão de genes relacionados à expansão da parede celular, o qual, juntamente com a hidrólise de pectinas na lamela média, consiste nos primeiros eventos que levam à formação do aerênquima (Grandis *et al.*, 2014; Gunawardena *et al.*, 2001b; Leite *et al.*, 2017).

A auxina regula diversos processos através do controle da expressão gênica de fatores de transcrição ARFs (Li *et al.*, 2016). Estes fatores de transcrição fazem isso se ligando a elementos de resposta à auxina chamados como AuxREs (*Auxin-Responsive Elements*), os quais se ligam aos promotores dos genes regulados por auxina através do reconhecimento da sequência motivo (5' - TGTCTC - 3'). Estes fatores podem tanto ativar como reprimir a transcrição dos seus genes alvos (Ulmasov *et al.*, 1997; Ulmasov *et al.*, 1999; Tiwari *et al.*, 2003; Li *et al.*, 2016). Foi identificado um fator de transcrição ARF durante o desenvolvimento do aerênquima em *O. sativa* (Yoo *et al.*, 2015). Neste trabalho, também foram observados dois genes da família de fatores de transcrição ARF na rede de regulação gênica em *S. bicolor* que se referem aos transcritos *ScARF1* e *ScARF3* em cana-de-açúcar (Figura 18), os quais parecem estar envolvidos na formação do aerênquima lisígeno em cana-de-açúcar, como será descrito abaixo.

Os transcritos *ScARF1* e *ScARF3* são possíveis homólogos de ARF2 (Tabela 9). Este fator de transcrição funciona como um repressor transcricional na sinalização de auxina em *A.*

thaliana (Tiwari *et al.*, 2003). Ele atua em vários processos biológicos análogos à formação do aerênquima, como por exemplo: na floração - na senescência e na abscisão do órgão floral, na maturação de frutos, na deiscência de síliquas e na senescência foliar, agindo nestes processos como um regulador positivo (Ellis *et al.*, 2005; Okushima *et al.*, 2005; Lim *et al.*, 2010; Hao *et al.*, 2015). Porém, ARF2 também pode agir como regulador negativo na divisão celular, na expansão celular, no crescimento de órgãos e na formação do gancho apical em *A. thaliana* (Li *et al.*, 2004; Schruff *et al.*, 2006).

O perfil de expressão de *ScARF1* e *ScARF3* mostrou que estes transcritos apresentaram um acúmulo da sua expressão principalmente nos S3 e S4 (Figura 19), onde se inicia e se completa o aerênquima em cana-de-açúcar (Leite *et al.*, 2017), respectivamente. O perfil de expressão destes transcritos contrastam com os altos níveis de auxina que foram encontrados na ponta da raiz (S1 e S2) e que diminuem em direção ao (S5) na raiz de cana-de-açúcar (Tavares *et al.*, 2018). Segundo Lim *et al.* (2010), o hormônio auxina funciona como supressor da senescência foliar, e ARF2 é um repressor da sinalização de auxina. Os autores sugeriram que a auxina participa suprimindo a senescência foliar, mas esta supressão é diminuída com a idade da folha, possivelmente através do aumento da expressão de ARF2, o qual é responsável por mediar a sinalização de auxina regulando-a temporalmente e espacialmente durante a senescência foliar. Este dado corrobora com o perfil de expressão para os transcritos *ScARF1* e *ScARF3* (Figura 19) em relação aos níveis de auxina nos diferentes segmentos da raiz de cana-de-açúcar encontrados por Tavares *et al.* (2018), sugerindo uma possível regulação negativa na expressão destes transcritos em relação a este hormônio e possivelmente na expansão celular, levando assim ao favorecimento da formação do aerênquima na raiz de cana-de-açúcar.

Além disso, verificamos que *ScRAV1* e *ScARF1* e *ScARF3* apresentam um padrão de expressão inverso. *ScRAV1* apresenta maior expressão nos S1 e S2 (Figura 19), e foi confirmada por qPCR por Tavares *et al.* (2019). Isto ocorreu a despeito de *ScARF1* e *ScARF3* apresentarem maior expressão nos S3 e S4 (Figura 19). Os fatores transcrição RAV e ARF2 respondem aos hormônios etileno e auxina, respectivamente (Alonso *et al.*, 2003; Ellis *et al.*, 2005). Foi visto por Tavares *et al.* (2018), que os níveis de etileno nas raízes de cana-de-açúcar durante a formação do aerênquima são maiores nos S1 e S2 e diminui em direção ao S5, acompanhando o padrão de expressão de *ScRAV1* do RNA-Seq (Figura 19) e por qPCR (Tavares *et al.*, 2019). Porém, os níveis de etileno contrastam com o perfil de expressão dos transcritos de *ScARF1* e *ScARF3* nas raízes de cana-de-açúcar durante a formação do aerênquima.

Neste trabalho, verificou-se que *SbRAV1* ortólogo de *ScRAV1* tem sítios de ligação para RAV na região promotora dos homólogos *ScARF1* e *ScARF3* em *S. bicolor* (Tabela 10). Assim, é possível inferir que haja um papel regulatório negativo de *ScRAV1* via etileno sobre a expressão destes transcritos nos dois primeiros segmentos que antecede a formação do aerênquima na raiz de cana-de-açúcar (Figura 18). Corroborando com isso, já foi visto que os fatores de transcrição RAVs (RAV1, TEM1 e TEM2) e ARF2 participam na senescência foliar e na floração em *A. thaliana* (Hu *et al.*, 2004; Ellis *et al.*, 2005; Castillejo & Pelaz, 2008; Zhao *et al.*, 2008; Lim *et al.*, 2010; Woo *et al.*, 2010). A atuação destes fatores de transcrição nos mesmos processos biológicos leva a acreditar que haja uma possível interação/regulação entre eles, pois no caso da senescência foliar, assim como no do aerênquima, há módulos funcionais em comum (Grandis *et al.*, 2014).

O cross-talk entre etileno e auxina via ARF2 tem sido observado em alguns processos biológicos. Por exemplo, na maturação de frutos de *S. lycopersicum* por SIARF2 ortólogo de ARF2 em *A. thaliana*, observou-se que o silenciamento de SIARF2 resultou na produção de

menos etileno e na regulação negativa de reguladores importantes na maturação dos frutos como RIN, CNR, NOR e TAGL1 (Hao *et al.*, 2015). Estes resultados apontaram que SlARF2 como um novo componente da rede regulatória do processo de maturação em *S. lycopersicum*, sugerindo que auxina junto com etileno podem ser hormônios essenciais na maturação dos frutos (Hao *et al.*, 2015).

Outro exemplo de processo biológico em que módulos similares estão envolvidos é a formação do gancho apical em plântulas de *A. thaliana*. Li *et al.* (2004) verificaram que o nível da proteína ARF2 foi diminuído quando as plântulas foram expostas ao etileno, sugerindo que o acúmulo da proteína ARF2 é negativamente regulado por etileno, o qual é dependente de HLS1 (*HOOKLESS1*), um gene de resposta ao etileno que é responsável pelas interações cruzadas entre etileno e auxina durante a formação do gancho apical nas plântulas (Lehman *et al.*, 1996). Além disso, também foi sugerido que etileno poderia mediar a degradação da proteína ARF2 via proteossomo (Li *et al.*, 2004). Este dado corrobora com a hipótese deste trabalho, sugerindo que em diferentes processos biológicos que ocorrem em diferentes espécies, há módulos em comum (Grandis *et al.*, 2004; Tavares *et al.*, 2015).

Finalmente, já foi observado anteriormente que a maioria dos genes pertencentes à família ACS, uma enzima chave na biossíntese de etileno, pode ser induzida por auxina (Yamagami *et al.*, 2003; Tsuchisaka & Theologis, 2004). De acordo com Okushima *et al.* (2005), a expressão de ACS2, ACS6 e ACS8 foi inibida no mutante *arf2-6* no tempo de floração, possivelmente devido à inibição da atividade de ACS. Além disso, verificou-se que os genes ACS6 e ACS8 apresentaram nas suas regiões promotoras 4 a 6 AuxREs, respectivamente, sugerindo uma possível regulação de ARF2 sobre os genes de biossíntese de etileno (Okushima *et al.*, 2005).

Os dados acima corroboram a hipótese de que há uma interação entre auxina e etileno durante o desenvolvimento do aerênquima em cana-de-açúcar (Tavares *et al.*, 2018). Assim, avaliar a interação *in vivo* de *ScRAV1* com os transcritos *ScARF1* e *ScARF3*, seria importante para entender a interação entre etileno e auxina via estes transcritos durante a formação do aerênquima na raiz de cana-de-açúcar.

4.3. *ScRAV1* possivelmente coordena a expressão de fatores de transcrição relacionados à MCP no córtex da raiz de cana-de-açúcar

Acredita-se que a sinalização de etileno é um aspecto importante para regular a MCP durante a formação do aerênquima lisígeno (He *et al.*, 1996; Drew *et al.*, 2000), pois, este hormônio induz a produção ROS no início do desenvolvimento do aerênquima lisígeno de *O. sativa* (Parlanti *et al.*, 2011; Steffens *et al.*, 2011; Yamauchi *et al.*, 2017), *Z. mays* (Yamauchi *et al.*, 2011) e *T. aestivum* (Yamauchi *et al.*, 2014). Corroborando com isto, uma proteína chamada RAV (EDF3, AT3G25730) responsável na sinalização de etileno (Alonso *et al.*, 2003), é superexpressa no transcriptoma quando há a produção de ROS em *A. thaliana* (Gadjev *et al.*, 2006). Um possível homólogo desta proteína, *ZmRAV1L* (GRMZM2G169654) é superexpresso somente nas células corticais que irão formar o aerênquima em *Z. mays* (Rajhi *et al.*, 2011). Com base na árvore filogenética do fator de transcrição RAV, é possível inferir que *ScRAV1* é ortólogo de ambos RAV (EDF3, AT3G25730) em *A. thaliana* e de *ZmRAV1L* (GRMZM2G169654) em *Z. mays* (Figura 8), e que a expressão do seu transcrito é maior nos segmentos iniciais (S1 e S2) que antecede a formação do aerênquima em cana-de-açúcar (Tavares *et al.*, 2019). Assim, é possível sugerir que *ScRAV1* via etileno pode estar envolvido na sinalização de ROS na MCP de células específicas do córtex durante o desenvolvimento do aerênquima lisígeno na raiz de cana-de-açúcar.

Uma das famílias fatores de transcrição relacionada à MCP é família NAC devido à sua participação principalmente na formação da parede secundária (Zhong *et al.*, 2007) e na senescência foliar (Guo *et al.*, 2004; Buchanan-Wollaston *et al.*, 2005; Liu *et al.*, 2009). Contudo, alguns autores também mostraram que a superexpressão de OsNAC5 e OsNAC9 aumentou o eustelo e o aerênquima em *O. sativa* cv *Nipponbare* (Redillas *et al.*, 2012; Jeong *et al.*, 2013). Neste trabalho, foram encontrados vários fatores de transcrição da família NAC que se co-expressam com *ScRAV1* em todos os segmentos da raiz de cana-de-açúcar (Figura 15 e 16). Já na rede regulatória gênica de *SbRAV1* com fatores de transcrição (Figura 18), foram observados dois fatores de transcrição da família NAC que podem estar envolvidos com MCP e ROS durante a formação do aerênquima. Entre os genes homólogos de *S. bicolor* encontrados em cana-de-açúcar estão os transcritos *ScNAC2* e *ScNAC8* (Figura 18), como será descrito abaixo.

O transcrito *ScNAC2* é um possível homólogo de ANAC087 em *A. thaliana* (Tabela 9). Recentemente, observou-se que ANAC087 orchestra a degradação da cromatina *post-mortem* da região distal da coifa lateral (*Lateral Root Cap*) na zona de alongamento, por promover indiretamente a expressão da nuclease BFN1 (*Bifunctional Nuclease1*) e de enzimas hidrolíticas, bem como controlar o tempo de MCP na região proximal da coifa lateral e na columela na raiz de *A. thaliana* (Huysmans *et al.*, 2018). Segundo os autores, este fator de transcrição pode estar envolvido na regulação da MCP sendo ativado em vários tecidos em diferentes circunstâncias como na diferenciação celular, na senescência e em estresses. É possível que o nível de expressão possa determinar a função deste fator de transcrição em relação ao seu papel na MCP, ou através da interação com outros fatores de transcrição especificamente para ajustar a MCP nas células (Huysmans *et al.*, 2018). Os dados destes

autores mostram que ANAC087 é preferencialmente expresso em células preparando-se para MCP.

ANAC087 também foi expresso à jusante do regulador ORE1 (*ORESARA1/NAC2/ANAC092*) durante a senescência foliar (Kim *et al.*, 2014). Foi observada alta expressão para ANAC087 no transcriptoma de *A. thaliana* de folhas senescentes, conforme Buchanan-Wollaston *et al.* (2005). Outro estudo mais recente observou-se que o fator de transcrição (BnaNAC87) em *Brassica napus* L., um possível ortólogo de ANAC087 em *A. thaliana*, obteve uma alta expressão em folhas senescentes medida por qRT-PCR (Yan *et al.*, 2018). A superexpressão de BnaNAC87 induziu a expressão de alguns genes como RbohB (*Respiratory burst oxidase homolog B*) e VPE1a (*Vacuolar Processing Enzyme a*), implicados na geração de ROS e morte celular, respectivamente. Conforme os autores, este fator de transcrição está presente no núcleo, e mostrou ser um ativador dos seus genes alvos, portanto, eles concluíram que BnaNAC87 é um regulador positivo do metabolismo de ROS e morte celular (Yan *et al.*, 2018).

Assim como na senescência foliar tem-se verificado que a expressão de RBOH, enzima responsável pela geração de ROS (Li *et al.*, 2017), aumenta em condições que favorecem a formação aerênquima em raízes de gramíneas. Em *Z. mays*, observou-se a superexpressão de uma RBOH nas células corticais da raiz em condições de alagamento (Rajhi *et al.*, 2011, Yamauchi *et al.*, 2011). Já o pré-tratamento com ACC levou ao aumento da expressão de três genes TaRBOH em *T. aestivum* (Yamauchi *et al.*, 2014). No caso da cana-de-açúcar, foi observado o acúmulo do transcrito referente à RBOH no S5 nas raízes quando tratadas com 1-MCP durante a formação do aerênquima (Tavares, 2015). Estes dados sugerem que possivelmente *ScNAC2* pode estar envolvido na MCP e na produção de ROS no córtex da raiz de cana-de-açúcar.

O homólogo BnaNAC87 de *ScNAC2* induziu a morte celular em folhas senescentes, o qual foi confirmada através da técnica de TUNEL pela fragmentação nuclear do DNA que é um indício de morte celular (Yan *et al.*, 2018). No perfil de expressão do RNA-Seq de cana-de-açúcar para *ScNAC2*, pode-se verificar que este transcrito apresenta maior expressão no S3 (Figura 19), aonde observou-se um maior pico de núcleos marcados pela técnica de TUNEL, os quais diminuem gradativamente em direção ao S4 e S5 em células que irão formar o aerênquima em cana-de-açúcar (Leite, 2013). Em *Z. mays*, também se observou que os núcleos positivos pela técnica de TUNEL diminuem com o aumento das células corticais que morrem para formar o aerênquima (Gunawardena *et al.*, 2001a).

Verificando a expressão *ScRAV1* em relação à *ScNAC2* nos diferentes segmentos da raiz da cana-de-açúcar no RNA-Seq, pode-se observar que *ScRAV1* apresenta uma alta expressão no S1 e S2, enquanto *ScNAC2* apresenta uma menor expressão nestes dois segmentos iniciais (Figura 19). O gene homólogo de *ScNAC2* em *S. bicolor* apresentou sítios de ligação de RAV ao DNA (Tabela 10), e a rede regulatória mostrou que *SbRAV1* ortólogo de *ScRAV1* pode ser um repressor da transcrição de *SbNAC2* (Figura 18). Isto pode sugerir que *ScRAV1* modula negativamente a produção de ROS e a MCP por uma possível regulação negativa sobre o transcrito *ScNAC2* nos dois primeiros segmentos iniciais da raiz que antecede a formação do aerênquima em cana-de-açúcar. Corroborando com o fato de que o fator de transcrição RAV1 foi descrito como um regulador negativo do desenvolvimento do vegetal (Hu *et al.*, 2004).

O outro transcrito *ScNAC8* é um possível homólogo de SOG1 (*Supressor of Gamma Response 1*) em *A. thaliana* (Tabela 9), que contribui para a estabilidade do genoma depois do dano do DNA, devido ao fato de induzir várias vias como a parada do ciclo celular, reparo do DNA, MCP e endoreduplicação através da regulação de vários genes (Yoshiyama *et al.*, 2015).

As quebras de duplas fitas no DNA (DSBs; *DNA Double Strand Breaks*) são um sinal ao dano ao DNA que pode ser causado por ROS (Chen & Umeda, 2015), as quais são moléculas altamente reativas e prejudiciais as células vegetais, por danificarem os componentes celulares como o DNA, proteínas e lipídeos, e ainda levar a morte celular (Apel & Hirt, 2004; Mittler *et al.*, 2004), se os níveis de ROS forem suficientemente altos (Tsukagoshi, 2016). Chen & Umeda (2015) sugeriu que a produção de ROS pode ser responsável pelas DSBs mediadas pelo fator de transcrição SOG1, o qual é responsável por ligar-se diretamente na região promotora de FMO1, a qual codifica uma flavina-monooxygenase que está associada em particular na produção de (H₂O₂) (Chen & Umeda, 2015). Os autores acreditam que as DSBs mediadas por SOG1 induzem a expressão de FMO1 levando ao acúmulo de (H₂O₂) na zona de alongamento celular (EDZ, *ELONGATION ZONE*) nas raízes de *A. thaliana*.

No caso, de cana-de-açúcar a EDZ possivelmente refere-se aos segmentos S2 ao S4 da raiz de cana-de-açúcar (Leite *et al.*, 2017). A fragmentação do DNA é uma das características da MCP durante a formação do aerênquima. O início da MCP foi observado a partir do segmento S2 através da técnica de TUNEL devido à fragmentação do DNA nas células do córtex durante o desenvolvimento do aerênquima em cana-de-açúcar (Leite, 2013), e a mesma técnica também foi utilizada no aerênquima de *Z. mays* (Gunawardena *et al.*, 2001a). Contudo não há dados sobre a produção de espécies reativas de oxigênio em cana-de-açúcar durante o desenvolvimento do aerênquima.

SOG1 ainda induz diretamente a expressão dos genes SMR5 e SMR7 (*SIAMESE-RELATED*), que são inibidores de quinase dependentes de ciclina que levam à parada do ciclo celular devido às DSBs pelo estresse oxidativo (Yi *et al.*, 2014). A alta expressão nos S1, S2 e S4 do seu homólogo *ScNAC8* em cana-de-açúcar (Figura 19), pode sugerir que este transcrito atue na parada do ciclo celular no S1, e o início da MCP das células específicas do córtex no S2

que se tornarão o aerênquima em cana-de-açúcar (Leite, 2013). O aumento da expressão deste transcrito no S4 (Figura 19), possivelmente deve favorecer ainda mais a formação do aerênquima em cana-de-açúcar, levando mais células a serem em MCP, uma vez que já foi observado um maior número de espaços de gasosos neste segmento comparado com os três segmentos iniciais (Leite, 2013). Isto possivelmente indica a participação de *ScNAC8* na MCP via ROS no desenvolvimento do aerênquima em cana-de-açúcar.

Acredita-se que SOG1 pode mediar a MCP por vias que controlam a expressão de fatores chave que levam a morte celular, pois recentemente foi demonstrado por imunoprecipitação de cromatina (ChIP) que ANAC087 pode ser um possível alvo de SOG1 (Ogita *et al.*, 2018). Como visto no parágrafo 4 e 5 desta seção, ANAC087 e BnaNAC87 estão relacionados à MCP e à produção de ROS, de acordo com os autores Huysmans *et al.* (2018) e Yan *et al.* (2018). Este dado é condizente com a expressão inversa dos seus homólogos *ScNAC2* e *ScNAC8* no RNA-Seq de cana-de-açúcar (Figura 19), sugerindo talvez uma regulação entre eles na homeostase da produção de ROS e MCP de células específicas que irão formar o aerênquima na raiz de cana-de-açúcar.

Apesar da expressão de *ScNAC8* não ser condizente com a regulação direta do ortólogo em *S. bicolor* para *ScRAV1* como proposto na rede regulatória (Figura 18), a sua expressão precisa ainda ser validada e pode sugerir que este fator de transcrição pode ser regulado por outros fatores que regulam a transcrição deste gene em cana-de-açúcar. Além disso, foi reportado a participação de SOG1 na resistência à patógeno (Ogita *et al.*, 2018). Sabe-se que as proteínas RAVs estão envolvidas vários estresses bióticos e abióticos como resposta ao frio (Fowler *et al.*, 2005), na defesa a patógenos, à seca e ao estresse salino (Sohn *et al.*, 2006), os quais induzem a formação de ROS e consequentemente podem levar à MCP (Mittler *et al.*, 2004; Mittler *et al.*, 2011). Além disso, um gene envolvido na resposta na defesa vegetal

também está envolvido na formação do aerênquima lisígeno nos hipocótilos em *A. thaliana* (Mühlenbock *et al.*, 2007). Isto sugere que diferentes processos possivelmente são regulados pelos mesmos genes, mas com vias regulatórias diferentes. Assim, sugere ainda que haja uma possível interação de *ScRAV1* e *ScNAC8* durante a formação do aerênquima, porém esta interação precisa ser melhor averiguada.

4.4. *ScRAV1* possivelmente coordena a expressão de genes relacionados à biossíntese de lignina no córtex durante a formação do aerênquima na raiz de cana-de-açúcar

Observou-se que houve uma predominância de transcritos de parede celular para biossíntese de lignina (Figura 10A a D) e para os fatores de transcrição como NACs, MYB e HOMEBOX em todas as redes de co-expressão de *ScRAV1* em cana-de-açúcar (Figura 16A a D) e também nas redes de regulação gênica com os seus genes homólogos em *S. bicolor* (Figura 13 e 18). Dentre os transcritos de parede celular estão: *ScCAD1*, *ScCCoAOMT4*, *ScCCR2*, *ScCOMT2*, *ScC3H1*, *ScC4H1* e *ScPAL2* em cana-de-açúcar, os quais são homólogos de *ATCAD6*, *CCoAOMT1*, *AT1G76470.1*, *COMT1*, *REF8*, *C4H* e *PAL2* em *A. thaliana*, respectivamente (Tabela 4). Estes genes estão envolvidos na via de biossíntese de lignina (Vanholme *et al.*, 2010; Barros *et al.*, 2015; Yoon *et al.*, 2015; Zhong & Ye, 2015) ou no metabolismo secundário nas plantas (Vanholme *et al.*, 2012; Labeeuw *et al.*, 2015).

Em relação aos transcritos para fatores de transcrição em cana-de-açúcar verificou-se a presença dos transcritos: *ScNAC3*, *ScNAC5*, *ScNAC7*, *ScHB6* e *ScMYB2*, os quais são possíveis homólogos de *NAC075*, *ANAC083*, *NAC073*, *HB-6* e *MYB4* em *A. thaliana*, respectivamente (Tabela 9). Estes fatores de transcrição *NAC075*, *ANAC083* e *NAC073* estão na primeira camada de *masters switches* da parede secundária, enquanto o *MYB4* está na segunda camada de *master switch* para a deposição de lignina em *A. thaliana* (Nakano *et al.*, 2015).

Estes fatores de transcrição desempenham diferentes papéis na formação da parede secundária. Os fatores de transcrição NAC075 e ANAC083 são reguladores positivo e negativo na formação dos vasos do xilema em *A. thaliana*, respectivamente (Nakano *et al.*, 2015; Yamaguchi *et al.*, 2010). Referente ao fator de transcrição NAC073 ou SND2 (*SECONDARY WALL-ASSOCIATED NAC DOMAIN2*), este está envolvido na ativação da expressão de síntese de celulose, de hemicelulose e de lignina na parede secundária de fibras em *A. thaliana* (Hussey *et al.*, 2011). Já o fator de transcrição MYB4 regula negativamente os genes na biossíntese de lignina como o 4CL1 e C4H em *A. thaliana* (Jin *et al.*, 2000). No que diz a respeito do fator de transcrição HB-6 ou também conhecido como BLR (*BELLRINGER*), observou-se em mutantes triplos contendo *blr*, *brevipedicellus (bp)* e o fator de transcrição *revoluta (rev)* que o desenvolvimento do xilema foi notavelmente afetado, resultando quase que na falta de vasos e um espessamento secundário das células parenquimáticas do xilema (Etchells *et al.*, 2012). É possível que os transcritos homólogos em cana-de-açúcar atuem de forma similar durante o desenvolvimento do cilindro vascular.

