UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

RAFAEL HENRIQUE GALLINARI

# UNDERSTANDING SUGARCANE CELL WALL TO INCREASE 2G ETHANOL PRODUCTION

# COMPREENDENDO A PAREDE CELULAR DE CANA DE AÇÚCAR PARA O AUMENTO DA PRODUÇÃO DE ETANOL 2G

Campinas

2020

**RAFAEL HENRIQUE GALLINARI**

# UNDERSTANDING SUGARCANE CELL WALL TO INCREASE 2G ETHANOL PRODUCTION

# COMPREENDENDO A PAREDE CELULAR DE CANA DE AÇÚCAR PARA O AUMENTO DA PRODUÇÃO DE ETANOL 2G

Thesis presented to the Institute of Biology of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Genetics and Molecular Biology, in the area of Plant Genetics and Breeding.

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Genética e Biologia Molecular, na área de Genética Vegetal e Melhoramento.

Advisor: PROF. DR. MARCELO MENOSSI TEIXEIRA

ESTE ARQUIVO DIGITAL CORRESPONDE A VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO RAFAEL HENRIQUE GALLINARI E ORIENTADO PELO PROF. DR. MARCELO MENOSSI TEIXEIRA.

**Campinas**

**2020**

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

Gallinari, Rafael Henrique, 1992-
G137u Understanding sugarcane cell wall to increase 2G ethanol production / Rafael Henrique Gallinari. – Campinas, SP : [s.n.], 2020.

Orientador: Marcelo Menossi.
Coorientador: Pedro Araújo.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Cana-de-açúcar. 2. Etanol 2G. 3. Interferência de RNA. 4. Ácido glucurônico. 5. Hemicelulose. I. Menossi, Marcelo, 1968-. II. Araújo, Pedro, 1985-. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

Informações Complementares

**Título em outro idioma:** Compreendendo a parede celular de cana de açúcar para o aumento da produção de etanol 2G
**Palavras-chave em inglês:**
Sugarcane
2G ethanol
RNA Interference
Glucuronic acid
Hemicellulose
**Área de concentração:** Genética Vegetal e Melhoramento
**Titulação:** Doutor em Genética e Biologia Molecular
**Banca examinadora:**
Marcelo Menossi [Orientador]
Andre Luís Ferraz
Igor Cesarino
Jennifer Charlotte Mortimer
Marcos Silveira Buckeridge
**Data de defesa:** 31-10-2020
**Programa de Pós-Graduação:** Genética e Biologia Molecular

**Identificação e informações acadêmicas do(a) aluno(a)**
- ORCID do autor: https://orcid.org/0000-0003-0223-0304
- Currículo Lattes do autor: http://lattes.cnpq.br/6229858286385466

Campinas, 30th October 2020

# Certificate of Approval

## Examining Committee

Certificate of approval, giving visibility to the Examining Committee with the information that the Minutes of Defence, signed by the members of the Examining Committee, appear in the SIGA / Thesis Flow System and in the Unit's Program Secretariat;

Prof. Dr. Marcelo Menossi (Advisor)

Prof. Dr. Andre Luís Ferraz

Prof. Dr. Igor Cesarino

Dr. Jennifer Charlotte Mortimer

Prof. Dr. Marcos Silveira Buckeridge

*Os membros da Comissão Examinadora acima assinaram a Ata de defesa, que se encontra no processo de vida acadêmica do aluno.*

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa de Genética e Biologia Molecular do Instituto de Biologia

**ACKNOWLEDGMENTS**

In first of all I would like to thank my advisor Marcelo Menossi, whom was extremely important for the development of my scientifical background and logical reasoning. Beside that he was always there ready to help and support with technical problems and possible solutions. Another very important thing that I learned with him, is that we need to learn how to plan everything in our lives, in order to be ready even for the most unlikely futures. Despite the fact that he is a person very busy with all his tasks as a full professor, he will book a time for each one of his students and will want to know what is happening in order to guarantee that the lab runs most smoothly possible.

Another person I would like to show my gratitude is Pedro Araújo, whom was of extremely importance for my growth and development as a scientist during my whole PhD. He was responsible to break a paradigm that the protocols need to be strictly followed showing that the importance to understand the main objective of each analysis is most important, and it can be done in different ways, our role is to find out which is the best one for us. Besides that, he was really important teaching me how to behave in a professional way and giving psychological and moral support during every chosen choice during my work. I consider him as a really reliable and a great friend.

I would like to thanks my lab friends, starting with Giovanna Guidelli, whom was my first mentor and a really good friend inside the lab, even before starting my PhD. Rafael Della was another important personality, because in a time which I wasn´t able to proceed with some phylogenetic analysis, he was the person that showed me a beautiful way that facilitated everything and together with professor Mariana Nery helped my work to keep going forward. And of course, all the others lab mates and friends: Aline Yochikawa, Daniel Neri, William, Maria Carolina, Denis Bassi, Max Tsunada, Jose Sergio, Lucia Matiello, Ana Laura, Nicolas, Rafael Tavares, Thais Ferreira, Valter Alessio, Giane Bueno, Camila Purchatti, Isabella Poletto, Germanna Righetto, Wilson, Camila Cunha, Marcelo Menossi, Paulo de Lucca and Juliana.

Thanks, from the bottom of my heart to all my family who has always supported me. For my home family: Edmilson, Eidilene, Felipe and Beatriz, Bruno and Thais, and Fernanda. For my nephews and niece: Miguel, Matheus and Giovana. For my cousins: Thábata and Xande, Thais, Nicole, Lucas, Gabriela (Juninha), Gabriela and Carlos, Izabela, Natalia, Iohran and Illana. For my grandparents: Aparecida Rocha, Jurandir and Odila.

For all the collaborations and the friendship that we started with indefinite period: Mariana Nery, Sarita Rabelo and Juliana Mayer.

I am so grateful for all the time spend together with all my friends, that everytime I needed they were there and even new friends that appeared just in the right time: Guilherme Bersi, Jessica Stefany do Vale**,** Caio Vasconcelos, João Campos, Lucas Dalava, Marcia Passariello, Juan Llerena, Rosa Lorizolla Cordeiro, Aldair, Michelle Muffato, Breno Barbosa, Mariane Sarraipa, Juliana Boldrin, Renato Petitto Netto, Bruna and Fernanda Eleuterio. In addition to them, everybody who contributed with to me to have better days as a teacher: Perry and Marry Krassner, Marcia Luders, Tigrinhos and Golden tiger's staff, Guilherme Francisco, Francisco (Gui´s dad) Sophia Rovere, Rita Delicato, Kiki Stier, Ligia Zaiden, Paulo Tonhasolo, Sabrina Siqueira and everybody who was my students.

Thank so much for my GBMigos: Elisa, Marina, Renna, Livia, Ludimilla, Joao, Maria Carolina and Lidia. And thanks a lot, to all my friends from Cambridge: Deusa and Leonardo, Ian Wood, Henry Temple, Oliver Terret, Li Yu, Alberto Echevarria-Poza, Kathy, Yoshi, Jan Lyczakowski, Keisuke, Dora, Wendy Gibson, Nadine, Xiaolan, Liam Jones, Louis Wilson, Loubab Zedane, Paul Dupree, Yusuf. In memory of a friend that received me every lunch at the department of Biochemistry with a cozy "Hello Rafel, how are you going? How can I help you?" thanks for making my stays a little bit warmer with those small gestures Kevin.

Another person, whom was of extremely importance for my development as a scientist is Jan Lyczakowski. Since the first time we met each other, he embraced me as a little pupil and taught me everything he could during his last semester in Cambridge. And even after that he worked as an advisor and guided me through critical moments. If that wasn´t enough he encouraged me to move on since the beginning until the last words typed in this thesis. I consider you as a brother for everything you did and still do. Thank you very much.

I am really grateful for all the help and support that prof. Paul Dupree gave me. He gave me a vote of confidence and believed that since the beginning our collaboration would work. Not only that, but he tried everything he could to be able to bring me to Cambridge and materialize this collaboration. Besides that, he was an excellent leader and advisor exceeding all my expectations in the best possible way. I have learned a lot with him in a professional and personal way. I consider him a real friend.

And thanks to everybody whom I might have forgotten to put here, but helped me in some way to arrive here.

There are people that just a thank you will not be enough, so a special thanks to Pedro Araujo, Jan Lyczakowski and Paul Dupree who were responsible for giving me light and happiness even in the darkest days.

For National Council for Scientific and Technological Development (CNPq: 140868/2016-0) and São Paulo Foundation Research (FAPESP: 2017/15895-4) for financial support. This study was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Finance Code 001 (88887.364139/2019-00).

**ABSTRACT (ENGLISH)**

There is an increasing demand for renewable energy sources. Bioethanol from sugarcane is among the best alternatives due to its cost effectiveness. The sugarcane bagasse, which has a high energy potential, recently became a target for the production of second-generation ethanol (2G). However, the structure of the plant cell walls evolved in such a way that the release of monosaccharides used to produce 2G ethanol is a major challenge. In model organisms it was shown that the *GUX2* gene is involved in the structure of hemicellulose, one of the main polysaccharides in plant cell walls. When *GUX2* is silenced, the amount of glucuronic acid (GlcA) deposited on xylan is reduced, which ultimately decreases the cell wall recalcitrance for the saccharification process. In this work, we identified and characterized two sugarcane GUX enzymes (ScGUX1 and ScGUX2) through *in vitro* and *in vivo* assays. Complementation studies using Arabidopsis mutants for *GUX1* or *GUX2* indicated that the ScGUX1 and ScGUX2 activities are conserved between mono and dicots. Regarding the GlcA deposition patterning over the xylan in these plants, we observed that ScGUX1 has a preference to add GlcA every six xyloses, while ScGUX2 adds GlcA every five and seven xyloses. Moreover, the *ScGUX2* gene from sugarcane was silenced, resulting in plants with a significantly reduction in GlcA deposition over xylan. In these *ScGUX2*-silenced plants the GlcA addition pattern – every five and seven xyloses, disappeared, corroborating the previous *in vivo* and *in vitro* assays. This change resulted in a bagasse with a higher saccharification: up to 22% more glucose and up to 43% more xylose were released. Interestingly, the silenced plants did not show any abnormality in their development compared to wild type controls. These results indicate that the recalcitrance of the cell wall in the silenced events is reduced in comparison to wild type plants.

**ABSTRACT (PORTUGUESE)**

Existe uma demanda crescente por fontes renováveis de energia. O bioetanol de cana-de-açúcar está entre as melhores alternativas devido ao seu baixo custo. O bagaço da cana-de-açúcar, de alto potencial energético, tornou-se recentemente alvo para a produção de etanol de segunda geração (2G). No entanto, a estrutura das paredes celulares das plantas evoluiu de tal forma que a liberação dos monossacarídeos usados para produzir o etanol 2G é um grande desafio. Em organismos modelo, foi demonstrado que o gene *GUX2* está envolvido na estrutura da hemicelulose, um dos principais polissacarídeos da parede celular vegetal.

Quando o *GUX2* é silenciado, a quantidade de ácido glucurônico (GlcA) depositado no xilano é reduzida, o que acaba diminuindo a recalcitrância da parede celular para o processo de sacarificação. Neste trabalho, identificamos e caracterizamos duas enzimas GUX da cana-de-açúcar (ScGUX1 e ScGUX2) por meio de ensaios *in vitro* e *in vivo*. Estudos de complementação usando mutantes de Arabidopsis para *GUX1* ou *GUX2* indicaram que as atividades de ScGUX1 e ScGUX2 são conservadas entre mono e dicotiledôneas. Em relação ao padrão de deposição de GlcA sobre o xilano nessas plantas, observamos que ScGUX1 tem preferência em adicionar GlcA a cada seis xiloses, enquanto ScGUX2 adiciona GlcA a cada cinco e sete xiloses. Além disso, o gene *ScGUX2* da cana-de-açúcar foi silenciado, resultando em plantas com redução significativa na deposição de GlcA sobre o xilano. Nessas plantas silenciadas com *ScGUX2*, o padrão de adição de GlcA - a cada cinco e sete xiloses, desapareceu, corroborando os ensaios *in vivo* e *in vitro* anteriores. Essa alteração resultou em um bagaço com maior sacarificação: foram liberados até 22% mais glicose e até 43% mais xilose. Curiosamente, as plantas silenciadas não mostraram qualquer anormalidade no seu desenvolvimento em comparação com os controles do tipo selvagem. Estes resultados indicam que a recalcitrância da parede celular nos eventos silenciados é reduzida em comparação com as plantas do tipo selvagem.

## ABBREVIATIONS

| | |
|---|---|
| **AIR:** | alcohol insoluble residues |
| **AmAc:** | ammonium acetate |
| **ANOVA**: | Analysis of variances |
| **ANTS:** | 8-aminonapthalene-1,3,6-trisulphonic acid |
| **Ara:** | Arabinose |
| **At:** | *Arabidopsis thaliana* |
| **DUF:** | Domain of Unknown Function |
| **Gal:** | Galactose |
| **GalA:** | Galacturonic acid |
| **GAX:** | Glucuronoarabinoxylan |
| **GH:** | glycosyl hydrolase |
| **GlcA:** | Glucuronic acid |
| **Glu:** | Glucose |
| **GT:** | Glycosyl Transferase |
| **GUX:** | GlucUronic acid substitution of Xylan |
| **GX:** | Glucuronoxylan |
| **Hz:** | Hertz |
| **IRX:** | irregular xylem |
| **LB:** | Lysogeny broth |
| **Man:** | Mannose |
| **MeGlcA:** | 4-O-methylglucuronic acid |
| **MES:** | 4-Morpholineethanesulfonic acid |
| **mRNA:** | messenger Ribonucleic acid |
| **MS:** | Murashige and Skoog |
| **NEB:** | New England Biolabs |
| **PACE:** | polysaccharide analysis by carbohydrate gel electrophoresis |
| **RGI:** | Rhamnogalacturonan-I |
| **RGII:** | Rhamnogalacturonan-II |
| **Rha:** | Rhamnose |
| **Sc:** | *Saccharum spp* |

| | |
|---|---|
| **UDP:** | Uridine diphosphate |
| **UDP-GlcA:** | Uridine diphosphate glucose Glucuronic Acid |
| **UX:** | Uridine diphosphate |
| **UX4:** | glucurono-xylotetraose |
| **WB:** | Western Blot |
| **WT:** | wild-type |
| **XXT:** | Xyloglucan xylosyltransferase |
| **XyG:** | Xyloglucan |
| **Xyl:** | Xylose |

# SUMMARY

# CHAPTER 1: INTRODUCTION

## 1. PLANT CELL WALL

Plant cell walls are composed of a matrix of complex polysaccharides, proteins and phenolic compounds (Figure 1). Cell walls are directly linked to growth, cell differentiation, intercellular communication and water movement in plants (Cosgrove, 2005). Two distinct types of cell walls are differentiated: primary and secondary (Figure 2A).

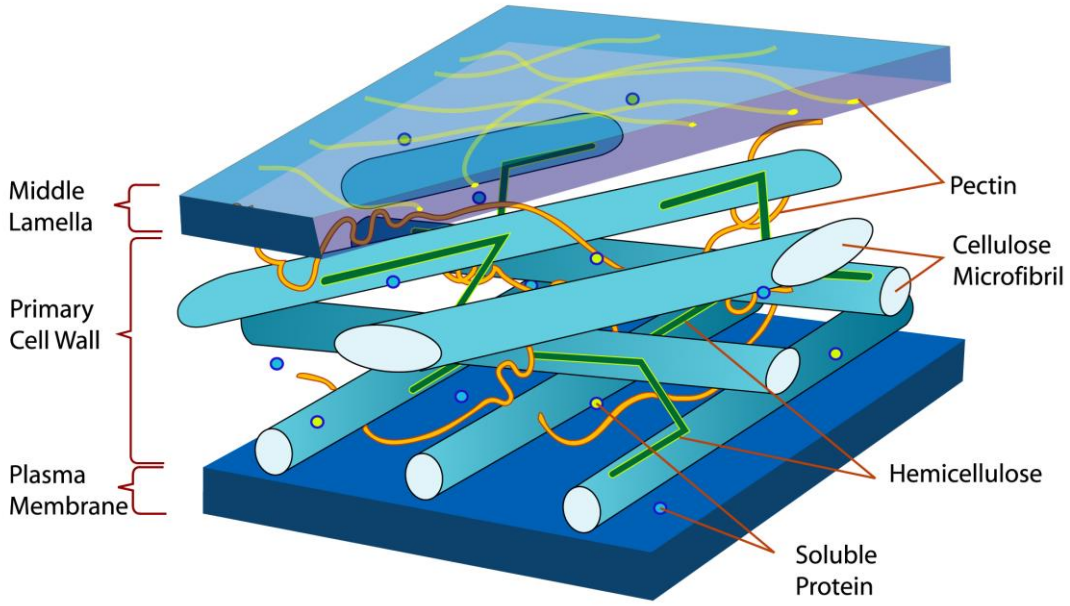



**Figure 1 Plant primary cell wall contain cellulose microfibrils, hemicelluloses, pectin and soluble proteins**. "Plant cell wall diagram" by LadyofHats. Licensed under Public Domain via Wikimedia Commons

The primary cell wall (Figure 2B) is composed of microfibrils of cellulose, which are incorporated into a matrix of complex polysaccharides that are divided into two classes: pectins and hemicelluloses. Pectins are a diverse group of polysaccharides, likely to be linked via covalent bonds, which are soluble in aqueous buffers, acidic solutions and calcium chelators. Hemicelluloses comprise polysaccharides that require alkaline solutions for solubilization (Zhong & Ye, 2009). The secondary cell walls (Figure 2C) are mainly present in tracheary elements and wood fibers, providing mechanical strength to withstand the negative pressure generated during transpiration and to enable the entire plant to upright growth. Cellulose, hemicellulose and lignin are the main components of plant secondary cell walls (Zhong & Ye, 2009). Lignin is a term that covers a large group of aromatic polymers resulting

from the oxidative coupling of 4-hydroxyphenylpropanoids (monolignols). Lignins are generally deposited during secondary cell walls formation, making the thickened walls rigid and impermeable to water (Vanholme et al., 2010). The cellulose microfibrils form a crosslinked network with hemicellulose providing more rigidity and hydrophobicity to plant cell wall (Zhong & Ye, 2009).

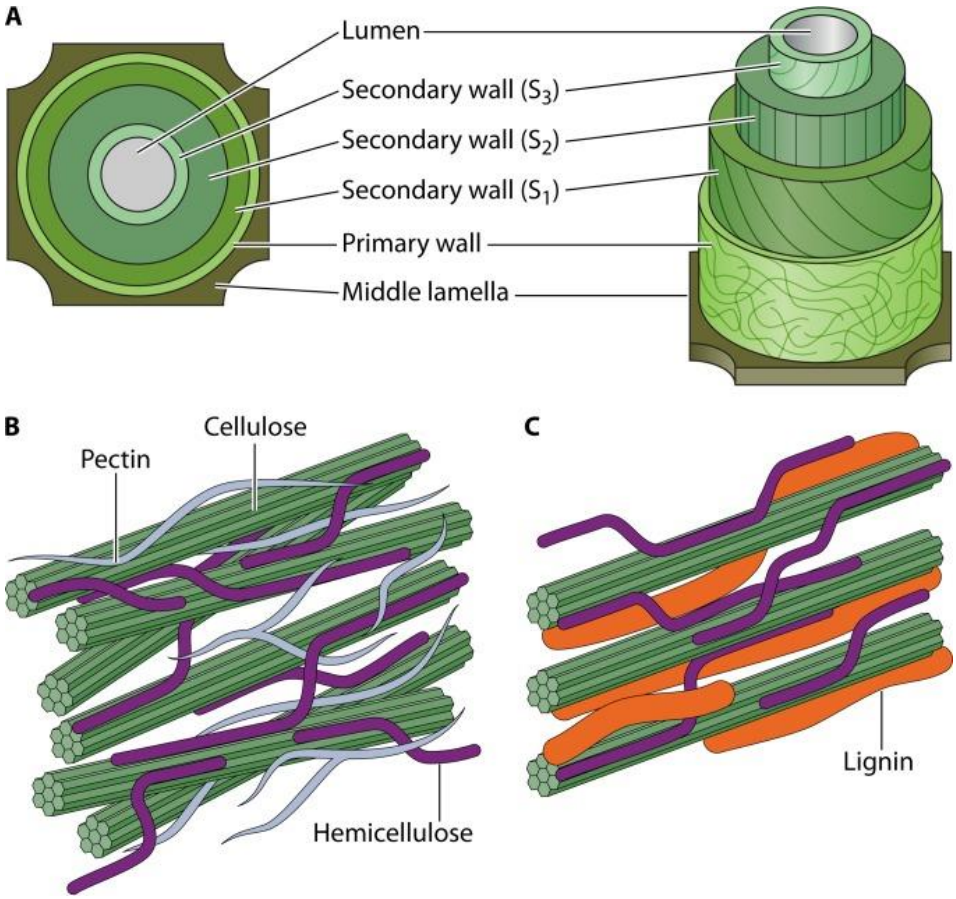



**Figure 2 Simplified model of plant cell wall structure.** (A) The structure consists of three main layers: the middle lamella and the primary and secondary walls. (A). Additional layers ($S_1$to $S_3$) are identified for the secondary cell wall. The main classes of compounds forming the primary (B) and secondary wall (C) are also shown (Rytioja et al., 2014).

## 1.1. CELLULOSE

Cellulose is the main organic compound and the most abundant polysaccharide derived from biomass. Not only of great importance for paper production industry, this polymer can be used for many approaches for different purposes as in textile, materials, chemical and biotechnology industries (Samir et al., 2005; Ummartyotin & Manuspiya, 2015).

Because of the great abundance of plant biomass worldwide, cellulose is frequently known as the most abundant bio-polymer in the world (Sommervile, 2006). The simple structure of cellulose is composed by a structure of a homo-polymer of β-1,4-linked glucose units (Figure 3; Hallac & Ragauskas, 2011). During cellulose microfibril formation process, chains of glucan acquire a twofold helical conformation and assemble into thin layers through edge-to-edge hydrogen bonding (Nishiyama et al., 2002). The hydrogen bonding between glucan results into the formation of cellulose microfibrils that can have a variation in the amount and size of individual glucan chains among different species (Jarvis, 2018). Cellulose microfibrils are formed by the association of 18 or 24 glucan chains (Wingren et al., 2003). The crystallinity, diameter and length of cellulose microfibrils are particularly variable and strictly related to the age and type of plant tissue.

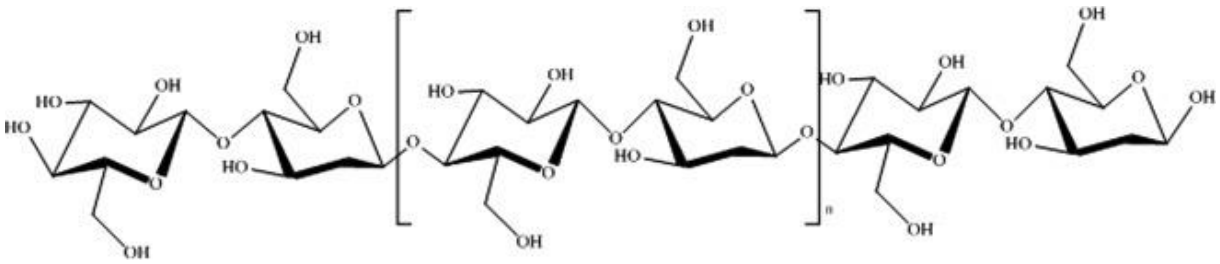


**Figure 3 Scheme representing cellulose structure** (Ummartyotin & Manuspiya, 2015)

There are multiple cellulose synthase (CESA) catalytic subunits that together are responsible to synthesise cellulose glucan chains in plants and organize into the microfibril, which is catalysed by mobile plasma membrane localized cellulose synthesis complexes (CSCs) (Kimura et al., 1999). Many CESAs are responsible for the biosynthesis of primary and secondary cell wall cellulose (Holland et al., 2000, Fagard et al., 2000). In Arabidopsis, CESA family is composed by 10 members (Kumar and Turner, 2015). In secondary cell wall, more specifically in the xylem, the cellulose synthesis is catalysed by the activity of CESA4, CESA7 and CESA8 (Doblin et al., 2002). Mutations in *CESAs*, which encodes secondary cell wall specific proteins, results in a drastically reduction in stem cellulose amount and in xylem vessels collapsed. Since CESAs have been identified through discovery of irregular xylem (IRX) mutants (Turner and Somerville, 1997), the genes *CESA4*, *CESA7* and *CESA8* can be also referred as *IRX5*, *IRX3* and *IRX1* respectively (Taylor et al., 2003). In addition to CESAs, there are auxiliary proteins, such as STELLO and KORRIGAN that supports the activity and trafficking of CSCs,

having some relevance into microfibril formation process (Zhang et al., 2016, Vain et al., 2014). Despite being known the importance of those auxiliary proteins, the specific pathway how it occurs is still unclear.

## 1.2. LIGNIN

Lignin is a phenylpropanoid biopolymer and has the role of giving mechanical strength to the plant as well as microorganisms resistance (Ponnusamy et al., 2019). It is believed to be connected with cellulose and hemicellulose through covalent and hydrogenic linkages becoming extraordinarily strong and recalcitrant to pretreatment approaches commonly used to increase saccharification (Sun & Cheng, 2002). Lignin is made mainly from three building blocks: p-coumaryl, coniferyl and sinapyl alcohols. These building blocks, also known as monolignols, have differences in the quantity of methoxy groups on phenolic nucleus and are generally abbreviated as S (syringil), G (guaiacyl) and H (p-hydroxyphenyl) (Vanholme et al., 2010) monomers respectively. The relative amount of S, G and H in lignin varies among different species of plants. Generally, hardwood lignin (e.g. eucalyptus, poplar, oak) contains both G and S units, whereas softwood lignin (e.g. pine, spruce) is composed almost entirely of G units. Herbaceous biomass is made from all three units, but like in hardwood and softwood their H content is also very low. (Vanholme et al., 2012).

The process of lignin deposition in secondary cell walls is denominated lignification and is frequently related to apoptosis (Meents et al., 2018). Lignification is of great importance for the maintenance plant biomass properties. Plants with lower lignin content or modified monolignol composition frequently have xylem vessels collapsed and are seriously dwarfed (Bonawitz and Chapple, 2010)

## 1.3. PECTIN

Pectin is a very complex group of polysaccharides with different roles in plant development, growth, morphology and defense against microbial and herbivores attack (Mohnen, 2008a). It is composed by approximately 70% of D-galacturonic acid (GalA) and all the pectic polysaccharides have GalA acid linked at the O-1 and the O-4 position (Mohnen, 2008a). The pectic groups include homogalacturonan (HG), xylogalacturonan (XGA), rhamnogalacturonan I (RGI) and the substituted rhamnogalacturonan II (RGII) (Figure 4,

Harholt et al., 2010). A minimum of 67 transferases are required for the biosynthesis of pectin (Mohnen, 2008b). Pectin represents a significantly amount of primary cell walls, corresponding to approximately 35% in dicots and non-graminaceous monocots, 2 to 10% of grass and other commelinoids and up to 5% in woody plants biomass (Mohnen, 2008a).

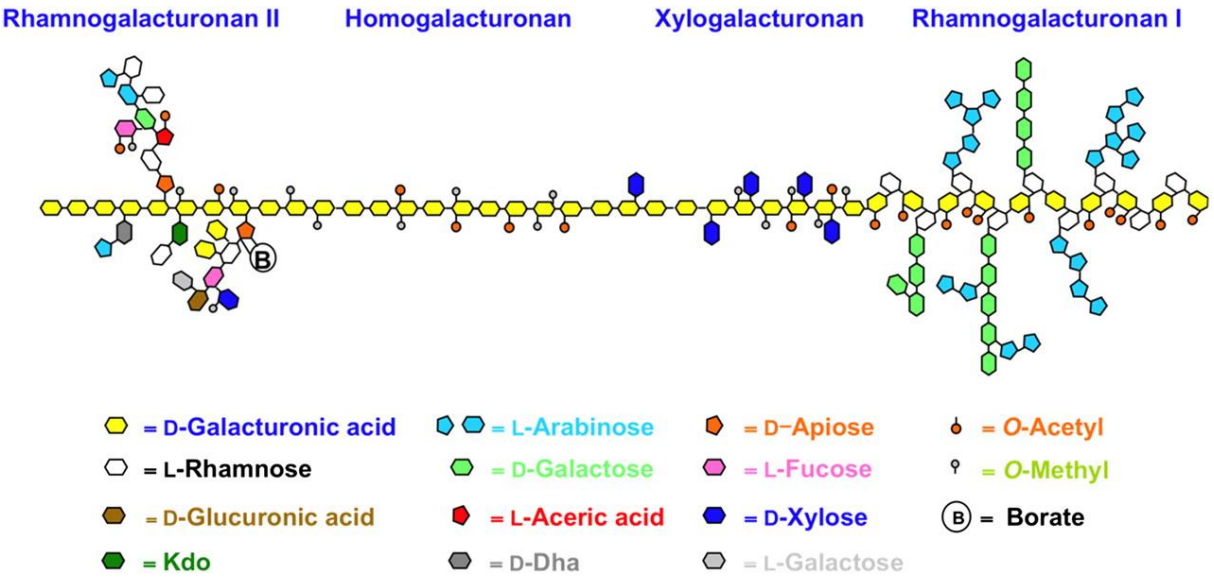



**Figure 4 Scheme of pectin structure.** This representative figure shows the four pectic groups (RGII, HG, XGA and RGI) with its respectively branches (Harholt et al., 2010).

In food and cosmetic industries, pectin is commonly used as stabilizing and gelling agent. Besides that, many studies show several beneficial effects on human health like decreasing cholesterol and serum glucose levels, reducing cancer (Jackson et al., 2007) and stimulating immune system (Inngjerdingen et al., 2007)

## 2. HEMICELLULOSE

Hemicelluloses are a class of non-cellulosic and non-pectic plant cell wall polysaccharides with β-1,4-linked backbones composed of glucose, mannose or xylose monomers. The main biological role of hemicelluloses is to contribute to the cell wall strengthening by the interaction with cellulose and other cell wall components (Scheller & Ulskov, 2010). Hemicelluloses are divided into: xyloglucans, xylans and heteromannans.

## 2.1. XYLOGLUCAN

Xyloglucan (XyG) is the dominant hemicellulose in primary cell walls of most spermatophytes, with a notable exception for grasses. Xyloglucans (XyG), like cellulose, are made of β-1,4-linked glucans, but the backbone is heavily substituted with with α-1,6-xyloses. Those branches can be further substituted with β-1,2-linked Galactopyranose, which may have α-1,2-linked L-Fucopyranose substitutions (Scheller & Ulskov, 2010). Branching of xyloglucan prevents its bonding to hydrogen along the length of the molecules, meaning that XyG is a highly flexible polysaccharide (Peña et al., 2008).

## 2.2. MIXED-LINKAGE GLUCAN

Mixed-linkage glucan (MLG), sometimes referred as to β-glucan, are also similar to cellulose, but instead of only β-1,4-linked glucosides, single β-1,3- linked glucosides are present between blocks of β-1,4-linked glucosides introducing irregularities in the chain (Fincher, 2009). MLG is largely found in endosperm cell walls and it is suggested to be a storage carbohydrate. In vegetative tissues, it was shown to be accumulated in primary cell walls of young and organs through elongation process. (Vega-Sánchez et al., 2013)

## 2.3. MANNAN

Heteromannans, including mannans, glucomannans, galactomannans and galactoglucomannans, are a family of polysaccharides which encompasses polymers based on a backbone with a high content of β-1,4-linked mannose subunits (Matheson, 1990). Mannan is one of the main components present in cell wall of vascular plants (Moreira & Filho, 2008). Heteromannans are composed of a β-1,4-linked backbone of mannose, forming mannans or galactomannans, or a combination of glucose and mannose, forming glucomannans and galactoglucomannans (Liepman et al., 2007; Ebringerova et al., 2005). Galactoglucomannans are the main hemicelluloses of gymnosperm secondary cell walls (Ebringerova et al., 2005). There is also a significant heteromannan content in cell walls of early land plants such as mosses and lycophytes (Harholt et al., 2008). In spermatophyte secondary cell walls, mannans and glucomannans are less prevalent and seem to be replaced by other hemicelluloses (Goubet et al., 2003). In this work we will focus mainly on xylan. The main polysaccharides found in plants are represented in Figure 5.

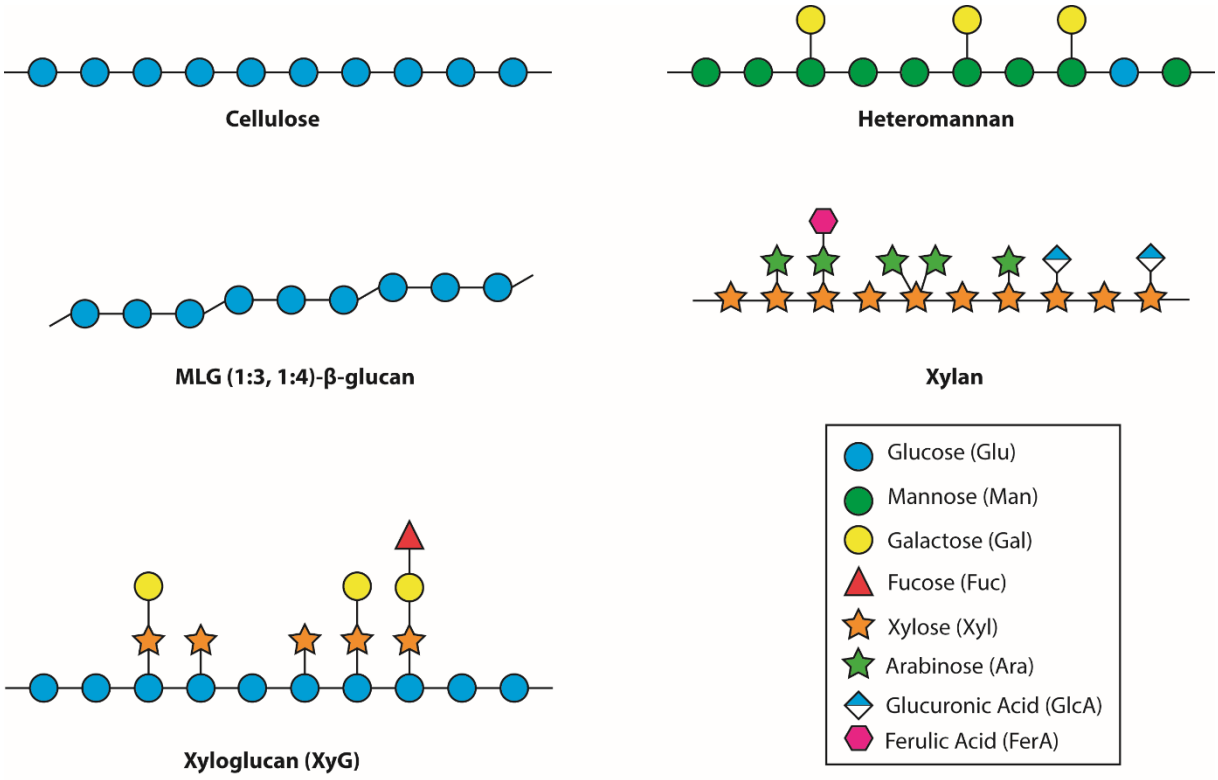



**Figure 5 Main polysaccharides found in plants**

## 2.4. XYLAN

Xylan is a polymer composed of β-1,4-linked xylose (Xyl) residues decorated with glucuronic acid (GlcA), acetyl, 4-O-methylglucuronic acid (MeGlcA) and arabinose (Ara) residues (Rennie & Scheller, 2014). The great majority of xylans that have been studied hold different amounts of α-1–2-linked GlcA branches that can be further modified. A very frequent modification of GlcA includes methylation on carbon four resulting in the formation of 4-O-methyl-glucuronic acid (MeGlcA) (Wilkie, 1979; Yoo et al., 2012; Scheller and Ulvskov, 2010).

There is a variation in the degree of GlcA methylation among different plants (Ebringerova et al., 2005, Peña et al., 2007, Kulkarni et al., 2012). For xylans that bear both methylated and non-methylate GlcA, the branching is designated as [Me]GlcA, representing both forms (Mortimer et al., 2010). Besides methylation, GlcA can also be decorated by neutral sugars, which include, in primary cell walls of *Arabidopsis thaliana* and in monocots, α-1,2-linked arabinofuranose residues (Mortimer et al., 2015) and galactose in *Eucalyptus grandis* (Peña et al., 2016). In addition to glucuronidation, the backbone can be branched with α-D-glucuronic

acid (GlcpA), α-L-arabinofuranosyl (Araf) residues, or a combination of both to give glucuronoxylan (GX), arabinoxylans (AX), and glucoronarabinoxylans (GAX). C2-O and C3-O can furthermore be acetylated (Wilkie, 1979; Yoo et al., 2012).

Different species of plants, and even tissues of the same species, present variation in xylan structures (Carpita, 1996). In grasses, primary and secondary cell walls contain considerable amounts of xylan. Xylan can be also found in primary cell walls of dicots with but the amount is significantly lower than in grass primary cell walls. On the other hand, in secondary cell wall of dicots, xylan is the predominant hemicellulose, but unlike in grasses it has no arabinosyl side chains (Anders et al., 2012; Scheler & Ulskov, 2010). In grasses arabinose branches of xylan can be substituted by xylose on position O-2 (Chiniquy et al., 2012) or coumaric and ferulic acid attached to position O-5 (Scheller and Ulvskov, 2010). In figure 6 we can see the different structures that can be attached to the xylan in dicots and grasses.

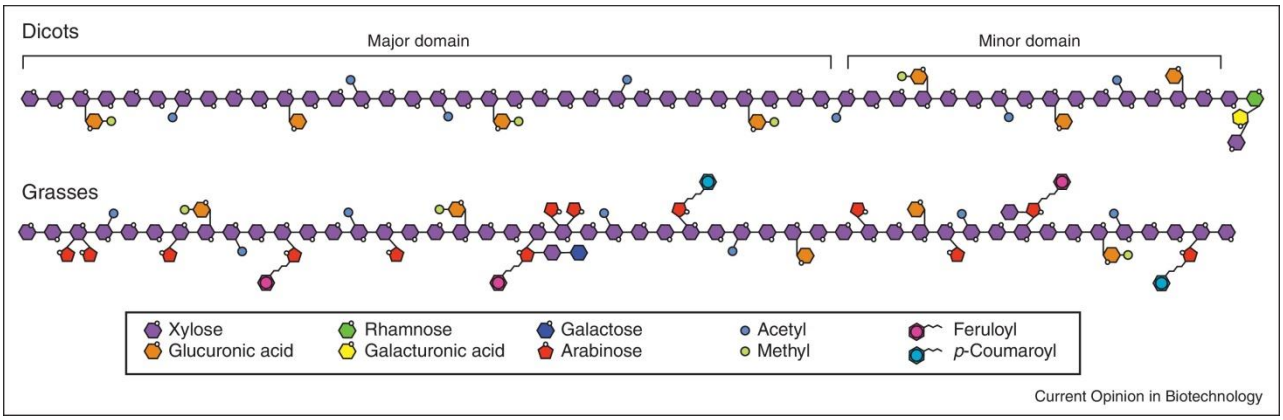



**Figure 6. Differences between dicots and grasses xylan decoration** (Rennie E & Scheller, 2014).

## 2.5. XYLAN BIOSYNTHESIS

Xylan biosynthesis happens inside the cisternae of the Golgi apparatus of plant cells and is believed to occur over several stages (Scheller and Ulvskov, 2010). Xylan backbone biosynthesis is catalysed by proteins of GT43 and GT47 families. In *Arabidopsis thaliana, it* has been reported that secondary cell wall xylan synthesis is catalysed mainly by IRX9, IRX10 and IRX14 proteins and for primary cell walls IRX9-Like (IRX9L), IRX10L and IRX14 are involved in the process (Mortimer et al., 2015, Wu et al., 2010). It was suggested that GT43 family members (IRX9 and IRX14) can combine with a GT47 family member (IRX10) forming a Xylan Synthase Complex (XSC) (Zeng et al., 2016). Urbanowicz et al. (2014) demonstrated that IRX10L is capable of synthetizing the xylan backbone *in vitro* and suggested that GT43 proteins

IRX9 and IRX14 increases the synthesis of xylan either by being responsible to initiate a role in xylan chain or forming a multimeric xylan synthase complex. *TaIRX9* knockouts showed shorter chain lengths of xylan in wheat grains (Pellny et al., 2020). Mutations in XSC proteins result in a decrease in xylan content, collapsed vessels and plant dwarfism (Wu et al., 2010, Brown et al., 2009).

The incorporation of neutral sugar branches to xylan is believed to be catalysed by GT61 family members (Anders et al., 2012, Zhong et al., 2018). In grasses, the xylan is widely substituted with arabinofuranose and in wheat, a GT61 responsible for adding α-(1,3)–linked arabinose was characterized with mutagenesis and gain of function studies (Anders et al., 2012). BAHD acyltransferase family enzymes may be responsible for the addition of ferulic acid to arabinose ahead of its positioning on the xylan backbone (de Souza et al., 2018). In Arabidopsis, a GT61 family member known as MUCI21 was suggested to be responsible for the addition of xylose to the xylan backbone in mucilage cell walls (Voiniciuc et al., 2015). In rice, a XAX GT61 glycosyltransferase was proposed to add xylosyl side chains onto xylan (Chiniquy et al., 2012).

The GT8 family which *GUX* genes belong is represented by five members (*GUX1* to *GUX5*) in *A. thaliana,* and *in vitro* experiments have confirmed that *A. thaliana GUX1*, *GUX2* and *GUX4* are active glucuronosyltransferases, being able to transfer GlcA from UDP-GlcA onto a xylohexaose acceptor (Rennie et al., 2012). Furthermore, an *in vivo* activity of *At*GUX3 has been reported in primary cell wall and it is responsible for its glucuronidation (Mortimer et al., 2015).

