# UNIVERSIDADE DE SÃO PAULO

## ESCOLA DE ENGENHARIA DE LORENA

GUSTAVO MITSUNORI AOYAGI

# IDENTIFICAÇÃO, ANOTAÇÃO E ANÁLISE FILOGENÉTICA DAS FAMÍLIAS GÊNICAS ENVOLVIDAS NA VIA DE BIOSSÍNTESE DE HEMICELULOSE EM CANA-DE-AÇÚCAR (*Saccharum spp*.)

LORENA

### GUSTAVO MITSUNORI AOYAGI

# IDENTIFICAÇÃO, ANOTAÇÃO E ANÁLISE FILOGENÉTICA DAS FAMÍLIAS GÊNICAS ENVOLVIDAS NA VIA DE BIOSSÍNTESE DE HEMICELULOSE EM CANA-DE-AÇÚCAR (*Saccharum spp*.)

Dissertação apresentada à Escola de Engenharia de Lorena da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-Graduação em Biotecnologia Industrial

Área de Concentração: Conversão de Biomassa

Orientador: Prof. Dr. Elisson Antonio da Costa Romanel

Versão Original

LORENA

2016

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE

> Ficha catalográfica elaborada pelo Sistema Automatizado da Escola de Engenharia de Lorena, com os dados fornecidos pelo(a) autor(a)

Aoyagi, Gustavo Mitsunori Identificação, anotação e análise filogenética das famílias gênicas envolvidas na via de biossíntese de hemicelulose em cana-de-açúcar (Saccharum spp.) / Gustavo Mitsunori Aoyagi; orientador Elisson Antonio da Costa Romanel - Versão Original. - Lorena, 2016. 77 p.

Dissertação (Mestrado em Ciências - Programa de Pós Graduação em Biotecnologia Industrial na Área de Conversão de Biomassa) - Escola de Engenharia de Lorena da Universidade de São Paulo. 2016 Orientador: Elisson Antonio da Costa Romanel

1. Hemicelulose. 2. Glicosiltransferase. 3. Açúcar nucleotídeo. 4. Cana-de-açúcar. 5. Bioinformática. I. Título. II. Romanel, Elisson Antonio da Costa, orient.

#### **RESUMO**

AOYAGI, G. M. Identificação, anotação e análise filogenética das famílias gênicas envolvidas na via de biossíntese de hemicelulose em cana-de-açúcar (*Saccharum spp.*). 2016. 77p. Dissertação (Mestrado em Ciências) – Universidade de São Paulo, Escola de Engenharia de Lorena, Lorena, 2016.

A parede celular de plantas é formada basicamente por celulose, hemicelulose e lignina. A formação dos polímeros de hemicelulose depende do suprimento de precursores chamados de açúcares-nucleotídeos. A biossíntese das diferentes estruturas de hemicelulose da parede celular envolve a participação de enzimas pertencentes às famílias das glicosiltransferases (GTs). Estudos feitos em Arabidopsis thaliana, Brachypodium distachyon, Oryza sativa (arroz) e Zea mays (milho) auxiliaram na descoberta de 11 enzimas da via de interconversão nucleotídeo-açúcar e de enzimas da família das glicosiltransferases (GTs), como as GT2, GT8, GT43, GT47, GT61 e GT75, envolvidas na biossíntese de hemicelulose. O presente trabalho visa a identificação de genes da via de biossíntese de hemicelulose da parede celular de cana-de-açúcar (Saccharum spp.) e análise filogenética entre Arabidopsis thaliana (planta modelo de eudicotiledôneas), Oryza sativa, Brachypodium distachyon, Zea mays, Sorghum bicolor e Saccharum spp. Foram identificados os genes das famílias GT2, GT8, GT43, GT47, GT61, GT75, CSL, Epimerase e UDPG em cana-de-açúcar a partir da busca em sete bibliotecas de RNA-Seq utilizando as sequências de O. sativa, Z. mays e S. bicolor como referência. Os domínios específicos de cada família gênica foram confirmados através do programa PFAM e consequentemente anotados. A identificação e anotação das sequencias possibilitou a construção de bancos de sequências das famílias envolvidas na biossíntese de hemicelulose para as espécies A. thaliana, B. distachyon, O. sativa, Z. mays e S. bicolor. Foram identificadas para cada espécie, respectivamente, um total de 67, 49, 49, 60 e 56 genes bona fides. O presente trabalho, além da identificação de genes nas diferentes espécies, permitiu a identificação e seleção de 27 genes candidatos envolvidos na biossíntese de hemicelulose em cana-de-açúcar e possivelmente envolvidos na recalcitrância da parede celular nas diferentes bibliotecas de RNA-Seq de cana-de-açúcar

Palavras-Chave: Hemicelulose. Glicosiltransferase. Açúcar-nucleotídeo. Cana-deaçúcar. Bioinformática

#### ABSTRACT

AOYAGI, G. M. Identification, annotation and phylogenetic analysis of gene families involved in hemicellulose biosynthesis pathway in sugarcane (*Saccharum spp.*). 2016. 77p. Dissertation (Master of Science) – Escola de Engenharia de Lorena, Universidade de São Paulo, Lorena, 2016.

The plant cell wall is mainly composed of cellulose, hemicellulose and lignin. The formation of hemicellulose polymers lies on the supply of the so-called sugar-nucleotide precursors. The diverse hemicellulose structures biosynthesis of cell wall involves the participation of enzymes belonging to the families of glycosyltransferases (GTs). Studies in Arabidopsis thaliana, Brachypodium distachyon, Oryza sativa (rice) and Zea mays (corn) aid the discovery of 11 enzymes of the nucleotide sugar interconversion pathway and enzymes of the GT family, as GT2, GT8, GT43, GT47, GT61 e GT75, involved in the hemicelluloses biosynthesis. This study aims to the identification of hemicellulose biosynthesis pathway genes from the cell wall of sugarcane (Saccharum spp.) and phylogenetic analysis of Arabidopsis thaliana (eudicotyledonous plant model), Oryza sativa, Brachypodium distachyon, Zea mays, Sorghum bicolor and Saccharum spp. The genes of the GT2, GT8, GT 43, GT47, GT61, GT75, CSL, epimerase and UDPG families were identified in sugarcane from a search in seven RNA-Seq libraries using the sequences of O. sativa, Z. mays and S. bicolor as reference. The specific domains of each gene family have been confirmed through the PFAM program and consequently noted. The identification and annotation of the sequences enabled the construction of sequences banks of the families involved in hemicellulose biosynthesis in the species A. thaliana, B. distachyon, O. sativa, Z. mays and S. bicolor. It was identified for each species, respectively, a total of 67, 49, 49, 60 and 56 bona fides genes. This work, in addition to the identification of genes in different species, allowed the identification and selection of 27 candidate genes involved in the biosynthesis of hemicelluloses in sugarcane and possibly involved in cell wall recalcitrance in the different sugarcane RNA-Seq libraries.

Keywords: Hemicellulose. Glycosyltransferases. Sugar-nucleotide. Sugarcane. Bioinformatics

### LISTA DE FIGURAS

Figura 1 - Modelo estrutural da parede celular de plantas. LM: lamela média; P: parede
primária; S1+S2+S3: parede secundária9
Figura 2 - Representação simplificada da parede celular tipo I e tipo II. A parede está
sendo ilustrada com corte transversal11
Figura 3 - Representação da estrutura de xilanas
Figura 4 - Representação da estrutura de xiloglucana (XyG) de plantas14
Figura 5 - Representação da estrutura das glucanas de cadeia mista (MLG)
Figura 6 - Representação das estruturas de heteromananas16
Figura 7 - Diagrama simplificado demonstrando os subtipos de homologia (eventos de
ortologia e paralogia)
Figura 8 - Esquema das 11 enzimas da via de interconversão açúcar-nucleotídeo
envolvidas na síntese de hemiceluloses da parede celular de plantas
Figura 9 - Famílias gênicas envolvidas da síntese dos polissacarídeos da parede25
Figura 10 - Árvore filogenética das proteínas do clado A com domínio Epimerase das
espécies A. thaliana, B. distachyon, O. sativa, S. bicolor e Z. mays
Figura 11 - Árvores filogenéticas das proteínas das espécies A. thaliana, B. distachyon, O.
sativa, S. bicolor e Z. mays das famílias gênicas GT2, UGD, GT43 e GT75
Figura 12 - Árvore filogenética das proteínas com domínio GT8 das espécies A. thaliana,
B. distachyon, O. sativa, S. bicolor e Z. mays
Figura 13 - Árvore filogenética das proteínas com domínio GT47 das espécies A. thaliana,
B. distachyon, O. sativa, S. bicolor e Z. mays
Figura 14 - Árvores filogenéticas das proteínas das espécies A. thaliana, B. distachyon, O.
sativa, S. bicolor e Z. mays das famílias gênicas CSL e GT6155
Figura 15 - Representação de uma árvore filogenética de representantes da subfamília
Panicoideae
Figura 16 - Alinhamento de sequências contendo o domínio epimerase e UGD dos
diferentes bancos de cana-de-açúcar com genes bona fides de S. bicolor
Figura 17 - Alinhamento de sequências GT75 dos diferentes bancos de cana-de-açúcar com
o gene <i>bona fide</i> Sobic.001G173300 de <i>S. bicolor</i>
Figura 18 - Alinhamento de sequências GT43 dos diferentes bancos de cana-de-açúcar com
o gene bona fide Sobic.009G229300 de S. bicolor

Figura 19 -	Alinhamento	de sequências	GT61 dos	s diferentes	bancos de	e cana-de-açú	car com
o gene bon	a fide Sobic.01	0G144400 de	S. bicolo	r			60

# LISTA DE TABELAS

Tabela 1 - Composição das paredes celulares dos diferentes tipos de grupos de plantas
(Eudicotiledôneas e Monocotiledôneas comelinídeas e não comelinídeas) 12
Tabela 2 - Composição e características agronômicas da cana-de-açúcar, cana energia e
sorgo
Tabela 3 - Número de genes das famílias gênicas relacionadas com a biossíntese de
hemicelulose da parede celular
Tabela 4 - Número de sequências das famílias gênicas relacionadas com a biossíntese de
hemicelulose em distintas bibliotecas de cana-de-açúcar
Tabela 5 - Quantidade de genes bona fides das enzimas de interconversão açúcar
nucleotídeo nas diferentes espécies
Tabela 6 - Número de genes candidatos relacionados com a biossíntese de hemicelulose e
possivelmente envolvidas com a recalcitrância da parede celular em distintas bibliotecas de
cana-de-açúcar

# SUMÁRIO

1	INT	RODUÇÃO9
	1.1	Características e composição da parede celular9
	1.2	Estrutura das principais classes de hemiceluloses
	1.3	Bioinformática, filogenia, genômica evolutiva e comparativa
	1.4 açúcai	Importância econômica, estrutura da parede celular e genética da cana-de-
	1.5	Genética da biossíntese de hemicelulose
	1.5.	A via de interconversão açúcar-nucleotídeo
	1.5.	2 Biossíntese de xilana
	1.5.	Biossíntese de xiloglucana
	1.5.	4 Biossíntese de MLG
	1.5.	5 Biossíntese de heteromanana
	1.6	Justificativa
2	OB.	IETIVOS
	2.1	Objetivos Gerais
	2.2	Objetivos específicos
3	MA	TERIAL E MÉTODOS
	3.1 thalia	Identificação dos genes da via de biossíntese de hemicelulose em <i>Arabidopsis</i> <i>na</i> e gramíneas
	3.2 biossíi	Anotação das sequências gênicas das famílias relacionadas ao processo de ntese de hemicelulose
	3.3	Identificação e anotação dos genes em cana-de-açúcar (Saccharum sp.) 34
	3.4	Alinhamento e análise filogenética
4	RES	SULTADOS E DISCUSSÃO
	4.1	Anotação das famílias gênicas envolvidas na biossíntese de hemicelulose 36
	4.2 Arabio	Filogenia e Identificação dos genes <i>bona fides</i> da biossíntese de hemicelulose em lopsis e gramíneas
	4.2.	1 Análise dos genes envolvidos na via de interconversão açúcar-nucleotídeo38
	4.2.	2 Análise dos genes da família GT8 39
	4.2.	Análise dos genes da família GT43 40
	4.2.	4 Análise dos genes da família GT47 41
	4.2.	5 Análise dos genes da família GT61
	4.2.	6 Análise dos genes da família GT75 44
	4.2.	7 Análise dos genes da família GT2 e CSL

4.3	Identificação de genes candidatos envolvidos na biossíntese de hemicelulose	e em
cana	a-de-açúcar e possivelmente envolvidos na recalcitrância da parede celular	47
5. C	ONCLUSÃO	61
REFE	RÊNCIAS	62

#### 1 INTRODUÇÃO

#### 1.1 Características e composição da parede celular

A parede celular é uma estrutura muito complexa que possui diversas funções. Além de proporcionar às células rigidez mecânica, define a sua morfologia, controla a expansão celular e transporte intercelular, protege a célula contra uma grande quantidade de organismos potencialmente patogênicos e predadores e também participa na comunicação intercelular (DHUGGA, 2001). Como muitas destas funções são exigidas ao mesmo tempo em vários estágios de desenvolvimento do vegetal, a parede celular possui simultaneamente, grande flexibilidade com um máximo de resistência, devido a um sofisticado controle metabólico, com regiões de extrema complexidade (CARPITA; GIBEAUT, 1993; GIBEAUT; CARPITA, 1994). Morfologicamente, a parede celular é dividida em três regiões distintas, a lamela média, parede primária e parede secundária (Figura 1). A lamela média é a região de interseção entre duas células adjacentes. Esta camada intercelular é rica em substâncias pécticas e lignina (BUCHANAN; GRUISSEM; JONES, 2000).

Figura 1 - Modelo estrutural da parede celular de plantas. LM: lamela média; P: parede primária; S1+S2+S3: parede secundária



Fonte: Adaptado de (FENGEL; WEGENER, 1989)

A parede celular é uma estrutura complexa, diversa e dinâmica, sofrendo modificações ao longo de processos de divisão celular, crescimento e diferenciação. Os tipos de polissacarídeos presentes na parede celular variam de acordo com a espécie, tipo celular e localização, estágio de desenvolvimento e de respostas a estresses bióticos ou abióticos (CARPITA; TIERNEY; CAMPBELL, 2001) A parede primária vegetal é composta por três estruturas independentes, as microfibrilas de celulose (~30%),

polissacarídeos não-celulósicos como as hemiceluloses (~30%) e pectinas (~30%) e ainda proteínas estruturais (~10%). A fina parede primária (100 nm ou menos) e altamente hidratada (~60% de peso úmido) é produzida por células em expansão podendo assim, se alongar com crescimento difuso ou orientado (CARPITA; GIBEAUT, 1993; HAYASHI, 1989; O'NEIL; YORK, 2003).

A parede primária de eudicotiledôneas, gimnospermas e monocotiledôneas nãocomelinídeas (conhecida também como parede celular tipo I) consiste de uma rede celulósica inserida em uma matriz complexa de polissacarídeos, onde as xiloglucanas e pectinas são os mais abundantes (CARPITA; GIBEAUT, 1993; MCCAN; ROBERTS, 1994). A parede primária das Poales e monocotiledôneas comelinídeas relacionadas (conhecida também como parede celular tipo II) é organizada do mesmo modo, exceto que os polissacarídeos predominantes são as glucuronoarabinoxilanas (GAX) e glucanas de cadeia mistas ((1,3;1,4)-β-D-glucanas). Os níveis de pectina e xiloglucana são relativamente baixos (CARPITA; GIBEAUT, 1993; MCCAN; ROBERTS, 1994). A Figura 2 ilustra a diferença entre a parede celular tipo I e tipo II. A maneira como os polímeros estão arranjados na parede de cada tecido pode variar consideravelmente. Esse arranjo de polímeros, chamado de arquitetura da parede celular, está intimamente relacionado com a estrutura desses carboidratos (SOUZA et al., 2013).

Quando o crescimento celular é finalizado, inicia-se o processo de diferenciação celular com início da formação da parede secundária. A transição da parede celular primária para formação da parede secundária é marcada pela redução da síntese de pectina e um grande aumento na síntese de celulose, hemiceluloses e lignina (BOUDET et al., 2003).Durante o desenvolvimento da parede celular secundária, há uma redução da flexibilidade e consequente aumento da força da célula. A celulose e a matriz de polissacarídeos com baixo grau de substituição na cadeia principal, como as heteroxilanas e heteromananas, são depositadas em um padrão altamente ordenado durante o desenvolvimento da parede celular, tornando-a altamente hidrofóbica. Ligninas hidrofóbicas encobrem as microfibrilas de celulose e a matriz de polissacarídeos, podendo ser covalentemente complexada com os polissacarídeos da parede. Essa associação lignina-polissacarídeos (juntamente com o processo de desidratação) são fatores que contribuem para a chamada recalcitrância da parede celular lignificada a digestão enzimática (DOBLIN; PETTOLINO; BACIC, 2010).

Devido ao interesse na biomassa de plantas para produção de biocombustíveis, é importante reconhecer que o grau de recalcitrância não é influenciado apenas pela química da parede celular, mas também com a anatomia estrutural das células e tecidos.

Figura 2 - Representação simplificada da parede celular tipo I e tipo II. A parede está sendo ilustrada com corte transversal. Em vermelho está destacado as principais diferenças. As Poales e monocotiledôneas comelinídeas possuem como principal hemicelulose da parede primária a glucuronoarabionxilana (GAX), juntamente com as glucanas de cadeia mista (MLG), com baixo conteúdo de pectina. Eudicotiledôneas, gimnospermas e monocotiledôneas não-comelinídeas possuem a xiloglucana como principal hemicelulose da parede primária e uma matriz contendo pectina e proteínas estruturais. Note que na parede secundária a MLG e pectinas se tornam componentes minoritários (reduzem em massa) e há um aumento na quantidade de lignina.



Fonte: Adaptado de (FERREIRA et al., 2013)

A parede celular secundária contém tipicamente três camadas distintas, S1, S2 e S3 (Figura 1), as quais são diferenciadas, em espessura, composição e orientação das microfibrilas de celulose. A camada exterior S1 é a mais fina e representa apenas 5-10% da espessura total da parede celular. Nesta camada as microfibrilas de celulose formam um ângulo de 60-80° em relação ao eixo da célula. A camada intermediária S2 é a mais espessa e representa 75-85% da espessura total da parede celular, e a orientação das microfibrilas forma um ângulo de 5-30° em relação ao eixo da célula. As características relacionadas à espessura e à orientação das microfibrilas influenciam diretamente as propriedades físicas e mecânicas da célula. Já a camada S3 é relativamente fina, e forma um ângulo de 60-90° em relação ao eixo da célula (FENGEL; WEGENER, 1989; EMONS; MULDER, 2000). A Tabela 1, demonstra a diferença de composição da parede celular (primária e secundária) entre gramíneas e eudicotiledôneas.

Crunos do plantas	Parede primária		Parede Secundária		
Gi upos de plantas	Fase fibrilar	Fase matricial*	Fase fibrilar	Fase matricial*	
	Celulose	Polissacarídeos		GX	
		pécticos	Celulose		
Fudicatiladônas		Xiloglucana			
Euulcotheuoneas		Heteroxilana		Glucomanana	
		Galactoglucomanas			
		(minoritário)			
Monocotiledôneas:					
	Celulose	Polissacarídeos		GX	
		pécticos	Calulosa		
Comolinídoos		Xiloglucana			
Comennueas		Heteroxilana	Celulose	Glucomanana	
		Galactoglucomanas			
		(minoritário)			
	Celulose	GAX		GAX	
Não-Comelinídeas		Polissacarídeos			
(gramíneas)		pécticos	Celulose		
(Stummens)		Xiloglucana		Galactoglucomanas	
		MLG		(minoritário)	

Tabela 1 - Composição das paredes celulares dos diferentes tipos de grupos de plantas (Eudicotiledôneas e Monocotiledôneas comelinídeas e não comelinídeas). \*Em ordem de abundância.

Os avanços tecnológicos são enormes no sentido de elucidar os mecanismos de formação da parede celular, envolvendo um grande número de genes, relacionados à biogênese, organização e também a mudanças estruturais. Na planta modelo, *Arabidopsis thaliana* (eudicotiledônea), aproximadamente 15% do seu genoma está relacionado ao processo de formação e modificação da parede celular, revelando a complexidade da estrutura e formação da parede celular das plantas (CARPITA; TIERNEY; CAMPBELL, 2001).

#### 1.2 Estrutura das principais classes de hemiceluloses

As hemiceluloses são classes de polissacarídeos não-celulósicos que contém substituintes heterogêneos interligadas às microfibrilas de celulose. As hemiceluloses são divididas em quatro (4) classes principais: Heteroxilanas, Xiloglucanas, Glucanas de cadeia mista (MLG) e Mananas (LEROUXEL et al., 2006; PAULY et al., 2013).

## (1) Heteroxilanas

Heteroxilanas são hemiceluloses majoritárias na parede secundária de eudicotiledôneas e em monocotiledôneas comelinídeas. Elas são essenciais no

Fonte: Adaptado de (DOBLIN; PETTOLINO; BACIC, 2010)

desenvolvimento de tecidos com alto conteúdo de parede secundária e conferindo fortalecimento a esses tecidos (PAULY et al., 2013).

A cadeia principal das heteroxilana (Figura 3) consiste de xiloses (Xylp) ligadas linearmente por ligações do tipo  $\beta$ -1,4 na qual aproximadamente 10% da cadeia sofre substituição no C(O)3 e C(O)2 com resíduos de L-arabinofuranosil (Araf), ácido-α-D glucurônico (GlcpA e MeGlcpA - ácido 4-O-metil-glucurônico) ou uma combinação dos dois arabinoxilanas (AXs), (metil)glucuronoxilanas formando (GXs) e glucuronoarabinoxilanas (GAXs), respectivamente ((KEPPLER; SHOWALTER, 2010; EBRINGEROVÁ, 2005; O'NEIL; YORK, 2003). A cadeia principal de Xylp pode, ainda, ser acetilada nas posições C(O)3 e C(O)2 e em alguns casos em ambas (YORK; O'NEILL, 2008). Acredita-se que os substituintes de Araf na cadeia de xilana promovem o impedimento das ligações de hidrogênio entre moléculas adjacentes de AXs ou com as microfibrilas de celulose, levando uma redução na hidrofobicidade (URAHARA et al., 2004). O resíduo de Araf em GAX ou AX de gramíneas pode ser decorado pela adição de substituintes  $\beta$ -1,2 Xylp ou  $\alpha$ -1,3 Araf (FAIK, 2010). Além disso, substituintes de feruloil e coumaroil presentes na posição C-5 da cadeia lateral de Araf pode sofrer esterificação e levando a ligações covalentes entre a heteroxilana e a lignina em gramíneas (FAIK, 2010).





Fonte: Adaptado de (PAULY et al., 2013).

GXs são os polissacarídeos não-celulósicos mais abundantes na parede secundária de eudicotiledôneas, enquanto que as GAXs são componentes minoritários da parede secundária de coníferas (*softwoods*). Em comparação, AXs são as xilanas predominantes na parede celular das gramíneas, com as GAXs sendo menos abundantes (YORK; O'NEILL, 2008). Além da cadeia lateral, foi observado a existência de um tetrassacarídeo de  $\beta$ -Xyl-1-3- $\alpha$ -Rha-1-2- $\alpha$ -GalA-1-4- $\beta$ -Xyl no terminal redutor de GX (Figura 3) em espécies de eudicotiledôneas e gimnospermas (PENA et al., 2007; YORK; O'NEILL, 2008). Acredita-se que desempenha um papel na biossíntese de xilana. Entretanto, tal tetrassacarídeo não foi observado em xilanas de gramíneas (PAULY et al., 2013).

#### (2) Xiloglucanas (XyG)

XyG é a hemicelulose mais abundante na parede primária de eudicotiledôneas com 20-30% de proporção massa/massa (w/w) (SCHELLER; ULVSKOV, 2010). Acredita-se que as XyGs formam ligações cruzadas entre as microfibrilas de celulose formando uma forte rede de XyG-celulose, funcionando como o principal componente da parede celular primária (HAYASHI, 1989; PAULY et al., 1999; SOMERVILLE et al., 2004). XyGs podem, também, se ligar de forma covalente com pectinas. Por isso, acredita-se que o metabolismo de XyG desempenha papel importante na elongação celular (PAULY et al., 2001). Paralelamente, as XyGs podem desempenhar papel na prevenção da formação de agregados de celulose microfibrilar (ANDERSON et al., 2010; THOMPSON, 2005) ou permitindo a interação da celulose com outros componentes da matriz celular (HA; APPERLEY; JARVIS, 1997; KEEGSTRA et al., 1973; CAVALIER et al., 2008).

Figura 4 - Representação da estrutura de xiloglucana (XyG) de plantas. fucogalactoXylG é estrutura predominante de eudicotiledôneas e arabinogalactoXyG de monocotiledôneas. As ligações glicosídicas entre os monossacarídeos estão representadas pela sua configuração



Fonte: Adaptado de (PAULY et al., 2013).

Xiloglucanas possuem a cadeia principal constituída de ligações  $\beta$ -1,4 de glucanas, substituídas parcialmente com unidades de xilosil na posição O-6 (LEROUXEL et al., 2006; PAULY et al., 2013). XyG de várias eudicotiledôneas, como *Arabidopsis*, e monocotiledôneas não-gramíneas, normalmente possuem de uma maneira regular três resíduos de xilosil a cada quatro resíduos de glicose da cadeia principal (Figura 4), enquanto que em plantas da ordem das *Poales* (GIBEAUT et al., 2005) e diversos outros clados das monocotiledôneas (HSIEH; HARRIS, 2009), a cadeia principal das XyGs é menos xilosilada com apenas dois resíduos de xilosil a cada quatro ou mais de glicose (Figura 4) (HOFFMAN et al., 2005).

Na maioria das espécies de plantas, os grupos xilosil são substituídos na posição O-2, geralmente com galactosil, galacturonosil, arabinosil ou outros resíduos de glicosil (HOFFMAN et al., 2005; HANTUS et al., 1997). Ainda, os resíduos de galactosil podem ser decorados, na posição O-2, com fucosil e/ou substituintes de O-acetil (LEROUXEL et al., 2006; PAULY et al., 2013). As xiloglucanas são sintetizadas no complexo de Golgi e então, transportadas para a parede celular (HAYASHI, 1989).

#### (3) Glucanas de cadeia mista (*MLG – Mixed linked glucans*)

Em plantas, as MLG são restritivas da parede celular das gramíneas (Poaceae). A quantidade de MLG é regulada durante o desenvolvimento sofrendo modificações durante a elongação celular. Há um acúmulo de aproximadamente 20% de massa seca de MLG em tecidos jovens (CARPITA; MCCANN, 2010; CARPITA, 1996). A MLG é então degradada por hidrólise pelas liquenases e  $\beta$ -glucosidades de plantas (KIM; OLEK; CARPITA, 2000; HRMOVA; FINCHER, 2001), concomitantemente com a expansão celular. Devido à redução dessa hemicelulose ao longo do processo de expansão celular, foi proposto que, possivelmente, a MLG seja um polímero de armazenamento de energia no momento de elongação celular (PAULY et al., 2013).



Fonte: Adaptado de (PAULY et al., 2013).

A MLG (1,3; 1,4- $\beta$ -glucana) é um homopolímero de glicose, sem substituintes e sem ramificações, com distribuição aleatória de unidades  $\beta$ -1,4 de celotriosil e celotetraosil conectadas por ligações  $\beta$ -1,3 (Figura 5). A maioria dos segmentos encontrados em MLG consiste de celotrioses e celotetraoses. Entretanto, pode-se encontrar celooligômeros maiores, com resíduos de glucopiranosil variando de 5-20 unidades. Esses celooligômeros correspondem a 10% da cadeia das MLGs (FINCHER, 2009). A razão entre as unidades de celotriose e celotetraose bem como a abundância das MLGs varia consideravelmente entre as diferentes espécies de gramíneas e entre tecidos diferentes (BRENNAN; CLEARY, 2005)

#### (4) Heteromananas

As heteromananas são hemiceluloses encontradas em grandes quantidades na parede celular de gimnospermas. As heteromananas, segundo a sua função biológica, conferem rigidez (RODRÍGUEZ-GACIO et al., 2012; BUCKERIDGE, 2010), além de outros papéis fisiológicos, como sinalização da parede, embriogênese e diferenciação tecidual (GOUBET et al., 2009; RODRÍGUEZ-GACIO et al., 2012; BENOVA-KAKOSOVA et al., 2006).