O perfil de expressão para os transcritos de biossíntese de lignina mostrou que a maioria destes transcritos apresentou uma maior expressão entre o S2 ao S4, com exceção de *ScC4H1* que apresentou maior expressão no S1 da raiz de cana-de-açúcar (Figura 14), respectivamente. Já os fatores de transcrição *ScNAC3*, *ScNAC5*, *ScNAC7* e *ScMYB2* apresentaram maior expressão entre o S3 e S4, enquanto o fator de transcrição *ScHB6* apresentou uma alta expressão no S1 ao S3, e diminui no S4 (Figura 19). A expressão destes transcritos pode estar relacionada com a formação da parede secundária nos elementos de vaso do xilema e fibras na raiz de cana-de-açúcar, como visto por Leite *et al.* (2017). Já comparando com a expressão de *ScRAV1* observou-se uma maior expressão nos dois segmentos iniciais da raiz (Figura 19), enquanto a expressão da maioria dos transcritos de

parede celular e fatores de transcrição tendem apresentar pico de expressão a partir do S2 ao S4, com exceção dos transcritos *ScC4H1* (Figura 14) e *ScHB6* (Figura 19), como visto acima.

Além disso, foi observada a presença de sítios putativos de ligação ao DNA para RAV para os genes homólogos para biossíntese de lignina (Tabela 5) e para os fatores de transcrição em *S. bicolor* (Tabela 10). Contribuindo com esta informação, foi reportado anteriormente que a expressão do ortólogo de *A. thaliana* RAV2 em *E. grandis* é *down-regulated* durante a indução de elementos traqueais (preferencialmente em células do xilema) (Cassan-Wang *et al.*, 2013). O fator de transcrição RAV2 em *A. thaliana* e quatro sequências em *E. grandis* são ortólogos de *ScRAV1*, *SbRAV1* e *ZmRAV1L*, conforme a Figura 8. Para este último fator de transcrição não se observou a sua expressão por PCR semi-quantitativa no tecido vascular da raiz em condições de alagamento na formação do aerênquima em *Z. mays* (Rajhi *et al.*, 2011). Corroborando com a hipótese que os fatores de transcrição RAV provavelmente agem negativamente no desenvolvimento da parede secundária em plantas.

Devido *SbRAV1* ser ortólogo de *ScRAV1* (Figura 8), os dados da regulação gênica em *S. bicolor* podem sugerir que os genes homólogos para transcritos de parede celular (Figura 13) e para os fatores de transcrição (Figura 18) em cana-de-açúcar podem ser coordenados transcricionalmente negativamente por *ScRAV1* nos dois segmentos iniciais do córtex da raiz de cana-de-açúcar durante a formação do aerênquima e da parede secundária. Isto pode ocorrer tanto através de uma possível regulação direta sobre os genes de parede celular de biossíntese de lignina, ou através dos fatores de transcrição que representam a primeira e a segunda camada de *master switches* para formação da parede secundária, como descrito nos parágrafos 1 e 2 desta seção.

No primeiro caso, é possível que *ScRAV1* talvez module a expressão destes transcritos negativamente, pois estes transcritos são o último passo na regulação da via de síntese de lignina/fenilpropanóides (Hussey *et al.*, 2011). Em relação aos fatores de transcrição, a modulação negativa de *ScRAV1* sob a expressão destes transcritos dependerá de qual papel eles desempenham na síntese de lignina e em qual nível hierárquico de regulação transcricional eles estão neste processo, isto é, se eles agem ativando ou reprimindo outros fatores de transcrição e/ou genes de parede celular relacionados à biossíntese de lignina na parede celular, como visto para *A. thaliana* no parágrafo 3 desta seção.

Sabe-se que no processo de formação da parede secundária assim como na formação do aerênquima ocorre MCP (Fukuda, 1996). De acordo com Kozela & Regan (2003), se compararmos as células do xilema que apresentam parede secundária (lignificação) com as células do aerênquima pode-se verificar que ambas passam por morte celular e autólise celular, resultando na produção de um espaço funcional dentro do corpo da planta (Kozela & Regan, 2003). A única diferença entre os dois tipos de células é a presença da parede secundária nos tecidos condutores de água (Kozela & Regan, 2003). Ainda pode-se observar que ambas as células aparentam uma semelhança morfológica em forma de tubo (lacuna), aonde as primeiras células são especializadas pelo transporte de água, enquanto as últimas são responsáveis pelo transporte de oxigênio, respectivamente (Kozela & Regan, 2003).

Assim, entender como *ScRAV1* modula ambos os processos pode ser essencial para diminuir a lignificação nas paredes celulares do colmo de cana-de-açúcar futuramente. Pois, a biossíntese de lignina tornou-se um dos focos principais, uma vez que se sabe que este polímero é um dos principais fatores que levam à recalcitrância da parede celular para um processamento eficiente de obtenção de etanol de 2G (Vanholme *et al.*, 2012). Este polímero impede a liberação de monossacarídeos na hidrólise enzimática dos polissacarídeos da parede

celular durante o processo de sacarificação, que é necessário para a produção de bioetanol de 2G a partir da biomassa lignocelulósica (Chen & Dixon, 2007; Yoshida *et al.*, 2008).

VI. CONCLUSÕES

Este trabalho verificou na árvore filogenética que há quatro sequências para os fatores de transcrição da família RAV em cana-de-açúcar, sendo três destas sequências são parálogos de *ScRAV1* e apenas uma sequência podendo ser uma isoforma ou uma duplicação gênica deste fator de transcrição. Na caracterização dos domínios de *ScRAV1* observou-se que os domínios AP2 e B3 são conservados, bem como o motivo BRD – RLFGV que pode ser essencial na atividade repressiva de *ScRAV1*. Além disso, nas regiões promotoras dos homólogos para os genes de parede celular e fatores de transcrição em *S. bicolor* observou-se os motivos AP2 (CAACA) e B3 (CACCTG) como sítios de ligação para RAV, sugerindo que este fator de transcrição pode atuar como repressor independente de possuir apenas um ou dois motivos na região promotora destes genes.

As redes de co-expressão de *ScRAV1* mostraram que este fator de transcrição tem co-expressão tanto com genes de biossíntese e degradação da parede celular, e com várias famílias de fatores de transcrição. Já as centralidades de grau das redes de co-expressão de *ScRAV1* permitiram identificar transcritos que possivelmente são essenciais no desenvolvimento do aerênquima. A anotação dos transcritos de cana-de-açúcar em *S. bicolor*, e este último em *A. thaliana* possibilitaram a anotação funcional e a escolha dos transcritos de parede celular e fatores de transcrição que possivelmente agem durante o desenvolvimento do aerênquima. A busca de sítios de ligação para RAV para os motivos AP2 (CAACA) e B3 (CACCTG) em *S. bicolor* para genes homólogos de parede celular e para fatores de transcrição em cana-de-açúcar sugeriram possíveis alvos este fator de transcrição durante a formação do aerênquima em cana-de-açúcar.

O fator de transcrição *ScRAV1* está associado a uma série de processos durante o desenvolvimento da raiz de cana-de-açúcar. Aparentemente, este fator de transcrição atua

tanto em eventos meristemáticos assim como controlando o metabolismo de carboidratos que geram os substratos para a síntese de parede celular. Ainda ele possivelmente controla também processos de hidrólise de hemiceluloses e pectinas na porção superior ao ápice da raiz, levando às modificações que formam o aerênquima. Nossas redes de correlação também permitiram vislumbrar o que parece ser o início do processo que irão completar o tecido vascular (deposição de lignina e celulose no xilema e fibras).

Nossos resultados, em conjunto com o que foi encontrado na literatura e com os trabalhos prévios do grupo, dão suporte à hipótese de que a formação do aerênquima em cana de açúcar ilustra a existência de módulos funcionais que ocorrem em diferentes espécies vegetais. O mecanismo de controle do metabolismo da parede celular envolve a sinalização por etileno (via RAV) e auxina (via ARF). Nas porções S2 e S3 das raízes de cana-de-açúcar, ocorre um processo de modificação nas paredes celulares fortemente controlado, cujo contexto molecular e metabólico parece ser parte de um padrão fisiológico que se replica como parte de processos fisiológicos distintos (senescência foliar, floral, amadurecimento de frutos, p.ex.) e nos vegetais de diferentes grupos taxonômicos.

Dentre os candidatos de parede celular possivelmente alvos da regulação de *ScRAV1* estão dois transcritos relacionados à hidrólise de pectina - (*ScPAE2* e *ScPGase1*), dois transcritos relacionados à hidrólise de hemicelulose/celulose (*ScGH3 2* e *ScEGase4*) e vários transcritos associados com a biossíntese de lignina. A degradação de pectina nas células do córtex da raiz de cana-de-açúcar é evidente, e um passo importante no processo bioetanol de 2G uma vez que elas diminuem a adesão celular e aumentam a porosidade celular, facilitando a ação de outras enzimas na hidrólise de outros polissacarídeos da parede celular na produção de bioetanol 2G. Em relação à provável regulação deste fator de transcrição sobre os transcritos de parede celular e fatores de transcrição na biossíntese de lignina, pode abrir

caminho para produzir uma cana-de-açúcar menos lignificada e com maior acessibilidade aos outros polissacarídeos da parede celular, uma vez que este polímero é um dos gargalos para a produção de bioetanol 2G.

Entre outros candidatos possivelmente alvos de *ScRAV1* estão fatores de transcrição de resposta à auxina (*ScARF1*, *ScARF3*) e MCP (*ScNAC2*, *ScNAC8*), e podem estar envolvidos na formação dos espaços gasosos no córtex da raiz de cana-de-açúcar. Assim, *ScRAV1* pode ser é um ótimo candidato a interagir com vários transcritos de parede celular e fatores de transcrição durante a formação do aerênquima. Este pode ser o caminho para entender como este processo biológico de degradação da parede celular pela própria planta pode ser utilizado para a produção de bioetanol de 2G em um futuro próximo.

VII. REFERÊNCIAS BIBLIOGRÁFICAS

- Abel S, Theologis A. 1996. Early genes and auxin action. **Plant Physiology** 111 (1), 9-17.
- Aden A, Ruth M, Ibsen K, Jechura J, Neeves K, Sheehan J, Wallace B, Montague L, Slayton A, Lukas J. 2002. Lignocellulosic biomass to ethanol process design and economics utilizing co-current dilute acid prehydrolysis and enzymatic hydrolysis for corn stover. **National Renewable Energy Laboratory Technical Report** NREL/TP-510-32438.
- Adie B, Chico JM, Rubio-Somoza I, Solano R. 2007. Modulation of Plant Defenses by Ethylene. **Journal of Plant Growth Regulation** 26 (2), 160-177.
- Aitken KS, McNeil MD, Berkman PJ, Hermann S, Kilian A, Bundock PC, Li J. 2014. Comparative mapping in the Poaceae family reveals translocations in the complex polyploid genome of sugarcane. **BMC Plant Biology** 14, 190.
- Albert R, Barabási AL. 2002. Statistical mechanics of complex networks. **Reviews of Modern Physics** 74 (1), 47-97.
- Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, Suzuki M. 1998. A novel mode of DNA recognition by a β -sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. **The EMBO Journal** 17 (18), 5484-5496.
- Alonso JM, Hirayama T, Roman G, Nourizadeh S, Ecker JR. 1999. EIN2, a bifunctional transducer of ethylene and stress responses in *Arabidopsis*. **Science** 284 (5423), 2148-2152.
- Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Mayers CC, Parker H, Prednis L, Ansaris Y, Choy N, Deen H, Geralt M, Hazari N, Hom E, Karnes M, Mulholland C, Ndubaku R, Schmidt I, Guzman P, Aguilar-Henonin L, Schmid M, Weigel D, Carter DE, Marchand T, Risseuw E, Brogden D, Zeko A, Crosby WL, Berry CC, Ecker JR. 2003.

- Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. **Science** 301 (5633), 653–657.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. **Journal of Molecular Biology** 215 (3), 403-410.
- Apel K, Hirt H. 2004. Reactive Oxygen Species: metabolism, oxidative stress, and signal transduction. **Annual Review of Plant Physiology and Plant Molecular** 55,373-399.
- Arber A. 1920. **Water plants, a study of aquatic angiosperms**. Cambridge, UK: Cambridge University Press.
- Arikado H, Adachi Y. 1955. Anatomical and ecological responses of barley and some forage crops to the flooding treatment. **Bulletin of the Faculty of Agriculture**, Mie University 11, 1–29.
- Armstrong, W. 1979. Aeration in higher plants. **Advances in Botanical Research** 7, 236–332.
- Ashtiani M, Salehzadeh-Yazdi A, Razaghi-Moghadam Z, Henning H, Wolkenhauer O, Mirzaie M, Jafari M. 2018. A systematic survey of centrality measures for protein-protein interaction networks. **BMC Systems Biology** 12 (1), 80.
- Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, Albrecht M. 2008. Computing topological parameters of biological networks. **Bioinformatics** 24 (2), 282-284.
- Atkinson RG, Schröder R, Hallett IC, Cohen D, MacRae EA. 2002. Overexpression of polygalacturonase in transgenic apple trees leads to a range of novel phenotypes involving changes in cell adhesion. **Plant Physiology** 129 (1), 122–133.
- Atkinson RG, Sutherland PW, Johnston SL, Gunaseelan K, Hallett IC, Mitra D, Brummell DA, Schröder R, Johnston JW, Schaffer RJ. 2012. Down-regulation of POLYGALACTURONASE1 alters firmness, tensile strength and water loss in apple (*Malus x domestica*) fruit. **BMC Plant Biology** 12, 129.

- Banerjee G, Scott-Craig JS, Walton JD. 2010. Improving Enzymes for Biomass Conversion: A Basic Research Perspective. **Bioenergy Research** 3 (1), 82-92.
- Barabási AL, Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. **Nature Reviews Genetics** 5 (2), 101–113.
- Baron-Epel O, Gharyal PK, Schindler M. 1988. Pectins as mediators of wall porosity in soybean cells. **Planta** 175 (3), 389–395.
- Barros J, Serk H, Granlund I, Pesquet E. 2015. The cell biology of lignification in higher plants. **Annals of Botany** 115 (7), 1053-1074.
- Barry CS, McQuinn RP, Thompson AJ, Seymour GB, Grierson D, Giovannoni JJ. 2005. Ethylene insensitivity conferred by the Green-ripe and Never-ripe 2 ripening mutants of tomato. **Plant Physiology** 138 (1), 267-275.
- Begum MK, Miah MAS, Islam MS, Hossain MA, Alam MR. 2008. Effects of periodic flood stress on sugarcane yield and yield contributing characters. **Sugar Tech** 10(4), 355-358.
- Belesini AA, Carvalho FMS, Telles BR, de Castro GM, Giachetto PF, Vantini JS, Carlin SD, Cazetta JO, Pinheiro DG, Ferro MIT. 2017. De novo transcriptome assembly of sugarcane leaves submitted to prolonged water-deficit stress. **Genetics and Molecular Research** 16(2), 1-20.
- Bienz M. 2006. The PHD finger, a nuclear protein-interaction domain. **Trends in Biochemical Sciences** 31 (1), 35-40.
- Bizet F, Hummel I, Bogeat-Triboulot MB. 2015. Length and activity of the root apical meristem revealed *in vivo* by infrared imaging. **Journal of Experimental Botany** 66 (5), 1387-1395.

- Bouranis DL, Chorianopoulou SN, Siyiannis VF, Protonotarios VE, Hawkesford MJ. 2003. Aerenchyma formation in roots of maize during sulphate starvation. **Planta** 217 (3), 382–391.
- Bragina TV, Martinovich LI, Rodionova NA, Bezborodov AM, Grineva GM. 2001. Ethylene induced activation of xylanase in adventitious roots of maize as a response to the stress effect of root submersion. **Applied Biochemistry and Microbiology** 37 (6), 722–725.
- Bragina TV, Rodionova NA, Grineva GM. 2003. Ethylene production and activation of hydrolytic enzymes during acclimation of maize seedlings to partial flooding. **Russian Journal of Plant Physiology** 50 (6), 794–798.
- Buanafina MMO. 2009. Feruloylation in grasses: current and future perspectives. **Molecular Plant** 2 (5), 861–872.
- Buanafina MMO, Cosgrove DJ. 2014. Cell walls: Structure and Biogenesis. In: Moore PH, Botha FC. (Ed.). **Sugarcane: Physiology, Biochemistry & Functional Biology**. 1st ed. New Delhi: Wiley Blackwell. pp. 307-329.
- Buchanan M, Burton RA, Dhugga KS, Rafalski AJ, Tingey SV, Shirley NJ, Fincher GB. 2012. Endo-(1,4)- β -glucanase gene families in the grasses: temporal and spatial co-transcription of orthologous genes. **BMC Plant Biology** 12 (235), 1-19.
- Buchanan-Wollaston V, Page T, Harrison E, Breeze E, Lim PO, Nam HG, Lin JF, Wu SH, Swidzinski J, Ishizaki K, Leaver CJ. 2005. Comparative transcriptome analysis reveals significant differences in gene expression and signalling pathways between developmental and dark/starvation-induced senescence in *Arabidopsis*. **The Plant Journal** 42 (4), 567–585.

- Buckeridge MS, De Souza AP. 2014. Breaking the “Glycomic Code” of cell wall polysaccharides may improve second-generation bioenergy production from biomass. **Bioenergy Research** 7 (4), 1065–1073.
- Buckeridge MS, De Souza AP, Tavares EQP, Cambler AB. 2017. Sugarcane cell wall structure and degradation: from monosaccharide analyses to the glycomic code. In: Buckeridge MS, De Souza AP (Ed.). **Advances of Basic Science for Second Generation Bioethanol from Sugarcane**. New York: Springer. pp. 7-19.
- Buckeridge MS, Dos Santos WD, Tiné MAS, De Souza AM. 2015. The cell wall architecture of sugarcane and its implications to cell wall recalcitrance. In: Lam E, Carrer H, Silva JA. (Ed.). **Compendium of Bioenergy Plants: Sugarcane**. 1st ed. New York: CRC Press. pp.31-501.
- Buckeridge MS, Grandis A, Tavares EQP. 2019. Disassembling the Glycomic Code of Sugarcane Cell Walls to Improve Second-Generation Bioethanol Production. In: Ray RC, Ramachandran S (Ed.). **Bioethanol Production from Food Crops – Sustainable Sources, Interventions, and Challenges**. pp. 31-43.
- Buckeridge MS, Rayon C, Urbanowicz B, Tiné MAS, Carpita NC. 2004. Mixed linkage (1,3), (1,4) - β -D-glucans of grasses. **Cereal Chemistry** 81 (1),115–127.
- Buckeridge MS, Santos WD, De Souza AP. 2010. Routes for cellulosic ethanol in Brazil. In: Cortex LAB. (Ed.). **Sugarcane bioethanol: R&D for productivity and sustainability**. 1st ed. São Paulo: Edgar Blucher. pp. 365-390.
- Buckeridge MS, Souza AP, Arundale RA, Anderson-Teixeira KJ, DeLucia E. 2012. Ethanol from sugarcane in Brazil: a “midway” strategy for increasing ethanol production while maximizing environmental benefits. **Global Change Biology and Bioenergy** 4 (2), 119-126.

- Caffall KH, Mohnen D. 2009. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. **Carbohydrate Research** 344 (14), 1879–1900.
- Calvo-Flores F, Dobado JÁ, Isac-García J. 2015. **Lignin and Lignans as Renewable Raw Materials – Chemistry, Technology and Applications**. John Wiley & Sons. pp. 9-42.
- Camagna M, Takemoto D. 2018. Hypersensitive Response in Plants. In: **Encyclopedia of Life Sciences**. John Wiley & Sons pp. 1-7.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active Enzymes database (CAZy): an expert resource for Glycogenomics. **Nucleic Acids Research** 37 (Database issue), D233-D238.
- Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, Amigo Hub, Web Presence Working Group. 2009. AmiGO: online access to ontology and annotation data. **Bioinformatics** 25 (2), 288-289.
- Cardona CA, Quintero JA, Paz IC. 2010. Production of bioethanol from sugarcane bagasse: status and perspectives. **Bioresource Technology** 101(13), 4754-4766.
- Carpita NC. 1996. Structure and biogenesis of the cell walls of grasses. **Annual Review of Plant Physiology and Plant Molecular Biology** 47 (1), 445–476.
- 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. **The Plant Journal** (1), 1-30.
- Carpita NC, McCann MC. 2008. Maize and sorghum: genetic resources for bioenergy grasses. **Trends in Plant Science** 13 (8), 415-420.
- Carpita NC, Ralph J, McCann MC. 2015. The Cell Wall. In: Buchanan BB, Gruissem W, Jones RL. (Ed.). **Biochemistry & Molecular Biology of Plants**. 2nd ed. UK: John Wiley & Sons Ltd. pp. 45-110.

- Carrasco JL, Ancilo G, Mayda E, Vera P. 2003. A novel transcription factor involved in plant defense endowed with protein phosphatase activity. **The EMBO Journal** 22 (13), 3376-3384.
- Cassan-Wang H, Goué N, Saldi MN, Legacy S, Sivadin P, Goffner D, Grima-Pettenati J. 2013. Identification of novel transcription factors regulating secondary cell wall formation in *Arabidopsis*. **Frontiers in Plant Science** 4 (189), 1- 14.
- Castillejo C, Pelaz S. 2008. The balance between CONSTANS and TEMPRANILLO activities determines FT expression to trigger flowering. **Current Biology** 18 (17),1338–1343.
- Causier B, Ashworth M, Guo W, Davies B. 2012. The TOPLESS Interactome: a framework for gene repression in *Arabidopsis*. **Plant Physiology** 158 (1), 423–438.
- Chao Q, Rothenberg M, Solano R, Roman G, Terzaghi W, Ecker JR. 1997. Activation of the ethylene gas response pathway in *Arabidopsis* by the nuclear protein ETHYLENE-INSENSITIVE3 and related proteins. **Cell** 89 (7), 1133-1144.
- Chen F, Dixon RA. 2007. Lignin modification improves fermentable sugar yields for biofuel production. **Nature Biotechnology** 25 (7), 759–761.
- Chen P, Umeda M. 2015. DNA double-strand breaks induce the expression of Flavin-containing monooxygenase and reduce root meristem size in *Arabidopsis thaliana*. **Genes to Cells** 20 (8), 636-646.
- Chen R, Li Q, Tan H, Chen J, Xiao Y, Ma R, Gao S, Zerbe P, Chen W, Zhang L. 2015. Gene-to-metabolite network for biosynthesis of lignans in MeJA-elicited *Isatis indigotica* hairy root cultures. **Frontiers in Plant Science** 6, 952.
- Chikina MD, Troyanskaya OG. 2011. Accurate quantification of functional analogy among close homologs. **PLoS Computational Biology** 7 (2), e1001074.