Glucuronic acid (GlcA) is found on all plant xylan molecules (Rennie & Scheller, 2014). GlucUronic acid substitution of Xylan (GUX) enzymes are GT8 family members responsible for the addition of GlcA onto the xylan backbone. In *A. thaliana*, two GUX enzymes were identified and characterized (AtGUX1 and AtGUX2) and demonstrated to be involved in the xylan synthesis process (Mortimer et al., 2010). The *gux1/2* double mutants showed no detectable [Me]GlcA branches on secondary cell wall xylan (Mortimer et al., 2010), while *gux1/2/3* triple mutant showed a complete loss of GlcA and MeGlcA side chains in primary and secondary cell walls (Lee et al., 2012). Complete loss of GlcA in *gux1/2/3* plants does not result in any significant growth phenotype when compared to WT plants (Mortimer et al., 2015; Lyczakowski et al., 2017).

## 3. DIFFERENCES BETWEEN PLANT CELL WALLS

Monocots and dicots are composed mainly of four polymers: cellulose, hemicellulose, lignin and pectic polysaccharides. The principal difference between the walls of monocots and dicots is in the amount and nature of those polymers. As a result of evolution process, cell walls are naturally resistant to deconstruction by chemical or biological reactions and this property is generally designed to as recalcitrance (Pattathil et al., 2015). One of the major hurdles to cost-effective cellulosic biofuel production is the recalcitrance of cell wall (Himmel et al., 2007; Chundawat et al., 2011a).

Some studies demonstrated by chemical analyses that most species groups of economic interest for sustainable biofuel production have great differences in their composition (Pauly and Keegstra, 2008; Vogel, 2008). For example, the walls of dicots, non-commelinoid monocots and gymnosperms in comparison with commelinoid monocots (e.g. grasses) have a great difference in the relative amount of non-cellulosic components and their interaction between themselves (Carpita and Gibeaut, 1993; Carpita, 1996; Vogel, 2008).

In grasses, primary walls are characterized by the great amount of hemicellulosic polysaccharides, glucuronoarabinoxylans (GAX) and mixed-linkage glucans (MLG), with small amounts of xyloglucans (Xyg), pectic polysaccharides and glycoproteins. On the other hand, primary walls of dicot and gymnosperm are mainly characterized by a greater number of xyloglucans and then by mannans and glucomannans as the principal hemicelluloses and also have significantly more pectic polysaccharides and glycoproteins (Pattathil et al., 2015; Vogel, 2008).

There is a reasonable difference in structure and composition from primary and secondary walls, the latter generally has proportionally more cellulose and less pectic polysaccharides, with mainly non-cellulosic polysaccharides in grasses secondary walls, while woody dicots have more xylan and lignin, with glucomannans that are also available in smaller amounts in woody dicots (Vogel, 2008). In woody gymnosperms lignin and hemicelluloses are the most present components of secondary wall as in dicots, but galactomannan and glucomannan are relatively more abundant than xylan (Pauly and Keegstra, 2008)

In general, the amount of lignin is also variable among different plants. In both woody dicot and monocot grasses, lignin content is mainly from G and S units with traces of H units. However, the grasses generally contain higher amounts of H units. Differently, lignin from

woody gymnosperms is mainly composed of G units with smaller amounts of H units (Pattathil et al., 2015).

## 4. CELL WALL AND ENZYMATIC SACCHARIFICATION

The plant cell wall is mainly composed of cellulose that is formed by the association of β-1, 4-linked glucosides. The hydrolysis of cellulose produces glucose, which can be used in the production of 2G ethanol. As plants grow, cells in several organs expand and become substantially rigid and are reinforced by the deposition of the secondary cell walls. For this reason, plants at the maturation stage have high strength and resilience. Although advantageous for the plant to protect against pests and diseases, these changes hinder the release of sugars that are trapped in the cell (Wingren et al., 2003). Along plant development several biochemical modifications controlling the cell wall loosening take place, allowing adjustments in the mechanical and growth properties of cell walls from different tissues (Cosgrove, 2005). Some of these natural changes propitiate hydrophilic regions on the cell wall, allowing sugar access, with clear implications for the saccharification process aiming the production of 2G ethanol.

Saccharification is the entire degradation of a complex carbohydrate (e.g. cellulose or starch) into simpler sugars components such as glucose, maltose, xylose, maltotriose and sometimes dextrin (Kudus, 2019). In the saccharification process (hydrolysis of plant biomass), usually a pretreatment with a strong base, in the case of grasses, is used. This makes the process more expensive and requires several steps to recover the alkaline reagents (neutralisation with acids) and to provide conditions that will allow yeasts to ferment the released sugar (Gomez et al., 2008). After the chemical treatment, cocktails containing hydrolytic enzymes are used to release the sugar that is still trapped inside the lignocellulose complexes. However, enzymes are only able to perform the hydrolysis in a solution containing a maximum of 8% of bagasse, due to the presence of inhibitory substance from various cellular processes, reducing the enzymatic activity (Wingren et al., 2003). Therefore, strategies aiming changes the cell wall to facilitate saccharification process are key to overcome the intrinsic recalcitrance to hydrolysis, achieving a highly efficient and viable production of 2G ethanol (De Souza et al., 2013).

## 5. SUGARCANE GENOME COMPLEXITY

Sugarcane is a monocot species from the Poacea family (Gramineae), belonging to the genus *Saccharum*. The crossing of *Saccharum* species produced hybrids that are considered the pillars of modern sugarcane cultivars. The hybrids resulting from crosses between *S. officinarum* and *S. spontaneum* are considered the parents of all varieties of sugarcane currently (Irvine, 1999). While *S. officinarum* can accumulate sucrose in the stem, *S. spontaneum* has higher vigor and is resistant to diseases and different stresses (Gianotto et al., 2011). Due to the result of many interspecific hybridizations that occurred during the last century between those two high polyploidy species, the modern sugarcane cultivars have an extremely complex genome, maybe the most complex across all the other crops (Garsmeur et al., 2018).

Breeding programs allowed an increase in production during the last 30 years and there are high expectations that the use of biotechnology will leverage this progress (Gianotto et al., 2011). However, the lack of knowledge about genetics from sugarcane has delayed further studies in comparison with other major crops (Gouy et al., 2013). As a way to illustrate, the sorghum genome is diploid, which means it has only two copies of each genome, while modern sugarcane cultivars comprise between 10 to 12 copies of each chromosome (Figure 7) (Garsmeur et al., 2018). Those cultivars are highly heterozygous and have around 100 to 130 chromosomes, mainly derived from *S. officinarum*, around 10 to 20% from *S. spontaneum* and another approximately 10% from interspecific recombinants (D´Hont et al., 1996; Piperidis et al., 2010).

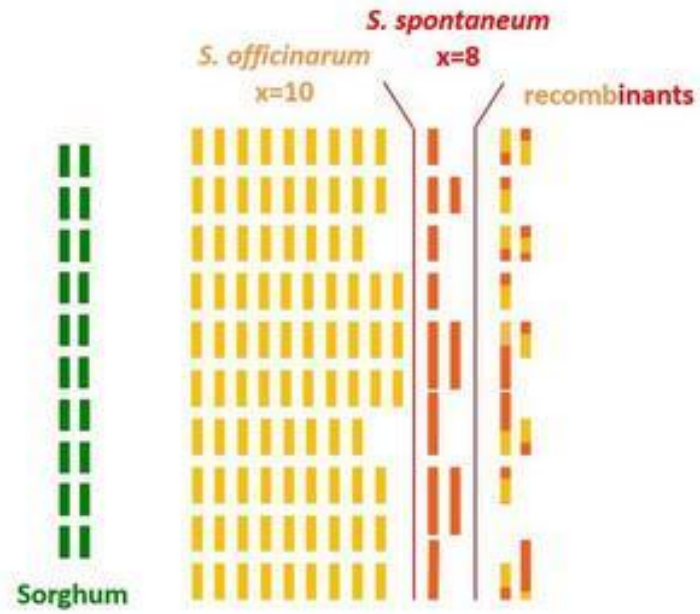



**Figure 7 Schematic representation of the genome of a typical modern sugarcane cultivar in comparison to sorghum genome.** Each bar represents a chromosome, in orange or red when originating from *S. officinarum* or *S. spontaneum*, respectively. In green is represented the sorghum genome, the closest genome to sugarcane from a diploid plant. Chromosomes aligned on the same row are hom(oe)ologous chromosomes (HG). The key characteristics of this genome are the high polyploidy, aneuploidy, bispecific origin of the chromosomes, the existence of structural differences between chromosomes of the two origins, and the presence of interspecific chromosome recombinants (Cirad website; Garsmeur et al., 2018)

In the last two decades, several studies involving molecular genetics allowed the development of resources, increasing the understanding of sugarcane genome (Garsmeur et al., 2018). The use of single-dose markers was very informative in a high polyploid organism (Wu et al., 1992) revealing that the mixture of chromosomes in those cultivars was the result from general polysomy (an organism with at least one more chromosome than normal) and in some cases with some preference for pairing (D´Hont et al., 1994; Grivet et al., 1996; Aitken et al., 2005; Aitken et al., 2014) whereas meiosis of modern cultivars principally was involved in the first meiotic division as bivalents (Bremer, 1923; Price, 1963; Burner et al., 1994).

Another important step was the use of comparative mapping involving other Poaceae species, which revealed a large-scale genome-wide collinearity with sorghum, plant that became a model organism for studies with sugarcane (Dufour et al., 1997; Guimaraes et al., 1997; Ming et al., 1998; Le Cunff et al., 2008; Aitken et al., 2014). For the identification of genomic regions related with relevant agronomic traits QTL studies were conducted revealing smaller effects (Aitken et al., 2008; Ming et al., 2001; Hoarau et al., 2002) and some major

resistance genes (Raboin et al., 2006; Piperidis et al., 2008; Al-Janabi et al., 2007). A different approach under development is the use of genomic selection (Gouys et al., 2013) and genome-wide association (GWA) (Debibakas et al., 2014; Gouys et al., 2015). In both cases, the low rate of linkage equilibrium detected in modern sugarcane cultivar is used for the analysis (Janoo et al., 1999; Raboin et al., 2008)

The first sugarcane genomic resources came as Bacterial Artificial Chromosomes (BAC) libraries (Tomkins et al., 1999; Figueira et al., 2012) and large Expressed Sequences Tag (EST) datasets (Vettore et al., 2003), being the latter of greater importance since it provided a database and the largest collection of Sugarcane ESTs (SUCEST) at that time (Vettore et al., 2003). In an attempt to unravel the whole sugarcane genome, many works partially sequenced it (Grativol et al., 2014; de Setta et al., 2014; Okura et al., 2016; Miller et al., 2017). Through the use of Illumina Seq, long reads were generated and a sugarcane draft genome was published (Riaño-Pachón & Matiello, 2017).

Recently, the first monoploid sugarcane genome came out (Sugarcane Genome Hub) containing 382 Mb of sequence with high quality contigs (Gasmeur et al., 2018) and more recently an assembly of 4.26 Gb, holding approximately 374k putative genes and promoter regions predicted, being the biggest step until now to elucidate the whole and very complex polyploidy sugarcane genome and its data was updated into SUCEST database (Souza et al., 2019; Vettore et al., 2003). Both databases were used for all sugarcane genes searches in the experiments in this thesis and the genes were compared between them to have more confidence, avoiding possible future mistakes.

## 6. SUGARCANE AND BIOETHANOL

Plant biomass is an important renewable source of clean energy, without depleting raw materials such as fossil fuels that contribute to global warming. As the transport sector is responsible for a significant fraction of the emission of greenhouse gases, the replacement of fossil fuels with biofuels, such as bioethanol, can reduce significantly the environmental impact and contribute to socio-economic gains (Dias et al., 2012). The first-generation (1G) ethanol is obtained from the fermentation of sucrose or starch-based products using yeast. Currently, both bioethanol and biodiesel are produced from sources that are also used for feeding, which may confer changes in the price suddenly according to the more profitable product: food or fuel (Gomez et al., 2008).

Second-generation (2G) ethanol is obtained from plant biomass, mainly lignocellulosic materials. The conversion of lignocellulosic biomass requires conditioning and/or pretreatment of raw materials for fermenting organisms that convert the sugar into alcohol (Cardona & Sanchez, 2007). Plant biomass that can be used for 2G ethanol is inexpensive, non-edible and comprises one of the biological materials most abundant on the planet, yet underutilized. It is worth mentioning that the competition for areas for food production can be avoided by planting in areas not conducive to food (Haberl et al., 2007; Gomez et al., 2008). The production of biomass-derived biofuels has great potential in countries with limited oil resources, decreasing the dependence on fossil fuels, besides being able to be produced by a wide range of materials (Hoang et al., 2015).

### 6.1. ETHANOL PRODUCTION AROUND THE WORLD

The most frequently crops used to produce ethanol are corn, wheat, sugarbeet and sugarcane (Goldemberg & Guardabassi, 2010). Ethanol is produced mainly in the USA (from maize) and Brazil (from sugarcane). In 2019, the USA produced 59.8 billion liters, Brazil 32.6 billion liters, European Union 5.4 billion liters (mainly from sugarbeet) and China 3.4 billion liters (mainly from maize) with a grand total of 110.1 billion liters of ethanol per year (RFA, 2019). In 2018 the land in use for ethanol production in USA was 154 million hectares and in Brazil 79 million hectares (RFA 2019; IBGE, 2020).

One of the main advantages of producing sugarcane ethanol is the reduction in greenhouse gases (GHG) emissions in comparison with petrol. Sugarcane ethanol generates

80% lower greenhouse gases (GHG) emissions, while maize generates 30-52% and sugar beet 40% lower GHG emissions than petrol (SCOPE, 2015). Unlike other crops that might have the problem of competition between fuel and food, the use of sugarcane has the advantage of allowing the production of both ethanol and sugar with substantial flexibility in the decision on the final product. Moreover, if necessary, the ethanol can be produced from molasses only, a coproduct of sugar production (Souza et al., 2018). The use of lignocellulosic biomass as a substrate is very attractive because of its affordability and abundancy as agricultural residues (Katahira et al., 2017).

## 6.2. SECOND GENERATION ETHANOL INDUSTRY

Brazil has two second-generation (2G) facilities operating commercially, while the US has one and Norway another. In addition to that, there are other 2G facilities under construction or planning stage: Romania, US, China, Finland, and Slovakia (Padella et al., 2019). Besides that, many reports with different aims on biological, chemical, physical and physicochemical pretreatment methods have been described emphasizing alkali, diluted acid, and steam pretreatment as a way to increase biofuel production (Sarker et al., 2017; Meghana & Shastri, 2020).

The bioethanol industry has an important role in the current economy, promoting energy sustainability, diversifying energy matrices, and reducing GHG emissions. In Brazil, renewable energy represented 45.3% of the total energy supply (versus a worldwide representative of 14.3%) in 2018. Ethanol and sugarcane bagasse represented 38.4% of Brazilian energy supply among renewable energy (Brazilian Ministry of Mines and Energy, 2019). Sugarcane bagasse is a solid byproduct that remains after sugarcane juice extraction at the amount of 280 kg/ton of sugarcane (Molina et al., 1995). The energy content of bagasse corresponds from 30% to 40% of the plant's total energy (Braunbeck & Cortez, 2005), and is commonly used for electricity and heat production. Regardless of its potential as a raw material for 2G ethanol production, the main hurdle is the cost of the pretreatments to reach commercial scales (Rosales-Calderon & Arantes, 2019). Sugarcane bagasse hydrolysates are mainly formed by a considerable amount of xylan followed by minor amounts of glucan and arabinan. Besides that, it has small amounts of inhibitors such as acetic acid, furfural, formic acid, and 5-hydroximethylfurfural (Li et al., 2017). The most prominent hemicellulosic component from

sugarcane bagasse is xylan, accounting from 73,1% to 82,6% of the total hemicellulose (Carvalho et al., 2015), while glucose, arabinose, and glucuronic acid were present in minor amount. One ton of glucan yields 1.11 tons of six-carbon sugars, which in theory, could be fermented into 651 L of ethanol, while one ton of xylan or arabinan yields 1.14 tons of five-carbon sugars, which in theory could be fermented into 666 L of ethanol. With the use of microorganisms, all lignocellulose-derived sugars can be fermented into bioethanol (Balat, 2011). For 2G ethanol production, the higher amount of fermentable sugars released the higher fermentation yield.

## 7. CHALLENGES OF THE SUGARCANE CELL WALL

In angiosperms, the primary cell walls are divided into two main groups: type I walls and type II walls, classified according to their wall architecture, chemical structur of their components and the way they are synthesised (Carpita, 1996, Yokoyama & Nishitani, 2004). Dicot plants and non-commelinoid monocot plants have type I cell walls, which are composed by a structure of cellulose-xyloglucan with roughly the same amount of cellulose microfibrils and xyloglucans (Buckeridge et al., 2004). While commelinoid monocots (e.g. rice, maize, sugarcane) have type II walls, which differ from type I walls, presenting more cellulose than xyloglucan. The main glycans that cross-link the cellulose microfibrils in cereals are glucuronoarabinoxylan (GAX) (Nishitani and Nevins 1991, Carpita and Gibeaut 1993) and (1→3),(1→4)-β-D-glucan that are exclusive from poales (Buckeridge et al., 2004). When comparing the amount of pectin between type I and type II cell walls, the type II possesses less pectin, but higher quantity of phenylpropanoids (Iiyama et al. 1990). In type II cell walls, the most abundant hydroxycinnamate compound found is ferulic acid, which is esterified to the fifth carbon of arabinosyl side chains of arabinoxylans (Yokoyama & Nishitani, 2004).

Sugarcane biomass can be divided into four main fractions: fiber (organic solid fraction), soluble solids (waxes, sucrose and other chemicals), non-soluble solids (inorganic substances) and water (Canilha et al., 2012; Shi et al., 2013). The main aim of second generation of biofuels is the use of the fiber fraction, more specifically the constituents of the plant cell wall to generate biofuel (Schubert, 2006; Henry, 2010). There are several approaches to optimize the composition of the biomass source for biofuel production, like advances in

pre-treatment methods or by biotechnologically engineering cell wall synthesis pathways, creating a biomass that can be more easily processed (Sims et al., 2006; Simpson, 2009; Viikari et al., 2012; Buckeridge et al., 2016).

The fiber fraction of sugarcane is composed by three main components: cellulose, hemicellulose and lignin. Cellulose constitutes around 50% of the dry weight sugarcane bagasse while hemicellulose and lignin each account for about 25% (Loureiro et al., 2011). Sub fractionation of the cell wall in large classes of components is the method most used in to quantify the amount of their components, as was the case of Masarin et al. (2011), who observed that the cell wall components have a range of glucan (38% to 43%), hemicellulose (25% to 32%), lignin (17% to 24%) and extractive content (2% to 7%). A recent study characterised the composition of sugarcane bagasse among many laboratories in Brazil and showed an average of cell wall components of 42.3% glucan, 22.3% xylan, 21.3% lignin, 6.7% total extractives, 3.5% acetyl, 2.2% arabinan, 1.5% whole ash and 0.6% galactan (Sluiter et al., 2016).

De Souza et al. (2013) showed that the structural features of sugarcane cell walls largely agree with what is known from other grasses. In the case of leaves and stems, most of the pectins and β-glucans are accessible to enzymatic treatments, while hemicelluloses (arabinoxylan and xyloglucan) require a pretreatment for removal of the acetyl and/or feruloil-esters radicals to turn these polysaccharides susceptible to hydrolysis by exo- and endo-enzymes (De Souza et al., 2013). Therefore, decreasing the number of decorations present in the hemicelluloses may facilitate saccharification and even reduce the reagents for the pretreatment.

There are some factors that contribute to hinder the access of enzymes to the secondary wall. First, the lignification of primary cell wall is the main factor of cell wall recalcitrance entangling the disruption of forage tissues (Wilson & Hatfield, 1997). Second, in some type of cell like the sclerenchyma fibers and the bundle sheath around vascular bundles, the lumen is mainly composed by the secondary wall thickening. Moreover, in grasses, specifically in vegetative tissues all cell types have primary and secondary walls lignification and thickening (Wilson & Hatfield, 1997). An extra barrier to cell wall digestibility is the presence of layers of suberin in primary cell walls of C4 subtropical grass crops (Wilson &

Hattersley, 1983), such as sugarcane, where suberization is commonly found over cell walls in the stems (Figueiredo et al., 2019).

Another agent of cell wall recalcitrance are the polysaccharides biopolymers themselves, especially hemicelluloses (Figueiredo et al., 2019). Hemicelluloses provide support for cellulose microfibrils, making the cell wall more robust and stronger (de Souza et al., 2013). Sugarcane hemicellulose is composed mainly of arabinoxylan, xyloglucan, mixed-linkage glucans and mannan traces (de Souza et al., 2013). These polysaccharides are complexed in a branched pattern and may interact with each other as well with cellulose domain. Xylan is composed of a β-1,4 linear chain linked to xylosil residues that are usually substituted by sugar chains, such as GlcA and MeGlcA (Buckeridge et al., 2016). The complexity of oligosaccharide branch arrangement of hemicelluloses and the proportion of methylayion, acetylation and feruloylation limit the hydrolysis of the dominant chain by cellulolytic enzymes. This limitation is even bigger at regions of linkage between polysaccharides or with other cell wall biopolymers (Pauly et al., 2013). Furthermore, during plant growth the concentration of phenolics in cell wall is increased, which is responsible for the cross-linking of suberin and lignin biopolymers with the polysaccharides, making the structure of the cell wall even stronger and decreasing it digestibility (Jung & Allen, 1995).

Some side groups with carboxylic groups, such as ferulic acid and p-coumaric acid, establish an ester bond with arabinoxylans hemicelluloses, with suberin fatty acids and also with lignin monolignols (Graça et al., 2015). Ferulic acid can still form hydroxycinnamates ester dimers, or oligomers of higher complexity, reinforcing the cross-link inside the carbohydrate fraction of the cell wall, as well as within the suberin lamella. In addition to that, hydroxycinnamic acids form covalent ether bonds with lignin monolignols, increasing the recalcitrance all over the cell wall (de Oliveira et al., 2015; Ralph, 2010).

The cell wall of taller grasses with thicker stems, such as corn, sorghum and sugarcane, the more abundant hydroxycinnamic acid is the more *p*-coumaric acid is inside (Costa et al., 2013; Xu et al., 2005). The accumulation of *p*-coumaric acid in the cell wall from grasses has been shown to be correlated with lignin deposition and higher cell wall recalcitrance (Lygin et al., 2011; Zhang et al., 2011).

One of the best ways to go over all the hurdles mentioned before are through genetic approaches involving genetic enhancement, plant breeding and molecular biology, with the goal to improve biomass sources by having crops with less or modified lignin, crops that self-produce enzymes or crops with increased cellulose and biomass overall (Sticklen, 2006). Nowadays one of the biggest bottlenecks for second generation biofuel industry is the cost of the enzymatic pretreatment of cellulosic biomass (approximately 25% of total processing expenses), which, together with the costs of biomass conversion and microbial tanks, decreases the competitiveness of 2GE biofuels when comparing to fossil fuel (Gnansounou & Dauriat, 2010; Macrelli et al., 2014; Van Der Weijde et al., 2013). This highlights the importance of genetic improvement of biomass composition to reduce processing costs.

## 8. BIOTECHNOLOGY AND THE OPTIMIZATION OF CELL WALL TO INCREASE SACCHARIFICATION

Several works explore the extensive network of genes responsible for the composition and architecture of the cell wall. The fundamental knowledge is largely described in model plants such as *Arabidopsis thaliana* and *Brachypodium distachyon*, and the identification and characterization of genes aiming a commercial use is a challenge. A translational approach, taking advantage of this knowledge from model plants, is vital to sugarcane studies.

Many studies correlate changes in lignin amount and composition with an increase in biomass processing, with these modifications frequently resulting in a biomass yield penalty (Chanoca et al., 2019; Wadenbäck et al., 2008; Wagner et al., 2009; Voelker et al., 2010; Stout et al., 2014)**.** The gene Cinnamoyl-Coa Reductase (*CCR1*) expresses an enzyme responsible for the first lignin-specific biosynthesis and when knocked-out, a phenotype with severe growth problems and significant reduction in total lignin content in different species of plants is observed (Jones et al., 2001; Leplé et al., 2007; Derikvand et al., 2008; Tamasloukht et al., 2011). A recent study showed that maize *ccr1* mutants had an increase in the amount of monolignol ferulate conjugates (ML-FA), and it might be related to an increase in saccharification due to a higher digestibility of the cell wall (Smith et al., 2017). Because the vast majority of works in the area resulted in problems in plant development, we preferred to focus on a gene related to another component of plant cell wall, with higher agronomical and commercial potential. So, we decided to prospect works with some gene related to

hemicellulose that, when silenced, produced an increase in the saccharification yield, but with no harm to the plant.

The *SAC1* gene from *Brachypodium* encodes a glycosyltransferase of the GT61 family, being responsible for replacing arabinoxylan in the cell wall. Null mutants of this gene do not make this substitution, and their cell wall is 167% more susceptible to saccharification in relation to wild type (WT) plants. Additionally, the mutants show no reduction in resistance to physical stress and stiffness of the stem (Marriott et al., 2014). Mutants of the *XAX1* gene from rice, which also encodes a glycosyltransferase GT61, showed an increase of 62% in the saccharification compared to WT (Chiniquy et al., 2012).

In *Arabidopsis GUX1* and *GUX2* encode a glycosyltransferase of GT8 family, and null mutants had a reduction in xylan decoration pattern of approximately 60% and 30%, respectively (Mortimer et al., 2010). These mutants had increased saccharification, without significantly altering the physiological characteristics of the plant. In the case of the double mutant *gux1/2* there was no substitution, with 100% xylan availability (Mortimer et al., 2010). Recently, an *in vitro* analysis of the Arabidopsis *gux1/2* mutant, 30% more glucose and 700% more xylose were release during saccharification, compared to WT plants. These changes were due to the reduced glucuronosyl substitutions of xylan, since the lack of branches in xylan reduced the recalcitrance to the enzymatic treatment. The ethanol yields obtained through enzymatic saccharification and fermentation were two times higher than in WT plants (Lyczakowski et al., 2017).

## 9. SUGARCANE GENES SELECTION

Thinking about the high complexity of cell wall and its association in the plant development in addition to the data obtained from previous works with model organisms, we decided to use other criterion to select the gene to this work. Another relevant point at the time that this gene was being selected was the number of references and information about the gene as well as the availability of its orthologues sequences in sugarcane databases.

According to data from Sun et al. (2004), sugarcane bagasse is composed of 43.6% cellulose, 33.5% hemicellulose and lignin 18.1%. This information fostered the relevance of genes related to hemicellulose metabolism, which could contribute to the knowledge in the

area focusing on the increase of productivity of second-generation ethanol without any penalties to sugarcane development. After a careful analysis of all the works cited above and rationalising about all the return that this work could contribute in the future, the genes *GUX1* and *GUX2* were chosen to be further studied.

The *GUX* genes are glycosyltransferases localized in Golgi apparatus that belong to the GT8 family and are responsible for adding GlcA into the xylan, which confers high recalcitrance to the plant cell wall that makes more difficult to release sugars during the saccharification, hindering the production of biofuel from biomass (Mortimer et al., 2010; Lee et al., 2012).

A model was built to understand the interactions between cellulose and xylan, where the xylan hydrogen bonds with the cellulose microfibril hydrophilic surfaces, forming a flattened two-fold helical screw. It is believed that this fold blocks the accessibility of cellulases to cellulose, which is used to produce fermentable sugars (Simmons et al., 2016). Additionally, the acetyl radicals released from xylan during the pretreatment of plant biomass can inhibit the microorganisms used to ferment sugars (Lee et al., 2012; Himmel et al., 2007, Helle et al., 2003, Simmons et al., 2016).

The *GUX* genes are responsible for the production of glycosyltransferases, which add GlcA in the side chains, hindering the access of cellulases to biomass. However, when silenced, the efficiency of the biomass fermentation process may increase significantly without any interference in plant growth (Mortimer et al., 2010; Mortimer et al., 2015; Lyczakowski et al., 2017).

# CHAPTER 2: MATERIALS AND METHODS

## 1. PHYLOGENETICS ANALYSIS

In order to reconstruct the phylogenetic relationship of the GUX family, we selected 16 angiosperm species (including six monocots and ten dicots) that are either model plants or important crops to serve as outgroups in the phylogenetics analysis. The phylogenetic analysis was done according to Gallinari et al. (2020) and served as the basis for all the study done through this work. We were able to identify the orthologues of Arabidopsis *GUX* genes in different species and as our main focus here is sugarcane, we identified *GUX* orthologues for this species. All the sequences were aligned with MAFFT (Katoh and Standley, 2013) using the iterative refinement method L-INS-I and no treatment were done in the aligned sequences. Maximum likelihood phylogenetic analysis of the *GUX* multiple sequence alignment was performed using IQ-Tree v1.6.1 (Trifinopoulos et al., 2016). Branch support was acquired by 1,000 ultrafast bootstraps pseudoreplicates (Minh et al., 2013), under JTT+I+G4 model identified by ModelFinder (Kalyaanamoorthy et al., 2017). For the Bayesian phylogenetic analysis, we used MrBayes v3.2.6 (Ronquist and Huelsenbeck, 2003), using 1,000,000 generations, sample frequency of 500 and diagnostic frequency of 5,000, under JTT+I+G model of evolution. Phylogenetic trees were visualized and edited in FigTree v1.4.3.

## 2. ARABIDOPSIS PLANTS

A*rabidopsis thaliana* seeds from Columbia-0 ecotype (col-0) were sowed in ½ MS medium (2.2g/L MS salts including vitamins, 1% sucrose, 0.5g/L MES and 0.8% agar with pH adjusted to 5.8 using KOH) and vernalized at 4 °C for three days to break dormancy. Then the plantlets were transferred to soil with substrate and vermiculite (ratio 2:1) in a cabinet maintained at 21°C, with a 16-h light, 8-h dark photoperiod. Mutant insertion lines without any [Me]GlcA sugar decorations for *gux1/*2 and Wild Type (WT) were used for complementation and xylan analyses experiments. Alcohol insoluble residue (AIR) was prepared from 5-cm-long sections of mature *Arabidopsis thaliana* stem. All AIR preparation was carried out as described in (Mortimer et al., 2010).

For sugarcane analysis, mature culms (10 and 11) from 10 months old plants grown in a greenhouse were used. The culms were snap frozen and milled in a horizontal ball mill (TECNAL, TE-8100-FZ). Then alcohol insoluble residue (AIR) was prepared from the ball milled material. All AIR preparation was carried out as described (Mortimer et al., 2010).

## 2.1. Preparation of Alcohol Insoluble Residues (AIR)

For AIR preparation, stems from plants (25 – 50 mg wet weight) were homogenised in ethanol absolute by ball milling with steel balls in a Teflon vessel at 15 Hz for 5 minutes. The milling cycle was repeated four times with 5-minute intervals between each milling step. Homogenised material was washed with ethanol absolute and incubated over-night in a 2:3 (v:v) mixture of methanol and chloroform. On the next day, another methanol: chloroform wash was repeated for 1h in a fresh solution. Thereafter, the sample underwent consecutive wash steps in ethanol absolute, 60% ethanol (v/v), 80% ethanol (v/v) and ethanol absolute. Following the final wash, a pellet of alcohol insoluble residues (AIR) was dried in an oven set at 50°C for 48 hours. The same protocol was used for both Arabidopsis and sugarcane stems.

## 2.2. POLYSACCHARIDE ANALYSIS BY CARBOHYDRATE GEL ELECTROPHORESIS (PACE)

For the PACE analysis, a xylanase treatment was performed on AIR material, releasing oligosaccharides and dried *in vacuo* and derivatised with 8-aminonapthalene-1,3,6-trisulphonic acid (ANTS; Invitrogen). The derivatisation process was performed by reductive amination of oligosaccharides released from 0.5 mg of AIR in acidified DMSO (0.3 µL glacial acetic acid, 3 µL water and 3.3 µL DMSO) in the presence of ANTS (3.3 µL of 0.2 M solution in DMSO) at 37°C overnight. Following ANTS derivatisation the solution was dried *in vacuo* and re-suspended in 100 µL of 6 M urea.

PACE gel casting and running was performed following a customised protocol. A single gel mix was poured into a pre-assembled glass gel mould. The gel mix was composed from water (20 mL), 1M Tris-Borate buffer pH = 8.2 (3 mL), 40% acrylamide solution (29:1 feed ratio, 8.4 mL), 10% ammonium persulphate (0.18 mL) and 1,2-Bis(dimethylamino)ethane (TEMED, 0.04 mL). Following casting the gel was cross-linked at 4°C over-night prior to the run. Gel running was performed in 0.1 M Tris-Borate pH = 8.2 buffer at 1000 V for 45 minutes. Sample amount was customised to allow for adequate band intensity and varied between 1 and 5 µL

of urea re-suspended ANTS labelled oligosaccharides. Gel visualization was performed using a G: Box UV transluminator (Syngene). Band intensities were quantified using a volume integration script built into ImageJ software (National Institute of Health, USA).

## *3.* MOLECULAR CLONING AND GENERATION OF TRANSGENIC LINES OF *ARABIDOPSIS THALIANA* AND *NICOTIANA BENTHAMIANA*

### 3.1. GOLDEN GATE VECTORS PREPARATION

For all Arabidopsis and tobacco expression experiments GoldenGate Moclo technology was used to assemble binary vectors. In this technique L2 binary vectors are assembled from L1 transcriptional units which are made from L0 parts such as promoters, coding sequences, tags and terminators (Weber et al., 2011; Patron et al., 2015). GoldenGate assembly relies on the use of Type IIS restriction enzymes together with T4 DNA ligase to digest and ligate L0 modules (or L1 transcriptional units) in one pot reaction. The order of assembled units is guided by the fact that specific linkers, designed during part synthesis, are released upon BsaI (L0 modules) or BpiI (L1 transcriptional units) digestion. These form sticky ends which ligate with the backbone donor provided.

Specifically, 100 ng of each L0 module need (or L1 transcriptional unit) was mixed together with 100 ng of backbone donor, 1 µL of BsaI (or BpiI enzyme9NEB), 1µL of T4 DNA ligase (NEB), 0.15 µL BSA protein (NEB), 1.5 µL x10 T4 DNA ligase buffer with ATP (NEB) and diluted up to 15 µL in DNAse free water. This mix was incubated in a thermocycler in the following program: 25x cycles at 37**°**C for three minutes and for all the next steps one cycle at 16°C for four minutes, 50°C for five minutes and 80°C for five minutes. After finished the program, 2 µL of the resulting product were used for NEB DH5α *E. coli* transformation. The genes encoding the sugarcane *ScGUX1* and *ScGUX2* enzymes were synthesised by GeneWiz with Golden Gate compatible BsaI sites.

### 3.2. VECTORS CLONING

For multiplication of the plasmids and sub-cloning, 60 µL aliquots of NEB 5α competent *E. coli* cells were used. The transformation was performed by incubating the cells with plasmid DNA (100 – 200 ng) for 30 minutes on ice. This was followed by a heat shock at 42°C carried

out for 30-45 seconds. After 5 minutes of recovery on ice the cells suspension was amended with 300 µL of LB medium (1.5 % agar, 25g/L of LB). After 1 hour of outgrowth at 37°C the suspension was spun at 10000 RCF for 2 minutes and the cell pellet was plated on LB agar amended with specific antibiotic. Table 1 shows all the types of antibiotics used for selection of different GoldenGate assembly components.

**Table 1. Antibiotics used for screening of transformed bacteria with its respective concentration and its use**

| Antibiotics | Concentration | Selection of |
|---|---|---|
| Spectinomycin | 50 ug / mL | L0 modules |
| Ampicilin | 100 ug / mL | L1 transcriptional units and binary partners for AGL-1 |
| Kanamycin | 50 ug / mL | L2 binary vectors and pEAQ |
| Gentamycin | 50 ug / mL | Binary partner for GV3101 |

Transcriptional units made were composed from promoter modules (p), coding sequences, tags and terminators. All binary vectors used for *A. thaliana* transformation were composed from three distinct L1 transcriptional units:

Position 1: p35S – Kanamycin resistance cassette – 35S terminator

Position 2: p*IRX5* – *ScGUX* – 3xMyc – Nos terminator

Position 3: p*Oleosin* – *OLE1* – eGFP – *A. thaliana* Actin2 terminator

Unit used at position two varied according to the ScGUX enzyme over-expressed (*ScGUX1* or *ScGUX2*). Both position one and position three transcriptional units were used to screen for transgenic *Arabidopsis thaliana* plants as described in the next section. The p*IRX5* sequence used in these assembly reactions was the same as one described by Pear et al. (1996).

### 3.3. COMPLEMENTATION EXPERIMENTS EXPRESSING *SCGUX1* AND *SCGUX2* IN *ARABIDOPSIS THALIANA*

The gene encoding the *Saccharum spp* enzyme was synthesised by GeneWiz. Golden Gate MoClo technology was used to assembly binary vectors. All binary vectors used for *A. thaliana* transformation were composed from three L1 transcriptional units: Position 1: p35S-KanR-35S terminator/ Position 2: pIRX5-***ScGUX***-3xMyc-Nos terminator/ Position 3: p*Oleosin*-*OLE1*-eGFP- *A. thaliana* Actin2 terminator. Unit used at position 2 varied according to the *ScGUX* enzyme expressed (*ScGUX1* or *ScGUX2*). Both position one and position three

transcriptional units were used to screen for transgenic *A. thaliana* plants as described in the section below. Representative L2 binary vectors used for *A. thaliana* transformation generated using GoldenGate assembly is presented on Figure 8. Protein expression was driven by a 666 bp promoter sequence of *A. thaliana IRX5* gene (Taylor et al., 2003). *A. thaliana gux1/2* plants were transformed using the floral dip method (Clough et al., 1998). Three independent *gux1/2* lines homozygous for the pIRX5-***ScGUX***-Myc-NosT construct were derived and analysed in this study.

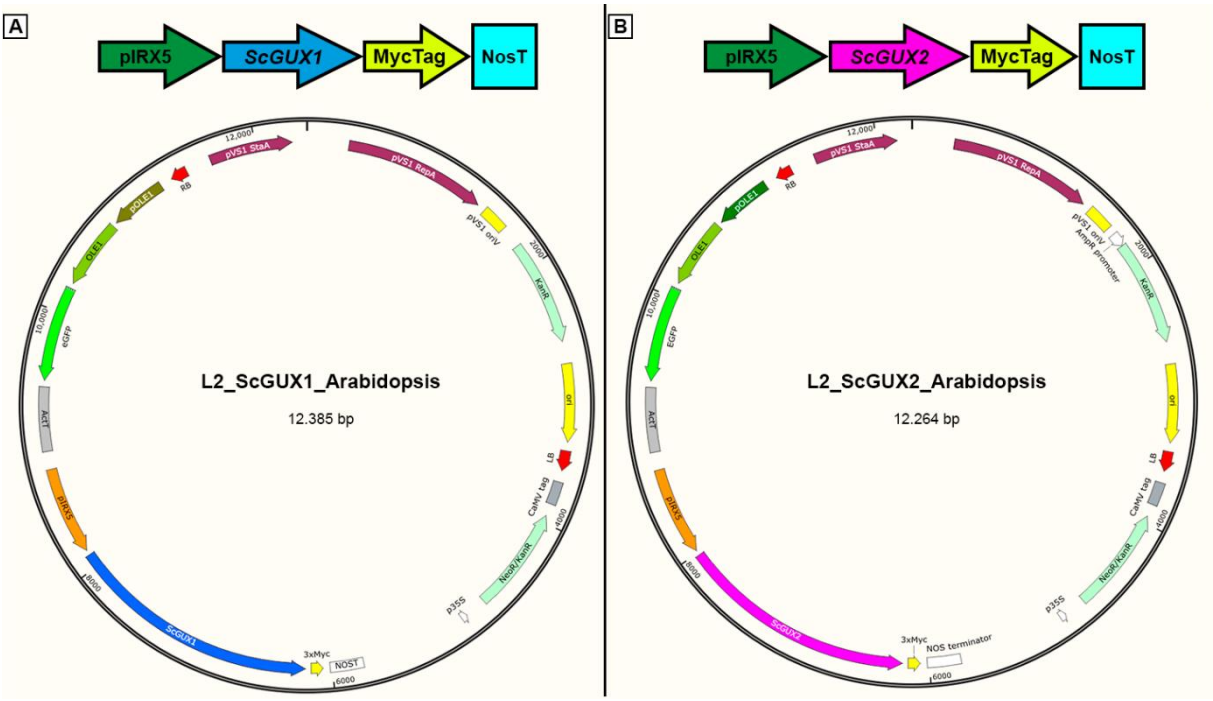



**Figure 8 Vector maps of constructs for expression of *ScGUX1* (A) and *ScGUX2* (B) in Arabidopsis *gux1/2* mutants.**

Protein expression was driven by a 2 kbp promoter sequence of *A. thaliana IRX5* gene. *A. thaliana gux1/2* plants were transformed using the floral dip method (Clough & Bent, 1998). Kanamycin resistant plants were screened for the construct using PCR. Three independent *gux1/2* lines homozygous for the pIRX5-*ScGUX*-Myc-NosT construct were derived and analysed in this study.

### 3.3.1. Tobacco vectors

The synthesised genes by GENEWIZ were used to assembly binary vectors. Both binary vectors were used for tobacco (*Nicotiana benthamiana).* A level 2 assembly was used in order to produce the vector for agroinfiltration through leaves of *N. benthamiana*. The vector L2_***ScGUX***_Tobacco has 2 constructs: Position 1: Kanamicine Plant Resistance, Position 2: Protein expression (*ScGUX1*, Figure 9A or *ScGUX2*, Figure 9B) was driven by a promoter sequence of cauliflower mosaic virus 4x35S gene, with a Myc tag and 35S terminator in a pQA backbone vector.