Baseada na composição da cadeia principal e nos substituintes da cadeia lateral, as heteromananas são agrupadas em quatro classes distintas: mananas, glucomananas, galactomananas, e galactoglucomananas (PAULY et al., 2013; SCHELLER; ULVSKOV, 2010). As mananas e galactomananas possuem a cadeia principal constituída de ligações  $\beta$ -1,4 de manoses (Figura 6); as glucomananas e galactoglucomananas contêm unidades de manose e glicose em sua cadeia principal ligadas pela configuração de  $\beta$ -1,4 (Figura 6). Nos casos das galactomananas e das galactoglucomananas, o resíduo de manosil pode ser substituído com um resíduo de  $\alpha$ -1,6 galactosil (SCHELLER; ULVSKOV, 2010). Além disso, o resíduo de manosil pode sofrer O-acetilação nas posições de O-2 e O-3 da molécula de manose. O alto grau de substituição da cadeia principal está associado com o aumento da solubilidade do polímero (PAULY et al., 2013).

Figura 6 - Representação das estruturas de heteromananas.



Fonte: Adaptado de (PAULY et al., 2013)

#### 1.3 Bioinformática, filogenia, genômica evolutiva e comparativa

A genômica, análise da sequência completa de DNA de um organismo, tem sido de grande importância em estudos biológicos. O genoma de um organismo é de fundamental importância para a compreensão das funções individuais dos genes e das redes de interação entre eles, pois possibilita inferir relações evolutivas podendo, ainda, revelar mecanismos regulatórios até então não estudados (BEVAN; UAUAY, 2013).

A abordagem com sistemas modelos é de importância para o entendimento de processos biológicos a níveis genéticos e moleculares. A Arabidopsis thaliana, primeiro

vegetal superior com o genoma sequenciado, é uma eudicotiledônea da família Brassicaceae que apesar de não ser de importância econômica, é considerada uma planta modelo para estudos em biologia de plantas e tem contribuído para a identificação de genes relacionados com a biossíntese de hemicelulose na parede celular (LIEPMAN et al., 2010; KOORNEEF; MEINKE, 2010; MEINKE et al., 1998). Grande quantidade de recursos genômicos, como banco de dados de ESTs e cDNA, bem como um conjunto de dados de expressão abordando diversas condições ambientais e estágios de desenvolvimento tornaram a Arabidopsis uma excelente referência para realização de genômica comparativa e funcional (SPANNAGL et al., 2011). Desde a liberação do genoma de Arabidopsis no ano 2000 (Arabidopsis Genome Initiative), um número crescente de espécies de plantas vem tendo seu genoma sequenciado, revolucionando nosso entendimento da biologia vegetal e facilitando o melhoramento das espécies de interesse (VARSHNEY et al., 2012; BOLGER et al., 2014, BEVAN; UAUAY, 2013). Novas metodologias biológicas vêm se desenvolvendo devido às tecnologias "ômicas". Primeiramente a análise de DNA (genômica), seguido pelo RNA (transcriptômica), proteínas (proteômica) e pequenos metabólitos (metabolômica) (WANG; WANG; XIA, 2015).

Por sua vez, *Oryza sativa* (arroz), juntamente com *Brachypodium distachyon* são considerados organismos modelos em gramíneas, grupo este que contém espécies de grande importância agronômica como o próprio arroz, milho, trigo e a cana-de-açúcar. Em ambas espécies, o genoma já está sequenciado com alta qualidade e ainda, existe uma rica quantidade de estudos genômicos realizados (CHRISTENSEN et al., 2009; DRAPER et al., 2001; KONISHI et al., 2007; OIKAWA et al., 2010; THE INTERNATIONAL BRACHYPODIUM INITIATIVE, 2010). Devido à importância agronômica de cultivares do grupo das gramíneas, uma estratégia interessante é aplicar o conhecimento que se tem em arroz, Brachypodium como também em Arabidopsis para permitir o entendimento de processos bioquímicos de culturas de interesse econômico e agrícola. Isso permite uma maior precisão em programas de melhoramento genético.

Os recentes avanços no sequenciamento de genomas através, por exemplo, das tecnologias de sequenciamento de nova geração, permitem o estudo de espécies vegetais não consideradas modelos, como a busca de genes de interesse agronômico. A identificação desses genes em diferentes espécies permite um entendimento de como algumas características vegetais de interesse são reguladas, conhecimento este que pode ser utilizado diretamente no melhoramento vegetal (EDWARDS; BATLEY 2010).

O conceito de homologia de sequencias é definida como a relação entre dois genes que divergiram a partir de um ancestral comum (FITCH, 2000). A noção de ortologia versus paralogia foi introduzida por Walter Fitch em 1970 e são termos essênciais para definir relações evolutivas distintas. Genes considerados ortólogos são aqueles que possuem uma relação de homologia na qual descendem de um gene ancestral em comum a partir de um evento de especiação (KOONIN, 2001). Portanto, pode-se dizer que genes ortólogos são genes homólogos de diferentes espécies. Em contraste, genes parálogos são aqueles que dentro do mesmo genoma sofreram duplicação (KOONIN, 2001). Apesar da paralogia ocorrer dentro do mesmo genoma, existem casos de genes parálogos em espécies diferentes. A figura 7 demonstra em um diagrama os eventos de paralogia e ortologia.

Figura 7 - Diagrama simplificado demonstrando os subtipos de homologia (eventos de ortologia e paralogia). Os eventos de especiação levaram a formação das espécies A, B e C. Os genes A1, B1, B2, C1, C2 e C3 descenderam de um gene ancestral comum a partir de eventos evolutivos de especiação e de duplicação gênica. Existem dois eventos de especiação (linhas tracejadas) e dois eventos de duplicação (barras horizontais). Dois genes com o ancestral comum a partir das linhas tracejadas (especiação) são ortólogos. E dois genes com o ancestral comum a partir de barras horizontais (duplicação gênica) são parálogos. Assim, os genes C2 e C3 são parálogos entre si e ortólogos do gene B2. E os três genes são parálogos do gene B1, mas ortólogos ao gene A1.



Fonte: Adaptado de (FITCH, 2000)

A identificação de regiões codificantes de proteínas em sequências de nucleotídeos permanece um desafio na interpretação de genomas (STATES; GISH, 1994). A descoberta de homologia entre uma determinada sequência de ácido nucleico e/ou uma proteína é frequentemente o primeiro indício sobre a função de um gene recém-sequenciado. Como os bancos de dados de DNA e de aminoácidos estão em constante expansão, eles se tornam cada vez mais úteis na análise de novas sequências, pois é cada vez maior a chance de se encontrar alguma homologia (ALTSCHUL et al., 1990). Existe um grande número de *softwares* disponíveis para a busca em bases de dados, porém todos utilizam algum método de análise de similaridade entre sequências para distinguir relações de significado

19

biológico daquelas similaridades ao acaso (NEEDLEMAN; WUNSH, 1970; SELLERS, 1974; SANKOFF; KRUSKAL, 1983; WATERMAN, 1984).

O BLAST (*Basic Local Alignment Search Tool*) é um programa de busca de similaridade de sequências que pode ser usado via interface na *web*, ou como ferramenta individual. Existem vários tipos específicos de blast, que comparam aminoácidos e nucleotídeos com o banco de dados. É um programa heurístico que busca pequenas regiões de similaridade entre duas sequências e tenta iniciar um alinhamento a partir desses *—hot spots*. Além do alinhamento em si, o blast oferece dados estatísticos para ajudar a interpretação do significado biológico desse alinhamento (MCGINNIS; MADDEN, 2004).

O HMMER é outro software de busca de similaridade, que utiliza o método probabilístico das cadeias ocultas de Markov (*Hidden Markov Models, HMM*). Pode ser utilizado na busca em um banco de dados a partir de uma única sequência proteica, um grupo de sequências, ou ainda um perfil HMM, que pode ser criado pelo próprio programa. Inclui alguns algoritmos de busca específicos, como o *jackhmmer* e o *hmmrsearch* (FINN; CLEMENTS; EDDY, 2011).

#### 1.4 Importância econômica, estrutura da parede celular e genética da cana-de-açúcar

A cana-de-açúcar é uma das mais importantes culturas do mundo, sendo a fonte primária de açúcar para alimentos e bebidas e para a produção de biocombustíveis. As dificuldades e limitações associadas ao melhoramento tradicional de cana-de-açúcar, além do grande potencial do uso de engenharia genética, tornam a biotecnologia uma abordagem interessante para o melhoramento de aspectos relevantes da cana-de-açúcar, como conteúdo de açúcar e tolerância a estresse biótico e abiótico (LAKSHMANAN et al., 2005). Ainda, a biotecnologia pode ser utilizada para diversificação da indústria de cana-de-açúcar, criando linhagens que produzam produtos alternativos de alto valor agregados, como compostos farmacêuticos, antibióticos, produtos industriais e açúcares alternativos (WANG et al., 2005; BASNAYAKE et al., 2012; HARRISON et al., 2011; PETRASOVITS et al., 2012; KINKEMA et al., 2014).

Atualmente, o Brasil produz intensivamente bioetanol a partir da cana-de-açúcar (AMORIM et al., 2011), entretanto, está cada vez mais evidente que a produção de bioetanol tanto a partir do amido de milho armazenado nas sementes quanto da sacarose de

cana-de-açúcar armazenada nos colmos, o chamado etanol de primeira geração (1G), não serão suficientes para atender a demanda futura de combustível (SOUZA et al., 2013).

Nesse contexto, diversos grupos de pesquisa estão buscando alternativas eficientes para hidrolisar os polissacarídeos da parede celular do bagaço a fim de fornecer fundamento científico para o desenvolvimento de tecnologia para produção do chamado etanol de segunda geração (2G). As gramíneas com fotossíntese C4 são caracterizadas pela sua alta produtividade e rendimento de biomassa (VAN DER WEIJDE et al., 2013). Assim, a biomassa das gramíneas está entre as matérias-primas consideradas promissoras para produção de biocombustíveis 2G (SIMS et al., 2010). Ademais, a biomassa da cana-de-açúcar pode ser utilizada para produção de bioenergia através da queima do bagaço para gerar eletricidade (AMORIM et al., 2011).

Uma vez que os biocombustíveis lignocelulósicos utilizam-se da parede celular da planta como fonte para geração de açúcares fermentáveis, é importante compreender a composição, arquitetura e a genética da parede celular para desenvolver estratégias para uma degradação eficiente (FERREIRA et al., 2013). A recalcitrância dos materiais lignocelulósicos está relacionada com diversos fatores, incluindo as associações entre a celulose, hemicelulose e a lignina da parede celular, o que dificulta a ação das celulases (HIMMEL et al., 2007).

Descobrir a complexidade estrutural dos polímeros da parede celular, bem como suas interações é essencial para contornar os problemas de recalcitrância do material (SOUZA et al., 2013). Assim, estudos sobre a composição química da parede celular representam uma etapa importante na elucidação dessa complexidade estrutura. Já existem alguns estudos sobre a composição química confinando o fracionamento da parede celular em grandes classes de compostos em cana-de-açúcar. Relata-se que o bagaço da cana-de-açúcar possui 38–43 % de glucanas, 25–32 % de hemicelulose, 17–24 % de lignina e 1.6–7.5 % de extrativos (MASARIN et al., 2011). Outros autores (SANJUÁN et al., 2001) relatam a composição da parede celular sendo de 48.6 % de celulose, 31.1 % hemiceluloses (principalmente xilose e galactose), 19.1 % lignina, e 1.2 % cinzas.

Durante o processamento da cana-de-açúcar, grandes quantidades de bagaço de cana-de açúcar são geradas após a maceração desta para retirada do caldo dos colmos. A quantidade de bagaço disponível varia entre 250 a 300 kg por tonelada de cana-de-açúcar (REIN, 2007). Em geral, o bagaço é queimado para produzir vapor e eletricidade como

parte do esquema de co-geração nas usinas. O bagaço residual é descartado como dejeto industrial de baixo valor. O conteúdo dos componentes no bagaço varia, sendo em média 41,6% celulose, 25,1% hemicelulose e 20,3% lignina (KIM; DAY, 2011). O bagaço oferece muitas vantagens no uso para bioconversão por micro-organismo devido a sua relativa baixa concentração de cinzas (4,8%) comparada a outras matérias-primas como palha de arroz (17,5%) e palha de milho (11%) (PANDEY et al., 2000; GOH et al., 2010; KIM; DAY, 2011).

Espécies do complexo *Saccharum* (cana-de-açúcar) fazem parte da família das gramíneas (Poaceae) que realizam fotossíntese C4, e juntamente com sorgo, milho e *Miscanthus* formam a superfamília Panicoidae as quais possuem importância econômica (KELLOGG, 2001). O uso da fotossíntese C4 pelas plantas, evita a fotorrespiração levando a uma maior eficiência na conversão de energia do que a via C3 utilizada por arroz, trigo e outras gramíneas (ZHU; LONG; ORT, 2008). Assim, a fotossíntese C4 leva a um maior acúmulo de biomassa.

As variedades comerciais de cana-de-açúcar possuem a notável capacidade de armazenar altos níveis de sacarose, chegando a 40% de sua massa seca (FERREIRA et al., 2013). Essas variedades modernas de cana-de-açúcar são derivadas da hibridização interespecífica entre as espécies Saccharum officinarum e Saccharum spontaneum. S. officinarum possui colmos grossos e suculentos com alto conteúdo de sacarose enquanto que a espécie selvagem S. spontaneum possui colmos finos e fibrosos com baixa sacarose, mas com alta tolerância a estresse (GRIVET; ARRUDA, 2002). Essas variedades modernas resultantes da hibridização, são altamente poliplóides e aneuplóides com contribuição desigual dos genomas parentais de S. officinarum (80-90%) e S. spontaneum (10-20%) e 10% são recombinantes das duas espécies (PIPERIDIS; PIPERIDIS; HONT, 2010). Ainda, o genoma dessas variedades é muito maior (cultivar R570, 10,000 Mb e 2n = 115) que milho (4600 Mb, 2n = 20), sorgo (1500 Mb, 2n = 20) ou arroz (775 Mb, 2n = 24) (D'HONT; GLASZMANN, 2001; INTERNATIONAL RICE GENOME SEQUENCING PROJECT, 2005; PATERSON et al., 2009; SCHNABLE et al., 2009). A alta ploidia e a complexidade do genoma geram desafios para a expressão de transgenes e para o desenvolvimento de marcadores moleculares (LAKSHMANAN et al., 2005).

Recentemente, os esforços em melhoramento estão visando aumentar o conteúdo energético (GJ/ha) utilizando o genótipo de *S. spontaneum* para produção da cana energia, uma cana-de-açúcar com altos rendimentos, com maior conteúdo de fibras e alta tolerância

a seca e pouca sacarose (STICKLEN, 2008; FERREIRA et al., 2013) A cana energia é, portanto, um híbrido de cana-de-açúcar comercial e selvagem. Tem elevada tonelagem (capacidade de transporte em um determinado volume fixo) quando comparado com as canas comerciais, porém com reduzida graduação Brix (% de sólidos solúveis). O ciclo de vida é um pouco mais extenso do que o da cana-de-açúcar comercial, e apresenta aproximadamente o dobro de conteúdo de fibras (KIM; DAY, 2011) (Tabela 2).

Propriedades	Cana-de-açúcar	Cana energia	Sorgo
Ciclo da cultura (meses)	10-12	10-15	3.5
Número de ciclos/ano	Um	Um	Um
Rendimento (t.ha <sup>-1</sup> .ano <sup>-1</sup> )	70	100	60
Brix (% caldo)	13-15	10-12	11-13
Fibras (% cana)	13.5	26.7	13

Tabela 2 - Composição e características agronômicas da cana-de-açúcar, cana energia e sorgo.

Fonte: Adaptado de (KIM; DAY, 2011)

Recentemente, vêm sendo gerado uma grande quantidade de estudos genéticos, químicos e composicionais da parede celular com as variedades cana-de-açúcar (CARDOSO-SILVA et al., 2014; COSTA et al., 2013; COSTET et al., 2012; GARCIA et al., 2013; MASARIN et al., 2011; PALHARES et al., 2012; SOUZA et al., 2013; VICENTINI et al., 2015). Apesar de sua importância econômica, o genoma da cana-de-açúcar ainda não foi sequenciado, diferentemente de outras espécies de importância como o sorgo, milho e arroz. Está em progresso um esforço multinacional para produção do genoma referência de cana-de-açúcar (sugarcanegenome.org), mas muitos obstáculos devem surgir durante o caminho, particularmente devido ao alto número de homólogos e alelos intimamente relacionados que colapsam em um único *contig* durante a montagem (NISHIYAMA JR et al., 2014).

Até o momento, estudos de genômica funcional e transcriptoma contam com os conjuntos de dados de *expressed sequence tag* (EST) e de *RNA-Seq* (CARDOSO-SILVA et al., 2014; FERREIRA et al., 2013; VICENTINI et al., 2015). Apesar da ausência do genoma da cana-de-açúcar, esses estudos produzem os primeiros ensaios, isto é, disponibilização massiva de sequencias relevantes para a identificação de sequencias homólogas com alta identidade a genes previamente caracterizados funcionalmente nas diversas vias metabólicas de interesse para o melhoramento da cana-de-açúcar (FERREIRA et al., 2013; MANNERS; CASU, 2011; MENOSSI; VINCENTZ; SOUZA, 2008).

#### 1.5 Genética da biossíntese de hemicelulose

Os diferentes tipos de carboidratos sintetizados pelo processo de fotossíntese podem ser utilizados como precursores de vários componentes celulares (aminoácidos, lipídios e ácidos nucléicos) e também como componentes estruturais da parede celular. A biossíntese dos componentes da parede celular, tais como celulose, hemiceluloses, pectina e lignina dependem do fornecimento dos açúcares que foram fixados pelo processo fotossintético (TAIZ; ZEIGER, 2004).

A parede celular é constituída de vários polissacarídeos formando uma rede estrutural complexa. Para organizar essa rede de polissacarídeos, a célula necessita de um amplo mecanismo biossintético. Assim, estima-se que aproximadamente 2000 genes estão envolvidos no processo de biossíntese e manutenção da parede celular (CARPITA; TIERNEY; CAMPBELL, 2001).

Análises de bioinformática das sequências genômicas revelaram que a maioria das proteínas que realizam a biossíntese da parede celular são codificadas por famílias multigênicas. A síntese de polissacarídeos pode ser dividida em quatro estágios distintos: (1) produção dos doadores de açúcar-nucleotídeos, (2) iniciação da síntese do polímero, (3) elongação do polímero e (4) terminação da síntese (DELMER; STONE, 1988).

#### 1.5.1 A via de interconversão açúcar-nucleotídeo

A maioria do carbono fixado pela fotossíntese de plantas é incorporado aos polissacarídeos da parede celular. Os monossacarídeos utilizados para montagem de carboidratos são altamente diversos (SEIFERT, 2004). Açúcares-nucleotídeos são unidades de monossacarídeos ativados que são diretamente utilizados pelas glicosiltransferases (GTs) para síntese de diversos glicoconjugados e polissacarídeos. Em plantas, existem pelo menos 30 açúcares-nucleotídeos diferentes, muitos dos quais possuem papel na síntese de diferentes polissacarídeos da parede celular (REITER, 2008; YIN et al., 2011).

As enzimas envolvidas nas reações que levam a formação dos diversos açúcaresnucleotídeos são denominadas de enzimas de interconversão açúcar nucleotídeos (NSE), como mostra a Figura 8 (YIN et al., 2011). A formação de GDP-L-Fucose é realizada pela GDP-D-manose-4,6-desidratase (GMD) (BONIN et al., 1997) juntamente com a enzima bifuncional GDP-4-ceto-6-deoxi-D-manose-3,5-epimerase-4-redutase (GER) (BONIN; REITER, 2000). Já a GDP-L-Galactose é formada pela GDP-D-manose 3,5-epimerase (GME).

Figura 8 - Esquema das 11 enzimas da via de interconversão açúcar-nucleotídeo envolvidas na síntese de hemiceluloses da parede celular de plantas



Fonte: Adaptado de (YIN et al., 2011).

A UDP-D-Glicose pode ser convertida diretamente em UDP-D-Galactose, UDP-D-Ácido Glucurônico ou UDP-L-Ramnose. É provável que a *RHM* (gene da Ramnose Sinthase) catalisa a reação que leva a conversão de UDP-D-Glicose em UDP-L-Ramnose em plantas (REITER; VANZIN, 2001), com o auxílio da nucleotídeo-ramnose sinthase/epimerase-redutase (NRS/ER, conhecida também como UER) (YIN et al., 2011).

O UDP-D-Ácido Glucurônico é irreversivelmente formado a partir da UDP-D-Glicose pela UDPD-glucose desidrogenase (UGD). A biossíntese do UDP-D-Ácido Glucurônico pode ainda ocorrer por um caminho alternativo que envolve a enzima inositol oxigenase (ARNER et al., 2001, KANTER et al., 2003).

A base molecular da biossíntese de UDP-D-Xilose foi elucidada no fungo patogênico *Cryptococcus neoformans* através da abordagem de gene candidato utilizando uma UDP-D-Ácido Glucurônico decarboxilase bacteriana (BAR-PELED; GRIFFITH; DOERING, 2001) e da UDP-D-xilose sintase (UXS) purificada de ervilha (KOBAYASHI et al., 2002). Alternativamente, genes parálogos codificam para duas isoformas de UDP-D-xilose 4-epimerase (UXE) cineticamente distintas. A UXE promove a interconversão de UDP-D-Xilose em UDP-L-Arabinose (SEIFERT, 2004). Além da UXS e UXE, existe ainda uma terceira enzima que leva formação de UDP-D-Xilose: UDP-D-apiose/UDP-D-

xilose sintase (AXS). AXS1 recombinantes purificados levaram a formação de UDP-D-Apiose e UDP-D-Xilose utilizando UDP-D-Ácido Glucurônico como substrato (MOLHOJ; VERMA; REITER, 2003).

UDP-D-Glicose, UDP-D-Ácido Glucurônico e UDP-D-Xilose são interconvertidas reversivelmente em UDP-D-Galactose, UDP-D-Ácido Galacturônico e UDP-L-Arabinose respectivamente (REITER; VANZIN, 2001; FEINGOLD, 1982). A enzima UGE promove a interconversão da UDP-D-Glicose em UDP-D-Galactose. E por fim, o UDP-D-Ácido Galacturônico é gerado pela UDP-D-Ácido glucurônico 4-epimerase (GAE) (FEINGOLD, 1982).

Figura 9 - Famílias gênicas envolvidas da síntese dos polissacarídeos da parede celular. A figura mostra o arranjo e localização propostos para algumas das principais enzimas envolvidas na biossíntese de polissacarídeos. A cadeia de (1,4)-β-glucanas (celulose) são produzidas pelo Complexo Enzimático da Celulose Sintase (CSC), ou pelas rosetadas na membrana plasmática (MP) que se associam entre si formando as microfibrilas. Membros da família das CESAs e das CSLC e CSLD estão envolvidos na síntese de celulose e na formação do CSC.A celulose utiliza a UDP-D-Glicose no lado citoplasmático e libera os produtos nos apoplastos. Foi proposto que as CSLs, GAUT/GATLs e GTs participam da síntese dos polissacarídeos não-celulósicos. As cadeias laterais encontradas em polissacarídeos ramificados são adicionadas pelas GTs localizadas no aparato de Golgi.



Fonte: Adaptado de (DOBLIN; PETTOLINO; BACIC, 2010).

A biossíntese dos polímeros da parede celular ocorre em diferentes regiões da célula. As hemiceluloses e as pectinas são sintetizadas no complexo de Golgi (GIBEAUT; CARPITA, 1994; SCHEIBLE; PAULY, 2004). Nesse caso, os açúcares-nucleotídeos devem ser transportados para dentro do lúmen do complexo de Golgi para que os polissacarídeos sejam sintetizados. No interior do complexo de Golgi, as glicosiltransferases (GTs) transferem um resíduo de açúcar para a cadeia de polissacarídeos em construção, formando o complexo polimérico, que é então transportado via vesículas secretoras do complexo de Golgi, para a superfície da parede celular (BAR-PELED; O'NEILL, 2011; SCHEIBLE; PAULY, 2004). Os diferentes tipos de glicosiltransferases envolvidas na biossíntese de polissacarídeos depende tipicamente do tipo de açúcar-nucleotídeo que será utilizado como substrato (DOBLIN; PETTOLINO; BACIC, 2010). A Figura 9 demonstra onde os diferentes polissacarídeos são sintetizados em uma parede celular esquematizada.

O processo central da biossíntese de polissacarídeos é realizado pela ação das glicosiltransferases (GTs). Uma GT catalisa a formação de ligações entre monossacarídeos adjacentes através da transferência de açúcar de um doador ativado (nucleotídeos-açúcares) a um apropriado aceptor (cadeias crescentes de polissacarídeos) levando a formação de ligações glicosídicas ocasionando na extensão do polissacarídeo (DOBLIN; PETTOLINO; BACIC, 2010). A molécula alvo aceptora é específica para cada enzima GT, mas como um todo, há uma grande diversidade incluindo todos os tipos de macromoléculas e um grande número de classes de compostos de baixo peso molecular (LIU; MUSHEGIAN,2003). A biossíntese de cada estrutura de hemicelulose bem como as GTs envolvidas serão abordadas nos próximos tópicos.

#### 1.5.2 Biossíntese de xilana

Uma grande quantidade de genes está envolvida na biossíntese de xilana (DOERING; LATHE; PERSSON, 2012; SCHELLER; ULVSKOV, 2010). Diferente das outras hemiceluloses não existe evidências de que genes das famílias CSLs atuam na formação da cadeia de xilana. Em contra-partida diversas GTs foram identificadas e caracterizadas na síntese da cadeia de xilana, IRX9/IRX9L e IRX14/IRX14L (GT43) (CHINIQUY et al., 2013; KEPPLER; SHOWALTER, 2010; WU et al., 2010), IRX10/IRX10L (GT47) (BROWN et al., 2009). IRX10L, assim como IRX9L e IRX14L atuam juntamente com IRX10, IRX9 e IRX14, respectivamente, na elongação da cadeia xilana das heteroxilanas (JENSEN et al., 2013). Foi verificado que essas proteínas–*like* 

independentes não desempenham atividade de biossíntese de xilana. Entretanto atuam complementado a atividade realizada pelas IRX10, IRX9 e IRX14 (BROWN et al., 2009; CHINIQUY et al., 2013).

Diversas glicosiltransferases, FRA8 (PENA et al., 2007) e F8H (GT47) (LEE et al., 2009). IRX8 (ou GAUT12) (PENA et al., 2007; PERSSON et al., 2007) e PARVUS (GT8) (LEE et al., 2007), desempenham um papel na formação de um tetrassacarídeo, 4- $\beta$ -D-Xyl*p*- (1 $\rightarrow$ 4) - $\beta$ -D-Xyl*p*- (1 $\rightarrow$ 3) - $\alpha$ -L-Rha*p*- (1 $\rightarrow$ 2) - $\alpha$ -D-Gal*p*A- (1 $\rightarrow$ 4) -D-Xyl*p* na extremidade redutora de GX (PENA et al., 2007). O tetrassacarídeo da extremidade redutora possivelmente possui papel importante na biossíntese de xilana. Foi proposto que ele possa atuar como um iniciador (*primer*), ou então que ele atue como um terminador. Esse tetrassacarídeo não é encontrado na GAX de gramíneas (YORK; O'NEILL, 2008).