- Choi H, Hong J, Ha J, Kang J, Kim SY. 2000. ABFs, a family of ABA-responsive element binding factors. **The Journal of Biological Chemistry** 275(3),1723–1730.
- Colmer TD. 2003. Long-distance transport of gases in plants: a perspective on internal aeration and radial oxygen loss from roots. **Plant, Cell & Environment** 26 (1), 17–36.
- Colmer TD, Pedersen O. 2008. Oxygen dynamics in submerged rice (*Oryza sativa*). **The New Phytologist** 178 (2), 326–334.
- Colmer TD, Peeters AJM, Wagemaker CAM, Vriezen WH, Ammerlaan A, Voeseek LACJ. 2004. Expression of α -expansin genes during root acclimations to O₂ deficiency in *Rumex palustris*. **Plant Molecular Biology** 56 (3), 423–437.
- Cosgrove DJ. 2003. Expansion of the plant cell wall. In: Rose JKC (Ed.). **The Plant Cell Wall – Annual Plant Reviews**, Volume 8. USA: Blackwell Publishing Ltd. pp:1-381.
- Costa MC, Righetti K, Nijveen H, Yazdanpanah F, Ligterink W, Buitink J, Hilhorst HW. 2015. A gene co-expression network predicts functional genes controlling the re-establishment of desiccation tolerance in germinated *Arabidopsis thaliana* seeds. **Planta** 242 (2), 435–449.
- Cregg JM. 1999. Expression in methylotrophic yeast *Pichia pastoris*. In: Fernandez JM, and Hoeffler JP. (Ed). **Gene Expression Systems: Using Nature for the Art of Expression**. New York: Academic Press. pp 157–191.
- Crivellari AC. 2012. Caracterização estrutural das hemiceluloses de paredes celulares de cana-de-açúcar. **Tese (Mestrado)**. Instituto de Biociências da Universidade de São Paulo. Departamento de Botânica. 95p.
- Crombie HJ, Chengappa S, Hellyer A, Reid JSG. 1998. A xyloglucan oligosaccharide-active, transglycosylating beta-D-glucosidase from the cotyledons of nasturtium (*Tropaeolum*

- majus* L) seedlings: purification, properties and characterization of a cDNA clone. **The Plant Journal** 15 (1), 27–38.
- Crowell EF, Gonneau M., Vernhettes S., Hofte H. 2010. Regulation of anisotropic cell expansion in higher plants. **Comptes Rendus Biologies** 333 (4), 320–324.
- Dai X, Wang Y, Yang A, Zhang WH. 2012. OsMYB2P-1, an R2R3 MYB transcription factor, is involved in the regulation of phosphate-starvation responses and root architecture in rice. **Plant Physiology** 159 (1), 169–183.
- Daniels J, Smith P, Panton N, Williams CA. 1975. The origin of the genus *Saccharum*. **Sugarcane Breeding Newsletter** 36, 24–39.
- Das RC, Shultz JL. 1987. Secretion of heterologous proteins from *Saccharomyces cerevisiae*. **Biotechnology Progress** 3 (1), 43–48.
- Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, Grotewold E. 2003. AGRIS: *Arabidopsis* gene regulatory information server, an information resource of *Arabidopsis* cis-regulatory elements and transcription factors. **BMC Bioinformatics** 4, 25.
- De Bary A. 1884. **Comparative anatomy of the vegetative organs of phanerogams and ferns**. Oxford, UK: Oxford University Press.
- de Setta N, Monteiro-Vitorello CB, Metcalfe CJ, Cruz GMQ, Del Bem LE, Vicentini R, Nogueira FTS, Campos RA, Nunes SL, Turrini PCG, Vieira AP, Cruz AO, Corrêa TCS, Hotta CT, Varani AM, Vautim S, Trindade AS, Vilela MM, Lembke CG, Sato PM, Andrade RF, Nishiyama Jr MY, Cardoso-Silva CB, Scortecci KC, Garcia AAF, Carneiro MS, Kim C, Paterson AH, Bergès H, D' Hont A, Souza AP, Souza GM, Vicentz M, Kitajima JP, Van Sluys MA. 2014. Building the sugarcane genome for biotechnology and identifying evolutionary trends. **BMC Genomics** 15, 540.

- De Siqueira Ferreira S, Nishiyama MY Jr., Paterson AH, Souza GM. 2013. Biofuel and energy crops: high-yield *Saccharinae* take center stage in the post-genomics era. **Genome Biology** 14 (6), 210-222.
- De Souza AP. 2011. Mecanismos fotossintéticos e relação fonte-dreno em cana-de-açúcar cultivada em atmosfera enriquecida de CO₂. **Tese (Doutorado)**. Instituto de Biociências da Universidade de São Paulo. Departamento de Botânica. 208p.
- De Souza AP, Grandis A, Leite DCC, Buckeridge MS. 2014. Sugarcane as a bioenergy source: history performance and perspectives for second-generation bioethanol. **Bioenergy Research** 7 (1),24-35.
- De Souza AP, Leite DCC, Pattathil S, Hahn MG, Buckeridge MS. 2013. Composition and structure of sugarcane cell wall polysaccharides: implications for second generation bioethanol. **Bioenergy Research** 6 (2), 564-579.
- De Vries RP, Battaglia E, Coutinho PM, Henrissat B, Visser J. 2011. (Hemi-) Cellulose Degrading Enzymes and Their Encoding Genes from *Aspergillus* and *Trichoderma*. In: Hofrichter M. (Ed). **The Mycota (A Comprehensive Treatise on Fungi as Experimental Systems for Basic and Applied Research) - Industrial Applications X**. 2nd ed. Berlin: Springer. pp. 341-355.
- Dos Santos WD, Gómez EO, Buckeridge MS. 2011. Bioenergy and the sustainable revolution. In: Buckeridge MS, Goldman GH (Ed.). **Routes to cellulosic ethanol**. 1st ed. New York: Springer. pp 15-26.
- Drew MC, He CJ, Morgan PW. 1989. Decreased ethylene biosynthesis, and induction of aerenchyma, by nitrogen-starvation or phosphate-starvation in adventitious roots of *Zea mays* L. **Plant Physiology** 91 (1), 266–271.

- Drew MC, He CJ, Morgan PW. 2000. Programmed cell death and aerenchyma formation in roots. **Trends in Plant Science** 5 (3), 123-127.
- Drew MC, Jackson MB, Giffard SC, Campbell R. 1981. Inhibition by silver ions of gas space (aerenchyma) formation in adventitious roots of *Zea mays* L. subjected to exogenous ethylene or to oxygen deficiency. **Planta** 153 (3), 217–224.
- Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. 2010. MYB transcription factors in *Arabidopsis*. **Trends in Plant Science** 15 (10), 573-581.
- Dufour P, Deu M, Grivet L, D'Hont A, Paulet F, Bouet A, Lanaud C, Glaszmann JC, Hamon P. 1997. Construction of a composite *Sorghum* genome map and comparison with sugarcane, a related complex polyploid. **Theoretical and Applied Genetics** 94 (3-4), 409–418.
- Ebanks R, Dupont M, Shareck F, Morosoli R, Kluepfel D, Dupont, C. 2000. Development of an *Escherichia coli* expression system and thermostability screening assay for libraries of mutant xylanase. **Journal of Industrial Microbiology & Biotechnology** 25 (6), 310–314.
- Ellis CM, Nagpal P, Young JC, Hagen G, Guilfoyle TJ, Reed JW. 2005. Auxin Response Factor 1 and Auxin Response Factor 2 regulate senescence and floral organ abscission in *Arabidopsis thaliana*. **Development** 132 (20), 4563-74.
- Etchells JP, Moore L, Jiang WZ, Prescott H, Capper R, Saunders NJ, Bhatt AM, Dickinson HG. 2012. A role for BELLRINGER in cell wall development in supported by loss-of-function phenotypes. **BMC Plant Biology** 12, 212.
- Evans DE. 2003. Aerenchyma formation. **The New Phytologist** 161 (1), 35-49.
- Fagerstedt KV. 2010. Programmed cell death and aerenchyma formation under hypoxia. In: Mancuso S, Shabala S (Ed.). **Waterlogging Signalling and Tolerance in Plants**. Berlin: Springer. pp. 99–118.

- Fang G, Bhardwaj N, Robillotto R, Gerstein MB. 2010. Getting started in gene orthology and functional analysis. **PloS Computational Biology** 6 (3), e1000703.
- Faria JA, Reis PA, Reis MT, Rosado GL, Pinheiro GL, Mendes GC, Fontes EP. 2011. The NAC domain-containing protein, GmNAC6, is a downstream component of the ER stress- and osmotic stress-induced NRP- mediated cell-death signaling pathway. **BMC Plant Biology** 11, 129.
- Feller A, Machemer K, Braun EL, Grotewold E. 2011. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. **The Plant Journal** 66 (1), 94-116.
- Feng CZ, Chen Y, Wang C, Kong YH, Wu WH, Chen YF. 2014. *Arabidopsis* RAV1 transcription factor, phosphorylated by SnRK2 kinases, regulates the expression of ABI3, ABI4, and ABI5 during seed germination and early seedling development. **The Plant Journal** 80 (4), 654-668.
- Fincher GB. 2009. Exploring the evolution of (1,3;1,4)-beta-D-glucans in plant cell walls: comparative genomics can help!. **Current Opinion in Plant Biology** 12 (2), 140–147.
- Fincher GB, Stone BA. 2004. Chemistry of nonstarch polysaccharides. In: Wrigley C, Corke H, Walker CE (Ed.). **Encyclopedia of grain science**. 1st ed. Oxford: Elsevier.pp. 206–222.
- Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. **Nucleic Acids Research** 44 (D1), D279-285.
- Fowler SG, Cook D, Thomashow MF. 2005. Low temperature induction of *Arabidopsis* CBF1, 2, and 3 is gated by the circadian clock. **Plant Physiology** 137 (3), 961–968.
- Fukuda H. 1996. Xylogenesis: initiation, progression, and cell death. **Annual Review of Plant Physiology and Plant Molecular Biology** 47, 299-325.

- Gadjev I, Stone JM, Gechev TS. 2008. Programmed cell death in plants: new insights into redox regulation and the role of hydrogen peroxide. **International Review of Cell and Molecular Biology** 270, 87-144.
- Gadjev I, Vanderauwera S, Gechev TS, Laloi C, Minkoi IN, Shulaev V, Apel K, Inzé D, Mitter R, Van Breusegem F. 2006. Transcriptomic Footprints Disclose Specificity of Reactive Oxygen Species Signaling in *Arabidopsis*. **Plant Physiology** 141 (2), 436-445.
- Garsmeur O, Charron C, Bocs S, Jouffe V, Samain S, Couloux A, Droc G, Zini C, Glaszmann JC, Van Sluys MA, D' Hont A. 2011. High homologous gene conservation despite extreme autopolyploid redundancy in sugarcane. **New Phytologist** 189 (2), 629-642.
- Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, Jenkins J, Martin G, Charron C, Hervouet C, Costet L, Yahiaoui N, Healey A, Sims D, Cherukuri Y, Sreedasyam A, Killian A, Chan A, Van Sluys MA, Swaminathan K, Town C, Bergès, Simmons B, Glaszmann JC, van der Vossen E, Henry R, Schmutz J, D'Hont A. 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. **Nature Communications** 9 (2638), 1-10.
- Gaston K, Jayaraman PS. 2003. Transcriptional repression in eukaryotes: repressors and repression mechanisms. **Cellular and Molecular Life Sciences CMLS** 60 (4), 721–741.
- Gechev TS, Van Breusegem F, Stone JM, Denev I, Laloi C. 2006. Reactive oxygen species as signals that modulate plant stress responses and programmed cell death. **Bioessays** 28 (11), 1091–1101.
- Ghasemi M, Seidkhani H, Tamimi F, Rahgozar M, Masoudi-Nejad A. 2014. Centrality Measures in Biological Networks. **Current Bioinformatics** 9 (4), 1-17.

- Gibeaut DM, Carpita NC. 1993. Synthesis of (1→3), (1→4)-beta-D-glucan in the Golgi apparatus of maize coleoptiles. **Proceedings of the National Academy of Sciences of the United States of America** 90 (9),3850–3854.
- Gilbert RA, Rainbolt CR, Morris DR, Bennett AC. 2007. Morphological Responses of Sugarcane to Long-Term Flooding. **Agronomy Journal** 99, 1622-1628.
- Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM. 1992. Isolation of the *Arabidopsis* ABI3 gene by positional cloning. **The Plant Cell** 4 (10), 1251-1261.
- Glover NM, Redestig H, Dessimoz C. 2016. Homoeologs: What Are They and How Do We Infer Them?. **Trends in Plant Science** 21 (7), 609-621.
- Goda H, Sawa S, Asami T, Fujioka S, Shimada Y, Yoshida S. 2004. Comprehensive comparison of auxin-regulated and brassinosteroid-regulated genes in *Arabidopsis*. **Plant Physiology** 134 (4), 1555–1573.
- Goldemberg J. 2008. The Brazilian biofuel industry. **Biotechnology for Biofuels** 1(6), 1-7.
- Gonzalez, DH. 2016. Introduction to Transcription Factor Structure and Function. In: Gonzalez, DH (Ed.). **Plant Transcription Factors – Evolutionary, Structural and Functional Aspects**. 1st ed. Londres: Academic Press. pp.3-11.
- González-Carranza ZH, Elliott KA, Roberts JA. 2007. Expression of polygalacturonases and evidence to support their role during cell separation processes in *Arabidopsis thaliana*. **Journal of Experimental Botany** 58 (13), 3719–3730.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS. 2012. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research** 40, D1178–D1186.

- Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. **Nucleic Acids Research** 36 (10), 3420-3435.
- Grandis A. 2015. Papel das enzimas de degradação da parede celular na formação do aerênquima em raízes de cana-de-açúcar. **Tese (Doutorado)**. Instituto de Biociências da Universidade de São Paulo. Departamento de Botânica. 122p.
- Grandis A, De Souza AP, Tavares EQP, Buckeridge MS. 2014. Using Natural Plant Cell Wall Degradation Mechanisms to Improve Second Generation Bioethanol. In: McCann MC, Buckeridge MS, Carpita NC. (Ed.). **Plants and BioEnergy**. 1st ed. New York: Springer. pp. 211-230.
- Grativol C, Regulski M, Bertalan M, McCombie WR, da Silva FR, Zerlotini Neto A, Vicentini R, Farinelli L, Hemerly AS, Martienssen RA, Ferreira PC. 2014. Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus *Saccharum*. **The Plant Journal** 79(1), 162–172.
- Guimaraes CT, Sills GR, Sobral BW. 1997. Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. **Proceedings of the National Academy of Sciences of the United States of America** 94 (26), 14261–14266.
- Gunawardena AHLAN, Greenwood JS, Dengler NG. 2004. Programmed Cell Death Remodels Lace Plant Leaf Shape during Development. **The Plant Cell** 16 (1), 60-73.
- Gunawardena A, Pearce DM, Jackson MB, Hawes CR, Evans DE. 2001a. Characterisation of programmed cell death during aerenchyma formation induced by ethylene or hypoxia in roots of maize (*Zea mays* L.). **Planta** 212 (2), 205–214.
- Gunawardena A, Pearce DM, Jackson MB, Hawes CR, Evans DE. 2001b. Rapid changes in cell wall pectic polysaccharides are closely associated with early stages of aerenchyma

- formation, a spatially localized form of programmed cell death in roots of maize (*Zea mays* L.) promoted by ethylene. **Plant, Cell & Environment** 24 (12), 1369–1375.
- Guo H, Ecker JR. 2003. Plant responses to ethylene gas are mediated by SCF (EBF1/EBF2)-dependent proteolysis of EIN3 transcription factor. **Cell** 115 (6), 667-677.
- Guo Y, Cai Z, Gan S. 2004. Transcriptome of *Arabidopsis* leaf senescence. **Plant Cell and Environment** 27 (5), 521–549.
- Hall BP, Shakeel SN, Schaller GE. 2007. Ethylene receptors: ethylene perception and signal transduction. **Journal of Plant Growth Regulation** 26 (2), 118-130.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symposium Series** 41, 95-98.
- Hansen BO, Vaid N, Musialak-Lange M, Janowski M, Mutwil M. 2014. Elucidating gene function and function evolution through comparison of co-expression networks of plants. **Frontiers in Plant Science** 5 (394), 1-9.
- Hao Y, Hu G, Breitel D, Liu M, Mila I, Frasse P, Fu Y, Aharoni A, Bouzayen M, Zouine M. 2015. Auxin Response Factor SIARF2 is an Essential Component of the Regulatory Mechanism Controlling Fruit Ripening in Tomato. **Plos Genetics** 11 (12), 1-23.
- Hayashi, T. 1989. Xyloglucans in the primary cell wall. **Annual Review of Plant Physiology and Plant Molecular Biology** 40, 139-168.
- He CJ, Drew MC, Morgan PW. 1994. Induction of Enzymes Associated with Lysigenous Aerenchyma Formation in Roots of *Zea mays* during Hypoxia or Nitrogen Starvation. **Plant Physiology** 105 (3), 861-865.
- He CJ, Morgan PW, Drew MC. 1992. Enhanced sensitivity to ethylene in nitrogen-starved or phosphate-starved roots of *Zea mays* L. during aerenchyma formation. **Plant Physiology** 98 (1), 137–142.

- He CJ, Morgan PW, Drew MC. 1996. Transduction of an ethylene signal is required for cell death and lysis in the root cortex of maize during aerenchyma formation induced by hypoxia. **Plant Physiology** 112 (2),463–472.
- Held MA, Jiang N, Basu D, Showalter AM, Faik A. 2015. Plant Cell Wall Polysaccharides: Structure and Biosynthesis. In: Ramawat KG, Mérillon JM. (Ed.). **Polysaccharides Bioactivity and Biotechnology**. 1st ed. New York: Springer Reference. pp. 3-54.
- Higo K, Ugawa Y, Iwamoto M, Korenaga T.1999. Plant *cis*-acting regulatory DNA elements (PLACE) database:1999. **Nucleic Acids Research** 27 (1), 297-300.
- Himi E, Maekawa M, Noda K. 2011. Differential Expression of Three Flavanone 3-Hydroxylase Genes in Grains and Coleoptiles of Wheat. **International Journal of Plant Genomics** 369460, 1-11.
- 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.
- Hirayama T, Kieber JJ, Hirayama N, Kogan M, Guzman P, Nourizadeh S, Alonso JM, Dailey WP, Dancis A, Ecker JR. 1999. RESPONSIVE-TO-ANTAGONIST1, a Menkes/Wilson disease-related copper transporter, is required for ethylene signaling in *Arabidopsis*. **Cell** 97 (3), 383-393.
- Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, Botha FC, Henry RJ. 2017. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. **BMC Genomics** 18(1), 395.
- Hofmann NR. 2016. A Structure for Plant-Specific Transcription Factors: The GRAS Domain Revealed. **The Plant Cell** 28 (5), 993-994.

- Hong JC. 2016. General Aspects of Plant Transcription Factor Families. In: Gonzalez, DH (Ed.). **Plant Transcription Factors – Evolutionary, Structural and Functional Aspects**. 1st ed. Londres: Academic Press. pp.35-56.
- Hrmova M, Harvey AJ, Wang J, Shirley NJ, Jones GP, Stone BA, Hø PB, Fincher GB. 1996. Barley beta-D-glucan exohydrolases with beta-D-glucosidase activity. Purification, characterization, and determination of primary structure from a cDNA clone. **The Journal of Biological Chemistry** 271 (9), 5277-5286.
- Hu YX, Wang YX, Liu XF, Li JY. 2004. *Arabidopsis* RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development. **Cell Research** 14 (1), 8–15.
- Huang J, Wang JF, Zhang HS. 2004. Structure and function of plant C2H2 zinc finger protein. **Yi Chuan** 26 (3), 414-418.
- Hussey SG, Mizrahi E, Spokevicius AV, Bossinger G, Berger DK, Myburg AA. 2011. SND2, a NAC transcription factor gene, regulates genes involved in secondary cell wall development in *Arabidopsis* fibres and increases fibre cell area in *Eucalyptus*. **BMC Plant Biology** 11, 173.
- Huysmans M, Buono RA, Skorzinski N, Radio MC, De Winter F, Parizot B, Mertens J, Karimi M, Fendrych M, Nowack MK. 2018. NAC Transcription Factors ANAC087 and ANAC46 Control Distinct Aspects of Programmed Cell Death in the *Arabidopsis* Columella and Lateral Root Cap. **The Plant Cell** 30 (9) 2197-2213.
- Iida K, Seki M, Sakurai T, Satou M, Akiyama K, Toyoda T, Konagaya A, Shinozaki K. 2005. RARTF: database and tools for complete sets of *Arabidopsis* transcription factors. **DNA Research: an international journal for rapid publication of reports on genes and genomes** 12 (4), 247–256.

- Ikeda M, Ohme-Takagi M. 2009. A novel group of transcriptional repressors in *Arabidopsis*. **Plant and Cell Physiology** 50 (5), 970–975.
- Ishii, S. 1984. Cell wall cementing materials of grass leaves. **Plant Physiology** 76 (4), 959–961.
- Jackson MB, Armstrong W. 1999. Formation of aerenchyma and the processes of plant ventilation in relation to soil flooding and submergence. **Plant Biology** 1 (3), 274–287.
- Jackson MB, Fenning TM, Jenkins W. 1985. Aerenchyma (gas-space) formation in adventitious roots of rice (*Oryza sativa* L.) is not controlled by ethylene or small partial pressures of oxygen. **Journal of Experimental Botany** 36,1566–1572.
- Jacquet G, Pollet B, Lapierre C. 1995. New ether-linked ferulic acid-coniferyl alcohol dimers identified in grass straws. **Journal of Agricultural and Food Chemistry** 43 (10), 2746–2751.
- Jaglo-Ottosen KR, Gilmour SJ, Zarka DJ, Schabenberger O, Thomashow MF. 1998. *Arabidopsis* CBF1 overexpression induces COR genes and enhance freezing tolerance. **Science** 280 (5360), 104-106.
- Jaiswal D, De Souza AP, Larsen S, LeBauer D, Miguez FE, Sparovek G, Bollero G, Buckeridge MS, Long SP. 2017. Brazilian sugarcane ethanol as an expandable green alternative to crude oil use. **Nature Climate Change** 7, 788-792.
- Jakoby M, Weisshaar B, Dröge-Laser W, Vicente- Carbajosa J, Tiedemann J, Kroj T, Parcy F, bZIP Research Group. 2002. bZIP transcription factors in *Arabidopsis*. **Trends in Plant Science** 7 (3), 106-111.
- Jalili M, Salehzadeh-Yazdi A, Shailendra G, Wolkenhauer O, Yaghmaie M, Resendis-Antonio O, Alimoghaddam K. 2016. Evolution of Centrality Measurements for the Detection of Essential Proteins in Biological Networks. **Frontiers in Physiology** 7, 375.

- Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, D'Hont A. 2007. Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. **The Plant Journal** 50 (4), 574–585.
- Jarvis MC, Briggs SP, Knox JP. 2003. Intercellular adhesion and cell separation in plants. **Plant, Cell & Environment** 26 (7), 977–989.
- Jenkins ES, Paul W, Craze M, Whitelaw CA, Weigand A, Roberts JA. 1999. Dehiscence-related expression of an *Arabidopsis thaliana* gene encoding a polygalacturonase in transgenic plants of *Brassica napus*. **Plant, Cell & Environment** 22, 159–167.
- Jeong JS, Kim YS, Redillas MC, Jang G, Jung H, Bang SW, Choi YD, Ha SH, Reuzeau C, Kim JK. 2013. OsNAC5 overexpression enlarges root diameter in rice plants leading to enhanced drought tolerance and increased grain yield in the field. **Plant Biotechnology Journal** 11 (1), 101-114.
- Jin H, Cominelli E, Bailey P, Parr A, Mehrtens F, Jones J, Tonelli C, Weisshaar B, Martin C. 2000. Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. **The EMBO Journal** 19 (22), 6150–6161.
- Jin JP, Tian F, Yang DC, Meng YQ, Kong L, Luo JC, Gao G. 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. **Nucleic Acids Research** 45 (D1), D1040-D1045.
- Jofuku KD, den Boer BGW, Van Montagu M, Okamura JK. 1994. Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. **The Plant Cell** 6 (9), 1211-1225.
- Justin SHFW, Armstrong W. 1987. The anatomical characteristics of roots and plant response to soil flooding. **The New Phytologist** 106 (3), 465–495.

- Justin SHF, Armstrong W. 1991. A reassessment of the influence of NAA on aerenchyma formation in maize roots. **New Phytologist** 117(4), 607-618.
- Kagaya Y, Ohmiya K, Hattori T. 1999. RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. **Nucleic Acids Research** 27 (2), 470–478.
- Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. **Molecular Biology and Evolution** 30 (4), 772-780.
- Kaufmann K, Pajoro A, Angenent GC. 2010. Regulation of transcription in plants: mechanisms controlling developmental switches. **Nature Reviews Genetics** 11 (12), 830–842.
- Keller T, Damude HG, Werner D, Doemer P, Dixon RA, Lamb C. 1998. A plant homolog of the neutrophil NADPH oxidase gp91phox subunit gene encodes a plasma membrane protein with Ca²⁺ binding motifs. **The Plant Cell** 10 (2), 255-286.
- Kellogg EA. 2013. **Phylogenetic Relationships of Saccharinae and Sorghinae. Genomics of the Saccharinae**. Paterson, AH (Eds.). New York, Berlin: Springer. pp. 3-22.
- Kieffer M, Stern Y, Cook H, Clerici E, Maulbetsch C, Laux T, Davies B. 2006. Analysis of the transcription factor WUSCHEL and its functional homologue in *Antirrhinum* reveals a potential mechanism for their roles in meristem maintenance. **The Plant Cell** 18 (3), 560–573.
- Kim HJ, Hong SH, Kim YW, Lee IH, Jun JH, Phee BK, Rupak T, Jeong H, Lee Y, Hong BS, Nam HG, Woo HR, Lim PO. 2014. Gene regulatory cascade of senescence-associated NAC transcription factors activated by ETHYLENE-INSENSITIVE2-mediated leaf senescence signalling in *Arabidopsis*. **Journal of Experimental Botany** 65 (14), 4023–4036.