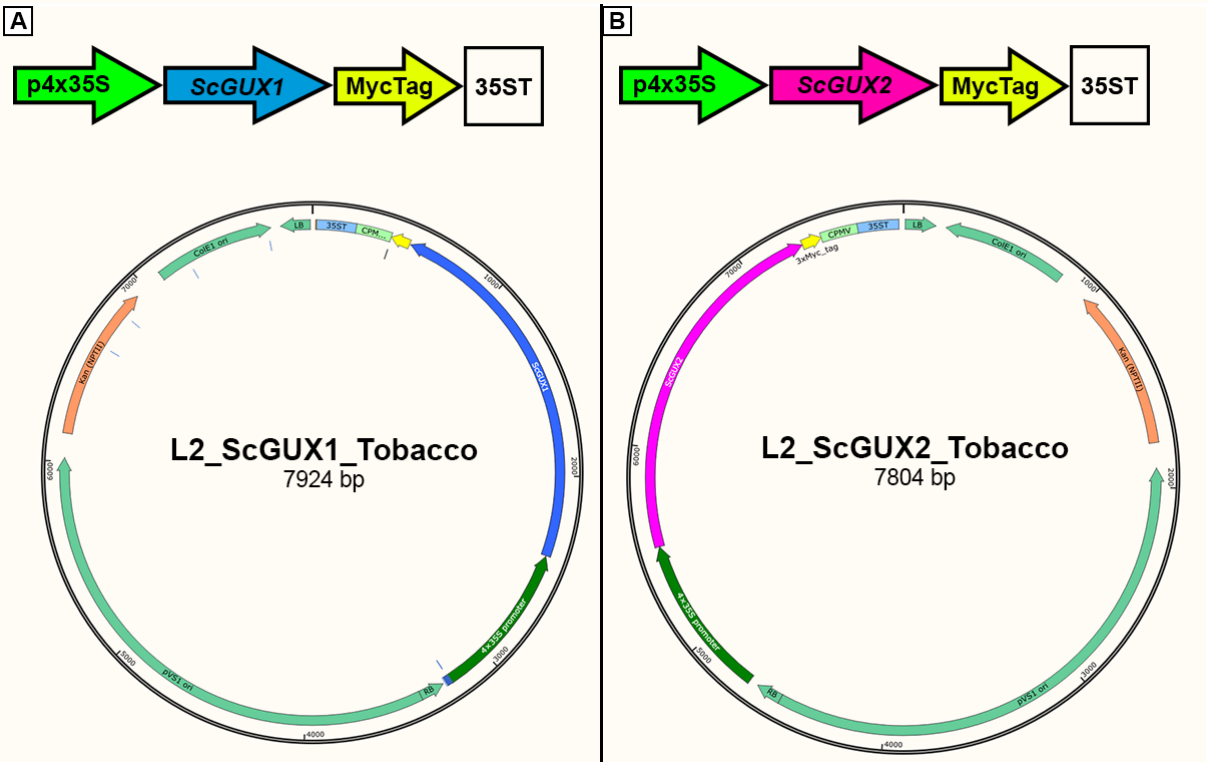



**Figure 9 Vector maps of constructs for expression of *ScGUX1* (B) and *ScGUX2* (C) in tobacco (*N. benthamiana*).**

## 3.4. *Arabidopsis thaliana* transformation via *Agrobacterium tumefaciens*

The *Agrobacterium* strain used for Arabidopsis transformation was GV3101 strain, AGL-1 strain was used for expression in *N. benthamiana* and EHA105 for sugarcane calli transformation. The same protocol was used for preparation of chemically competent cells for all strains. For this, a single colony of *A. tumefaciens* was picked and grown in LB (5mL) overnight at 30°C in the presence of required antibiotic for the maintenance of the binary

partner plasmid. The next day this starter culture was used to inoculate 50 mL LB (with antibiotic). This culture was incubated in a shaker with 800 rpm at 30°C until OD600 reached an OD of approximately 0.6. Then, the culture was chilled on ice, spun at 3000 RCF for 5 minutes at 4°C and the pellet was re-suspended in 1 mL of sterilised LB. This suspension was aliquoted (100 µL / aliquot) and stored at -80°C until needed.

For transformation of *A. tumefaciens* an aliquot of chemically competent cells was thawed on ice. Liquid suspension was amended with 1 µg of plasmid DNA and incubated on ice for 5 minutes. Thereafter, the cells were frozen in liquid nitrogen for 5 minutes and heat shocked in a water bath at 37°C for 5 minutes. This was followed by the addition of 250 µL of LB, transferred to a 15 mL falcon tube and outgrowth at 30°C shaker for 2 hours. Following that, the cell suspension was spun (3000 RCF, 2 minutes) and the entire pellet was re-suspended in a small volume of liquid LB and plated on a solid LB agar amended with antibiotics required for selection of the binary vector and the partner plasmid (see Table 1). Colonies of transformed *Agrobacterium tumefaciens* were detected following 48h of plate incubation at 30°C and grew up to 72h.

For *A. thaliana* transformation the floral dip protocol was followed (Clough and Bent, 1998). An overnight culture (50 mL) of *A. tumefaciens* with the binary vector and the partner plasmid was spun (3000 RCF, 5 minutes) and the pellet was re-suspended in 50 mL of the dipping solution (5% sucrose, 10 mM $MgCl_2$, 0.001% Silwett® L-77). Flowers of 4-week-old *A. thaliana* were dipped in the suspension and following that the plant was bagged for 72 hours. The seeds were collected from dipped *A. thaliana* plants after 4 more weeks of growth.

For selection of transgenic *A. thaliana* seeds, a screening of green fluorescence seeds was used adapted after (Shimada et al., 2010). Specifically, the L1 transcriptional unit encoding OLE1-eGFP protein under the control of p*Oleosin* enables seed specific expression of a stable GFP tagged protein. In consequence, transformed seeds could be differentiated from non-transformed seeds through bright green under the UV light. The green seeds were collected and grown to generate between 15 and 20 T1 plants per construct. Seeds from T1 plants were collected and lines for which 75% of seeds were fluorescent were identified as the mono-insertional ones and used for generation of stable transgenic lines. To do that, T2 plants were grown and their seeds were screened for 100% fluorescence signal which means homozygosity.

### 3.5. DNA EXTRACTION AND GENOTYPING

Additionally, to the GFP seed screening, for each transgenic line, a PCR was done to confirm insertion of the construct. In order to do that, a small circular section of a leaf was placed in 100 µL of TNE-SDS genomic DNA extraction buffer (250 mM NaCl, 25 mM EDTA, 0.5% (w/v) SDS, 200 mM Tris-HCl, pH 7.5). Leaf material was ground in the buffer using a sterile micro pestle and vortexed. The solution was spun at 13000 rpm for 1 minute and the supernatant (300 to 400 µL) was mixed with the same volume of isopropanol in a new tube. This mix was incubated at R.T. for 2 minutes and then spun at 13000 rpm for 5 minutes. The supernatant was discarded and the pellet was dried *in vacuo* using speedvac. Following it, the pellet was resuspended in in 20 µL sterile MilliQ and 1 µL of it was used as a template for PCR. Each genotyping PCR was run using RedTaq PCR Master Mix supplied by Sigma-Aldrich and used according to manufacturer's instructions. To summarise, for each reaction the template was mixed with 12.5 µL RedTaq master mix, 10 µL water and 1.25 µL of 10 µM forward and reverse primer solutions (**see Table 2 for primers used**). Following initial melting step (5 minutes, 94°C), the RedTaq reaction was run for 40 cycles with 94°C melting for 1 minute, 55°C annealing for 2 minutes and 72°C extension for 3 minutes. This was followed by final elongation at 72°C for 4 minutes. Products were analysed on 1% agarose gel prepared with GelRed DNA labelling reagent (Biotium) and visualised on a UV transluminator. PCR product size was estimated by comparison to a DNA standard (GeneRuler 1 kB DNA ladder, Thermo Scientific).

**Table 2 Primers used for sequencing and genotyping**

| Primers | Sequence (5' -> 3') | Primer purpose |
|---|---|---|
| *ScGUX1*_France_Fw | ATGGGTTCCTTGGAGGCCCGGTACC | Sequencing and genotyping |
| *ScGUX2*_France_FW | ATGGGGGTGACCACCGCCGGGGAGG | Sequencing and genotyping |
| Myc_Rv | TCACAGATCTTCCTCAGAGA | Sequencing and genotyping |
| p*IRX5* sense Forward | TCCACCAAATCTTGTTGCTG | Sequencing and genotyping |
| pEAQ backbone | TTCTTGTCGGTGTGGTCTTG | Screening for correct insertion |

# 4. HETEROLOGOUS PROTEIN EXPRESSION AND SUGARCANE GUX ACTIVITY ASSAYS

## 4.1. *NICOTIANA BENTHAMIANA* EXPRESSION USING THE PEAQ-HT SYSTEM

The construct L2_***ScGUX***_Tobacco (*ScGUX1* or *ScGUX2*) was transformed into competent AGL-1 *Agrobacterium tumefaciens* following protocol detailed in section 3.4. Bacterial culture (OD600 between 0.6 and 0.8) was pelleted by centrifugation (3200 RCF for 10 minutes) and resuspended to the same bacterial cell concentration in infiltration medium (0.5% D-glucose, 50 mM MES and 2 mM $Na_3PO_4$, 1 mM acetosyringone), prepared after (Sparkes et al., 2006). This bacterial solution was infiltrated into *N. benthamiana* leaves. Leaves were harvested three days following the infiltration and the membranes fraction enriched for ScGUX were collected as described in Rennier et al. (2012) with some adaptations. In brief, leaves were homogenised in a microsome buffer by grinding with mortar and pestle in a cold room. Homogenate was strained through a nitrocellulose mesh and a cleared solution was pelleted by centrifugation (3200 RCF for 10 minutes). Cleared supernatant was aliquoted into 2 mL tubes and membranes were collected by two rounds of centrifugation (21000 RCF for 1h and 0.5h). Obtained membranes were frozen in liquid nitrogen and stored at -80°C for further use in the *in vitro* activity assays.

## 4.2. WESTERN BLOT DETECTION OF *SCGUX* AND ENRICHED MEMBRANES FRACTION

To determine the protein amount in the membranes, fraction a modified Bradford reagent (Expedeon) was used for quantification. Each well of SDS-PAGE (10–15% gradient, BioRad) was loaded with 5 µg of *ScGUX1*, *ScGUX2* or another protein used as a control for enriched *N. benthamiana* leaf membrane protein. Following the run, the gel was transferred onto nitrocellulose membrane using iBlot system (Life Technologies). The membrane was blocked overnight in 5% milk in TBS solution. The following day it was probed with 1:2000 anti-Myc primary antibody (rabbit polyclonal, Santa-Cruz, A14) and with 1:10000 mouse anti-rabbit HRP linked secondary antibody (Bio-Rad, 170-6515). Amersham ECL prime HRP substrate (GE-Lifesciences) was used to obtain signal from membrane bound antibodies.

## 4.3. EXTRACTION OF ARABIDOPSIS ACETYLATED HETEROXYLAN

AIR (100 mg) from *gux1/2* plants were used as a starting material for extraction of the acetylated heteroxylan lacking [Me]GlcA decorations. In first place, the material was incubated in ammonium oxalate solution (0.5%; 10 mL) at 85°C for 2 h for depectination. Thereafter, the suspension was washed with water and incubated with per acetic acid (11%; 5 mL) for 30 minutes at 85°C, for delignification. To remove Per acetic acid from the suspension it has been washed three times, one with acetone and two with water. Following the washes, the pellet was dried in an oven set at 50°C for 48 h, remaining the holocellulose.

For xylan extraction, 25 mg of holocellulose was used. The pellet was homogenised in 5 mL DMSO and incubated at 80°C for 24 h. This was followed by exchange of the DMSO medium and another 24 h extraction period at 80°C. DMSO fractions, containing acetylated xylan, from both days were combined. Acetylated xylan was washed into water by exchanging the solvent using PD-10 desalting columns (GE Healthcare). Xylan suspension was dried *in vacuo* and stored at room temperature for further use. From 25 mg of holocellulose 25 aliquots of acetylated heteroxylan were obtained following extraction and PD-10 purification.

For estimation of xylan extraction efficiency fractions were collected for the ammonium-oxalate extracted xylan, post-ammonium oxalate extraction residues and the final DMSO extract. Each fraction was exchanged into 0.1 M ammonium acetate buffer using PD-10 desalting columns.

## 4.4. SCGUX1 AND SCGUX2 ACTIVITY ASSAYS

Each reaction mix was prepared as described by Rennie et al. (2012), except for UDP-xylose. To summarise, dried aliquots of acetylated heteroxylan were resuspended in 30 µL of a reaction master-mix (0.5 mM DTT, 10 mM MnCl2, 10 mM $MgCl_2$, 2% Triton-X100, 10 mM UDP-GlcA). UDP-GlcA concentration was changed for some reactions or the sugar nucleotide was replaced with water in certain reactions to control for non-specific glucuronosylation. The reaction mix was amended with 30 µL of undiluted microsomal proteins extracted from *N. benthamiana* leaves. The reaction was performed overnight at room temperature and terminated with a 100°C heat treatment for 10 minutes.

Following termination of the reaction the polysaccharide product was extracted with a methanol: chloroform lipid removal protocol adapted after (Bligh and Dyer, 1959). The

reaction mix was amended with 450 µL of 1:2 (v:v) methanol: chloroform mix and shaken vigorously. This was followed by addition of 150 µL of chloroform and 150 µL of water. Each liquid addition step was followed by vigorous shaking. Obtained mixture was phase-separated by centrifugation at 160 RCF for 10 minutes. Top aqueous layer was collected and amended with ethanol absolute to the final concentration of 80%. To precipitate polysaccharides, this mix was incubated at -20°C for 1 h and spun at 21000 RCF for 15 minutes. This was followed by two wash steps with 1 mL of ethanol absolute. Obtained polysaccharide pellet was dried *in vacuo* and stored for further analysis at -20°C. In most cases, the pellet was deacetylated with 4 M NaOH and digested with arabinofuranosidase GH62 plus *Neocallimastix patriciarum* xylanase GH11 (endo-1,4-β-Xylanase, Megazyme) or BoGH30 (Nzytech).

## 5. SUGARCANE

### 5.1. GENE SELECTION

*ScGUX2* gene was selected as the target for using RNAi, using a selected using an online program (http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::sirna) a specific region was identified presenting with high scores and consequentially higher probability for gene silence and were highlighted in yellow (Figure 10). The RNAi whole sequence is represented in bold. Each specific region was select and checked against SUCEST and Sugarcane Genome Hub database to avoid non-specific silencing and undesirable targets.

**>ScGUX2_CDS**

ATGGGGGTGACCACCGCCGGGGAGGCCGTCTGCAAGTCCCCGGTGCGGGCCTCGGTCATCG-
TCAAGATGAACGCCGCGTTCCTCGCCTTCTTCCTCTTCGCCTACCTGGCGCTCCTCCTCCACCCCAAGTAC
TCGGACATCCTCGACCGCGGCGCCACCTCCCTCGTCCGCTGCACCTTCCGCGACTCCTGTCCGCCACCGTC
GTCGACGACGACCCAGCTCTCACGGAAGCTGGGAGGCGTGGCGGCGAACAAGGTGGCGGCGGCGGAGCGCA
TCGTGAACGCGGGCCGCGCGCCGGCCATGTTCGAGGAGCTCCGTGGCCGGCTGCGGATGGGCCTGGTGAAC
ATCGGGCGCGACGAGGTGCTGGCGCTGGGCGTGGAGGGCGACGCGGTGCGCGTGGACTTCGAGCGCGTCTC
CGAGACGTTCCGGTGGTCGGACCTGTTCCCGGAGTGGATCGACGAGGAGGAGGACGACGAGGGCCCGTCCT
GCCCGGAGCTCCCCATGCCGGACTGGTCCCGGTACGCCGACGGCGGCGGCGACGTGGACGTGGTGGTGGCG
TCGGTGCCGTGCAACCGCAGCGCGCCCGGGTGGAACCGCGACGTGTTCAGGCTGCAGGTGCACCTGGTGGC
GGCGCACGTGGCGGCGCGGAAGGGCCGGCGCGACGGGGGCGGGGCCGTGCGCGTGGTGCTGCGGAGCCAGT
GCGAGCCCATGATGGACCTGTTCCGCTGCGACGAGGCGGTGGGACGGGAGGGGGACTGGTACATGTACAGG
GTCGACGTGCAGCGCCTGGAGGAGAAGCTCCGCCTGCCCGTGGGGTCATGCAATCTCGCTATGCCGCTCTG
GGGAGCAGGAGGGATCCAGGAGGTGTTGAACGCGTCGTCGTGGTCGGACGGGCGTCCCCGGCGCGAGGCGT
ACGCGACGGTGCTGCACTCGTCGGACACGTACCTGTGCGGCGCGATCGTACTGGCGCAGAGCATCCGCCGC
TCGGGGTCCACCCGCGACCTGGTCCTCCTCCACGACCACACGGTGTCGAAGCCGGCGCTCCGCGCGCTGAC
GGCGGCCGGGTGGACGCCGCGCCGGATCAAGCGCATCCGCAACCCTCGCGCGGCGCGGGGCACCTACAACG
AGTACAACTACAGCAAGTTCCGGCT**GTGGCAGCTGTCCGACTACGACCGCGTGGTGTTCGTGGACGCCGAC**
**ATCCTGGTGCTCCGCAACCTGGACGCGCTGTTCGCGTTCCCGCAGCTCACGGCGGTGGGCAACGACGGCTC**
**CCTCTTCAACTCCGGGGTGATGGTGATCGAGCCGTCGGCATGCACGTTCGACGCGCTGATCCGGGACCGCC**
**GGACCATCCGGTCGTACAACGGCGGCGACCAGGGGTTCCTGAACGAGGTGTTCGTGTGGTGGCACCGGCTG**
**CCGCGGCGGGTG**AACTACCTCAAGAACTTCTGGGC**CAACACCACCGGGGAGCGCGCGCTCAAGGAGCGGAT**
**GTTCCGGGCGGACCCGGCCGAGGTGTGGTCCATCCACTACCTGGGGATGAAGCCCTGGACGTGCTACAGGG**
**ACTACGACTGCAACTGGAACGTGGCGGACCAGCGGGTGTACGCCAGCGACGAGGCGCACAAGCGCTGGTGG**
**CAGGTGTACGACCAGATGGGGGAGAC**CATGCGCGGGCCATGCCGCCTCTCGGAGCGGAGGAAGGTGGAGAT
CGCCTGGGACAGGCACGTCGCCGAGGAGATCGGGTACGCGGACCAGCACTGGAAGATCAACATCACGGACC
CAAGGAAATGGGACTGA

**Figure 10. *ScGUX2* coding sequence (CDS).** The highlighted area in yellow represent the area with higher probability for gene silencing, while the whole area highlighted in bold represent the RNAi whole sequence.

## 5.2. GENE CLONING AND GATEWAY VECTOR CONSTRUCTION

Primers were designed to amplify the sense and antisense fragments each gene and the complete CDS from *ScGUX2* (Table 3). Restriction sites were added in each primer for further cloning orientation (in bold): BamHI: ggatcc; EcoRV: gatatc; MluI: acgcgt and KpnI: ggtacc.

**Table 3. Primers used for *ScGUX2* RNAi**

| | | |
|---|---|---|
| *ScGUX2* sense fragment | gux_rnai_sFw | **cgacgcgtc**GTGGCAGCTGTCCGACTAC |
| | gux_rnai_sRv | **ggggtaccc**cGTCTCCCCCATCTGGTCGTA |
| *ScGUX2* antisense fragment | gux_rnai_asFw | **cgggatcc**GTCTCCCCCATCTGGTCGTA |
| | gux_rnai_asRv | **ggccagatatc**GTGGCAGCTGTCCGACTAC |

The selected region of *ScGUX2* gene most likely to form a hairpin was amplified from sugarcane cDNA. Bands obtained by PCR were cut, purified and analyzed in agarose gel. The genes expected size corresponds to bands size observed in the gel. Then the purified PCRs (*ScGUX2* sense and *ScGUX2* antisense) were cloned into a modified pCR8/GW/TOPO with restriction sites compatible to BamHI and EcoRV (for antisense) and MluI with KpnI (for sense) as showed in the Figure 11. The vector modified pCR8::as*ScGUX2*:intron:s*ScGUX2* (pCR8::*ScGUX2*) was digested with restriction enzymes, purified and analysed in agarose gel to confirm the real size expected for each digestion. After the size of bands was confirmed, the vector was sequenced to check if the selected region to form the hairpin for *ScGUX2* had no mutations or deletions.

Then, the entry vector pCR8::*ScGUX2* was recombined with the destination vector pGVG (Guidelli et al., 2018) via Gateway system. The resulting vectors were digested with restriction enzymes. Observed and expected bands presented the same size**.** These results were confirmed by sequencing and the vector with the construct named pGVG::*ScGUX2*_RNAi.

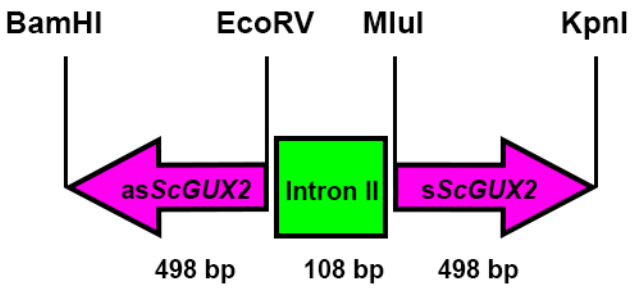



**Figure 11. Scheme of the vector containing both antisense and sense sequence of *ScGUX2* with an intron between them with the four restriction sites used to link all of them together.**

## 5.3. PLANT MATERIAL AND CULTIVATION

The SP80-3280 variety of sugarcane was grown in a greenhouse with drip irrigation for six months. After this, the meristematic region was used to initiate the *in vitro* tissue culture. The top-stalks with three and six months of age passed through an aseptic procedure with 70% alcohol before its entry into the lab (to reduce the risk of contamination). After cleaned

aseptically the leaves were removed to reach the younger leaves, known as palm hearts. These palm hearts were cut into small explants which were passed into the culture medium.

### 5.4. CALLI GROWTH AND INDUCTION

For the growth of the wild type explants, MS medium was used (Murashige and Skoog, 1962) containing sucrose, agar and phytohormones as cytokinin and / or auxin. The explants in contact with the culture medium formed tissues or undifferentiated cells with high potential for germination and formation of new plants, known as embryogenic calli, that were used in transformation.

### 5.5. PREPARATION OF AGROBACTERIUM FOR SUGARCANE CALLI TRANSFORMATION

The Agrobacterium is able to mobilize and selectively transfer the T-DNA to a plant cell. In our case we used competent cells of strain EHA105 containing the binary vector pGVG (Guidelli et al., 2018), containing the maize Ubi-1 promoter controlling a hairpin coding region with part of the *ScGUX2* gene. This vector is based in the NPTII gene for the selection of transformed cells. Two Agrobacterium EHA105 strain culture (pGVG empty vector and pGVG::*ScGUX2*_RNAi) were inoculated from a glycerol stock and streaked in different plates, where it was cultivated overnight. The bacterial cells were then washed and resuspended in a culture medium suitable for inoculation of sugarcane tissue.

### 5.6. SUGARCANE MATERIAL TRANSFORMATION

PangeiaBiotech (Campinas, Brazil) transformed sugarcane calli with the Agrobacterium bearing the vectors described in the previous section and delivered all the transgenic lines and the controls: empty vector (EV) and non-transformed calli that passed through all the tissue culture process (WT) in the sugarcane cultivar SP80-3280. After reaching approximately 10-15 cm, the plants were transferred to culture media containing phytohormones (auxin and gibberellin), which promoted the rooting of the same. They delivered around 15 independent events with 5-8 plants each plus the controls EV and WT. The plants had around 10 cm and were transferred to 500 mL pots with substrate for acclimatization for 1 month **(Figure 12).** After that, they were cultivated in pots of 18 L containing a mix of sand, substrate and clay soil

(1:1:1) until the plants became ready to be harvested. **Figure 13** shows a scheme of the plants from pots until the proper stage for experimental sampling. We cut the plants after seven months and allow them to sprout again in order to reduce heterogeneities in the developmental stage, avoiding variation in the cell wall analyses.

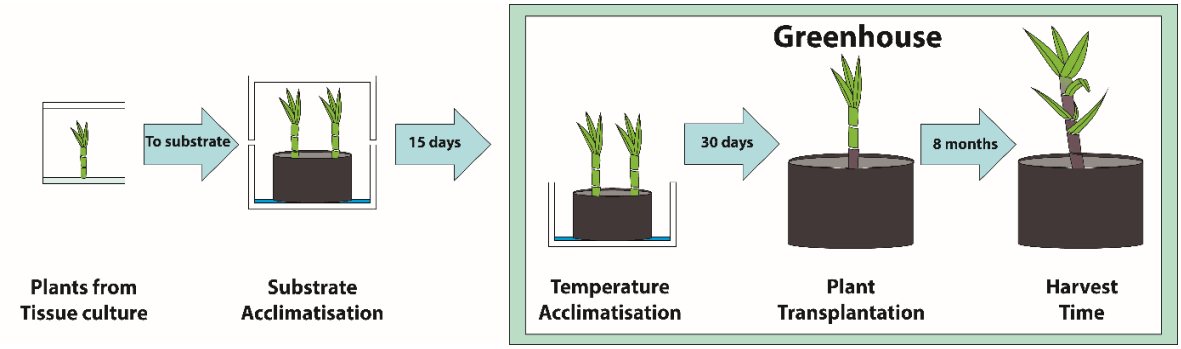



**Figure 12 Scheme of acclimatization of plants from tissue culture to green house with each step until the sugarcanes were ready to be harvested.**

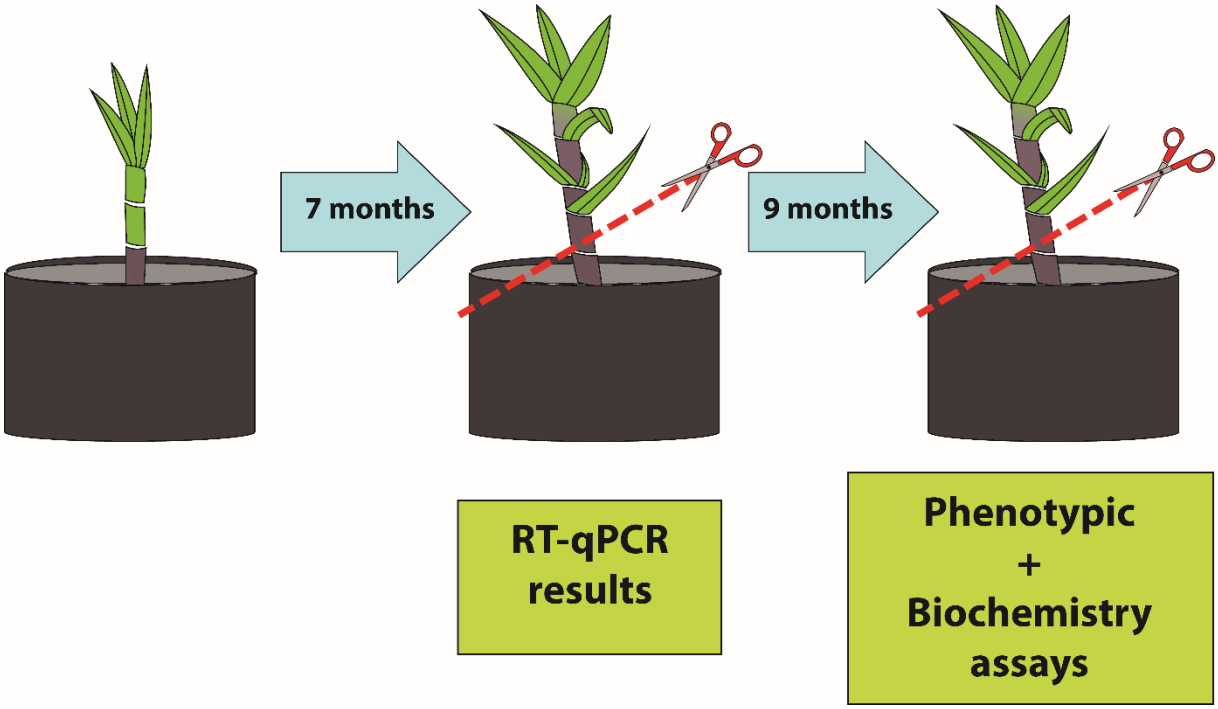



**Figure 13. Scheme representing the time line during the growth (7 months) and regrowth (9 months) period of sugarcane plants.** The plants were harvested in the last stage for cell wall analyses

# 6. SUGARCANE PHENOTYPIC AND BIOCHEMICAL ANALYSIS

## 6.1. PHENOTYPIC ANALYSIS

All the events and controls had the following agronomic parameters: number of culms, mature culm diameter, height, fresh weight and brix according to the methodology employed by Jung et al. (2013).

# 7. BIOCHEMICAL ASSAYS

## 7.1. CELLULOSE, HEMICELLULOSE AND PECTIN CONTENT ANALYSIS

For sugarcane analysis, mature culms (10 and 11) from 10 months old grown in a greenhouse were used (Figure 14). The culms were snap frozen and ball milled in a horizontal ball mill (TECNAL, TE-8100-FZ) ball miller. To determine cellulose, hemicellulose and lignin content the ground material was used.

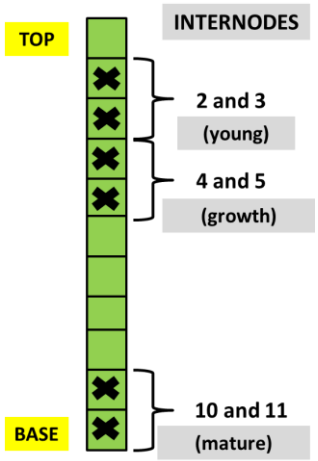



**Figure 14. Scheme of sugarcane and the culms chosen to the analyses.** The "x" represents the internodes of culms collected.

The mature internodes were ground and homogenized inside a solution of liquid nitrogen using a ball mill (TE-350, TECNAL) for 1 minute, repeating three times for each sample. After obtain a thin powder, 30 mg of each sample were homogenised with 1 mL of cold distilled water using a vortex (AP56, Phoenix) and then centrifuged for 14000 rpm for 10 min. After that the supernatant was discarded. The pellet was washed according to the following sequence: 1 mL of cold distilled water, 1 mL of acetone and 1 mL of methanol/

chloroform (1:1). Every wash was followed by a vigorously shaking using a vortex and centrifuged at 14000 rpm for 10 minutes. The supernatant was always discarded after each centrifugation. Then the samples were left to dry on a speedvac at 37°C for 30 min. Following this, to remove the starch, the residue obtained was treated with 1 mL of 2U/mL of Pancreatic Amylase in a 0.1M buffer solution of Sodium acetate pH=6,5 for 37°C for 3 h. Thereafter, a centrifugation at 14000 rpm for 10 min was performed to recover the solid material.

The solid material was washed with 0.6 mL of 20mM of Ammonium Oxalate (pH = 4.0) and left in a dry cabinet at 70°C for 1 hour. After that the solution was centrifuged at 14000 rpm for 10 min. This wash was repeated three times and the three supernatants were combined for pectin analysis.

After centrifuged, the pellet was washed with 1 mL of 0.1M NaOH and left inside a desiccator at RT under vacuum pressure for 24h in the dark. Then the solution was centrifuged and the supernatant (HEM) was saved for hemicellulose analysis. Next, the pellet was washed with 0.4 mL of NaOH 17.5% (w/w) and left inside a desiccator at RT for 8h. Then the solution was centrifuged and the supernatant (HEM) was saved again. Both washes were repeated three times, and at the end all the HEM supernatants were combined for hemicellulose analysis.

The alkaline residue was washed according to the following sequence: 1 mL of distilled water, 1 mL of 1mM of Acetic Acid, 1 mL of absolute ethanol. Every wash was followed by a vigorously shaking using a vortex and centrifuged at 14000 rpm for 10 minutes and the supernatant was always discarded after each centrifugation

Following the washes, the samples were left to dry on a speedvac at 37°C for 30 min. After dried, the samples were dissolved in 1 mL of $H_2SO_4$ 72% for 1 hour at RT. Next the solution was diluted 30 times in distilled water to obtain the corresponding fraction of cellulose. The total sugar content in each fraction was determined according to the sulfuric phenol method using glucose as standard (DuBois et al., 1956).

## 7.2. DETERMINATION OF THE PERCENTAGE OF SACCHARIFICATION

Before starting saccharification analysis, the excess sugar was removed through the following washing: topped up tube with distilled water (repeated for two times), topped up with acetone PA, topped up with methanol/chloroform (1:1). Every wash was followed by a vigorously shaking using a vortex and centrifuged at 14000 rpm for 10 minutes and the supernatant was always discarded after each centrifugation. Then the samples were left to dry on a speedvac at 37°C for 30 min. It is important to do all the washing steps to avoid having problems with saccharification results due to the excess of sugar. The first supernatant after washing with distilled water must have a syrup color and after the second water washing it should be much clearer.

In Eppendorf tubes, an amount of grounded biomass equivalent to 5 mg of cellulose, previously calculated according to total sugar content protocol described in previous section, was weighed. Then, the following reagents were added: 500 µL of 0.1M sodium citrate buffer (pH=4.8), 10 µL of sodium azide and topped up with water for 1000 µL. The solution was heated to 50°C and then, added 6.08 µL of a mixture (1:4, v/v) of cellulase enzymes (1.2 FPU/10 mg of cellulose) and cellobiase (1.26 U pNPGU/ 10 mg of cellulose). With the cocktail ready, prepared from *Aspergillus niger* cellobiose and *Trichoderma reesei* (Sigma-Aldrich), the tubes with the samples and the blanks were tightly closed and placed inclined to 45° in a shaker at 50°C, for 160 rpm during five days.

For the next step all the samples were centrifuged at 14000 rpm for 10 min and DuBois et al. (1956) sulfuric phenol method protocol was performed with some modifications using Glucose as standard. The solution was centrifuged for 14000 rpm for 10 min. Then 6.5 µL of supernatant was added to another Eppendorf tube with 993.5 µl of distilled water, so the sample was diluted 153.8x. From the sample 153.8x diluted, triplicates were prepared. For each triplicate was added 100 µL of diluted sample, 400 µL of $H_2SO_4$ and 100 µL of phenol. The solution was resuspended and put in an ELISA plate, following the order of triplicate as shown in Figure 15. After that the ELISA plate was read in a multi-mode microplate reader (SpectraMax M3) at 490 ηm.

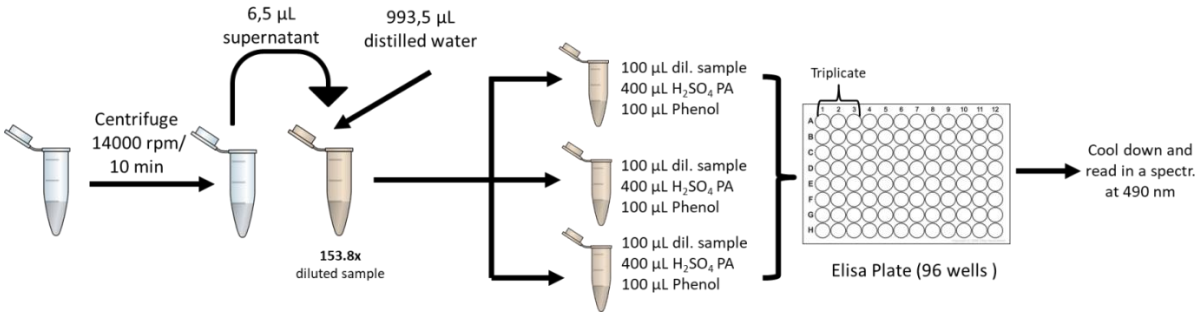



**Figure 15. Scheme of total sugar content analysis.**

## 7.3. ANALYSIS OF SUGARCANE TRANSFORMANTS

Events that survived to the selective agents were confirmed by PCR with primers for *ScGUX2* gene to evaluate gene silencing, and primer for a sugarcane *UBI* gene, (SCCCST2001G02.g) as internal control for normalization. RTqPCR was used to evaluate the level of *ScGUX2* expression in different independent events. The transcription levels of *ScGUX2* were quantified in wild-type plants, in different tissues (young and mature culm and leaves) and used as a guide to select the best transgenic events. The levels of the transcripts from *ScGUX2* were quantified by qPCR. Sugarcane Mature Stem samples were macerated in liquid nitrogen. Total RNA was extracted using Trizol, quantified by spectrophotometry in a NanoDrop 2000 (ThermoScientific) and further analysed in 1% agarose gel electrophoresis. To remove any DNA contamination, samples were treated with DNase I (Thermo Scientific,) and then reverse transcription was performed using iScript™ cDNA Synthesis Kit (Bio-rad, 1708891). The cDNA was subjected to qPCR using GoTaq® qPCR Master Mix (Promega, A6002). For each reaction, 3 µL of cDNA (1:100 dilution) and 0.3 µL of the respective forward and reverse gene-specific primers (10 µM) were used in a final volume of 15 µL. Reactions were conducted on the 7500 Real Time PCR System (Applied Biosystems) under the follow conditions:

- Holding stage - 50°C (2 minutes) and 95°C (10 minutes);
- Cycling stage (40 cycles) 95°C (15 seconds) and 60°C (1 minute).

Gene relative quantification was performed using $2^{-\Delta\Delta CT}$ method. The endogenous polyubiquitin gene (SCCCST2001G02.g; NCBI accession CA179923.1) was used as internal control (Papini-Terzi et al., 2005) for data normalization. Expression level of target genes in the silenced plants (*ScGUX2*) was compared with the control groups (WT and EV) expressed

as fold change and qRT-PCR was calculated using fold change average from the three biological replicates.

Primers functionality was tested by evaluating their CT and melting curve, and their efficiency was measured by analysing the efficiency coefficient of standard curve.

**Table 4. Primers used in the qPCR**

| Gene | Primer Forward (5'-3') | Primer Reverse (5'-3') |
|---|---|---|
| *ScGUX2* | GGGTCATGCAATCTCGCTAT | GGTACGTGTCCGACGAGTG |
| *ScUBI* | CCGGTCCTTTAAACCAACTCAGT | CCCTCTGGTGTACCTCCATTTG |

## 7.4. CELL WALL FRACTIONATION

Aliquots of AIR material (0.5 mg and 1 mg) were deacetylate with 20 µL of 4M NaOH and let react for 1 hour at room temperature (RT). Followed by a neutralisation with 80 µL of 1M HCl and completed up to 1000 µL with 0.1 M of ammonium acetate pH 5.5. For 0.5 mg AIR aliquots it was added GH11 enzymes and for 1 mg AIR aliquots it was added GH30 enzymes as described below.

## 7.5. CELL WALL FRACTIONATION AND POLYSACCHARIDE ANALYSIS USING CARBOHYDRATE GEL ELECTROPHORESIS (PACE)

For PACE the ball milled material obtained from section 7.1 was used to prepare Alcohol Insoluble Residue (AIR). All AIR preparation was carried out as described in (Mortimer et al., 2010). This method consists in fractionating the plant cell wall with the use of an alkaline reaction three times and analyses the product using ANTS PACE (Mortimer et al., 2010; Brown et al., 2007). After the PACE run, the polysaccharide will move according to their molecular weight and will be compared with a specific ladder that will serve as a parameter to the analysis. For extraction of sugarcane acetylated heteroxylan, the same protocol in section 4.3 was followed with WT sugarcane and *ScGUX2* silenced plants.

## 7.6. CELL WALL FRACTIONATION

Aliquots of AIR material (0.5 mg and 1 mg) were deacetylate with 20 µL of 4M NaOH and left to react for 1 hour at Room Temperature (RT). Followed by a neutralisation with 80

µL of 1M HCl and completed up to 1000 µL with 0.1 M of Ammonium Acetate pH 5.5. For 0.5 mg AIR aliquots it was added GH11 plus GH62 enzymes and for 1 mg AIR aliquots it was added GH30 plus GH62 enzymes as described below.

### 7.7. MEASUREMENT OF XYLOSE (XYL) RELEASED BY PACE

To examine if the structure of xylan has been affected by RNAi silencing, we will use xylanases to digest the xylan, and examine the oligosaccharides produced by PACE or DASH. The different events will have distinct oligosaccharide profiles in the case their structure has been altered due to the RNA interference of the xylan biosynthetic enzyme *ScGUX2* (Bromley et al., 2013).

To understand if there is a reduction in the frequency of xylan glucuronic acid substitutions in *ScGUX2* silenced sugarcane, we digested alkali-extracted xylan with the enzyme GH11. This xylanase produces just three xylan oligosaccharides, one of which contains glucuronic acid. Running the samples on a PACE gel allowed us to analyse the relative amounts of the different oligosaccharides and establish if there is a reduction in glucuronic acid substitutions on xylan in the *ScGUX2* RNAi lines. To test if the spacing between consecutive glucuronic acid substitutions has been altered by *ScGUX2* silencing, we digested alkali-extracted xylan with BoGH30. This enzyme cuts the xylan backbone whenever it recognizes a glucuronic acid, thus the length of the released oligosaccharides will give information about the number of xylosyl residues between consecutive glucuronic acid substitutions. The GUX2 enzyme in Arabidopsis produces tightly packed glucuronic acid substitutions that are normally 5, 6 or 7 xylosyl residues apart. These may be reduced in frequency in the *ScGUX2* silenced sugarcane. As the BoGH30 is known to also occasionally cut the xylan backbone when there is an arabinose substitution, instead of a glucuronic acid, we can combine BoGH30 (xylanase) with GH62 (arabinofuranosidase) to remove the arabinoses from the xylan. We expected different relative amounts of xylan oligosaccharides in the *ScGUX2* RNAi line.

### 7.8. POLYSACCHARIDE ANALYSIS BY CARBOHYDRATE GEL ELECTROPHORESIS (PACE) OF B-XYLANASE GH11 PLUS GH62 DIGESTION PRODUCTS

AIR material (0.5 mg) was digested with *N. patriciarum* GH11 enzyme plus GH62 overnight as adapted from Mortimer et al. (2010). Released oligosaccharides were dried and

derivatised with 8-aminonapthalene-1,3,6-trisulphonic acid (ANTS; Invitrogen). ANTS derivitasation, PACE running and visualisation were performed as previously described (Goubet et al., 2002; Bromley et al., 2013).