A FRA8 pertence à família da glicosiltransferase GT47. Não se sabe ao certo a função bioquímica da enzima. Foi proposto que FRA8 promove a ligação  $\beta$  de xilose (Xyl) ao O3 da ramnose utilizando o açúcar-nucleotídeo UDP- $\alpha$ -D-xilose como substrato ou promove a ligação  $\alpha$  da ramnose ao O2 do ácido galacturônico (GalA) utilizando UDP- $\beta$ -L-ramnose como substrato (PENA et al., 2007).

O gene da família GT47 que codifica para F8H, é um parálogo funcional de FRA8, estando envolvido na biossíntese do tetrassacarídeo da extremidade redutora de GX. O F8H é expresso preferencialmente em células de xilema, onde a parede secundária possui grande quantidade de GX (LEE et al., 2009).

A enzima IRX8 (ou denominada GAUT12), pertencente à GT8, é um homólogo da GAUT1 (transferase de GalA) e afeta os níveis de glucuronoxilana em plantas superiores. Desempenha, também, papel na biossíntese do tetrassacarídeo da extremidade redutora. A IRX8 é um forte candidato na formação da ligação de GalA com Xil (STERLING et al., 2006). IRX8 promove uma ligação importante entre o polímero de xilana e a matriz da parede secundária da célula, afetando diretamente a integridade dessa parede secundária (PERSSON et al., 2007).

O gene *PARVUS*, pertencente à GT8, é expresso durante a biossíntese da parede secundária em fibras e vasos. A enzima PARVUS possui atividade caracterizada na sequencia iniciadora do tetrassacarídeo da extremidade redutora de GX e também na adição de ácido glucurônico da cadeia lateral. Assim, a PARVUS possui um papel na iniciação da biossíntese da sequência iniciadora do tetrassacarídeo de GX (LEE et al.,

2007). As proteínas FRA8, F8H e IRX8 são encontradas no Golgi, enquanto que PARVUS é co-localizada tanto no Golgi quanto no retículo endoplasmático (LEE et al., 2009).

Com relação a substituição nas xilanas, as enzimas GUX1 e GUX2, da família GT8, são requeridas para a adição das ramificados de ácido glucurônico e de ácido 4-Ometilglucurônico em xilana de Arabidopsis na parede celular de caule (MORTIMER et al., 2010). A existência de diversas enzimas GUX levam a diferenças no padrão de substituição das moléculas de GlcA na cadeia de xilana. Essas enzimas estão localizadas no Golgi. A ação da GUX1 e GUX2 é independente da ação das enzimas envolvidas na biossíntese da cadeia principal de xilana. Ou seja, a síntese da cadeia principal é desacoplada e independente da atividade de substituição de cadeia (MORTIMER et al., 2010).

Membros da família GT61 também possuem papel na biossíntese das heteroxilanas (JENSEN et al., 2013). As proteínas da GT61 desempenham papel na biossíntese de arabinoxilana e consequentemente revelam uma divergência evolutiva na parede celular de gramíneas. Essas proteínas atuam na adição de  $\alpha$ -(1,3)-arabinosil na cadeia de xilana, sendo nomeadas de xilana arabinosiltransferases (XATs) (ANDERS et al., 2012; MARRIOTT et al., 2014). Entretanto, existem membros da GT61 que atuam também com atividade de xilosil transferase. A proteína de arroz XAX1 possui atividade de  $\beta$ -1,2-xilosil transferase, promovendo a transferência de xilose a partir do açúcar-nucleotídeo UDP-xilose ao resíduo de arabinose. A adição dessa unidade de xilose leva a formação da cadeia lateral  $\beta$ -Xylp-1-2- $\alpha$ -Araf (CHINIQUY et al., 2012).

Diversas outras GTs da família GT75, foram identificados como potencialmente envolvidos na biossíntese de heteroxilana (ZENG et al., 2010). Membros da família GT75 conhecidos como RGP ou UAM catalisam a interconversão de UDP-Ara*f* e UDP-Ara*p* (KONISHI et al., 2007). As GTs, provavelmente, se organizam em complexos enzimáticos para sintetizar a cadeia de xilana (FAIK, 2010). Evidência comprovando a existência de um complexo da xilana sintase foi obtida em experimento com trigo na qual o complexo proteico contendo membros das famílias GT43, 47 e 75 sofreram co-imunoprecipitação (ZENG et al., 2010).

#### 1.5.3 Biossíntese de xiloglucana

Diversas proteínas codificadas por famílias genicas da celulose sintase-*like* são conhecidas por atuar na síntese de cadeias de glucanas envolvidas na biossíntese dos

polissacarídeos da parede celular. A cadeia principal de glucanas da XyG é sintetizada por um ou mais genes membros da família de celulose sintase-*like* C (CSLC) (COCURON et al., 2007). A expressão heteróloga da proteína CSLC4 de Arabidopsis na levedura *Pichia pastoris* resultou na produção de cadeia de 1,4-β-glucana, confirmando a atividade de glucana sintase da CSLC. Foi demonstrado, também, que a CSLC4 possui localização no aparato de Golgi, o sítio da biossíntese de XyG (DAVIS et al., 2010). O domínio catalítico da proteína CSLC4 está localizado do lado citosólico da membrana do Golgi. Durante a biossíntese de xiloglucana, a cadeia nascente de glucana está conectada no lúmen do Golgi pelo domínio transmembrana da CSLC (DAVIS et al., 2010).

Os substituintes de xilosil da cadeia lateral da XyG podem sofrer substituições, também. Proteínas pertencentes a família GT47, MUR3 e XLT2, foram identificadas com atividade  $\beta$ -1,2 galactosiltransferases. MUR3 atua como XyG:GalT, ou seja, realiza a transferência de um resíduo de galactosil ao terceiro resíduo de xilosil (MADSON et al., 2003). A XLT2 também possui atividade de galactose transferase, entretanto ela é específica da posição 2 da molécula de xilosil (JENSEN et al., 2012). O gene que codifica XUT1 também pertence à família GT47. A XUT1 é uma galacturonosiltransferase específica de xiloglucana, responsável pela formação de ligações de (1,2) ácido  $\beta$ -Dgalactosilurônico- $\alpha$ -D-xilosil (PENA et al., 2012). Acredita-se que outros genes da GT47 podem atuar na diversidade de decorações da cadeia de XyG (PAULY et al., 2013).

#### 1.5.4 Biossíntese de MLG

As glucanas de cadeia mista são de grande importância durante o desenvolvimento da parede primária e possui uma função estrutural transiente no suporte a elongação da parede celular (CHRISTENSEN et al., 2009). Unidades de glicose da MLG desempenha papel no mecanismo de reciclagem de açúcar durante o desenvolvimento de sementes em milho (GIBEAUT; CARPITA, 1991). Isso indica que a MLG atua como um composto de armazenamento de fácil conversão, paralelamente ao seu papel estrutural na parede celular, explicando a alta quantidade de MLG na parede celular de endosperma.

Recentemente, genes envolvidos na biossíntese de MLG foram identificados (BURTON et al., 2006; DOBLIN et al., 2009). Assim, duas classes de genes foram identificadas como atuantes da síntese das glucanas de cadeias mistas, *CSLF* e *CSLH*, ambas representam ramificações específicas da superfamília gênica das CSLs (celulose sintase-*like*) de gramíneas (CHRISTENSEN et al., 2009; VOGEL, 2008). Ainda não se

sabe qual é exatamente o papel que essas proteínas possuem na biossíntese de MLG, entretanto ensaios biológicos demonstraram que a biossíntese de MLG ocorre no aparato de Golgi (PAULY et al., 2013) e que tanto a CSLF quanto CSLH estão localizadas no Golgi (DOBLIN et al., 2009).

Foi proposto um mecanismo de duas etapas para a síntese de MLG (BURTON; FINCHER, 2012). Primeiramente, celodextrinas são sintetizadas no Golgi e seguidamente são conectadas por ligações  $\beta$ -1,3 pela CSLF ou CSLH ou por uma transglicosilase. Foi proposto a existência de uma possível transglicosilase de MLG, devido a identificação de atividade de transglicosilase na montagem de MLG (FRY et al., 2008).

A CSLF6 é uma proteína integral de membrana e considerada o componente majoritário do complexo enzimático (1-3,1-4)- $\beta$ -glucana sintase. Um único aminoácido dentro do domínio transmembrana da CSLF controla a estrutura da MLG (JOBLING, 2015). Assim, um novo mecanismo de controle da estrutura do polissacarídeo foi proposto: a arquitetura do poro da membrana plasmática e a translocação do polissacarídeo crescente através da membrana controla como o monômero de glucana é inserido e coordenado no sítio ativo. Esse controle afeta a proporção de ligações  $\beta$ 1-3 e  $\beta$ 1-4 do polissacarídeo (JOBLING, 2015).

#### 1.5.5 Biossíntese de heteromanana

Assim como todos os polissacarídeos da parede celular, as heteromananas são sintetizadas a partir de seus açúcares-nucleotídeos ativados. Para as mananas, essas moléculas são GDP-manose, GDP-glicose e UDP-galactose (LIEPMAN; WILKERSON; KEEGSTRA, 2005). Inicialmente, foi descoberto a existência da manana sintase (ManS), membro da família celulose sintase-like A (CSLA) (DHUGGA et al., 2004). Essa família é responsável pela síntese de manana e glucomanana tanto em eudicotiledôneas quanto em gramíneas (VOGEL, 2008). A atividade da ManS foi demonstrada pela proteína CSLA7 de *Arabidopsis thaliana*, que juntamente com evidências genéticas acredita-se que essa proteína utiliza a GDP-manose como aceptor único (GOUBET et al., 2009; LIEPMAN; WILKERSON; KEEGSTRA, 2005).

Abordagem de genética reversa e bioquímica utilizadas na planta modelo *Arabidopsis*, comprovaram que a família CSLA atua na síntese da cadeia principal de glucomanana (LIEPMAN; WILKERSON; KEEGSTRA, 2005). A partir de análises de microarranjo e qRT-PCR, os transcritos dos genes de *CSLA* de Arabidopsis apresentaram

um padrão de expressão tecido-específico em tecidos vegetativos e florais. (LIEPMAN et al., 2007). Isso suporta a hipótese de que as heteromananas possuem funcionalidade em processos metabólicos e na estrutura da parede celular além do armazenamento de carboidrato pela planta.

Foi demonstrado que as enzimas CSLA2, CSLA3 e CSLA9 são responsáveis pela síntese de glucomanana em tecidos do caule, enquanto que a CSLA7 está envolvida na síntese de glucomanana em tecido embrionário (GOUBET et al., 2009). Acredita-se que proteínas da família CSLD também participem da biossíntese de manana (VERHERTBRUGGEN et al., 2011).

#### 1.6 Justificativa

Informações sobre como o fluxo de metabólitos dos diversos açúcares-nucleotídeos afetam a biossíntese de polissacarídeos da parede celular são importantes para o entendimento da regulação das complexas rotas metabólicas, de forma a permitir assim, modificações do conteúdo da parede celular. A bioinformática, análise evolutiva e genômica funcional podem auxiliar na compreensão do funcionamento dos genes, e assim, na modulação da expressão de alguns genes envolvidos no processo de síntese da parede celular regulando os diferentes tipos de hemiceluloses (REITER; VANZIN, 2001; HARPER; BAR-PELED, 2002). A disponibilidade de informações a respeito dos genes envolvidos na biossíntese da hemicelulose em espécies modelo, como a *Arabidopsis thaliana, Brachypodium distachyon* e *Oryza sativa*, serve de instrumento para identificar genes *bona fides* em cana-de-açúcar de maneira a identificar candidatos interessantes a análise genética funcional.

Até o momento, apesar de já existir uma grande quantidade de dados de EST, poucos genes da via de biossíntese de hemicelulose foram identificados em cana-de-açúcar (CASU et al., 2007; VETTORE et al., 2003; VICENTINI et al., 2015). Portanto, a análise filogenética comparada a gramíneas relacionadas e a identificação destes genes em cana-de-açúcar contribuem para a seleção de genes candidatos envolvidos na biossíntese de hemicelulose e possivelmente envolvidos na recalcitrância da parede celular. Tais resultados poderão ser usados para definir os potenciais candidatos gênicos para análise de expressão gênica em amostras contrastantes de teor de hemicelulose cultivados na EEL-USP.

#### **2 OBJETIVOS**

#### 2.1 Objetivos Gerais

O presente trabalho visa a identificação, anotação e filogenia dos genes das famílias gênicas das glicosil<u>t</u>ransferases, celulose sintase-*like* e interconversão açúcar-nucleotídeo envolvidos na biossíntese de hemicelulose em gramíneas e cana-de-açúcar.

#### 2.2 Objetivos específicos

Usar os genes das famílias gênicas glicosiltransferase (GT2, GT8, GT43, GT47, GT61, GT75), da via de interconversão açúcar-nucleotídeo (*GME*, *GMD*, *GER*, *UGE*, *RHM*, *NRS/ER*, *UGD*, *UAXS*, *UXS*, *UGlcAE*, *UXE*) e *cellulose-sintase-like* (*Csl*) caracterizados funcionalmente como envolvidos na biossíntese de hemicelulose em *Arabidopsis thaliana* para identificar e anotar os genes homólogos em *Sorghum bicolor* (sorgo), *Zea mays* (milho), *Oryza sativa* (arroz) e *Brachypodium distachyon*;

Usar os genes anotados das gramíneas *Sorghum bicolor*, *Zea mays* e *Oryza sativa* para identificar as sequencias homólogas em cana-de-açúcar nos bancos do SUCEST e distintas bibliotecas de RNA-Seq;

Realizar análises filogenéticas para identificar os genes ortólogos e parálogos *bonafides* envolvidos na biossíntese de hemicelulose nas distintas gramíneas e em cana-deaçúcar.

#### **3 MATERIAL E MÉTODOS**

# 3.1 Identificação dos genes da via de biossíntese de hemicelulose em *Arabidopsis thaliana* e gramíneas

Os genes de Arabidopsis thaliana e suas respectivas famílias gênicas envolvidas na biossíntese de hemicelulose, tais como, At3g46440.1 (Epimerase), At1g26570 (UDPG), At4g16590 (GT2), At1g32180 (CESA), At1g77130 (GT8), At1g27440 (GT47) (PENNING et al., 2009), At2g37090 (GT43) (OIKAWA et al., 2010), At2g03370 (GT61) (MITCHELL; DUPREE; SHEWRY, 2007) e At3g02230 (GT75) (ZENG et al., 2010) foram baixados no TAIR database (The Arabidopsis Information Resource) versão 10 (LAMESCH et al., 2011) (http://www.arabidopsis.org/index.jsp) e usados como isca inicial para identificar seus homólogos em cada família gênica. Essa busca foi realizada através de blastp local usando como base o proteoma de Arabidopsis disponíveis no Phytozome v10.1 (http://phytozome.jgi.doe.gov/pz/portal.html) (GOODSTEIN et al., 2012) e complementada pelo método de busca probabilístico HMMER (hmmsearch e jackhmmer).

Para as gramíneas, os genes de Arabidopsis foram usados como isca através de blast*p* a para a busca de ortólogos e parálogos em *Brachypodium distachyon v2.1, Oryza sativa v7.0, Zea mays v6a e Sorghum bicolor v2.1*. Foram utilizadas como sequências de entrada para o blast*p*, seqüências de aminoácidos de Arabidopsis contra um banco de dados de proteínas. Todas as sequências que retornaram do blast*p* foram baixadas para posterior anotação.

Com a finalidade de buscar o maior número de sequências de aminoácidos de cada família gênica, empregou-se uma estratégia complementar de busca utilizando o modelo de cadeias ocultas (HMMs-<u>Hidden Markov Models</u>) através do programa HMMER v3.1 (http://hmmer.janelia.org/). Foram construídos perfis de HMMs para cada família gênica de cada espécie (*Arabidopsis thaliana*, *Brachypodium distachyon*, *Oryza sativa*, *Zea mays* e *Sorghum bicolor*), utilizando o *script hmmerbuild* (FINN; CLEMENTS; EDDY, 2011). A busca entre HMMs e o banco de proteínas foi realizada pelo *script hmmsearch* utilizando o nível de significância (*e-value*) padrão do programa. As novas sequências de aminoácidos foram identificadas, enquanto que as redundantes foram removidas.

# 3.2 Anotação das sequências gênicas das famílias relacionadas ao processo de biossíntese de hemicelulose

As sequências protéicas de Arabidopsis e das monocotiledôneas baixadas do *Phytozome v10.1* foram anotadas com as seguintes informações: código da sequência, sequência primária, nome (ou *alias*, quando houver), nome do gene, localização cromossômica, sentido da sequência (adotou-se sinal positivo (+) para a orientação 5'->3'), tamanho da sequência proteica, domínio da proteína e seu *e-value*, e por fim, se a sequência proteica está completa ou incompleta. A anotação e confirmação da presença de domínios específicos de cada família gênica foi validada através do programa *PFAM* (http://pfam.sanger.ac.uk) (FINN et al., 2014), fornecido pelo *EMBL European Bioinformatics Institute*. O nome do gene foi anotado de acordo com revisões na literatura, enquanto que as demais informações foram retiradas do próprio *Phytozome v10.1*. Todas as sequências protéicas validadas foram baixadas e armazenadas levando a montagem de um banco de sequências protéicas de cada família gênica.

#### 3.3 Identificação e anotação dos genes em cana-de-açúcar (Saccharum sp.)

Para a identificação das sequencias homólogas em cana-de-açúcar, utilizou-se as sequências CDS (*Codifying DNA Sequence*) de *Sorghum bicolor, Zea mays* e *Oryza sativa* como iscas para a realização do blast*n* local. Para a busca das sequencias em cana-de-açúcar, foram utilizadas sete listas distintas de sequencias de DNA de cana-de-açúcar: [1] sequencias ESTs construídas pelo SUCEST (VETTORE et al., 2003); [2] sequencias de RNA-Seq montadas a partir de amostras de raiz disponibilizadas pelo colaborador Marcos Buckeridge (dados não publicados); [3] sequencias de RNA-Seq montadas a partir de amostras de raiz disponibilizadas pelo colaborador Marcos Buckeridge (dados não publicados); [3] sequencias de RNA-Seq montadas a partir de amostras de plântula e disponibilizadas pela colaboradora Dra. Adriana Hemerly (dados não publicados); [4] sequencias de RNA-Seq montados a partir de amostras de folha (CARDOSO-SILVA et al., 2014); [5] sequencias da de RNA-Seq geradas a partir de amostras de folhas jovens (GRATIVOL et al., 2014); [6] sequencias de RNA-Seq montadas a partir de amostras de folhas jovens (VICENTINI et al., 2015); [7] sequencias de RNA-Seq montadas partir dos dados de amostras de folha e entrenó (NISHIYAMA JR et al., 2014).

Para todas as bibliotecas, o processo de anotação da lista foi realizado inicialmente pela conversão das sequências de DNA de cana-de-açúcar em sequências de aminoácidos a partir de seis fases de leitura aberta (*open reading frames - ORF*), três fases no sentido 5'-
>3' e as outras três no sentido oposto, pela ferramenta *ExPASy translate* (http://web.expasy.org/translate/) (GASTEIGER et al., 2003). A confirmação da presença de domínios específicos de cada família gênica foi validada através do software online *PFAM*. Nos casos em que as sequências de DNA não puderam ser deduzidas pelo *ExPASy translate*, realizou-se a dedução pelo *FGENESH*+ fornecido pelo *Softberry* (http://www.softberry.com/berry.phtml), que faz comparação entre a sequência CDS de cana-de-açúcar de entrada com a sequência protéica homóloga em *Sorghum bicolor* (do banco de dados do próprio *software*) identificada por blast*n*, retornando a sequência de proteína codificada na região de DNA submetida.

#### 3.4 Alinhamento e análise filogenética

Para a Arabidopsis e as espécies de gramíneas analisadas, após a obtenção da lista completa das proteínas de cada família gênica, procedeu-se com o alinhamento das sequencias proteícas a partir do *MUSCLE* (http://www.ebi.ac.uk/Tools/msa/muscle/) (EDGAR, 2004). A construção das árvores filogenéticas foi realizada pelo método de *Neighbor-Joining* usando o *MEGA6 (Molecular Evolutionary Genetics Analysis)* (TAMURA et al, 2013). Foram utilizados os parâmetros de análise estatística por *Bootstrap* com 1000 replicações, método de distância *p* e com deleção par a par.

Em vista da dificuldade e limitação para montagem de árvores filogenéticas com as sequências incompletas de cana-de-açúcar no programa MEGA, foi utilizado programa *Seaview v.4.5.4* (GOUY; GUINDON; GASCUEL, 2010) para contornar essa limitação e proceder com a análise de homologia das sequências de cana-de-açúcar em relação as sequencias de sorgo e milho previamente anotadas. Inicialmente, procedeu-se com o uso de um algoritmo no MEGA para encontrar o melhor modelo para cada família gênica analisada. As árvores filogenéticas de cada família gênica foram realizadas com o uso das sequencias das sete bibliotecas de cana-de-açúcar juntamente com as sequencias de *Sorghum bicolor* e *Zea mays*. A construção das árvores foi feita pelo *PhyML* do *Seaview v4.5.4* utilizando a análise por máxima verossimilhança. Após definição dos prováveis genes ortólogos, os genes alvo de cana-de-açúcar foram analisados por meio de alinhamento múltiplo por *ClustalW* (THOMPSON et al., 1994) disponível no software *BioEdit Sequence Alignment Editor v.7.2.5* para verificar se as sequências estavam completas, e assim, determinar os genes candidatos de cana-de-açúcar.

#### **4 RESULTADOS E DISCUSSÃO**

#### 4.1 Anotação das famílias gênicas envolvidas na biossíntese de hemicelulose

Fazendo uso de genes chaves de *Arabidopsis thaliana* membros de cada família gênica envolvida na biossíntese de hemicelulose e usando os métodos de busca por similaridade e probabilidade, Blast*p* e HMMER, respectivamente, foram encontrados 178, 182, 185, 187 e 273 genes membros das famílias gênicas envolvidas na via de biossíntese de hemicelulose em *Arabidopsis thaliana*, *Brachypodium distachyon*, *Oryza sativa*, *Sorghum bicolor* e *Zea mays*, respectivamente (Tabela 3). Para simplificar a explicação destas famílias gênicas, as mesmas foram separadas em três grupos distintos de acordo com o seu envolvimento funcional. O primeiro grupo representa as GTs, com atividade de glicosiltransferase, constituindo as famílias GT8, GT43, GT47, GT61 e GT75. O segundo grupo contém os genes que codificam para as NSE (<u>Nucleotide Sugar Interconversion Enzymes-</u> Epimerase e UDPG). E o terceiro grupo possui as famílias GT2 e CSL que codificam para proteínas com atividade celulose sintase-*like*. A família GT2 no presente trabalho abrange apenas as CSLA e CSLC, enquanto que a CSL contém as CSLF e CSLH.

Para *A. thaliana*, dentro desses 178 genes, foram encontrados 26 pertencem ao grupo CSL, já identificados anteriormente (RICHMOND; SOMERVILLE, 2000) que são aqueles que possuem o domínio celulose sintase e a GT2 que possui 14 genes envolvidos na via de biossíntese de hemicelulose. No grupo das GTs foram encontrados 99 genes, e 39 genes foram identificados como membros do grupo das NSE (Tabela 3).

No caso de *B distachyon* (Tabela 3), foram encontrados 108 genes do grupo das GTs. O grupo das NSE possui 34 genes encontrados. E por fim, o terceiro grupo (celulose sintase-*like*) possui 40 genes identificados, número semelhante ao encontrado em *A. thaliana*. Em *O. sativa* (Tabela 3), foram encontrados 100, 37 e 50 genes correspondentes aos grupos das GTs, NSE e celulose sintase-*like*, respectivamente.

A busca em *S. bicolor* (Tabela 3) resultou em 100 genes do grupo das GTs encontrados, 37 das NSE e 50 do grupo das Celulose sintase-*like*. Já em *Z. mays* (Tabela 3) identificou-se 146 sequências do grupo das GTs, 54 e 73 dos grupos das NSE e das Celulose sintase-*like*, respectivamente.

As sequencias anotadas em *Oryza sativa*, *Zea mays* e *Sorghum bicolor* foram subsequentemente utilizados para identificação de genes em cana-de-açúcar a partir das

sete bibliotecas disponibilizadas por [1] (VETTORE et al., 2003); [2] (BUCKERIDGE et al., in press); [3] (HEMERLY et al., in press); [4] (CARDOSO-SILVA et al., 2014); [5] (GRATIVOL et al., 2014); [6] (VICENTINI et al., 2015); [7] (NISHIYAMA JR et al., 2014).

Tabela 3 - Número de genes das famílias gênicas relacionadas com a biossíntese de hemicelulose da parede celular. A letra "B" representa o número de genes encontrados de acordo com a metodologia de busca. "BF" representa a quantidade de genes *bona fides*. O tamanho do genoma representado entre parênteses (\*) corresponde ao tamanho do genoma haplóide. Todas as espécies são diplóides. *At: Arabidopsis thaliana* 2n = 10 (THE ARABIDOPSIS GENOME INITIATIVE, 2000); *Bd: Brachypodium distachyon* 2n = 10 (THE INTERNATIONAL BRACHYPODIUM INITIATIVE, 2010); *Os: Oryza sativa* 2n = 24 (INTERNATIONAL RICE GENOME SEQUENCING PROJECT, 2005); *Sb: Sorghum bicolor* 2n = 20 (PATERSON et al., 2009); *Zm: Zea mays* 2n = 20 (SCHNABLE et al., 2009).

Espécies/	A (125	t Mb)	B (272	ed Mb)	(389	)s Mb)	(750	b Mb)	Zz (2300	m Mb)
Familia	В	BF	В	BF	В	BF	В	BF	В	BF
GT8	41	6	37	4	16	3	17	5	29	6
GT43	4	4	10	5	14	5	10	5	16	9
GT47	39	7	35	4	39	5	39	6	52	9
GT61	10	0	22	4	26	4	30	3	35	4
GT75	5	5	4	4	3	3	4	4	13	5
Sub-total 1	99	22	108	21	98	20	100	23	146	33
Epimerase	34	32	30	19	33	21	33	22	51	17
UDPG	5	4	4	3	5	2	4	3	3	2
Sub-total 2	39	36	34	22	38	23	37	24	54	19
GT2	14	6	14	2	16	2	15	2	21	3
CSL	26	3	26	4	33	4	35	6	52	5
Sub-total 3	40	9	40	6	49	6	50	8	73	8
TOTAL:	178	67	182	49	185	49	187	56	273	60

Fonte: Autoria própria

A Tabela 4 sintetiza os resultados da busca de sequências das famílias gênicas envolvidas com a biossíntese de hemicelulose nas sete bibliotecas de cana-de-açúcar analisadas. Para as bibliotecas que utilizaram amostras de folha [4] e [5], foram identificadas, respectivamente, 171 e 85 sequências, com domínio anotado, membros das famílias gênicas envolvidas na biossíntese de hemicelulose. Para a biblioteca de raíz [2], foram identificadas um total de 146 sequências e para a biblioteca de plântula [3], 190 sequências foram identificadas. A busca de sequências no banco do SUCEST [1], retornou um total de 204 sequências com domínio. O banco de entrenó [6] possui 144 sequências identificadas e a biblioteca que utilizou amostras de entrenó e folha [7] permitiu a identificação de 511 sequências com domínio.

Tabela 4 - Número de sequências das famílias gênicas relacionadas com a biossíntese de hemicelulose em distintas bibliotecas de cana-de-açúcar. Em parênteses representa o número total de sequências e o número fora dos parênteses representa a quantidade de sequências com domínio característico. [1] SUCEST (VETTORE et al., 2003); [2] Raiz (BUCKERIDGE et al., in press); [3] Plântula (HEMERLY et al., in press); [4] Folha (CARDOSO-SILVA et al., 2014); [5] Folha (GRATIVOL et al., 2014); [6] Entrenó (VICENTINI et al., 2015); [7] Folha e Entrenó (NISHIYAMA JR et al., 2014).