- Kim JB, Olek AT, Carpita NC. 2000. Plasma membrane and cell wall exo- β -D-glucanases in developing maize coleoptiles. **Plant Physiology** 123 (2), 471–485.
- Knox JP. 1992. Cell adhesion, cell separation and plant morphogenesis. **The Plant Journal** 2 (2), 137–141.
- Konishi M, Yanagisawa S. 2008. Ethylene signaling in *Arabidopsis* involves feedback regulation via the elaborate control of EBF2 expression by EIN3. **The Plant Journal** 55 (5), 821–831.
- Kosugi S, Hasebe M, Tomita M, Yanagawa H. 2009. Systematic identification of yeast cell cycle-dependent nucleocytoplasmic shuttling proteins by prediction of composite motifs. **Proceedings of the National Academy of Sciences of the United States of America** 106 (25), 10171–10176.
- Kozela C, Regan S. 2003. How plants make tubes. **Trends in Plant Science** 8(4), 159–164.
- Kranthi T, Rao SB, Manimaran P. 2016. Prediction of essential genes in human using graph centrality measures and a machine learning approach. **Indian Journal of Discrete Mathematics** 1 (2), 139–150.
- Kummerfeld SK, Teichmann SA. 2006. DBD: a transcription factor prediction database. **Nucleic Acids Research** 34, D74–D81.
- Labeeuw L, Martone PT, Boucher Y, Case RJ. 2015. Ancient origin of the biosynthesis of lignin precursors. **Biology Direct** 10 (23), 1–21.
- Lam E. 2004. Controlled cell death, plant survival and development. **Nature Reviews Molecular Cell Biology** 5 (4), 305–315.
- Latarullo MBG, Tavares EQP, Maldonado GP, Leite DCC, Buckeridge MS. 2016. Pectins, Endopolygalacturonases, and Bioenergy. **Frontiers in Plant Science** 7 (1401), 1–7.

- Le Cunff L, Garsmeur O, Raboin LM, Pauquet J, Telismart H, Selvi A, Grivet L, Philippe R, Begum D, Deu M, Costet L, Wing R, Glaszmann JC, D'Hont A. 2008. Diploid / polyploid syntenic shuttle mapping and haplotype-specific chromosome walking toward a rust resistance gene (Bru1) in highly polyploid sugarcane (2n _ 12x _ 115). **Genetics** 180 (1), 649–660.
- Lee BH, Kwon SH, Lee SJ, Park SK, Song JT, Lee S, Lee MM, Hwang YS, Kim JH. 2015 The *Arabidopsis thaliana* NGATHA transcription factors negatively regulate cell proliferation of lateral organs. **Plant Molecular Biology** 89 (4-5), 529-538.
- Lee D, Polisensky DH, Braam J. 2005. Genome-wide identification of touch- and darkness-regulated *Arabidopsis* genes: a focus on calmodulin-like and XTH genes. **The New Phytologist** 165 (2), 429–444.
- Lehman A, Black R, Ecker JR. 1996. HOOKLESS1, an ethylene response gene, is required for differential cell elongation in the *Arabidopsis* hypocotyl. **Cell** 85 (2), 183–194.
- Leite DCC. 2013. Modificações da parede celular durante a formação de aerênquima em raízes de cana-de-açúcar. **Dissertação (Mestrado em Ciências)**. Instituto de Biociências, Universidade de São Paulo, São Paulo.
- Leite DCC, Grandis A, Tavares EQP, Piovezani AR, Pattathil S, Avci U, Rossini A, Cambler A, De Souza AP, Hahn MG, Buckeridge MS. 2017. Cell wall changes during the formation of aerenchyma in sugarcane roots. **Annals of Botany** 120 (5), 693-708.
- Li CW, SU RC, Cheng CP, Sanjava, You SJ, Hsieh TH, Chao TC, Chan MT. 2011. Tomato RAV transcription factor is a pivotal modulator involved in the AP2/EREBP – mediated defense pathway. **Plant Physiology** 156 (1), 213-227.
- Li H, Johnson P, Stepanova A, Alonso JM, Ecker JR. 2004. Convergence of signaling pathways in the control of differential cell growth in *Arabidopsis*. **Developmental Cell** 7(2),193–204.

- Li M, Zheng R, Zhang H, Wang J, Pan Y. 2014a. Effective identification of essential proteins based on priori knowledge, network topology and gene expressions. **Methods** 67 (1), 325–333.
- Li S, Bashline L, Lei L, Gu Y. 2014b. Cellulose Synthesis and Its Regulation. **The Arabidopsis Book** 12, e0169.
- Li SB, Xie ZZ, Hu CG, Zhang JZ. 2016. A Review of Auxin Response Factors (ARFs) in Plants. **Frontiers in Plant Science** 7 (47), 1-7.
- Li WY, Chen BX, Chen ZJ, Gao YT, Chen Z, Liu J. 2017. Reactive Oxygen Species Generated by NADPH Oxidases Promote Radicle Protrusion and Root Elongation during Rice Seed Germination. **International Journal of Molecular Science** 18 (1), E110.
- Liang C, Wang Y, Zhu Y, Tang J, Hu B, Liu L, Ou S, Wu H, Sun X, Chu J, Chu C. 2014. OsNAP connects abscisic acid and leaf senescence by fine-tuning abscisic acid biosynthesis and directly targeting senescence-associated genes in rice. **Proceedings of the National Academy of Sciences of the United States of America** 111 (27) 10013–10018.
- Libertini E, Li Y, McQueen-Mason S. 2004. Phylogenetic analysis of the plant endo-beta-1,4-glucanase gene family. **Journal of Molecular Evolution** 58 (5), 506–515.
- Lim PO, Lee IC, Kim J, Kim HJ, Ryu JS, Woo HR, Nam HG. 2010. Auxin response factor 2 (ARF2) plays a major role in regulating auxin-mediated leaf longevity. **Journal Experimental Botany** 61 (5), 1419-1430.
- Liu YZ, Baig MNR, Fan R, Ye JL, Cao CY, Deng XX. 2009. Identification and expression pattern of a novel NAM, ATAF, and CUC-like gene from *Citrus sinensis* Osbeck. **Plant Molecular Biology Reporter** 27, 292–297.

- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. **Nucleic Acids Research** 42 (Database issue), D490-D495.
- Long JA, Ohno C, Smith ZR, Meyerowitz EM. 2006. TOPLESS regulates apical embryonic fate in *Arabidopsis*. **Science** 312 (5779), 1520–1523.
- Lopez-Casado G, Urbanowicz BR, Damasceno CM, Rose JK. 2008. Plant glycosyl hydrolases and biofuels: a natural marriage. **Current Opinion in Plant Biology** 11(3), 329-337.
- Macedo IC. 2005. **Sugar Cane's Energy – Twelve studies on Brazilian sugar cane agribusiness and its sustainability**. UNICA – União da Agroindústria Canavieira do Estado de São Paulo. Berlendis & Vertecchia, São Paulo. pp.1-233.
- Magnani E, Sjölander K, Hake S. 2004. From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. **The Plant Cell** 16 (4), 2265–2277.
- Manners JM, Casu RE. 2011. Transcriptome Analysis and Functional Genomics of Sugarcane. **Tropical Plant Biology** 4 (1),9-21.
- Mano Y, Omori F. 2013. Relationship between constitutive root aerenchyma formation and flooding tolerance in *Zea nicaraguensis*. **Plant and Soil** 370 (1-2), 447-460.
- Marmorstein R. 2003. Structure of SET domain proteins: a new twist on histone methylation. **Trends in Biochemical Sciences** 28 (2), 59-62.
- Martel C, Giovannoni J. 2007. Fruit ripening. In: Roberts JA, Gonzalez-Carranza Z (Ed.). **Plant Cell Separation and Adhesion**. 1st ed. UK: Blackwell Publishing Ltd. pp. 164-182.
- Matías-Hernández L, Aguilar-Jaramillo AE, Marín-González E, Suárez-López P, Pelaz S. 2014. RAV genes: regulation of floral induction and beyond. **Annals of Botany** 114 (7), 1459–1470.

- Matsukura C, Kawai M, Toyofuku K, Barrero RA, Uchimiya H, Yamaguchi J. 2000. Transverse vein differentiation associated with gas space formation—fate of the middle cell layer in leaf sheath development of rice. **Annals of Botany** 85 (1), 19–27.
- Mattiello L, Riaño-Pachón DM, Martins MC, da Cruz LP, Bassi D, Marchiori PE, Ribeiro RV, Labate MT, Labate CA, Menossi M. 2015. Physiological and transcriptional analyses of developmental stages along sugarcane leaf. **BMC Plant Biology** 15, 300.
- McCarty DR, Carson CB, Stinard PS, Robertson DS. 1989. Molecular analysis of viviparous-1: An abscisic acid-insensitive mutant of maize. **The Plant Cell** 1 (5), 523–532.
- McCarty DR, Hattori T, Carson CB, Vasil V, Lazar M, Vasil IK. 1991. The Viviparous-1 developmental gene of maize encodes a novel transcriptional activator. **Cell** 66 (5), 895-905.
- McGrath CL, Lynch M. 2012. Evolutionary significance of whole-genome duplication. In: Soltis PS, Soltis DE. (Ed.). **Polyploidy and genome evolution**. Berlin: Springer. pp. 1–20.
- Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M. 1998. *Arabidopsis thaliana*: a model plant for genome analysis. **Science** 282 (5389), 662-682.
- Mendes GC, Reis PA, Calil IP, Carvalho HH, Aragão FJ, Fontes EP. 2013. GmNAC30 and GmNAC81 integrate the endoplasmic reticulum stress- and osmotic stress-induced cell death responses through a vacuolar processing enzyme. **Proceedings of the National Academy of Sciences of the United States of America** 110 (48), 19627–19632.
- Miller JR, Dilley KA, Harkins DM, Torralba MG, Moncera KJ, Beerli K, Goglin K, Stockwell TB, Sutton GG, Shabman RS. 2017. Initial genome sequencing of the sugarcane CP 96-1252 complex hybrid. **F1000Research** 1, 6: 688.
- Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrels ME, Irvine JE, Paterson AH. 1998. Detailed alignment of

- Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. **Genetics** 150 (4), 1663–1682
- Minh BQ, Nguyen MAT, Von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. **Molecular Biology and Evolution** 30 (5), 1188–1195.
- Mittler R, Vanderauwera S, Gollery M, Breusegem FV. 2004. Reactive oxygen gene network of plants. **Trends in Plant Science** 9 (10), 490–498.
- Mittler R, Vanderauwera S, Suzuki N, Miller G, Tognetti VB, Vandepoele K, Gollery M, Shulaev V, Van Breusegem F. 2011. ROS signaling: the new wave?. **Trends in Plant Science** 16 (6), 300–309.
- Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. 2011. AP2/ERF family transcription factors in plant abiotic stress responses. **Biochimica et Biophysica Acta** 1819 (2), 86–96.
- Mosammaparast N, Shi Y. 2010. Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. **Annual Review of Biochemistry** 79, 155–179.
- Mosier N, Wyman C, Dale B, Elander R, Lee YY, Holtzapple M, Ladisch M. 2005. Features of promising technologies for pretreatment of lignocellulosic biomass. **Bioresource Technology** 96 (6), 673–686.
- Movahedi S, Van Bel M, Heyndrickx KS, Vandepoele K. 2012. Comparative co-expression analysis in plant biology. **Plant, Cell & Environment** 35 (10), 1787–1798.
- Mühlenbock P, Plaszczyca M, Plaszczyca M, Mellerowicz E, Karpinski S. 2007. Lysigenous aerenchyma formation in *Arabidopsis* is controlled by LESION SIMULATING DISEASE1. **The Plant Cell** 19(11), 3819–3830.
- Murtagh F, Legendre P. 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?. **Journal of Classification** 31 (3), 274–295.

- Mutwil M, Ruprecht C, Giorgi FM, Bringmann M, Usadel B, Persson S. 2009. Transcriptional wiring of cell wall-related genes in *Arabidopsis*. **Molecular Plant** 2 (5), 1015-1024.
- Nakano T, Suzuki K, Fujimura T, Shinshi H. 2006. Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. **Plant Physiology** 140 (2), 411–432.
- Nakano Y, Yamaguchi M, Endo H, Rejab NA, Ohtani M. 2015. NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. **Frontiers in Plant Science** 6 (288), 1-18.
- Ogita N, Okushima Y, Tokizawa M, Yamamoto YY, Tanaka M, Seiki M, Makita Y, Matsui M, Okamoto YK, Sakamoto T, Kurata T, Hiruma K, Saijo Y, Takahashi N, Umeda M. 2018. Identifying the target genes of SUPPRESSOR OF GAMMA RESPONSE 1, a master transcription factor controlling DNA damage response in *Arabidopsis*. **The Plant Journal** 94 (3), 439-453.
- Ohme-Takagi M, Shinshi H. 1995. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. **The Plant Cell** 7 (2), 173-82.
- Okamuro JK, Caster B, Villarroel R, Van Montagu M, Jofuku KD. 1997. The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in *Arabidopsis*. **Proceedings of the National Academy of Sciences of the United States of America** 94 (13), 7076–7081.
- Okura VK, de Souza RS, de Siqueira Tada SF, Arruda P. 2016. BAC-Pool Sequencing and Assembly of 19 Mb of the Complex Sugarcane Genome. **Frontiers in Plant Science** 7 (342), 1-8.
- Okushima Y, Mitina I, Quach HL, Theologis A. 2005. AUXIN RESPONSE FACTOR (ARF2): a pleiotropic developmental regulator. **The Plant Journal** 43 (1), 29-46.

- Olmedo G, Guo H, Gregory BD, Nourizadeh SD, Aguilar- Henonin L, Li H, An F, Guzman P, Ecker JR. 2006. ETHYLENE-INSENSITIVE5 encodes a 5'→3' exoribonuclease required for regulation of the EIN3-targeting F-box proteins EBF1/2. **Proceedings of the National Academy of Sciences of the United States of America** 103 (36),13286-13293.
- Pakarinen A, Zhang J, Brock T, Maijala P, Vikari L. 2012. Enzymatic accessibility of fiber hemp is enhanced by enzymatic or chemical removal of pectin. **Bioresource Technology** 107, 275-281.
- Park YB, Cosgrove DJ. 2012. A revised architecture of primary cell walls based on biomechanical changes induced by substrate-specific endoglucanases. **Plant Physiology** 158 (4), 1933–1943.
- Parlanti S, Kudahettige NP, Lombardi L, Mensuali-Sodi A, Alpi A, Perata P, Pucciariello C. 2011. Distinct mechanisms for aerenchyma formation in leaf sheaths of rice genotypes displaying a quiescence or escape strategy for flooding tolerance. **Annals of Botany** 107 (8), 1335–1343.
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash T, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS. 2009. The *Sorghum bicolor* genome and the diversification of grasses. **Nature** 457 (7229), 551-556.

- Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. **Proceedings of the National Academy of Sciences of the United States of America** 101 (26), 9903-9908.
- Pavlopoulos GA, Secrier M, Moschopoulos C, Soldatos TG, Kossida S, Aerts J, Schneider R, Bagos PG. 2011. Using Graph theory to analyze biological networks. **BioData Mining** 4, 10.
- Peaucelle A, Braybrook S, Höfte H. 2012. Cell wall mechanics and growth control in plants: the role of pectins revisited. **Frontiers in Plant Science** 3, 121.
- Philippe F, Pelloux J, Rayon C. 2017. Plant pectin acetyltransferase structure and function: new insights from bioinformatic analysis. **BMC Genomics** 18 (1), 456.
- Piovezani AR. 2017. SIT: uma ferramenta para integração e visualização de dados em larga escala e sua aplicação em cana-de-açúcar. **Tese (Doutorado)**. Programa Interunidades de Pós-Graduação em Bioinformática. 149 p.
- Podzimska-Sroka D, O' Shea C, Gregersen PL, Skriver K. 2015. NAC transcription factors in Senescence: From Molecular Structure to Function in Crops. **Plants (Basel)** 4 (3), 412-448.
- Polizeli MLTM, Corrêa ECP, Polizeli AM, Jorge JA. 2011. Hydrolases from Microorganisms used for Degradation of Plant Cell Wall and Bioenergy. In: Buckeridge M, Goldman G. (Ed). **Routes to Cellulosic Ethanol**. 1st ed. New York: Springer. pp.115-134.
- Potuschak T, Lechner E, Parmentier Y, Yanagisawa S, Grava S, Koncz C, Genschik P. 2003. EIN3-dependent regulation of plant ethylene hormone signaling by two *Arabidopsis* F box proteins: EBF1 and EBF2. **Cell** 115 (6), 679-689.

- Potuschak T, Vansiri A, Binder BM, Lechner E, Vierstra RD, Genschik P. 2006. The exoribonuclease XRN4 is a component of the ethylene response pathway in *Arabidopsis*. **The Plant Cell** 18 (11), 3047-3057.
- Promkhambut A, Polthanee A, Akkasaeng C, Younger A. 2011. Growth, yield and aerenchyma formation of sweet and multipurpose sorghum (*Sorghum bicolor* L. Moench) as affected by flooding at different growth stages. **Australian Journal of Crop Science** 5 (8), 954-965.
- Puranik S, Sahu PP, Srivastava OS, Prasad M. 2012. NAC proteins: regulation and role in stress tolerance. **Trends in Plant Science** 17 (6), 369–381.
- Qiao H, Chang KN, Yazaki J, Ecker JR. 2009. Interplay between ethylene, ETP1/ETP2 F-box proteins, and degradation of EIN2 triggers ethylene responses in *Arabidopsis*. **Genes & Development** 23 (4), 512-521.
- Rajhi I, Yamauchi T, Takahashi H, Nishiuchi S, Shiono K, Watanabe R, Mliki A, Nagamura Y, Tsutsumi N, Nishizawa NK, Nakazono M. 2011. Identification of genes expressed in maize root cortical cells during lysigenous aerenchyma formation using laser microdissection and microarray analyses. **The New Phytologist** 190 (2), 351-368.
- Rambaut A. **FigTree version 1.4.3.** Disponível em <<http://tree.bio.ed.ac.uk/software/figtree/>>. Acesso em 01 Aug. 2016.
- Raven JA. 1996. Into the voids: the distribution, function, development and maintenance of gas spaces in plants. **Annals of Botany** 78, 137–142.
- Redillas MC, Jeong JS, Kim YS, Jung H, Bang SW, Choi YD, Ha SH, Reuzeau C, Kim JK. 2012. The overexpression of OsNAC9 alters the root architecture of rice plants enhancing drought resistance and grain yield under field conditions. **Plant Biotechnology Journal** 10 (7), 792-805.

- Renewable Fuels Association. **Industry Statistics.** Disponível em <<https://ethanolrfa.org/resources/industry/statistics/>>. Acesso em 13 Aug. 2018.
- Rhee S, Mutwil M. 2014. Towards revealing the functions of all genes in plants. **Trends in Plant Science** 19 (4), 212-221.
- Riaño-Pachón DM, Mattiello L. 2017. Draft genome sequencing of the sugarcane hybrid SP80-3280. **F1000Research** 2, 6-861.
- Riaño-Pachón DM, Ruzicic S, Dreyer I, Mueller-Roeber B. 2007. PlnTFDB: an integrative plant transcription factor database. **BMC Bioinformatics** 8, 42.
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G. 2000. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. **Science** 290 (5499), 2105–2110.
- Riechmann JL, Meyerowitz EM. 1998. The AP2/EREBP family of plant transcription factors. **Biological Chemistry** 379 (6), 633–646.
- Riechmann JL, Ratcliffe OJ. 2000. A genomic perspective on plant transcription factors. **Current Opinion in Plant Biology** 3 (5), 423–434.
- 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. **Frontiers in Plant Science** 2 (23), 1-13.
- Ruprecht C, Persson S. 2012. Co-expression of cell-wall related genes: new tools and insights. **Frontiers in Plant Science** 3 (83), 1-7.
- Rushton PJ, Somssich IE, Ringler P, Shen QJ. 2010. WRKY transcription factors. **Trends in Plant Science** 15 (5), 247-258.

- Saab IN, Sachs MM. 1996. A Flooding-Induced Xyloglucan Endo-Transglycosylase Homolog in Maize is Responsive to Ethylene and Associated with Aerenchyma. **Plant Physiology** 112, 385-391.
- Saballos A. 2008. Development and utilization of sorghum as a bioenergy crop. In: Vermerris W (Ed.). **Genetic Improvement of Bioenergy Crops..** New York: Springer. pp. 211-248.
- Sachs JA. 1882. **A text book of botany.** Oxford, UK: Oxford University Press.
- Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K. 2002. DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. **Biochemical and Biophysical Research Communications** 290 (3), 998–1009.
- Sampedro J, Valdivia ER, Fraga P, Iglesias N, Revilla G, Zarra I. 2017. Soluble and Membrane-Bound β -Glucosidases Are Involved in Trimming the Xyloglucan Backbone. **Plant Physiology** 173 (2), 1017-1030.
- Sander L, Child R, Ulvskov P, Albrechtsen M, Borkhardt, B. 2001. Analysis of a dehiscence zone endo-polygalacturonase in oilseed rape (*Brassica napus*) and *Arabidopsis thaliana*: evidence for roles in cell separation in dehiscence and abscission zones, and in stylar tissues during pollen tube growth. **Plant Molecular Biology** 46 (4), 469–479.
- Schruff MC, Spielman M, Tiwari S, Adams S, Fenby N, Scott RJ. 2006. The AUXIN RESPONSE FACTOR 2 gene of *Arabidopsis* links auxin signaling, cell division, and the size of seeds and other organs. **Development** 133 (2), 251-261.
- Schussler EE, Longstreth DJ. 1996. Aerenchyma develops by cell lysis in roots and cell separation in leaf petioles in *Sagittaria lancifolia* (Alismataceae). **American Journal of Botany** 83 (10), 1266–1273.

- Sebastian J, Lee JY. 2013. Root Apical Meristems. In: **Encyclopedia of Life Science**. Chichester: John Wiley & Sons, Ltd. pp. 1-9.
- Serin EA, Nijveen H, Hilhorst HW, Ligterink W. 2016. Learning from Co-expression Networks: Possibilities and Challenges. **Frontiers in Plant Science** 7, 444.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. **Genome Research** 13 (11), 2498-2504.
- Silva AM. 2005. Caracterização da parede celular de *Saccharum officinarum* L. (cana-de-açúcar) e *Brachiaria decumbens* Stapf (braquiária). **Tese (Doutorado)**. Instituto de Biologia. Universidade Estadual de Campinas.105p.
- Silva AT, Ribone PA, Chan RL, Ligterink W, Hilhorst HW. 2016. A predictive co-expression network identifies novel genes controlling the seed-to-seedling phase transition in *Arabidopsis thaliana*. **Plant Physiology** 170 (4), 2218–2231.
- Sindhu R, Binod P, Pandey A. 2016. Biological pretreatment of lignocellulosic biomass – An overview. **Bioresource Technology** 199, 76-82.
- Smirnoff N, Crawford RMM. 1983. Variation in the structure and response to flooding of root aerenchyma in some wetland plants. **Annals of Botany** 51 (2), 237–249.
- Socol CR, Vanderberghe LPS, Medeiros ABP, Karp SG, Buckeridge M, Ramos LP, Pitarelo AP, Ferreira-Leitão V, Gottschalk LMF, Ferrara MA, Bon EPS, Moraes LMP, Araújo JA, Torres FAG. 2010. Bioethanol from lignocelluloses: Status and perspectives in Brazil. **Bioresource Technology** 101 (13), 4820-4825.
- Sohn KH, Lee SC, Jung HW, Hong JK, Hwang BK. 2006. Expression and functional roles of the pepper pathogen-induced transcription factor RAV1 in bacterial disease resistance, and drought and salt stress tolerance. **Plant Molecular Biology** 61 (6), 897–915.