### 7.9. POLYSACCHARIDE ANALYSIS BY CARBOHYDRATE GEL ELECTROPHORESIS (PACE) OF GLUCURONOXYLANASE GH30 PLUS GH62 DIGESTION PRODUCTS

AIR material (1 mg) was digested with *Bacteroides ovatus* GH30 enzyme (Glucuronoxylanase 30A, Nzytech) for 30 minutes in a thermomixer at 800 rpm at 37°C and to denature the enzymes, samples were put straight into a heat block at 100°C for 10 min. Then GH62 was added overnight at 800 rpm at 37°C. Released oligosaccharides were dried and derivatised with 8-aminonapthalene-1,3,6-trisulphonic acid (ANTS; Invitrogen). ANTS derivitasation, PACE running and visualisation were performed as previously described (Goubet et al., 2002; Bromley et al., 2013).

## 8. SACCHARIFICATION OF SUGARCANE BIOMASS

Novozymes Cellic® CTec2 (also available from Sigma-Aldrich/Merck) was used for all saccharification and fermentation experiments. Enzyme stock (350 µL) was diluted to a total volume of 2.5 mL with 0.1M ammonium acetate pH = 5.0 (AmAc) buffer. The enzyme sample was cleared from residual sugars using PD-10 desalting column (GE Healthcare) and eluted using 3.5 mL AmAc buffer, generating 1:10 (v/v) Cellic® CTec2 solution. Dried biomass (1 mg) was homogenised in 1 mL of AmAc buffer pH 5.0. Homogenised AIR was amended with 20 µL 1:10 Cellic® CTec2 working solution.

For saccharification of dried biomass 20 µL of 1:10 (v:v) Cellic® CTec2 solution was used. The enzyme solution was added to 1 mg biomass material/ml 0.1 M AmAc buffer pH=5.0. Biomass suspension was generated by ball milling 10 mg of the biomass in 10 mL of ethanol 70% for 10 min at 25 Hz, with 10-min intervals between each ball milling for 5 times.

Saccharification was carried out for 24 h at 50°C with 800 rpm. The reaction was terminated by heat-treating the suspension at 100°C for 10 min. D-Glucose and D-Xylose release from the biomass was quantified using commercial kits (Megazyme). Sugar

concentration for each experiment was standardised with readings obtained from biomass and enzyme only controls.

## 9. HISTOLOGICAL ANALYSES

Stem material from internodes 3 and 7 was fixed in formalin/acetic acid/ethanol/water (5:5:60:30, v/v/v/v) (Johansen, 1940) and histochemical tests were made with hand cut sections of ~0.5 mm thickness using a steel blade. The sections were stained with Mäule reagent (Mäule, 1901) and for lignin histochemical analysis it was used phloroglucinol (Johansen, 1940). For Mäule, the sections were treated with $KMnO_4$ for 3 min, washed with water, maintained in 10% HCl for 1 min, and mounted in ammonia. For phloroglucinol, the sections were treated for 3 min with 1% phloroglucinol in 95% ethanol and mounted in 25% HCl. Photomicrographs were taken with an Olympus BX 51 photomicroscope equipped with an Olympus DP71 camera.

## 10. STATISTICAL ANALYSIS, SAMPLING AND THESIS PREPARATION

For all quantitative experiments, unless otherwise stated, three biological replicates of plant material were grown and analysed. For each biological replicate, 3 technical replicates were analysed for the quantitation of sugar release efficiency. Statistical analysis was performed using packages available with R Studio software (R Foundation). Statistical tests used to compare average measurements for samples are mentioned in Figure legends and mostly include Student's T test and ANOVA.

# CHAPTER 3: RESULTS

## 1. Identifying and characterizing sugarcane GUX genes

There are five GUX protein sequences for *Arabidopsis thaliana* already characterized by Mortimer et al. (2010) and Rennie et al. (2012), and their sequences were retrieved from GenBank (https://www.ncbi.nlm.nih.gov). In order to reconstruct the phylogenetic relationship of the GUX family, we selected 16 angiosperm species (including six monocots and ten dicots) that are either model plants or important crops and the common liverwort to serve as outgroup in the phylogenetics analysis (Gallinari et al. 2020).

Using putative GUX proteins identified *in silico* for each species and their aligned sequences, we reconstructed the phylogenetic trees. Maximum likelihood and Bayesian phylogenies arranged the GUX family into well-supported clades, allowing us to define the orthologous and paralogous relationships. The only exception was for the clade called GUX 'X', which is composed of few monocots GUX proteins arranged in different places of the tree depending on the dataset used (nucleotides or aminoacids), and hence we could not establish with complete confidence whether these genes are duplications originated from GUX 4 or GUX 1/3. However, the tree derived from aminoacids alignment (Figure 16) presented a stronger support for a relationship with GUX 1/3 (99.6 from bootstrap and 1.0 of posterior probability). Therefore, with caution, we will consider GUX 'X' a specific monocot duplication from GUX 1 or 3 gene.

Regarding the evolution pattern observed in the GUX gene family, it can be attributed to a mixture of divergent, concerted and birth-and-death evolutionary models. The divergent model, *i.e.* accumulation of differences between groups that may ultimately lead to the formation of new species/groups (Nei and Rooney, 2005) can be observed in the GUX2 clade. In this case, there is a division between genes from monocots and dicots (Figure 16), indicating that *GUX2* originated before the split between monocots and dicots, and that during evolution they accumulated changes specifics to each group. A similar divergent model was observed

in *PHO1* genes, which are involved in phosphate absorption in plants, and where Class II genes from monocots and dicots are separated (He et al., 2013).

The concerted evolution, *i.e.*, members of a gene family evolving in a concerted manner instead of independently (Nei and Rooney, 2005), can be observed in the relationship between the GUX1 and GUX3 clades (Figure 16). Regarding these genes, the phylogenetic tree recovered paralogous clades instead of orthologous clades, indicating that paralogous genes (*e.g.* GUX1 and GUX3 of monocots) are more similar to one another than they are to their true orthologs in closely related species (*e.g.* GUX1 of both monocots and dicots). The clade GUX1 monocot was named this way because BLAST analyses of most of its sequences show the *Arabidopsis* GUX1 as top hits. The same reasoning applies for the clade GUX3 monocot, where most sequences are more similar to *Arabidopsis* GUX3 than GUX1. However, further functional analyses of these proteins are necessary to corroborate the paralog relationship of GUX1 and GUX3 in monocots. The concerted evolution model has also been observed among rice genes from chromosome 11 and 12 that went through a series of genomic modification events until they became more similar among their paralogs than their orthologs (Wang et al., 2007). Further explanations about the phylogenetic tree and the other GUX genes can be found at Gallinari et al. (2020).

Since our main goal with this tree was to identify GUX sugarcane orthologues, after analysing the phylogenetic tree and interpreting the relationships between species, we chose two putative sugarcane *GUX* sequences to be further studied (Saccharum_sp1 and Saccharum_sp2 in Figure 16).

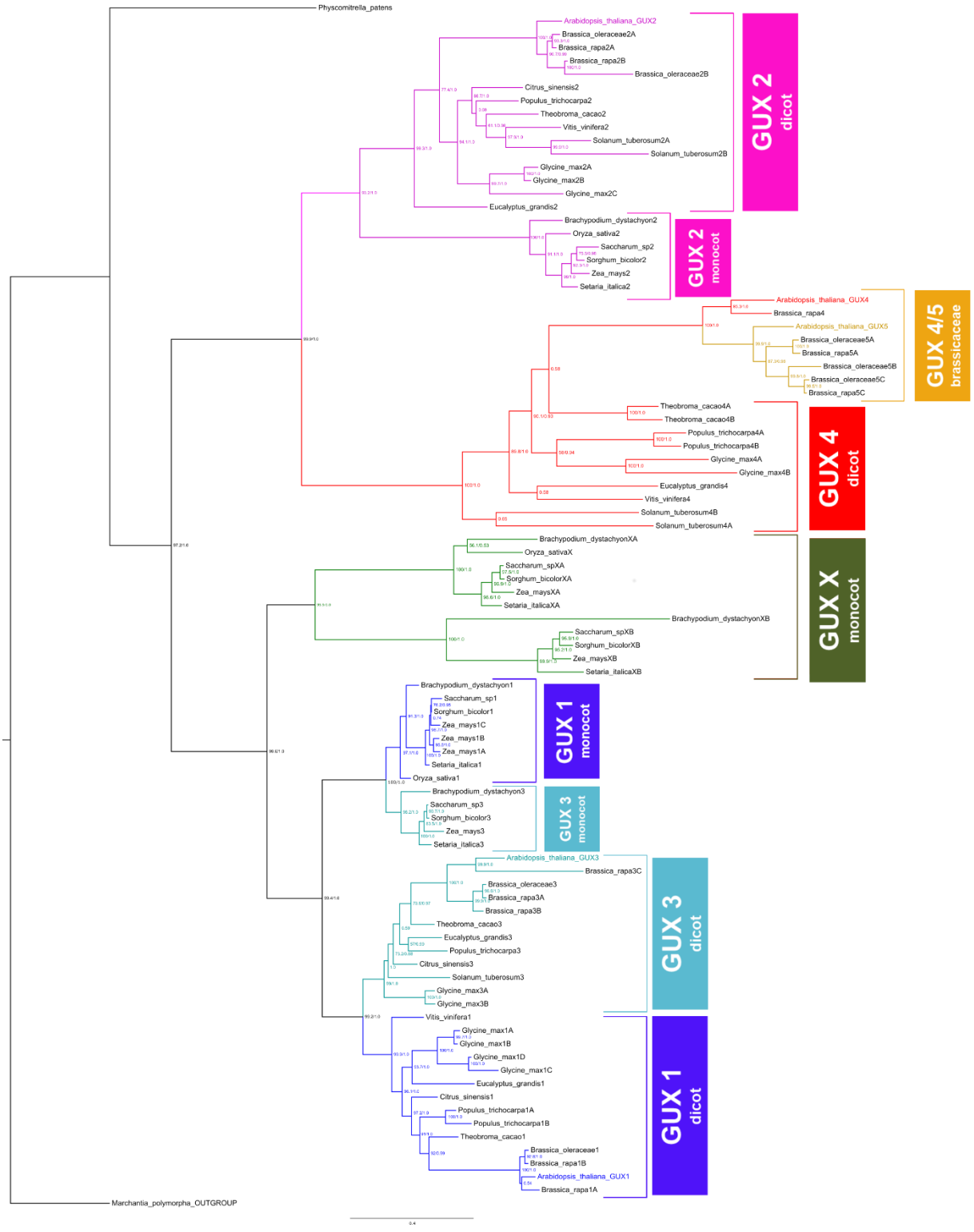



**Figure 16 Phylogenetic tree of GUX proteins in plants.** Numbers on nodes correspond to the maximum likelihood (ML) ultrafast bootstrap support values followed by Bayesian posterior probabilities. The colored branches are represented by: GUX1 (dark blue), GUX2 (pink), GUX3 (light blue), GUX4 (red), GUX4/5 (yellow) and GUX 'X' (green). The GUX sequence from *Marchantya polymorpha* was used as the outgroup (Gallinari et al., 2020)

## 2. ScGUX1 and ScGUX2 are active enzymes *in vitro*

Having identified sugarcane homologues of known GUX enzymes we wanted to establish if putative ScGUX are active glucuronosyltransferases onto xylan. The putative *ScGUX* genes: *ScGUX1* and *ScGUX2*, from *Saccharum_sp1* and *Saccharum_sp2* respectively were obtained from phylogenetic tree analysis (Figure 16). To detect if the putative ScGUX enzymes have xylan glucuronosyltransferase activity, the enzymes were expressed in the tobacco *N. benthamiana* (Sparkes et al., 2006) using a GoldenGate vector (Figure 17A and 17B). As a control for assaying glucuronosyltransferase activity, a family GT61 enzyme, which is a putative conifer xylan arabinosyltransferases, was also expressed. Expression of the enzymes was confirmed using western blot against the Myc tag which is placed at the C terminus of the proteins of interest (Figure 17C).

Following successful expression of the proteins, we wanted to study the *in vitro* glucuronosyltransferase activity of ScGUX. For that, intact polymeric xylan from *gux1/2* mutant Arabidopsis, lacking any [Me]GlcA decorations, was used as an acceptor. Since this xylan is insoluble without acetylation, acetylated polymeric *gux1/2* xylan was used in all *in vitro activity reactions.* Microsomes extracted from tobacco leaves expressing ScGUX1, ScGUX2 or the control GT61 were incubated with acetylated *gux1/2* xylan and the reaction products were extracted after 5 hours of reaction and deacetylated. Reaction products were analysed by PACE using digestion with a GH11 xylanase (endo-1,4-β-Xylanase, Megazyme). The product GlcA-xylotetraose (X**U**XX) that indicates xylan glucuronosyltransferase activity was observed for reactions using microsomes from ScGUX1 and ScGUX2 expressing plants incubated with the xylan acceptor in the presence of UDP-GlcA (Figure 18). This result is consistent with the activity observed for PgGUX plus UDP-GlcA reaction on *gux1/2* acetylated xylan (Lyczakowski et al., 2017)

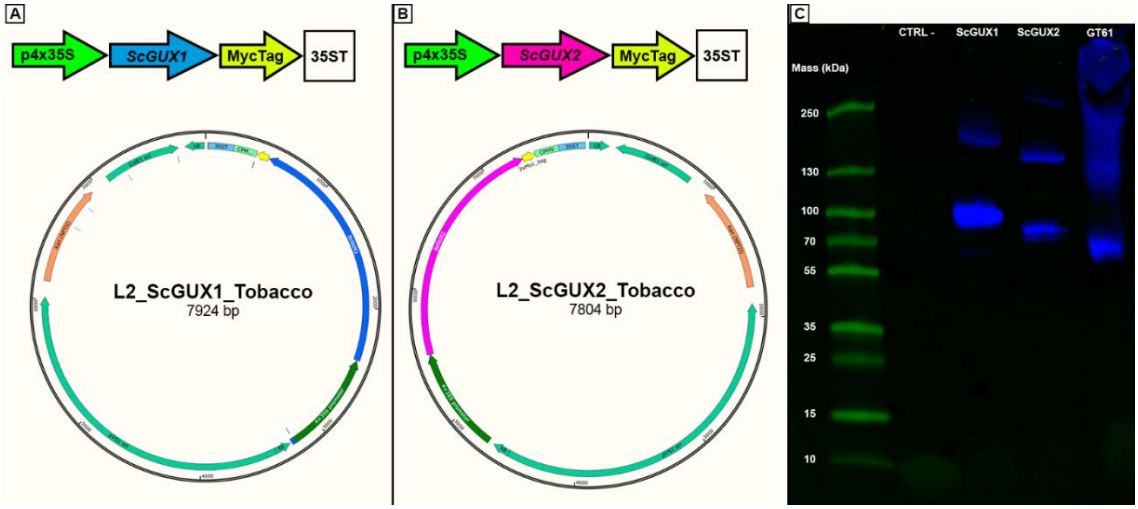



**Figure 17 Vector maps of constructs for expression of ScGUX1 (A) and ScGUX2 (B) in tobacco (*N. benthamiana*).** C) Western blot from enriched microsomes from *N. benthamiana* of no infiltrated tobacco leaves (CTRL -), ScGUX1, ScGUX2 and a GT61 enzyme with Myc-Tag.

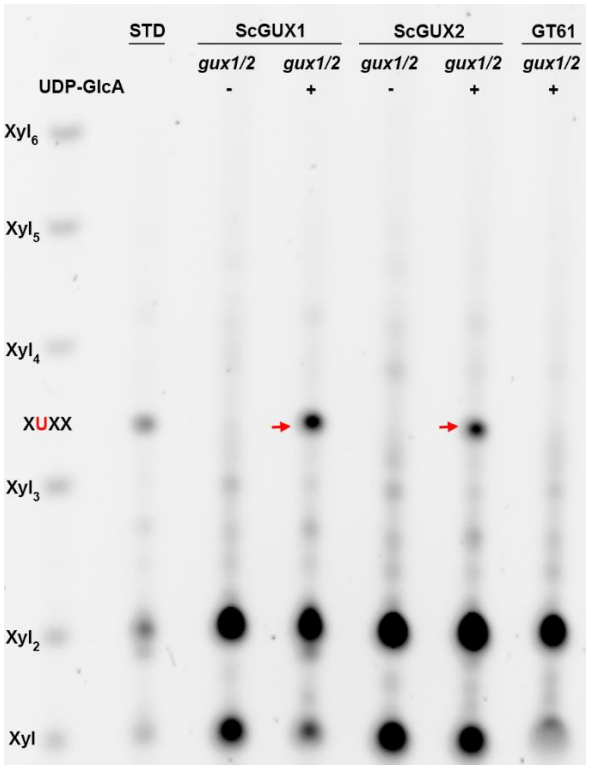



**Figure 18 ScGUX1 and ScGUX2 have xylan glucuronosyltransferase activity in vitro.** The assays were performed with UDP-GlcA, acetylated xylan without Me[GlcA] decorations from Arabidopsis *gux1/2* mutants and microsomes from *N. benthamiana* enriched for ScGUX1, ScGUX2 or control GT61 protein. Products of the *in vitro* activity reaction were digested with β-xylanase GH11 (Megazyme) and analysed by PACE. The enzyme generates xylose, xylobiose and X**U**XX (highlighted in red) oligosaccharide if any GlcA is present on the xylan. The positive (+) symbol means that 10mM of UDP-GlcA was added to the assay, while the negative (-) symbol means that no UDP-GlcA was added. *A. thaliana* WT was used as a Standard (STD) for the digestion.

### 3. Complementation of xylan structural phenotype in *gux1/2* Arabidopsis mutant

Having established that ScGUX1 and ScGUX2 are active on acetylated xylan *in vitro*, we tested whether we could use these enzymes to introduce xylan decorations *in vivo*. This will provide further validation that the enzymes are active glucuronosyltransferases. The ScGUX coding sequences were placed under the control Arabidopsis *IRX5* promoter, active in lignifying tissues. The constructs were transformed into Arabidopsis *gux1/2* plants which lack [Me]GlcA branches on secondary cell wall xylan (Mortimer et al.,2010). These mutant plants were chosen in order to ensure that most part of [Me]GlcA detected on the xylan in the transgenic plants would be introduced by the ScGUX enzymes. For the screening of the homozygous seeds the gene *Oleosin* was tagged with eGFP driven by promoter Oleosin and Nos terminator (Hellens et al., 2000; Cormack et al., 1996; Depicker et al., 1982). This system allows identification of transgenic seeds by presence of GFP within the seed-coat.

After obtaining Arabidopsis *gux1/2* plants expressing either *ScGUX1* or *ScGUX2*, we sampled cell wall material from their basal stems, deacetylated it and digested with a xylanase GH11 (Megazyme) and analysed resulting oligosaccharides using PACE (Figure 19).

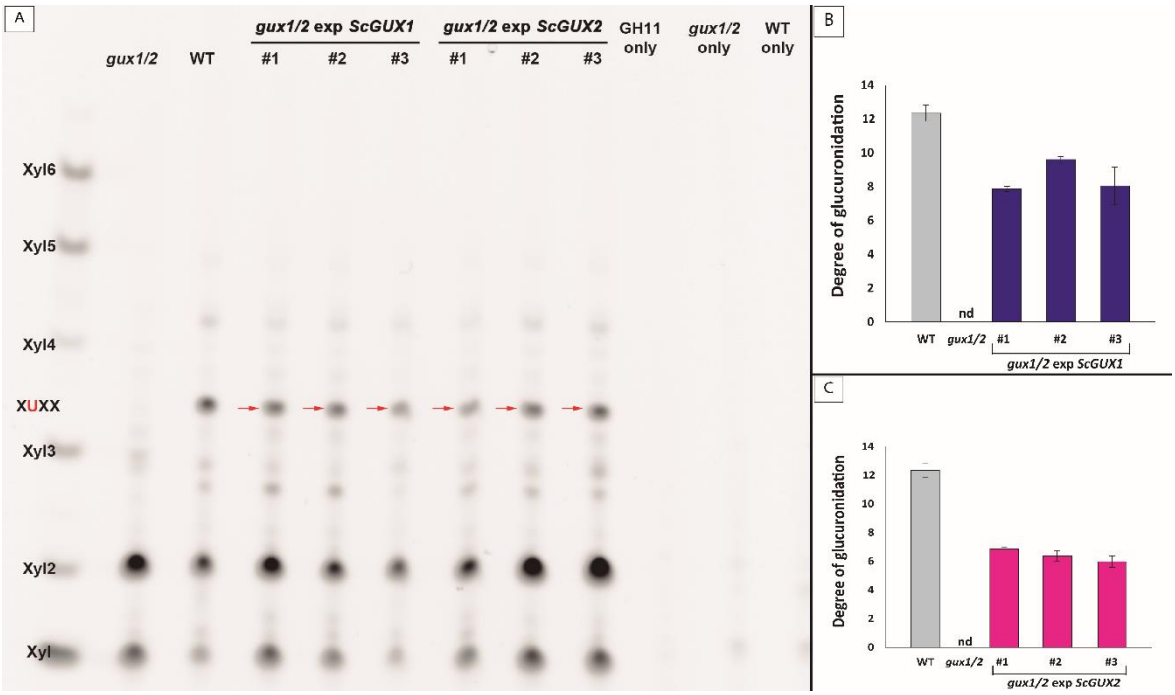



**Figure 19 A) Complementation experiment of *gux1/2* Arabidopsis plants expressing *ScGUX1* or *ScGUX2***.PACE analysis after xylanase GH11 digestion from 500 µg of AIR. *A. thaliana* was used as a standard (STD) for the

digestion. *gux1/2*: double mutant Arabidopsis; WT: *A. thaliana* Wild Type; *gux1/2* expressing *ScGUX1* and gux1/2 expressing *ScGUX2*: *gux1/2* double mutant overexpressing *ScGUX1* and *ScaGUX2*, respectively (#1, #2, #3 indicate (three independent lines), *gux1/2* expressing *ScGUX2* (three independent lines) and controls (enzyme control: GH11 only, biomass controls: *gux1/2* and WT without the enzyme). The glucuronidated product of the digestion: XUXX is highlighted in red and the bands which correspond to it were marked with red arrows. Quantification of the degree of xylan glucuronidation in all plants analysed using ImageJ for *gux1/2* expressing *ScGUX1* (B) and *gux1/2* expressing *ScGUX2* (C) is also presented.

In this analysis (Fig 19A), plants with GUX enzymes active released the X**U**XX oligosaccharide. The band is present in digestion products from WT plants but absent in digests performed using *gux1/2* cell wall material. Interestingly, we have observed that all three independent transgenic lines expressing each gene separately (*ScGUX1* or *ScGUX2*) in *gux1/*2 Arabidopsis plants have detectable XUXX. This demonstrates that in *gux1/2* plants expressing *ScGUX1* and *ScGUX2* the presence of GlcA on xylan is recovered. This provides further evidence that *ScGUX1* and *ScGUX2* are glucuronosyl transferases and are able to add GlcA onto the xylan, as observed with the *in vitro* activity experiments using *gux1/2* acetylated heteroxylan (Figure 18). From the PACE experiments the amount of glucuronidation on xylan in transgenic plants was calculated using imageJ for each of the genes (*ScGUX1* – Figure 19B; *ScGUX2* – Figure 19C) expressed in Arabidopsis *gux1/2* mutants. For *ScGUX1*, the obtained degree of glucuronidation varied between 8 to 10%, while for *ScGUX2* it amounted varied between 5 and 7%. This is less than 12% reported previously for WT Arabidopsis (Mortimer et al., 2010) and measured in this study.

### 4. Activity of ScGUX2 in sugarcane plants

After demonstrating that the ScGUX enzymes are active xylan glucuronosyltransferases we wanted to study the physiological role of xylan glucuronidation in sugarcane plants. To achieve this, we used the pGVG vector, a pCAMBIA modified with Gateway recombination sites (Guidelli et al., 2018), to silence *ScGUX2* in sugarcane. We have decided to silence ScGUX2 based on the expression profile of the homologs in a model monocot species, *Brachypodium dystachion*. Although *BdGUX1* expression is higher in stem (Figure 20A) it is also expressed across all the plant. On the other hand, *BdGUX2* expression is more specific to the stem material (Figure 20B). Since we were primarily interested in modifying xylan in the sugarcane culm, this potential stem specificity can be beneficial, as we

would avoid changing xylan structure in other tissues that may cause unwanted phenotypes. All the genes cited here were used based on our work (Gallinari et al., 2020).

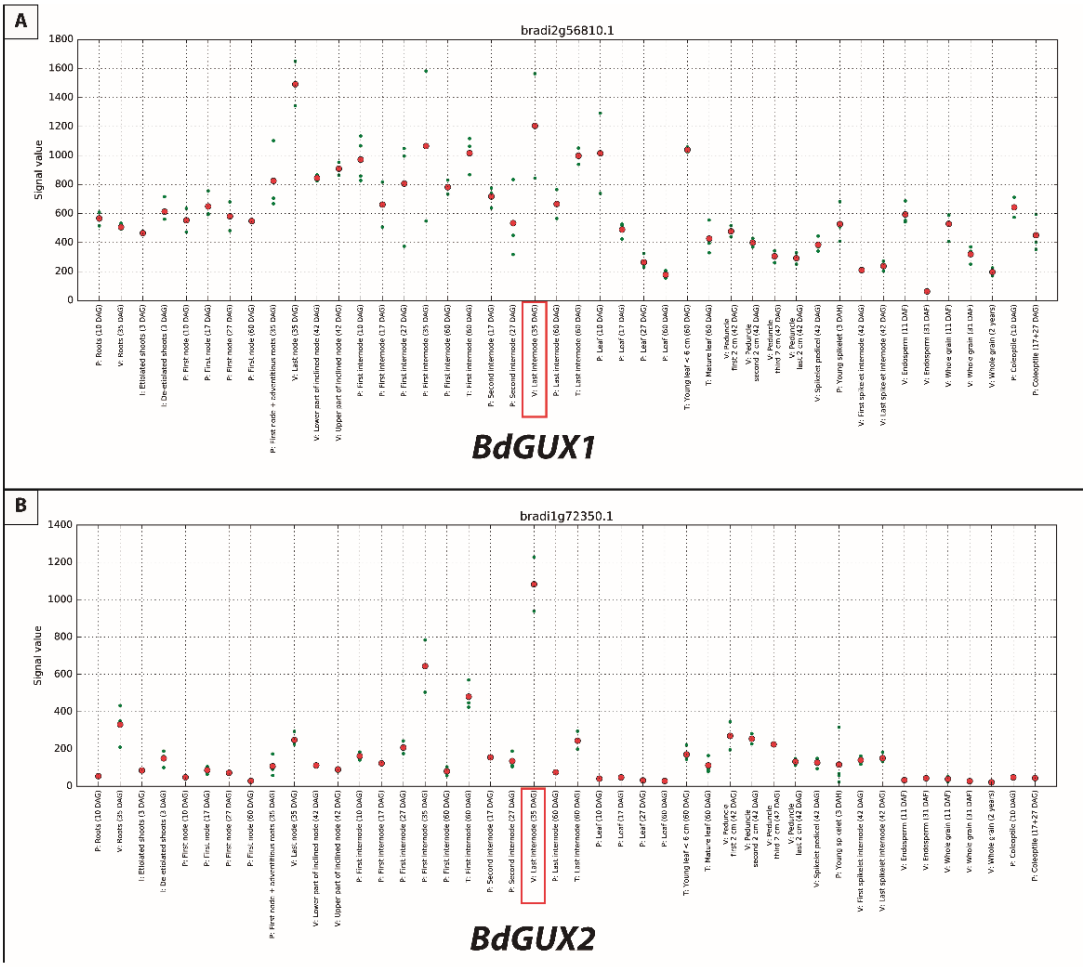



**Figure 20 Profile expression of Bradi2g56810.1 (*BdGUX1* - A) and Bradi1g72350.1 (*BdGUX2* - B).** DAG: Days after germination, days after fertilization (DAF), days after heading (DAH). Growth facilities are indicated: P, growth chamber in Versailles; V, glasshouse in Versailles; I, *in vitro* culture in Versailles; T, growth chamber in Toulouse. Adapted from Sibout et al., 2017

Following the choice of *ScGUX2*, the coding sequence (CDS) was determined and specific primers were built to amplify this region in two distinct senses: anti-sense *ScGUX2* (as*ScGUX2*) and sense *ScGUX2* (s*ScGUX2*). After amplified both senses of *ScGUX2*, they were linked to an entry vector modified pCR8/GW/TOPO, with restriction sites compatible to the construct, Gateway compatible sites and then recombined into the destination vector pGVG (Guidelli et al., 2018), resulting into the final vector pGVG: *ScGUX2*_RNAi (Figure 21). All the steps were confirmed by restriction enzymes reactions followed by the expected bands and

sequencing to ensure that the final vector was exactly as planned. Then, the agrobacterium EHA105 was transformed with pGVG: *ScGUX2*_RNAi and it was given to Pangeia Biotech to transform sugarcane calli material, that delivered plantlets of transgenic sugarcane.

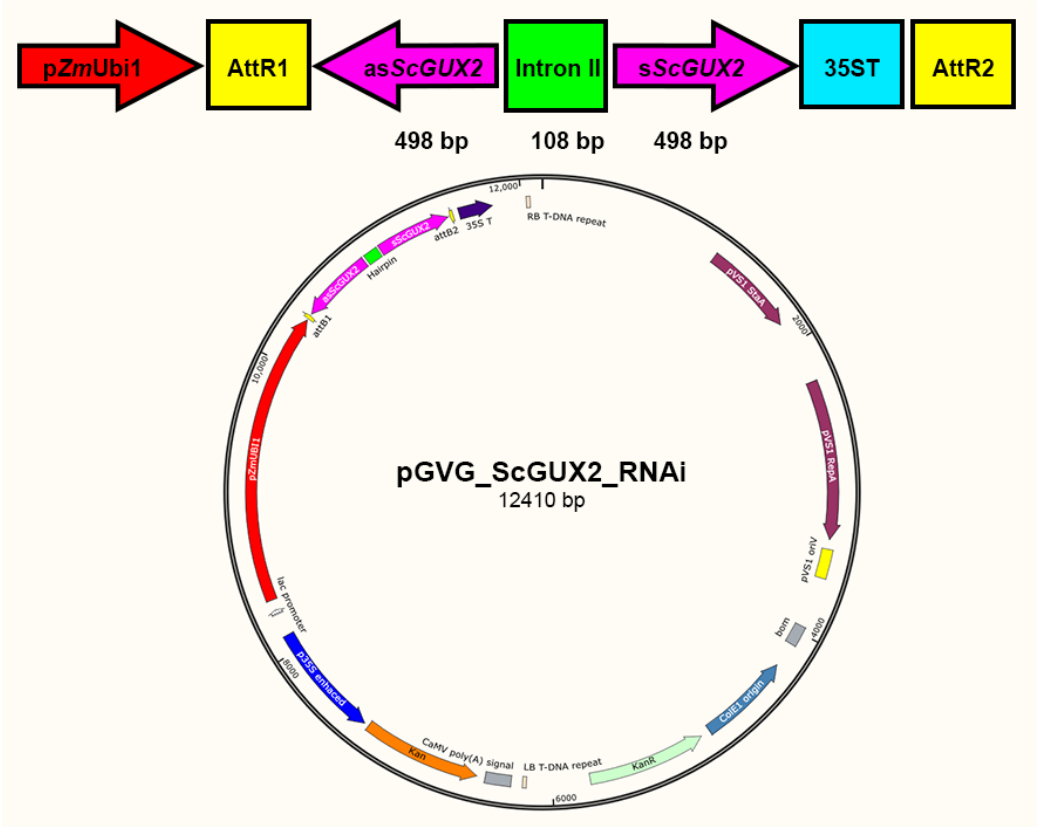



**Figure 21 Vector map for silencing of *ScGUX2* in sugarcane.** pZmUbi1 promoter, AttR1/R2 (gateway cassette), *ScGUX2* sequence (as: anti sense and s: sense), Intron II: RNAi hairpin, 35ST: CaMV 35S terminator.

Since *GUX* genes are related with secondary growth, we had to wait the plants to be mature enough to be able to measure it expression through RT-qPCR. The plants were transplanted to bigger pots and regular maintenance (fertilizer, irrigation and weed removal) was done. In order to evaluate the *ScGUX2* expression, we have analised three parts of the 9 months old WT sugarcane plant: leaves, young culm and mature culm (Figure 22A). From this profile expression we noticed that young culms and leaves have a similar expression of *ScGUX2*, but since our main goal is to understand the relation of this gene in the bagasse of sugarcane, young culms were chosen to be further analysed. The *ScGUX2* expression was measured from young culms (internodes 3 to 5) from 7 months old plants (Figure 22B).

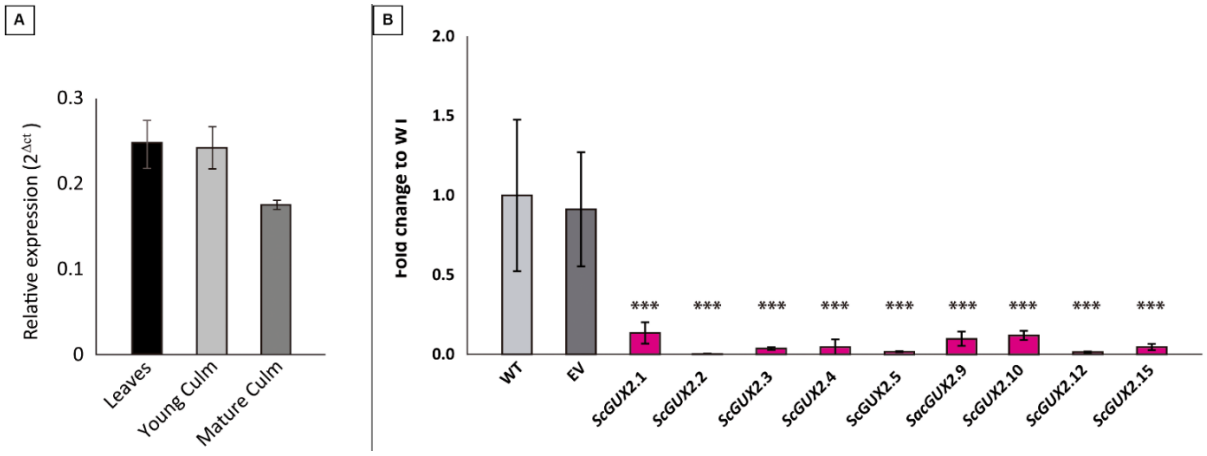



**Figure 22 A) *ScGUX2* profile expression from 9 months old WT sugarcane plants.** B) *ScGUX2* expression in transgenic lines (internodes 3-5 from 7 months old plants). WT: No transformed WT sugarcane, EV: Empty Vector and *ScGUX2*: silenced sugarcane RNAi lines. All the transgenic lines have lower expression in comparison to control and empty vector (n=3). Error bars represent standard deviation of three technical replicates of biomass *p value <0.05; **p value <0.01; ***p value <0.001 (pairwise T-test).

The transcription levels of *ScGUX2* were quantified in wild-type plants and plants transformed with the silencing vector. As a control, plants transformed with the empty vector (EV) were also analysed. In the experiment young culms (internodes 3 to 5) were used for RNA extraction and expression profiling. The levels of the transcripts from *ScGUX2* were quantified by qPCR and compared to select the transgenic events with lowest *ScGUX2* expression. In all transgenic lines we observed reduction in *ScGUX2* expression compared to WT levels. Most lines transformed with the silencing vector had lower expression level than the EV control. This indicates successful silencing of *ScGUX2* in the transgenic material.

After confirming that we were able to successfully reduce the expression of *ScGUX2*, we analysed the phenotype of plants with reduction in gene expression. In our analysis, we have observed no difference when number of culms (Figure 23A), culm diameter (Figure 23B), plant height (Figure 23C), plant fresh weight (Figure 23D) or brix yield (Figure 23E) of the WT and transgenic plants were compared. Therefore, downregulation of *ScGUX2* expression does not have any significant effect on plant growth and agronomic performance.

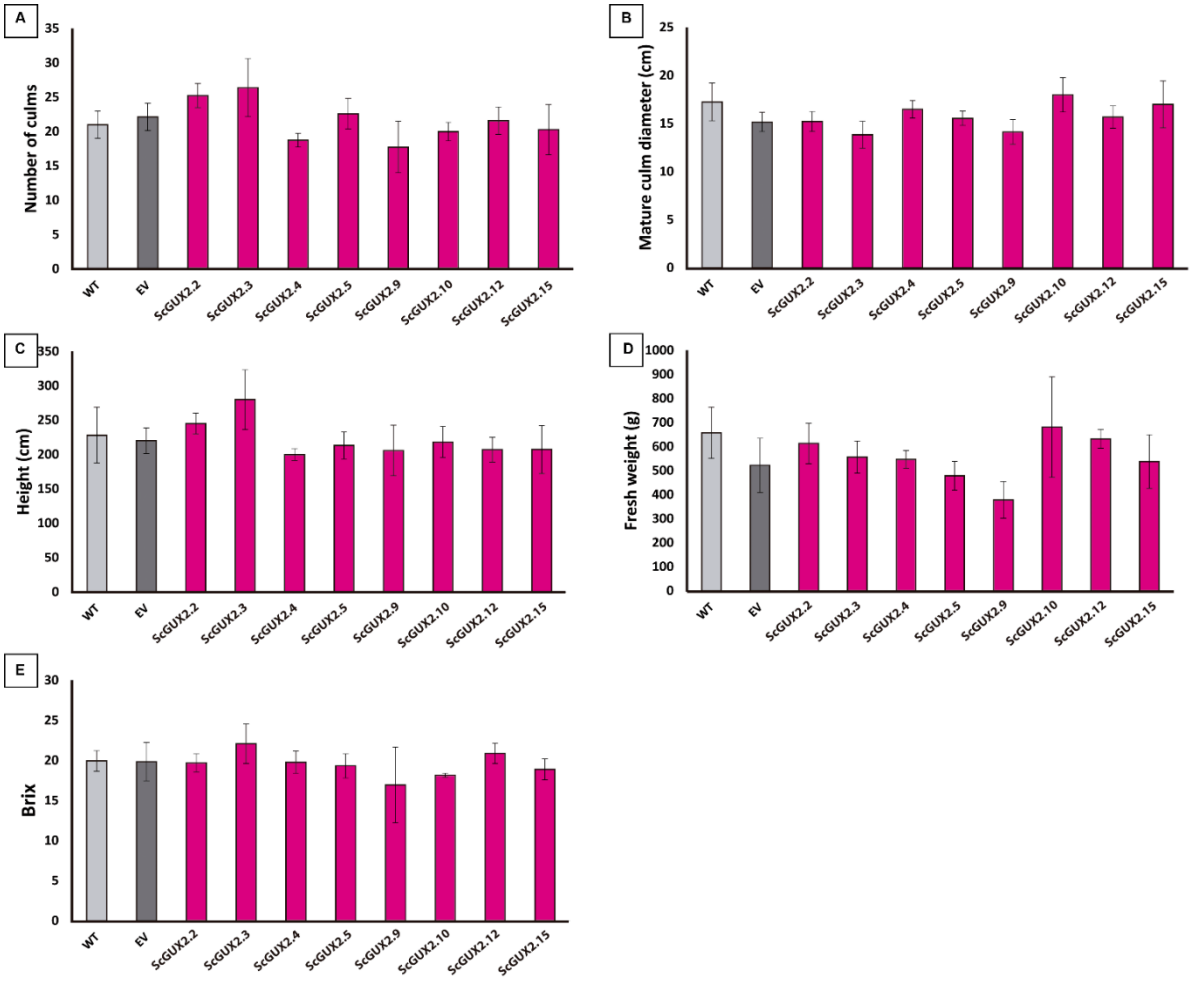



**Figure 23 Compilation of agronomic parameters for *ScGUX2* gene in sugarcane.** (A) Number of culms among the transgenic plants and controls. (B) Mature culm diameter from the base – internodes 10 and 11. (C) Height of the plants from the first internode to base – without leaves. (D) Fresh weight of the culms, the straw from old leaves was removed. (E) Brix levels from the base: internodes 10 and 11. All the measurements were evaluated using ANOVA (n=5).

## 5. Properties of the biomass from *ScGUX2*-silenced plants

To investigate if the silencing of *ScGUX2* has an effect on xylan structure in sugarcane biomass, we digested culm AIR with a combination of arabinofuranosidase GH62 (Figure 24A) and β-xylanase GH11 (Figure 24B). Since sugarcane xylan is highly arabinosylated the arabinofuranosidase treatment is necessary to see clear degradation of xylan to glucuronidated and non-glucuronidated oligosaccharides. This will allow quantification of the degree of xylan glucuronidation in WT and transgenic lines. We hypothesised that in the transgenic lines the degree of xylan glucuronidation will be reduced due to *ScGUX2* silencing.

Released oligosaccharides (Figure 24C) were derivatised with ANTS fluorophore and analysed with PACE (Figure 24D). In order to make it easier to understand the enzymes choice and how to interpret the results for the next paragraphs of the thesis we have done a representative scheme.