Famílias	[1]	[2]	[3]	[4]	[5]	[6]	[7]
GT8	28 (61)	33 (44)	36 (79)	27 (32)	6 (11)	22 (110)	72 (152)
GT43	9 (15)	5 (7)	10 (88)	7 (9)	2 (6)	8 (87)	27 (36)
GT47	35 (49)	15 (23)	24 (41)	17 (23)	6 (10)	25 (87)	60 (83)
GT61	17 (43)	10 (35)	20 (121)	14 (42)	5 (17)	18 (67)	50 (118)
GT75	7 (8)	6 (6)	3 (82)	3 (3)	2 (3)	2 (21)	6 (6)
Epimerase	50 (75)	30 (44)	32 (46)	53 (70)	26 (59)	29 (215)	94 (184)
UDPG	5 (5)	2 (2)	5 (17)	1 (1)	2 (2)	2 (9)	6 (6)
GT2	10 (83)	11 (19)	48 (155)	14 (17)	4 (7)	13 (31)	37 (265)
CSL	43 (83)	34 (42)	12 (155)	35 (44)	32 (41)	25 (124)	159 (265)
TOTAL	204 (422)	146 (222)	190 (784)	171 (241)	85 (156)	144 (751)	511 (1115)
	(122)	(222)	(101)	(211)	(100)	(101)	(1110)

Fonte: Autoria própria

# 4.2 Filogenia e Identificação dos genes *bona fides* da biossíntese de hemicelulose em Arabidopsis e gramíneas

#### 4.2.1 Análise dos genes envolvidos na via de interconversão açúcar-nucleotídeo

Foi demonstrado que genes que codificam proteínas com domínios epimerase e UDPG, estão relacionadas as enzimas de interconversão de açúcar-nucleotídeo (YIN et al., 2011). A partir de análises filogenéticas (YIN et al., 2011), observou-se que a superfamília Epimerase pode ser subdividida em três grandes clados (A, B e C). As 11 enzimas das NSE, apesar de serem originadas a partir de diferentes genes progenitores, estão todas presentes no clado A. Assim, a partir da revisão bibliográfica, determinou-se que das 34 sequências com domínio epimerase, apenas duas (At4g33030; At4g00560) não foram classificadas como *bona fides* em *A. thaliana* (Tabela 3). Das cinco sequências anotadas com o domínio UDPG, quatro foram caracterizadas como UGD (SEITZ et al., 2000) e, portanto, classificadas como *bona fides* totalizando 36 *bona fides* em *A thaliana*. A Tabela 5 revela a quantidade de genes *bona fides* para cada uma das 11 NSE.

Prévios trabalhos (BAR-PELED; O'NEILL, 2011; RANCOUR et al., 2015; REITER, 2008; YIN et al., 2011) identificaram genes envolvidos na biossíntese de hemicelulose em Arabidopsis e Brachypodium. Baseando-se nestas informações e juntamente com análises filogenéticas, foram identificados 23 e 19 genes são *bona fides* 

em *O. sativa* e *Z. mays*, respectivamente (Tabela 5). Os genes em *B. distachyon* foram anotados segundo revisão na literatura (RANCOUR et al., 2015), juntamente com a análise das árvores filogenéticas das NSE com domínio Epimerase e UDPG (Figura 10 e Figura 11B). Assim, foram identificados 22 genes *bona fides* em *B. distachyon* (Tabela 5). Os genes em *S. bicolor* foram identificados a partir de análise das árvores filogenéticas, obtendo-se 24 genes (Tabela 5).

Tabela 5 - Quantidade de genes *bona fides* das enzimas de interconversão açúcar nucleotídeo nas diferentes espécies. *At: Arabidopsis thaliana; Bd: Brachypodium distachyon; Os: Oryza sativa; Sb: Sorghum bicolor; Zea mays.* 

NSE	At	Bd	Os	Sb	Zm
GMD	2	1	1	2	1
GER	2	1	1	1	1
GME	1	2	2	2	1
RHM	3	1	1	1	1
UER	1	0	1	1	1
UGD	4	3	2	3	2
UXS	6	5	5	5	4
UXE	4	3	3	3	3
AXS	2	1	1	1	2
UGE	5	2	2	2	0
GAE	6	3	4	4	3
Total	36	22	23	25	19

Fonte: Autoria própria

#### 4.2.2 Análise dos genes da família gênica GT8

Os membros da família glicosiltransferase 8 (GT8) atuam na biossíntese de diversas moléculas do metabolismo da planta (YIN et al., 2010), incluindo as hemiceluloses. A árvore filogenética da GT8 (Figura 12) foi dividida em cinco grandes clados de A a E seguindo o proposto em (PENNING et al., 2009). Os clados *bona fides* são os clados A, C e D, onde os genes identificados e caracterizados funcionalmente com atividade na biossíntese das hemiceluloses estão presentes. O clado A, possui os genes GUX1, GUX2 e GUX4 de Arabidopsis que desempenham papel na biossíntese de glucuronoxilana, realizando a adição dos substituintes glucuronosil e 4-O-metil-glucuronosil na cadeia principal de xilana (RENNIE et al., 2012). Os clados D e C possuem os genes GAUT e GAUT-like (GATL), respectivamente que codificam enzimas com atividade de galacturonosil transferase (YIN et al., 2010) e foi proposto que essas enzimas estão envolvidas na biossíntese de pectina e/ou xilana (CAFFALL et al., 2009; MOHNEN,

2008). Essa família de galacturonosil transferase possui 25 membros, o que é consistente com o observado na literatura (YIN et al., 2010). Dentre os 25 membros, destaca-se os genes *PARVUS/GATL1* e *IRX8/GAUT12* envolvidos na biossíntese do tetraoligossacarídeo da extremidade redutora de GXs de eudicotiledôneas. Ao analisar a árvore filogenética da GT8 (Figura 12) nenhum homólogo ao *IRX8/GAUT12* é observado em gramíneas, sugerindo que este gene possui função especializada na síntese de GX apenas em eudicotiledôneas (YIN et al., 2011). Foi proposto que um outro gene do grupo da GAUT, *GAUT8/QUA1*, está envolvido na biossíntese de xilana e/ou pectina (BOUTON et al., 2002).

Não existem, até o momento, estudos de caracterização funcional dos genes da família GT8 envolvidos na biossíntese de hemicelulose nas espécies de gramíneas analisadas. Assim o uso de bioinformática e a análise filogenética torna-se uma importante ferramenta para fazer inferências acerca dos genes das espécies.

A partir da árvore filogenética (Figura 12), existem dois (Bd2g56810; Bd2g24737), um (Os01g65780), três (Sb009G144000; Sb009G144200; Sb003G376700) e quatro genes (Zm2G058472; Zm2G002023; Zm2G365544; Zm2G135743) ortólogos a *GUX1* em *B. distachyon, O. sativa, S. bicolor* e *Z. mays* respectivamente. Existe apenas um gene ortólogo a *GUX2* e *GUX4* em cada espécie (Bd1g72350; Os03g08600; Zm2G109431; Sb001G479800). Apesar de não existir evidências da existência de um tetraoligossacarídeo na extremidade redutora de xilanas de gramíneas, o gene *PARVUS* possui homologia com quatro genes das espécies de gramíneas, sendo um de cada espécie (Bd1g12560; Os03g47530; Zm2G149024; Sb001G138200).

#### 4.2.3 Análise dos genes da família gênica GT43

Os genes de *A thaliana*, *IRX9*, *IRX9L*, *IRX14*, *IRX14L* identificados e caracterizados estão envolvidos na elongação da cadeia de xilana (KEPPLER; SHOWALTER, 2010; PENA et al., 2007), juntamente com genes da família GT47, como *IRX10* e *IRX10L* (BROWN et al., 2009). Assim, a família GT43 possui enzimas relacionadas com a elongação da cadeia principal de xilana das moléculas de heteroxilanas (KEPPLER; SHOWALTER, 2010; PENA et al., 2007). A metodologia desenvolvida pelo presente trabalho, possibilitou a busca dos quatro membros da família GT43 sendo os quatro membros *bona fides* da biossíntese de xilana.

Estudos prévios permitiram a identificação de 10 genes pertencentes a família GT43 de arroz (OIKAWA et al., 2010). A anotação realizada pelo presente trabalho, permitiu encontrar 14 genes pertencentes a GT43, ou seja, quatro a mais (Os03g57910; Os04g58040; Os07g39990; Os01g05400) da quantidade já identificada (OIKAWA et al., 2010). Além disso, os quatro genes (At2G37090; At1G27600; At5G67230; At4G36890) da GT43 de Arabidopsis, são ortólogos a outros quatro genes em arroz, demonstrando que *Oryza sativa* também possui os genes *IRX9*, *IRX9L*, *IRX14* e *IRX14L* (OIKAWA et al., 2010). *B. distachyon*, possui 10 genes, o que está de acordo com (THE INTERNATIONAL BRACHYPODIUM INITIATIVE, 2010).

A árvore filogenética (Figura 11C) foi dividida e classificada em cinco clados: clado IRX14, clado IRX9 e clado IRX9L e dois clados específicos de gramíneas. A partir da análise da árvore filogenética juntamente com o auxílio de estudos de atividade enzimática dos membros em arroz (LEE et al., 2014), cinco genes de arroz *OsGT43A*, *OsGT43C* (clado *IRX9*), *OsGT43E*, *OsGT43F* (clado IRX9L) e *OsGT43J* (clado *IRX14*) possuem papel na biossíntese da cadeia de xilana. Os outros genes anotados em arroz (*OsGT43B*, *OsGT43G*, *OsGT43H*), apesar de estarem em clados *bona fides*, não demonstraram atividade na biossíntese da cadeia de xilana.

Os cinco genes *bona fides* (Os05g03174; Os07g49370; Os05g48600; Os01g48440; Os06g47340) de *O. sativa* foram utilizados como referência para anotar os genes, por homologia, nas outras espécies. Assim, no clado IRX9, existem dois genes *bona fides* (Bd1g16540; Bd2g37970; Zm2G076307; Zm2G012874; Sb009G026100; Sb002G430700) para cada uma das três espécies. Para o clado IRX9L, foram anotados dois (Bd2g16840; Bd2g46410), cinco (Zm2G025302; Zm2G356579; Zm2G150264; Zm2G103628; Zm2G037469) e dois (Sb009G229300; Sb003G254700) genes *bona fides* em *B. distachyon*, *Z mays* e *S. bicolor*, respectivamente. E por fim, no clado IRX14, existem dois genes *bona fides* (Zm2G150302; Zm2G113655) para *Z. mays* e um para *B. distachyon* e *S. bicolor* (Bd1g33320; Sb010G238800).

#### 4.2.4 Análise dos genes da família gênica GT47

De todas as GTs analisadas, os membros da GT47 aparentam estar envolvidos em uma maior quantidade de processos de biossíntese das hemiceluloses. A partir de estudos de caracterização gênica, constatou-se que os genes da família GT47 desempenham diversas funções na síntese dos componentes da hemicelulose: elongação da cadeia de xilana das heteroxilanas (BROWN et al., 2009); biossíntese do tetraoligossacarídeo da extremidade redutora de GX (LEE et al., 2009; PENA et al., 2007) e, ainda, participa na decoração da cadeia lateral de xiloglucanas (MADSON et al., 2003). A árvore filogenética da família gênica GT47 (Figura 13) foi dividida em cinco grupos distintos: clados A, B, C, D e E (PENNING et al., 2009). Os genes dos cinco clados codificam enzimas com pelo menos quatro tipos diferentes de atividade de transferase: galactosil, arabinosil, xilosil e glucuronosil transferases. Acredita-se que os membros do clado C e B estejam envolvidos na biossíntese de pectinas (JENSEN et al., 2008; HARHOLT et al., 2006).

O clado A inclui o gene MUR3 (At2g20370), que codifica uma galactosil transferase de xiloglucana, adicionando a unidade de (1,2)- $\beta$ -D-Gal no primeiro resíduo de xilose da extremidade redutora da unidade de heptassacarídeo de xiloglucanas (MADSON et al., 2003). O gene MUR3 pertence a um grupo ortólogo a outros genes em gramíneas, havendo ainda uma expansão em milho (Figura 13). Isso está de acordo com o observado na literatura, que demonstra atividade de MUR3 em arroz (Os03g05110) e Brachypodium (Bd1g75450) (LIU; PAULITZ; PAULY, 2015). O gene XLT2 não demonstra redundância funcional com MUR3, mas também é necessário para a substituição de galactosil em xiloglucanas em uma posição diferente (JENSEN et al., 2012). Uma mutação de perda de função na XUT1, gene do mesmo clado que MUR3 e XLT2, resultou na síntese de xiloglucana com ausência de ácido galactunônico (PENA et al., 2012). Entretanto, a XUT1 não possui ortologia com nenhum gene de gramíneas. Assim, percebe-se que membros do clado A desempenham atividade de galactosil transferase. Ao analisar a árvore filogenética (Figura 13), verifica-se que XLT2 de Arabidopsis possui ortologia com um grupo de dois genes de cada espécie de gramínea (Bd5g18927; Bd5g18940; Os06g23420; Os04g48480; Zm2G305146; Zm2G303135; Sb006G186100; Sb006G186200). O grupo ortólogo de XLT2 é consistente com o observado na literatura (LIU; PAULITZ; PAULY, 2015), onde Os04g48480 é classificado como um parálogo de AtXLT2.

Os genes *FRA8* e *F8H*, juntamente com os genes *PARVUS* e *IRX8* da GT8, estão envolvidos na biossíntese do tetraoligossacarídeo da extremidade redutora de GX (LEE et al., 2009; PENA et al., 2007). O gene *F8H*, é um homólogo de *FRA8* sendo um parálogo funcional, estando envolvido, também, na biossíntese do tetraoligossacarídeo da extremidade redutora de GX (LEE et al., 2009). Isso é consistente com a árvore filogenética (Figura 13), onde *FRA8* é parálogo a *F8H* no clado E.

*IRX10* e *IRX10L* também pertencem ao clado E. Foi evidenciado que existe uma considerável redundância funcional desses genes (BROWN et al., 2009). Esses genes redundantes atuam na elongação da cadeia principal de xilana das heteroxilanas, devido a presença de atividade de (1,4)- $\beta$ -xilosiltransferase pelas enzimas codificadas por esses genes (BROWN et al., 2009). A sequência proteica OsGT47A de arroz exibe uma identidade de 93.49% com IRX10 de Arabidopsis possuindo uma conservação funcional com essa proteína, estando envolvida com a síntese de GAX em *O. sativa* (ZHANG et al., 2014). Existe um outro ortólogo a IRX10 (Os01g70200), consistente com o observado na literatura (MOLINARI et al., 2013). O gene Os01g70200 possui ortologia com dois genes de milho (Zm2G152029; Zm2G000581) e um de sorgo (Sb003G410800) (Figura 13).

O gene *OsGUT1* codifica uma arabinosil transferase de AX em arroz (MITCHELL; DUPREE; SHEWRY, 2007). O gene *OsGUT1* é ortólogo a um grupo de gramíneas, onde existem dois genes de milho e sorgo (Zm2G059845; Zm2G056702; Sb009G220200; Sb003G410700) e um de *B. distachyon* (Bd2g59400) e ao próprio gene *OsGT47A* de arroz (Figura 13).

#### 4.2.5 Análise dos genes da família gênica GT61

A partir da anotação dos genes da família GT61, foi quantificado 10, 22, 26, 30 e 35 genes encontrados em *A. thaliana*, *B distachyon*, *O. sativa*, *S. bicolor e Z. mays*, respectivamente (Tabela 3). As enzimas envolvidas na biossíntese de hemicelulose codificadas pelos membros da GT61 possuem atividade de xilana arabinosiltransferase (ANDERS et al., 2012) e também estão envolvidas na transferência de resíduos de xilose da UDP-xilose na cadeia lateral das heteroxilanas (CHINIQUY et al., 2012).

O gene específico de gramíneas designado como *XAX1* foi caracterizado em arroz (Os02g22380) (CHINIQUY et al., 2012). Essa glicosiltransferase da família GT61 possui atividade de  $\beta$ -1,2-xilosil transferase, transferindo resíduo de xilose da UDP-xilose para a cadeia de xilana. Acredita-se que genes da família GT61, em arroz, estejam relacionados com a inserção de cadeias laterais na xilana, como é o caso da atividade de  $\alpha$ -(1,3)-arabinosil transferase (proteínas XAT – <u>x</u>ilana <u>a</u>rabinosil<u>t</u>ransferase) (ANDERS et al., 2012). Estudo de caracterização gênica em mutantes de Brachypodium, tendo como candidato uma GT61, demonstrou alteração na composição da parede celular (MARRIOTT et al., 2014) promovendo um maior rendimento no processo sacarificação.

Esse estudo fornece mais indícios do envolvimento da GT61 na substituição de AXs. O gene caracterizado em Brachypodium é um *XAX1*, similar ao de arroz.

A separação dos clados da GT61 (Figura 14B) em três grupos distintos foi feita seguindo a abordagem de classificação realizada pelo trabalho de (ANDERS et al., 2012). Os clados são divididos em A, B e C. Enquanto que o clado C se caracteriza por conter  $\beta$ -(1,2) xilosiltransferases envolvidas na N-glicosilação, o clado A é específico de plantas superiores e para a nossa análise é considerado o clado *bona fide*, por conter os genes *bona fides XAT* e *XAX1* de arroz e *XAX1* de *B. distachyon*.

A partir de análises de ortologia de genes pela árvore filogenética (Figura 14B), existem quatro genes *bona fides* para arroz, três já caracterizados em estudos (Os02g22480; Os03g37010; Os02g22380) e um ortólogo (Os01g02940) ao gene caracterizado em Brachypodium (Bd2g01480). Para Brachypodium existem quatro genes *bona fides*, sendo um já caracterizado e quatro ortólogos (Bd2g25450; Bd3g11310; Bd1g06560; Bd3g11337) aos genes de *O. sativa*. Os genes *bona fides* em *S. bicolor* e *Z. mays* foram definidos a partir de ortologia com os genes de *O. sativa* e *B. distachyon*. Assim, foram obtidos quatro genes (Sb003G094600; Sb004G007600; Sb010G144400; Sb004G134100; Zm2G074896; Zm2G096946; Zm2G130800; Zm5G875864) em *S. bicolor* e *Z. mays*.

#### 4.2.6 Análise dos genes da família gênica GT75

As enzimas codificadas pelos genes pertencentes a família GT75 realizam a interconversão de UDP-L-Ara*p* e UDP-L-Ara*f* (KONISHI et al., 2007). Essas enzimas são conhecidas como UAM (UDP-Arabinopiranose mutase) ou RGP (*Reversibly Glycosylated Protein*). Elas estão relacionadas com a biossíntese de GAX e XyG, devido a existência de arabinose na cadeia lateral dessas moléculas (ZENG et al., 2010) (DHUGGA et al., 1997). A UDP-L-Ara*p* é formada a partir de UDP-D-Xilose a partir da UXE, enzima da via de interconversão açúcar-nucleotídeo.

A família RGP em *A. thaliana* consiste de cinco membros altamente relacionados (RAUTENGARTEN et al., 2011) o que é consistente com a busca realizada. A busca feita retornou cinco sequências pertencentes a família GT75 (RGP) (Tabela 3). Os cinco genes da GT75 de *A. thaliana* estão relacionados com a interconversão de UDP-L-Arap e UDP-L-Araf (RAUTENGARTEN et al., 2011), sendo assim considerados *bona fides*. Abordagens com RNAi visando a geração de linhagens transgênicas mutantes para os

genes da família GT75 de *B. distachyon*, resultaram na alteração na composição de carboidratos da parede celular com diminuição significante de arabinose (RANCOUR et al., 2015). Esses resultados suportam a ideia de que a RGP/UAM é uma UDP-arabinopiranose mutase. O presente trabalho permitiu definir, a partir da anotação gênica, que os quatro membros encontrados em *B distachyon* são importantes candidatos da GT75. Em arroz, existem três genes *bona fides* da família GT75: *UAM1*, *UAM2* e *UAM3*, catalisando a interconversão de UDP-Araf e UDP-Arap (KONISHI et al., 2007), estando de acordo com o observado em *A. thaliana* e *B distachyon*. Segundo a busca feita por blast e hmmer, os três membros da família GT75 foram encontrados (Tabela 3), consistente com o observado em (KONISHI et al., 2007).

No caso de Z. mays e S. bicolor a determinação dos genes bona fides foi feita a partir da análise filogenética da família GT75 (Figura 11D) que permitiu a identificação de uma expansão com cerca de 13 genes em Z. mays, enquanto que nas outras espécies a quantidade de genes varia de três a cinco. Desses 13 genes, oito formam um grupo (*cluster*) entre si sugerindo que são genes específicos dessa espécie. Portanto, existem cinco genes *bona fides* em Z. mays. Segundo a árvore filogenética da GT75 (Figura 11D), os quatro genes de S. bicolor são genes *bona fides*.

#### 4.2.7 Análise dos genes da família gênica GT2 e CSL

Os genes CSL são agrupados em nove famílias denominadas como CSLA, B, C, D, E, F, G, H e J (VERHERTBRUGGEN et al., 2011). A biossíntese de heteromananas recruta enzimas CSLA e CSLD. Os genes CSLD2, CSLD3 e CSLD5 foram identificados e caracterizados em Arabidopsis thaliana e acredita-se que possuem atividade de manana além de formarem sintase. possivelmente um complexo enzimático (VERHERTBRUGGEN et al., 2011). Esses três genes estão inclusos dentro das CSL (Tabela 3) que possui 26 genes anotados. As enzimas CSLA1, CSLA2, CSLA3 e CSLA9 são responsáveis pela síntese de glucomanana em tecidos do caule, enquanto que a CSLA7 está envolvida na síntese de glucomanana em tecido embrionário (GOUBET et al., 2009; LIEPMAN et al., 2007). Existem evidências de que essas enzimas estão localizadas na membrana do Golgi (DAVIS et al., 2010). A família CSLA faz parte das GT2, que em Arabidopsis existem 14 genes anotados (Tabela 3).

As famílias CSLA e CSLC são altamente relacionadas entre si e são as famílias mais divergentes das outras CSLs. Assim, juntamente com a CSLA, a família GT2 possui a

CSLC. A proteína CSLC4 de Arabidopsis provavelmente está envolvida na síntese da cadeia principal de glucana da xiloglucana, demonstrando que a família CSLC esteja envolvida na biossíntese de xiloglucana (COCURON et al., 2007). A CSLC também possui localização na membrana do Golgi (DAVIS et al., 2010). Assim, *A. thaliana* possui nove genes *bona fides* do grupo das GT2 e CSL, sendo seis GT2 e três CSL.

Foram anotados 45 genes da superfamília CESA/CSL (GT2 e domínio celulose sintase) em arroz (WANG et al., 2010). O processo de busca do presente trabalho possibilitou a identificação de 49 proteínas de *O. sativa* contendo os domínios característicos dessa superfamília, quatro sequencias a mais identificadas (Os09g30105; Os07g36680; Os06g11050; Os09g26770).

O perfil de expressão dos genes da superfamília CESA/CSL de arroz foi determinado e verificou-se que houve expressão de pelo menos um gene de seis famílias CSL (*CSLA1, CSLC9, CSLD2, CSLE1, CSLF6* e *CSLH1*) em todos os tecidos examinados, sugerindo que o grupo das CSLs é essencial para e biossíntese da parede celular (WANG et al., 2010).

Os membros pertencentes a família CSLF e CSLH, são encontrados apenas em gramíneas acredita-se estejam envolvidos síntese de MLG e que na (VERHERTBRUGGEN et al., 2011). Um novo mecanismo de controle da estrutura de MLG foi proposto (JOBLING, 2015): A arquitetura do poro da membrana plasmática e a translocação do polissacarídeo crescente através da membrana controla como o monômero de glucana é inserido e coordenado no sítio ativo. Esse controle afeta a proporção de ligações  $\beta$ 1-3 e  $\beta$ 1-4 do polissacarídeo. Esse mecanismo de controle foi comprovado com a proteína CSLF6 das espécies de B distachyon, O. sativa, Z. mays e S. bicolor (JOBLING, 2015). A análise do perfil de expressão dos genes das famílias CSLF e CSLH em Brachypodium comprovou que BdCSLF6 está relacionado com a biossíntese de MLG, juntamente com os genes da família CSLH (CHRISTENSEN et al., 2009). Foi confirmado que além do CSLF6, os genes CSLF2, CSLF4, CSLF7, CSLF9 de O. sativa estão envolvidos com a biossíntese de MLG (BURTON et al., 2006; WANG et al., 2010) (VEGA-SÁNCHEZ et al., 2012).

A anotação dos genes *bona fides* em gramíneas foi feita através da revisão na literatura juntamente com a análise de ortologia a partir das árvores filogenéticas GT2 (Figura 11A) e CSL (Figura 14A). Para a CSLA (Figura 11A), apenas um gene

(Bd3g06740; Os02g09930; Zm2G105631; Sb004G075900) de cada espécie foi anotado, ortólogos ao *CSLA2* de Arabidopsis. Para a CSLC (Figura 11A), um gene (Bd1g07277; Sb001G075600) de cada espécie foi anotado por pertencerem ao grupo da *CSLC9* de *O*. *sativa*, com exceção de milho que possui dois genes no grupo (ZmAC183932.3 FGT007; Zm2G028286).

No caso da árvore filogenética CSL (Figura 14A), a família CSLD possui dois grupos ortólogos: um grupo ortólogo ao *CSLD5* de Arabidopis, com um gene de cada espécie (Bd4g05027; Os12g36890; Zm2G015886; Sb008G125700), e um grupo ortólogo aos genes *CSLD2* e *CSLD3* de Arabidopsis, contendo, também, um gene de espécie (Bd1g50170; Os06g02180; Zm5G870176; Sb010G008600), estando contido o *CSLD2* de *O. sativa*, As famílias CSLF e CSLH são específicas de gramíneas, não havendo a presença de genes de *A. thaliana*. Um grupo ortólogo ao gene *CSLH1* de *O. sativa* foi identificado, contendo um gene de *B. distachyon* (Bd5g10130) e *Z. mays* (Zm2G074546) e três de *S. bicolor* (Sb006G080600; Sb006G080700; Sb006G080800). No caso da CSLF, os genes CSLF6 foram anotados de acordo com (JOBLING, 2015) e sendo comprovada a relação de proximidade evolutiva com a (Figura 14A). Para milho, existem dois genes do grupo CSLF6, sendo classificado como *CSLF6* e *CSLF6-2* (JOBLING, 2015).

# 4.3 Identificação de genes candidatos envolvidos na biossíntese de hemicelulose em cana-de-açúcar e possivelmente envolvidos na recalcitrância da parede celular

Os genes *bona fides* de *Sorghum bicolor* e de *Zea mays* foram utilizados para identificar os genes candidatos na biossíntese de hemicelulose e possivelmente envolvidos na recalcitrância da parede celular nas sete bibliotecas de cana-de-açúcar. Foram obtidos 27 genes candidatos em cana-de-açúcar, a partir dessa abordagem. A Tabela 6 apresenta o resultado da quantidade de genes candidatos nas diferentes bibliotecas de cana-de-açúcar.

A diversidade genética de milho e de sorgo e a divergência evolutiva entre ambos (Figura 15) possibilita a busca de genes em cana-de-açúcar que possuem homologia com uma espécie e não com outra. É proposta a utilização de milho e sorgo como modelos chave para a descoberta de genes relacionados com a biomassa da parede celular e para estudos de melhoramento genético de gramíneas com potencial para produção de bioenergia (CARPITA; MCCANN, 2008).