- Solano R, Stepanova A, Chao Q, Ecker J. 1998. Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. **Genes & Development** 12 (23), 3703-3714.
- Steffens B, Geske T, Sauter M. 2011. Aerenchyma formation in the rice stem and its promotion by H₂O₂. **The New Phytologist** 190 (2), 369–378.
- Stepanova AN, Alonso JM. 2009. Ethylene signaling and response: where different regulatory modules meet. **Current Opinion in Plant Biology** 12 (5), 548-555.
- Supratim C. 2014. Fundamentals of Molecular Evolution. In: Supratim C. (Ed.). **Bioinformatics for Beginners – Genes, Genomes, Molecular Evolution, Databases and Analytical Tools**. Elsevier. pp.27-53.
- Swaminathan K, Peterson K, Jack T. 2008. The plant B3 superfamily. **Trends in Plant Science** 13(12), 647- 655.
- Tabuchi A, Li LC, Cosgrove DJ. 2011. Matrix solubilization and cell wall weakening by beta-expansin (group-1 allergen) from maize pollen. **The Plant Journal** 68 (3), 546–559.
- Taiz L, Zeiger E, Møller IM, Murphy A. 2017. Arquitetura da Célula e do Vegetal. In: Taiz L, Zeiger E, Møller IM, Murphy A. (Ed.). **Fisiologia e Desenvolvimento vegetal**. 6th ed. Porto Alegre: Artmed. pp. 1-49.
- Takahashi H, Yamauchi T, Colmer TD, Nakazono M. 2014. Aerenchyma Formation in Plants. In: van Dongen J, Licausi F. (Ed.). **Low-Oxygen Stress in Plants: Oxygen Sensing and Adaptive Responses to Hypoxia**. 1st ed. Vienna: Springer. pp. 247-265.
- Talbott LD, Ray PM. 1992. Molecular size and separability features of pea cell wall polysaccharides: implications for models of primary wall structure. **Plant Physiology** 98 (1), 357–368.

- Tavares EQP. 2015. Regulação da degradação da parede celular durante a formação do aerênquima em raízes de cana-de-açúcar. **Tese (Doutorado)**. Instituto de Biociências da Universidade de São Paulo. Departamento de Botânica. 153p.
- Tavares EQP, Buckeridge MS. 2015. Do plant cell walls have a code?. **Plant Science** 241, 286–294.
- Tavares EQP, De Souza AP, Buckeridge MS. 2015. How endogenous plant cell-wall degradation mechanisms can help achieve higher efficiency in saccharification of biomass. **Journal of Experimental Botany** 66 (14), 4133–4143.
- Tavares EQP, De Souza AP, Romim GH, Grandis A, Plasencia A, Gaiarsa JW, Grima-Pettenati J, de Setta N, Van Sluys MA, Buckeridge MS. 2019. The control of endopolygalacturonase expression by the sugarcane RAV transcription factor during aerenchyma formation. **Journal of Experimental Botany** 70 (2), 497-506.
- Tavares EQP, Grandis A, Lembke CG, Souza GM, Purgatto E, De Souza AP, Buckeridge MS. 2018. Roles of auxin and ethylene in aerenchyma formation in sugarcane roots. **Plant Signaling & Behavior** 13(3), e1422464-3.
- Tetsushi H, Karim MA. 2007. Flooding Tolerance of Sugarcane in Relation to Growth, Physiology and Root Structure. **South Pacific Studies** 28(1), 9-22.
- The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. **Nucleic Acids Research** 45 (D1), D158-D169.
- Tirosh I, Bilu Y, Barkai N. 2007. Comparative biology: beyond sequence analysis. **Current Opinion in Biotechnology** 18 (4), 371–377.
- Tiwari SB, Hagen G, Guilfoyle T. 2003. The roles of auxin response factor domains in auxin-responsive transcription. **The Plant Cell** 15 (2), 533–543.

- Torres MA, Dongl JL. 2005. Functions of the respiratory burst oxidase in biotic interactions, abiotic stress and development. **Current Opinion in Plant Biology** 8 (4), 397-403.
- Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. **Nucleic Acids Research** 44 (W1), W232-W235.
- Trought MCT, Drew MC. 1980. The development of waterlogging damage in young wheat plants in anaerobic solution cultures. **Journal of Experimental Botany** 31 (6), 1573–1585.
- Tsang DL, Edmond C, Harrington JL, Nühse TS. 2011. Cell Wall Integrity Controls Root Elongation via a General 1-Aminocyclopropane-1-Carboxylic Acid-Dependent, Ethylene-Independent Pathway. **Plant Physiology** 156 (2), 596 – 604.
- Tsuchisaka A, Theologis A. 2004. Unique and overlapping expression patterns among the *Arabidopsis* 1-amino-cyclopropane-1-carboxylate synthase gene family members. **Plant Physiology** 136 (2), 2982–3000.
- Tsukagoshi H. 2016. Control of root growth and development by reactive oxygen species. **Current Opinion in Plant Biology** 29, 57-63.
- Tsukagoshi H, Busch W, Benfey PN. 2010. Transcriptional regulation of ROS controls transition from proliferation to differentiation in the root. **Cell** 143 (4), 606–616.
- Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J. 2006. A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. **Science** 314 (5803), 1298–1301.
- Ulmasov T, Hagen G, Gulfoyle TJ. 1997. ARF1, a transcription factor that binds to auxin response elements. **Science** 276 (5320), 1865-1868.

- Ulmasov T, Hagen G, Guilfoyle TJ. 1999. Activation and repression of transcription by auxin-response factors. **Proceedings of the National Academy of Sciences of the United States of America** 96 (10), 5844–5849.
- União da Indústria da Cana-de-Açúcar. Disponível em <<http://www.unicadata.com.br/>>. Acesso em 3 jun. 2017.
- Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K. 2000. *Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. **Proceedings of the National Academy of Science of the United States of America** 97(21), 11632–11637.
- Urbanowicz BR, Bennett AB, Del Campillo E, Catalá C, Hayashi T, Henrissat B, Höfte H, McQueen-Mason SJ, Patterson SE, Shoseyov O, Teeri TT, Rose JK. 2007. Structural Organization and a Standardized Nomenclature for Plant Endo-1,4- β -Glucanases (Cellulases) of Glycosyl Hydrolase Family 9. **Plant Physiology** 144 (4), 1693-1696.
- Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D, Persson S, Provart NJ. 2009. Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. **Plant, Cell & Environment** 32 (12), 1633–1651.
- Van de Velde J, Heyndrickx KS, Vandepoele K. 2014. Inference of transcriptional networks in *Arabidopsis* through conserved noncoding sequence analysis. **The Plant Cell** 26 (7), 2729-2745.
- Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. 2010. Lignin Biosynthesis and Structure. **Plant Physiology** 153 (3), 895-905.
- Vanholme R, Morreel K, Darrah C, Oyarce P, Grabber JH, Ralph J, Boerjan W. 2012. Metabolic engineering of novel lignin in biomass crops. **New Phytologist** 196 (4), 978-1000.

- van Loon LC, Geraats BP, Linthorst HJ. 2006. Ethylene as a modulator of disease resistance in plants. **Trends in Plant Science** 11 (4), 184-191.
- Varghese JN, Hrmova M, Fincher GB. 1999. Three-dimensional structure of a barley beta-D-glucan exohydrolase, a family 3 glycosyl hydrolase. **Structure** 7 (2), 179-190.
- Verbančič J, Lunn JE, Stitt M, Persson S. 2018. Carbon Supply and the Regulation of the Cell Wall Synthesis. **Molecular Plant** 11 (1), 75-94.
- Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MIT, Henrique-Silva F, Giglioti EA, Lemos MVF, Coutinho LL, Nobrega MP, Carrer H, França SC, Bacci Júnior M, Goldman MHS, Gomes SL, Nunes LR, Camargo LEA, Siqueira WJ, Van Sluys MA, Thiemann OH, Kuramae EE, Santelli RV, Marino CL, Targon MLPN, Ferro JA, Silveira HCS, Marini DC, Lemos EGM, Monteiro-Vitorello CB, Tambor JHM, 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 MMA, Rosa Jr VE, Siviero F, Coscrato VE, Machado MA, Grivet L, Di Mauro SMZ, Nobrega FG, Menck CFM, Braga MDV, Telles GP, Cara FAA, Pedrosa G, Meidanis J, Arruda P. 2003. Analysis and Functional Annotation of an Expressed Sequence Tag Collection for Tropical Crop Sugarcane. **Genome Research** 13 (12), 2725-35.
- Waltner JK, Peterson FC, Lytle BL, Volkman BF. 2005. Structure of the B3 domain from *Arabidopsis thaliana* protein At1g16640. **Protein Science** 14(9), 2478–2483.
- Wang KLC, Li H, Ecker JR. 2002. Ethylene Biosynthesis and Signaling Networks. **The Plant Cell** 14 (1), S131-S151.
- Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B. 2009. gplots: Various R programming tools for plotting data. **R package version** 2 (4).

- Weigel D. 1995. The APETALA2 domain is related to a novel type of DNA binding protein. **The Plant Cell** 7 (4), 388-389.
- Whitmarsh AJ, Davis RJ. 2000. Regulation of transcription factor function by phosphorylation. **Cell and Molecular Life Sciences** 57(8-9) 1172-1183.
- Willats WG, McCartney L, Mackie W, Knox JP. 2001. Pectin: cell biology and prospects for functional analysis. **Plant Molecular Biology** 47 (1-2), 9-27.
- Wilson DB, Urbanowicz B. **Glycoside hydrolase family 9**. Disponível em <<http://www.cazypedia.org/>>. Acesso em 15 Aug. 2018.
- Wilson K, Long D, Swinburne J, Coupland G. 1996. A dissociation insertion causes a semidominant mutation that increases expression of TINY, an *Arabidopsis* gene related to APETALA2. **The Plant Cell** 8 (4), 659-671.
- Wolfe CJ, Kohane IS, Butte AJ. 2005. Systematic survey reveals general applicability of “guilt-by-association” within gene coexpression networks. **BMC Bioinformatics** 6, 227.
- Woo HR, Kim JH, Kim J, Kim J, Lee U, Song IJ, Kim JH, Lee HY, Nam HG, Lim PO. 2010. The RAV1 transcription factor positively regulates leaf senescence in *Arabidopsis*. **Journal of Experimental Botany** 61 (14), 3947–3957.
- Woo HR, Kim HJ, Nam HG, Lim PO. 2013. Plant leaf senescence and death – regulation by multiple layers of control and implications for aging in general. **Journal of Cell Science** 126 (Pt 21), 4823-4833.
- Xiao C, Somerville C, Anderson CT. 2014. POLYGALACTURONASE INVOLVED IN EXPANSION1 functions in cell elongation and flower development in *Arabidopsis*. **The Plant Cell** 26 (3), 1018–1035.
- Yamagami T, Tsuchisaka A, Yamada K, Haddon WF, Harden LA, Theologis A. 2003. Biochemical diversity among the 1-amino-cyclopropane-1-carboxylate synthase isozymes encoded

- by the *Arabidopsis* gene family. **The Journal of Biological Chemistry** 278 (49), 49102–49112.
- Yamaguchi M, Ohtani M, Mitsuda N, Kubo M, Ohme-Takagi M, Fukuda H, Demura T. 2010. VND-INTERACTING2, a NAC Domain Transcription Factor, Negatively Regulates Xylem Vessel Formation in *Arabidopsis*. **The Plant Cell** 22 (4),1249-1263.
- Yamasaki K. 2015. Structures, Functions, and Evolutionary Histories of DNA- Binding Domains of Plant-Specific Transcription Factors. In: Gonzalez DH (Ed.). **Plant Transcription Factors: Evolutionary, Structural and Functional Aspects**. 1 st ed. London: Academic Press. pp. 57- 72.
- Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seiki E, Matsuda T, Tomo Y, Hayami N, Terada T, Shirouzu M, Osanai T, Tanaka A, Seiki M, Shinozaki K, Yokoyama S. 2004. Solution structure of the B3 DNA binding domain of the *Arabidopsis* cold-responsive transcription factor RAV1. **The Plant Cell** 16 (12), 3448–3459.
- Yamasaki K, Kigawa T, Seki M, Shinozaki K, Yokoyama S. 2013. DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. **Trends in Plant Science** 18 (5), 267-276.
- Yamauchi T, Rajhi I, Nakazono M. 2011. Lysigenous aerenchyma formation in maize root is confined to cortical cells by regulation of genes related to generation and scavenging of reactive oxygen species. **Plant Signaling & Behavior** 6 (5), 759–761.
- Yamauchi T, Shimamura S, Nakazono M, Mochizuki T. 2013. Aerenchyma formation in crop species: A review. **Field Crops Research** 152, 8-16.
- Yamauchi T, Watanabe K, Fukazawa A, Mori H, Abe F, Kawaguchi K, Oyanagi A, Nakazono M. 2014. Ethylene and reactive oxygen species are involved in root aerenchyma formation

- and adaptation of wheat seedlings to oxygen-deficient conditions. **Journal of Experimental Botany** 65 (1), 261–273.
- Yamauchi T, Yoshioka M, Fukazawa A, Mori H, Nishizawa NK, Tsutsumi N, Yoshioka H, Nakazono M. 2017. An NADPH Oxidase RBOH Functions in Rice Roots during Lysigenous Aerenchyma Formation under Oxygen-Deficient Conditions. **The Plant Cell** 29 (4), 775-790.
- Yan J, Tong T, Li X, Chen Q, Dai M, Niu F, Yang M, Deyholos MK, Yang B, Jiang YQ. 2018. A Novel NAC-Type Transcription Factor, NAC87, from Oilseed Rape Modulates Reactive Oxygen Species Accumulation and Cell Death. **Plant & Cell Physiology** 59 (2), 290-303.
- Yang ZT, Wang MJ, Sun L, Lu SJ, Bi DL, Sun L, Song ZT, Zhang SS, Zhou SF, Liu JX. 2014. The membrane-associated transcription factor NAC089 controls ER-stress induced programmed cell death in plants. **PLoS Genetics** 10 (3), e1004243.
- Yi D, Alvim Kamel CL, Cools T, Vanderauwera S, Takahashi N, Okushima Y, Eekhout T, Yoshiyama KO, Larkin J, Van den Daele H, Conklin P, Britt A, Umeda M, De Veylder L. 2014. The *Arabidopsis* SIAMESE-RELATED cyclin-dependent kinase inhibitors SMR5 and SMR7 regulate the DNA damage checkpoint in response to reactive oxygen species. **The Plant Cell** 26 (1), 296-309.
- Yoo SD, Cho YH, Tena G, Xiong Y, Sheen J. 2008. Dual control of nuclear EIN3 by bifurcate MAPK cascades in C2H4 signalling. **Nature** 451 (7180), 789-795.
- Yoo YH, Choi HK, Jung KH. 2015. Genome-wide identification and analysis of genes associated with lysigenous aerenchyma formation in rice roots. **Journal of Plant Biology** 58 (2), 117–127.
- Yoon J, Choi H, An G. 2015. Roles of lignin biosynthesis and regulatory genes in plant development. **Journal of Integrative Plant Biology** 57 (11), 902-912.

- Yoshida K, Komae K. 2006. A rice family 9 glycoside hydrolase isozyme with broad substrate specificity for hemicelluloses in type II cell walls. **Plant & Cell Physiology** 47 (11), 1541-54.
- Yoshida M, Liu Y, Uchida S, Kawarada K, Ukagami Y, Ichinose H, Kaneko S, Fukuda K. 2008. Effects of cellulose crystallinity, hemicellulose, and lignin on the enzymatic hydrolysis of *Miscanthus sinensis* to monosaccharides. **Bioscience, Biotechnology, and Biochemistry** 72 (3), 805–810.
- Yoshiyama KO. 2015. SOG1: a master regulator of the DNA damage in plants. **Genes Genetic & Systems** 90 (4), 209-216.
- Young TE, Gallie DR. 2000. Programmed Cell Death during endosperm development. **Plant Molecular Biology** 44 (3), 283-301.
- Zhao L, Luo Q, Yang C, Han Y, Li W. 2008. A RAV-like transcription factor controls photosynthesis and senescence in soybean. **Planta** 227 (6), 1389–1399.
- Zhao Y, Chan Z, Gao J, Xing L, Cao M, Yu C, Hu Y, You J, Shi H, Zhu Y, Gong Y, Mu Z, Wang H, Deng X, Wang P, Bressan RA, Zhu JK. 2016. ABA receptor PYL9 promotes drought resistance and leaf senescence. **Proceedings of the National Academy of Sciences of the United States of America** 113(7), 1949-1954.
- Zhong J, Wang J, Peng W, Zhang Z, Pan Y. 2013. Prediction of essential proteins based on gene expression programming. **BMC Genomics** 14 (Suppl 4), S7.
- Zhong R, Richardson EA, Ye ZH. 2007. The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. **The Plant Cell** 19 (9), 2776–2792.
- Zhong R, Ye ZH. 2015. Secondary Cell Walls: Biosynthesis, Patterned Deposition and Transcriptional Regulation. **Plant & Cell Physiology** 56 (2), 195-214.

Zhu Z, Xu F, Zhang Y, Cheng YT, Wiermer M, Li X, Zhang Y. 2010. *Arabidopsis* resistance protein SNC1 activates immune responses through association with a transcriptional corepressor. **Proceedings of the National Academy of Sciences of the United States of America** 107 (31), 13960–13965.

APÊNDICE A. Relação de sequências de proteínas usadas para a árvore filogenética do fator de transcrição RAV, e seus respectivos domínios proteicos com o valor de E-value.

Identificação	Espécie	Banco de Dados	Domínio Proteico		E-value
			AP2	B3	
AT1G13260.1	<i>Arabidopsis thaliana</i>	Phytozome	60 - 109	188 - 292	1,4e-07/ 1,7e-29
AT1G25560.1	<i>Arabidopsis thaliana</i>	Phytozome	70-119	195-306	3,8e-06/ 8,4e-30
AT1G68840.1	<i>Arabidopsis thaliana</i>	Phytozome	63-112	188-293	2,9e-08/ 1,5e-29
AT3G25730.1	<i>Arabidopsis thaliana</i>	Phytozome	60-109	183-289	3,8e-08/ 3,1e-28
Bradi2g02710.1	<i>Brachypodium distachyon</i>	Phytozome	98 - 147	228 - 335	6,3e-06/ 2,4e-27
Bradi2g02720.1	<i>Brachypodium distachyon</i>	Phytozome	53-102	172-293	1,4e-06/ 1e-26
Bradi2g17610.1	<i>Brachypodium distachyon</i>	Phytozome	90 - 139	216 - 326	5,4e-08/ 3,9e-26
Bradi2g47220.1	<i>Brachypodium distachyon</i>	Phytozome	73-122	215-237	7,5e-08/ 3,3e-26
Brara.B01818.1	<i>Brassica rapa</i>	Phytozome	66 - 115	187 - 290	1,9e-06/ 2e-30
Brara.B03202.1	<i>Brassica rapa</i>	Phytozome	65 - 114	193 - 298	3,2e-08/ 3,2e-28
Brara.E01662.1	<i>Brassica rapa</i>	Phytozome	42 - 91	173 - 282	5,4e-07/ 1,7e-25
Brara.F00268.1	<i>Brassica rapa</i>	Phytozome	43 - 92	176 - 287	7,7e-07/ 5,4e-25
Brara.F00294.1	<i>Brassica rapa</i>	Phytozome	44 - 93	171 - 282	0,00033/ 1,1e-24
Brara.F00801.1	<i>Brassica rapa</i>	Phytozome	58 - 107	185 - 289	3,8e-07/ 6,8e-29
Brara.F03325.1	<i>Brassica rapa</i>	Phytozome	64 - 113	189 - 294	1,2e-08/ 9,7e-28
Brara.H00217.1	<i>Brassica rapa</i>	Phytozome	47 - 95	170 - 272	4,8e-06/ 9,8e-26
Brara.H02115.1	<i>Brassica rapa</i>	Phytozome	65 - 114	191 - 297	1,8e-06/ 1,4e-29
Brara.I03099.1	<i>Brassica rapa</i>	Phytozome	68 - 117	196 - 309	1,1e-06/ 1,1e-28
Brara.I05049.1	<i>Brassica rapa</i>	Phytozome	58 - 107	182 - 287	4,3e-07/ 3,6e-29
Eucgr.B03048.1	<i>Eucalyptus grandis</i>	Phytozome	65-114	197-303	8,5e-09/ 1e-28
Eucgr.C00358.1	<i>Eucalyptus grandis</i>	Phytozome	48 - 97	171 - 289	2,8e-07/ 1,8e-23
Eucgr.D00053.1	<i>Eucalyptus grandis</i>	Phytozome	63 - 112	185 - 291	3,9e-08/ 5,4e-24
Eucgr.L01673.1	<i>Eucalyptus grandis</i>	Phytozome	63 - 112	195 - 301	3,6e-08/ 3,2e-26
Glyma.01G087500.1	<i>Glycine max</i>	Phytozome	75-124	205-316	2,3e-07/ 2,5e-28
Glyma.02G099500.1	<i>Glycine max</i>	Phytozome	81-130	210-328	1,1e-08/ 2,1e-27
Glyma.10G204400.1	<i>Glycine max</i>	Phytozome	52 - 101	172 - 285	8,1e-09/ 9,8e-27
Glyma.20G186200.1	<i>Glycine max</i>	Phytozome	58 - 107	178 - 291	2,3e-08/ 1e-25
Gohir.A03G096800.1	<i>Gossypium hirsutum</i>	Phytozome	58 - 107	189 - 293	1,8e-08/ 2,6e-30
Gohir.A12G170100.1	<i>Gossypium hirsutum</i>	Phytozome	64 - 113	193 - 297	1e-08/ 1,9e-29
Gohir.A13G053300.1	<i>Gossypium hirsutum</i>	Phytozome	48 - 97	179 - 282	3,2e-08/ 1,3e-28
Gohir.A13G173000.1	<i>Gossypium hirsutum</i>	Phytozome	43 - 90	166 - 278	6,9e-06/ 9,8e-23
Gohir.A13G214600.1	<i>Gossypium hirsutum</i>	Phytozome	46 - 95	172 - 283	6,4e-05/ 8,7e-25
Gohir.D02G115900.1	<i>Gossypium hirsutum</i>	Phytozome	58 - 107	189 - 293	6,5e-09/ 3,8e-30
Gohir.D12G173400.1	<i>Gossypium hirsutum</i>	Phytozome	64 - 113	193 - 297	1e-08/ 2,1e-29
Gohir.D13G069800.1	<i>Gossypium hirsutum</i>	Phytozome	48 - 97	179 - 279	3e-08/ 5e-28
Gohir.D13G178800.1	<i>Gossypium hirsutum</i>	Phytozome	43 - 90	166 - 278	4,3e-05/ 6,6e-23
HORVU1Hr1G082670.1	<i>Hordeum vulgare</i>	Phytozome	7 - 56	138 - 248	2,6e-07/ 2e-26
HORVU1Hr1G082670.6	<i>Hordeum vulgare</i>	Phytozome	65 - 114	202 - 312	2,4e-08/ 2,1e-27
HORVU1Hr1G082670.7	<i>Hordeum vulgare</i>	Phytozome	71 - 120	190 - 300	2,2e-08/ 1,9e-27
HORVU3Hr1G010060.1	<i>Hordeum vulgare</i>	Phytozome	79 - 128	203 - 312	1e-06/ 2,9e-27
HORVU3Hr1G010100.1	<i>Hordeum vulgare</i>	Phytozome	47 - 96	171 - 280	9e-07/ 2,4e-27
HORVU3Hr1G010190.5	<i>Hordeum vulgare</i>	Phytozome	91 - 140	205 - 314	1,1e-05/ 3,1e-27
HORVU3Hr1G010240.1	<i>Hordeum vulgare</i>	Phytozome	55 - 104	169 - 277	1,4e-06/ 4,9e-27
HORVU3Hr1G064640.1	<i>Hordeum vulgare</i>	Phytozome	54 - 103	186 - 293	6,5e-08/ 1,1e-26
HORVU3Hr1G064640.4	<i>Hordeum vulgare</i>	Phytozome	71 - 120	203 - 310	7e-08/ 1,2e-26
HORVU3Hr1G064640.6	<i>Hordeum vulgare</i>	Phytozome	71 - 120	203 - 310	6e-08/ 1,1e-26
HORVU4Hr1G082090.2	<i>Hordeum vulgare</i>	Phytozome	66 - 115	185 - 294	1,3e-06/ 5,8e-28
HORVU4Hr1G082090.3	<i>Hordeum vulgare</i>	Phytozome	54 - 103	173 - 282	1,2e-06/ 5,1e-28
KFL00094_0070	<i>Klebsormidium flaccidum</i>	PlantTFDB	296-345	416-517	9,9e-06/ 5,4e-26
Medtr1g093600.1	<i>Medicago truncatula</i>	Phytozome	62 - 111	192 - 308	1,4e-08/ 1,4e-25
Medtr5g053920.1	<i>Medicago truncatula</i>	Phytozome	64-113	195-306	7,5e-09/ 7,7e-29
Misin06G229100.1	<i>Miscanthus sinensis</i>	Phytozome	81 - 130	219 - 320	1,6e-07/ 3,2e-28
Misin16G052800.1	<i>Miscanthus sinensis</i>	Phytozome	75 - 124	190 - 297	3,5e-07/ 7,2e-22
Misin16G063600.1	<i>Miscanthus sinensis</i>	Phytozome	75 - 124	190 - 297	4e-07/ 3,7e-21
Misin16G223800.1	<i>Miscanthus sinensis</i>	Phytozome	79 - 128	212 - 323	6e-08/ 6,8e-27
Misin17G229300.1	<i>Miscanthus sinensis</i>	Phytozome	75 - 124	208 - 319	5,9e-08/ 5,6e-27
MisinT148700.1	<i>Miscanthus sinensis</i>	Phytozome	78 - 127	193 - 304	1,7e-07/ 3,8e-28

APÊNDICE A. Conclusão.