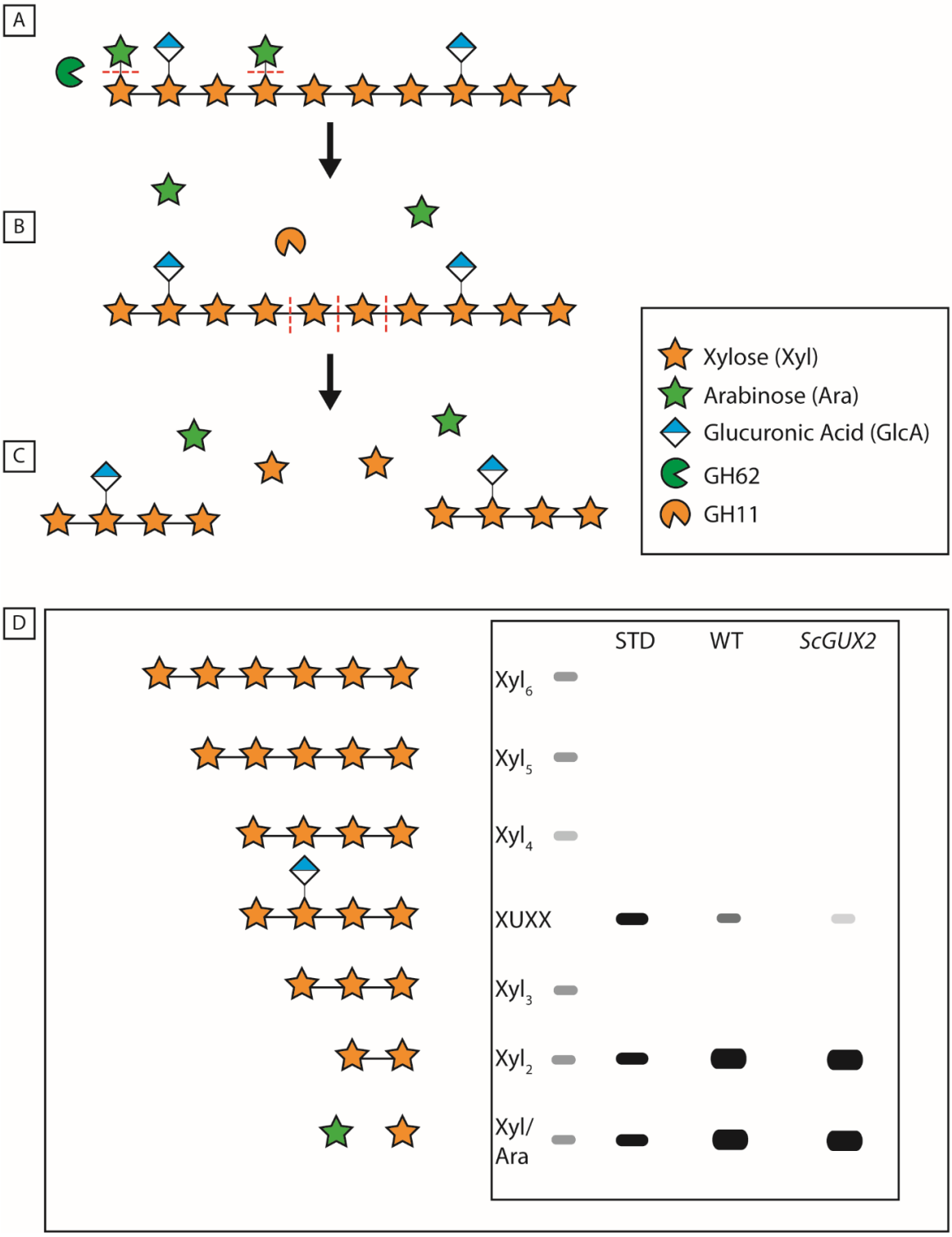



**Figure 24 Scheme of xylan digestion with GH62 and GH11 followed by a PACE gel result.** GH62 (A) was added to remove arabinose making the xylan more acessible for GH11 (B) digestion, resulting in the products (C): xylotetraose (XUXX), xylobiose ($Xyl_2$) and xylose (Xyl) plus arabinose (Ara). (D) It is a representation of the result after running the digestions in a PACE gel, where we can see a column which represent the xylose ladder ($Xyl_6$, $Xyl_5$, $Xyl_4$, $Xyl_3$, $Xyl_2$ and Xyl). The next three columns in the gel correspond to a standard (STD), wild type sugarcane (WT) and a sugarcane silenced line *ScGUX2*.

The GH11 + GH62 digestion of sugarcane culm AIR resulted in the release of three main products: Xylose (Xyl), Xylobiose (Xyl2) and xylotetrasaccharide with one GlcA branch (XUXX) (Figure 25A and 25B). The red arrows are showing the xylooligosaccharides with GlcA (X**U**XX). Interestingly, the silencing of sugarcane *ScGUX2* resulted in a reduction of XUXX band intensity in some of the transgenic lines. This is similar to what was observed in Arabidopsis *gux* mutants (Mortimer et al., 2010; Bromley et al., 2013). The reduction in XUXX band intensity suggests that silencing of *ScGUX2* results in a reduction in the degree of xylan glucuronidation. To quantify this change, we have determined the degree of xylan glucuronidation in all the analysed lines (Figure 25C) and observed that in the transgenic material, the degree of xylan glucuronidation ranged from 0.4 to 3.9%. This is less than for WT and EV controls where approximately 4% and 3.5% of xyloses respectively had the GlcA branch. The use of arabinofuranosidase GH62 was to remove the arabinoses and make the xylan easier to be digested by GH11 and result in fewer bands to be analysed. In addition to that, it worth to say that the degree of xylan glucuronidation is an approximation, since arabinose runs together with xylose monosaccharide on PACE and became the same band, but even though the reduction is evident. As long as there is no difference in any agronomic traits measured for the transgenic, WT and EV plants, the observed reduction in the degree of glucuronidation is the first detected phenotype associated with *ScGUX2* silencing in sugarcane.

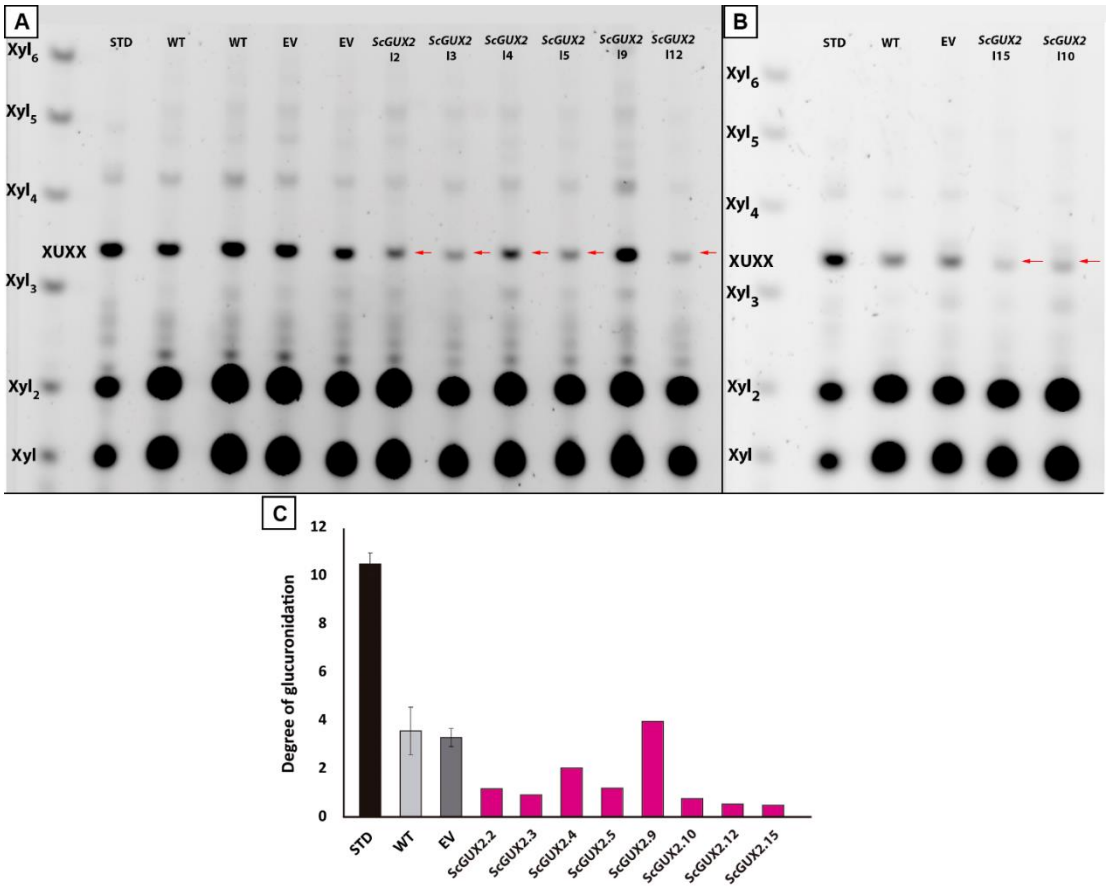



**Figure 25 Analysis of the degree of glucuronidation on xylan from different sugarcane plants with *ScGUX2* expression silenced**. A) and B) PACE analysis of arabinofuranosidase GH62 plus xylanase GH11 digestion from 500 µg of AIR biomass. *A. thaliana* was used as a standard (STD) for the digestion. WT: Sucargane Wild Type, EV: Empty Vector and *ScGUX2* (lines 2, 3, 4, 5, 9, 12, 15 and 10). The product of the digestion XUXX is highlighted in red and the bands which correspond to it were signalled by red arrows. The last band might have a mix of Ara and Xyl due to the use of GH62 enzyme. C) Quantification of % of xylose residues with GlcA using ImageJ. The error bars correspond to standard deviation of three replicates.

After screening all the transgenic lines to check for the potential GlcA reduction, the three lines with highest reduction (Figure 25C): *ScGUX2.*10, *ScGUX2.*12 and *ScGUX2.*15 were chosen for further analysis. As explained previously, the products expected from the digestion with GH11 plus GH62 for sugarcane are xylotetraose (X**U**XX), xylotriose ($Xyl_3$), xylobiose ($Xyl_2$) and xylose (Xyl). We wanted to use the GH11 + GH62 digestion to assay the degree of xylan glucuronidation across a larger number of biological replicates for the selected lines.

For each biological replicate, three technical replicates were digested and analysed separately (Figure 26A and 26B). Using the ImageJ, we are able to quantify the amount of

glucuronidation in each digestion (Figure 26C). Our analysis demonstrates that the reduction of GlcA in the silenced lines is effective and consistent across many biological replicates. For lines *ScGUX2*.10, *ScGUX2*.12 and *ScGUX2*.15 the average amount of glucuronidation was 0.5%, 0.7% and 0,6% respectively, while the average amount for WT and EV were 4% and 3.4% respectively.

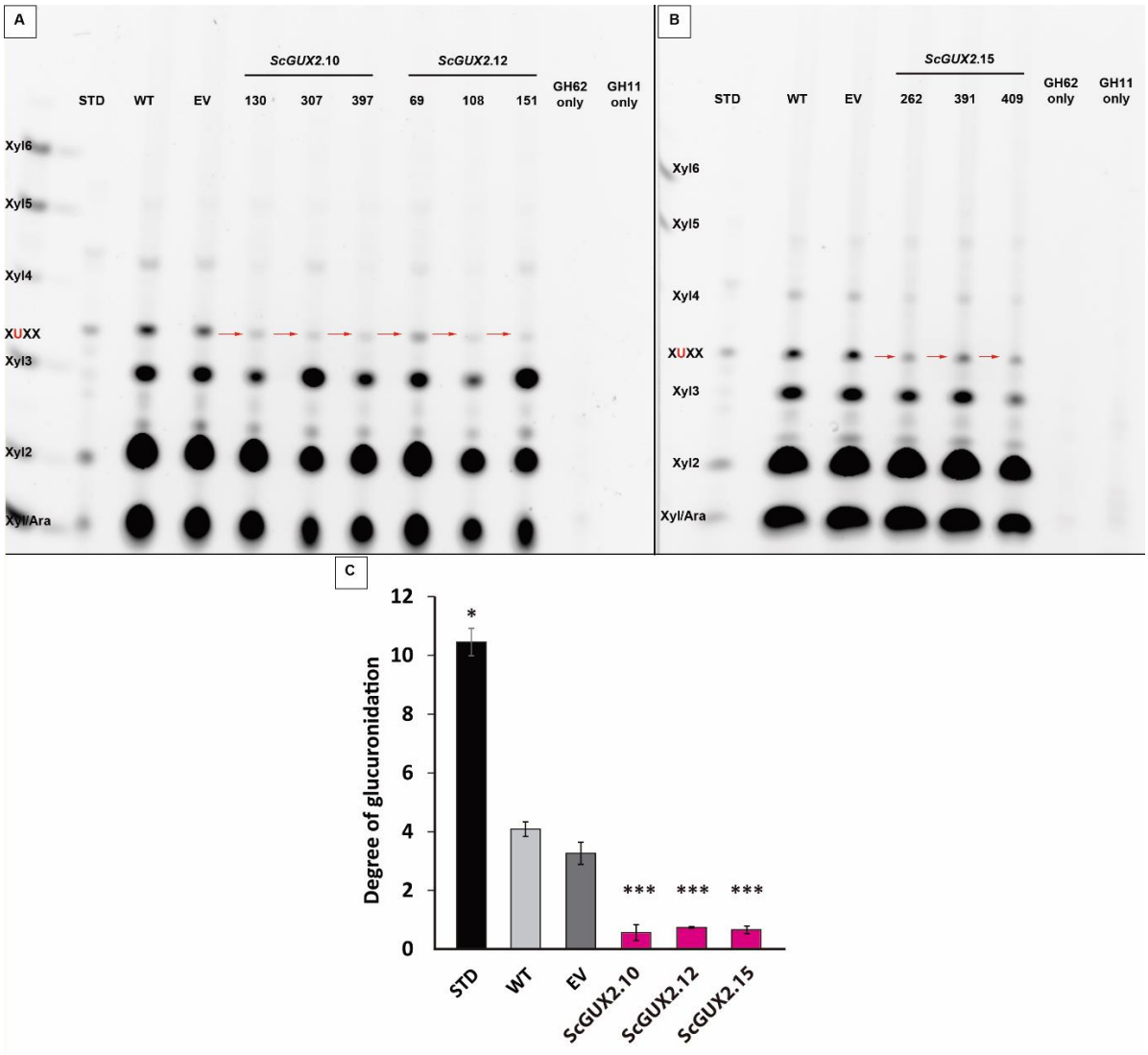



**Figure 26 Analysis of the degree of glucuronidation on xylan from the three independent events with higher GlcA reduction.** PACE analysis of arabinofuranosidase GH62 plus xylanase GH11 digestion from 500 µg of AIR biomass. *A. thaliana* was used as a standard (STD) for the digestion. WT: Sucargane Wild Type, EV: Empty Vector and silenced lines *ScGUX2* (lines 10 and 12 (A) and lines 15 (B)). The product of the digestion XUXX is highlighted in red and the bands which correspond to it were signalled by red arrows. C) Quantification of % of xylose residues with GlcA using ImageJ. The error bars correspond to standard deviation of three replicates. The last band might have a mix of Ara and Xyl due to the use of GH62 enzyme. *p-value<0.05, **p-value< 0.001, ***p-value<0.0001.

To further analyse the transgenic lines demonstrating the greatest reduction in the degree of xylan glucuronidation, a qPCR was performed in the three silenced lines to check if the phenotypical results were directly proportional to the gene expression level (Figure 27). The *ScGUX2* expression was reduced approximately five times in *ScGUX2*.10, ten times in *ScGUX2*.12 and seven times for *ScGUX2*.15 lines compared to WT expression levels. Since *ScGUX2.12* had the greatest reduction in the degree of xylan glucuronidation, yet the gene expression was reduced the least for this line, the amount of reduction of *ScGUX2* activity is not strictly related with the reduction of its gene expression. This difference between most silenced lines and phenotypic trait was also observed in sugarcane silenced *BAHD* lines, where the lines 3 and 4 with biggest silencing level were posteriorly discarded, because they presented characteristics not that relevant as the other three silenced lines analysed with less silencing level (de Souza et al., 2019). However, it is important to remember that the *ScGUX2* expression is clearly reduced in all the lines showing the structural phenotype on xylan. Therefore, expression profiling can help as an additional way to screen the silencing over many transgenic lines.

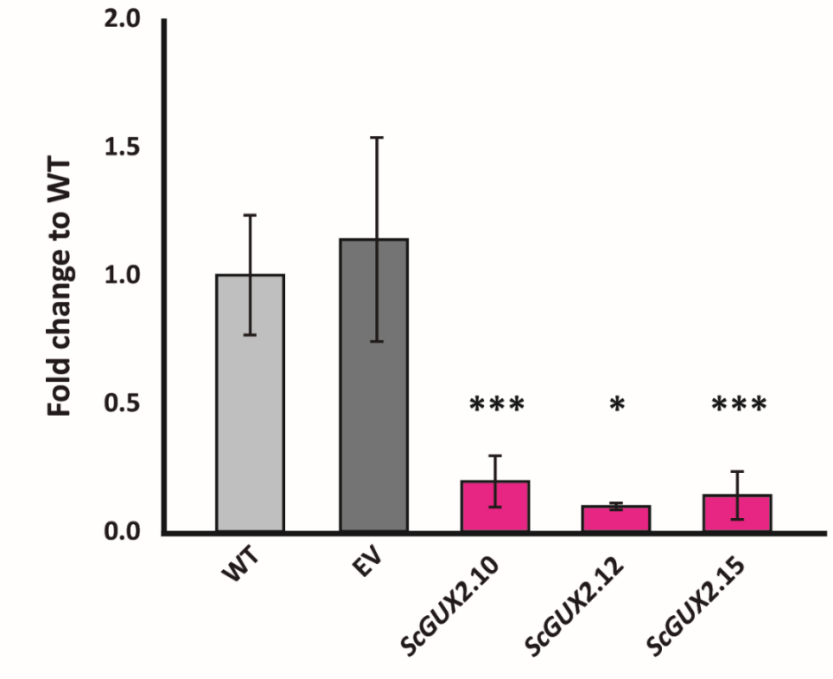



**Figure 27 Sugarcane selected silenced lines had a significantly reduction in *ScGUX2* expression**. Relative expression to WT from the three lines with highest reduction in GlcA. Error bars show standard deviation of three technical replicates and at least three biological replicates. *p-value<0.05, **p-value<0.005, ***p-value<0.0005 (pairwise T-test).

### *6. ScGUX2* events show a reduction in recalcitrance and increase in saccharification

Since there was a reduction in the degree of glucuronidation and gene expression in the transgenic lines with *ScGUX2* expression silenced, we decided to check any possible changes in the biomass recalcitrance related with modifications on xylan decorations. Previous experiments in Arabidopsis (Lyczakowski et al., 2017) indicated that removal of GlcA from xylan, in *gux1/2* mutant, results in significant increase in both glucose and xylose release form Arabidopsis stem biomass. We performed saccharification experiments on mature culms (internodes 10 and 11) of sugarcane plants (WT, EV and *ScGUX2* silenced lines) (Figure 28) using a commercial saccharification enzymatic cocktail Cellic CTec2. We have not used any pre-treatment on our biomass and measured the release of glucose (Figure 28A) and xylose (Figure 28B) from the plant material.

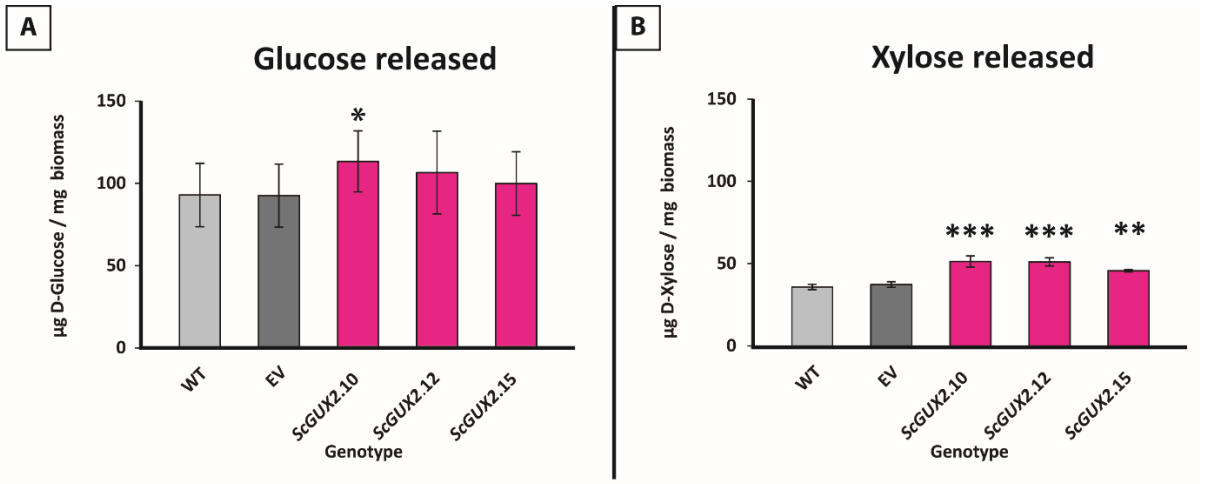



**Figure 28 The reduction of xylan glucuronidation in sugarcane results in reduced recalcitrance and increased saccharification.** Glucose (A) and Xylose (B) release from WT, EV and silenced events biological replicates following 24h of saccharification. Saccharification was performed using Cellic CTec2®. The error bar represents the standard deviation of three biological replicates of biomass. Error bars represent standard deviation of three biological and three technical replicates of biomass, *p value <0.07; **p value <0.01; ***p value <0.005(pairwise T-test)

For all the lines analysed, we can see an increase in the release of glucose (up to 22% more) and xylose (up to 43% more), which means that the recalcitrance of the silenced events is reduced in comparison to WT plants. In addition to that, the amount of lignin was measured by Dr. Pedro Araújo and we obtained the following results (Figure 29).

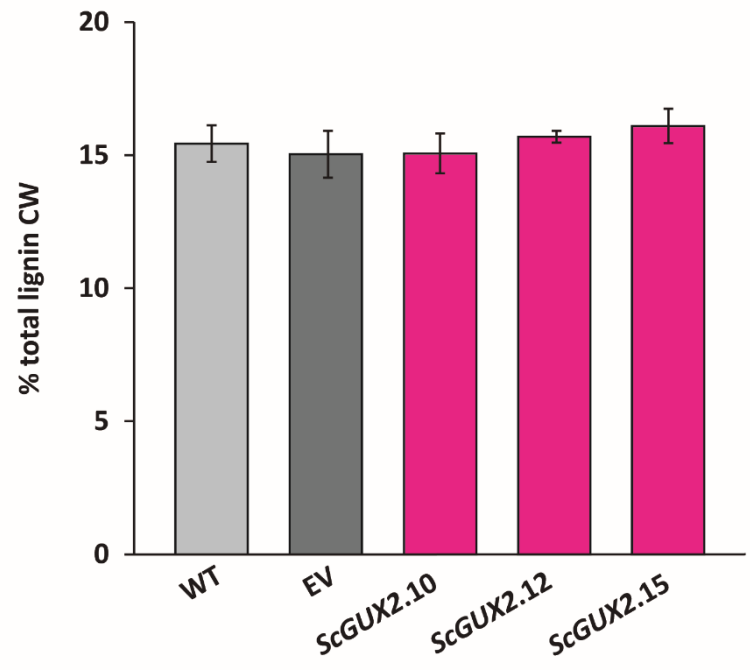



**Figure 29 Amount of lignin of WT, EV and *ScGUX2-silenced* sugarcane samples.** Pairwise T-test was performed and there is no significantly difference in lignin amount for all the samples. The error bar represents the standard deviation of three biological replicates of biomass. Error bars represent standard deviation of three biological and three technical replicates of biomass.

There is no difference in lignin amount among all the samples tested, probably because the silence of *ScGUX2* in sugarcane does not affect the final amount of lignin inside the culm of the plant.

To go further in the understanding of lignin in the material, we started a collaboration with Prof. Dr. Juliana Mayer (University of Campinas), whom analysed internodes 3 and 7 of the sugarcane stems to try to see any possible difference in its cellular development (Figure 30).

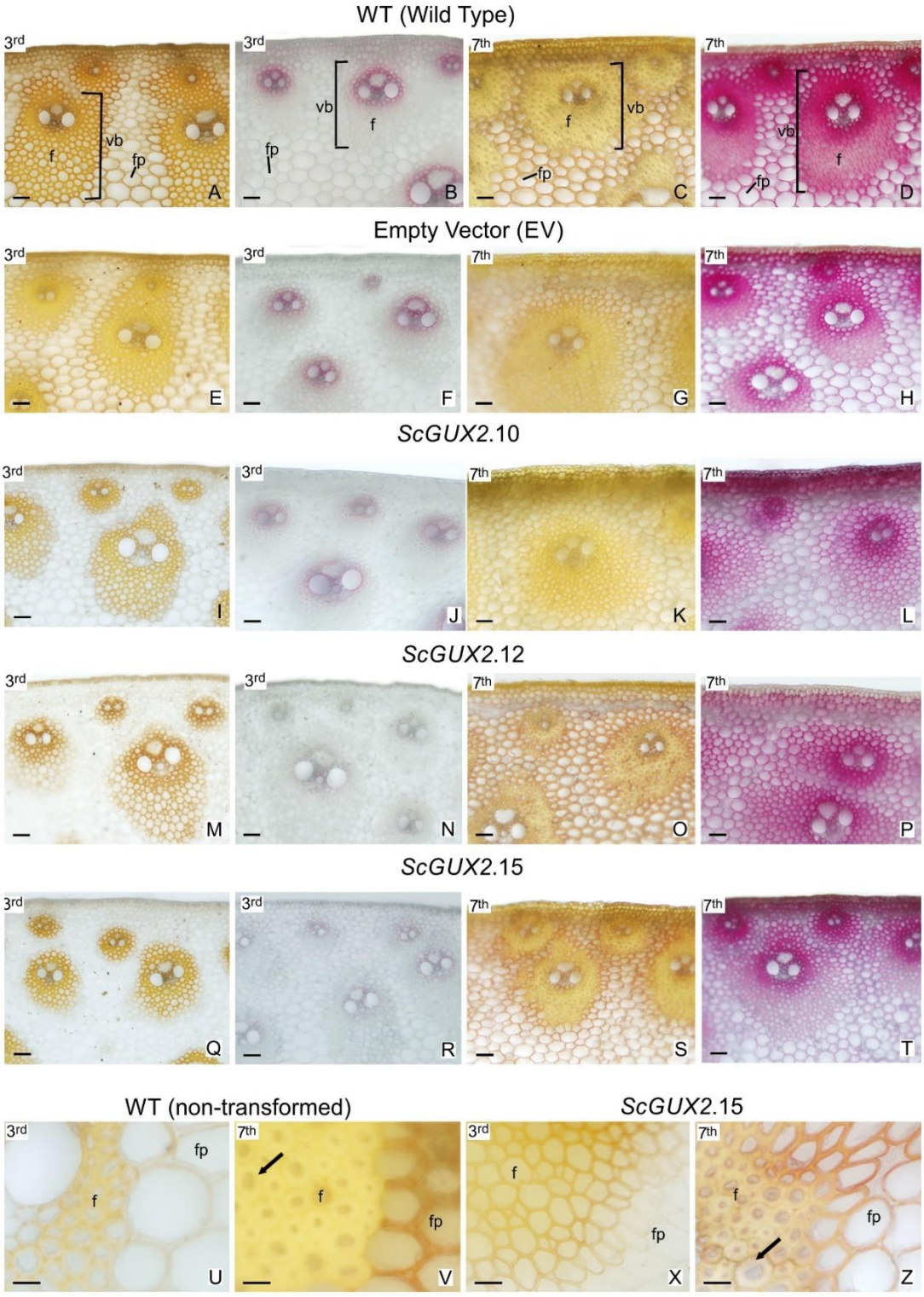



**Figure 30: Histochemical analysis of the *ScGUX2*-silenced line, Empty Vector (EV) and Wild Type (WT) plants.** Transverse sections of peripheral region of the culm in the third (immature internode) and seventh internode (mature internode) submitted to Maüle reaction for detection of lignin S and G (A, C, E, G, I, K, M, O, Q, S, U-Z) and stained with phloroglucinol-HCl for total lignin (B, D, F, H, J, L, N, P, R, T). U-Z. Detail of fibers and fundamental parenchyma. f = fibers; fp = fundamental parenchyma; vb = vascular bundle; arrow = fiber cell wall. Scale bars: A-T = 50 µm; U-Z = 20 µm.

In all samples analysed, the tissue in the third internode still is in differentiation stage. The histochemical tests with Maüle and Phloroglucinol-HCl reagents showed different results. Maüle reagent indicated on the third internode on the peripheric region of the culm that, for WT and EV, the fibers and tracheary elements from vascular bundle and fundamental parenchyma cells are lignified, showing a golden coloration, which means accumulation of H and/or H/G units (Figure 30 A, E). While for *ScGUX2* plants only vascular bundle cells showed lignification (Figure 30, I, M, Q). Despite the fact that phloroglucinol-HCl dyes total lignin from plant cell wall in a range from red to pink, the positive result was pointed in WT and EV only in a few fibers and tracheary elements of protoxylem and metaxylem of the same area analysed (Figure 30 B, F), and in tracheary elements in *ScGUX2* plants (Figure 30 F, J, R).

In seventh internode all the tissue is already differentiated, being possible to verify independently of the test used that the fibers inside the vascular bundle presents thicker cell wall in WT and EV (Figure 30 C-D, G-H), while in *ScGUX2* they present thinner cell walls (Figure 30 K-L, O-P, S-T, Z). It worth to note the difference of thickness of the cell wall of fibers in the details of figure 30 U-Z. On this internode the fibers present golden coloration, while the parenchymal cells a red-brownish coloration with Mäule reagent, indicating the presence of S units.

In the culm, sections of central region of third internode, the tissue is in the differentiation stage, but is possible to observe that in WT and EV samples, the parenchymal cells are already expanded and with lignified cell wall (Figure 31 A, E). While in *ScGUX2* samples they are less differentiated (Figure 31 I, M, Q). On the seventh internode the coloration of parenchymal cell walls of WT and EV are more intense and present a red-brownish coloration (Figure 31 C-D, G-H, arrow). On *ScGUX2* plants, the same parenchymal cell walls are thinner and with a low amount of lignification (Figure 31 K-L, O-P, S-T, arrow). Similar to that found in the peripheral region, the vascular bundle fibers in the central region have the thinnest cell wall in *ScGUX2* individuals (Figure 31 L, P, T) when compared to WT and EV individuals (Figure 31 D, H).

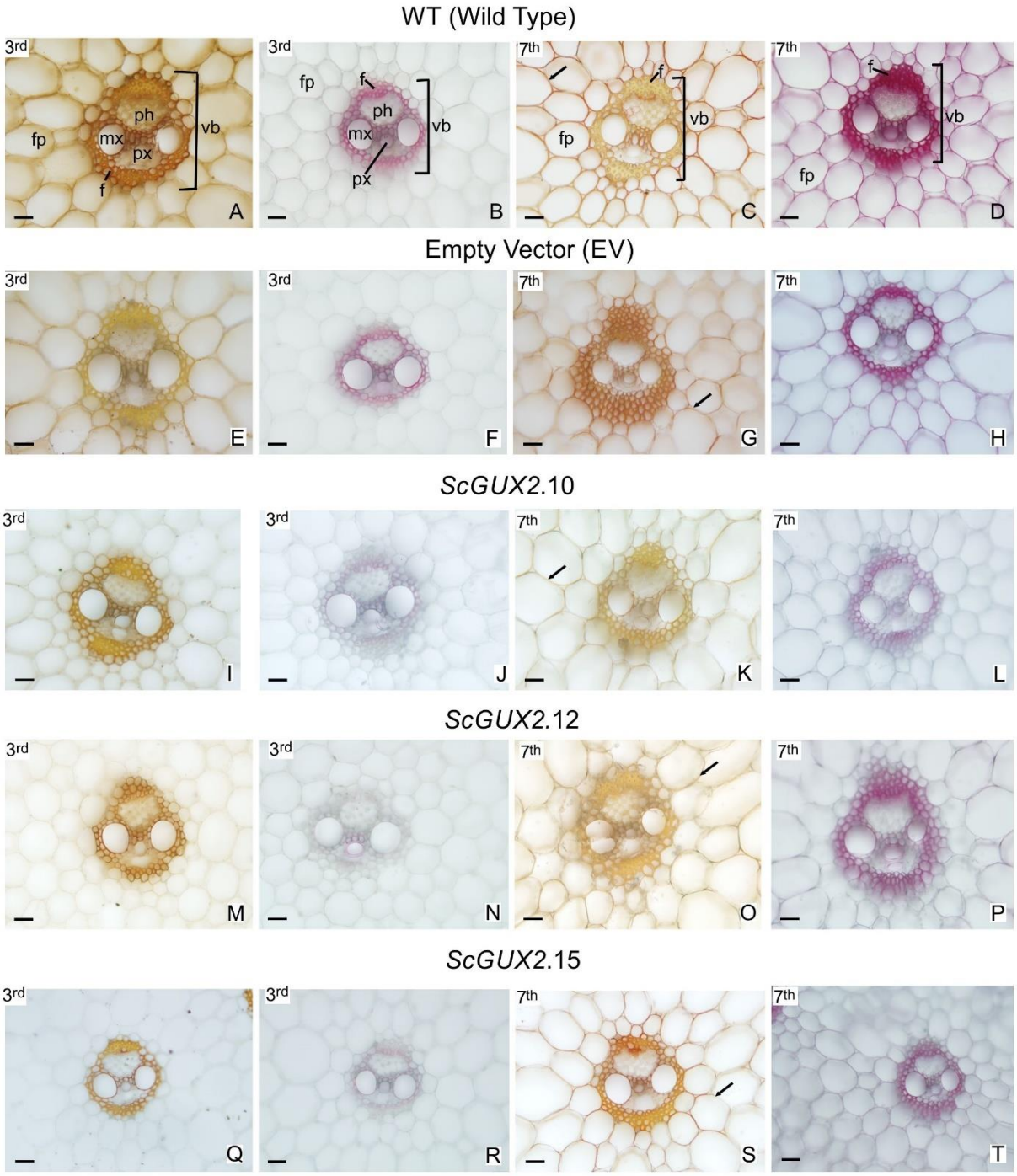



**Figure 31: Histochemical analysis of the *ScGUX2* transgenic line, Empty Vector (EV) and Wild Type (WT) plants.** Transverse sections of central region of the culm in the third (immature internode) and seventh internode (mature internode) submitted to Maüle reaction for detection of lignin S and G and stained with phloroglucinol-HCl for total lignin. Third and seventh internode: Left Columns = Maüle reaction and Right columns: phloroglucinol-HCl. f = fibers; fp = fundamental parenchyma; (mx) tracheary elements of metaxylem; (ph): phloem; (px) tracheary elements of protoxylem; vb = vascular bundle; arrow = fundamental cell wall. Scale bars: A-T = 50 µm.

We would like to know why the recalcitrance is reduced in sugarcane biomass with reduced levels of xylan glucuronidation. It is possible that similarly to what was suggested in dicots (Giummarella & Lawoko, 2016), the GlcA might be involved in lignin binding. Reduction in the amount of this lignin carbohydrate complex (LCC) in *gux* mutants was suggested to be responsible for the decrease in biomass recalcitrance (Lyczakowski et al., 2017). Alternatively, it is possible that xylan-cellulose interaction changes in plants with reduced amount of xylan glucuronidation. This could be due to changes in the pattern of GlcA branches in the *ScGUX2* silenced lines, which was previously observed to influence the xylan-cellulose interaction (Grantham et al., 2017). To investigate this hypothesis, we would like to study the pattern of GlcA in the plants and see if the capacity of xylan to bind cellulose changes upon *ScGUX2* silencing.

## 7. Pattern of GlcA branches on xylan added by ScGUX2

In order to make easier to interpret the results below, I have done another scheme with the use of GH30 and the expected results (Figure 32). To determine any regularity in the position of the glucuronic acid side chains on the xylan backbone, we performed a digestion of sugarcane AIR with glucuronoxylanase BoGH30 (Nzytech, Figure 32A). Xylanase GH30 can digest the xylan backbone only when the GlcA branch is present so the degree of polimerisation of the resulting oligosaccharides matches the distance between subsequent decorations (Bromley et al., 2013), followed by GH62 (Figure 32B) to remove the arabinoses and make it easy to interpret the results. In our experiments we have digested deacetylated AIR from *ScGUX2* silenced lines and from WT and EV controls. The resulting oligosaccharides were derivatized with ANTS and analysed on PACE (Figure 32C and 32D).

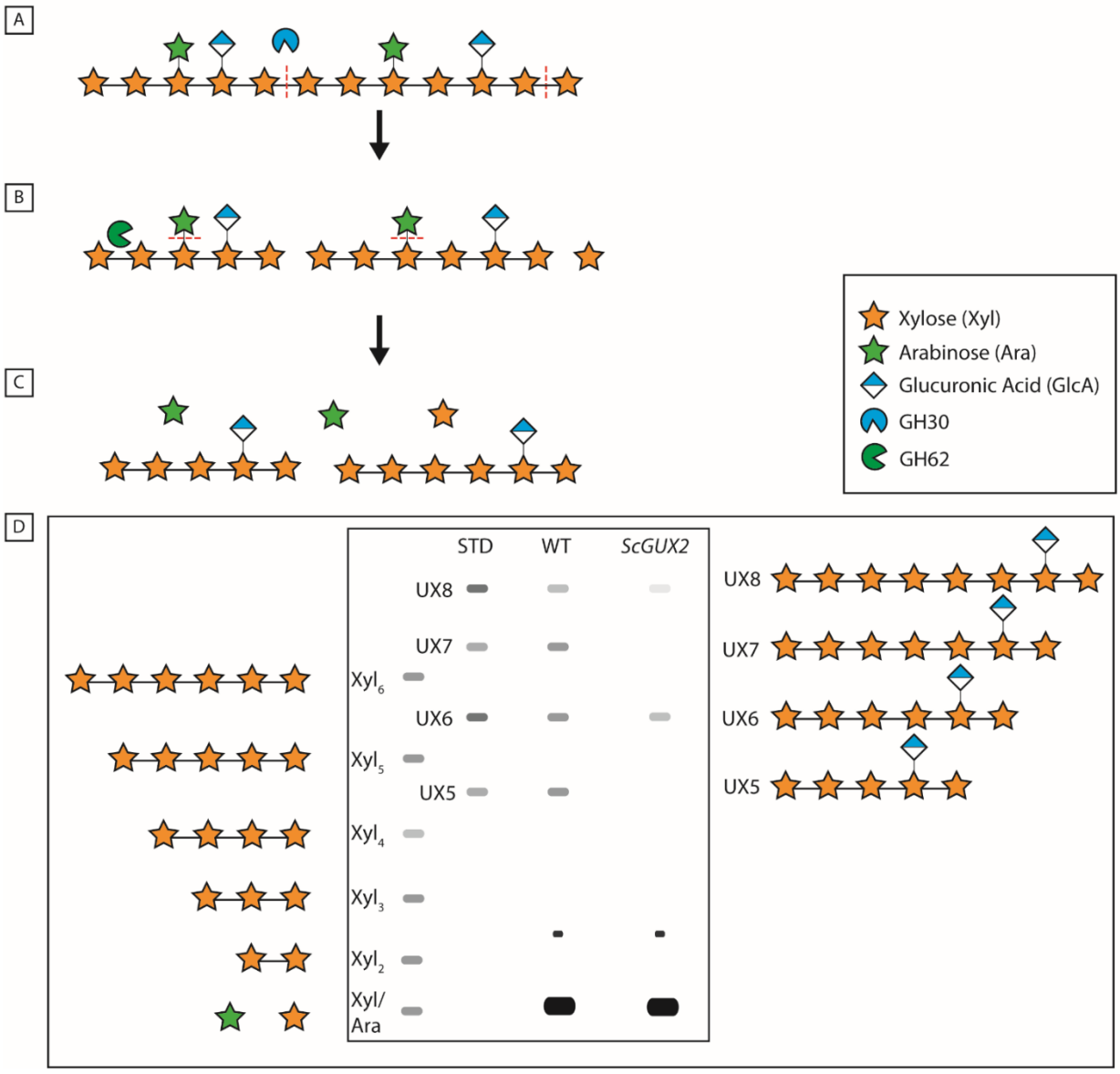



**Figure 32 Scheme of xylan digestion with GH30 and GH62 followed by a PACE gel result.** To avoid any possible star activity of long period with GH30, we performed 30 min of GH30 (A), killed the enzyme at 100°C for ten minutes and then added GH62 (B) to remove arabinoses, resulting in the products (C): xylopentaose (UX5), xylohexaose (UX6), xyloheptaose (UX7) and xylooctaose (UX8) and xylose (Xyl) plus arabinose (Ara). (D) It is a representation of the result after running the digestions in a PACE gel, where we can see a column which represent the xylose ladder ($Xyl_6$, $Xyl_5$, $Xyl_4$, $Xyl_3$, $Xyl_2$ and Xyl). The next three columns in the gel correspond to a standard (STD), wild type sugarcane (WT) and a sugarcane silenced line *ScGUX2*.

Following the scheme of an expected result from GH30 digestion, below we can see the real results (Figure 33).