O *Sorghum bicolor* possui um genoma considerado pequeno (Tabela 3), e é a espécie mais próxima evolutivamente de cana-de-açúcar com o genoma já sequenciado (Figura 15). Assim, *S. bicolor* torna-se um modelo atrativo para estudos de genômica

funcional das Saccharinae e outras gramíneas C4. Nesse contexto, a busca de sequências em *Sorghum bicolor e Z. mays* se torna de suma importância para estudar a cana-de-açúcar de forma indireta.

Tabela 6 - Número de genes candidatos relacionados com a biossíntese de hemicelulose e possivelmente envolvidas com a recalcitrância da parede celular em distintas bibliotecas de cana-de-açúcar. Em parênteses representa o número total de sequências ortólogas aos diversos genes *bona fides* e o número fora dos parênteses representa apenas as sequências completas. [1] SUCEST (VETTORE et al., 2003); [2] Raiz (BUCKERIDGE et al., in press); [3] Plântula (HEMERLY et al., in press); [4] Folha (CARDOSO-SILVA et al., 2014); [5] Folha (GRATIVOL et al., 2014); [6] Entrenó (VICENTINI et al., 2015); [7] Folha e Entrenó (NISHIYAMA JR et al., 2014)

Genes\Bibliotecas	[1]	[2]	[3]	[4]	[5]	[6]	[7]	Total
UXS (Epimerase)	-	1 (4)	2 (3)	1 (2)	(2)	-	-	4 (11)
AXS (Epimerase)	(1)	-	(1)	(1)	-	-	(1)	(4)
UGD (Epimerase)	2 (3)	2 (2)	(2)	-	-	1 (1)	(2)	5 (10)
UXE (Epimerase)	(2)	1 (3)	1 (2)	(2)	(1)	(1)	-	2 (11)
<i>RGP/UAM</i> (GT75)	2 (4)	1 (4)	(2)	1 (2)		(1)	1 (1)	5 (14)
GUX (GT8)	(1)	-	-	-	(2)	-	(1)	(4)
<i>IRX9</i> (GT43)	(1)	-	(1)	(1)	-	(1)	(4)	(8)
<i>IRX9L</i> (GT43)	1 (1)	1 (1)	(1)	1 (1)		1 (1)	(1)	4 (6)
<i>IRX14</i> (GT43)	(1)	(1)	(1)	(1)	-	(1)	1 (1)	1 (6)
<i>IRX10</i> (GT47)	-	-	-	-	(2)	(1)	(1)	(4)
<i>XAT</i> (GT61)	1 (2)	1 (2)	(2)	1 (2)	(1)	(2)	-	3 (11)
XAX (GT61)	1 (2)	(1)	(2)	(1)	-	(2)	(3)	1 (11)
CSLF (CSL)	1 (1)	(1)	(1)	(1)	-	1 (1)	-	2 (5)
CSLH (CSL)	(2)	(1)	(5)	(1)	-	(2)	-	(11)
Total	8 (21)	7 (20)	3 (23)	4 (15)	(8)	3 (14)	2 (15)	27 (116)
		Fonte	· Autori	nrónria				

Fonte: Autoria própria

Foi demonstrado que manipular a expressão de genes relacionados apenas com a via de interconversão açúcar-nucleotídeo não leva a alterações significativas na composição da parede celular (GONDOLF et al., 2014). Entretanto, a simultânea alteração da expressão de um gene *NSE* junto com uma transferase leva a resultados significativos. Assim, no caso da gramínea cana-de-açúcar os genes que codificam as enzimas UXS, AXS, UGD e UXE são candidatos do grupo das *NSE*, pois levam a formação de moléculas de UDP-D-xilose, UDP-D-xilose e UDP-D-Apiose, UDP-D-Ácido Glucurônico e UDP-L-Arabinose, respectivamente. Os genes da família GT75 codificam proteínas que também atuam na interconversão, realizando a interconversão de UDP-L-Arap e UDP-L-Araf, sendo, portanto, genes alvos. Obteve-se, através de análises filogenéticas e alinhamentos

com *S. bicolor* e *Z. mays*, 16 genes candidatos da via de interconversão açúcar-nucleotídeo e da família GT75 em cana-de-açúcar.

A UXS possui três genes completos ortólogos a Sb001G418100 (Figura 16A), apenas um gene completo ortólogo a Sb003G348800, e três genes ortólogos a Sb003G149500 sendo que nenhum dos três está completo. Existem quatro genes ortólogos ao gene AXS, Sb003G442300, entretanto nenhum possui anotação completa. Para a UGD, 10 genes foram identificados em cana-de-açúcar, sendo três completos ortólogos a Sb001G459800 (Figura 16B) e dois completos ortólogos a Sb001G084100, totalizando cinco genes *bona fides* para UGD. No caso da UXE, existem sete genes ortólogos a Sb001G400000, mas nenhum completo e quatro ortólogos a Sb006G217300, com dois genes completos.

Da família GT75, foram anotados 14 genes de cana-de-açúcar, sendo um gene completo ortólogo a Sb002G368600 e quatro genes completos ortólogos a Sb001G173300 (Figura 17).

As principais hemiceluloses presentes na parede celular de gramíneas e, portanto, em cana-de-açúcar são as heteroxilanas (GAX, majoritariamente) e MLG. Assim, a seleção de genes que codificam enzimas envolvidas na biossíntese da cadeia principal de xilana das heteroxilanas de cana-de-açúcar para futura alteração na expressão é uma estratégia que possivelmente possibilita o melhoramento na sacarificação da biomassa lignocelulósica. A inativação de um gene envolvido na biossíntese da cadeia de xilana em arroz, demonstrou uma redução no conteúdo de xilana e uma melhora na sacarificação (CHEN; ALONSO; SHACHAR-HILL, 2013).

Assim, é interessante a identificação e seleção de genes das famílias GT43 e GT47 em cana-de-açúcar. A partir da análise das árvores filogenéticas via ortologia de sequencias pelo PhyML, foram identificados, para a família GT43, seis genes ortólogos ao Sb009G229300, dentre os quais quatro estão com as suas sequencias completas (Figura 18). O gene Sb009G026100 possui ortologia com oito sequências de cana-de-açúcar, entretanto nenhuma está completa, já o gene Sb010G238800 possui apenas uma sequência completa dentro do grupo com seis sequências ortólogas. No caso da família GT47, existem quatro sequências ortólogas a Sb003G410800, entretanto nenhuma possui anotação completa.

Genes envolvidos na decoração da cadeia de xilana também são candidatos envolvidos na recalcitrância da parede celular, devido a formação de ligações cruzadas na

parede celular através das cadeias laterais da hemicelulose (CAFFALL et al., 2009; CHINIQUY et al., 2012; SOMERVILLE et al., 2004). Assim, os genes das famílias GT8 e GT61 são candidatos na recalcitrância da parede celular de cana-de-açúcar. Existem quatro sequências de cana-de-açúcar ortólogas a Sb001G479800 (GT8) de *S. bicolor*, gene possivelmente envolvido na adição dos substituintes glucuronosil e 4-O-metil-glucuronosil na cadeia principal de xilana, entretanto nenhuma delas está completa.

Foram identificados 11 genes possivelmente relacionados com a atividade de  $\alpha$ -1,3arabinosil transferase (XAT-GT61) em cana-de-açúcar a partir de ortologia com os genes de *S. bicolor*. Três genes ortólogos a Sb010G144400 (Figura 19), dentre os 11 identificados, estão completos e nenhum gene ortólogo a Sb004G134100 está completo. Ainda, para a família GT61, foram identificados 11 genes possivelmente relacionados com a atividade  $\beta$ -1,2-xilosil transferase (XAX) em cana-de-açúcar, sendo seis ortólogos a Sb004G007600 com uma única sequência completa, SCJRT1015F06.g, do banco do SUCEST [1] e cinco genes incompletos ortólogos a Sb03G094600 totalizando quatro genes *bona fides* da GT61 em cana-de-açúcar.

Existe uma grande quantidade de MLG na parede celular de gramíneas e consequentemente na parede celular de cana-de-açúcar tornando, portanto, os genes responsáveis pela biossíntese de MLG, candidatos na recalcitrância da parede celular. Foi demonstrado que os genes das famílias CSLF e CSLH são responsáveis por codificar enzimas envolvidas na biossíntese de MLG (CHRISTENSEN et al., 2009; JOBLING, 2015; VEGA-SÁNCHEZ et al., 2012; VERHERTBRUGGEN et al., 2011). Foram identificadas cinco sequencias ortólogas ao gene *CSLF6* (Sb007G050600) de sorgo, tendo duas sequências completas dos bancos do SUCEST [1] e biblioteca de entrenó [6]. Existem 11 sequencias de cana-de-açúcar ortólogas aos genes *CSLH* Sb006G080700 de *S. bicolor*, entretanto nenhuma sequência está completa.

Figura 10 - Árvore filogenética das proteínas do clado A com domínio epimerase das espécies *A. thaliana, B. distachyon, O. sativa, S. bicolor* e *Z. mays.* Os códigos dos genes extraídos do Phytozome para as diferentes espécies sofreram alteração para se adaptarem a figura. Para Brachypodium, o código "Bradi" foi modificado para "Bd"; para *Oryza sativa* o código "LOC\_Os" foi modificado para "Os"; para *Z. mays* o código "GRMZM" foi modificado para "Zm" e para *S. bicolor* o código foi alterado de "Sobic." para "Sb".



Fonte: Autoria pópria

Figura 11 - Árvores filogenéticas das proteínas das espécies *A. thaliana, B. distachyon, O. sativa, S. bicolor e Z. mays* das famílias gênicas GT2, UGD, GT43 e GT75. A) Análise da família gênica GT2 e B) Família gênica UGD. C) Família gênica GT43 D) Família gênica GT75. Os códigos dos genes extraídos do Phytozome para as diferentes espécies sofreram alteração para se adaptarem a figura. Para Brachypodium, o código "Bradi" foi modificado para "Bd"; para *Oryza sativa* o código "LOC\_Os" foi modificado para "Os"; para *Z. mays* o código "GRMZM" foi modificado para "Zm" e para *S. bicolor* o código foi alterado de "Sobic." para "Sb".



Fonte: Autoria própria

Figura 12 - Árvore filogenética das proteínas com domínio GT8 das espécies *A. thaliana, B. distachyon, O. sativa, S. bicolor* e *Z. mays.* Os códigos dos genes extraídos do Phytozome para as diferentes espécies sofreram alteração para se adaptarem a figura. Para Brachypodium, o código "Bradi" foi modificado para "Bd"; para *Oryza sativa* o código "LOC\_Os" foi modificado para "Os"; para *Z. mays* o código "GRMZM" foi modificado para "Zm" e para *S. bicolor* o código foi alterado de "Sobic." para "Sb".



Fonte: Autoria própria

Figura 13 - Árvore filogenética das proteínas com domínio GT47 das espécies *A. thaliana, B. distachyon, O. sativa, S. bicolor* e *Z. mays.* Os códigos dos genes extraídos do Phytozome para as diferentes espécies sofreram alteração para se adaptarem a figura. Para Brachypodium, o código "Bradi" foi modificado para "Bd"; para *Oryza sativa* o código "LOC\_Os" foi modificado para "Os"; para *Z. mays* o código "GRMZM" foi modificado para "Zm" e para *S. bicolor* o código foi alterado de "Sobic." para "Sb".



Fonte: Autoria própria

Figura 14 - Árvores filogenéticas das proteínas das espécies *A. thaliana, B. distachyon, O. sativa, S. bicolor* e *Z. mays* das famílias gênicas CSL e GT61. A) Análise da família gênica CSL e B) Família gênica GT61. Os códigos dos genes extraídos do Phytozome para as diferentes espécies sofreram alteração para se adaptarem a figura. Para Brachypodium, o código "Bradi" foi modificado para "Bd"; para *Oryza sativa* o código "LOC\_Os" foi modificado para "Os"; para *Z. mays* o código "GRMZM" foi modificado para "Zm" e para *S. bicolor* o código foi alterado de "Sobic." para "Sb".



Fonte: Autoria própria



Figura 15 - Representação de uma árvore filogenética de representantes da subfamília Panicoideae. A árvore filogenética revela a origem e o período de divergência das espécies em milhões de anos.

Fonte: Adaptado de (SETTA et al., 2014).

Figura 16 - Alinhamento de sequências contendo o domínio epimerase e UGD dos diferentes bancos de canade-açúcar com genes *bona fides* de *S. bicolor*. A) TCONS\_00008763 (Raíz), R1\_L412\_T1 (Plântula) e comp195955 (Folha) estão completas comparadas com Sobic.001G418100. B) SCJFLR1013H10 (SUCEST), TCONS\_00014608 (Raíz) e Locus1344.4 (Entrenó) estão completas comparadas com Sobic.001G459800.

A)

Sobic.001G418100 TCONS_00008763 TCONS_00026762 R1_L412_T1 comp195955	MAQKETNG MAQKETNG MAQKETNG MAQKETNG	10 SNGEHIS SNGEHIS SNGEHIS SNGEHIS	20 T R P P P T P S T R P P P T P S T R P P P T P S T R P P P T P S	P L R F S I P L R F S I P L R F S I P L R F S I	30 CFFQAN CFFQAN CFFQAN	 L R I L V L R I L V  L R I L V L R I L V	40 T G G A G I T G G A G I T G G A G I T G G A G I	50 F I G S H L V F I G S H L V F I G S H L V F I G S H L V	DKLMENE DKLMENE DKLMENE DKLMENE	60 KH F - KH
Sobic.001G418100 TCONS_00008763 TCONS_00026762 R1_L412_T1 comp195955	E V I VA D N L E V I VA D N F E V I VA D N F E V I VA D N F	70 FTGSKDN FTGSKDN FTGSKDN FTGSKDN	80 L K K W I GH P L K K W I GH P L K K W I GH P L K K W I GH P	R F E L I F R F E L I F R F E L I F		P L L V E P L L V E  P L L V E P L L V E		110 1 L A C P A S 1 L A C P A S 1 L A C P A S 1 L A C P A S	PIFYKHN PIFYKHN PIFYKHN PIFYKHN	120 - I P V P V  P V P V
Sobic.001G418100 TCONS_00008763 TCONS_00026762 R1_L412_T1 comp195955	KT I KT NVI KT I KT NVI KT I KT NVI KT I KT NVI KT I KT NVI	130 GTLNMLG GTLNMLG GTLNMLG GTLNMLG	140 L A K R V G A R L A K R V G A R L A K R V G A R L A K R V G A R	ILLTSI ILLTSI ILLTSI ILLTSI	150 SEVYG SEVYG SEVYG	D P L E H D P L E H D P L E H D P L E H D P L E H	160 PQTEA PQTEA PQTEA PQTEA	170 YWGNVNP YWGNVNP YWGNVNP YWGNVNP	I G V R S C Y	180 D E D E  D E D E
Sobic.001G418100 TCONS_00008763 TCONS_00026762 R1_L412_T1 comp195955	G KR VA ET L G KR VA ET L G KR VA ET L G KR VA ET L	190 M F D Y H R Q M F D Y H R Q M F D Y H R Q M F D Y H R Q	H G I E I R I A H G I E I R I A H G I E I R I A H G I E I R I A	R I F NT Y R I F NT Y R I F NT Y R I F NT Y	GPRMN GPRMN GPRMN GPRMN GPRMN	I DDGR I DDGR DGR I DDGR I DDGR	220 VVSNF1 VVSNF1 VVSIF1 VVSNF1 VVSNF1 VVSNF1	230 I A Q A V R G I A Q A V R G	E P L T VQR E P L T VQR E P L T VQR E P L T VQR E P L T VQR	240 P G P G P G P G P G
Sobic.001G418100 TCONS_00008763 TCONS_00026762 R1_L412_T1 comp195955	TQTRSFCY TQTRSFCY TQTRSFCY TQTRSFCY TQTRSFCY	250 V A DM V D G V A DM V D G	260 LIKLMNGN LIKLMNGN LIKLMNGN LIKLMNGN LIKLMNGN	NT GP II NT GP II NT GP II NT GP II NT GP II	270 .   N L G N P G N L G N P G	E FTML E FTML E FTML E FTML E FTML	280 E L A E N V E L A E N V	290     . V K E L I N P V K E L I N P	D V T V T M T D V T V T M T	300 E N E N E N E N E N
Sobic.001G418100 TCONS_00008763 TCONS_00026762 R1_L412_T1 comp195955	T P DD P R Q R T P DD P R Q R	310 K P D I T KA K P D I T KA	320 KEVLGWEP KEVLGWEP KEVLGWEP KEVLGWEP KEVLGWEP	KIVLKI KIVLRI KIVLRI KIVLRI KIVLRI	330 GLVLM GLVLM GLVLM GLVLM GLVLM GLVLM 10	E D D F R E D D F R E D D F R E D D F R E D D F R	340 E R L A V I E R L A V I 20	350 P K K T K A P K K T K A	30	40 50 60
		B) Sob SCI TCO R1_ Loca Trini Trini	ic.001G459800 FLR1013H10 NS_00014608 L8219_T1 Js1344.4 ity_comp65759 ity_comp20649	MVKIC MVKIC MVKIC MVKIC MVKIC	CLGAGY CLGAGY CLGAGY CLGAGY CLGAGY CLGAGY CLGAGY	V GG P TI V GG P TI	MAVIAL MAVIAL MAVIAL MAVIAL MAVIAL MAVIAL	KCPAIEN KCPAIEN KCPAIEN KCPAIEN KCPAIEN Q - PMSPI	I       I       I       I         I       V       V       I       S         I       V       V       I       S         I       V       V       I       S         I       V       V       I       S         I       V       V       I       S         I       V       V       I       S         I       V       V       I       S         I       V       V       I       S         I       V       V       I       S         I       V       V       I       S         I       V       V       I       S         I       V       V       I       S	PRIAAWNSDQLPIYEPGLDDVVKQ         PRIAAWNSDQLPIYEPGLDVVKQ         PRIAAWNSDQLPIYEPGLDVVKQ         PRIAAWNSDQLPIYEPGLDVVKQ         PRIAAWNSDQLPIYEPGLDVVKQ         PRIAAWNSDQLPIYEPGLDVVKQ         PRIAAWNSDQLPIYEPGLDVVKQ         Torverence         PRIAAWNSDQLPIYEPGLDVVKQ         PRIAAWNSDQLPIYEPGLDVVKQ         PRIAWNSDQLPIYEPGLDVVKQ         Torverence         Torvere         Torverence
		Sob SCJI TCO R1_ Loct Trin Trin	ic.001G459800 FLR1013H10 NS_00014608 L8219_T1 is1344.4 ity_comp65759 ity_comp20649	CRGRNI CRGRNI CRGRNI CRGRNI CRGRNI CRGRNI CRGRNI DEFKAI	70 LFFSND LFFSND LFFSND LFFSND LFFSND LFFSND LFFSND	I E K H V I E K H V	A E A D I I A E A D I I Q K P A F L	FVSVNT FVSVNT FVSVNT FVSVNT FVSVNT FVSVNT FVSVNT	90 TKTRGLG TKTRGLG TKTRGLG TKTRGLG TKTRGLG TKTRGLG TKTRGLG	100 A G K A A D L T Y WE SA A R M I A D V A K S D A G K A A D L T Y WE SA A R M I A D V A K S D A G K A A D L T Y WE SA A R M I A D V A K S D A G K A A D L T Y WE SA A R M I A D V A K S D A G K A A D L T Y WE SA A R M I A D V A K S D A G K A A D L T Y WE SA A R M I A D V A K S D A G K A A D L T Y WE SA A R M I A D V A K S D A G K A A D L T Y WE SA A R M I A D V A K S D A G K A A D L T Y WE SA A R M I A D V A K S D A G K A A D L T Y WE SA A R M I A D V A K S D
		Sob SCJI TCO R1_ Loci Trin Trin	ic.001G459800 FLR1013H10 NS_00014608 L8219_T1 Is1344.4 ity_comp65759 ity_comp20649	KIVVE KIVVE KIVVE KIVVE KIVVE MPAVA	130 C ST V P V C ST V P V C ST V P V C ST V P V C ST V P V	KTAEA KTAEA KTAEA KTAEA KTAEA		HNSKGII HNSKGI HNSKGI HNSKGI	150 NFQILSNP NFQILSNP NFQILSNP	EFLAEGTAIEDLFKPDRVLIGGRE EFLAEGTAIEDLFKPDRVLIGGRE EFLAEGTAIEDLFKPDRVLIGGRE EFLAEGTAIEDLFKPDRVLIGGRE
		Sob SCJI TCO R1_ Loca Trini Trini	ic.001G459800 FLR1013H10 NS_00014608 L8219_T1 is1344.4 ity_comp65759 ity_comp20649	T P E GQ T P E GQ T P E GQ	190 KAVKAL KAVKAL KAVKAL	K D V Y A K D V Y A K D V Y A K D V Y A		RILTTNI HILTTNI HILTTNI HILTTNI	210 WSAELSK WSAELSK WSAELSK	220 220 240 270 270 270 270 270 270 270 270 270 27
		Sob SCJI TCO R1_ Locc Trin Trin	ic.001G459800 FLR1013H10 NS_00014608 L8219_T1 is1344.4 ity_comp65759 ity_comp20649	GANVA GANVS GANVS GANVS	250 VAYAV VAYAV VAYAV	G KD S R G KD S R G KD S R G KD S R		NASVGFO NASVGFO NASVGFO	270 G S C F Q KD G S C F Q KD G S C F Q KD G S C F Q KD	280 290 300 ILNLVYICECNGLPEVANVWKQVI ILNLVYICECNGLPEVANVWKQVI ILNLVYICECNGLPEVANVWKQVI ILNLVYICECNGLPEVANVWKQVI ILNLVYICECNGLPEVANVWKQVI
		Sob SCJI TCO R1_ Loca Trini Trini	ic.001G459800 FLR1013H10 NS_00014608 L8219_T1 is1344.4 ity_comp65759 ity_comp20649	K I NDYO K I NDYO K I NDYO	SR FV KSR FV KSR FV	NRVVS NRVVS NRVVS	320 SMFNTV SMFNTV SMFNTV SMFNTV	S G K K I A Y S G K K I A Y S G K K I A Y	330 /LGFAFKK /LGFAFKK /LGFAFKK	340 350 360 360 360 360 360 360 360 360 360 36
		Sob SCJI TCO R1_ Loc Trin Trin	ic.001G459800 FLR1013H10 NS_00014608 L8219_T1 Is1344.4 ity_comp65759 ity_comp20649	SIYDPO SIYDPO SIYDPO SIYDPO		I Q R D L I Q R D L I Q R D L I Q R D L	380 A M N K F D A M N K F D A M N K F D	WDHPIHI WDHPIHI WDHPIHI WDHPIHI	390 QPMSPTA QPMSPTA QPMSPTA	400 410 420 420 420 420 420 420 420 420 420 42
		Sob SCJ TCO R1_ Loc Trin Trin	ic.001G459800 FLR1013H10 NS_00014608 L8219_T1 us1344.4 ity_comp65759 ity_comp20649	EWDEFI EWDEFI EWDEFI	430 KALDYK KALDYK KALDYK KALDYK	K I Y D SI K I Y D SI K I Y D SI K I Y D SI K I Y D SI	440 MOKPAF MQKPAF MQKPAF	L FD GR N L FD GR N L FD GR N L FD GR N	450 / I D A E KM R / I D A E KM R / I D A E KM R	460 E I G F I V Y S I G K P L D PWL K DM PA VA E I G F I V Y S I G K P L D PWL K DM PA VA E I G F I V Y S I G K P L D PWL K DM PA VA E I G F I V Y S I G K P L D PWL K DM PA VA

Fonte: Autoria própria

Figura 17 - Alinhamento de sequências GT75 dos diferentes bancos de cana-de-açúcar com o gene *bona fide* Sobic.001G173300 de *S. bicolor*. SCAGLR1021B05.g (SUCEST), TCONS\_00001314 (Raíz), comp179435 (Folha) e Trinity\_comp76780 (Folha e entrenó) estão completas e são genes candidatos.