MisinT241800.1	<i>Miscanthus sinensis</i>	Phytozome	83 - 132	219 - 320	1,6e-07/ 1,9e-28
MisinT530500.1	<i>Miscanthus sinensis</i>	Phytozome	13 - 62	128 - 235	5,6e-06/ 5,1e-28
Niben101Scf00872g03005.1	<i>Nicotiana benthamiana</i>	PlantTFDB	63-112	227-330	5,2e-09/ 4,8e-29
Niben101Scf03150g06002.1	<i>Nicotiana benthamiana</i>	PlantTFDB	63-112	198-301	5,7e-09/ 3,9e-30
Niben101Scf03488g06004.1	<i>Nicotiana benthamiana</i>	PlantTFDB	40 - 89	166 - 262	1,5e-05/ 2e-27
Niben101Scf10767g03010.1	<i>Nicotiana benthamiana</i>	PlantTFDB	59 - 108	215 - 316	1,7e-05/ 3,4e-29
Niben101Scf16806g00003.1	<i>Nicotiana benthamiana</i>	PlantTFDB	29 - 78	159 - 258	5e-07/ 1,6e-28
LOC_Os01g04750.1	<i>Oryza sativa</i>	Phytozome	65 - 114	178 - 285	4,6e-08/ 7,5e-26
LOC_Os01g04800.1	<i>Oryza sativa</i>	Phytozome	67-116	182-294	2,4e-08/ 1,6e-27
LOC_Os01g49830.1	<i>Oryza sativa</i>	Phytozome	80-129	216-321	1,7e-07/ 7,4e-28
LOC_Os05g47650.1	<i>Oryza sativa</i>	Phytozome	67-116	193-313	5,2e-08/ 4,6e-26
Pavir.Ca01826.1	<i>Panicum virgatum</i>	Phytozome	29 - 78	160 - 267	4,6e-08/ 2,3e-26
Pavir.Ea00549.1	<i>Panicum virgatum</i>	Phytozome	71 - 120	186 - 292	2,8e-08/ 2,9e-27
Pavir.Ea02092.1	<i>Panicum virgatum</i>	Phytozome	107 - 156	240 - 347	1,5e-07/ 7,2e-26
Pavir.Eb00701.1	<i>Panicum virgatum</i>	Phytozome	51 - 100	166 - 275	7,9e-06/ 7,3e-28
Pavir.Eb00702.1	<i>Panicum virgatum</i>	Phytozome	79 - 128	194 - 300	2,9e-08/ 3,8e-28
Pavir.Eb02845.1	<i>Panicum virgatum</i>	Phytozome	72-121	200-307	2e-08/ 9,6e-27
Pavir.J00384.1	<i>Panicum virgatum</i>	Phytozome	30 - 79	145 - 228	2,2e-07/ 7,3e-18
Pavir.J23476.1	<i>Panicum virgatum</i>	Phytozome	72-121	187-294	2,4e-08/ 8,7e-28
Phvul.003G111800.1	<i>Phaseolus vulgaris</i>	Phytozome	72 - 121	204 - 312	1,1e-08/ 3e-28
Phvul.007G102800.1	<i>Phaseolus vulgaris</i>	Phytozome	53 - 102	174 - 287	8,9e-09/ 2,6e-26
Potri.003G212800.1	<i>Populus trichocarpa</i>	Phytozome	43 - 92	169 - 275	3,6e-06/ 3,1e-24
Potri.008G117100.1	<i>Populus trichocarpa</i>	Phytozome	107-156	243-348	8,2e-09/ 8,2e-29
Potri.010G129200.1	<i>Populus trichocarpa</i>	Phytozome	104-153	240-344	8,4e-09/ 1,2e-29
*ScRAV1 (SHCRBa_021_A10)	<i>Saccharum</i> spp.	Sucest	80-129	218-316	1,4e-07/ 4,7e-28
**ScRAV2 (SCEZHR1055B01.G)	<i>Saccharum</i> spp.	Sucest	78-126	-	8,8e-05
**ScRAV3 (SCRRLRZ3042H10.G)	<i>Saccharum</i> spp.	Sucest	75-125	-	9,2e-09
**ScRAV4 (SCEZRT2023C02.G)	<i>Saccharum</i> spp.	Sucest	-	16-128	8.00E-29
Seita.3G158300.1	<i>Setaria italica</i>	Phytozome	76-125	206-312	1,7e-07/ 1,5e-26
Seita.5G096700.1	<i>Setaria italica</i>	Phytozome	34 - 83	146 - 254	3,6e-07/ 8,4e-25
Seita.5G096800.1	<i>Setaria italica</i>	Phytozome	35 - 84	150 - 256	8,7e-05/ 2,1e-26
Seita.5G096900.1	<i>Setaria italica</i>	Phytozome	75-124	190-297	8e-09/ 1,1e-28
Seita.5G280700.1	<i>Setaria italica</i>	Phytozome	79-128	214-320	4,5e-07/ 5,1e-28
Seita.7G266200.1	<i>Setaria italica</i>	Phytozome	31 - 80	131 - 218	9,1e-05/ 4,3e-14
Solyc04g007000.1.1	<i>Solanum lycopersicum</i>	Phytozome	65 - 114	203 - 307	2,2e-08/ 7,5e-27
Solyc05g009790.1.1	<i>Solanum lycopersicum</i>	Phytozome	71-120	207-310	1,5e-08/ 2,4e-30
Sobic.003G078100.1	<i>Sorghum bicolor</i>	Phytozome	82-131	197-308	1,3e-07/ 3,1e-28
Sobic.003G265200.1	<i>Sorghum bicolor</i>	Phytozome	80-129	222-326	1,4e-07/ 2e-28
Sobic.009G221400.1	<i>Sorghum bicolor</i>	Phytozome	81-130	215-325	5,5e-08/ 1,6e-26
Traes_3B_68BFA25F1.1	<i>Triticum aestivum</i>	Phytozome	55 - 104	169 - 277	1,3e-06/ 3,7e-27
Traes_3B_90BE2090B.1	<i>Triticum aestivum</i>	Phytozome	53-102	167-275	2e-07/ 8,4e-28
Traes_3B_D9C941D4C.1	<i>Triticum aestivum</i>	Phytozome	30 - 79	154 - 261	1,6e-05/ 1,7e-26
Traes_3B_DF7E92499.1	<i>Triticum aestivum</i>	Phytozome	48 - 97	162 - 271	2,7e-06/ 4,2e-27
Traes_3B_DF7E92499.2	<i>Triticum aestivum</i>	Phytozome	56 - 105	170 - 279	2,8e-06/ 4,4e-27
GRMZM2G059939_T02	<i>Zea mays</i>	Phytozome	72 - 121	206 - 318	5,9e-08/ 1,7e-26
GRMZM2G169654_T01	<i>Zea mays</i>	Phytozome	76-125	213-319	1,3e-07/ 7,3e-29

* Sequência proteica completa do fator de transcrição ScRAV1 usada na árvore filogenética derivada do BAC (cromossomo artificial bacteriano) SHCRBa_021_A10 (Tavares *et al.*, 2019), sendo a mesma sequência referente ao SAS (SCAGLR2011B02.g) que está presente no banco de dados do Sucest.

** SAS (*Sugarcane Assembled Sequences*).

APÊNDICE B. Valores das correlações de ScRAV1 com os transcritos de parede celular com p-valor $\leq 0,05$ nos diferentes segmentos das raízes de cana-de-açúcar.

SAS	Identificação	Sub_category1	Segmento	Correlação	p-valor
SCACCL6006A07.g	ScPAE1	Pectinacetyltransferase	S1	0,999910	0,008520
SCAGLR1064B01.g	ScEXP1	Expansin	S1	0,998893	0,029960
SCBFAD1089C07.g	ScGT8 1	Glycosyl transferase family 8 (cell wall metabolism, starch metabolism or lipopolysaccharide biosynthesis)	S1	0,999976	0,004398
SCBFLB2091E01.g	ScCalS2	1,3-beta-glucan synthase (putative callose synthase)	S1	-0,999989	0,002967
SCBGRT1047A07.g	ScBGN1	endo-1,3-beta-glucosidase	S1	-0,999399	0,022076
SCCCCL3001G09.b	ScCesA1	Cellulose synthase (CesA)	S1	0,999988	0,003118
SCCCCL3080C11.g	ScUGD1	UDP-D-glucuronate decarboxylase	S1	0,996963	0,049629
SCCCCL5004D04.g	ScCCR1	Cinnamoyl CoA reductase (CCR)	S1	0,998696	0,032511
SCCCLB1023H08.g	ScEXP2	Expansin	S1	0,997610	0,044019
SCCCLB1024D03.g	ScPAE2	Pectinacetyltransferase family protein	S1	1,000000	0,000003
SCCCLR1001A05.g	ScSUS1	sucrose synthase	S1	0,998095	0,039305
SCCCLR1069B09.g	ScCCoAOMT2	caffeoyl-CoA 3-O-methyltransferase	S1	0,999304	0,023758
SCCCLR1C06G07.g	ScSPS2	sucrose-phosphate synthase	S1	0,999289	0,024000
SCCCRZ1001H03.g	ScGalT1	Galactosyltransferase	S1	0,999987	0,003263
SCCCRZ1002A09.g	ScGT47 2	glycosyltransferase family 47	S1	0,999890	0,009448
SCCCRZ3097C07.g	ScCalS4	1,3-beta-glucan synthase (putative callose synthase)	S1	0,998787	0,031361
SCEPLB1044C07.g	ScXYN1	Glycoside hydrolase family 10 protein (Xylanase)	S1	0,999017	0,028232
SCEPLR1051B04.g	ScFLA1	Fasciclin-like arabinogalactan protein	S1	0,999988	0,003117
SCEQLB1064E10.g	ScEXP3	Expansin	S1	0,999112	0,026836
SCEQRT2091A08.g	ScXTH3	Xyloglucan endotransglycosylase/hydrolase (XTH)	S1	-0,998270	0,037448
SCEQSD1077C11.g	ScCW1	cell wall invertase	S1	-0,999596	0,018099
SCEZLB1007A09.g	ScGH3 2	glycosyl hydrolase family 3 protein	S1	0,999860	0,010649
SCJFHR1C03E12.b	ScGH1 3	Glycosyl hydrolase family 1	S1	0,999821	0,012056
SCJFLR1013A11.g	ScDIR-like1	dirigent-like protein	S1	-0,999652	0,016801
SCJFLR1013E04.g	ScUGDH1	UDP-glucose dehydrogenase	S1	0,997583	0,044274
SCJFLR1017B11.g	ScPAL2	phenylalanine ammonia-lyase (PAL)	S1	0,997706	0,043129
SCJFLR1074D02.g	ScBGN3	beta-1,3-glucanase Glycosyl hydrolases family 17	S1	0,999695	0,015733
SCJFRT1007G04.g	ScPGase3	Glycoside hydrolase family 28 protein (polygalacturonase)	S1	0,998930	0,029456
SCJFRT1060D04.g	ScGT8 3	Glycosyl transferase family 8 (cell wall metabolism, starch metabolism or lipopolysaccharide biosynthesis)	S1	0,999993	0,002455
SCJFRT2055A09.b	ScGlcT2	glucosyltransferase	S1	0,999068	0,027494
SCJFRZ2010G10.g	ScIFR/PCBER/PrR1	isoflavone reductase/Phenylcoumaran benzylic ether reductase/pinoresinol reductase (lignan biosynthesis)	S1	0,998777	0,031487
SCJLRT1013C01.g	ScPAL3	phenylalanine ammonia-lyase (PAL)	S1	0,999266	0,024389
SCJLRT1018H09.g	ScCW2	cell wall invertase	S1	0,997515	0,044891
SCJLRT1020C03.g	ScMAN2	Alpha-mannosidase (GH family 38)	S1	0,999676	0,016200
SCJLST1019F09.g	ScCHI4	Glycosyl hydrolases family 18 (chitinase; endo-N-acetylglucosaminidase; xylanase inhibitor; concanavalin B; narbonin)	S1	0,999295	0,023909
SCQGRT1040D08.g	ScFLA3	Fasciclin-like arabinogalactan protein	S1	0,998389	0,036141
SCQGRT1041B06.g	ScGlcT3	glucosyltransferase	S1	0,998940	0,029316
SCQSFL3032B02.g	ScGT47 4	glycosyltransferase family 47	S1	0,999688	0,015904
SCQSRT1034C08.g	ScFUT1	Fucosyltransferase	S1	-0,999863	0,010553
SCQSRT1034C09.g	ScCCoAOMT4	caffeoyl-CoA 3-O-methyltransferase	S1	0,999792	0,012982
SCQSRT1035C08.g	ScEGase3	endo-1,4-beta-glucanase Glycosyl hydrolase family 9	S1	0,999960	0,005688
SCQSRT1036E09.g	ScC3H1	putative p-Coumarate 3-hydroxylase (C3H)	S1	-0,999958	0,005833
SCRLR1016F07.g	ScGT8 4	Glycosyl transferase family 8 (cell wall metabolism, starch metabolism or lipopolysaccharide biosynthesis)	S1	0,997641	0,043733
SCRURT2006B10.g	ScBGN6	beta-1,3-glucanase Glycosyl hydrolases family 17	S1	0,999341	0,023118
SCSBLB1033B10.g	ScDIR-like2	dirigent-like protein	S1	0,999796	0,012844
SCSBS1057C12.g	ScCCR2	Cinnamoyl-coA reductase (CCR)	S1	0,997292	0,046860
SCSGFL4191C01.b	ScGT8 5	Glycosyl transferase family 8 (cell wall metabolism, starch metabolism or lipopolysaccharide biosynthesis)	S1	0,999363	0,022733
SCSGFL4C06E05.g	ScMAN3	Mannosyl-oligosaccharide 1,2-alpha-mannosidase	S1	0,997715	0,043041
SCSGHR1071F01.g	ScCesA3	Cellulose synthase (CesA)	S1	0,998283	0,037309
SCUTSD2027H09.g	ScCsl3	cellulose synthase like (CSL)	S1	-0,999802	0,012675
SCVPCL6041E07.g	ScC3H2	putative p-Coumarate 3-hydroxylase (C3H)	S1	0,998733	0,032054
SCVPFL1138B04.g	Sc4CL1	4-coumarate-CoA ligase	S1	0,998601	0,033678
SCVPRT2075H10.g	ScEXP5	Expansin	S1	0,999773	0,013557
SCVPRT3089B12.g	ScGH1 5	Glycosyl hydrolase family 1	S1	0,997442	0,045544
SCCCCL4009H01.g	ScC4H1	cinnamate 4-hydroxylase (C4H)	S2	-0,999753	0,014146

APÊNDICE B. Continuação.

SCCCHR1004C07.g	ScCCoAOMT1	caffeoyl-CoA 3-O-methyltransferase	S2	-0,999982	0,003788
SCCCLR1069D11.g	ScEGase1	endo-1,4-beta-glucanase Glycosyl hydrolase family 9	S2	-0,999272	0,024288
SCCCLR1079A02.g	ScCCoAOMT3	caffeoyl-CoA 3-O-methyltransferase	S2	-0,998734	0,032036
SCCCLR1C01A03.g	ScSUS2	sucrose synthase	S2	-0,999446	0,021191
SCCCLR2002A07.g	ScXTH2	Xyloglucan endotransglycosylase/hydrolase (XTH)	S2	-0,997224	0,047450
SCEQRT2028C06.g	ScEGase2	endo-1,4-beta-glucanase Glycosyl hydrolase family 9	S2	-0,999120	0,026711
SCEQSB1017F08.g	ScPGase2	Glycoside hydrolase family 28 protein (polygalacturonase)	S2	0,999440	0,021308
SCEZLB1007A09.g	ScGH3 2	glycosyl hydrolase family 3 protein	S2	-0,997859	0,041670
SCJFAD1010G05.b	ScABF1	Alpha-L-arabinofuranosidase	S2	-0,999944	0,006723
SCJFRZ2015C05.g	ScANN1	Annexin	S2	-0,997343	0,046416
SCJLRT1023B09.g	ScCOMT2	caffeic acid 3-O-methyltransferase	S2	0,999274	0,024264
SCMCRT2087B09.g	ScKOB1	kobito	S2	-0,998833	0,030761
SCQSHR1022C02.g	ScBGN4	beta-1,3-glucanase Glycosyl hydrolases family 17	S2	-0,999272	0,024286
SCRFL4004D05.g	ScUGD3	UDP-D-glucuronate decarboxylase	S2	-0,999923	0,007902
SCRUSB1061D02.g	ScGalT2	Galactosyltransferase	S2	-0,997648	0,043668
SCSBRZ3120B08.g	ScCHI5	chitinase	S2	-0,999151	0,026228
SCSFAD1125C08.g	ScPGI1	Polygalacturonase inhibitor	S2	-0,999345	0,023049
SCSFRT2069F06.g	ScXYN2	Glycoside hydrolase family 10 protein (Xylanase)	S2	-0,997666	0,043500
SCUTFL3079A10.g	ScCsl2	cellulose synthase like (CSL)	S2	-0,999684	0,016014
SCVPRT2076C12.g	ScXYN3	Glycoside hydrolase family 10 protein (Xylanase)	S2	-0,999695	0,015734
SCACLR1014A07.g	ScCalS1	putative callose synthase	S3	0,999991	0,002721
SCCCCL3080E10.b	ScPGase1	Glycoside hydrolase family 28 protein (polygalacturonase)	S3	0,998745	0,031897
SCCCRZ2002D05.g	ScGT47 3	glycosyltransferase family 47	S3	-0,996997	0,049348
SCEPFL3083G08.g	ScEPG1	Glycoside hydrolase family 28 protein (polygalacturonase)	S3	-0,997152	0,048057
SCEQRT1024E12.g	ScPAL1	phenylalanine ammonia-lyase (PAL)	S3	0,999285	0,024078
SCEQRT2029E12.g	ScCHI2	chitinase	S3	-0,999174	0,025885
SCEZLB1007H04.g	ScGT8 2	Glycosyl transferase family 8 (cell wall metabolism, starch metabolism or lipopolysaccharide biosynthesis)	S3	0,998114	0,039109
SCEZLB1013D07.g	ScBGN2	beta-1,3-glucanase Glycosyl hydrolases family 17	S3	0,999243	0,024768
SCEZLB2068G12.g	ScGlcT1	glucosyltransferase like protein	S3	0,998552	0,034264
SCEZLR1009C10.g	ScGalTI/XXT11	glycosyltransferase family 34 (Galactosyltransferase-like; Xylosyltransferase-like protein)	S3	-0,999562	0,018849
SCEZRZ3015E11.g	ScCHI3	chitinase	S3	-0,999938	0,007084
SCJFRZ3C05E03.b	ScIFR/PrR1	isoflavone reductase/pinoresinol reductase (lignan biosynthesis)	S3	0,998586	0,033858
SCQSST1037F05.g	ScUGPase1	UDP-glucose pyrophosphorylase	S3	0,999837	0,011481
SCRFHR1008D11.g	ScBGN5	beta-1,3-glucanase Glycosyl hydrolases family 17	S3	0,999517	0,019785
SCRUAD1064C10.g	ScF5H/F3H1	putative ferulate 5 hydroxylase (F5H); flavonoid 3 hydroxylase (F3H)	S3	0,999621	0,017518
SCRUSB1061D02.g	ScGalT2	Galactosyltransferase	S3	0,999964	0,005383
SCSBAD1054A07.g	ScCAD1	Cinnamyl alcohol dehydrogenase (CAD)	S3	-0,999836	0,011523
SCSBLB1032E01.g	ScFLA5	Fascilin-like arabinogalactan protein	S3	-0,997380	0,046098
SCSBSD1057C12.g	ScCCR2	Cinnamoyl-coA reductase (CCR)	S3	-0,999640	0,017089
SCVPLB1020E05.g	ScIFR/PrR2	isoflavone reductase/pinoresinol reductase (lignan biosynthesis)	S3	0,999991	0,002651
SCVPRZ2042G03.g	ScGH1 6	Glycosyl hydrolase family 1	S3	-0,998813	0,031027
SCACAD1036B04.g	ScSPS1	sucrose-phosphate synthase	S4	0,999491	0,020314
SCCCCL3140B03.g	ScXI1	Xylose isomerase	S4	-0,999676	0,016217
SCCCCL4013E04.g	ScCalS3	1,3-beta-glucan synthase (putative callose synthase)	S4	0,998584	0,033884
SCCCCL5001F06.g	ScXTH1	Xyloglucan endotransglycosylase/hydrolase (XTH)	S4	0,999588	0,018272
SCCCLB1025F10.g	ScGT47 1	glycosyltransferase family 47	S4	0,999903	0,008864
SCCCLB1C06E01.g	ScCsl1	Cellulose synthase like (CSL) putative mannan synthase	S4	0,999990	0,002840
SCCCLR1048B05.g	ScCHI1	Glycosyl hydrolases family 18 (probable chitinase)	S4	-0,999552	0,019066
SCCCLR1066B10.g	ScCesA2	Cellulose synthase (CesA)	S4	0,999950	0,006335
SCCCLR1C06G07.g	ScSPS2	sucrose-phosphate synthase	S4	0,997744	0,042774
SCCCRZ2001D09.g	ScUGD2	UDP-D-glucuronate decarboxylase	S4	0,998130	0,038936
SCCCST1006H12.g	ScGH1 1	Glycosyl hydrolase family 1	S4	-0,999998	0,001385
SCEPRT2046G07.g	ScLAC1	laccase-like protein	S4	0,999620	0,017557
SCEQHR1082B01.g	ScGH1 2	Glycosyl hydrolase family 1	S4	0,999802	0,012660
SCEQLR1093F09.g	ScGH3 1	glycosyl hydrolase family 3 protein	S4	0,997224	0,047444
SCEZLB1005G10.g	ScEXP4	Expansin	S4	0,998163	0,038597
SCEZLB1012C12.g	ScPAE3	Pectinacetylesterase	S4	0,999623	0,017487
SCEZLR1009C10.g	ScGalTI/XXT11	glycosyltransferase family 34 (Galactosyltransferase-like; Xylosyltransferase-like protein)	S4	0,998168	0,038542

APÊNDICE B. Conclusão.