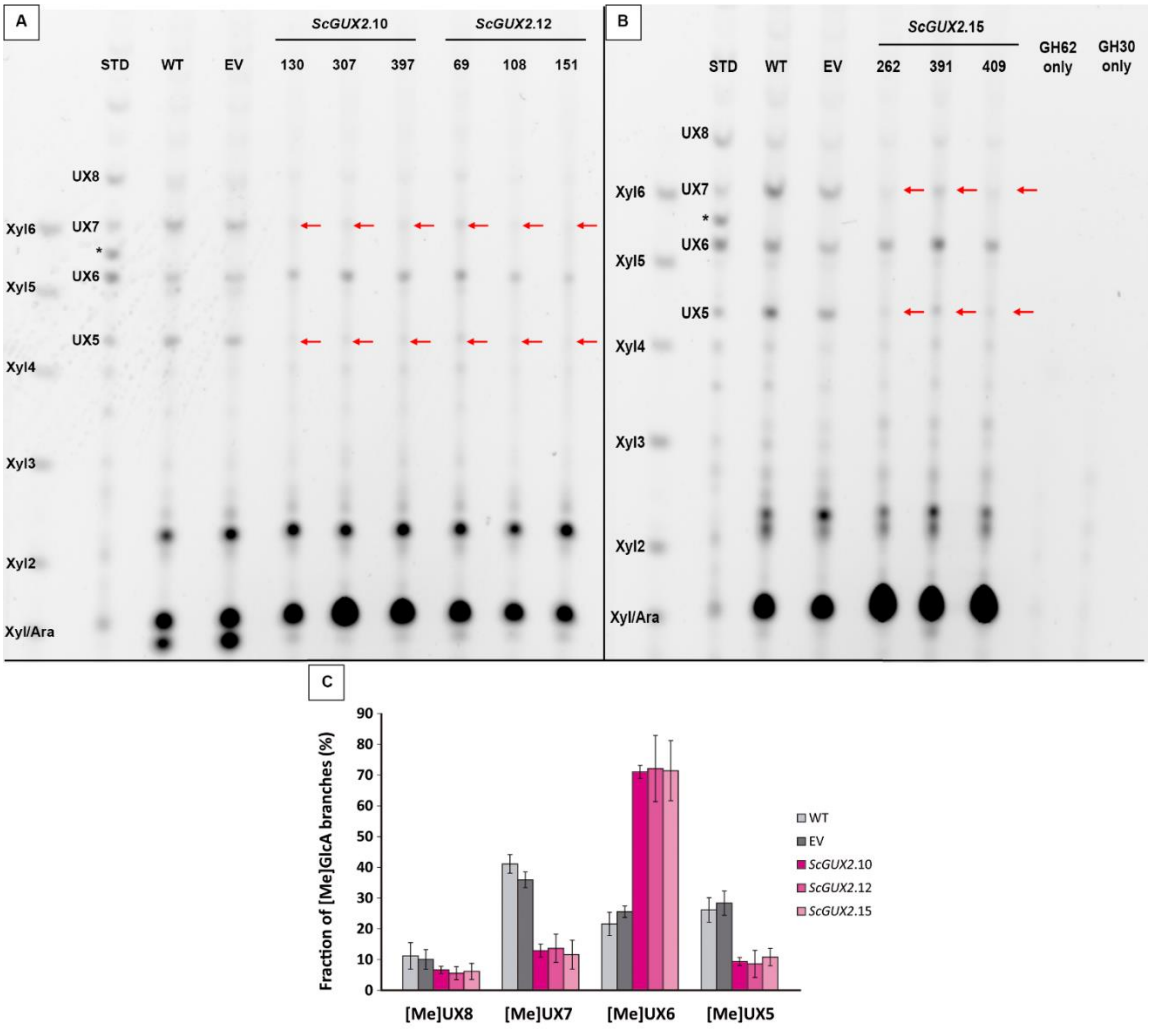



**Figure 33 [Me]GlcA distribution on xylan in sugarcane samples.** *A. thaliana* was used as standard (STD), WT sugarcane (WT), Empty Vector (EV) and *ScGUX2* silenced independent lines (lines 10, 12 (A) and 15(B)). The red arrows highlight the bands that correspond to UX5 and UX7. 1 mg of AIR material was hydrolysed with glucuronoxylanase BoGH30, followed by arabinofuranosidade GH62 analysed by PACE. C) [Me]GlcA pattern quantitation on xylan in sugarcane samples. Error bars show standard deviation of three replicates. * PUX structure

After the digestion of WT and EV sugarcane biomass with BoGH30 and GH62, we can see a ladder of oligosaccharides from UX5 up to UX8. In this digestion UX5, UX6 and UX7 are the most abundant products. In the silenced *ScGUX2* lines the UX6 oligosaccharide is dominant across all the lines and a reduction in the amount of UX5 and UX7 (highlighted by red arrows), compared to WT and EV, is evident. These results suggest that *ScGUX2* has a preference for adding GlcA every five and seven xyloses of the xylan backbone and in its absence this specific pattern is reduced. Which might suggest that the silencing of *ScGUX2* affect mainly the

incompatible domain, remaining the compatible domain made mostly by UX6 keeping the interaction of xylan with cellulose without any loss to the plant cell wall.

## 8. *ScGUX1* and *ScGUX2* complementation in Arabidopsis have different activities in xylan patterning

We can also see that the total amount of UX6 is almost the same when *ScGUX2* is silenced as in the WT and EV lines, which suggests that another *ScGUX* is responsible for this patterning. In order to establish if *ScGUX1* is responsible for this patterning, we want to understand in the end what is the specific activity of the two sugarcane GUX enzymes. To study that we performed BoGH30 digestion on Arabidopsis *gux1/*2 expressing *ScGUX1* or *ScGUX2*.

As a way to try to understand which *ScGUX* was responsible for the UX6 patterning, which was the main remaining pattern type in lines with *ScGUX2* silenced, and in order to get stronger evidences that *ScGUX2* has a preference for adding GlcA onto every 5$^{th}$ and 7$^{th}$ xylose of the xylan, BoGH30 digestion was performed on biomass from the *gux1/2* Arabidopsis complemented with *ScGUX1* or *ScGUX2* (Figure 34).

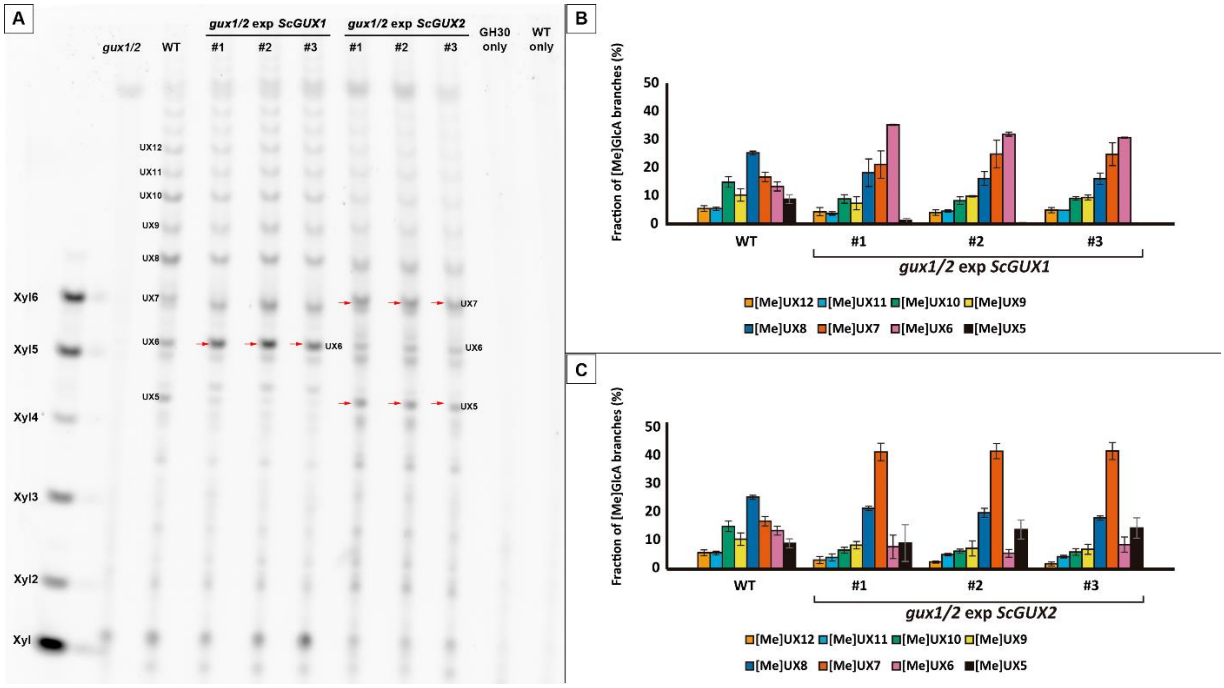



**Figure 34 A) PACE from BoGH30 digestion from Arabidopsis *gux1/2* complemented lines with *ScGUX1* or *ScGUX2*.** [Me]GlcA distribution on xylan in *A. thaliana* samples: *gux1/2*, WT: Wild Type, *ScGUX1*: *gux1/2* expressing *ScGUX1* and *ScGUX2*: gux1/2 expressing *ScGUX2*. One mg of AIR material was hydrolysed with glucuronoxylanase BoGH30 and analysed by PACE. [Me]GlcA pattern quantitation on xylan in gux1/2 plants expressing *ScGUX1* (B) or *ScGUX2* (C). Error bars show standard deviation of three replicates.

When we digest AIR from WT Arabidopsis with BoGH30, the oligossacharides will have a GlcA on it, resulting mainly in the following products: UX5, UX6, UX7, UX8, UX9, UX10, UX11 and UX12. Since BoGH30 requires the presence of GlcA to perform the digestion, when we add BoGH30 to *gux1/2* biomass, which contain no GlcA there is no digestion and consequently no products (*gux1/2* – first well from figure 34A).

Based on the previous observation, we can see that *gux1/2* expressing *ScGUX1* can recover the patterning of GlcA deposition from UX6 to UX12, but it has a preference for adding GlcA every six xyloses (UX6 bands are highlighted by red arrows in Figure 34A and are represented as pink bars in Figure 34B), while *gux1/2* expressing *ScGUX2* can recover the patterning of GlcA deposition from UX5 to UX12. In addition to that *ScGUX2* has a preference to add GlcA every five and seventh xyloses (UX5 and UX7 are highlighted by red arrows in Figure 34A and represented as black and orange bars respectively in Figure 34C) as we have seen previously in sugarcane *ScGUX2* silenced lines (Figure 33A/B).

## 9. ScGUX1 and ScGUX2 enzymes perform different activities in different acetylated xylan acceptors

We wanted to further investigate the activity of ScGUX1 and ScGUX2 to understand if the enzymes have preference towards type of substitutions present on the backbone of the xylan acceptor. To do that we have prepared three different heteroxylan acetylated acceptors (*gux1/2*, WT sugarcane and *ScGUX2* sugarcane), incubated with either *ScGUX1* or *ScGUX2* (with [+] or without [-] UDP-GlcA) and digested reaction products with GH11 plus GH62. The digestion products were analysed on PACE. In the analysis four main bands: X**U**XX, xylotriose ($Xyl_3$) xylobiose ($Xyl_2$) and xylose (Xyl) were observed (Figure 35).

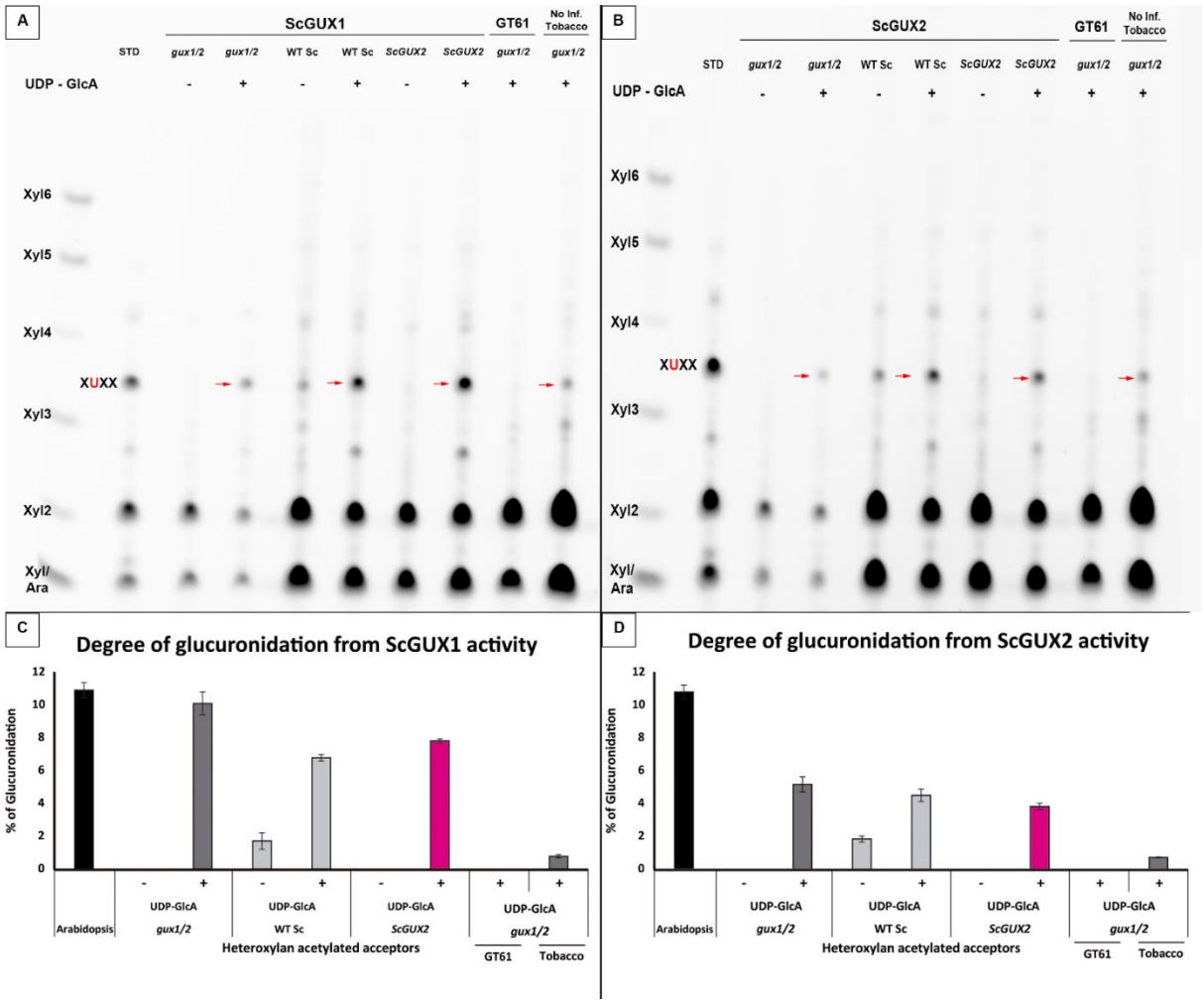



**Figure 35 ScGUX1 (A) and ScGUX2 (B) have xylan glucuronosyltransferase activity *in vitro* and different activity in different acceptors.** The assays were performed with UDP-GlcA, acetylated xylan without Me[GlcA] decorations from Arabidopsis *gux1/2* mutants, WT sugarcane, *ScGUX2* sugarcane silenced events and microsomes from *N. benthamiana* enriched for ScGUX1 (A) or ScGUX2 (B), control GT61 protein and Non

infiltrated tobacco. Products of the *in vitro* activity reaction were digested with β-xylanase GH11 (Megazyme) and analysed by PACE. The enzyme generates xylose, xylobiose, xylotriose (only in sugarcane) plus the UX4 (highlighted in red), oligosaccharide if any GlcA is present on the xylan. The positive (+) symbol means that 10mM of UDP-GlcA was added to the assay, while the negative (-) symbol means that no UDP-GlcA was added. Arabidopsis thaliana was used as a Standard (STD) for the digestion. Quantification of % of xylose residues with GlcA using ImageJ for ScGUX1 (C) and ScGUX2 (D) in vitro activity.

From Figure 35, we can see that when added UDP-GlcA with ScGUX1 to the acetylated xylan acceptors (*gux1/2*, WT sugarcane or *ScGUX2*), *ScGUX1* can add GlcA to all of them, including to the WT sugarcane, which increases the percentage of glucuronidation. This applies to *ScGUX2*, but when we compare ScGUX2 activity with that of ScGUX1 the amount of GlcA it deposits is reduced for all xylan acceptors. It is important to highlight that the microsomes from tobacco leaves not infiltrated with Agrobacterium and those expressing a control GT61 enzyme both were able to add a small amount of GlcA onto the xylan backbone. This activity might come from endogenous activity from GUX that are present in tobacco plants.

As an alternative way to test the specificity of both enzymes, we performed BoGH30 on *in vitro* products of ScGUX1 and ScGUX2 glucuronidation on the same set of acceptors and analysed resulting oligosaccharides with PACE (Figure 36).

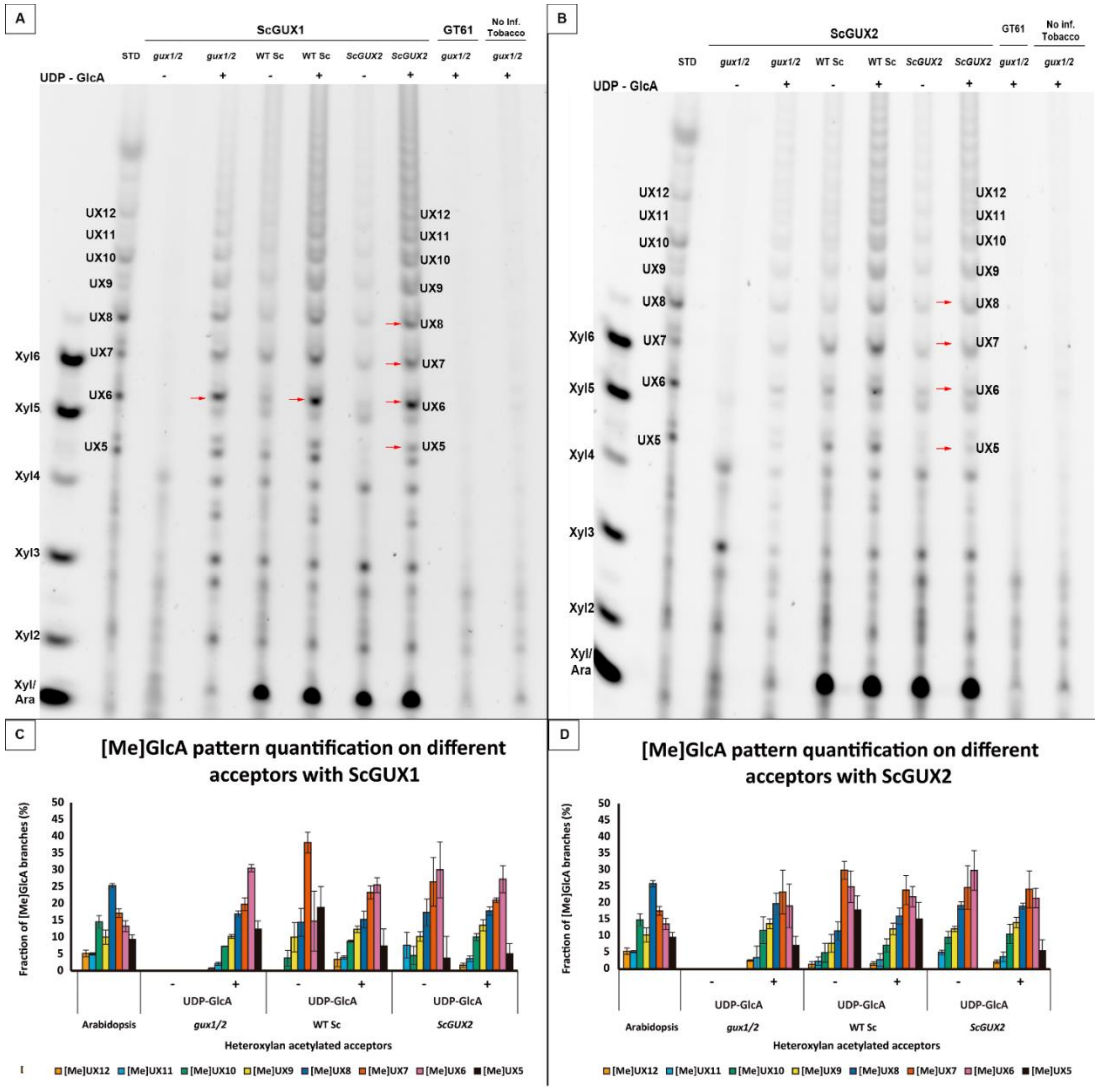



**Figure 36 *In vitro* activity of ScGUX1 (A) and ScGUX2(B) on different heteroxylan acetylated acceptors.** [Me]GlcA distribution on different xylan acceptors: *A. thaliana*: *gux1/2*, and sugarcane acceptors: WT Sac: Wild Type sugarcane and *ScGUX2*: RNAi *ScGUX2* lines. 1 mg of AIR material was hydrolysed with glucuronoxylanase BoGH30 and analysed by PACE. The assays were performed with UDP-GlcA, acetylated xylan without Me[GlcA] decorations from Arabidopsis *gux1/2* mutants and acetylated xylan from WT sugarcane and *ScGUX2* sugarcane silenced events and microsomes from *N. benthamiana* enriched for ScGUX1 (A), ScGUX2 (B), control GT61 protein and Non infiltrated tobacco. Products of the in vitro activity reaction were digested with BoGH30 (Nzytech) plus GH62 and analysed by PACE. The positive (+) symbol means that 10mM of UDP-GlcA was added to the assay, while the negative (-) symbol means that no UDP-GlcA was added. *A. thaliana* was used as a Standard (STD) for the digestion. [Me]GlcA pattern quantification on xylan from ScGUX1 (C) or ScGUX2 (D). Error bars show standard deviation of three replicates.

We can see that both ScGUX1 and ScGUX2 are able to add GlcA from UX5 to UX12 in all the acetylated xylan acceptors chosen (*gux1/2*, WT sugarcane and *ScGUX2*). Nonetheless for ScGUX1 there is still a preference for adding GlcA every six xyloses, which gives stronger

evidences in addition with sugarcane silenced lines (Figure 33) and *gux1/2* complementation assay with ScGUX1 (Figure 34A) that ScGUX1 has a preference for adding GlcA every six xyloses, while for *ScGUX2* it is not clear if there is a preference for adding GlcA into the xylan looking only to this assay. These results are similar with what was suggested in studies with Arabidopsis that *AtGUX1* has a preference for adding GlcA into a major domain, also described as an even pattern, adding GlcA from UX6 to UX20, in a compatible domain, and *AtGUX2* has a preference for adding GlcA into a minor domain, also described as an odd pattern, adding GlcA mainly in UX5, UX6 and UX7 in an incompatible domain (Bromley et al., 2013; Simmons et al., 2016).

# CHAPTER 4: DISCUSSION

This thesis sought to identify and characterise genes encoding sugarcane GUX enzymes. The aim of the work was to establish:

- The number of genes encoding GUX enzymes in the sugarcane genome and the phylogenetic classification of sugarcane GUX with respect to other known xylan glucuronosyl transferases.
- If the sugarcane GUX enzymes are able to transfer the GlcA onto the xylan acceptor *in vitro* and *in vivo* and what pattern of GlcA decorations do they generate.
- Phenotypes of sugarcane plants with reduced degree of xylan glucuronidation.

*GUX* genes comprise a multigene family, with five homologous genes annotated in the Arabidopsis genome (*AtGUX1-5*; Mortimer et al. 2010; Rennie et al. 2012), and at least one gene in the conifer *Picea glauca* (*PgGUX*; Lyczakowski et al. 2017). Recently *GUX* genes were also localized in other crops such as populus (*Populus trichocarpa,* Porth et al., 2018), eucalyptus (*Eucalyptus grandis,* Wierzbicki et al., 2019), rice (*Oryza sativa,* Gao et al., 2020) and in several other plants and crops of economic importance, being the one of most importance for this thesis, the sugarcane (*Saccharum spp,* Gallinari et al., 2020).

We have done an analysis using all the data available from transcriptome for each of the crops analysed with a more reliable method to infer the number of *GUX* genes for each species. From 18 plant species, seven of them (*Brachypodium distachyon*, *Setaria italica, Solanum tuberosum*, *Theobroma cacao*, *Arabidopsis thaliana*, *Sorghum bicolor* and *Saccharum* spp) have five GUX proteins in their genome, whereas five species have more than five orthologs: *Zea mays* and *Brassica oleraceae* have seven GUX, *Brassica rapa* has 10, *Glycine max* has 11, and *Populus trichocarpa* has six. On the other hand, six out of 18 species have less than five GUX: *Eucalyptus grandis* has four GUX proteins, *Oryza sativa*, *Vitis vinifera* and *Citrus sinensis* have three, and both *Marchantya polymorpha* and *Physcomitrella patens* have only one GUX.

All the *GUX* genes identified were aligned and surveyed through phylogenetic methods and arranged in different clades, based on the previously identified and characterized *Arabidopsis thaliana* GUX *genes* (*AtGUX1-5*; Mortimer et al. 2010; Rennie et al. 2012). We were able to arrange them into six main clades: GUX1, GUX2, GUX3 (all of them have both monocot and dicot subdivision) GUX4/5 brassicaceae and GUX X monocot. After the classification of all the GUX in clades, we identified the sugarcane *GUX* genes to start to understand their functions. From figure 16, we observed that GUX2 species generate a monophyletic clade, including both monocots and dicots plants, making it easy to identify the sugarcane *GUX2* orthologue from Arabidopsis. While for GUX1, a process called concerted evolution happened together with GUX3 species, forming two paraphyletic clades, one containing GUX1 and GUX3 dicot species and the other one with GUX1 and GUX3 monocot species. Inside the GUX1 monocot clade, we found the sugarcane *GUX1* orthologue from Arabidopsis. It is worth to mention that the phylogenetic analysis was the first step done before all the work in this thesis and we had some difficulties in identifying the whole sequence of sugarcane *GUX* until the second half of my PhD, which were solved when the sugarcane genome became available (Garsmeur et al., 2018). With the full sequences of the identified and correspondent *ScGUX1* and *ScGUX2* from our phylogenetic analyses (Gallinari et al. 2020) we were able to start working into the vectors to start to characterize them, using tobacco as an expression system.

Tobacco has been used to study the activity of many plant glycosyl transferases (GTs, eg. beta1-4 galactan synthase, Linawag et al., 2012; mannan galactosyltransferase: MAGT, Yu et al., 2018), and other GUXs (*At*GUX1-5, Rennie et al., 2012; *Pg*GUX, Lyczakowski et al*.,* 2017) and it is considered a reliable method to analyse plant GTs. The expression of *ScGUX1* and *ScGUX2* in tobacco (Figure 17) showed one more time that such system is a viable way to analyse plant GTs. In addition to that, after purifying the microsomes and doing an *in vitro*

activity with Arabidopsis *gux1/2* mutants, we observed that both *ScGUX1* and *ScGUX2* are active glucosyl transferases and are able to add GlcA to the *gux1/2* acetylated xylan (Figure 18).

After confirming that both *ScGUX1* and *ScGUX2* worked on *in vitro* assay, we build new vectors under the control of the secondary cell wall specific Arabidopsis *IRX5* promoter with *ScGUX1* or *ScGUX2* CDS to transform Arabidopsis *gux1/2* mutants, which lacks xylan GlcA decoration to understand if ScGUX enzymes would have the same behaviour *in vitro* and *in vivo* assays. After obtaining homozygotes of both *gux1/2* expressing either *ScGUX1* or *ScGUX2* we could see that all lines recovered the lack of GlcA, showing that the complementation experiment (*in vivo*) worked. It is important to note that *ScGUX1* has a slightly higher activity than *ScGUX2*, with the first one presenting a degree of glucuronidation that varied between 8 to 10%, while the second one the amounts varied between 5 and 7%.

With these results we were able to identify and characterise both *ScGUX1* and *ScGUX2*, which are active glucuronosyltransferases that can complement the lack of GlcA phenotype of Arabidopsis *gux1/2* mutants *in vivo* on the xylan. In order to understand the specific activity of each ScGUX we performed the same assays, but digesting with BoGH30 instead of GH11, resulting in a specific pattern of decoration resulted from each enzyme role.

For ScGUX1, in both Arabidopsis complementation assays (Figure 34) there was a preference for adding GlcA every six xyloses. While for ScGUX2 we observed a preference for adding GlcA every five and seven xyloses (Figure 34), where these bands were predominant over the others. These results go according what had been observed previously for Arabidopsis GUX, where AtGUX1 has a preference for adding GlcA in an even pattern, producing a compatible xylan, while *AtGUX2* has a preference for adding GlcA in an odd pattern, generating an incompatible xylan (Bromley et al., 2013).

An important point to highlight is that there are few studies from literature that manage to produce transgenic sugarcanes with the main focus to understand the biochemistry behind it with the present depth and they are generally focused on expressing genes that confers any type of resistance to the plant against diseases (Wang et al., 2017; Gao et al., 2016; Riaz et al., 2020) or involves silencing/expressing lignin pathway genes to obtain

a higher saccharification (Jungh et al., 2012; Bewg et al., 2016; Poovaiah et al., 2016) without characterising the plant cell wall as performed in this work.

Another crucial point is that even though most studies have reported no phenotype in Arabidopsis *gux* mutants (Mortimer et al., 2010; Mortimer et al., 2015; Lyczakowski et al., 2017) there were isolated reports of some growth reduction in Arabidopsis mutants (Lee et al., 2012). It is possible that the phenotype may be linked to growth conditions and only manifest under specific stress cases.

In our work, sugarcane *ScGUX2* silenced plants had no agronomical differences, but had a specific biochemical phenotype, which was a decrease in GlcA that might made the xylan more accessible to the enzymes and resulted in an increase for saccharification. The increase in the release of glucose, which had up to 22% more and up to 43% more xylose released in comparison to WT plants, shows that *ScGUX2* plants could be of extreme importance for 2GE ethanol production industries. As observed in *gux1/2* Arabidopsis mutants, xylose is also the main sugar increased in saccharification (Lyczakowski et al., 2017) as in *ScGUX2* silenced lines. This shows the importance of [Me]GlcA decorations for recalcitrance of biomass, playing a critical role and having a significantly impact on release of xylose (Lyczakowski et al., 2017).

The silence of *ScGUX2* in sugarcane showed that this enzyme has a preference for adding GlcA every five and seven xyloses, where those bands basically disappeared from the gel (Figure 33A and B) and goes according to the complementation result previously discussed here for Arabidopsis (Figure 34), which corresponds to an incompatible pattern, without any harmful effect to the plant. The pattern is actually more compatible in these plants, so the xylan is likely to interact still very well with the cellulose.

Focusing on *in vitro* assays, ScGUX1 (Figure 36A) appears to be more active than ScGUX2 (Figure 36B). On the other hand, ScGUX2 apparently works in a constant way both on acetylated (Arabidopsis) and arabinosylated (sugarcane) xylan, while ScGUX1 appears to be acting better only on acetylated xylan. Since in arabinosylated (sugarcane) xylan ScGUX1 is less active than on acetylated xylan, this may suggest that the presence of arabinose hinder the activity of ScGUX1. This might suggest that ScGUX1 work before xylan arabinosylation by GT61s (Scheller & Rennie, 2013).

The result from Arabidopsis complementation with *ScGUX2* in addition with the results from sugarcane *ScGUX2* lines give us strong evidences to suggest that *ScGUX2* has a preference to add GlcA every five and seven xyloses maintaining a specific patterning, which corresponds to an incompatible pattern previously described in Arabidopsis by other authors (Bromley et al., 2013; Simmons et al., 2017)

Finally, but not less important, the observed saccharification phenotypes in the silenced *ScGUX2* lines might be explained by at least two hypotheses:

1. The xylan in the sugarcane plants interacts less well with the cellulose and this leads to an increase in saccharification. This hypothesis is unlikely to be correct because in the *ScGUX2* silenced sugarcane the xylan appears to be more compatible with cellulose interaction (GH30 results from the mutant in Figure 33 and *in* vitro activity from Figure 36). This hypothesis is further disproved by the observation that a nearly complete loss of xylan cellulose interaction in Arabidopsis *esk1* mutant (Grantham et al., 2017) does not result in increase in glucose or xylose release in this plant (Lyczakowski et al., 2017).
2. The GlcA binds to lignin (Terret & Dupree, 2018) and this is the linkage that is reduced in silenced *ScGUX2* plants. The loss of this linkage and the lack of GlcA is also proposed to affect lignin deposition (Reis & Vian, 2004), which may explain the difference in deposition of lignin between the 3$^{rd}$ and 7$^{th}$ internode (Figure 30 and 31). Also, it is possible that the incompatible xylan, made by *ScGUX2*, as we have presented in this work, is closer to lignin in grasses than the compatible xylan (Kang et al., 2019), so the loss of *ScGUX2* may give stronger phenotypes in sugarcane than it would give in single mutants of Arabidopsis, for example.

# CHAPTER 5: CONCLUSION AND FUTURE WORK

- Sugarcane genome presents four GUX proteins, classified among five clades: GUX1, GUX2, GUX3 and GUX X.
- From the four GUX proteins, we were able to identify and characterize two of them: ScGUX1 and ScGUX2.
- Both ScGUX1 and ScGUX2 are active glucuronosyl transferases and able to add GlcA to xylan.
- ScGUX1 has a preference for adding GlcA every six xyloses (compatible xylan), while ScGUX2 has a preference for adding GlcA every five and seven xyloses (incompatible xylan)
- Silenced *ScGUX2* presented no adverse growth phenotype and a significantly increase in saccharification of both glucose and xylose.

The reduction of *ScGUX2* in sugarcane plants caused a decrease in the glucuronidation over the xylan, which resulted in a biomass more accessible to enzymes and consequently higher yield in ethanol production. These transgenic lines could represent a significantly increase for the second-generation ethanol industry. It might still be early to make assumptions, but with the increase in the release of fermentable sugars and consequently higher 2GE production those silenced lines might justify a decrease in the pre-treatment or even the absence, actually the biggest bottleneck in 2GE production.

Three new sugarcanes, considered elite varieties are being transformed under the same construction pGVG: *ScGUX2*_RNAi and will be evaluated during the next years as possible strong alternatives for 2GE industry. Fermentation experiments were in the pipeline to be done, but due to coronavirus pandemic we were not able to finish these experiments in time for thesis writing.

Despite the fact that this work has a high value for biofuel industry, which might represent a substantial increase in the yield of ethanol production, it also contributed for a better understanding of basic biology of grass cell wall biosynthesis. We were able to analyse

agronomical traits, plant anatomy, biochemistry of plant cell wall with focus on the xylan and the importance of GlcA for biomass recalcitrance. Therefore, insights from this work can be used to transform other important biomass species such as for example maize, miscanthus, sorghum and even energy cane.

## REFERENCES

Aitken KS, et al. (2008) Genetic control of yield related stalk traits in sugarcane. Theor. Appl. Genet. 117: 1191–1203

Aitken KS, et al. (2014) A comprehensive genetic map of sugarcane that provides enhanced map coverage and integrates high-throughput Diversity Array Technology (DArT) markers. BMC Genomics 15(152): 1-12.

Aitken KS, Jackson PA & McIntyre CL (2005) A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo) logous linkage groups in a sugarcane cultivar. Theor. Appl. Genet. 110: 789–801

Al-Janabi SM, et al. (2007) Identification of a major quantitative trait locus (QTL) for yellow spot (Mycovellosiella koepkei) disease resistance in sugarcane. Mol. Breed. 19: 1–14.

Anders N, Wilkinson MD, Lovegrove A, et al. Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses (2012). *Proc Natl Acad Sci.* 109(3):989-993. doi:10.1073/pnas.1115858109

Balat M (2011) Production of bioethanol from lignocellulosic materials via the biochemical pathway: A review. Energy conversion and Management. 52(2):858-875.

Bewg WP, Poovaiah C, Lan W, Ralph J, Coleman HD (2016) RNAi downregulation of three key lignin genes in sugarcane improves glucose release without reduction in sugar production. *Biotechnol Biofuels*. 9(270):1-13 doi:10.1186/s13068-016-0683-y

Bligh EG & Dyer WJ (1959) A RAPID METHOD OF TOTAL LIPID EXTRACTION AND PURIFICATION *Canadian Journal of Biochemistry and Physiology*. 37(8): 911-917

Bonawitz ND and Chapple C (2010) 'The Genetics of Lignin Biosynthesis: Connecting Genotype to Phenotype', *Annual Review of Genetics, Vol 44,* 44, pp. 337-363.

Braunbeck O & Cortez, LAB. O cultivo da Cana-de-açúcar e os Uso de Resíduos. In ROSILLO-CALE, BAJAY E ROTHMAN (org) “Uso da Biomassa para Produção de Energia na Indústria Brasileira”. ISBN 85-268-0685-8. Campinas, São Paulo. Editora da UNICAMP, 2005.

Brazilian Ministry of Mines and Energy, 2019. Brazilian Energy Review - 2018 (Resenha Energética Brasileira - 2018).

Bremer GA (1923) cytological investigation of some species and species-hybrids of the genus Saccharum. Genetica 5: 273–326.

Bromley JR, Busse-Wicher M, Tryfona T, Mortimer JC, Zhang Z, Brown DM, Dupree P (2013) GUX1 and GUX2 glucuronyltransferases decorate distinct domains of glucuronoxylan with different substitution patterns. The Plant Journal. 74: 423-434

Brown DM, Goubet F, Wong VW, Goodacre R, Stephens E, Dupree P, Turner SR (2007) Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. The Plant Journal. 52:1154-1168.

Brown L & Torget R (1996) Enzymatic Saccharification of Lignocellulosic Biomass, in NREL Ethanol Project, LAP 009.

Brutnell TP, Wang L, Swatwood K, Goldschimdt A, Jacksn D, Zhu X, Kellogg E, Ecka JV (2010) Setaria viridis: A Model for C4 Photosynthesis. The Plant Cell. 22: 2537-2544.

Buckeridge MS, Santos WD, Tiné MAS, Souza AP (2016) The Cell Wall Architecture of Sugarcane and its Implications to Cell Wall Recalcitrance. In book: Compendium of bioenergy plants: SUGARCANE. Publisher: CRC Press – Taylor and Francis Group. Editors: Lam E, Carrer H, da Silva JA, Kole C.

Burner DM & Legendre BL (1994) Cytogenetic and fertility characteristics of elite sugarcane clones. Sugar Cane 1: 6–10.

Busse-Wicher M, Gomes TCF, Tryfona T, Nikolovski N, Stott K, Grantham NJ, Bolam DN, Skaf MS and Dupree P (2014) The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of *Arabidopsis thaliana*. The Plant Journal. 79: 492–506**.**

Cafall KH & Mohnen D (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydrate Research. 344: 1879-1900.

Canilha L, Kumar CA, Dos Santos MTS, Fernandes AFA, Da Costa FWL, Das Gracas AFM, et al. (2012). Bioconversion of sugarcane biomass into ethanol: an overview about composition, pretreatment methods, detoxification of hydrolysates, enzymatic saccharification, and ethanol fermentation. J. Biomed. Biotechnol. 989572.

Cardona CA & Sánchez OJ (2007) Fuel ethanol production: Process design trends and integration opportunities. Bioresource Technology. 98: 2415-2457.

Carpita NC & Gibeaut DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. The Plant Journal 3: 1–30.

Carpita NC (1996) Structure and biogenesis of the cell walls of grasses. Annual Review of Plant Physiology and Plant Molecular Biology 47: 445–476.

Carvalho DMD, Sevastyanova O, Penna LS, Silva BPD, Lindström ME, Colodette JL (2015) Assessment of chemical transformations in eucalyptus, sugarcane bagasse and straw during hydrothermal, dilute acid, and alkaline pretreatments. Ind. Crops Prod. 73: 118–126. https://doi.org/10.1016/j.indcrop.2015.04.021.

Chanoca A, de Vries L and Boerjan W (2019) Lignin Engineering in Forest Trees. *Front. Plant Sci.* 10:912. doi: 10.3389/fpls.2019.00912

Chiniquy D, Sharma V, Schultink A, Baidoo EE, Rautengarten C, Cheng K, Carroll A, Ulvskov P, Harholt J, Keasling JD, Pauly M, Scheller HV and Ronald PC (2012) XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. PNAS. 109(42): 17117-17122.

Chundawat SPS, Beckham GT, Himmel ME, Dale BE (2011) Deconstruction of lignocellulosic biomass to fuels and chemicals. Annual Review of Chemical and Biomolecular Engineering 2: 121–145.

Clough SJ & Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant Journal. 16(6): 735-743.

Cormack BP, Valdivi RH & Falkow S (1996) FACS-optimized mutants of the green fluorescent protein (GFP). Gene. 173: 33-38.

Cosgrove DJ (2005) Growth of the plant cell wall. Nature. 6: 850-861.

Costa THF, Masarin F, Bonifácio TO, Milagres AMF & Ferraz A (2013) The enzymatic recalcitrance of internodes of sugar cane hybrids with contrasting lignin contents. Industrial Crops and Products, 51, 202–211. https://doi.org/10.1016/j.indcrop.2013.08.078

D'Hont, A. et al. (1994) A molecular approach to unraveling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. Genome 37: 222–230.

D'Hont, A. et al. (1996) Characterisation of the double genome structure of modern sugarcane cultivars (Saccharum spp.) by molecular cytogenetics. Mol. Gen. Genet. 250: 405–413.

de Oliveira DM, Finger-Teixeira A, Rodrigues MT, Salvador VH, Moreira-Vilar FC, Molinari HB, et al. (2015) Ferulic acid: A key component in grass lignocellulose recalcitrance to hydrolysis. Plant Biotechnology Journal. 13: 1224–1232. https://doi.org/10.1111/pbi.12292

de Setta N, Monteiro-Vitorello CB, Metcalfe CJ, et al.: (2014) Building the sugarcane genome for biotechnology and identifying evolutionary trends. BMC Genomics. 15(1): 540.

de Souza WR, Pacheco TF, Duarte KE et al. (2019) Silencing of a BAHD acyltransferase in sugarcane increases biomass digestibility. Biotechnol Biofuels 12(111):1-14. https://doi.org/10.1186/s13068-019-1450-7

Debibakas S, et al. (2014) Prospecting sugarcane resistance to Sugarcane yellow leaf virus by genome-wide association. Theor. Appl. Genet. 127: 1719–1732.

Depicker A, Stachel S, Dhaese P, Zanbryski P, Goodman HM (1982) Nopaline synthase: transcript mapping and DNA sequence. J Mol Appl Genet 1(6):561–573

Derikvand MM, Sierra JB, Ruel K, Pollet B, Do CT, Thevenin J, Bufard D, Jouanin L, Lapierre C. (2008) Redirection of the phenylpropanoid pathway to feruloyl malate in Arabidopsis mutants defcient for cinnamoyl-CoA reductase 1. Planta. 227(5):943–56

Dias MOS, Junqueira TL, Cavalett O, Cunha MP, Jesus CDF, Rossell CEV, Filho RM, Bonomi A (2012) Integrated versus stand-alone second-generation ethanol production from sugarcane bagasse and trash. Bioresource Technology. 103: 152-161.

Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F (1956) Colorimetric Method for Determination of Sugars and Related Substances. Anal. Chem. 28 (3): 350-356

Dufour P, et al. (1997) Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. Theor. Appl. Genet. 94, 409–418.

Ebringerova A, Hrom AZ, Heinze T (2005) Hemicellulose. Adv. Polym. Sci. 186:1–67.