							1	10						2	0						3	0						40						5	D						60
Sobic.001G173300	M	A	т	v.	٢N	P	Ġ	AS	v	PS	Ť.	PL	Ľ.	ĸ	È	É.	DI	v	i	ΡT	i	R N	Ú.	DF	ιù,	EM	Ŵ	P	FF	QF	Ŷ	Ĥ I	Ĺİ	i١	Q	D	G D	PT	ĸ	ΓI	ĸ
SCAGLR1021B05	м	A (	Т	۷.	٢N	P	G	s s	۷	P S	т	PL	. Ц	K	) E	L	DI	۷	I	РТ	I	RN	L	DF	L	EM	W	Р	FF	QF	P Y	н	LI	I١	Q	D	G D	РТ	. <b>K</b> .	r I	ĸ
SCEZHR1049B08			-			-	-		÷.		2		Ξ.	2.	1	2		5	23		2.2		5.		Ξ.			-				2.2	1	2.7						1	5
TCONS_0001314		A (		Ľ.		٢	G			r 5	2	P L	5			5			1		11				5	E M			55	<u>ب</u>					Q		3 <b>D</b>	P 1	<u>.</u>	11	2
R1 L153 T3	-		-			-			-		-		-			-		-							-			-							-						-
comp179435	м	A (	т	V.	r١	P	G	s s	۷	PS	т	PL	ь.	ĸ	) E	L	DI	v	I	РТ	I	RN	L	DF	L.	EM	W	Р	FF	QF	P Y	н	LI	I١	l Q	D	G D	РТ	' <b>K</b> '	ΓI	ĸ
Locus327.10			-	5.1							2			2.2	1	7		-	2.2						Ξ.			-						2.7	1						5
Trinity_comp76780	м	A G	Т	V.	r v	P	G	s s	v	PS	т	PL	· L	ĸ	) E	L	DI	v	I	РТ	I	RN	L	DF	L	EM	W	Р	FF	QF	P Y	н	I	I	Q	D	G D	РТ	. <b>к</b>	r I	ĸ
				1			. '	1.			1			. 8	U   .			1			. 9	1 .			1			100 ·			- 1			. 11	0			1			120
Sobic.001G173300	۷	PE	G	F	D١	Ē	Ľ	ÝN	R	ND	I	NR	I	L	G P	ĸ	A S	Ċ	15	5 F	ĸ	) S	A	CR	Ċ	FG	Y١	١V	s ĸ	ĸ	¢Ý	I	ΥT	IC	D	D	CF	V A	K	D P	т
SCAGLR1021B05	v	PE	G	F	D١	E	L,	YN	R	ND	I	NR	I	L	G P	ĸ	A S	С	1 5	5 F	ĸ	D S	A	CR	С	FG	YN	۱V	s ĸ	ĸ	C Y	I	ΥT	IC	D	D	CF	V A	K	P	Ţ
TCONS 00001314	v			E.		Ē	1.5	<b>v</b> N	Ē	ND	÷	ND	÷	1	- D		<u>,</u>	ċ	Ŧ		× I				è	FO	v	i v	с <b>к</b>		~~	-	<u>и</u> т	T P	. n	<b>n</b> (		v			÷
TCONS_00013055	1		-			-	Ξ.		-		2		-	-		2		-	-		-		2		-					-		2.									2
R1_L153_T3	-		-			-			-		-		-			-		-							-			-							-						-
comp179435	v	PE	G	F	D١	Ē	г,	YN	R	ND	I	NR	I	L	G P	ĸ	A S	С	1 5	5 F	ĸ	DS	A	CR	С	FG	YN	۱V	s ĸ	ĸ	C Y	I	ΥT	IC	D	D	CF	V A	K	P	Τ
Locus327.10	5		-	21		-	73				÷		Ξ.	1		÷		÷.	12		2.		-		÷					2.		23						V A	K	P	Ţ.
mincy_comp/0/80	•					-	۰.		ĸ		1				3 P		<b>M</b> 3	C			~	3	~		C	r e			- <b>n</b>	~ '											÷.,
							1	30						14	40						1	50						160						17	0						180
	2			I.		2	÷.	Ĩ.	÷		J.				ĭ.	÷	12	1	2		÷	ī .	2		1	1.5		· ï		1	• 1	2.			í.	÷.,	• •	1			I
Sobic.001G173300	G	KC	ļ	N	1	. E	Q	ΗI	ĸ	NI	M	SP	S	11	? <b>F</b>	E	FA	IT.	53	D	P	Y R	E	GA	D	FV	R	Ŷ	PF	SL	. R	E	σŤ	83	A	VS	5 H	GL	.wi	L N	I
SCAGLR1021B05	G	KL	Ì	N		Ē	à	HT	ĸ	NI	M	5 1	S	Ť		4	FN	H	23	ע אין אין	P	T R Y R	튍	G A	D	FV	R	÷.	PF	SI	R	EC		нт	Â	v	5 H	GI	wi		÷.
TCONS_00001314	Ğ	KC	ī	N	Ň	Ē	õ	нî	ĸ	NI	м	S F	ŝ	τi	Ì	÷Ē.	FN	iτ	ĩ١	rD	P	YR	Ē	GA	D	FΫ	R	Ŷ	PF	s i	R	Ē	GТ	нi	Â	v s	5 H	ĞL	w	L N	î
TCONS_00013055	-		-			-	-		-		-		-			-		-							-			-							-						-
R1_L153_T3	-		1	1		2	-		5		ċ.		-			2	11	2	1				20		-			-			-	2.2									2
COMP1/9435	G	KL	1	N		Ē	8 N		ĸ	NI	M	5 1	5	+ 1		4		#	23		P	7 K	닅		D D		R C	÷		SI	- K	E			•	v		GL	. w i		÷
Trinity comp76780	Ğ	ĸc	ī	N	ì	Ē	ŏ	нî	ĸ	NI	м	SF	ŝ	τi		÷.	FN	ίŤ	ĩ	rĎ	P	YR	Ē	GA	Ď	FΫ	R	Ŷ	PF	si	R	Ē	ĞТ	ні	Â	v s	ŚН	ĞL	w	L N	î.
							1	90						2	00						2	10						220						23	0						240
Sobic 0016173300	Þ		'n		 т с	0	ć,	, .	D	K F		NE	Þ	÷,	. / n		v N	i÷.	÷.		ċ	ł.	É.	 РМ		G M	N I	, L	E D	, P r	51	÷,	 D		v	Ê,		MO			P
SCAGLR1021B05	P		D	Â	ΡT	ŏ	Ъ	νĸ	P	KE	R	N	R	÷		Â	V N	ι÷.	ii	PK	G	έč	F	PM	č	GM	N	č	FD	RC	δĒ	i	GP	AN	Ý	FC	GL	M	D	GÕ	P
SCEZHR1049B08	P	D١	D	AI	РТ	ò	Ē	V K	Ρ	ΚE	R	NE	R	Ŷ	/ D	A	٧M	İΤ	Ī	ΡK	G	τĽ	F	PM	Ċ	GM	N	č	FD	RC	Ĺ	I	G P	AN	İΫ	F	G L	M	D	۶õ	P
TCONS_00001314	Р	D١	D	A I	РТ	Q	Ľ	V K	Ρ	K E	R	NE	R	Y١	/ D	A	٧M	Т	I	P K	G	T L	FI	ΡM	С	GM	N	c	FD	RC	L	I	G P	AM	Y	F	GL	M	D	GQ	Ρ
TCONS_00013055	-		-			-	L '	V K	Р	KE	R	NE	R	Y	/ D	A	VM	Т	I	P K	G	r L	FI	РМ	С	GM	N	c	FD	RC	L	1 (	G P	AN	I Y	FO	GL	MG	D	۶Q	P
comp179435	P	с. р у	n		 Р Т	0	ī.	v <b>k</b>	P	K F	R	NE	R	v	/ n	Ā	VN	ιŤ.	T	P K	G	r i	E I	РМ	i c	GM		c	FD	RF	51	T	G P	A N	v	FC	31	м 6		- Q	P
Locus327.10	P	D	D	A	РТ	ŏ	ĩ١	vк	P	ĸΕ	R	NE	R	Ϋ́́	/ D	A	٧M	iτ	ÎÌ	ΡK	Ğ	τĩ	Ē	РМ	č	GM	N	č	FD	RC	δĒ	Î	G P	AN	İŶ	F	ΞĒ	M	D	σõ	P
Trinity_comp76780	Р	D١	D	A I	РТ	Q	Ľ	v ĸ	Ρ	K E	R	NE	R	Y١	/ D	A	٧M	Т	I	P K	G	T L	FI	ΡM	C	GM	N	c	FD	RC	L	I	G P	AM	I Y	FO	GL	M (	D	GQ	Ρ
							2	50						21	50						2	70						280						25	10			τ.			300
Sobic.001G173300	I	GR	Y	DI	DN	w	A	Ġw	IC.	vк	v	IC	D	H I	s	É.	G۷	ĸ	T (	GL	P	ΎΙ	w	I S	ĸ	AS	NI	F	V N	Ĺ.F	ίĸ	Ē١	ĸκ	G 1	F	w	2 E	D I	I	PF	É.
SCAGLR1021B05	I	GR	Y	DI	DM	W	A	GW	C	V K	۷	IC	D	нı	. s	L,	G۷	K	т	GL	P	ΥI	w	I S	ĸ	A S	N	F	V N	LF	١ĸ	E	ĸκ	G 1	F	wq	ξE	D 1	I	PF	F
SCEZHR1049B08	÷	GR	Ľ	D		W	A	GW	c	VK	v	ič	D	H	. s	÷	GV	ĸ	Ţ	GL	P	Ϋ́Ι	w	S	ĸ	AS	N	. <u>F</u>	VN	15	ιĸ	E )	Υĸ	GI	E	w	S E	DI	Ţ	PF	5
TCONS_0001314	÷	GR	Ŷ	Ы	אינ	w	A	GW	ic.	vk	v			21	. 3 S	v	TA	S	÷	3 L	P .		~			A 3		5	• •	2.5		5		v	F	VI	2 6	NI	M	5	5.
R1_L153_T3	î	GR	Ŷ	D	DN	iw	A	GW	ič	v ĸ	v	ĩc	D	нi	s	i.	Ġν	ĸ	τċ	GL	P	ΥI	w	I S	ĸ	A S	N	F	V N	LB	١к	E١	ĸк	GI	F	ŵ	Ē	DI	I	P F	E.
comp179435	I	GR	L Y	D	DM	W	A	GW	C	V K	۷	IC	D	н	. s	L	G۷	ĸ	т	GL	P	ΥI	w	S	ĸ	A S	N	F	V N	LF	١ĸ	E١	ĸκ	G 1	F	wq	ξE	D 1	I	PF	F.
Locus327.10	ï	GR	Y	D	DM	W	A	GW	C	VK	v	ic	D	H	. s	Ŀ.	G -								Ξ.			1		2.2		2.3			1		E	DI	I	PF	£ .
Trinity_comp76780	-	GR	· ·	וט	ייי	1 W	A	GW	C	VK	•	10	. D	н	. 5	-	GV	ĸ		3 L	Р	11	w	15	ĸ	AS			V IN		сĸ	E 1	rĸ	GI		w	ξĒ	נט		PF	P
							2	10						2	20						2	20						240						25	0						260
	2			T			:	ĵ.	÷		Ţ				ï.	÷		1	÷			ĩ.	÷		Т			.			• 1				ĩ			1			1
Sobic.001G173300	Q	NN	<u>'</u> <u>T</u>	1	P		C	DT	v	QK	ç	Y I	Y	LS	G	Q	V	E	ĸ	G	1	ID	P	Y F	v	KL		A	мv	TV	VI.	E /	w	DE	÷	N	<b>P A</b>	ŢĘ	A	Ē	N
SCAGLK1021805	Q K	N N	1	1	-	U	CI.			ų K	. U	11	1	13	9 G	5	V 8		ĸ	G	ĸ	IJ	2	ľ	v	ĸL	AL	A	M V	1.4	VI.	- /	• vV	D E	1	N	A	1.0	A	ι E	N
TCONS 00001314	ò	NN	/ т	ī	P	D	c	DT	v	O K	c	ΥL	Ŷ	Ĺ	G	Ē	VK	Ē	ĸ	G	ĸ	IP	P	YF	v	KL	A	A	мv	τv	VI	Ê.	w	DE	ī.	N	PA	TP	A	E	N
TCONS_00013055	Ĩ	GL	. М	I	LÌ	N	Ċ	кŤ	۷	Ť -	-		-			-		-							-										-						-
R1_L153_T3	Q	NN	<u>'</u> T	I	P	D	C	DT	۷	QK	c	YL	. Y	LS	6 <b>G</b>	E	VK	E	KI	L G	K	ID	P	YF	V	KL	A	A	чv	TV	VI	E/	w	DE	Ļ	NF	<b>P A</b>	TP	v	E	Ν
comp179435	9	N N	1 T	1			C	TU	v	Ο K	č	۲L	÷	1.5	G	E	VK	E	ĸ	G	K	ID	P	TF	v	K L			M V M V	TW	V Í V T	=	W	DE	÷	N	A	F P T P		É	N N
Trinity comp76780	ŏ	NN	ίt	i	P	Ď	č	D T	v	δĸ	č	Ϋ́ι	Ý	Ë ŝ	6 G	Ē	v	Ē	Ř I	G	ĥ	ĪD	P	ŶÊ	v	K L	Â	A	чv	τv	v i	Ê	w	DE	ĩ	N	P A	ŤP	Â	È	N
-,	~	1	1	- 1	1	1			1	÷.,		1	1			-	1	-		1		1	1	1	-	1					-	- 1								1	
Sobic.001G173300	6	ĸ	Å	Ċ																																					
SCAGLR1021B05	G	ĸ	A	C																																					
SCEZHR1049B08	-	÷.		,																																					
TCONS_0001314	-			<b>`</b>																																					

Fonte: Autoria própria

TCONS\_000\_\_ R1\_L153\_T3 comp179435 Locus327.10 Figura 18 - Alinhamento de sequências GT43 dos diferentes bancos de cana-de-açúcar com o gene *bona fide* Sobic.009G229300 de *S. bicolor*. SCCCRZ3001G10 (SUCEST), TCONS\_00009175 (Raíz), comp195498 (Folha) e Locus4936.3 (Entrenó) são genes candidatos de cana-de-açúcar da GT43.

			. 1			10						20			. 1			30				1.			40   -			į .		!	50   .			I.			60 
Sobic.009G229300	MN	S	5 R	RN	S	V	IL		G	S V	R	DW	SE	Ę	ND	P	S P	SP	K		÷	s Q	S	v	A M	RO	L	L A	SI	, I	ŚL	D	FF	Ĥ	S	SK	Ľ.
TCONS_00009175	M١	s	5 R	RN	s	v	ΪĹ	RE	G	s v	R	DW	S	F	ND	P	S P	SP	ĸ	ĩĩ	Y	sų	s	v	AM	R	L	LA	s	/1	s L	D	FF	ũ	s	s K	ĩ.
R1_L28897_T1	M			 • N		v	÷.			 s v		 • •		÷.	N D		 5 D	 5 D	K	11	÷	 5 0		v		- · ·			5	, <del>,</del>	 			11			÷.
Locus4936.3	ΜV	s	5 R	RN	s	s v	ĪĹ	RE	G	s v	R	DW	SE	F	ND	P	5 P	S P	ĸ	Ē	Ŷ	s õ	s	v	A M	R	Ē	LA	s I	í i	s L	D	FF	ũ	s	s ĸ	Ē.
Trinity_comp88357	M	s	5 <b>R</b>	RN	S (	G V	IL	RE	G	s v	R	DW	SE	F	ND	P	S P	S P	K		Y :	s Q	S	v	AM	R (	L	LA	s١	/ I	S L	D	FF	LI	. s	s ĸ	L
						70						80						90	)					1	00					1	10						120
Sobic.009G229300	KS	A		A M	Ť S	.   5 0	R H	SR	s	 O E	R	·   sk	s	G	LS	ċ	R	· I V A	v	Ĥ L	i.	FF	ÊN	iv	GI	Ê	G	F M	P I	Ē	sv	D	v Y	ĸ	ĊI.	vs	È
SCCCRZ3001G10	KS	A	C V	AM	TS	Q	RH	SR	s	Q E	R	s ĸ	SK	G	LS	Ċ	RR	VA	V	H L	Ŀ.	EE	FN	11	ĠI	E I	G	FM	PN	1 E	s v	D	YY	ĸ	(1)	vs	E
R1_L28897_T1	KS	A	C V .	<u>АМ</u>	1	Q	<b>к</b> н	5 8	s	QE	R	5 K	51	G	LS	C	R R	V A			5	FF	FN		G I		G	FM	P N	F	5 V	D	V Y	ĸ		v s	Ε.
comp195498	KS	A	C V	AM	TS	Q	RH	SR	S	QE	R	S K	SK	G	LS	C	RR	VA	V	H L	÷	FF	FN	I	GI	E1	G	FM	PI	÷Ē	S V	D	YY	K	(1)	vs	E
Trinity_comp88357	KS	Â	c v	AM	T S	Q	RH	SR	S	Q E	R	SK	S	G	LS	č	R	VA	v	ΗĽ	÷.	FF	FN	iv	GI	Fi	G	FM	P	i F	s v s v	D	νΥ	ĸ	à	v s v s	Ē
			- 1			130 .			. 1			14	D		. 1			15	0.			į .			160			į .		1	170			I.			180
Sobic.009G229300 SCCCR73001G10	N	R	P	탪	EC	V	IE	TE	Ŧ	M G M G	Ţ	ΚV	KE	÷	ET	v	vv	EK	E		÷	ID ID	EF	Q Q	V Q V F	ES	P	P V P V	P /	M	LD	DI		D	FA	ES	S
TCONS_00009175	NE	RI	P	FH	Ē	s v	I D	ŤĒ	Ť	MG	T	κv	KE	ĩ	ĒŤ	v	vv	EK	Ē	V E	ĩ	I D	Ē	۰õ	V E	Ē	P	ΡV	P /	м	LD	D	EA	D	FA	ES	s
R1_L28897_T1 comp195498	N	R	P	ĒĤ	Ē	v	T F	÷.	т	 м с	Ŧ	ĸ v	N	÷.	ĒŤ	v	 	F K	Ē	 V F	÷.	 T D	01	0	VF	Ē	P	 P V	- · ·	м	 I D	D I	F A	<b>D</b>	FA	FS	ŝ
Locus4936.3	NE	RI	P	FH	Ē	v	IE	Ť	Ť	MG	Ť	κv	NE	ĩ.	ĒŤ	v	vv	EK	E	V K	Ē	I D	Ē	ò	V E	E	P	ΡV	Ρ/	м	LD	D	EĂ	D	FA	ES	s
Trinity_comp88357	N	RI	P	FH	E	G V	IE	TE	т	MG	T	κv	N	L	ЕТ	v	vv	EK	E	VE	L	ID	QF	Q	VE	ES	P	PV	P /	М	LD	DI	EA	D	FAI	ES	S
						190						20	D					21	0					:	20					2	230						240
Sobic.009G229300	P A	É.	-   P G	İĖ	É S	D	İV	AK	.   К.К.	ίĹ	I	iv	τi	т	s v	RI	PO	.   0 A	Ý	ÝĹ	N	I. RL	Å	v	LK	Â	, H	.   P	ċ.	w	ĽV	v	ĒW	P	0	s y	e l
SCCCRZ3001G10	P A	L.	P A	IE	ES	D	IV	AK	ĸ	LL	1	IV	I I	Τ	s v	R	PQ	Q A	Y	YL	N	RL	A	v	LK	A \	Q	P	Ŀ.	. w	LV	۷	EW	P	Q	S Y	E
R1_L28897_T1	P A		<b>A</b>					A 8	с к -		1		11	4	5 V	<b>K</b>	P Q	Q A	1		-	K L	-			A	iõ	A P	Ľ	w		v	EW	P	ŏ	SY	Ē
comp195498	P A	5	PA	IE	ES	D	IV	AK	ĸ	ĿĿ	I	IV	I.	Ţ	S V	R	PQ	QA	Y	YL	N	RL	<u>A </u>	v	LK	A \	Q	P	5	. W	LV	V	EW	P	Q	SY	E
Trinity_comp88357	PA	Ľ	PA	İĒ	ES	5 D	IV	AK	(K	11	i	IV	Η	÷	s v s v	R	PQ	Q A Q A	Y	YL	N	RL	Â	v	LK	A	iQ	A P	Ľ	w		v	EW	P	Q	SY	Ē
			. 1			250 .			. 1			26 .	D		.			27 .	0.			1.			280   -			·		1	290   -			I.			300 
Sobic.009G229300 SCCCRZ3001G10	T A	E		RS	s c	250 .   . VI	ч ү ч ү	R H		v c	R	26 .   K N		s	·   V R V R	ĸ	IA	27 ·   V C	° Q	R N		A I A I	Ŷ	IV	80   . K R	н	Ļ	. D G	IN	1н	190   . FA FA	DI	EE	 R S R S	Y	S A S A	300   D
Sobic.009G229300 SCCCRZ3001G10 TCONS_00009175	T A T A T A	E E E		R S R S R S	S S S	250 VI	4 Y 4 Y 4 Y 4 Y	R H R H R H	L L L	v c v c v c	R R R	26   K N K N K N		s s	VR VR VR	K K K		27 V C V C V C	Q Q Q	R N R N R N	I N I N	A I A I A I	Y I Y I Y I	I V V	80   . K R K R K R	нн	L L L		IN IN IN	4 H 4 H 4 H	90 FA FA	DI DI DI	E E E E E E	R S R S R S	Y Y Y	S A S A S A	300 D D
Sobic.009G229300 SCCCRZ3001G10 TCONS_00009175 R1_L28897_T1 comp195498		E E E E		R S R S R S R S R S	S S S S S	250                                     	4 Y 4 Y 4 Y 4 Y	R H R H R H R H			R R R R R	26 K N K N K N K N K N		s s s s s s	·   VR VR VR VR	K K K K K K		27 VC VC VC	9 . QQ QQ 0	R N R N R N R N R N			Y Y Y Y Y	I V V V	180   . KR KR KR KR	H H H H H H H H H H H H H H H H H H H	L L L L				90 FA FA FA FA			R S R S R S R S R S	S Y S Y S Y S Y	S A S A S A S A S A	300   D   D   D   D   D
Sobic.009G229300 SCCCRZ3001G10 TCONS_00009175 R1_L28897_T1 comp195498 Locus4936.3		EEEE		R S R S R S R S R S R S	S S S S S S S S S S S S S S S S S S S	250 VI VI VI VI	4 Y 4 Y 4 Y 4 Y 4 Y				R R R R R	26/ K N K N K N K N K N K N K N		\$ \$ \$ \$ \$ \$	VR VR VR VR				000000				Y Y Y Y Y	1 V V V V V V V	80   . KR KR KR KR						90 FA FA FA FA FA				S Y S Y S Y S Y S Y	S A S A S A S A S A S A	300 D D D D D D
Sobic.009G229300 SCCCRZ3001G10 TCONS_00009175 R1_L28897_T1 comp195498 Locus4936.3 Trinity_comp88357	T A T A T A T A T A T A	E E E E E E E		R S R S R S R S R S R S	S S S S S S S S S S S S S S S S S S S	250                                     	4 Y 4 Y 4 Y 4 Y 4 Y 4 Y	R H R H R H R H R H			R R R R R R	26 KN KN KN KN KN KN		5555555	V R V R V R V R V R V R V R	K K K K K K K K K			00000000	R N R N R N R N R N R N R N			Y Y Y Y Y Y	1 V V V V V V V V	80 KR KR KR KR KR KR						90 FA FA FA FA FA				S Y S Y S Y S Y S Y S Y	S A S A S A S A S A S A S A	300 D D D D D D D
Sobic.0096229300 SCCCR23001610 TCONS_0009175 R1_L28897_T1 comp195498 Locus4936.3 Trinity_comp88357	Т А Т А Т А Т А Т А Т А	E E E E E		R S R S R S R S R S R S R S		250 VI VI VI VI VI	4 Y 4 Y 4 Y 4 Y 4 Y 4 Y				R R R R R R	26/ K N K N K N K N K N K N K N		\$ \$ \$ \$ \$ \$ \$	VR VR VR VR VR VR	K K K K K K K K K K K K K K			°. • • • •	R N R N R N R N R N R N R N			Y Y Y Y Y Y	I V V V V V V V	180 KR KR KR KR KR KR						90 FA FA FA FA FA FA				S Y S Y S Y S Y S Y	S A S A S A S A S A S A S A	300 D D D D D D D 360
Sobic.0096229300 SCCCR23001610 TCONS_00009175 R1_L28897_T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.0096229300		EEEE		RSRRS RRS RRS RRS RRS RRS RRS RRS	S ( S ( S ( S ( S ( S ( S ( S ( S ( S (	250 VI VI VI VI VI VI VI VI VI VI VI VI VI	4 Y 4 Y 4 Y 4 Y 4 Y 4 Y 4 Y 4 Y 5 G	RH RH RH SV			R R R R R	26/ K N K N K N K N K N K N K N K N		S S S S S S S S S S S S S	VR VR VR VR VR VR VR VR	K K K K K K						A I A I A A A A A A A A A A A A A A A A	Y   Y   Y   Y   Y   Y   Y   Y		280   . KR KR KR KR KR KR KR (KR (KR) (40) - GW						190   . FA FA FA FA FA FA FA FA FA FA			RR RR RR RR RR RR RR RR RR RR RR RR RR	Y Y Y Y Y Y Y	S A S A S A S A S A S A S A S A S A S A	300 D D D D D D D D D D D D D D D D D D
Sobic.0096229300 SCCCR23001G10 TCONS_00009175 R1_L28897_T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.0096229300 SCCR2301G10		E		RSS RSS RSS RSS RSS RSS RSS RSS RSS RSS		250 VI VI VI VI VI VI VI VI VI VI VI VI VI	MYYYYYYY MYY MYY MYY MYY	RRRRR SV			R R R R R I I I	260 K N K N K N K N K N K N K N K N		SSSSSSS KKK		K K K K K K						A I A I A A I A A I A A A A A A A A A A			180   KR KR KR KR KR KR KR KR KR KR KR KR						P90 FA FA FA FA FA FA FA FA FA FA FA			RR RR RR RR RR RR RR RR RR RR RR RR RR	5 Y 5 Y 5 Y 5 Y 5 Y 5 Y		300   D D D D D D D D D D D D D D D D D D D
Sobic.009G229300 SCCCR23001G10 TCONS_00009175 R1_L28897 T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10 TCONS_00009175 R1_L28897_T1				RSS RRSS RRSS RRSS RRSS RRSS RRSS RRSS		250 VI VI VI VI VI VI VI VI VI VI	NYYYYY YYYYY FGGFFG	RRRRRR SVVSV	1 L 1 L 1 L 1 L 1 L 1 L 1 L 1 L 1 L 1 L		R R R R R I I I I I	26 K N K N K N K N K N K N K N K N		SSSSSS KKKK		K K K K K K K K K						A I A A A A A A A A A A A A A A A A A A			180 KR KR KR KR KR KR KR KR KR KR KR KR KR						90 FA FA FA FA FA FA FA FA FA KS KS KS			RRRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR	5 Y 5 Y 5 Y 5 Y 5 Y 5 Y 5 Y 5 Y 5 Y		300 D D D D D D D D D D D D D
Sobic.009G229300 SCCCR23001G10 TCONS_00009175 R1_L28897 T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR2301G10 TCONS_00009175 R1_L28897 T1 comp195498 Locus4936.3	T A T A T A T A T A T A T A T A T A T A			RRSSS R R R R RSSS R		250 ·   ·   ·   ·   ·   ·   ·   ·		RRRRRR SVV			R R R R R I I I I I I I	26 K N K N K N K N K N K N K N K N		SSSSSSS KKKKK																	PO FA FA FA FA FA FA FA FA FA FA FA FA FA			RRRRR RRRRR			300 D D D D D D D D D D D D D D D D D D
Sobic.009G229300 SCCCR2301G10 TCONS_0009175 R1_L28897 T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 R1_L28897_T1 COMS_0009175 R1_L28897_T1 Trinity_comp88353	A T A A A A			RRRRRR QQQQQQQQQQ		250 VI VI VI VI VI VI VI VI VI VI	MYYYYY MYYYY FFGGGGG	RRRRR SVVV			R R R R R I I I I I I I I I I I	26/ K N K N K N K N K N K N K N K N K N K N		SSSSSS KKKKKK								A I A A A A A A A A A A A A A A A A A A			KRRKRRKRRKRRWRW						PO FA FA FA FA FA FA FA FA FA FA FA FA FA			RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR			300 D D D D D D D D D D D D D D D D D D
Sobic.009G229300 SCCCR23001G10 TCOMS_0009175 R1_L28897 T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10 SCCCR2300 SCCCR23001G10 SCCCR2300	T A T A T A T A T A T A T A V F V F V F V F			RRRRRRR QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ		250 VI VI VI VI VI VI VI VI VI VI		RRRRR SSVV	·   L   L   L   L   L   L   L   V P V P V P V P V P V P V P		R R R R R I I I I I I I I I I I	26/ KN KN KN KN KN KN KN KN KN KN KN KN KN		SSSSSS SSSSSS	VVVVVVV VVVVVVV YYYYYR							A I I A A A A A A A A A A A A A A A A A	YYYYYY RRRRRR		KRRKRRKRRWRWRW						FAFFAFFA FAFFA FAKKSSKKS KKS			RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR	FP FP FP FP		300 D D D D D D D D D D D D D D D D D D
Sobic.009G229300 SCCCR23001G10 TCONS_00009175 R1_L28897 T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10 TCONS_00009175 SCCCR23001G10 TCONS_00009175 R1_L28897 T1 comp195498 Locus4936.3 Trinity_comp88357	T A A A A A A A A A A A A A A A A A A A			RRRRRRR QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ		250 ·   VI VI VI VI VI VI VI VI VI VI		RRRRR SVVSVV			R R R R R R I I I I I I I I I I I	26/ KN KN KN KN KN KN KN KN KN KN	GI GI GI GI GI GI	SSSSSS KXXXXX	VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV							AIIAAI AAI AAI AAI AAI AAI AAI AAI AAI			180 KRR KKRKKK 140 GW RRW RRW RRW						FAFFAFFA FAFFA FASSO SO KKSSKKS			RRRRR RRRRR	F P P P F P P P F P P P F P P P F P P P F P P P F P P P F P P P F P		300   D D D D D D D D D D D D D D D C C C C C
Sobic.009G229300 SCCCR23001G10 TCONS_00009175 R1_22897_T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10 TCONS_00009175 R1_22897_T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10	TAATAA TAATAA TAA TAA VFF VFF VFF VFF VFF VFF SAA			RRRRRR QQQQQQ NNS		250 VI VI VI VI VI VI VI VI VI VI		RRRRR SVV SVV SVV SVV	VP VP VP VP VP VP VP VP VP VP VP		R R R R R R R I I I I I I I I I I I I	26/ KN KN KN KN KN KN KN KN KN KN		SSSSSS KKKKKK SS									YYYYYY RRRRR SF		280 - KR KKR KKR KKR KKR KKR KKR KKR KKR KKR						190 FAFFAFFA SO SO KKSS KKSS IIIE IE			RRRRRR RRRRR PI		SASSASSA SSASSA SSA SSA SSA SSA SSA SSA	3000   D D D D D D D D D D D D D D D D F F F F
Sobic.009G229300 SCCCR23001G10 TCONS_00009175 R1_22897 T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10 TCONS_0009G229300 SCCCR2301G10 Sobic.009G229300 SCCCR2301G10 TCONS_0000175	TAAAAA TTAAAAA TTAAAAA VYFY VYFY VYFY SAAA			RRRRRRR QQQQQQ NNNN		250 VI VI VI VI VI VI VI VI VI VI		RRRRR SSVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV			R R R R R R R R R R R R R R R R R R R			SSSSSS XXXXXX SSSS	RRRRRR RRRRRRR IIII							AAIII .NYYY	YYYYYY RRRRR SSS								190 FAAFFAFFAFFAFFASS		EEEEEEE	RRRRRR RR PRRRR PDDD	F P F P F P F P F P F P	SAASSASSASSASSASSASSASSASSASSASSASSASSA	3000   D D D D D D D D D D D D D D D D D D D
Sobic.009G229300 SCCCR23001G10 TCONS_0009175 R1_L28897_T1 comp195498 Locus4936.3 TrinRy_comp88357 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 R1_L28897_T1 comp195498	TAAAAAA TTATTAA VVFFFF VVFFF SAAAAA SSAAA			RRRRRR QQQQQQQ NNNNN	SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	250 250 2 VI 3 VI 4 VI		RRRRRR SSSSSS PPPPP			R R R R R R R R R R R R R R R R R R R			22222222222222222222222222222222222222	IRRRRRR IRRRRRRR   IRRRRRRR IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII							-AAAAAA -GGGGGGGG -QQQQQ	YYYYYY RRRRRR SSSS		280 . KRKKKK 440 . KRRKKK 440 . KRW 440 . KRW 440 . KRW 440 . KRW 440 . KRW 440 . KRW 440 . KRW 440 . KRW 440 . KRW 440 . KRW 440 . KRW 440 . KRW 440 . KRW 440 . KRW 440 . KRW 440 . K						190 FAAFFA FAAFFA 150 KKSS KKSS 10 IEEIEI			RRRRR RR RRRRR DDDDD	FP FP FP FP FP FP FP FP FP FP FP FP FP F	SAASSAASSAASSAASSAASSAASSAASSAASSAASSA	300   D D D D D D D D D D D D D D D C D C
Sobic.009G229300 SCCCR23001G10 TCONS_00009175 R1_L28897 T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10 TCONS_00009175 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10	TTTTTT TTTTT VVFFFFF SSSSSS			RRRRRR QQQQQQ NNNNNN	SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	250 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 3 VI 4		RRRRRR SSSSSS PPPPPP			R R R R R R R R R R R R R R R R R R R			~~~~~	IRRRRRR IRRRRRR   IRRRRRR IRRRRRR   IVVVVVV IRRRRRR   IVVVVVV IVVVVVV							-AAAAAA -GGGGGGGG -QQQQQQQ	YYYYYYY RRRRRR SSSSSS								190 FAAFFA FFAFFA SS KKSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS		EEEEEEE . TTTTTTT . LLPPPPP	RRRRRR RR RR DDDDDD		SSSSSS IIIIII NNNNNNNNNNNNNNNNNNNNNNNNN	300   D D D D D D D D D D D D D D D D D D D
Sobic.009G229300 SCCCR23001G10 TCONS_0009175 R1_L28897 T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G2930 SCCCR2301G10 TCONS_0009175 Sobic.009G2930 SCCCR2301G10 TCONS_0009175 Sobic.009G2930 SCCCR2301G10 TCONS_0009175 Sobic.009G2930 SCCCR2301G10 TCONS_0009175 Sobic.009G2930 SCCCR2301G10 TCONS_0009175 Sobic.009G2930 SCCCR2301G10 TCONS_0009175 SCCCR2301G10 TCONS_0009175 SCCCR2301G10 TCONS_0009175 SCCCR2301G10 SCCCR2301G10 SCCCR2301G10 SCCCR2301G10 SCCCR230 SCCCR2301G10 SCCCR2300 SCCCR230	TTAAAAAAAA TTTTTTT VVFFFF SAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA			RRRRRR QQQQQQ NNNNNN	SSSSSS IIIIIII TTTTTT	250 VI VI VI VI VI VI VI VI VI VI		RRRRR SSSSSS PPPPP	IL IL IL IL IL IL IL IL IL IL IL IL IL I		R R R R R R R R R R R R R R R R R R R	26-INNNNNN 32-IVVVV 38-IPPPPPP		~~~~~	RRRRRR RRRRRR IIIIIII		IIIIII VVVVVV ISSSSS	27 - CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC				IAAAAAA IGGGGGGG IQQQQQQ	YYYYYY RRRRRR SSSSSS		180 RRKKKRR 140 RRWW				IIIIIII AAAAAA EEEEEEE		190 FFAFFF 10 SON KKSSSS 10 FFAFFF 10 SON KKSSSSS 10 FEEFFF		EEEEEEE . TTTTTTTT . LLLLLLLLL	RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR		· SAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	300 D D D D D D D D D D D D D D D D D D
Sobic.009G229300 SCCCR23001G10 TCONS_00009175 ST_1_22897_T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10 TCONS_00009175 Sobic.009G229300 SCCCR2301G10 TCONS_00009175 Sobic.009G229300 SCCCR2301G10 TCONS_00009175 Sobic.009G229300 TCONS_00009175 Trinity_comp88357	TTAAAAAAA TTTTTTT VVFFFFFAAAAAAAAAAAAAAA			RRRRRR QQQQQQ NNNNNN	SSSSSS IIIIII TTTTTT	250 VI VI VI VI VI VI VI VI VI VI		RRRRR SSSSSS PPPPP	IL IL IL IL IL IL IL IL IL IL IL IL IL I		R R R R R R R R R R R R R R R R R R R			22222222222222222222222222222222222222	-VVVVVVV -YYYYYYY -VVVVV-		IIIIII VULLUUVUU	27- VVCCCCC 33- EEGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG				-AAAAAAA -GGGGGGGG -QQQQQQQQ	YYYYYY RRRRR SSSSSS		180 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR				IIIIIII AAAAAA FEEEEEE		90 FFAAFFF 50 SSIKKKKK 10 IIIIIIIIIIIIIIIIIIIIIIIIIIIIII		EEEEEEE	RRRRR RR RRRRR DDDDDDDDDDD		SAAAA GGGGGGGG .RRRRRRNNNNNNNNNNNNNNNNNNNNNNN	300   D D D D D D D D D D D D D D D D C C C C
Sobic.009G229300 SCCCR23001G10 TCONS_0009175 R1_L28897_T1 comp195498 Locus4936.3 TriniRy_comp88357 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 R1_L28897_T1 comp195498 Locus4936.3 TriniRy_comp88357 Sobic.009G229300 SCCCR23001G10 SCCCR23001G10 SCCCR2300 SCCCR2301G10 SCCCR230	TTTTTTT TTTTTTT VVFFF SSSSSS M			RRRRRRR QQQQQQQ NNNNNN N	SSSSSS IIIIII TTTTTT	250 9 VI 9 VI 9 VI 9 VI 9 VI 9 VI 9 VI 9 VI		RRRRRR SSSSSS PPPPP N	IL IL IL IL IL IL IL IL IL IL IL IL IL I		R R R R R R R R R R R R R R R R R R R	26-INNNNNNN 21-IVVVV 38-PPPPPP 44-II			RRRRRRR RRRRRRR IIIIIII - E			27 - CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC				INTERPOLATION INTERPOLATIONI INTERPOLATION I	YYYYYY RRRRRR SSSSSSS		180 RRARRAR 141 GRANNWW IEEEEEEEE				IIIIIII AAAAAA FEEEEEE		90 FFAFFF 50 SSIKKSS 11 I I I I I I I I I I I I I I I I I		EEEEEEE TTTTTTTTTLLLLLL	RRRRRR RRRRRR DDDDDD		SAAAA GGGGGG RRRRRR	300 D D D D D D D D D D D D D D D D D D
Sobic.009G229300 SCCCR23001G10 TCONS_0009175 R1_L28897 T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 Sobic.009G229300 SCCCR23001G10 TCONS_0009175	TTTTTTT TTTTTTT VVVVVV SSSSSS MMM			RRRRRRR QQQQQQQ NNNNNN NNN	SSSSSS IIIIII TTTTTT EEE	250 VI VI VI VI VI VI VI VI VI VI VI VI VI		RRRRRR SSSSSSS PPPPPP NNN	IL IL IL IL IL IL IL IL IL IL IL IL IL I		R R R R R R R R R R R R R R R R R R R	26-INNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN			-VVVVVVV - PRRRRRR - IIIIIII			27 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ	RRRRR ILICOCC .GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG		I I I I I I I I I I I I I I I I I I I	YYYYYY RRRRRR SSSSSSS		180 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR				IIIIIII AAAAAA EEEEEEE		90 FFAAFFF S0 KKKSSS IIIIIIIIIIIIIIIIIIIIIIIIIIIII		· EEEEEEE · TTTTTTTT · LLLLLL	RRRRRR RRRRRR DDDDDD		SAAAA GGGGGGGG RRRRRR	300 D D D D D D D D D D D D D D D D D D
Sobic.009G229300 SCCCR23001G10 TCONS_00009175 R1_L28897_T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10 TCONS_00009175 Sobic.009G229300 SCCCR23001G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175 Sobic.009G229300 SCCCR2301G10 TCONS_0009175	TTTTTTT VVVVVV SSSSSS MMMM			RRRRRRR QQQQQQQ NNNNNN NNNN	SSSSSS IIIIIII TTTTTT EEEE	250 250 250 250 250 250 250 250		RRRRRR SSSSSS PPPPPP NNNN	ILLIL ILLIL ILLILILILILILILILILILILILIL		R R R R R R R R R R R R R R R R R R R	26-INNNNNN 22-IVVVVV 38-PPPPPP 44-IIII		SSSSSS KKKKKKK SSSSSS				VVVVVVV 31 GGGGGGGGGGGGGGGGGGGGGGGGGGGGG	0 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ	RRRRR I C C C C C G G G G G G G G G G G G G G		IIIIIII INYYNNNN EEEEEEE	YYYYYY RRRRRR SSSSSSS		180 · RRRRRRR AD · GRANNWARR AD · EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE				IIIIIII AAAAAA FEEEEEE		90 FFAAAAFFF 90 SSIKKKKK 91 IIIIIIIIIIIIIIIIIIIIIIIIIIIIII		· EEEEEEE · TTTTTTTT · LLLLLL	RRRRRR RRRRRR DDDDDD		SSAAAA GGGGGGG SSSSSSSSSSSSSSSSSSSSSSSS	300 D D D D D D D D D D D D D D D D D D
Sobic.009G229300 SCCCR23001G10 TCOMS_0009175 R1_L28897 T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10 TCOMS_0094798 Locus4936.3 Trinity_comp88357 Sobic.009G229300 SCCCR23001G10 TCOMS_0009175 R1_L28897 T1 comp195498 Locus4936.3 Trinity_comp88357 Sobic.009G229300	TTTTTTT VVVVVV SSSSSS MMMMM		ILLLLL IMMMMMM IFFFFFFFFFFFFFFFF	RRRRRRR QQQQQQQ NNNNNN NNNNNN	SSSSSSS IIIIIII TTTTTT EEEEEE	250 250 250 250 250 250 250 250		RRRRRR SSSSSS PPPPPP NNNNN	ILLILILILILILILILILILILILILILILILILILI	VVVVVV VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV	R R R R R R R R R R R R R R R R R R R	26-INNNNNNN 22-IVVVVV 38-IPPPPPP 44-IIIIII	GITTITITITITITITITITITITITITITITITITITI					27-CCCCCCC 33-GGGGGGGGGGGGGGGGGGGGGGGGGGGG	QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ	RRRRR I I I I I I I I I I I I I I I I I		IIIIIII INYYNNNN EEEEEEE	YYYYYY RRRRRR SSSSSSS		180 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR				IIIIIII AAAAAA EEEEEEE		90 FFAAAFFF 850 SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS		· EEEEEEE · TTTTTTTT · LLLLLL	RRRRRR RRRRRR DDDDDD		SSSSSS - IIIIII - NNNNNNNNNNNNNNNNNNNNNN	300 D D D D D D D D D D D D D D D D D D