SCEZST3147C12.g	ScPAE4	Pectinacetylesterase	S4	-0,999068	0,027485
SCJFLR1017B11.g	ScPAL2	phenylalanine ammonia-lyase (PAL)	S4	-0,999776	0,013475
SCJFLR1017E03.g	ScGH1 4	Glycosyl hydrolase family 1	S4	0,999597	0,018084
SCJFRT1061B07.g	ScCOMT1	caffeic acid 3-O-methyltransferase	S4	0,999325	0,023385
SCJFRZ2028D12.g	ScMAN1	Mannosyl-oligosaccharide 1,2- α -mannosidase	S4	0,997929	0,040976
SCMCSB1114E08.g	ScFLA2	Fasciclin-like arabinogalactan protein	S4	-0,999875	0,010048
SCQGF4080C07.g	ScCal55	1,3-beta-glucan synthase (putative callose synthase)	S4	0,998968	0,028919
SCQSLR1040G04.g	ScFLA4	Fasciclin-like arabinogalactan protein	S4	0,998471	0,035213
SCQSRT1034C08.g	ScFUT1	Fucosyltransferase	S4	-0,999938	0,007090
SCQSRT1034C09.g	ScCCoAOMT4	caffeoyl-CoA 3-O-methyltransferase	S4	-0,999979	0,004153
SCQSSB1077F07.g	ScXYL1	Glycosyl hydrolase family 3 protein (Xylosidase)	S4	-0,997763	0,042589
SCRFST1045G02.g	ScCal56	1,3-beta-glucan synthase (putative callose synthase)	S4	-0,997589	0,044215
SCUTAD1034F12.b	ScCesA4	Cellulose synthase (CesA)	S4	-0,997004	0,049292
SCUTAM2088D10.g	ScD-TDP-GD/UGD1	D-TDP-glucose dehydratase/UDP-D-glucuronate decarboxylase	S4	0,998978	0,028778
SCVPRZ3027D06.g	ScEGase4	endo-1,4-beta-glucanase Glycosyl hydrolase family 9	S4	0,999886	0,009594

APÊNDICE C. Valores das correlações de *ScRAV1* com os fatores de transcrição com p-valor $\leq 0,05$ nos diferentes segmentos das raízes de cana-de-açúcar.

SAS	Identificação	Sub_category	Segmento	Correlação	p-valor
SCACLR1127A11.g	ScbHLH1	BHLH	S1	0,999984	0,003626
SCACST3157E04.g	ScCCAAT1	CCAAT	S1	-0,999195	0,025550
SCAGHR1018G01.g	ScSET1	SET	S1	0,998689	0,032600
SCAGLR2033E10.g	ScALF1	ALFIN-LIKE	S1	-0,998774	0,031523
SCBGLR1118B02.g	ScARR1	ARR-B	S1	0,998866	0,030315
SCCCCL3001D02.g	ScNAC2	NAC	S1	0,999996	0,001837
SCCCCL3001G10.b	ScDBP1	DBP	S1	-0,999048	0,027787
SCCCCL4005B02.g	ScCSD2	CSD	S1	0,999263	0,024441
SCCCCLB1002H05.g	ScC3H2	C3H	S1	0,998556	0,034213
SCCCLR1048A12.g	ScPHD2	PHD	S1	0,999339	0,023147
SCCCLR1048B01.g	ScHMG1	HMG	S1	0,999307	0,023695
SCCCLR1C05B03.g	ScMYB2	MYB	S1	0,999996	0,001853
SCCCLR1C08B04.g	ScMT-I GTA1	METALLOTHIONEIN-I GENE TRANSCRIPTION ACTIVATOR	S1	-0,998738	0,031984
SCCCLR2003E10.g	ScNAC6	NAC	S1	0,999314	0,023582
SCCCRT1002F09.g	ScbZIP4	BZIP	S1	0,999941	0,006933
SCCCRZ2003B03.g	ScbZIP5	BZIP	S1	0,999462	0,020889
SCEPLB1042B08.g	ScCOL2	C2C2-CO-LIKE	S1	0,999329	0,023318
SCEPLR1030C04.g	ScLZ2	LEUCINE ZIPPER	S1	0,998740	0,031957
SCEPRZ1008A04.g	ScEREBP3	AP2/EREBP	S1	0,999901	0,008953
SCEQHR1081E06.g	ScEREBP4	AP2/EREBP	S1	0,998651	0,033070
SCEQRT1024G02.g	ScHB3	HOMEBOX	S1	0,998645	0,033149
SCEQRT1030H12.g	ScMYB3	MYB	S1	0,998427	0,035707
SCEQRT2028G08.g	ScMYB4	MYB	S1	0,999611	0,017766
SCEQRT2091B01.g	ScHB4	HOMEBOX	S1	0,999167	0,025980
SCEZLB1009D03.g	ScDBP4	DBP	S1	0,999105	0,026935
SCEZRZ1012H10.g	ScHB5	HOMEBOX	S1	0,999779	0,013370
SCEZRZ3015B06.g	ScGATA1	C2C2-GATA	S1	0,999862	0,010564
SCJFLR1035B03.g	ScARR4	ARR-B	S1	0,999196	0,025532
SCJFLR1035G11.g	ScbZIP6	BZIP	S1	0,999280	0,024151
SCJFRT1008G01.g	ScDBP5	DBP	S1	0,999230	0,024984
SCJFST1013H06.g	ScMADS2	MADS	S1	0,999158	0,026123
SCJLLR1011F10.g	ScC3H3	C3H	S1	0,997803	0,042211
SCJLLR1033A10.g	ScHAP1	CCAAT-HAP2	S1	0,999882	0,009775
SCJLRT1013D05.g	ScC3H4	C3H	S1	0,997456	0,045418
SCJLRT1015C10.g	ScC2H7	C2H2	S1	0,998167	0,038554
SCJLRT1015D07.g	ScMADS3	MADS	S1	0,998894	0,029949
SCJLRT1019E07.g	ScbZIP7	BZIP	S1	0,999885	0,009655
SCJLRZ1024E12.g	ScMYB5	MYB	S1	0,997279	0,046970
SCMCSB1109C11.g	ScNAC8	NAC	S1	0,999953	0,006204
SCQGHR1012C06.g	ScPHD5	PHD	S1	-0,999799	0,012771
SCQGLR1085G10.g	ScMADS4	MADS	S1	0,999708	0,015385
SCRFLR1012G12.g	ScSNF2	SNF2	S1	-0,999911	0,008508
SCRULB1060D10.g	ScDBP7	DBP	S1	0,997798	0,042259
SCSBAD1051C05.g	ScMYB8	MYB	S1	0,998127	0,038973
SCSGFL4C07B02.g	ScABI2	ABI3/VP1	S1	0,999958	0,005835
SCSGHR1070C04.g	ScBZR2	BZR	S1	-0,997686	0,043313
SCSGRZ3062B09.g	ScC3H6	C3H	S1	0,999692	0,015802
SCUTHR1065A04.g	ScC2H10	C2H2	S1	0,996952	0,049721
SCUTLR2023D06.g	ScHAP2	CCAAT-HAP2	S1	0,999582	0,018404
SCUTRZ2023F11.g	ScDBP9	DBP	S1	0,999878	0,009935
SCUTST3130A02.g	ScSBP2	SBP	S1	0,998686	0,032641
SCVPCL6042A04.g	ScMYB10	MYB	S1	0,999358	0,022818
SCVPLB1018G11.g	ScbZIP10	BZIP	S1	0,999997	0,001550
SCVPRZ2039D09.g	ScDBP11	DBP	S1	0,999900	0,008998
SCVPRZ2039D11.g	ScDBP12	DBP	S1	0,998198	0,038226
SCVPRZ2039E07.g	ScGEBP1	GEBP	S1	0,997351	0,046344
SCVPRZ2042G02.g	ScMYBR1	MYB-RELATED	S1	0,999977	0,004283
SCVPST1060E04.g	ScbHLH6	BHLH	S1	0,999977	0,004328
SCBFRZ2016F05.g	ScGRAS1	GRAS	S2	-0,999769	0,013677
SCBGRT1054F09.g	ScC2H1	C2H2	S2	0,997889	0,041374
SCCCCL7038B02.g	ScPHD1	PHD	S2	-0,998379	0,036257

APÊNDICE C. Continuação.

SCCCFL2002D10.g	ScNAC3	NAC	S2	-0,999712	0,015274
SCCCLR1C03C05.g	ScbZIP3	BZIP	S2	-0,999112	0,026831
SCCCLR2001F11.g	ScMADS1	MADS	S2	-0,999286	0,024053
SCCCRT2001B02.g	ScALF2	ALFIN-LIKE	S2	-0,998416	0,035842
SCCCRZ1004G09.g	ScC2H4	C2H2	S2	-0,997632	0,043824
SCCCRZ2001A12.g	ScbHLH4	BHLH	S2	-0,999999	0,001062
SCCCSB1002C05.g	ScABI1	ABI3/VP1	S2	-0,998236	0,037819
SCEPCL6029A11.g	ScC2H5	C2H2	S2	0,999613	0,017718
SCEQRT2025D01.g	ScJMJ1	JUMONJI	S2	-0,999823	0,011985
SCEZHR1047A03.g	ScWRKY1	WRKY	S2	-0,998959	0,029048
SCEZLR1009A01.g	ScARF3	ARF	S2	0,999781	0,013337
SCJFRZ2010A05.g	ScHB6	HOMEBOX	S2	0,999962	0,005523
SCJLRZ1019D02.g	ScARR5	ARR-B	S2	-0,998043	0,039831
SCQSAM1031B04.g	ScbHLH5	BHLH	S2	-0,999358	0,022806
SCQSFL3031C01.g	ScC3H5	C3H	S2	-0,999249	0,024677
SCRRLT3031B06.g	ScMYB7	MYB	S2	0,998199	0,038214
SCRUFL3068C10.g	ScbHSH1	BHSH	S2	-0,997077	0,048686
SCSFLR2009G10.g	ScSET4	SET	S2	-0,999998	0,001145
SCSFLR2016F09.g	ScbZIP8	BZIP	S2	0,999994	0,002216
SCSGRZ3061B02.g	ScGRAS4	GRAS	S2	-0,997951	0,040760
SCUTLR1058E01.g	ScEREBP6	AP2/EREBP	S2	-0,998170	0,038520
SCAGAD1077B04.g	ScEREBP1	AP2/EREBP	S3	-0,999992	0,002554
SCAGFL8004F02.g	ScJUN1	JUN ACTIVADOR	S3	-0,999966	0,005261
SCBGLR1003E02.g	ScDOF1	C2C2-DOF	S3	-0,999888	0,009519
SCBGLR1113F05.g	ScMYB1	MYB	S3	-0,999887	0,009563
SCBGLR1118D05.g	ScNAC1	NAC	S3	-0,999854	0,010888
SCCCCL2001C11.b	ScSET2	SET	S3	-0,999851	0,010992
SCCCCL4005C09.g	ScbZIP2	BZIP	S3	-0,999838	0,011448
SCCCCL4014B04.g	ScC2H2	C2H2	S3	-0,999804	0,012616
SCCCLB1024B02.g	ScZIM2	ZIM	S3	-0,999490	0,020327
SCCCLR1076F11.g	ScLIM1	LIM	S3	-0,999471	0,020701
SCCCLR1C04E10.g	ScPHD3	PHD	S3	-0,999353	0,022900
SCCCLR2C01C12.g	ScDBP2	DBP	S3	-0,998671	0,032826
SCCCRT1002F09.g	ScbZIP4	BZIP	S3	-0,998665	0,032900
SCCCST2002F02.g	ScEREBP2	AP2/EREBP	S3	-0,998656	0,033012
SCCCST3001B11.g	ScARR2	ARR-B	S3	-0,998473	0,035188
SCEPLR1008H10.g	ScDBP3	DBP	S3	-0,997603	0,044091
SCEPRZ1008F02.g	ScLIM2	LIM	S3	0,996998	0,049344
SCEPRZ1010C07.g	ScHB2	HOMEBOX	S3	0,997050	0,048914
SCEZAM2058F10.g	ScARR3	ARR-B	S3	0,997254	0,047187
SCEZLR1009A01.g	ScARF3	ARF	S3	0,997334	0,046499
SCEZRZ3095B02.g	ScSNF1	SNF2	S3	0,997407	0,045852
SCJLRT1006H10.g	ScC2H6	C2H2	S3	0,997736	0,042848
SCJLRZ1019G05.b	ScGRAS3	GRAS	S3	0,997949	0,040779
SCQGLR2017G12.g	ScDBP6	DBP	S3	0,997993	0,040345
SCQGLR2025A07.g	ScWRKY2	WRKY	S3	0,998331	0,036789
SCQGLR2025E03.g	ScWRKY3	WRKY	S3	0,998523	0,034610
SCQSFL3031C01.g	ScC3H5	C3H	S3	0,998532	0,034494
SCRLLR1059G12.g	ScMYB6	MYB	S3	0,998672	0,032813
SCRUHR1073H07.g	ScC2H8	C2H2	S3	0,998902	0,029831
SCSFRT2072F01.g	ScMYB9	MYB	S3	0,998938	0,029344
SCSFST1064B01.g	ScARF4	ARF	S3	0,999024	0,028131
SCUTFL1055E01.g	ScC3H7	C3H	S3	0,999710	0,015345
SCUTLR2023G02.g	ScWRKY4	WRKY	S3	0,999838	0,011469
SCVPLB1017D07.g	ScDBP10	DBP	S3	0,999874	0,010098
SCACLR1127A11.g	ScbHLH1	BHLH	S4	-0,999999	0,000918
SCACLR1130E07.g	ScbHLH2	BHLH	S4	-0,999973	0,004691
SCAGLR2011C01.g	ScCSD1	CSD	S4	-0,999956	0,005982
SCAGLR2018G01.g	ScZIM1	ZIM	S4	-0,999680	0,016108
SCBGLR1118D05.g	ScNAC1	NAC	S4	-0,999626	0,017413
SCCCCL2001C11.b	ScSET2	SET	S4	-0,999430	0,021503
SCCCCL3001H09.g	ScbZIP1	BZIP	S4	-0,999329	0,023319

APÊNDICE C. Conclusão.

SCCCCL4017E11.g	ScC3H1	C3H	S4	-0,999173	0,025897
SCCCFL2002D10.g	ScNAC3	NAC	S4	-0,998858	0,030424
SCCCLB1021E04.g	ScLZ1	LEUCINE ZIPPER	S4	-0,998825	0,030859
SCCCLR1022C05.g	ScHLH1	HLH (HELIX-LOOP-HELIX)	S4	-0,998716	0,032269
SCCCLR1022F12.g	ScARF1	ARF	S4	-0,998082	0,039440
SCCCLR1024F10.g	ScARF2	ARF	S4	-0,997173	0,047883
SCCCLR1048F08.g	ScJUN2	JUN ACTIVADOR	S4	0,996935	0,049854
SCCCLR1070C08.g	ScNAC4	NAC	S4	0,997321	0,046614
SCCCLR1075D03.g	ScC2H3	C2H2	S4	0,997524	0,044812
SCCCLR1076F07.g	ScbHLH3	BHLH	S4	0,997566	0,044430
SCCCLR1079C08.g	ScHMG2	HMG	S4	0,997588	0,044227
SCCCLR1C04A01.g	ScHB1	HOMEBOX	S4	0,997698	0,043202
SCCCLR1C04E10.g	ScPHD3	PHD	S4	0,997896	0,041309
SCCCLR1C04H01.g	ScNAC5	NAC	S4	0,997948	0,040790
SCCCLR1C05E08.g	ScCOL1	C2C2-CO-LIKE	S4	0,997987	0,040402
SCCCR21004G09.g	ScC2H4	C2H2	S4	0,998105	0,039196
SCCCSB1002C05.g	ScABI1	ABI3/VP1	S4	0,998293	0,037206
SCCCST3005F09.g	ScNAC7	NAC	S4	0,998944	0,029261
SCEPAM2011G11.g	ScJUN3	JUN ACTIVADOR	S4	0,998986	0,028667
SCEPRZ1011B04.g	ScPHD4	PHD	S4	0,999239	0,024845
SCEQRT2026B09.g	ScBZR1	BZR	S4	0,999267	0,024372
SCEZLB1009A09.g	ScHR1	HORMONE-RELATED	S4	0,999272	0,024285
SCJFFL1C07F02.b	ScLZ3	LEUCINE ZIPPER	S4	0,999317	0,023526
SCJLLR1054D12.g	ScGRAS2	GRAS	S4	0,999352	0,022925
SCQGLR1062D05.g	ScSBP1	SBP	S4	0,999563	0,018815
SCQSHR1023B08.g	ScEREBP5	AP2/EREBP	S4	0,999607	0,017843
SCQSRT1034D11.g	ScPHD6	PHD	S4	0,999627	0,017388
SCSBRZ2020C03.g	ScC2H9	C2H2	S4	0,999664	0,016497
SCSFLR2009C10.g	ScSET3	SET	S4	0,999666	0,016445
SCSFLR2009G10.g	ScSET4	SET	S4	0,999750	0,014249
SCSGFL4C06A07.g	ScPHD7	PHD	S4	0,999766	0,013764
SCSGFL4C07G09.g	ScbZIP9	BZIP	S4	0,999809	0,012456
SCSGSB1005A04.g	ScHB7	HOMEBOX	S4	0,999900	0,008998
SCUTAM2113B12.g	ScDBP8	DBP	S4	0,999917	0,008195
SCVPLB1017D07.g	ScDBP10	DBP	S4	0,999987	0,003278
SCVPLR2019D04.g	ScIAA1	AUX/IAA	S4	0,999991	0,002654
SCVPRZ2043G02.g	ScSET5	SET	S4	0,999997	0,001458
SCVPRZ2043G09.g	ScBZR3	BZR	S4	0,999999	0,000787

APÊNDICE D. Termos GOs para função molecular, processo biológico e componente celular para os transcritos de parede celular anotados em *A. thaliana* com base nas sequências completas de *S. bicolor*.

SAS	Identificação	Segmento	<i>S. bicolor</i>	<i>A. thaliana</i>	Símbolo	GO		
			Hit	Hit		Molecular Function	Biological Process	Cellular Component
SCJFRT1060D04.g	ScGT8 3	S1	Sobic.002G274700.1	AT3G61130.1	GAUT1, LGT1	GO:0047262 polygalacturonate 4-alpha-galacturonosyltransferase activity GO:0016757 transferase activity, transferring glycosyl groups GO:0016740 transferase activity	GO:0010289 homogalacturonan biosynthetic process GO:0045489 pectin biosynthetic process GO:0071555 cell wall organization	GO:0005794 Golgi apparatus GO:0000139 Golgi membrane GO:0005768 endosome GO:0005802 trans-Golgi network GO:0016020 membrane GO:0016021 integral component of membrane
SCCCLB1024D03.g	ScPAE2	S1	Sobic.006G204500.1	AT4G19410.1	PAE7, ATPAE7	GO:0016787 hydrolase activity GO:0052793 pectin acetylsterase activity	GO:0071555 cell wall organization	GO:0005618 cell wall GO:0005576 extracellular region GO:0009505 plant-type cell wall GO:0009506 plasmodesma GO:0016020 membrane
SCBFAD1089C07.g	ScGT8 1	S1	Sobic.001G338400.1	AT1G18580.1	GAUT11	GO:0016757 transferase activity, transferring glycosyl groups GO:0047262 polygalacturonate 4-alpha-galacturonosyltransferase activity GO:0016740 transferase activity	GO:0010192 mucilage biosynthetic process GO:0045489 pectin biosynthetic process GO:0071555 cell wall organization	GO:0005794 Golgi apparatus GO:0000139 Golgi membrane GO:0016020 membrane GO:0016021 integral component of membrane
SCEPLR1051B04.g	ScFLA1	S1	Sobic.009G055900.1	AT4G12730.1	FLA2		GO:0005886 plasma membrane GO:0031225 anchored component of membrane GO:0016020 membrane GO:0005774 vacuolar membrane GO:0046658 anchored component of plasma membrane	
SCCCRZ1001H03.g	ScGalT1	S1	Sobic.010G059700.2	AT4G32120.1	HPGT2	GO:0008378 galactosyltransferase activity GO:0016740 transferase activity GO:0016757 transferase activity, transferring glycosyl groups GO:1990714 hydroxyproline O-galactosyltransferase activity GO:0003824 catalytic activity GO:0004553 hydrolase activity, hydrolyzing O-glycosyl compounds GO:0008810 cellulase activity GO:0016787 hydrolase activity GO:0016798 hydrolase activity, acting on glycosyl bonds	GO:0006486 protein glycosylation GO:0010405 arabinogalactan protein metabolic process GO:0018258 protein O-linked glycosylation via hydroxyproline	GO:0016020 membrane GO:0000139 Golgi membrane GO:0005794 Golgi apparatus GO:0016021 integral component of membrane
SCQSRT1035C08.g	ScEGase3	S1	Sobic.006G282300.1	AT1G65610.1	ATGH9A2, KOR2, ATKOR2	GO:0003824 catalytic activity GO:0004553 hydrolase activity, hydrolyzing O-glycosyl compounds GO:0008810 cellulase activity GO:0016787 hydrolase activity GO:0016798 hydrolase activity, acting on glycosyl bonds	GO:0005975 carbohydrate metabolic process GO:0000272 polysaccharide catabolic process GO:0008152 metabolic process GO:0030245 cellulose catabolic process	GO:0016020 membrane GO:0016021 integral component of membrane GO:0071555 cell wall organization
SCQSRT1036E09.g	ScC3H1	S1	Sobic.003G327800.1	AT2G40890.1	CYP98A3, REF8	GO:0005515 protein binding GO:0004497 monooxygenase activity GO:0046409 p-coumarate 3-hydroxylase activity GO:0005506 iron ion binding GO:0016491 oxidoreductase activity GO:0016705 oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen GO:0016709 oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen GO:0020037 heme binding	GO:0009699 phenylpropanoid biosynthetic process GO:0005514 oxidation-reduction process GO:0009805 coumarin biosynthetic process GO:0009809 lignin biosynthetic process GO:0009813 flavonoid biosynthetic process	GO:0005783 endoplasmic reticulum GO:0016020 membrane GO:0005739 mitochondrion GO:0005886 plasma membrane GO:0016021 integral component of membrane GO:0043231 intracellular membrane-bounded organelle

APÊNDICE D. Continuação.

						GO:0042802 identical protein binding GO:0046872 metal ion binding		
SCEZLB1007A09.g	ScGH3 2	S1/S2	Sobic.001G089000.1	AT5G20950.2	BGLC1	GO:0004553 hydrolase activity, hydrolyzing O-glycosyl compounds GO:0008422 beta-glucosidase activity GO:0016787 hydrolase activity	GO:0005975 carbohydrate metabolic process GO:0009251 glucan catabolic process GO:0009969 xyloglucan biosynthetic process	GO:0005618 cell wall GO:0009505 plant-type cell wall GO:0005829 cytosol GO:0005886 plasma membrane GO:0009506 plasmodesma GO:0016020 membrane GO:0048046 apoplast
SCRLFL4004D05.g	ScUGD3	S2	Sobic.009G112600.1	AT2G47650.1	UXS4	GO:0048040 UDP-glucuronate decarboxylase activity GO:0016829 lyase activity GO:0016831 carboxy-lyase activity GO:0070403 NAD+ binding	GO:0033320 UDP-D-xylose biosynthetic process	GO:0005794 Golgi apparatus GO:0016020 membrane GO:0016021 integral component of membrane GO:0032580 Golgi cisterna membrane GO:0005578 proteinaceous extracellular matrix
SCJFAD1010G05.b	ScABF1	S2	Sobic.008G017800.1	AT3G10740.1	ARAF, ARAF1, ASD1	GO:0046556 alpha-L- arabinofuranosidase activity GO:0009044 xylan 1,4-beta-xylosidase activity GO:0016787 hydrolase activity	GO:0046373 L-arabinose metabolic process GO:0045493 xylan catabolic process	GO:0005576 extracellular region GO:0005773 vacuole GO:0009505 plant-type cell wall GO:0048046 apoplast
SCVPRT2076C12.g	ScXYN3	S2	Sobic.002G127400.2	AT1G10050.1	ATXYN2	GO:0004553 hydrolase activity, hydrolyzing O-glycosyl compounds GO:0016787 hydrolase activity GO:0016798 hydrolase activity, acting on glycosyl bonds GO:0031176 endo-1,4-beta-xylanase activity	GO:0005975 carbohydrate metabolic process	
SCSFAD1125C08.g	ScPGI1	S2	Sobic.002G343600.1	AT5G06860.1	ATPGIP1, PGIP1	GO:0090353 polygalacturonase inhibitor activity	GO:0043086 negative regulation of catalytic activity	GO:0005618 cell wall GO:0016020 membrane GO:0005576 extracellular region GO:0005794 Golgi apparatus GO:0005829 cytosol GO:0009505 plant-type cell wall GO:0009506 plasmodesma GO:0000139 Golgi membrane GO:0005794 Golgi apparatus GO:0009506 plasmodesma GO:0016020 membrane GO:0016021 integral component of membrane
SCUTFL3079A10.g	ScCsl2	S2	Sobic.003G308100.1	AT4G07960.1	ATCSLC12, CSLC12	GO:0016740 transferase activity GO:0016757 transferase activity, transferring glycosyl groups	GO:0071555 cell wall organization	
SCCCCL4009H01.g	ScC4H1	S2	Sobic.002G126600.1	AT2G30490.1	C4H	GO:0005515 protein binding GO:0016710 trans-cinnamate 4- monooxygenase activity GO:0004497 monooxygenase activity GO:0005506 iron ion binding GO:0016491 oxidoreductase activity GO:0016705 oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen GO:0020037 heme binding GO:0042802 identical protein binding GO:0046872 metal ion binding GO:0047763 caffeate O- methyltransferase activity GO:0008168 methyltransferase activity GO:0030744 luteolin O- methyltransferase activity	GO:0055114 oxidation-reduction process GO:0006744 ubiquinone biosynthetic process GO:0009555 pollen development GO:0009698 phenylpropanoid metabolic process GO:0009808 lignin metabolic process GO:0032502 developmental process GO:0040007 growth GO:0080167 response to karrikin GO:0030187 melatonin biosynthetic process GO:0009809 lignin biosynthetic process GO:0032259 methylation	GO:0005783 endoplasmic reticulum GO:0016020 membrane GO:0005886 plasma membrane GO:0005774 vacuolar membrane GO:0005794 Golgi apparatus GO:0009505 plant-type cell wall GO:0009506 plasmodesma GO:0016021 integral component of membrane
SCJLRT1023B09.g	ScCOMT2	S2	Sobic.007G047300.1	AT5G54160.1	ATOMT1, OMT1, COMT1, ATCOMT, OMT3			GO:0005634 nucleus GO:0005737 cytoplasm GO:0005886 plasma membrane GO:0009506 plasmodesma

APÊNDICE D. Continuação.