Fagard M, Desnos T, Desprez T, Goubet F et al. (2000) 'PROCUSTE1 encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of arabidopsis'. *Plant Cell.* 12(12): 2409-2423.

Figueira TR, et al. (2012) A BAC library of the SP80-3280 sugarcane variety (saccharum sp.) and its inferred microsynteny with the sorghum genome. BMC Res. Notes 5: 185

Figueiredo R, Araújo P, Llerena JPP, Mazzafera P (2018) Suberin and hemicellulose in sugarcane cell wall architecture and crop digestibility: A biotechnological perspective. Food and Energy Security. 8:4-24

Fincher GB (2009) Exploring the evolution of (1,3;1,4)-beta-D-glucans in plant cell walls: comparative genomics can help! Curr. Opin. Plant Biol. 12: 140–7.

Foust TD, Arent D, Macedo IC, Goldemberg J, Hoysala C, Filho RM, Nigro FEB, Richard TL, Saddler J, Samseth J, Somerville CR, Energy Security, in: Bioenergy & Sustainability: Bridging the Gaps, SCOPE 72, Scientific 387 Committee on Problems of the Environment (SCOPE), São Paulo, 2015: pp. 33-61

Gallinari RH, Coletta RD, Araújo P, Menossi M, Nery MF (2020) Bringing to light the molecular evolution of *GUX* genes in plants. Genetics and Molecular Biology. 43(1):1-9.

Gao X, Kumar R, Wyman CE (2014) Fast Hemicellulose Quantification Via a Simple One-Step Acid Hydrolysis. Biotechnology and Bioengineering. 111: 1088-1096.

Gao S, Yang Y, Wang C, Guo J, Zhou D, et al. (2016) Transgenic Sugarcane with a *cry1Ac* Gene Exhibited Better Phenotypic Traits and Enhanced Resistance against Sugarcane Borer. PLOS ONE 11(4): e0153929

Garsmeur O, Droc G, Antonise R, et al. (2018) A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat Commun* **9:** 2638. https://doi.org/10.1038/s41467-018-05051-5

Gianotto AC, Abreu HMC, Arruda P, Bespalhok JCF, Burnquist WL, Creste S, Ciero L, Ferro JA, Figueira AVO, Filgueiras TS, et al. (2011) Sugarcane (Saccharum X officinarum): A Reference Study for the Regulation of Genetically Modified Cultivars in Brazil. Tropical Plant Biology. 4:62-89.

Giummarella N & Lawoko M (2016) Structural Basis for the Formation and Regulation of Lignin–Xylan Bonds in Birch. *ACS Sustainable Chem. Eng.* 4(10): 5319–5326

Gnansounou E & Dauriat A (2010) Techno-economic analysis of lignocellulosic ethanol: a review. Bioresour. Technol. 101: 4980–4991. doi:10.1016/j.biortech.2010.02.009

Goldemberg J and Guardabassi P (2010) The potential for first-generation ethanol production from sugarcane. *Biofuels Bioprod Bioref* **4**:17–24.

Gomez LD, Clare GS, McQueen-Mason J (2008) Sustainable liquid biofuels from biomass: the writings on the walls. New Phytologist. 178: 473-85.

Goubet F, Jackson P, Deery MJ, Dupree P (2002) Polysaccharide Analysis Using Carbohydrate Gel Electrophoresis: A Method to Study Plant Cell Wall Polysaccharides and Polysaccharide Hydrolases. Analytical Biochemistry. 300: 53–68.

Goubet F, Misrahi A, Park SK, Zhang ZN, Twell D, et al. (2003) AtCSLA7, a cellulose synthase-like putative glycosyltransferase, is important for pollen tube growth and embryogenesis in Arabidopsis. Plant Physiol. 131:547–57

Gouy M et al. Experimental assessment of the accuracy of genomic selection in sugarcane. Theor. Appl. Genet. 126, 2575–2586 (2013).

Gouy M, et al. (2015) Genome wide association mapping of agro-morphological and disease resistance traits in sugarcane. Euphytica 202: 269–284.

Gouy M, Nibouch, S, Hoarau JH & Costet L (2013) in Translational Genomics for Crop Breeding: Abiotic Stress, Yield and Quality, Vol. 2, (eds. Varshney, R. & Tuberosa, R.) Ch. 13, 211–237 (John Wiley & Sons, Ames, Chichester and Oxford, 2013).

Graça J, Cabral V, Santos S, Lamosa P, Serra O, Molinas M, et al. (2015) Partial depolymerization of genetically modified potato tuber periderm reveals intermolecular linkages in suberin polyester. Phytochemistry, 117: 209–219. ttps://doi.org/10.1016/j. phytochem.2015.06.010

Grantham NJ, Wurman-Rodrich J, Terrett O, Lyczakowski J, Stott KM, Iuga D, Simmons TJ et al. (2017) An even pattern of xylan substitution is critical for interaction with cellulose in plant cell walls. Nature plants. 3 (11): 859-865.

Grativol C, Regulski M, Bertalan M, et al. (2014) Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus Saccharum. Plant J. 79(1): 162–172

Grivet L, et al. (1996) RFLP mapping in cultivated sugarcane (Saccharum spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid. Genetics 142: 987–1000.

Guidelli GV, Matiello L, Gallinari RH, de Lucca PC, Menossi M (2018) pGVG: a new Gateway-compatible vector for transformation of sugarcane and other monocot crops. Genetics and Molecular Biology. Genetics and Molecular Biology. 41(2):450-454.

Guilherme AA, Dantas PVF, Santos ES, Fernandes FAN, Macedo GR (2015) Evaluation of composition, characterization and enzymatic hydrolysis of pretreated sugar cane bagasse. Brazilian Journal of Chemical Engineering. 32: 23-33.

Guimaraes CT, Sills GR & Sobral BW (1997) Comparative mapping of Andropogoneae: Saccharum L. (sugarcane) and its relation to sorghum and maize. Proc. Natl Acad. Sci. USA 94: 14261–14266.

Haberl H, Erb KH, Krausmann F, Gaube V, Bondeau A, Plutzar C, Gingrich S, Lucht W, Fischer-Kowalski M (2007) Quantifying and mapping the human appropriation of net primary

production in earth's terrestrial ecosystems. Proceedings of the National Academy of Sciences. 104: 12942–12947.

Hallac BB and Ragauskas AJ (2011) Analyzing cellulose degree of polymerization and its relevancy to cellulosic ethanol, Biofuels Bioproducts & Biorefining-Biofpr*,* 5(2): 215-225.

Harholt J, Ulvskov P, Sørensen I, Willats W, Petersen BL, et al. (2008) Cell wall composition and glycosyltransferases involved in cell wall formation of *Selaginella moellendorffii*. Poster session presented at Third Conference on Biosynthesis of Plant Cell Walls, Asilomar, California, United States.

He L, Zhao M, Wang Y, Gai J and He C (2013) Phylogeny, structural evolution and functional diversification of the plant *PHOSPHATE1* gene family: a focus on *Glycine max*. BMC Evol Biol 13:103.

Helle S, Cameron D, Lam J, White B, Duff S (2003) Effect of inhibitory compounds found in biomass hydrolysates on growth and xylose fermentation by a genetically engineered strain of *S. cerevisae*. Enzyme Microb. Technol. 33: 786-792.

Hellens RP, Edwards EA, Leyland NR, Bean S and Mullineaux PM (2000) pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. Plant Molecular Biology. 42: 819-932.

Henry RJ (2010) Evaluation of plant biomass resources available for replacement of fossil oil. Plant Biotechnol. J. 8:288–293. doi:10.1111/j.1467-7652.2009. 00482.x

Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science. 315: 804-807.

Hoang NV, Furtado A, Botha FC, Simmons BA, Henry RJ (2015) Potential for Genetic Improvement of Sugarcane as a Source of Biomass for Biofuels. Frontiers. 3: 1-15.

Hoarau JY, et al. (2002) Genetic dissection of a modern sugarcane cultivar (Saccharum spp.). II. Detection of QTLs for yield components. Theor. Appl. Genet. 105, 1027–1037.

Holland N, Holland D, Helentjaris T, Dhugga KS, Xoconostle-Cazares B and Delmer DP (2000) A comparative analysis of the plant cellulose synthase (CesA) gene family. Plant Physiology. 123(4): 1313-1323.

Huang X & Madan A (1999) CAP3: A DNA Sequence Assembly Program. Genome Res. 9(9): 868-877.

Iiyama K, Lam TBT and Stone BA (1990) *Phytochemistry*29: 733–737.

Inngjerdingen KT, Trushar R Patel, Xinyong Chen, Lennart Kenne, Stephanie Allen, Gordon A Morris, Stephen E Harding, Tsukasa Matsumoto, Drissa Diallo, Haruki Yamada, Terje E Inngjerdingen KT, Patel TR, Chen X, Kenne L et al. (2007) Immunological and Structural Properties of a Pectic Polymer from *Glinus Oppositifolius*, *Glycobiology*. 17 (12):1299–1310. https://doi.org/10.1093/glycob/cwm088

Irvine JE (1999) Saccharum species as horticultural classes Theoretical and Applied. Genetics. 98: 186–194.

Jackson CL, Tina M Dreaden, Lisa K Theobald, Nhien M Tran, Tiffany L Beal, Manal Eid, Mu Yun Gao, Robert B Shirley, Mark T Stoffel, M Vijay Kumar, Debra Mohnen (2007) Pectin induces apoptosis in human prostate cancer cells: correlation of apoptotic function with pectin structure, *Glycobiology*. 17(8): 805–819, https://doi.org/10.1093/glycob/cwm054

Jannoo N et al. (1999) Molecular investigation of the genetic base of sugarcane cultivars. Theor. Appl. Genet. 99: 171–184.

Jarvis MC (2018) Structure of native cellulose microfibrils, the starting point for nanocellulose manufacture, *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences,* 376(2112).

Jensen JK, Busse-Wicher M, Poulsen CP, Fangel JU, Smith PJ, Yang JY, Peña MJ, Dinesen MH, Martens HJ, Melkonian M, et al. (2018) Identification of an algal xylan synthase indicates that there is functional orthology between algal and plant cell wall biosynthesis. New Phytol. 1-12

Johansen D A (1940) *Plant microtechnique*. (McGraw-Hill Publishing Company, Ltd., London, 1940).

Jones L, Ennos AR, Turner SR (2001) Cloning and characterization of irregular xylem4 (irx4): a severely lignin-defcient mutant of Arabidopsis. Plant Journal. 26(2):205–16.

Jung HG & Allen MS (1995) Characteristics of plant cell walls affecting intake and digestibility of forages by ruminants. Journal of Animal Science. 73: 2774–2790. https://doi.org/10.2527/1995.7392774x

Jung JH, Fouad WM, Vermerris W, Gallo M and Altpeter F (2012) RNAi suppression of lignin biosynthesis in sugarcane reduces recalcitrance for biofuel production from lignocellulosic biomass. Plant Biotechnol J, 10: 1067-1076. doi:10.1111/j.1467-7652.2012.00734.x

Jung JH, Vermerris W, Gallo M, Fedenko JR, Erickson JE, Altpeter F. (2013) RNA interference suppression of lignin biosynthesis increases fermentable sugar yields for biofuel production from field-grown sugarcane. *Plant Biotechnol J*. 11(6):709-716. doi:10.1111/pbi.12061

Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A and Jermiin LS (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Methods 14:587-589.

Kang X, Kirui A, Dickwella Widanage MC, et al. (2019) Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state NMR. *Nat Commun* 10(347):1-9 https://doi.org/10.1038/s41467-018-08252-0

Katahira S, Muramoto N, Moriya S, et al. (2017) Screening and evolution of a novel protist xylose isomerase from the termite Reticulitermes speratus for efficient xylose fermentation in Saccharomyces cerevisiae. Biotechnol Biofuels 10(203): 1-18. https://doi.org/10.1186/s13068-017-0890-1

Kato Y, Ito S, Iki K and Matsuda K (1982) *Plant Cell Physiol.*23: 351–364.

Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-780.

Klein BC, Sampaio ILM, Mantelatto PE, Filho RM, Bonomi A (2019) Beyond ethanol, sugar, and electricity: a critical review of product diversification in Brazilian sugarcane mills. Biofuels, Bioprod. Biorefining 13: 809–821.

Kudus M. Enzymes in Food Biotechnology: Production, Applications, and Future prospects, in: Enzymatic production of syrups. Elsevier. 2019. 155

Kulkarni AR, Peña MJ, Avci U, Mazumder K, Urbanowicz BR, Pattathil S, Yin Y, O'Neill MA, Roberts AW, Hahn MG, Xu Y, Darvill AG and York WS (2012) The ability of land plants to synthesize glucuronoxylans predates the evolution of tracheophytes, *Glycobiology,* 22(3):439-451.

Kumar, M. and Turner, S. (2015) Plant cellulose synthesis: CESA proteins crossing kingdoms, Phytochemistry. 112: 91-99.

Le Cunff L, et al. (2008) Diploid/polyploid syntenic shuttle mapping and haplotypespecific chromosome walking toward a rust resistance gene (Bru1) in highly polyploid sugarcane (2n approximately 12x approximately 115). Genetics 180: 649–660.

Lee C, Teng Q, Zhong R, Ye Z (2012) Arabidopsis GUX Proteins Are Glucuronyltrasnferases Responsible for the Addition of Glucuronic Acid Side Chains onto Xylan. Plant & Cell Physiology. 53(7): 1204-1216.

Leplé J-C, Dauwe R, Morreel K, Storme V, Lapierre C, Pollet B, Naumann A, Kang K-Y, Kim H, Ruel K, et al. (2007) Downregulation of cinnamoyl-coenzyme A reductase in poplar: multiple-level phenotyping reveals efects on cell wall polymer metabolism and structure. Plant Cell. 19(11):3669–91

Li H, Xiong L, Chen X, Wang C, Qi G, Huang C, Luo M, Chen X (2017) Enhanced enzymatic hydrolysis and acetone-butanol-ethanol fermentation of sugarcane bagasse by combined diluted acid with oxidate ammonolysis pretreatment. Bioresour. Technol. 228:257–263. https://doi.org/10.1016/j.biortech.2016.12.119.

Liepman AH, Nairn CJ, Willats WGT, Sørensen I, Roberts AW, Keegstra K (2007) Functional genomic analysis supports conservation of function among cellulose synthase-like A gene family members and suggest diverse roles of mannans in plants. Plant Physiol 143:1881–1893.

Liwanag AJ, Ebert B, Verhertbruggen Y, et al. (2012) Pectin biosynthesis: GALS1 in Arabidopsis thaliana is a β-1,4-galactan β-1,4-galactosyltransferase. *Plant Cell*. 24(12):5024-5036. doi:10.1105/tpc.112.106625

Loureiro M, Barbosa MP, Lopes FF and Silvério F (2011) “Sugarcane breeding and selection for more efficient biomass conversion in cellulosic ethanol,” in Routes to Cellulosic Ethanol, eds Buckeridge MS and Goldman GH (New York, NY: Springer), 199–239.

Lyczakowski JJ, Wicher KB, Terrett OM, Faria-Blanc N, Yu X, Brown D, Krogh KBRM, Dupree P, Busse-Wicher M (2017) Removal of glucuronic acid from xylan is a strategy to improve the conversion of plant biomass to sugars for bioenergy. Biotechnology for Biofuels. 10 (224): 1-11.

Lygin AV, Upton J, Dohleman FG, Juvik J, Zabotina OA, Widholm JM & Lozovaya VV (2011) Composition of cell wall phenolics and polysaccharides of the potential bioenergy crop – Miscanthus. GCB Bioenergy, 3, 333–345. https://doi.org/10.1111/j.1757-1707.2011.01091.x

Macrelli S, Galbe M and Wallberg O (2014) Effects of production and Market factors on ethanol profitability for an integrated first and second generation ethanol plant using the whole sugarcane as feedstock. Biotechnol. Biofuels 7: 26. doi:10.1186/1754-6834-7-26.

Marriot PE, Sibout R, Lapierre C, Fangel JU, Willats WGT, Hofte H, Gómez LD, Mcqueen-Mason SJ (2014) Range of cell-wall alterations enhance saccharification in Brachypodium distachyon mutants. PNAS. 111: 14601-14606.

Masarin F, Gurpilhares DB, Baffa DCF, Barbosa MHP, Carvalho W, Ferraz A, Milagres AMF (2011) Chemical composition and enzymatic digestibility of sugarcane clones selected for varied lignin content. Biotechnology for Biofuels. 4: 1-10.

Matache M, Vladut V, Danciu A, Voicea I, Pirnă I, Postelnicu E, Chirilă C (2011) Vegetal Biomass, A Renewable Source for Obtaining of Clean Energy. Computational Engineering in Systems Applications. 3: 96-100.

Matheson NK (1990) Carbohydrates. Methods Plant Biochem. 2, 371–413.

Mäule C (1901) Das verhalten verholzter membranen gegen kaliumpermanganat, eine holzreaktion neuer art. (A. Zimmer’s Verlag (Ernst Mohrmann), 1901).

Meents MJ, Watanabe Y and Samuels AL (2018) The cell biology of secondary cell wall biosynthesis, *Annals of Botany,* 121(6): 1107-1125.

Meghana M & Shastri Y (2020) Sustainable Valorization of Sugar Industry Waste: Status, Opportunities, and Challenges. Bioresource technology. 303: 1-53.

Miller JR, Dilley KA, Harkins DM, et al. (2017) Initial genome sequencing of the sugarcane CP 96-1252 complex hybrid. F1000 research. 6(688): 1-5.

Ming R, et al. (1998) Detailed alignment of saccharum and sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150: 1663–1682.

Ming R, Liu SC, Moore PH, Irvine JE & Paterson AH (2001) QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res. 11: 2075–2084.

Minh BQ, Nguyen MAT and Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol 30:1188–1195.

Mohnen D, Bar-Peled L & Somerville C (2008b) Cell Wall synthesis. Biomass Recalcitrance, Blackwell Publishing, Oxford

Mohnen, D. (2008a) Pectin structure and biosynthesis, Current Opinion in Plant Biology, 11(3): 266-277.

Molina JWF, Ripoli TC, Geraldi RN, Amaral JR. (1995) Aspectos econômicos e operacionais do enfardamento de resíduos de colheita de cana-de-açúcar para aproveitamento energético. STAB—Açúcar Álcool e Subprodutos. 13 (05): 28-31

Moreira LRS and Filho EXF (2008) An overview of mannan structure and mannan-degrading enzyme systems. Applied Microbiology and Biotechnology. 79(2):165-178.

Mortimer JC, Faria-Blanc N, Yu X, Tryfona T, Sorieul M, Ng YZ, Zhang Z, Stott K, Anders N and Dupree P (2015) An unusual xylan in Arabidopsis primary cell walls is synthesised by GUX3, IRX9L, IRX10L and IRX14. Plant Journal. 83(3): 413-527.

Mortimer JC, Miles GP, Brown DM, Zhang Z, Segura MP, Weimar T, Yu X, Seffen KA, Stephens E, Turner SR, Dupree P (2010) Absence of branches from xylan in Arabidopsis *gux* mutants reveals potential for simplification of lignocellulosic biomass. PNAS. 107: 17409-17414.

Murashige T & Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol. Plant., 15 (3): 473-497

Naik SN, Goud VV, Prasant KR, Ajay KD (2010) Production of first and second-generation biofuels: A comprehensive review. Renewable and Sustainable Energy Reviews. 14: 578–597.

Nei M and Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 39:121–152.

Nishitani K and Nevins DJ (1989) *Plant Physiol.*91: 242–248.Carpita NC (1996) *Annu. Rev. Plant Physiol. Plant Mol. Biol.*47: 445–476.

Okura VK, de Souza RS, de Siqueira Tada SF, et al. (2016) BAC-Pool Sequencing and Assembly of 19 Mb of the Complex Sugarcane Genome. Front Plant Sci. 7(342):1-8.

Padella M, O'Connell A, Prussi M (2019) What is still limiting the deployment of cellulosic ethanol? Analysis of the current status of the sector. Appl. Sci. 9(4523):1-13.

Papini-Terzi FS, Rocha FR, Vêncio RZ, Oliveira KC, Felix JM, Vicentini R, Rocha CS, Simões AC, Ulian EC, di Mauro SMZ et al. (2005) Transcription profiling of signal transduction-related genes in sugarcane tissues. DNA Research 12:27-38.

Pattathil S, Hahn MG, Dale BE, Chundawat SPS (2015) Insights into plant cell wall structure, architecture, and integrity using glycome profiling of native and AFEX$^{TM}$-pre-treated biomass. *Journal of Experimental Botany*. 66(14):4279–4294

Pauly M, Gille S, Liu L, Mansoori N, de Souza A, Schultink A, Xiong G (2013). Hemicellulose biosynthesis. Planta. 238: 627– 642. https://doi.org/10.1007/s00425-013-1921-1

Pauly M, Keegstra K. (2008) Cell-wall carbohydrates and their modification as a resource for biofuels. The Plant Journal 54: 559–568.

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996) Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. *Proc Natl Acad Sci.* 93(22):12637-12642. doi:10.1073/pnas.93.22.12637

Pellny TK, Patil A, Wood AJ, Freeman J, Halsey K, Plummer, A., Kosik, O., Temple, H., Collins, J. D., Dupree, P., Berry, S., Shewry, P. R., Lovegrove, A., Phillips, A. L. and Mitchell, R. A.

C. (2020) Loss of TaIRX9b gene function in wheat decreases chain length and amount of arabinoxylan in grain but increases cross-linking. *Plant Biotechnol. J. 1-12.* https://doi.org/10.1111/pbi.13393

Peña MJ, Darvill AG, Eberhard S, York WS & O'Neill MA (2008) Moss and liverwort xyloglucans contain galacturonic acid and are structurally distinct from the xyloglucans synthesized by hornworts and vascular plants. Glycobiology 18: 891–904.

Peña MJ, Kulkarni AR, Backe J, Boyd M, O'Neill MA and York WS (2016) Structural diversity of xylans in the cell walls of monocots, *Planta,* 244(3): 589-606.

Peña MJ, Zhong RQ, Zhou GK, Richardson EA, O'Neill MA, Darvill AG, York WS and Ye ZH (2007) Arabidopsis irregular xylem8 and irregular xylem9: Implications for the complexity of glucuronoxylan biosynthesis. *Plant Cell,* 19(2): 549-563.

Peng F, Ren JL, Xu F, Bian J, Peng P, Sun RC (2009) Comparative study of hemicelluloses obtained by graded ethanol precipitation of sugarcane bagasse. J Agr Food Chem. 57: 6305–6317.

Piperidis G, Piperidis N & D'Hont A (2010) Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. Mol. Genet. Genom. 284: 65–73.

Piperidis N, et al. (2008) Comparative genetics in sugarcane enables structured map enhancement and validation of marker-trait associations. Mol. Breed. 21: 233–247.

Ponnusamy VK, Nguyen DD, Dharmaraja J, Shobana S, Banu JR, Saratale RG, Chang SW, Kumar G (2019). Bioresource Technology. 271:462-472.

Poovaiah CR, Bewg WP, Lan W, et al. (2016) Sugarcane transgenics expressing MYB transcription factors show improved glucose release. *Biotechnol Biofuels* 9(143):1-18. https://doi.org/10.1186/s13068-016-0559-1

Price S (1963) Cytogenetics of modern sugarcanes. Econ. Bot. 17: 97–105

Raboin LM, et al. (2006) Genetic mapping in sugarcane, a high polyploid, using biparental progeny: identification of a gene controlling stalk colour and a new rust resistance gene. Theor. Appl. Genet. 112: 1382–1391.

Raboin LM, Pauquet J, Butterfield M, D'Hont A & Glaszmann JC (2008) Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. Theor. Appl. Genet. 116: 701–714.

Ralph J (2010) Hydroxycinnamates in lignification. Phytochemistry Reviews, 9: 65–83. https://doi.org/10.1007/s11101-009-9141-9

Reis D & Vian B (2004) Helicoidal pattern in secondary cell walls and possible role of xylans in their construction. *C R Biol*. 327(9-10):785-790. doi:10.1016/j.crvi.2004.04.008

Rennie E & Scheller HV (2014) Xylan biosynthesis. Current Opinion in Biotechnology. 26:100-107.

RFA (Renewables Fuel Association), Statistics - Annual World Fuel Ethanol Production. Available at: http://www.ethanolrfa.org/industry/statistics/annual-ethanol-production/ [may 15th, 2020]

Riaño-Pachón DM, Mattiello L. (2017) Draft genome sequencing of the sugarcane hybrid SP80-3280. *F1000Res*. 6:861. doi:10.12688/f1000research.11859.2

Riaz S, Nasir IA, Bhatti MU, et al. (2020) Resistance to *Chilo infuscatellus* (Lepidoptera: Pyraloidea) in transgenic lines of sugarcane expressing *Bacillus thuringiensis* derived Vip3A protein. *Mol Biol Rep* **47,** 2649–2658

Ronquist F and Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Rosales-Calderon O, Arantes V (2019) A review on commercial-scale high-value products that can be produced alongside cellulosic ethanol. Biotechnol. Biofuels 12(240):1-58. https://doi.org/10.1186/s13068-019-1529-1.

Rytioja J, Hildén K, Yuzon J, Hatakka A, de Vries RP, Mäkelä MR (2014) Plant-polysaccharide-degrading enzymes from Basidiomycetes. Microbiol Mol Biol Rev. 78(4):614-49. doi: 10.1128/MMBR.00035-14. PMID: 25428937; PMCID: PMC4248655.

Samir MASA, Alloin F & Dufresne A (2005) Review of recent research into cellulosic whiskers, their properties and their application in nanocomposite field. Biomacromolecule, 6: 612-626

Sarker TC, Azam SMGG, Bonanomi G (2017) Recent Advances in Sugarcane Industry Solid By-Products Valorization. Waste and Biomass Valorization 8: 241–266.

Schädel C, Blöchl A, Richter A, Hoch G (2010) Quantification and monosaccharide composition of hemicelluloses from different plant functional types. Plant Physiology and Biochemistry. 48: 1-8.

Scheller HV & Ulvskov P (2010) Hemicelluloses. Annual Review of Plant Biology. 61: 263-289.

Schubert C (2006) Can biofuels finally take center stage? Nat. Biotechnol. 24, 777–784. doi:10.1038/nbt0706-777

Shi J, Gladden JM, Sathitsuksanoh N, Kambam P, Sandoval L, Mitra D, et al. (2013). One-pot ionic liquid pretreatment and saccharification of switchgrass. Green Chem. 15, 2579–2589. doi:10.1039/C3GC40545A

Sibout R, Proost S, Hansen BO, Vaid N, Giorgi FM, Ho-Yue-Kuang S, et al. (2017) Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in Brachypodium distachyon. New Phytologist. 215: 1009–1025.

SIDRA IBGE LSPA May 15th 2020 https://sidra.ibge.gov.br/

Simmons TJ, Mortimer JC, Bernardinelli OD, Pöppler A, Brown SP, deAzevedo, ER, Dupree R, Dupree P (2016) Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR. Nature communications. 13902(7): 1-9.

Simpson T (2009) Biofuels: the past, present, and a new vision for the future. Bioscience 59: 926–927. doi:10.1525/bio.2009.59.11.2

Sims REH, Hastings A, Schlamadinger B, Taylor G and Smith P (2006) Energy crops: current status and future prospects. Glob. Chang. Biol. 12: 2054–2076. doi:10.1111/j.1365-2486.2006.01163.x

Sluiter JB, Chum H, Gomes AC, Tavares RPA, Azevedo V, et al. (2016) Evaluation of Brazilian Sugarcane Bagasse Characterization: An Interlaboratory Comparison Study. Journal of AOAC International. 99:579-585.

Smith RA, Cass CL, Mazaheri M, et al. (2017) Suppression of *CINNAMOYL-CoA REDUCTASE* increases the level of monolignol ferulates incorporated into maize lignins. Biotechnol Biofuels 10(109): 1-10. https://doi.org/10.1186/s13068-017-0793-1

Souza AP, Leite DCC, Pattathil S, Hahn MG, Buckeridge MS (2013) Composition and Structure of Sugarcane Cell Wall Polysaccharides: Implications for Second-Generation Bioethanol Production. Bioenerg. Res. 6: 564-579.

Souza GM, Sluys MAV, Lembke CG, Lee H, Margarido GRA, Hotta CT et al. (2019) Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop, *GigaScience*, 8(12) https://doi.org/10.1093/gigascience/giz129

Souza, S. P., Nogueira, L. A. H., Martinez, J., & Cortez, L. A. B. (2018). Sugarcane can afford a cleaner energy profile in Latin America & Caribbean. Renewable Energy. *121: 164–172.* doi:10.1016/j.renene.2018.01.024

Sparkes I, Runions J, Kearns A, et al. (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat Protoc* 1, 2019–2025 https://doi.org/10.1038/nprot.2006.286

Sticklen M (2006) Plant genetic engineering to improve biomass characteristics for biofuels. Curr. Opin. Biotechnol. 17: 315–319. doi:10.1016/j.copbio.2006.05.003

Stout AT, Davis AA, Domec JC, Yang CM, Shi R and King JS (2014) Growth under field conditions affects lignin content and productivity in transgenic Populus trichocarpa with altered lignin biosynthesis. Biomass Bioenergy. 68: 228–239. doi: 10.1016/j.biombioe.2014.06.008

Sun JX, Xu F, Sun XF, Sun RC, Wu SB (2004) Comparative study of lignins from ultrasonic irradiated sugar-cane bagasse. Polymer International. 53: 1711-1721.

Sun Y & Cheng J (2002) Hydrolysis of lignocellulosic materials for ethanol production: a review. Bioresour. Technol. 83 (1) (2002): 1-11

Tamasloukht B, Lam MSJWQ, Martinez Y, Tozo K, Barbier O, Jourda C, Jauneau A, Borderies G, Balzergue S, Renou JP, et al. (2011) Characterization of a cinnamoylCoA reductase 1 (CCR1)

mutant in maize: efects on lignifcation, fbre development, and global gene expression. J Exp Bot. 62(11):3837–48

Taylor NG, Howells RM, Huttly AK, Vickers K and Turner SR (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc. Natl. Acad. Sci. USA. 100: 1450–1455

Terrett OM & Dupree P (2018) Covalent interactions between lignin and hemicelluloses in plant secondary cell walls. *Curr Opin Biotechnol*. 56:97-104. doi:10.1016/j.copbio.2018.10.010

Tomkins JP et al. (1999) A bacterial artificial chromosome library for sugarcane. Theor. Appl. Genet. 99: 419–424.

Trifinopoulos J, Nguyen LT, von Haeseler A and Minh BQ (2016) W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. Nucl Acids Res 44:W232-W235.

Ummartyotin S & Manuspiya H (2015) A critical review on cellulose: From fundamental to an approach on sensor technology. Renewable and Sustainable Energy Review. 41:402-412

Urbanowicz BR, Peña MJ, Moniz HA, Moremen KW and York WS (2014) Two Arabidopsis proteins synthesize acetylated xylan in vitro. Plant J. 80: 197–206.

Vain T, Crowell EF, Timpano H, Biot E, Desprez T, Mansoori N, Trindade LM, Pagant S, Robert S, Hofte H, Gonneau M and Vernhettes S (2014) The Cellulase KORRIGAN Is Part of the Cellulose Synthase Complex, Plant Physiology. 165(4): 1521-1532.

Van Der Weijde T, Kamei ACL, Torres AF, Vermerris W, Dolstra O, Visser RG, et al. (2013) The potential of C4 grasses for cellulosic biofuel production. Front. Plant Sci. 4:107. doi:10.3389/fpls.2013.00107

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin Biosynthesis and Structure. Plant Physiology. 153: 895-905.

Vanholme R, Morreel K, Darrah C, Oyarce P, Grabber JH, Ralph J and Boerjan W (2012) Metabolic Engineering of Novel Lignin in Biomass Crops. New *Phytol.* 196**:** 978 —1000

Vega-Sanchez M, Verhertbruggen Y, Scheller HV, Ronald P (2013) Abundance of mixed linkage glucan in mature tissues and secondary cell walls of grasses. Plant signaling & behavior. 8(2): 1-4. 10.4161/psb.23143.

Vettore AL et al. (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res. 13: 2725–2735.

Viikari L, Vehmaanpera J and Koivula A (2012) Lignocellulosic ethanol: from science to industry. Biomass Bioeng. 46: 13–24. doi:10.1016/j.biombioe. 2012.05.008

Voelker SL, Lachenbruch B, Meinzer FC, Jourdes M, Ki CY, Patten AM, et al. (2010) Antisense down-regulation of 4cl expression alters lignification, tree growth, and saccharification potential of field-grown poplar. Plant Physiol. 154: 874–886. doi: 10.1104/pp.110.159269

Vogel J (2008). Unique aspects of the grass cell wall. Current Opinion in Plant Biology 11: 301–307.

Voiniciuc C, Gunl M, Schmidt MHW and Usadel B (2015) Highly Branched Xylan Made by IRREGULAR XYLEM14 and MUCILAGE-RELATED21 Links Mucilage to Arabidopsis Seeds. *Plant Physiology.* 169(4): 2481-2495.

Wadenbäck J, Von Arnold S, Egertsdotter U, Walter MH, Grima-Pettenati J, Goffner D, et al. (2008) Lignin biosynthesis in transgenic Norway spruce plants harboring an antisense construct for cinnamoyl CoA reductase (Ccr). Transgenic Res. 17: 379–392. doi: 10.1007/s11248-007-9113-z

Wagner A, Donaldson L, Kim H, Phillips L, Flint H, Steward D, et al. (2009) Suppression of 4-coumarate-CoA ligase in the coniferous gymnosperm Pinus radiata. Plant Physiol. 149: 370–383. doi: 10.1104/pp.108.125765

Wang WZ, Yang BP, Feng XY, et al. (2017) Development and Characterization of Transgenic Sugarcane with Insect Resistance and Herbicide Tolerance. *Front Plant Sci*. 8(1535):1-10. doi:10.3389/fpls.2017.01535

Wang X, Tang H, Bowers JE, Feltus FA and Paterson AH (2007) Extensive concerted evolution of rice paralogs and the road to regaining independence. Genetics 177:1753-63.

Wilkie KCB (1979) The hemicelluloses of grasses and cereals. Adv. Carbohyd. Chem. Biochem. 36:215–264. doi: 10.1016/S0065-2318(08)60237-1.

Wilson JR & Hatfield RD (1997) Structural and chemical changes of cell wall types during stem development: Consequences for fibre degradation by rumen microflora. Australian Journal of Agricultural Research 48: 165–180. https://doi.org/10.1071/A96051

Wilson JR, & Hattersley PW (1983) In vitro digestion of bundle sheath cells in rumen fluid and its relation to the suberized lamella and C4 photosynthetic type in Panicum species. Grass and Forage Science, 38: 219–223. https://doi.org/10.1111/j.1365-2494.1983.tb01643.x

Wingren A, Galben M, Zacchi G (2003) Techno-economic evaluation of producing ethanol from softwood: comparison of SSF and SHF and identification of bottlenecks. Biotechnology Process. 19: 1109-1117.

Wu AM, Hornblad E, Voxeur A, Gerber L, Rihouey C, Lerouge P and Marchant A (2010) 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 Physiology. 153(2): 542-554.

Wu K, et al. (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragment. Theor. Appl. Genet. 83: 294–300.

Yokoyama R, Nishitani K (2004) Genomic Basis for Cell-Wall Diversity in Plants. A Comparative Approach to Gene Families in Rice and *Arabidopsis*, *Plant and Cell Physiology*, 45(9): 1111–1121. https://doi.org/10.1093/pcp/pch151Buckeridge MS, Rayon C, Urbanowicz B, Tiné MAS, Carpita NC (2004) Mixed Linkage (1→3),(1→4)-β-d-Glucans of Grasses. Cereal Chem. 81(1)115-127

Yoo HD, Kim D, Paek SH. (2012) Plant cell wall polysaccharides as potential resources for the development of novel prebiotics. *Biomol Ther (Seoul)*. 20(4):371-379. doi:10.4062/biomolther.2012.20.4.371

Yu L, Lyczakowski J, Pereira CS, et al. (2018) The Patterned Structure of Galactoglucomannan Suggests It May Bind to Cellulose in Seed Mucilage. Plant physiology 178:1011-1026.

Zeng W, Lampugnani ER, Picard KL, Song LL, Wu AM, Farion IM, Zhao J, Ford K, Doblin MS and Bacic A (2016) Asparagus IRX9, IRX10, and IRX14A Are Components of an Active Xylan Backbone Synthase Complex that Forms in the Golgi Apparatus. Plant Physiology. 171(1): 93-109.

Zhang Y, Culhaoglu T, Pollet B, Melin C, Denoue D, Barrière Y, et al. (2011) Impact of lignin structure and cell wall reticulation on maize cell wall degradability. Journal of Agricultural and Food Chemistry. 59: 10129–10135. https://doi.org/10.1021/jf2028279

Zhang Y, Nikolovski N, Sorieul M, Vellosillo T, McFarlane HE, Dupree R, Kesten C, Schneider R, Driemeier C, Lathe R, Lampugnani E, Yu XL, Ivakov A, Doblin MS, Mortimer JC, Brown SP, Persson S and Dupree P (2016) Golgi-localized STELLO proteins regulate the assembly and trafficking of cellulose synthase complexes in Arabidopsis, Nature Communications. 7:1-14.

Zhong R and Ye ZH (2009) Secondary Cell Walls. eLS. 1-9p.

Zhong RQ, Cui DT, Phillips DR and Ye ZH (2018) A Novel Rice Xylosyltransferase Catalyzes the Addition of 2-O-Xylosyl Side Chains onto the Xylan Backbone, *Plant and Cell Physiology,* 59(3): 554-565.

# APPENDICES AND ANNEXES

# APPENDICES

UNICAMP

## INFORMAÇÃO

INFORMAMOS que o projeto **CIBio/IB No. 2007/03 – Genômica funcional de plantas**, cujo pesquisador responsável é o Prof. Dr. Marcelo Menossi Teixeira, sub-projeto "*UNDERSTANDING SUGARCANE CELL WALL TO INCREASE 2G ETHANOL PRODUCTION*", do pós graduando Rafael Henrique Gallinari, encontra-se devidamente aprovado e regularizado junto a CIBIO/IB-UNICAMP e a CTNBio, conforme legislação vigente.

Cidade Universitária "Zeferino Vaz",

10 de outubro de 2020

Prof. Dr. JOSÉ LUIZ PROENÇA MÓDENA

Presidente da CIBio

Instituto de Biologia- UNICAMP

**CIBio/IB-UNICAMP**
Comissão Interna de Biossegurança
Instituto de Biologia – Unicamp
Caixa Postal 6109 – 13083-970 Campinas SP
Tel.: (19) 3521-6359 – email: comisib@unicamp.br

Documento assinado eletronicamente por **JOSÉ LUIZ PROENÇA MODENA**, **PRESIDENTE DA CIBIO/IB/UNICAMP**, em 11/12/2020, às 08:59 horas, conforme Art. 10 § 2º da MP 2.200/2001 e Art. 1º da Resolução GR 54/2017.

A autenticidade do documento pode ser conferida no site: sigad.unicamp.br/verifica, informando o código verificador:
**2A319551 A05E429F 8BBC74A3 449FFD92**

UNICAMP
SIGAD

COORDENADORIA DE PÓS-GRADUAÇÃO
INSTITUTO DE BIOLOGIA
Universidade Estadual de Campinas
Caixa Postal 6109. 13083-970, Campinas, SP, Brasil
Fone: (19) 3521-6378. email: cpgib@unicamp.br

# DECLARAÇÃO

Em observância ao **§5° do Artigo 1° da Informação CCPG-UNICAMP/001/15**, referente a Bioética e Biossegurança, declaro que o conteúdo de minha tese de Doutorado, intitulada ***"UNDERSTANDING SUGARCANE CELL WALL TO INCREASE 2G ETHANOL PRODUCTION"***, desenvolvida no Programa de Pós Graduação em Genética e Biologia Molecular do Instituto de Biologia da Unicamp, não versa sobre pesquisa envolvendo seres humanos ou animais.

Assinatura:_____

Nome do(a) aluno(a): Rafael Henrique Gallinari

Assinatura:_____

Nome do(a) orientador(a): Marcelo Menossi Teixeira

Data: 10/10/2020

*Genetics and Molecular Biology* 43, 1, e20180208 (2020)

DOI: https://doi.org/10.1590/1678-4685-GMB-2018-0208

Short Communication
Evolutionary Genetics

# Bringing to light the molecular evolution of *GUX* genes in plants

Rafael Henrique Gallinari[1], Rafael Della Coletta[1], Pedro Araújo[1], Marcelo Menossi[1] and Mariana Freitas Nery[1]

[1]*Universidade Estadual de Campinas, Instituto de Biologia, Departamento de Genética, Evolução, Microbiologia e Imunologia, Campinas, SP, Brazil.*

## Abstract

Hemicellulose and cellulose are essential polysaccharides for plant development and major components of cell wall. They are also an important energy source for the production of ethanol from plant biomass, but their conversion to fermentable sugars is hindered by the complex structure of cell walls. The glucuronic acid substitution of xylan (GUX) enzymes attach glucuronic acid to xylan, a major component of hemicellulose, decreasing the efficiency of enzymes used for ethanol production. Since loss-of-function *gux* mutants of *Arabidopsis thaliana* enhance enzyme accessibility and cell wall digestion without adverse phenotypes, *GUX* genes are potential targets for genetically improving energy crops. However, comprehensive identification of GUX in important species and their evolutionary history are largely lacking. Here, we identified putative GUX proteins using hidden Markov model searches with the GT8 domain and a GUX-specific motif, and inferred the phylogenetic relationship of 18 species with Maximum likelihood and Bayesian approaches. Each species presented a variable number of GUX, and their evolution can be explained by a mixture of divergent, concerted and birth-and-death evolutionary models. This is the first broad insight into the evolution of *GUX* gene family in plants and will potentially guide genetic and functional studies in species used for biofuel production.

*Keywords*: *GUX*, sugarcane, phylogeny, angiosperms, biofuels.

Received: August 12, 2018; Accepted: May 6, 2019.