Fonte: Autoria própria

Figura 19 - Alinhamento de sequências GT61 dos diferentes bancos de cana-de-açúcar com o gene *bona fide* Sobic.010G144400 de *S. bicolor*. SCJFRZ2009A12 (SUCEST), TCONS\_00015532 (Raíz) e comp206869 (Folha) são sequências completas.

Sobic.010G144400 SCJFRZ2009A12. TCONS_00015532 R1_L846_T2 comp206869 Locus5234.3			S S S S S	G G G G G	EI EI EI EI	R I R I R I R I			10 V V V		. L L . L L	R ( R ( R ( R ( R (		S S S S S	RF RF RF		R R R R R R		VVVV			FIFIFI	30 FL FL FL FL	V V V V V	S L S L S L S L S L		F V F V F V F V F V			K   K   K   K   K   K			FFFFFF	S I S I S I S I		G G G S S		P P P P		000000	A P A P A P A P T P T P	60     T   T   T   T   T   T
Sobic.010G144400 SCJFR22009A12. TCONS_00015532 R1_L846_T2 comp206869 Locus5234.3				IIIIII	-000000				70 SI SI SI SI		- A A A A A A A			KRRRR		80   T T T T T			P P P P	GGGGG		RI RI RI RI RI		D D D D D		A A A A A		R R R R R R R		G G G G G G		G I G I G I G I G I	E E E E E	EEEEEE		v v v v v		E I E I E I E I E I E I		P P P P		120 G G G G G
Sobic.010G144400 SCJFRZ2009A12. TCONS_00015532 R1_L846_T2 comp206869 Locus5234.3			EEEEE						130 G G G G G G		RRRRR			E G G G G		14             					G G G G G	I KI KI EI KI	50   - R R R K R K R K R K	EEEEE	E R E R E R E R E R	GGGG		A T T T		H H H H H						Y Y Y Y Y		HI HI HI HI	DT DT DT DT DT	E E E E E		180 E E E E E
Sobic.010G144400 SCJFRZ2009A12. TCONS_00015532 R1_L846_T2 comp206869 Locus5234.3	ר 0 יי יי יי		KKKKK	KRRRR		D1 D1 D1 D1 D1	s s s s	YYYYYY	190                                     		GGGGG	S S S S	(P (P (P (P			20                                     	S   S   S   S   S   S	N N N N	RRRR	 A A A A		2 EP EP EP EP	10 M R M R M R M R	G		R R R R R		P P P P	22 NA NA NA			M M M M	E	P P P P P	230 5 H 5 H 5 H 5 H 5 H					W		240 K K K K K K
Sobic.010G144400 SCJFRZ2009A12. TCONS_00015532 R1_L846_T2 comp206869 Locus5234.3	0 F	P Y P Y P Y P Y P Y	P P P P P	RRRRR	- K K K K K			FFFFFF	250 C C C C C C		·HHHHH		EEEEE			26 K K K K	s s s s s s			 A A A A A	P P P P		70 KY KY KY KY	H H H H H		P P P P		IIIII	28 FS FS FS	L1 L1 L1 L1 L1	G G G G G G G G	Y 1 Y 1 Y 1 Y 1 Y 1 Y 1	N N N N	L  L  L	290 FH FH FH		F T F T F T F T				P L P L P L P L P L	300 F F F F
Sobic.010G144400 SCIFR22009A12. TCONS_00015532 R1_L846_T2 comp206869 Locus5234.3	ר 0 ר ר ר ר	T	A A A A A A A	555555555555555555555555555555555555555		FIFI		EEEEEE						M M M M		32/ W W W				 Y Y Y Y Y	H1 H1 H1	3 L( L( L( L( L(			5 K 5 K 5 K 5 K	Y Y Y Y Y Y	P V P V P V P V P V	I I I I I I I	34 DF DF DF DF	s s s s s			HO HO HO HO			H H H H			. G L G L G L G L G L	HHHH	A Y A Y A Y A Y A Y A Y	360 M M M M
Sobic.010G144400 SCJFR22009A12. TCONS_00015532 R1_L846_T2 comp206869 Locus5234.3	0	FFFFF	T T T T T T T	IIIIII					370 P P P P		YYYYY	SNSNSN	1V 1V 1V 1V		FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	38   R   R   R   R	FI				YSYS	3 G G G G G G	90 RD RD RD RD RD RD					EEEEE	40 Y P Y P Y P Y P Y P Y P			P P P P P	·LLLL		410 K K K K K K K K K K K K	R R R R R		T T T T T	RM RM RM RM RM	FFFFFF		420     L   L   L
Sobic.010G144400 SCJFR22009A12. TCONS_00015532 R1_L846_T2 comp206869 Locus5234.3			I I I I I I I	V I I I I I I	 A  A  A  A  A	M/ M/ M/ M/		EEEEE	430   L L L L	G F G F G F G F G F	EEEEEE					44                           				I I I I I I I	s		50 RL SL RL RL RL	V V V V V		- v v v v v v v v v v v v v v v v v v v		M	40 40 40 40 40 40 40 40	v v v v v v	HG HG HG					FI FI FI FI FI				TTTTT		480 - Q Q Q Q Q Q
Sobic.010G144400 SCIFR22009A12. TCONS_00015532 R1_L846_T2 comp206869 Locus5234.3	0 1 1 1 1 1 1	VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV	P P P P P	wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww		G I G I G I G I G I	E	W W W W			TTTTT		GGGGG	N N N N	P # P # P # P # P #		L! L! L! L!			- H H H H H H	Y	5 Y 9 Y 9 Y 9 Y 9 Y 9		6 6 6 6 6		EEEEE	S S S S S S S S S S S S			000000	Y P Y P Y P Y P	R I R I R I R I R I	E E E E E					A   A   A   A   A   A   A   A   A   A	FH FH FH FH FH	K K K K R		540 F F F F
Sobic.010G144400 SCJFR22009A12. TCONS_00015532 R1_L846_T2 comp206869 Locus5234.3	0		I I I I S	RRRRG					550 K K K K K S		VVVV	K I K I K I	DDDDDDDDD	CCCCC			R I R I R I R I K I	P 1 P 1 P 1 P 1				DI DI DI DI	70   . N L N L N L	N N N N	PPPP																	

Fonte: Autoria própria

### 5 CONCLUSÃO

Os diversos estudos realizados para identificação e caracterização gênica em Arabidopsis thaliana, Brachypodium distachyon e Oryza sativa possibilitou a seleção e anotação dos genes envolvidos na biossíntese de hemicelulose, tornando-se o ponto de partida para a busca de genes ortólogos em Z. mays e S. bicolor e consequentemente em cana-de-açúcar. Assim, a análise de ortologia das árvores filogenéticas entre as espécies foi essencial para identificação de genes bona fides em espécies com pouco ou nenhum estudo de caracterização gênica realizado. No caso das bibliotecas de cana-de-açúcar foram identificadas um total de mais de 1400 sequências com domínios característicos das famílias gênicas possuidoras de genes envolvidos na biossíntese de hemicelulose. Essas sequencias de cana-de-açúcar pertencem a sete bibliotecas analisadas dentre as quais está a do SUCEST e outras bibliotecas de RNA-Seq de diferentes órgãos da planta, como raíz, folha e entrenó. As sequências com domínio, das sete bibliotecas de cana-de-açúcar, foram utilizadas para construção de árvores filogenéticas juntamente com as sequências de S. bicolor e Z. mays, e a partir de ortologia com os genes bona fides dessas espécies foram identificadas 27 sequências completas de cana-de-açúcar e candidatas na biossíntese de hemicelulose e possivelmente envolvidos na recalcitrância da parede celular. Essas 27 sequências com anotação completa demonstram um grande passo na construção de oligonucleotídeos (primers) para análise de expressão e manipulação gênica da cana-deaçúcar. Assim, o presente trabalho contribuiu, não apenas para a seleção desses genes candidatos para futura caracterização funcional, como também no desenvolvimento de banco de genes envolvidos na biossíntese de hemicelulose para Arabidopsis thaliana, Brachypodium distachyon, Oryza sativa, Zea mays e Sorghum bicolor.

## REFERÊNCIAS

ALTSCHUL, S.F., et al. Basic local alignment search tool. **Journal of Molecular Biology**, v.215, p.403-410, 1990.

AMORIM, H., et al. Scientific challenges of bioethanol production in Brazil, Appl Microbiol Biotechnol, v.215, p.1267–1275, 2011.

ANDERS, N., et al. Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses, **PNAS**, v.109, p.989–993, 2012.

ANDERSON, C.T., et al. Real-Time Imaging of Cellulose Reorientation during Cell Wall Expansion in Arabidopsis Roots, **Plant Physiol**, v.152, p.787–796, 2010.

ARNER, R.J., et al. Myo-inositol oxygenase p. molecular cloning and expression of a unique enzyme that oxidizes myo-inositol and D-chiro-inositol, **Biochem J**, v.360, p.313-320, 2001;

BAR-PELED, M.; O'NEILL, M.A. Plant nucleotide sugar formation, interconversion, and salvage by sugar recycling, **Annu. Rev. Plant Biol**, v.62, p.127–155, 2011.

BAR-PELED, M.; GRIFFITH, C.L.; DOERING, T.L. Functional cloning and characterization of a UDP-glucuronic acid decarboxylase the pathogenic fungus *Cryptococcus neoformans* elucidates UDP- xylose synthesis, **Proc Natl Acad Sci USA**, v.98, p.12003-12008, 2001.

BASNAYAKE, S.W.V., et al. Field performance of transgenic sugarcane expressing isomaltulose synthase, **Plant Biotechnol J**, v.10, p.217–225, 2012.

BENOVA-KAKOSOVA, A., et al. Galactoglucomannans increase cell population density and alter the protoxylem/metaxylem tracheary element ratio in xylogenic cultures of Zinnia. **Plant Physiol,** v.142 p.696–709, 2006.

BEVAN, M.; UAUY, C. Genomics reveals new landscapes for crop improvement. **Genome Biology**, v.14 p.206, 2013.

BOLGER, M. E., et al. Plant genome sequencing-applications for crop improvement. **Curr. Opin. Biotechnol,** v.26 p.31–37, 2014.

BONIN, C., et al. The *MUR1* gene of *Arabidopsis thaliana* encodes an isoform of GDP-Dmannose-4,6-dehydratase, catalysing the first step in the de novo synthesis of GDP-Lfucose. **Proc. Natl. Acad. Sci. USA**, v.94, p.2085–2090, 1997. BONIN, C.P.; REITER, W.A bifunctional epimerase-reductase acts downstream of the *MUR1* gene product and completes the de novo synthesis of GDP- L -fucose in Arabidopsis. **The Plant Journal**, v.21, p.445–454, 2000.

BOUDET, A.M., et al. Lignins and lignocellulosics p. a better control of synthesis for new and improved uses. **Trends Plant Sci,** v.8, p.576-581, 2003.

BOUTON, S., et al. *QUASIMODO1* encodes a putative membrane-bound glycosyltransferase required for normal pectin synthesis and cell adhesion in Arabidopsis. **The Plant Cell,** v.14, p.2577–2590, 2002.

BRENNAN, C.S.; CLEARY, L.J. The potential use of cereal (1->3, 1->4)-b-D-glucans as functional food ingredients. **Journal of Cereal Science**, v.42, p.1–13, 2005.

BROWN, D.M., et al. Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. **The Plant Journal**, v.52, p.1154–1168, 2007.

BROWN, D.M., et al. Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in Arabidopsis. **Plant J**, v.57, p.732–746, 2009.

BUCHANAN, B.B.; GRUISSEM, W.; JONES, R.L. (2000). **Biochemistry & molecular biology of plants**. Rockville:American society of plant physiology, 2000. p.1408.

BUCKERIDGE, M.S. Seed cell wall storage polysaccharides p. models to understand cell wall biosynthesis and degradation. **Plant Physiol**, v.154, p.1017–1023, 2010.

BURTON, R.A.; FINCHER, G.B. Current challenges in cell wall biology in the cereals and grasses. **Front Plant Sci**, v.3, p.1-6, 2012.

BURTON, R.A.; GIDLEY, M.J.; FINCHER, G.B. Heterogeneity in the chemistry, structure and function of plant cell walls. **Nat Chem Biol**, v.6, p.724–732, 2010.

BURTON, R.A., et al. Cellulose Synthase–Like CslF Genes Mediate the Synthesis of Cell Wall (1,3;1,4)-beta-D-Glucans. **SCIENCE Reports,** v.311, p.2–5, 2006.

CAFFALL, K.H., et al. *Arabidopsis thaliana* T-DNA mutants implicate *GAUT* genes in the biosynthesis of pectin and xylan in cell walls and seed testa. **Molecular Plant,** v.2 p.1000–1014, 2009.

CAO, P.J., et al. Construction of a rice glycosyltransferase phylogenomic database and identification of rice-diverged glycosyltransferases. **Mol Plant,** v.1, p.858–77, 2008.

CARDOSO-SILVA, C.B., et al. De Novo Assembly and Transcriptome Analysis of Contrasting Sugarcane Varieties. **Plos One,** v.9, p.e88462, 2014.

CARPITA, N.C. Structure and biogenesis of the cell walls of grasses. Annu Rev Plant Physiol Plant Mol Biol, v.47, p.445–476, 1996.

CARPITA, N.C.; GIBEAUT, D.M. Structural models of primary cell walls in flowering plants consistency of molecular structure with the phisical properties of the walls during growth. **The Plant Journal**, v.3, p.1-30, 1993.

CARPITA, N.C.; MCCANN, M.C. Maize and sorghum p. genetic resources for bioenergy grasses. **Trends Plant Sci**, v.13, p.415-420, 2008.

CARPITA, N.; TIERNEY, M.; CAMPBELL, M. Molecular biology of the plant cell wall p. searching for the genes that define structure, architecture and dynamics. **Plant Mol. Biol,** v.47, p.1–5, 2001.

CASU, R.E., et al. Identification of transcripts associated with cell wall metabolism and development in the stem of sugarcane by Affymetrix GeneChip Sugarcane Genome Array expression. profiling. **Funct Integr Genomics,** v.7, p.153–167, 2007.

CAVALIER, D.M., et al. Disrupting two *Arabidopsis thaliana* xylosyltransferase genes results in plants deficient in xyloglucan, a major primary cell wall component. **Plant Cell**, v.20, p.1519–1537, 2008.

CHEN, X.; ALONSO, A.P.; SHACHAR-HILL, Y. Dynamic metabolic flux analysis of plant cell wall synthesis. **Metabolic Engineering**, v.18, p.78–85, 2013.

CHINIQUY, D., et al. XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. **Proc Natl Acad Sci USA**, v.109, p.17117–17122, 2012.

CHINIQUY, D., et al. Three novel rice genes closely related to the Arabidopsis *IRX9*, *IRX9L*, and *IRX14* genes and their roles in xylan biosynthesis. **Frontiers in Plant Science**, v.4, p.1-13, 2013.

CHRISTENSEN, U., et al. Characterization of the primary cell walls of seedlings of *Brachypodium distachyon* – A potential model plant for temperate grasses. **Phytochemistry**, v.71, p.1–8, 2009.

COCURON, J.C., et al. A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. **Proc Natl Acad Sci USA**, v.104, p.8550–8555, 2007. CORPET, F. Multiple sequence alignment with hierarchical clustering. Nucl. Acids Res, v.16, p.10881-10890, 1988.

COSTA, T.H.F., et al. The enzymatic recalcitrance of internodes of sugar cane hybrids with contrasting lignin contents. **Industrial Crops & Products,** v.51, p.202–211, 2013.

COSTET, L., et al. Haplotype structure around Bru1 reveals a narrow genetic basis for brown rust resistance in modern sugarcane cultivars. **Theoretical and Applied Genetics**, v.125, p.825–836, 2012.

DAVIS, J., et al. Arabidopsis mannan synthase CSLA9 and glucan synthase CSLC4 have opposite orientations in the Golgi membrane. **Plant J,** v.64 p.1028–1037, 2010.

DELMER, D.P.; STONE, B.A. (1988) **Biosynthesis of plant cell walls**. In: STUMPF P.K.; CONN E.E. The Biochemistry of plants. San Diego:Academic Press, 1988. p.373–420.

D'HONT, A.; GLASZMANN, J.C. Sugarcane genome analysis with molecular markers, a first decade of research. **Proc. Int. Soc. Sugarcane. Technol,** v.24 p.556-559, 2001.