						GO:0008171 O-methyltransferase activity GO:0016740 transferase activity GO:0017096 acetylserotonin O-methyltransferase activity GO:0030755 quercetin 3-O-methyltransferase activity GO:0033799 myricetin 3'-O-methyltransferase activity GO:0046983 protein dimerization activity GO:0102822 quercetin 3'-O-methyltransferase activity	GO:0051555 flavonol biosynthetic process	
SCVPLB1020E05.g	SciFR/PrR2	S3	Sobic.003G104600.1	AT1G75280.1	P3	GO:0016491 oxidoreductase activity	GO:0006979 response to oxidative stress GO:0046686 response to cadmium ion GO:0055114 oxidation-reduction process	GO:0005737 cytoplasm GO:0005886 plasma membrane
SCSBAD1054A07.g	ScCAD1	S3	Sobic.002G195600.1	AT4G37970.1	ATCAD6, CAD6	GO:0016491 oxidoreductase activity GO:0008270 zinc ion binding GO:0045551 cinnamyl-alcohol dehydrogenase activity GO:0046872 metal ion binding GO:0052747 sinapyl alcohol dehydrogenase activity	GO:0009809 lignin biosynthetic process GO:0055114 oxidation-reduction process GO:0009699 phenylpropanoid biosynthetic process	
SCQSST1037F05.g	ScUGPase1	S3	Sobic.002G291200.2	AT5G17310.2	UGP2	GO:0003983 UTP-glucose-1-phosphate uridylyltransferase activity GO:0016740 transferase activity GO:0016779 nucleotidyltransferase activity GO:0070569 uridylyltransferase activity	GO:0006011 UDP-glucose metabolic process GO:0005977 glycogen metabolic process GO:0008152 metabolic process GO:0009651 response to salt stress GO:0010942 positive regulation of cell death GO:0046686 response to cadmium ion GO:0052543 callose deposition in cell wall	GO:0005737 cytoplasm GO:0005829 cytosol GO:0005886 plasma membrane
SCRUSB1061D02.g	ScGalT2	S2/S3	Sobic.007G118500.1	AT5G53340.1	HPGT1	GO:0008378 galactosyltransferase activity GO:0016740 transferase activity GO:0016757 transferase activity, transferring glycosyl groups GO:1990714 hydroxyproline O-galactosyltransferase activity	GO:0006486 protein glycosylation GO:0010405 arabinogalactan protein metabolic process GO:0018258 protein O-linked glycosylation via hydroxyproline	GO:0005794 Golgi apparatus GO:0016020 membrane GO:0000139 Golgi membrane GO:0005768 endosome GO:0005802 trans-Golgi network GO:0016021 integral component of membrane
SCEZRZ3015E11.g	ScCHI3	S3	Sobic.001G516000.1	AT3G12500.1	ATHCHIB, PR3, PR-3, CHI-B, B-CHI, HCHIB	GO:0004568 chitinase activity GO:0008061 chitin binding GO:0016787 hydrolase activity GO:0016798 hydrolase activity, acting on glycosyl bonds	GO:0005975 carbohydrate metabolic process GO:0006032 chitin catabolic process GO:0000272 polysaccharide catabolic process GO:0006952 defense response GO:0008152 metabolic process GO:0009626 plant-type hypersensitive response GO:0016998 cell wall macromolecule catabolic process GO:0031640 killing of cells of other organism GO:0046686 response to cadmium ion GO:0050832 defense response to fungus	GO:0005773 vacuole GO:0005774 vacuolar membrane GO:0005829 cytosol

APÊNDICE D. Continuação.

SCEQRT2029E12.g	ScCHI2	S3	Sobic.009G130100.1	AT3G12500.1	ATHCHIB, PR3, PR-3, CHI-B, B-CHI, HCHIB	GO:0004568 chitinase activity	GO:0005975 carbohydrate metabolic process	GO:0005773 vacuole
						GO:0008061 chitin binding	GO:0006032 chitin catabolic process	
						GO:0016787 hydrolase activity	GO:0000272 polysaccharide catabolic process	GO:0005829 cytosol
						GO:0016798 hydrolase activity, acting on glycosyl bonds	GO:0006952 defense response	
							GO:0008152 metabolic process	
							GO:0009626 plant-type hypersensitive response	
							GO:0016998 cell wall macromolecule catabolic process	
							GO:0031640 killing of cells of other organism	
							GO:0046686 response to cadmium ion	
							GO:0050832 defense response to fungus	
SCS8SD1057C12.g	ScCCR2	S1/S3	Sobic.002G249900.1	AT1G76470.1		GO:0003824 catalytic activity		
						GO:0050662 coenzyme binding		
						GO:0004497 monooxygenase activity		
						GO:0005506 iron ion binding		
						GO:0016491 oxidoreductase activity	GO:0009813 flavonoid biosynthetic process	GO:0016020 membrane
						GO:0016705 oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	GO:0055114 oxidation-reduction process	GO:0005783 endoplasmic reticulum
						GO:0016711 flavonoid 3'-monooxygenase activity	GO:0009733 response to auxin	GO:0005789 endoplasmic reticulum membrane
						GO:0020037 heme binding	GO:0044550 secondary metabolite biosynthetic process	GO:0016021 integral component of membrane
						GO:0046872 metal ion binding		
						GO:0004650 polygalacturonase activity	GO:0005975 carbohydrate metabolic process	GO:0005576 extracellular region
						GO:0016787 hydrolase activity	GO:0008152 metabolic process	GO:0005773 vacuole
						GO:0016798 hydrolase activity, acting on glycosyl bonds	GO:0071555 cell wall organization	
						GO:0016829 lyase activity		
						GO:0042409 caffeoyl-CoA O-methyltransferase activity	GO:0009699 phenylpropanoid biosynthetic process	
						GO:0008168 methyltransferase activity	GO:0009805 coumarin biosynthetic process	
						GO:0008171 O-methyltransferase activity	GO:0009809 lignin biosynthetic process	
						GO:0016740 transferase activity	GO:0032259 methylation	
						GO:0046872 metal ion binding	GO:0046686 response to cadmium ion	
						GO:0005515 protein binding		
						GO:0016760 cellulose synthase (UDP-forming) activity	GO:0030244 cellulose biosynthetic process	GO:0005794 Golgi apparatus
						GO:0000977 RNA polymerase II regulatory region sequence-specific DNA binding	GO:0006952 defense response	GO:0005886 plasma membrane
						GO:0016740 transferase activity	GO:0009833 plant-type primary cell wall biogenesis	GO:0016020 membrane
						GO:0016757 transferase activity, transferring glycosyl groups	GO:0009834 plant-type secondary cell wall biogenesis	GO:0005768 endosome
						GO:0046872 metal ion binding	GO:0071555 cell wall organization	GO:0005802 trans-Golgi network
						GO:0008417 fucosyltransferase activity		GO:0009506 plasmodesma
						GO:0016740 transferase activity		GO:0016021 integral component of membrane
						GO:0016757 transferase activity, transferring glycosyl groups		
						GO:0046920 alpha-(1->3)-fucosyltransferase activity	GO:0036065 fucosylation	GO:0016020 membrane
							GO:0006486 protein glycosylation	GO:0000139 Golgi membrane
							GO:0010493 Lewis a epitope biosynthetic process	GO:0005794 Golgi apparatus
							GO:0071555 cell wall organization	GO:0016021 integral component of membrane
								GO:0032580 Golgi cisterna membrane
SCQSRT1034C08.g	ScFUT1	S1/S4	Sobic.008G053600.1	AT1G71990.1	FUT13, ATFUT13, FUCTC, FT4-M, ATFT4			

APÊNDICE D. Conclusão.

SCMCSB1114E08.g	ScFLA2	S4	Sobic.001G521850.1	AT2G45470.1	AGP8, FLA8			GO:0005886 plasma membrane GO:0031225 anchored component of membrane GO:0009505 plant-type cell wall GO:0016020 membrane GO:0046658 anchored component of plasma membrane GO:0048046 apoplast
SCVPRZ3027D06.g	ScEGase4	S4	Sobic.004G244600.1	AT2G32990.1	GH9B8	GO:0003824 catalytic activity GO:0004553 hydrolase activity, hydrolyzing O-glycosyl compounds GO:0008810 cellulase activity GO:0016787 hydrolase activity GO:0016798 hydrolase activity, acting on glycosyl bonds	GO:0005975 carbohydrate metabolic process GO:0000272 polysaccharide catabolic process GO:0008152 metabolic process GO:0030245 cellulose catabolic process GO:0071555 cell wall organization	GO:0005576 extracellular region
SCCCLB1025F10.g	ScGT47 1	S4	Sobic.009G220200.1	AT1G27440.1	GUT2, 10, ATGUT1	GO:0016757 transferase activity, transferring glycosyl groups GO:0016740 transferase activity GO:0042285 xylosyltransferase activity GO:0080116 glucuronoxylan glucuronosyltransferase activity	GO:0006486 protein glycosylation GO:0009834 plant-type secondary cell wall biogenesis GO:0010417 glucuronoxylan biosynthetic process GO:0071555 cell wall organization	GO:0000139 Golgi membrane GO:0005794 Golgi apparatus GO:0016020 membrane GO:0016021 integral component of membrane
SCJFLR1017B11.g	ScPAL2	S1/S4	Sobic.004G220300.1	AT3G53260.1	PAL2	GO:0005515 protein binding GO:0045548 phenylalanine ammonia-lyase activity GO:0003824 catalytic activity GO:0016829 lyase activity GO:0016841 ammonia-lyase activity	GO:0006559 L-phenylalanine catabolic process GO:0006979 response to oxidative stress GO:0009698 phenylpropanoid metabolic process GO:0009699 phenylpropanoid biosynthetic process GO:0009800 cinnamic acid biosynthetic process GO:0080167 response to karrikin	GO:0005737 cytoplasm GO:0005829 cytosol
SCCCCL3140B03.g	ScXI1	S4	Sobic.002G409800.1	AT5G57655.2		GO:0009045 xylose isomerase activity GO:0016853 isomerase activity GO:0046872 metal ion binding	GO:0005975 carbohydrate metabolic process GO:0042732 D-xylose metabolic process	GO:0005783 endoplasmic reticulum GO:0005737 cytoplasm GO:0005773 vacuole GO:0005774 vacuolar membrane GO:0005794 Golgi apparatus GO:0016020 membrane

APÊNDICE E. Termos GOs para componente celular, processo biológico e função molecular para os fatores de transcrição anotados em *A. thaliana* com base nas sequências completas de *S. bicolor*.

SAS	Identificação	Segmento	<i>S. bicolor</i> Hit	<i>A. thaliana</i> Hit	Símbolo	Cellular Component	GO	Molecular Function
SCVPRZ2042G02.g	ScMYBR1	S1	Sobic.004G222800.1	AT1G60670.2				
SCCCCL3001D02.g	ScNAC2	S1	Sobic.001G386000.1	AT5G18270.1	ANAC087	GO:0005634 nucleus GO:0005737 cytoplasm	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated GO:0009738 abscisic acid-activated signaling pathway	GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0003677 DNA binding GO:000976 transcription regulatory region sequence-specific DNA binding GO:0043565 sequence-specific DNA binding GO:0005515 protein binding
SCVPLB1018G11.g	ScbZIP10	S1	Sobic.010G081800.2	AT1G45249.1	ABF2, AREB1, ATAREB1	GO:0005634 nucleus		
SCCCLR1C05B03.g	ScMYB2	S1	Sobic.002G279100.1	AT4G38620.1	ATMYB4, MYB4	GO:0005634 nucleus	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated GO:0009751 response to salicylic acid GO:0009753 response to jasmonic acid GO:0010224 response to UV-B GO:0045892 negative regulation of transcription, DNA-templated GO:1903086 negative regulation of sinapate ester biosynthetic process GO:2000762 regulation of phenylpropanoid metabolic process	GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0005515 protein binding
SCVPRZ2039D09.g	ScDBP11	S1	Sobic.001G406300.1	AT1G07160.1			GO:0006470 protein dephosphorylation	GO:0003824 catalytic activity GO:0004721 phosphoprotein phosphatase activity GO:0004722 protein serine/threonine phosphatase activity GO:0016787 hydrolase activity
SCSGFL4C07B02.g	ScABI2	S1	Sobic.010G229000.1	AT1G30330.2	ARF6	GO:0005634 nucleus	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated GO:0009725 response to hormone GO:0009733 response to auxin GO:0009734 auxin-activated signaling pathway GO:0009908 flower development	GO:0046872 metal ion binding GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0005515 protein binding
SCMCSB1109C11.g	ScNAC8	S1	Sobic.010G111800.1	AT1G25580.1	NAC008, SOG1	GO:0005634 nucleus	GO:0000077 DNA damage checkpoint GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated GO:0010332 response to gamma radiation GO:0040020 regulation of meiotic nuclear division	GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0043565 sequence-specific DNA binding
SCSFLR2009G10.g	ScSET4	S2/S4	Sobic.003G241700.2	AT4G27910.1	ATX4, SDG16	GO:0005634 nucleus	GO:0016569 covalent chromatin modification GO:0032259 methylation GO:0034968 histone lysine methylation	GO:0008168 methyltransferase activity GO:0008270 zinc ion binding GO:0016740 transferase activity GO:0018024 histone-lysine N-methyltransferase activity

APÊNDICE E. Continuação.

								GO:0046872 metal ion binding
SCCCRZ2001A12.g	ScbHLH4	S2	Sobic.010G145200.1	AT1G29950.1	SACL3	GO:0005634 nucleus	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated	GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0046983 protein dimerization activity
SCJFRZ2010A05.g	ScHB6	S2	Sobic.009G159900.1	AT5G02030.1	LSN, PNY, HB-6, BLR, RPL, BLH9, VAN	GO:0005634 nucleus GO:0005829 cytosol	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated GO:0007389 pattern specification process GO:0010051 xylem and phloem pattern formation GO:0010076 maintenance of floral meristem identity GO:0010077 maintenance of inflorescence meristem identity GO:0010089 xylem development GO:0010154 fruit development GO:0010223 secondary shoot formation GO:0010228 vegetative to reproductive phase transition of meristem GO:0045892 negative regulation of transcription, DNA-templated GO:0048367 shoot system development GO:0048457 floral whorl morphogenesis GO:0080006 internode patterning GO:1905393 plant organ formation	GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0005515 protein binding GO:0043565 sequence-specific DNA binding
SCSFLR2016F09.g	ScbZIP8	S2	Sobic.001G380700.3	AT2G40620.1	BZIP18	GO:0005654 nucleoplasm GO:0005737 cytoplasm	GO:0006355 regulation of transcription, DNA-templated	GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0005515 protein binding GO:0043565 sequence-specific DNA binding
SCEZLR1009A01.g	ScARF3	S2/S3	Sobic.003G411900.1	AT5G62000.1	ARF1-BP, ARF2, HSS, ORE14	GO:0005634 nucleus	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated GO:0008285 negative regulation of cell proliferation GO:0009725 response to hormone GO:0009734 auxin-activated signaling pathway GO:0009737 response to abscisic acid GO:0009911 positive regulation of flower development GO:0010047 fruit dehiscence GO:0010150 leaf senescence GO:0010227 floral organ abscission GO:0045892 negative regulation of transcription, DNA-templated GO:0048481 plant ovule development	GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0005515 protein binding GO:0043565 sequence-specific DNA binding
SCEQRT2025D01.g	ScJM1	S2	Sobic.001G284400.1	AT5G46910.1		GO:0005634 nucleus	GO:0006355 regulation of transcription, DNA-templated GO:0010200 response to chitin	GO:0003700 transcription factor activity, sequence-specific DNA binding

APÊNDICE E. Continuação.

SCBFRZ2016F05.g	ScGRAS1	S2	Sobic.006G172900.1	AT4G00150.1	ATHAM3, HAM3, LOM3, SCL6-IV	GO:0005634 nucleus	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated GO:0007623 circadian rhythm GO:0030154 cell differentiation GO:0048768 root hair cell tip growth GO:0051301 cell division	GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0005515 protein binding GO:0043565 sequence-specific DNA binding
SCEPCL6029A11.g	ScC2H5	S2	Sobic.007G225100.1	AT1G22310.1	MBD8, ATM8D8	GO:0005634 nucleus	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated	GO:0003677 DNA binding GO:0008327 methyl-CpG binding
SCCCFL2002D10.g	ScNAC3	S2/S4	Sobic.003G035100.1	AT4G29230.1	anac075, NAC075	GO:0005634 nucleus	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated	GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0044212 transcription regulatory region DNA binding
SCCCCL7038B02.g	ScPHD1	S2	Sobic.005G034700.1	AT5G35210.2	PTM	GO:0005634 nucleus GO:0009507 chloroplast GO:0009527 plastid outer membrane GO:0009536 plastid GO:0009707 chloroplast outer membrane GO:0009941 chloroplast envelope GO:0016020 membrane GO:0016021 integral component of membrane	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated GO:0010019 chloroplast-nucleus signaling pathway GO:0045893 positive regulation of transcription, DNA-templated	GO:0000979 RNA polymerase II core promoter sequence-specific DNA binding GO:0005515 protein binding GO:0046872 metal ion binding
SCSFRT2072F01.g	ScMYB9	S3	Sobic.004G303600.1	AT5G52260.1	AtMYB19, MYB19	GO:0005634 nucleus	GO:0006355 regulation of transcription, DNA-templated	GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding
SCQGLR2025E03.g	ScWRKY3	S3	Sobic.001G332500.1	AT2G03340.1	WRKY3	GO:0005634 nucleus	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated	GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0043565 sequence-specific DNA binding
SCCCCL4014B04.g	ScC2H2	S3	Sobic.003G370700.2	AT1G34370.2	AtSTOP1, STOP1	GO:0005634 nucleus	GO:0006355 regulation of transcription, DNA-templated GO:0010044 response to aluminum ion GO:0010447 response to acidic pH GO:0006351 transcription, DNA-templated	GO:0003676 nucleic acid binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0046872 metal ion binding
SCCCCL2001C11.b	ScSET2	S3/S4	Sobic.006G080300.1	AT1G76710.1	ASHH1, SDG26	GO:0005634 nucleus GO:0005694 chromosome GO:0000775 chromosome, centromeric region	GO:0006281 DNA repair GO:0009908 flower development GO:0016571 histone methylation GO:0010224 response to UV-B GO:0010228 vegetative to reproductive phase transition of meristem GO:0016569 covalent chromatin modification GO:0032259 methylation GO:0034968 histone lysine methylation	GO:0018024 histone-lysine N-methyltransferase activity GO:0008168 methyltransferase activity GO:0016740 transferase activity

APÊNDICE E. Continuação.

SCQGLR2017G12.g	ScDBP6	S3	Sobic.001G309500.1	AT4G38520.2	APD6	GO:0006470 protein dephosphorylation	GO:0003824 catalytic activity GO:0004721 phosphoprotein phosphatase activity GO:0016787 hydrolase activity GO:0043169 cation binding GO:0046872 metal ion binding GO:0004722 protein serine/threonine phosphatase activity	
SCCCLB1024B02.g	ScZIM2	S3	Sobic.002G374100.1	AT1G74950.1	JAZ2, TIFY10B	GO:0005634 nucleus	GO:0006952 defense response GO:1903507 negative regulation of nucleic acid-templated transcription GO:0031347 regulation of defense response GO:2000022 regulation of jasmonic acid mediated signaling pathway GO:0006355 regulation of transcription, DNA-templated GO:0009753 response to jasmonic acid GO:0009611 response to wounding GO:0006351 transcription, DNA-templated	GO:0003714 transcription corepressor activity GO:0005515 protein binding
SCCCST3005F09.g	ScNAC7	S4	Sobic.003G251800.1	AT4G28500.1	ANAC073, NAC073, SND2	GO:0005634 nucleus	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated GO:0045893 positive regulation of transcription, DNA-templated GO:2000652 regulation of secondary cell wall biogenesis	GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0043565 sequence-specific DNA binding GO:0044212 transcription regulatory region DNA binding
SCCCLR1022F12.g	ScARF1	S4	Sobic.005G132000.1	AT5G62000.1	ARF1-BP, ARF2, HSS, ORE14	GO:0005634 nucleus	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated GO:0008285 negative regulation of cell proliferation GO:0009725 response to hormone GO:0009734 auxin-activated signaling pathway GO:0009737 response to abscisic acid GO:0009911 positive regulation of flower development GO:0010047 fruit dehiscence GO:0010150 leaf senescence GO:0010227 floral organ abscission GO:0045892 negative regulation of transcription, DNA-templated GO:0048481 plant ovule development	GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0005515 protein binding GO:0043565 sequence-specific DNA binding
SCCCLR1C05E08.g	ScCOL1	S4	Sobic.004G208400.1	AT1G06040.1	BBX24, STO	GO:0005634 nucleus GO:0005622 intracellular	GO:0006355 regulation of transcription, DNA-templated GO:0006351 transcription, DNA-templated GO:0009416 response to light stimulus GO:0010228 vegetative to reproductive phase transition of meristem GO:0048573 photoperiodism, flowering GO:0080167 response to karrikin	GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0003712 transcription cofactor activity GO:0005515 protein binding GO:0008270 zinc ion binding GO:0046872 metal ion binding

APÊNDICE E. Conclusão.

SCCCLR1048F08.g	ScJUN2	S4	Sobic.003G124800.1	AT5G23540.1		GO:0000502 proteasome complex GO:0005634 nucleus GO:0005829 cytosol	GO:0006508 proteolysis GO:0009651 response to salt stress GO:0030163 protein catabolic process	GO:0008233 peptidase activity GO:0008237 metallopeptidase activity GO:0016787 hydrolase activity GO:0046872 metal ion binding
SCCCLR1C04H01.g	ScNAC5	S4	Sobic.008G094700.1	AT5G13180.1	ANAC083, NAC083, VNI2	GO:0005634 nucleus	GO:0006351 transcription, DNA-templated GO:0006355 regulation of transcription, DNA-templated GO:0009651 response to salt stress GO:0009737 response to abscisic acid GO:0010089 xylem development GO:0010150 leaf senescence GO:0016032 viral process GO:0045892 negative regulation of transcription, DNA-templated	GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0005515 protein binding
SCACLR1130E07.g	ScbHLH2	S4	Sobic.002G054600.3	AT4G02590.3	UNE12	GO:0005634 nucleus		GO:0003677 DNA binding GO:0003700 transcription factor activity, sequence-specific DNA binding GO:0001046 core promoter sequence-specific DNA binding GO:0001228 transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding GO:0046983 protein dimerization activity
SCCCRZ1004G09.g	ScC2H4	S2/S4	Sobic.008G083000.3	AT5G63280.1		GO:0016020 membrane GO:0016021 integral component of membrane	GO:0008150 biological_process	GO:0043565 sequence-specific DNA binding GO:0046872 metal ion binding
SCCCLR1C04E10.g	ScPHD3	S3/S4	Sobic.001G522600.3	AT2G19260.1				GO:0008270 zinc ion binding GO:0046872 metal ion binding