DHUGGA, K.S. Building the wall p. genes and enzyme complexes for polysaccharide synthases. **Curr. Opin. Plant Biol,** v.4, p.488–493, 2001.

DHUGGA, K.S., et al. Guar seed beta-mannan synthase is a member of the cellulose synthase super gene family. **Science**, v.303, p.363–366, 2004.

DHUGGA, K.S.; TIWARI, S.C.; RAY, P.M. A reversibly glycosylated polypeptide (RGP1) possibly involved in plant cell wall synthesis p. purification, gene cloning, and trans-Golgi localization. **Proc Natl Acad Sci USA**, v.94, p.7679–7684, 1997.

DOBLIN M.S.D.; PETTOLINO F.; BACIC A. Plant cell walls p. the skeleton of the plant world. **Functional Plant Biology**, v.37, 357–381, 2010.

DOBLIN, M.S.D., et al. A barley cellulose synthase-like CSLH gene mediates (1,3;1,4)beta-d-glucan synthesis in transgenic Arabidopsis. **Proc Natl Acad Sci USA**, v.106, p.5996–6001, 2009.

DOERING, A.; LATHE, R.; PERSSON, S. An Update on Xylan Synthesis. Molecular Plant, v.5, p.769–771, 2012.

DRAPER, J., et al. *Brachypodium distachyon* p. A New Model System for Functional Genomics in Grasses. **Plant Physiology**, v.127, p.1539–1555, 2001.

EBRINGEROVÁ, A. Hemicellulose. Adv Polym Sci, v.186, p.1–67, 2005.

EDGAR, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput, **Nucleic Acids Research**, v.32, p.92-97, 2004.

EDWARDS, D.; BATLEY, J. Plant genome sequencing p. applications for crop improvement. **Plant Biotechnol J,** v.8, p.2–9, 2010.

EMONS, A.M.C.; MULDER, B.M. How the deposition of cellulose microfibrils builds, *5*(1), 35–40. **Trends in Plant Science**, v.5, p.35-40, 2000.

FAIK, A. Xylan Biosynthesis p. News from the Grass. **Plant Physiol**, v.153, p.396–402, 2010.

FENGEL, D.; WEGENER, G. **Wood-Chemistry, Ultrastructure, Reactions**. Berlim:Walter de Gruyter, 1989. p.613.

FEINGOLD, D.S. Aldo (and keto) hexoses and uronic acids. In: LOEWUS FA; TANNER W. Encyclopaedia of Plant Physiology, Berlim:Springer Verlag, 1982. p.3-76.

FERREIRA, S.S., et al. Biofuel and energy crops p. high-yield Saccharinae take center stage in the post-genomics era. **Genome Biology**, v.14, p.210, 2013.

FINCHER, G.B. Revolutionary Times in Our Understanding of Cell Wall Biosynthesis and Remodeling in the Grasses. **Plant Physiology**, v.149, p.27–37, 2009.

FINN, R.D., et al. The Pfam protein families database. Nucleic Acids Research, v.42, p.D222-D230, 2014.

FINN, R.D.; CLEMENTS, J.; EDDY, S.R. HMMER web server p. interactive sequence similarity searching. Nucleic Acids Research, v.39, p.W29-W37, 2011.

FITCH, W.M. Homology - a personal view on some of the problems. **Homology Reviews**, v.16, p.227–231, 2000.

FRY, S.C., et al. Mixed-linkage beta-glucan p. xyloglucan endotransglucosylase, a novel wall-remodelling enzyme from *Equisetum* (horsetails) and charophytic algae. **Plant J**, v.55, p.240–252, 2008.

GARCIA, A.A.F., et al. SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. **Scientific Reports,** v.3, p.1–10, 2013.

GASTEIGER, E., et al. ExPASy: the proteomics server for in-depth protein knowledge and

analysis. Nucleic Acids Res, v.31, p.3784-3788, 2003.

GIBEAUT, D.M.; CARPITA, N.C. Tracing Cell Wall Biogenesis in Intact Cells and Plants Selective Turnover and Alteration of Soluble and Cell Wall Polysaccharides in Grasses. **Plant Physiology**, v.97, p.551–561, 1991.

GIBEAUT, D.M.,; CARPITA, N.C. Biosynthesis of plant cell wall polysaccharides. **The FASEB Journal**, v.8, p.904–915, 1994.

GIBEAUT, D.M., et al. Changes in cell wall polysaccharides in developing barley ( Hordeum vulgare) coleoptiles. **Planta**, v.221, p.729–738, 2005.

GILLE, S.; PAULY, M. O-acetylation of plant cell wall polysaccharides. **Front Plant Sci**, v.31, p.3-12, 2012.

GOH, C.S., et al. Bio-ethanol from lignocelluloses p. status, perspectives and challenges in Malaysia. **Bioresour Technol,** v.101, p.4834–4841, 2010.

GOLDEMBERG, J. The Brazilian biofuels industry. **Biotechnol Biofuel**, v.1, p.6, 2008.

GONDOLF, V.M., et al. A gene stacking approach leads to engineered plants with highly increased galactan levels in Arabidopsis. **BMC Plant Biology**, v.14, p.1–11, 2014.

GOODSTEIN, D.M., et al. Phytozome p. a comparative platform for green plant genomics, **Nucleic Acids Res,** v.40, p.D1178-D1186, 2012.

GOUBET, F., et al. Cell wall glucomannan in Arabidopsis is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. **Plant Journal**, v.60, p.527–538, 2009.

GOUY, M.; GUINDON, S.; GASCUEL, O. SeaView version 4 p. a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. **Molecular Biology and Evolution,** v.27, p.221-224, 2010.

GRATIVOL, C., et al. Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus Saccharum. **The Plant Journal**, v.79, p.162–172, 2014.

GRIVET, L.; ARRUDA, P. Sugarcane genomics p. depicting the complex genome of an important tropical crop. **Curr. Opin. Plant Biol,** v.5, p.122–127, 2002.

HA, M.A.; APPERLEY, D.C.; JARVIS, M.C. Molecular rigidity in dry and hydrated

onion cell walls. Plant Physiol, v.115, p.593-598, 1997.

HAAS, B.J., et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, **Nature Protocols**, v.8, p.1494–1512, 2013.

HANTUS, S., et al. Structural characterization of novel l-galactose-containing oligosaccharide subunits of jojoba seed xyloglucans. **Carbohydr Res**, v.304, p.11–20, 1997.

HARHOLT, J., et al. ARABINAN DEFICIENT1 is a putative arabinosyltransferase involved in biosynthesis of pectic arabinan in Arabidopsis. **Plant Physiol**, v.140, p.49–58, 2006.

HARPER, A.D.; BAR-PELED, M. Biosynthesis of UDP-Xylose. Cloning and Characterization of a Novel Arabidopsis Gene Family, UXS, Encoding Soluble and Putative Membrane-Bound UDP-Glucuronic Acid Decarboxylase Isoforms. **Plant Physiology**, v.130, p.2188–2198, 2002.

HARRISON, M.D., et al. Accumulation of recombinant cellobiohydrolase and endoglucananse in the leaves of mature transgenic sugar cane. **Plant Biotechnol J**, v.9, p.884–896, 2011.

HAYASHI, T. Xyloglucans in the primary cell wall. Annu. Rev. Plant Physiol, v.40, p.139-168, 1989.

HIMMEL, M.E., et al. Biomass Recalcitrance. Science, v.315, p.804-808, 2007.

HOFFMAN, M., et al. Structural analysis of xyloglucans in the primary cell walls of plants in the subclass Asteridae. **Carbohydr Res**, v.340, p.1826–1840, 2005.

HORNBLAD, E., et al. Partial functional conservation of *IRX10* homologs in *Physcomitrella patens* and *Arabidopsis thaliana* indicates an evolutionary step contributing to vascular formation in land plants. **BMC Plant Biol**, v.13, p.3, 2013.

HRMOVA, M.; FINCHER, G.B. Structure-function relationships of beta-D-glucan endoand exohydrolases from higher plants. **Plant Mol Biol**, v.47, p.73–91, 2001.

HSIEH, Y.S.Y.; HARRIS, P.J. Xyloglucans of Monocotyledons Have Diverse Structures. **Molecular Plant,** v.2, p.943–965, 2009.

INTERNATIONAL RICE GENOME SEQUENCING PROJECT. The map-based

sequence of the rice genome. Nature, v.436, p.793-800, 2005.

JENSEN, J.K.; JOHNSON, N.; WILKERSON, C.G. Discovery of diversity in xylan biosynthetic genes by transcriptional profiling of a heteroxylan containing mucilaginous tissue. **Frontier in Plant Science,** v.4, p.1–15, 2013.

JENSEN, J.K., et al. Identification of a xylogalacturonan xylosyltransferase involved in pectin biosynthesis in Arabidopsis. **Plant Cell**, v.20, p.1289–1302, 2008.

JENSEN, J.K., et al. RNA-Seq Analysis of Developing Nasturtium Seeds (*Tropaeolum majus*) p. Identification Additional and Characterization of an additional Galactosyltransferase Involved in Xyloglucan Biosynthesis. **Molecular Plant,** v.5, p.984–992, 2012.

JOBLING, S.A. Membrane pore architecture of the CslF6 protein controls (1-3,1-4)-bglucan structure. **Science Advances**, v.1, p.e1500069, 2015.

KANTER, U., et al. Purification, characterization and functional cloning of inositol oxidase from Cryptococcus. **Yeast**, v.20, p.1317-1329, 2003.

KEEGSTRA, K., et al. The structure of plant cell walls p. III. A model of the walls of suspension- cultured sycamore cells based on the interconnections of the macromolecular components, **Plant Physiol**, v.51, p.188–197, 1973.

KELLOGG, E.A. Update on Evolution Evolutionary History of the Grasses. **Plant Physiol**, v.125, p.1198–1205, 2001.

KEPPLER, B.D.; SHOWALTER, A.M. IRX14 and IRX14-LIKE, Two Glycosyl Transferases Involved in Glucuronoxylan Biosynthesis and Drought Tolerance in Arabidopsis. **Molecular Plant,** v.3, p.834–841, 2010.

KIM, J.B.; OLEK, A.T.; CARPITA, N.C. Cell wall and membrane-associated exo-beta-dglucanases from developing maize seedlings. **Plant Physiol**, v.123, p.471–486, 2000.

KIM, M.; DAY, D. Composition of sugar cane, energy cane, and sweet sorghum suitable for ethanol production at Louisiana sugar mills. **J Ind Microbiol Biotechnol**, v.38, p.803-7, 2011.

KINKEMA, M., et al. Improved molecular tools for sugar cane biotechnology. **Plant Mol Biol**, v.84, p.497-508, 2014.

KOBAYASHI, M., et al. Purification and cDNA cloning of UDP-D-glucuronate carboxy-

lyase (UDP-D-xylose synthase) from pea seedlings. **Plant Cell Physiol,** v.43, p.1259-1265, 2002.

KONISHI, T., et al. A plant mutase that interconverts UDP-arabinofuranose and UDParabinopyranose. **Glycobiology**, v.17, p.345–354, 2007.

KOONIN, E.V. An apology for orthologs - or brave new memes. **Genome Biology,** v.2, p.1–2, 2001.

KOORNNEEF, M.; MEINKE, D. The development of Arabidopsis as a model plant. **The Plant Journal**, v.61, p.909–921, 2010.

KUREK, I., et al. Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. **Proc Natl Acad Sci USA**, v.99, p.11109–11114, 2002.

LAKSHMANAN, P., et al. Sugarcane biotechnology p. the challenges and opportunities. **In Vitro Cell Dev Biol Plant,** v.41, p.345–363, 2005.

LAMESCH, P., et al. The Arabidopsis Information Resource (TAIR) p. improved gene annotation and new tools. **Nucleic Acids Research**, v.40, p.D1202-10, 2011.

LAWRENCE, C.J.; WALBOT, V. Translational Genomics for Bioenergy Production from Fuelstock Grasses p. Maize as the Model Species. **The Plant Cell,** v.19, p.2091–2094, 2007.

LEE, C.H., et al. The F8H glycosyltransferase is a functional paralog of FRA8 involved in glucuronoxylan biosynthesis in Arabidopsis. **Plant Cell Physiol**, v.50, p.812–827, 2009.

LEE, C., et al. Functional roles of rice glycosyltransferase family GT43 in xylan biosynthesis. **Plant Signaling & Behavior,** v.9, p.e27809, 2009.

LEE, C., et al. The *PARVUS* gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. **Plant and Cell Physiology**, v.48, p.1659–1672, 2007.

LEMBKE, C.G., et al. Identification of sense and antisense transcripts regulated by drought in sugarcane. **Plant Mol Biol**, v.79, p.461–477, 2012.

LEROUXEL, O., et al. Biosynthesis of plant cell wall polysaccharides — a complex process. **Curr. Opin. Plant Biol,** v.9, p.621–630, 2006.
LIEPMAN, A.H., et al. Functional Genomic Analysis Supports Conservation of Function Among Cellulose Synthase-Like A Gene Family Members and Suggests Diverse Roles of Mannans. **Plant Physiol**, v.143, p.1881–1893, 2007.

LIEPMAN, A.H.; WILKERSON, C.G.; KEEGSTRA, K. Expression of celulose synthaselike (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. **Proc Natl Acad Sci USA**, v.102, p.2221–2226, 2005.

LIEPMAN, A.H., et al. Arabidopsis – a powerful model system for plant cell wall research. **The Plant Journal**, v.61, p.1107–1121, 2010.

LIU, J.; MUSHEGIAN, A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. **Protein Science p. A Publication of the Protein Society**, v.12, p.1418–1431, 2003.

LIU, L.; PAULITZ, J.; PAULY, M. The Presence of Fucogalactoxyloglucan and Its Synthesis in Rice Indicates Conserved Functional Importance in Plants. **Plant Physiology**, v.168, p.549–560, 2015.

MACEDO, I.C.; SEABRA, J.E.A.; SILVA, J.E.A.R. Green house gases emissions in the production and use of ethanol from sugarcane in Brazil p. the 2005/2006 averages and a prediction for 2020. **Biomass Bioenergy**, v.32, p.582–595, 2008.

MADSON, M., et al. The *MUR3* Gene of Arabidopsis Encodes a Xyloglucan Galactosyltransferase That Is Evolutionarily Related to Animal Exostosins. **Plant Cell**, v.15, p.1662–1670, 2003.

MANNERS, J.M.; CASU, R.E. Transcriptome Analysis and Functional Genomics of Sugarcane. **Tropical Plant Biology**, v.4, p.9–21, 2011.

MARRIOT, P.E., et al. Range of cell-wall alterations enhance saccharification in Brachypodium distachyon mutants. **Proc Natl Acad Sci USA**, v.111, p.14601–14606, 2014.

MASARIN, F., et al. Chemical composition and enzymatic digestibility of sugarcane clones selected for varied lignin content. **Biotechnology for Biofuels**, v.4, p.55, 2011.

MCCANN, M.C.; ROBERTS, K. Architecture of the primary cell wall. In: LLOYD C.W. The Cytoskeletal Basis of Plant Growth and Form. London:Academic Press, 1992. p. 109-129.

MCGINNIS, S.; MADDEN, T.L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. **Nucleic Acids Research**, v.32, p.20-25, 2004.

MEINKE, D.W., et al. *Arabidopsis thaliana*: A model plant for genome analysis. **Science**, v.282, p.662-682, 1998.

MENOSSI, M.; VINCENTZ, M.; SOUZA, G.M. Sugarcane Functional Genomics p. Gene Discovery for Agronomic Trait Development. **Int J Plant Genomics**, v.2, p.458732, 2008.

MITCHELL, R.A.C.; DUPREE, P.; SHEWRY, P.R. A Novel Bioinformatics Approach Identifies Candidate Genes for the Synthesis and Feruloylation of Arabinoxylan. **Plant Physiology**, v.144, p.43–53, 2007.

MOHNEN, D. Pectin structure and synthesis. Curr. Opin. Plant Biol, v.11, p.266–277, 2008.

MOLHOJ, M.; VERMA, R.; REITER, W.D. The biosynthesis of the branched-chain sugar D-apiose in plants p. functional cloning and characterization of a UDP-D-apiose/UDP-D-xylose synthase from Arabidopsis. **Plant J**, v.35, p.693-703, 2003.

MOLINARI, H.B.C., et al. Grass cell wall feruloylation p. distribution of bound ferulate and candidate gene expression in Brachypodium distachyon. **Frontiers in Plant Science**, v.4, p.50, 2013.

MORTIMER, J.C., et al. Absence of branches from xylan in Arabidopsis gux mutants reveals potential for simplification of lignocellulosic biomass. **Proc Natl Acad Sci USA**, v.107, p.17409–17414, 2010.

NEEDLEMAN, S.B.; WUNSCH, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. **J. Mol. Biol**, v.48, p.443-453, 1970.

NISHIYAMA JR, M.Y., et al. Full-Length Enriched cDNA Libraries and ORFeome Analysis of Sugarcane Hybrid and Ancestor Genotypes. **PlosOne**, v.9, p.e107351, 2014.

O'NEILL, M.A.; YORK, W.S. The composition and structure of plant primary cell walls. In: ROSE J.K.C. The plant cell wall. Oxford:Blackwell, 2003. 1-54.

OIKAWA, A., et al. An Integrative Approach to the Identification of Arabidopsis and Rice Genes Involved in Xylan and Secondary Wall Development. **PLoS ONE**, v.5, p.e15481, 2010.

PALHARES, A.C., et al. A novel linkage map of sugarcane with evidence for clustering of retrotransposon-based markers. **BMC Genetics**, v.13, p.51, 2012.

PANDEY, A.,et al. Biotechnological potential of agro-industrial residues. I p. sugarcane bagasse. **Bioresour Technol**, v.74 p.69–80, 2000.

PATERSON, A.H., et al. The *Sorghum bicolor* genome and the diversification of grasses. **Nature**, v.457 p.551–556, 2009.

PAULY, M., et al. Molecular domains of the cellulose/xyloglucan network in the cell walls of higher plants. **Plant J**, v.20, p.629–639, 1999.

PAULY, M., et al. Hemicellulose biosynthesis. Planta, v.238, p.627-642, 2013.

PAULY, M., et al. Changes in the structure of xyloglucan during cell elongation. **Planta**, v.212, p.842–850, 2001.

PENA, M.J., et al. A galacturonic acid- containing xyloglucan is involved in Arabidopsis root hair tip growth. **The Plant Cell**, v.24, p.4511–4524, 2012.

PENA, M.J.et al. Arabidopsis irregular xylem8 and irregular xylem9 p. Implications for the Complexity of Glucuronoxylan Biosynthesis. **The Plant Cell**, v.19, p.549-63, 2007.

PENNING, B.W., et al. Genetic resources for maize cell wall biology. **Plant Physiol**, v.151, p.1703–1728, 2009.

PERSSON, S., et al. The Arabidopsis irregular xylem8 mutant is deficient in glucuronoxylan and homogalacturonan, which are essential for secondary cell wall integrity. **Plant Cell**, v.19, p.237–255, 2007.

PETRASOVITS, L.A., et al. Enhanced polyhydroxybutyrate production in transgenic sugarcane. **Plant Biotechnol J**, v.10, p.569–578, 2012.

PIPERIDIS, G.; PIPERIDIS, N.; HONT, A.D. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. **Mol Genet Genomics**, v.284, p.65–73, 2010.

RANCOUR, D.M., et al. Cell wall composition and digestibility alterations in *Brachypodium distachyon* achieved through reduced expression of the UDP-arabinopyranose mutase. **Frontiers in Plant Science**, v.6, p.1–20, 2015.

RAUTENGARTEN, C., et al. The Interconversion of UDP-Arabinopyranose and UDP-

Arabinofuranose Is Indispensable for Plant Development in Arabidopsis. **The Plant Cell**, v.23, p.1373–1390, 2011.

REITER, W.D. Biochemical genetics of nucleotide sugar interconversion reactions. **Curr Opin Plant Biol**, v.11, p.236–243, 2008.

REITER, W.D.; VANZIN, G.F. Molecular genetics of nucleotide sugar interconversion pathways in plants. **Plant Mol Biol**, v.47, p.95–113, 2001.

RENNIE, E.A., et al. Three Members of the Arabidopsis Glycosyltransferase Family 8 Are Xylan Glucuronosyltransferases. **Plant Physiology**, v.159 p.1408–1417, 2012.

REIN, P.W. Prospects for the conversion of a sugar mill into a biorefinery. **Proc Inc Soc Sugar Cane Technol,** v.26, p.44–60, 2007.

RICHMOND, T.A.; SOMERVILLE, C.R. The cellulose synthase superfamily. **Plant Physiology**, v.124, p.495–498, 2000.

RODRÍGUEZ-GACIO, M.C., et al. Softening-up In Posidonia oceanica cadmium cell walls induces changes in DNA methylation and chromatin patterning. **Journal of Experimental Botany**, v.63, p.3975–3988, 2012.

SANJUAN, R., et al. Morphological and Chemical Composition of Pith and Fibers from Mexican Sugarcane Bagasse. **European Journal of Wood and Wood Products,** v.59, p.447-450, 2001.

SANKOF, D.; KRUSKAL, J.B. **Time warps, string edits and macromolecules**: The theory and practice of sequence comparision. Ontario:Addison-Wesley, 1983. p.1-44.

SCHELLER, H.V.; ULVSKOV, P. Hemicelluloses. Annu Rev Plant Biol, v.61, p.263–289, 2010.

SCHNABLE, P.S., et al. The B73 Maize Genome p. Complexity, Diversity, and Dynamics. Science, v.326, p.1112–1115, 2009.

SEIFERT, G.J. Nucleotide sugar interconversions and cell wall biosynthesis p. how to bring the inside to the outside. **Current Opinion in Plant Biology,** v.7, p.277–84, 2004.

SEITZ, B., et al. Matrix polysaccharide precursors in Arabidopsis cell walls are synthesized by alternate pathways with organ-specific expression patterns. **Plant Journal**, v.21, p.537–546, 2000.

SELLERS, P.H. On the theory and computation of evolutonary distances, **SIAM J. Appl. Math,** v.26, p.787-793, 1974.

SETTA, N., et al. Building the sugarcane genome for biotechnology and identifying evolutionary trends, **BMC Genomics**, v.30 p.1–17, 2014.

SCHEIBLE, W.R.; PAULY, M. Glycosyltransferases and cell wall biosynthesis p. novel players and insights, **Curr Opin Plant Biol**, v.7, p.285–295, 2004.

SIMS, R.E.H., et al. An overview of second generation biofuel technologies, **Bioresource Technology**, v.101, p.1570–1580, 2010.

SOMERVILLE, C., et al. Toward a systems approach to understanding plant cell walls, **Science**, v.306, p.2206–2211, 2004.

SOUZA, A.P., et al. Composition and Structure of Sugarcane Cell Wall Polysaccharides p. Implications for Second-Generation Bioethanol Productione, **BioEnergy Research**, v.6 p.564–579, 2013.

SPANNAGL, M., et al. Exploring the genomes: From Arabidopsis to crops, **Journal of Plant Physiology**, v.168, p.3–8, 2011.

STATES, D.J.; GISH, W. Combined use of sequence similarity and codon bias for coding region identification, **J Comput Biol**, v.1, p.39-50, 1994.

STERLING, J.D., et al. Functional identification of an Arabidopsis pectin biosynthetic homogalacturonan galacturonosyltransferase, **Proc Natl Acad Sci USA**, v.103, p.5236–41, 2006.

STICKLEN, M.B. Plant genetic engineering for biofuel production p. towards affordable cellulosic ethanol, **Nat Rev Genet**, v.9, p.433–443, 2008.

TAIZ, L.; ZEIGER, E. **Paredes celulares: estruturas, biogênese e expansão**. Fisiologia Vegetal. Porto Alegre:Artmed, 2004. Cap.15. p.339-364.

TAMURA, K., et al. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0, **Molecular Biology and Evolution,** v.30, p.2725-2729, 2013.

THE ARABIDOPSIS GENOME INITIATIVE. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, **Nature**, v.408, p.796–815, 2000.

THE INTERNATIONAL BRACHYPODIUM INITIATIVE. Genome sequencing and

analysis of the model grass Brachypodium distachyon, Nature, v.463, p.763-768, 2010.

THOMPSON, D.S. How do cell walls regulate plant growth. **J Exp Bot**, v56, p.2275–2285, 2005.

THOMPSON, J.D., HIGGINS, D.G., GIBSON, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, **Nucleic Acids Res**, v.22 p.4673-80, 1994.

URAHARA, T., et al. A  $\beta$ -(1->4)-xylosyltransferase involved in the synthesis of arabinoxylans in developing barley endosperms. **Physiologia Plantarum**, v.122, p.169–180, 2004.

VAN DER WEIJDE, T., et al. The potential of C4 grasses for cellulosic biofuel production. **Frontiers in Plant Science,** v.4, p.1–18, 2013.

VARSHNEY, R.K., et al. Can genomics boost productivity of orphan crops. **Nat. Biotechnol**, v.30, p.1172–1176, 2012.

VEGA-SÁNCHEZ, M.E., et al. Loss of Cellulose Synthase-Like F6 Function Affects Mixed-Linkage Glucan Deposition, Cell Wall Mechanical Properties, and Defense Responses in Vegetative Tissues of Rice. **Plant Physiology**, v,159, p.56–69, 2012.

VERHERTBRUGGEN, Y., et al. Mannan synthase activity in the CSLD family. **Plant** Signal Behav, v.6, p.1620–1623, 2011.

VETTORE, L., et al. Analysis and Functional Annotation of an Expressed Sequence Tag Collection for Tropical Crop Sugarcane. **Genome Research**, v.12, p.2725–2735, 2003.

VICENTINI, R., et al. Large-Scale Transcriptome Analysis of Two Sugarcane Genotypes Contrasting for Lignin Content. **PlosOne**, v.10, p.e0137698, 2015.

VOGEL, J. Unique aspects of the grass cell wall. **Curr Opin Plant Biol.** v.11, p.301–307, 2008.

WANG, M.L., et al. Production of biologically active GM-CSF in sugarcane p. a secure biofactory. **Transgenic Res**, v.14, p.167–178, 2005.

WANG, L., et al. Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. **BMC Plant Biology**, v.10, p.282, 2010.

WANG, M.L.; WANG S.; XIA G. From genome to gene p. a new epoch for wheat research. **Trends in Plant Science**, v.20 p.1360-1385, 2015.

WATERMAN, M.S. General methods of sequence comparison. Math Biol, v.46 p.473-500, 1984.

WU, A.M., et al. Analysis of the Arabidopsis IRX9/IRX9-L and IRX14/IRX14-L pairs of glycosyltransferase genes reveals critical contributions to biosynthesis of the hemicellulose glucuronoxylan. **Plant Physiol**, v.153, p.542–554, 2010.

YIN, Y., et al. Evolution and Function of the Plant Cell Wall Synthesis-Related Glycosyltransferase Family 8. **Plant Physiology**, v.153 p.1729–1746, 2010.

YIN, Y., et al. Evolution of Plant Nucleotide-Sugar Interconversion Enzymes. **PLoS ONE**, v.6, p.e27995, 2011.

YORK, W.S.; O'NEILL, M.A. Biochemical control of xylan biosynthesis - which end is up. **Curr Opin Plant Biol**, v.11, p.258–265, 2008.

ZENG, W., et al. A Glucurono(arabino)xylan Synthase Complex from Wheat Contains Members of the GT43, GT47, and GT75 Families and Functions Cooperatively. **Plant Physiology**, v.154, p.8–97, 2010.

ZHANG, B., et al. Functional conservation of the glycosyltransferase gene GT47A in the monocot rice. **Journal of Plant Research**, v.127, p.423–432, 2014.

ZHU, X.G.; LONG, S.P.; ORT, D.R. What is the maximum efficiency with which photosynthesis can convert solar energy into biomass. **Curr Opin Biotechnol**, v.19, p.153-159, 2008.