Plant evolution has been characterized by the development of complex organs and highly specialized cellular structures, including the complex plant cell wall (Sørensen *et al.*, 2010). This structure provides strength and support for the plant body, protects against pathogens and pests, regulates growth, minimizes water loss, and other mechanical and biochemical functions (Sarkar *et al.*, 2009). The cell wall, composed mainly by hemicellulose and cellulose, is very important to plant survival and accounts for most of their biomass (Park and Cosgrove, 2012; Loqué *et al.*, 2015). Consequently, from a technological perspective, the plant cell wall composed mainly of polysaccharides may serve as an important source of renewable energy. The problem is that its complexity decreases the efficiency of saccharification, *i.e.* the process of breaking down the polysaccharides into sugars that can be used as energy source (Jordan *et al.*, 2012; Yue *et al.*, 2014). For instance, the interaction between cellulose and xylan, one of the main components of hemicellulose, may impede the accessibility of enzymes that degrade cellulose to produce fermentable sugars (Simmons *et al.*, 2016).

Send correspondence to Mariana Freitas Nery. Universidade Estadual de Campinas, Departamento de Genética, Instituto de Biologia, 13083-862, Campinas, São Paulo, Brazil. E-mail: mariananery@gmail.com.

At the molecular level, several genes that control the deposition and arrangement of the plant cell wall have been reported in *Arabidopsis thaliana*, such as the irregular xylem (IRX) genes *IRX8*, *IRX9*, *IRX14*, the genes fragile fiber 8/ irregular xylem 7 **(***FRA8*), galacturonosyltransferase-like 1 (*PARVUS*) and glucuronic acid substitution of xylan (*GUX*) (Brown *et al.*, 2007; Lee *et al.*, 2007ab; Peña *et al.*, 2007; Mortimer *et al.*, 2010). While mutations in most of these genes only change the proportion of methylglucuronic acid (MeGlcA) and glucuronic acid (GlcA) attached to xylan, mutations on *GUX* genes were reported to reduce the presence of such residues that hinders the access of cellulases to biomass and to increase saccharification yield (Mortimer *et al.*, 2010; Lee *et al.*, 2012; Lyczakowski *et al.*, 2017). Importantly, these mutations did not interfere with plant development, making *GUX* genes potential targets for genetically engineering plant cell walls (Mortimer *et al.*, 2010; Lee *et al.*, 2012; Lyczakowski *et al.*, 2017). *GUX* genes comprise a multigene family, with five homologous genes annotated in the *Arabidopsis* genome (*AtGUX1-5*; Mortimer *et al.*, 2010; Rennie *et al.*, 2012), and at least one gene in the conifer *Picea glauca* (*PgGUX*; Lyczakowski *et al.*, 2017).

Accordingly, we performed *in silico* analyses to identify putative GUX proteins in different Angiosperm groups

to infer their phylogenetic relationships to ultimately unravel their evolution from a molecular standpoint. Our results can guide future applied research with GUX in economically important biofuel crops, since the first step towards the production of genetically modified plants is to understand how widespread these genes are in a phylogenetic context, and also in how many copies they are present within the genome.

In order to reconstruct the phylogenetic relationship of the GUX family, we selected 16 angiosperm species (including six monocots and ten dicots) that are either model plants or important crops: thale cress (*Arabidopsis thaliana*), purple false brome (*Brachypodium distachyon*), wild cabbage (*Brassica oleraceae*), turnip (*Brassica rapa*), sweet orange (*Citrus sinensis*), flooded gum (*Eucalyptus grandis*), soybean (*Glycine max*), rice (*Oryza sativa*), black cottonwood (*Populus trichocarpa*), sugarcane (*Saccharum* spp.), foxtail millet (*Setaria italica*), potato (*Solanum tuberosum*), sorghum (*Sorghum bicolor*), cocoa (*Theobroma cacao*), grape (*Vitis vinifera*) and maize (*Zea mays*). We also selected two bryophytes (the moss *Physcomitrella patens*, and the common liverwort *Marchantia polymorpha*) to serve as outgroups in the phylogenetics analysis. The accession numbers from each sequence are shown in Table 1.

Since the five GUX protein sequences for *Arabidopsis thaliana* were already characterized by Mortimer *et al.* (2010) and Rennie *et al.* (2012), we retrieved their sequences from GenBank. For the other 17 species described above (except for sugarcane), we developed a workflow to standardize the identification of GUX proteins based on gene search and protein domain/motif analyses described by Kumar *et al.* (2016) (Figure S1). For this purpose, we retrieved all protein sequences (only from primary transcripts) from the latest version of their reference genome available in Phytozome v12. All GUX enzymes have the glycosyl transferase family 8 (GT8) domain, which is responsible for the addition of glucuronosyl substitutions onto the xylan backbone (Rennie *et al.*, 2012). Therefore, we screened all protein sequences with a hidden Markov model (HMM) search (*hmmsearch* from HMMER v3.1b2) using the GT8 HMM available on PFAM (PF01501). Since not all proteins that have the GT8 domain are GUX proteins, we sought to identify a GUX specific motif. For this purpose, we performed MEME analysis (Bailey *et al.*, 2009) using the five GUX protein sequences described for *Arabidopsis* (Mortimer *et al.*, 2010; Rennie *et al.*, 2012) and two sequences of rice identified by HomoloGene (Database Resources of the National Center for Biotechnology Information, 2016) as input. The motif present in all those GUX sequences was used to screen all GT8 protein sequences in a subsequent HMMER analysis (Figures S2 and S3). Finally, we defined putative GUX sequences for each species when both GT8 domain and the GUX specific motif were present.

Among the 18 species surveyed, sugarcane is the only one that does not have a reference genome available in Phytozome. Thus, we identified its GUX proteins by performing BLAST searches in the SUCEST database (Vettore *et al.*, 2003) using the sorghum orthologs as queries. Then, we used the CAP3 contig assembly program (Huang and Madan, 1999) with the expressed sequence tags (ESTs) obtained from the BLAST search to assemble contigs for each GUX gene in sugarcane. For contigs with incomplete transcripts the closest sorghum ortholog was used to complete the sequence.

After identifying GUX protein sequences for each species, we aligned them with MAFFT (Katoh and Standley, 2013) using the iterative refinement method L-INS-I and no treatment were done in the aligned sequences. Maximum likelihood phylogenetic analysis of the GUX multiple sequence alignment was performed using IQ-Tree v1.6.1 (Trifinopoulos *et al.*, 2016). Branch support was acquired by 1,000 ultrafast bootstraps pseudoreplicates (Minh *et al.*, 2013), under JTT+I+G4 model identified by ModelFinder (Kalyaanamoorthy *et al.*, 2017). For the Bayesian phylogenetic analysis, we used MrBayes v3.2.6 (Ronquist and Huelsenbeck, 2003), using 1,000,000 generations, sample frequency of 500 and diagnostic frequency of 5,000, under JTT+I+G model of evolution. Phylogenetic trees were visualized and edited in FigTree v1.4.3 (Rambaut, 2009). We also tested for robustness of clade arrangements by performing the same analyzes with coding sequences (Figure S4), complete gene sequences (including introns, exons, and UTRs), and gene sequences plus 5' flanking 1 kb and 3' flanking 1 kb. In all cases cited above, the results were very similar.

We performed exhaustive HMM searches to identify GUX proteins in several representatives of plant groups, most with economic importance. The number of GUX varied from one to eleven among the species surveyed (Table 1), suggesting a gene family with a complex history of specific-lineages duplications. From 18 plant species, seven of them (*Brachypodium distachyon*, *Setaria italica, Solanum tuberosum*, *Theobroma cacao*, *Arabidopsis thaliana*, *Sorghum bicolor* and *Saccharum* spp.) have five GUX proteins in their genome, whereas five species have more than five orthologs: *Zea mays* and *Brassica oleraceae* have seven GUX, *Brassica rapa* has 10, *Glycine max* has 11, and *Populus trichocarpa* has six. On the other hand, six out of 18 species have less than five GUX: *Eucalyptus grandis* has four GUX proteins, *Oryza sativa*, *Vitis vinifera* and *Citrus sinensis* have three, and both *Marchantya polymorpha* and *Physcomitrella patens* have only one GUX.

Using putative GUX proteins identified *in silico* for each species and their aligned sequences, we reconstructed the phylogenetic trees. Maximum likelihood and Bayesian phylogenies arranged the GUX family into well-supported clades, allowing us to define the orthologous and paralogous relationships (Figure 1). The only exception was for

**Table 1** - Number of GUX proteins found by HMMER analysis in each species, scientific name, accession number, clade that each protein belongs and the name that appears on the phylogenetic tree.

| Scientific name (reference genome version) | # GUX | Accession number | Clade | Phylogenetic tree name |
|---|---|---|---|---|
| *Arabidopsis thaliana** (TAIR10) | 5 | At3g18660 | GUX 1 | Arabidopsis_thaliana_GUX1 |
| | | At4g33330 | GUX 2 | Arabidopsis_thaliana_GUX2 |
| | | At1g54940.1 | GUX 3 | Arabidopsis_thaliana_GUX3 |
| | | At1g77130.1 | GUX 4 | Arabidopsis_thaliana_GUX4 |
| | | At1g08990.1 | GUX 5 | Arabidopsis_thaliana_GUX5 |
| *Brachypodium distachyon†* (v3.1) | 4 | Bradi2g56810.1 | GUX 1 | Brachypodium_distachyon1 |
| | | Bradi1g72350.1 | GUX 2 | Brachypodium_distachyon2 |
| | | Bradi2g24737.4 | GUX 3 | Brachypodium_distachyon3 |
| | | Bradi3g45800.7 | GUX X | BrachypodiumdistachyonXA |
| | | Bradi5g27680.1 | GUX X | BrachypodiumdistachyonXB |
| *Brassica oleraceae†* (v1.0) | 7 | Bol030957 | GUX 1 | Brassica_oleraceae1 |
| | | Bol013572 | GUX 2 | Brassica_oleraceae2A |
| | | Bol017534 | GUX 2 | Brassica_oleraceae2B |
| | | Bol009658 | GUX 3 | Brassica_oleraceae3 |
| | | Bol006577 | GUX 4/5 | Brassica_oleraceae5A |
| | | Bol022153 | GUX 4/5 | Brassica_oleraceae5B |
| | | Bol022154 | GUX 4/5 | Brassica_oleraceae5C |
| *Brassica rapa†* (v1.3) | 10 | Brara.E02330.1 | GUX 1 | Brassica_rapa1A |
| | | Brara.A02917.1 | GUX 1 | Brassica_rapa1B |
| | | Brara.A00465.1 | GUX 2 | Brassica_rapa2A |
| | | Brara.H01273.1 | GUX 2 | Brassica_rapa2B |
| | | Brara.F01545.1 | GUX 3 | Brassica_rapa3A |
| | | Brara.H02280.1 | GUX 3 | Brassica_rapa3B |
| | | Brara.B02173.1 | GUX 3 | Brassica_rapa3C |
| | | Brara.I01695.1 | GUX 4 | Brassica_rapa4 |
| | | Brara.I05282.1 | GUX 4/5 | Brassica_rapa5A |
| | | Brara.H02850.1 | GUX 4/5 | Brassica_rapa5C |
| *Citrus sinensis†* (v1.1) | 3 | orange1.1g006648m | GUX 1 | Citrus_sinensis1 |
| | | orange1.1g007705m | GUX 2 | Citrus_sinensis2 |
| | | orange1.1g043696m | GUX 3 | Citrus_sinensis3 |
| *Eucalyptus grandis†* (v2.0) | 4 | Eucgr.H04942.1 | GUX 1 | Eucalyptus_grandis1 |
| | | Eucgr.F00232.1 | GUX 2 | Eucalyptus_grandis2 |
| | | Eucgr.F02737.1 | GUX 3 | Eucalyptus_grandis3 |
| | | Eucgr.L01540.1 | GUX 4 | Eucalyptus_grandis4 |
| *Glycine max†* (Wm82.a2.v1) | 11 | Glyma.04G214400.1 | GUX 1 | Glycine_max1A |
| | | Glyma.06G151900 | GUX 1 | Glycine_max1B |
| | | Glyma.05G060700.1 | GUX 1 | Glycine_max1C |
| | | Glyma.05G190200.1 | GUX 1 | Glycine_max1D |
| | | Glyma.17G242500.1 | GUX 2 | Glycine_max2A |
| | | Glyma.14G082500.1 | GUX 2 | Glycine_max2B |
| | | Glyma.04G038500.1 | GUX 2 | Glycine_max2C |
| | | Glyma.02G238200.1 | GUX 3 | Glycine_max3A |
| | | Glyma.14G122600.1 | GUX 3 | Glycine_max3B |
| | | Glyma.19G235600.1 | GUX 4 | Glycine_max4A |
| | | Glyma.10G154600.1 | GUX 4 | Glycine_max4B |
| *Marchantia polymorpha†* (v3.1) | 1 | Mapoly0120s0025.1 | OUTGROUP | Marchantia_polymorpha_OUTGROUP |

**Table 1 - cont.**

| Scientific name (reference genome version) | # GUX | Accession number | Clade | Phylogenetic tree name |
|---|---|---|---|---|
| *Oryza sativa†* (v7_JGI) | 3 | LOC_Os01g65780.2 | GUX 1 | Oryza_sativa1 |
| | | LOC_Os03g08600.1 | GUX 2 | Oryza_sativa2 |
| | | LOC_Os02g35020.1 | GUX X | Oryza_sativaX |
| *Physcomitrella patens†* (v3.3) | 1 | Pp3c1_28970V3.1 | OUTGROUP | Physcomitrella_patens |
| *Populus trichocarpa†* (v3.1) | 6 | Potri.007G107200.1 | GUX 1 | Populus_trichocarpa1A |
| | | Potri.005G061600.5 | GUX 1 | Populus_trichocarpa1B |
| | | Potri.014G029900.1 | GUX 2 | Populus_trichocarpa2 |
| | | Potri.005G187900.1 | GUX 3 | Populus_trichocarpa3 |
| | | Potri.005G033500.1 | GUX 4 | Populus_trichocarpa4A |
| | | Potri.013G022900.2 | GUX 4 | Populus_trichocarpa4B |
| *Saccharum spp* (Vettore *et al.*, 2003) | 5 | sugarcane_contig1 | GUX 1 | Saccharum_sp1 |
| | | sugarcane_contig2 | GUX 2 | Saccharum_sp2 |
| | | sugarcane_contig3 | GUX 3 | Saccharum_sp3 |
| | | sugarcane_contigXA | GUX X | Saccharum_spXA |
| | | sugarcane_contigXB | GUX X | Saccharum_spXB |
| *Setaria italica†* (v2.2) | 5 | Seita.5G402400.1 | GUX 1 | Setaria_italica1 |
| | | Seita.9G515500.1 | GUX 2 | Setaria_italica2 |
| | | Seita.3G235400.1 | GUX 3 | Setaria_italica3 |
| | | Seita.1G193600.1 | GUX X | Setaria_italicaXA |
| | | Seita.5G386200.1 | GUX X | Setaria_italicaXB |
| *Solanum tuberosum†* (v4.03) | 5 | PGSC0003DMT400020680 | GUX 2 | Solanum_tuberosum2A |
| | | PGSC0003DMT400020678 | GUX 2 | Solanum_tuberosum2B |
| | | PGSC0003DMT400063796 | GUX 3 | Solanum_tuberosum3 |
| | | PGSC0003DMT400048884 | GUX 4 | Solanum_tuberosum4A |
| | | PGSC0003DMT400048888 | GUX 4 | Solanum_tuberosum4B |
| *Sorghum bicolor†* (v3.1) | 5 | Sobic.003G376700.1 | GUX 1 | Sorghum_bicolor1 |
| | | Sobic.001G479800.1 | GUX 2 | Sorghum_bicolor2 |
| | | Sobic.009G144200.1 | GUX 3 | Sorghum_bicolor3 |
| | | Sobic.004G177000.1 | GUX X | Sorghum_bicolorXA |
| | | Sobic.003G360500.1 | GUX X | Sorghum_bicolorXB |
| *Theobroma cacao†* (v1.1) | 5 | Thecc1EG001429t2 | GUX 1 | Theobroma_cacao1 |
| | | Thecc1EG033846t1 | GUX 2 | Theobroma_cacao2 |
| | | Thecc1EG035450t1 | GUX 3 | Theobroma_cacao3 |
| | | Thecc1EG026564t1 | GUX 4 | Theobroma_cacao4A |
| | | Thecc1EG026565t1 | GUX 4 | Theobroma_cacao4B |
| *Vitis vinifera†* (Genoscope.12x) | 3 | GSVIVT01026525001 | GUX 1 | Vitis_vinifera1 |
| | | GSVIVT01009501001 | GUX 2 | Vitis_vinifera2 |
| | | GSVIVT01000046001 | GUX 4 | Vitis_vinifera4 |
| *Zea mays†* (Ensembl-18) | 7 | GRMZM2G365544_T01 | GUX 1 | Zea_mays1A |
| | | GRMZM2G135743_T02 | GUX 1 | Zea_mays1B |
| | | GRMZM2G002023_T02 | GUX 1 | Zea_mays1C |
| | | GRMZM2G109431_T01 | GUX 2 | Zea_mays2 |
| | | GRMZM2G058472_T02 | GUX 3 | Zea_mays3 |
| | | GRMZM2G031581_T01 | GUX X | Zea_maysXA |
| | | GRMZM2G441987_T01 | GUX X | Zea_maysXB |

* Accessions retrieved from TAIR database (https://www.arabidopsis.org/)
† Accessions retrieved from Phytozome v12 database (https://phytozome.jgi.doe.gov/)
°Accessions retrieved from SUCEST database (http://sucest-fun.org/); ESTs from sugarcane contigs are available in Table S1.

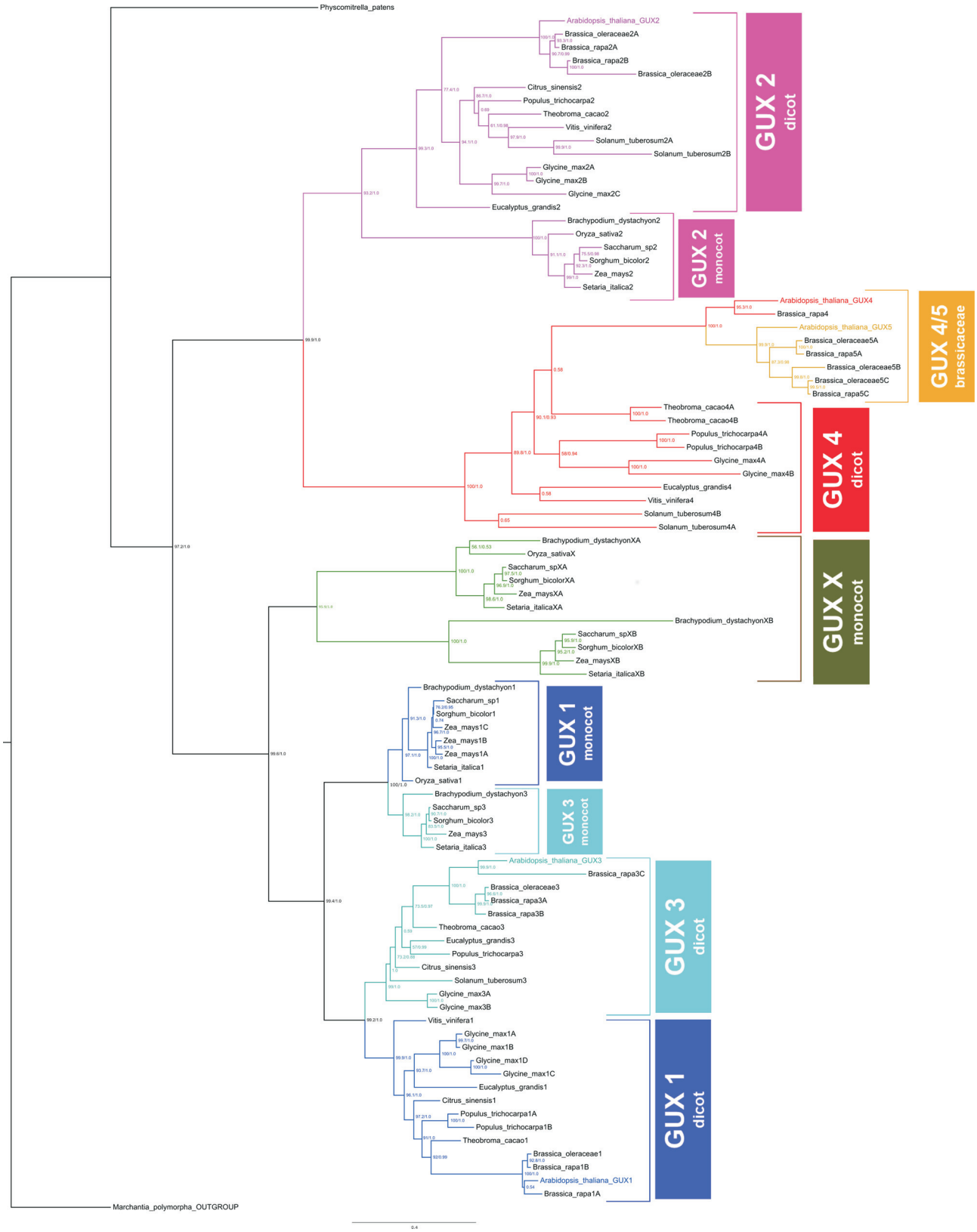



**Figure 1** - Phylogenetic tree of GUX proteins in plants. Numbers on nodes correspond to the maximum likelihood (ML) ultrafast bootstrap support values followed by Bayesian posterior probabilities. The colored branches are represented by: GUX1 (dark blue), GUX2 (pink), GUX3 (light blue), GUX4 (red), GUX4/5 (yellow) and GUX 'X' (green). The GUX sequence from *Marchantya polymorpha* was used as the outgroup.

the clade called GUX 'X', which is composed of few monocots GUX proteins arranged in different places of the tree depending on the dataset used (nucleotides or aminoacids), and hence we could not establish with complete confidence whether these genes are duplications originated from GUX 4 or GUX 1/3. However, the tree derived from aminoacids alignment (Figure 1) presented a stronger support for a relationship with GUX 1/3 (99.6 from bootstrap and 1.0 of posterior probability) than the tree derived from nucleotides alignment that placed this clade as sister of GUX 2 (less than 50 from bootstrap, and 0.5 of posterior probability). Therefore, with caution, we will consider GUX 'X' a specific monocot duplication from GUX 1 or 3 gene.

The GUX proteins are related to the growth and development of cell wall in plants (GUX1 and GUX2 are associated with secondary and GUX3 with primary cell wall development) and have economic importance for biotechnology industry (Lee *et al.*, 2012; Bromley *et al.*, 2013; Mortimer *et al.*, 2015). This highlights the importance of identifying the corresponding genes *in silico* among all species as we showed in our results. With our exhaustive search we were able to identify a great variation among the number of GUX genes in different species. The variation with more than five orthologs may be explained by both ancestral duplications and recent lineage-specific duplications in these plants. For example, at least two late whole-genome duplication events have occurred in *Glycine max* (Schmutz *et al.*, 2010), which can explain the highest number of GUX proteins in this species, with at least two copies of each GUX gene.

It is important to note that we cannot rule out the possibility that some GUX are not included in the genome assembly of these species. Although our description of GUX repertoire suggests a very dynamic evolutionary history, it is still necessary to corroborate these results with improved drafts of some species genomic sequences.

Regarding the evolution pattern observed in the GUX gene family, it can be attributed to a mixture of divergent, concerted and birth-and-death evolutionary models. The divergent model, *i.e.* accumulation of differences between groups that may ultimately lead to the formation of new species/groups (Nei and Rooney, 2005) can be observed in the GUX2 clade. In this case, there is a division between genes from monocots and dicots (Figure 1), indicating that GUX 2 originated before the split between monocots and dicots, and that during evolution they accumulated changes specifics to each group. A similar divergent model was observed in *PHO1* genes, which are involved in phosphate absorption in plants, and where Class II genes from monocots and dicots are separated (He *et al.*, 2013).

The concerted evolution, *i.e.*, members of a gene family evolving in a concerted manner instead of independently (Nei and Rooney, 2005), can be observed in the relationship between the GUX1 and GUX3 clades (Figure 1). Regarding these genes, the phylogenetic tree recovered paralogous clades instead of orthologous clades, indicating that paralogous genes (*e.g.* GUX1 and GUX3 of monocots) are more similar to one another than they are to their true orthologs in closely related species (*e.g.* GUX1 of both monocots and dicots). The clade GUX1 monocot was named this way because BLAST analyses of most of its sequences show the *Arabidopsis* GUX1 as top hits. The same reasoning applies for the clade GUX3 monocot, where most sequences are more similar to *Arabidopsis* GUX3 than GUX1. However, further functional analyses of these proteins are necessary to corroborate the paralog relationship of GUX1 and GUX3 in monocots. The concerted evolution model has also been observed among rice genes from chromosome 11 and 12 that went through a series of genomic modification events until they became more similar among their paralogs than their orthologs (Wang *et al.*, 2007). Furthermore, our analysis of the GUX family revealed characteristics consistent with the birth-and-death evolution model, *i.e.* new genes are originated by successive gene duplication, while some are deleted and others are maintained throughout evolution (Nei *et al.*, 1997), as we have identified lineage-specific patterns of duplication, deletion, and retention of genes among species (Nei and Hughes, 1992). As a result, some species possess fewer GUX (*e.g. Oryza sativa* has lost GUX3 gene), possibly due to deletion or loss-of-function mutations (Figure 2), whereas others possess specific paralogous duplications (*e.g. Solanum tuberosum* has two copies of GUX2 gene, and *Zea mays* has three copies of GUX1 gene). At the same time, we observe that GUX5 is exclusive to the Brassicaceae clade (highlighted in dark red in Figure 2), probably due to a recent duplication of GUX4 specific to this family. According to Blanc *et al.* (2003), *Arabidopsis* experienced two whole genome duplications during its evolution, with the earliest event occurring before the divergence of *Arabidopsis* and *Brassica rapa* (approximately 24-40 Mya). This event may explain the exclusivity of GUX5 in the Brassicaceae family (Figure 2). Accordingly, this Brassicaceae-specific clade was named GUX4/5. Moreover, the uncertainty regarding the GUX 'X' placement may indicate that those proteins arose independently from a monocot-specific duplication, and only functional studies will confirm if they belong to one of the five known GUX clades or if they indeed represent a novel GUX group.

Taken together, our results point to a history of ancestral and recent duplications. It is likely that a duplication event has occurred on a common ancestral of dicots and monocots, originating two copies: one that would give rise to GUX2 and one that would undergo another duplication event originating GUX1 and 3. These three genes seem to correspond to the gene set inherited from the common ancestral of monocots and dicots. After the split event around 140–150 Myr ago that gave rise to each group (Chaw *et al.*, 2004), GUX2 duplicated again only in dicots, originating

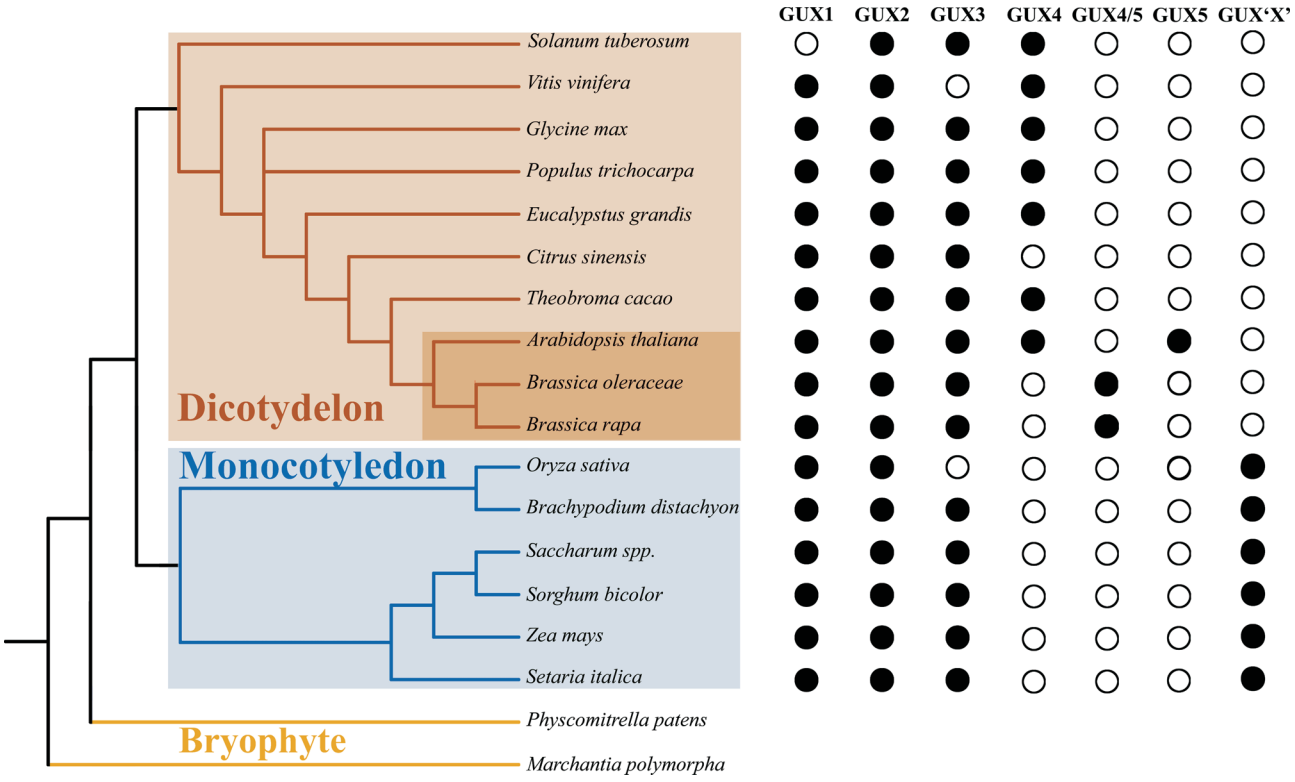



**Figure 2** - Cladogram representing the relationship among all the species surveyed in this study. Black circles represent the presence of gene(s) within a GUX clade and white circles represent the absence of genes within a clade. The distances do not correspond to phylogenetic distances. Orange box highlights dicotyledonous clade. Dark orange box highlights Brassicaceae family. Blue box highlights the monocotyledonous clade. Yellow lines highlight bryophytes as the outgroup.

GUX4, which later duplicated one more time only on the Brassicaceae clade, giving rise to GUX5, specific to this family. Monocots, on the other hand, maintained the ancestral set of GUXs 1, 2 and 3, and they are also likely to have a specific ancestral duplication from GUX1 or 3, named here as GUX 'X' as explained earlier. The functional differences of GUX 1, 2 and 3 shown in *Arabidopsis* by Bromley *et al.* (2013) and Mortimer *et al.* (2015) provide additional support to the evolutionary divergence demonstrated in this study. Figure S5 depicts this history inferred from our phylogenetic analyses.

Polyploidization followed by diploidization events have been frequent during the evolution of flowering plants, which often led to unpredictable and unexplained genomic variation. Consequently, gene loss, widespread modification of methylation patterns, and nonreciprocal chromosomal exchanges may have happened (Doyle *et al.*, 2008). This could explain part of the differences in the numbers of genes between the plants surveyed and also the dynamic history of this gene family, which shows a mixture of evolutionary models.

The first step towards understanding gene function is to know its evolutionary history in the group of interest. Knowing whether a gene is present in the genome as single or several copies, whether there were specific-lineage losses and gains, or whether the duplicates had evolved with an accelerated rate, can bring important inentendisights to better define the scope of further experimental studies. Our results provide a comprehensive overview of GUX proteins among land plants and also important information on their molecular evolutionary history, showing that this gene family has experienced a mixture of evolution models. This study serves as basis for future genetic engineering studies with the GUX family that aims to increase the efficiency of biofuels production.

## Acknowledgments

We thank Beatriz Ribeiro Gallinari for the support with editing images, Paul Dupree and Jan Lyczakowski for critical comments on the manuscript. This study was financially supported by the Comissão de Aperfeiçoamento de Pessoal do Nível Superior (CAPES – Brasil, Finance code 001) and by Fundação de Amparo à Pesquisa (FAPESP; grant number 2017/15895-4).

## Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

## Author Contributions

RHG, RDC and MFN designed and performed experiments. RHG, RDC, PA, MM and MFN analysed data and wrote the paper.

## References

Bailey TL, Bodén M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW and Noble WS (2009) MEME SUITE: Yools for motif discovery and searching. Nucleic Acids Res 37:W202-W208.

Blanc G, Hokamp K and Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis*. Genome Res 3**:**137–144.

Bromley JR, Busse-Wicher M, Tryfona T, Mortimer JC, Zhang Z, Brown DM and Dupree P (2013) GUX1 and GUX2 glucuronyltransferases decorate distinct domains of glucuronoxylan with different substitution patterns. Plant J 74:423-434.

Brown DM, Goubet F, Wong VW, Goodacre R, Stephens E, Dupree P and Turner SR (2007) Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. Plant J 52:1154-1168.

Chaw S, Chang C, Chen H and Li W (2004) Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. J Mol Evol 58:424-441.

Database Resources of the National Center for Biotechnology Information (2016) Nucleic Acids Res 44:D7–D19.

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS and Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet 42:443–61.

He L, Zhao M, Wang Y, Gai J and He C (2013) Phylogeny, structural evolution and functional diversification of the plant *PHOSPHATE1* gene family: a focus on *Glycine max*. BMC Evol Biol 13:103.

Huang X and Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9:868-877.

Jordan DB, Bowman MJ, Braker JD, Dien BS, Hector RE, Lee CC, Mertens A and Wagschal K (2012) Plant cell walls to ethanol. Biochem J 442:241-252.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A and Jermiin LS (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Methods 14:587-589.

Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-780.

Kumar I, Swaminathan K, Hudson K and Hudson ME (2016) Evolutionary divergence of phytochrome protein function in *Zea mays* PIF3 signaling. J Exp Bot 67:4231–4240.

Lee C, O'Neill MA, Tsumuraya Y, Darvill AG and Ye ZH (2007a) The irregular xylem9 mutant is deficient in xylan xylosyltransferase activity. Plant Cell Physiol 48:1624–1634.

Lee C, Zhong R, Richardson EA, Himmelsbach DS, McPhail BT and Ye Z (2007b) The *PARVUS* gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. Plant Cell Physiol 48:1659-1672.

Lee C, Teng Q, Zhong R and Ye Z (2012) *Arabidopsis* GUX proteins are glucuronyltrasnferases responsible for the addition of glucuronic acid side chains onto xylan. Plant Cell Physiol 53:1204-1216.

Loqué D, Scheller HV and Pauly M (2015) Engineering of plant cell walls for enhanced biofuel production. Curr Opin Plant Biol 25:151-161.

Lyczakowski JJ, Wicher KB, Terrett OM, Faria-Blanc N, Yu X, Brown D, Krogh KBRM, Dupree P and Busse-Wicher M (2017) Removal of glucuronic acid from xylan is a strategy to improve the conversion of plant biomass to sugars for bioenergy. Biotechnol Biofuels 10:1-11.

Minh BQ, Nguyen MAT and Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol 30:1188–1195.

Mortimer JC, Miles GP, Brown DM, Zhang Z, Segura MP, Weimar T, Yu X, Seffen KA, Stephens E, Turner SR *et al.* (2010) Absence of branches from xylan in *Arabidopsis gux* mutants reveals potential for simplification of lignocellulosic biomass. Proc Natl Acad Sci U S A 107:17409-17414.

Mortimer JC, Faria-Blanc N, Yu X, Tryfona T, Sorieul M, Ng YZ, Zhang Z, Stott K, Anders N and Dupree P (2015) An unusual xylan in *Arabidopsis* primary cell walls is synthesised by GUX3, IRX9L, IRX10L and IRX14. Plant J 83:413–426.

Nei M and Hughes AL (1992) Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In: 11th Histocompatibility Workshop and Conference, Oxford, p. 27.

Nei M and Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 39:121–152.

Nei M, Gu X and Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune-system. Proc Natl Acad Sci U S A 94:7799–7806.

Park YB and Cosgrove JD (2012) A revised architecture of primary cell walls based on biomechanical changes induced by substrate-specific endoglucanases. Plant Physiol 158:1933-1943.

Peña MJ, Zhong R, Zhou GK, Richardson EA, O'Neill MA, Darvill AG, York WS and Ye ZH (2007) *Arabidopsis* irregular xylem8 and irregular xylem9: Implications for the complexity of glucuronoxylan biosynthesis. Plant Cell 19:549–563.

Rennie EA, Hansen SF, Baidoo EEK, Hadi MZ, Keasling JD and Scheller HV (2012) Three members of the *Arabidopsis* glycosyltransferase family 8 are xylan glucuronosyltransferases. Plant Physiol 159:1408-1417.

Ronquist F and Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Sarkar P, Bosneaga E and Auer M (2009) Plant cell walls throughout evolution: Towards a molecular understanding of their design principles. J Exp Bot 60:3615-3635.

Schmutz J, Cannon S, Schlueter J, Ma J, Mitros T, Nelson W, Hyten D, Song Q, Thelen J, Cheng J *et al.* (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Simmons TJ, Mortimer JC, Bernardinelli OD, Pöppler A, Brown SP, de Azevedo ER, Dupree R and Dupree P (2016) Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR. Nat Commun 7:1-9.

Sørensen I, Domozych D and Willats WGT (2010) How have plant cell walls evolved? Plant Physiol 153:366-372.

Trifinopoulos J, Nguyen LT, von Haeseler A and Minh BQ (2016) W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. Nucl Acids Res 44:W232-W235.
Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MI, Henrique-Silva F, Giglioti EA, Lemos MV, Coutinho LL *et al.* (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res 13:2725-2735.
Wang X, Tang H, Bowers JE, Feltus FA and Paterson AH (2007) Extensive concerted evolution of rice paralogs and the road to regaining independence. Genetics 177:1753-63.
Yue D, You F and Snyder SW (2014) Biomass-to-bioenergy and biofuel supply chain optimization: overview, key issues and challenges. Comput Chem Eng 66:36-56.

## Internet Resources

Rambaut A (2009) FigTree version 1.4.2, http://tree.bio.ed.ac.uk.
HMMER v3.1b2, http://hmmer.org/.
PFAM - PF01501, http://pfam.xfam.org/family/PF01501.
Phytozome v12, https://phytozome.jgi.doe.gov.
TAIR database, https://www.arabidopsis.org/.
SUCEST database, http://sucest-fun.org/.

## Supplementary material

The following online material is available for this article?
Figure S1 – Scheme of the methodology used for the screening of genes.
Figure S2 – Conserved motif identified among the five GUX sequences described in *Arabidopsis* and two rice GUX sequences.
Figure S3 – GUX motif highlighted in the sequence alignment used to generate the phylogenetic tree.
Figure S4 – Phylogenetic tree of *GUX* coding sequences in plants.
Figure S5 – A hypothesis of the evolutionary history of GUX genes.
Table S1 – ESTs used to produce contigs for each GUX gene in sugarcane.

*Associate Editor: Marcia Pinheiro Margis*

INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL

31/05/2019 870190051190 12:04

29409161812001036

# Pedido nacional de Invenção, Modelo de Utilidade, Certificado de Adição de Invenção e entrada na fase nacional do PCT

**Número do Processo:** BR 10 2019 011260 3

## Dados do Depositante (71)

**Depositante 1 de 1**

**Nome ou Razão Social:** UNIVERSIDADE ESTADUAL DE CAMPINAS - UNICAMP

**Tipo de Pessoa:** Pessoa Jurídica

**CPF/CNPJ:** 46068425000133

**Nacionalidade:** Brasileira

**Qualificação Jurídica:** Instituição de Ensino e Pesquisa

**Endereço:** Cidade Universitária Zeferino Vaz

**Cidade:** Campinas

**Estado:** SP

**CEP:** 13084-971

**País:** Brasil

**Telefone:**

**Fax:**

**Email:** patentes@inova.unicamp.br

PETICIONAMENTO ELETRÔNICO

Esta solicitação foi enviada pelo sistema Peticionamento Eletrônico em 31/05/2019 às 12:04, Petição 870190051190

## Dados do Pedido

**Natureza Patente:** 10 - Patente de Invenção (PI)

**Título da Invenção ou Modelo de Utilidade (54):** PROCESSO PARA A PRODUÇÃO DE CANA-DE-AÇÚCAR TRANSGÊNICA COM MODIFICAÇÃO NA PAREDE CELULAR

**Resumo:** A presente invenção refere-se a um processo para a produção de cana-de-açúcar com modificação na parede celular, utilizando uma abordagem de RNAi para reduzir os níveis de expressão do gene homólogo de GUX em cana-de-açúcar (ScGUX2). Foi observado que a recalcitrância da biomassa nas plantas transgênicas foi reduzida e um aumento de pelo menos 15% no rendimento da sacarificação foi observado em plantas transgênicas de cana-de-açúcar. Descobriu-se também que nenhum efeito negativo foi observado em vários parâmetros agronômicos, como tamanho, peso, número de colmos e Brix. A parede celular das plantas transgênicas não apresentou redução dos níveis de celulose, hemicelulose e pectina. Os dados obtidos mostram que o processo sugerido na presente invenção apresenta um potencial comercial muito elevado, já que os eventos mostraram um aumento na sacarificação de pelo menos 15%, chegando até 30%, o que em larga escala representa um aumento na produção de etanol muito considerável.

**Figura a publicar:** 6

## Dados do Procurador

Procurador:

**Nome ou Razão Social:** Raquel Moutinho Barbosa

**Numero OAB:**

**Numero API:**

**CPF/CNPJ:** 30579878856

**Endereço:** Rua Roxo Moreira, 1831

**Cidade:** Campinas

**Estado:** SP

**CEP:** 13083-591

**Telefone:** 1935212620

**Fax:**

**Email:** raquel.barbosa@inova.unicamp.br

## Dados do Inventor (72)

### Inventor 1 de 3

**Nome:** PEDRO ARAÚJO
**CPF:** 33847130870
**Nacionalidade:** Brasileira
**Qualificação Física:** Biólogo, biomédico e afins
**Endereço:** Alexandre Cazelatto, 2689 casa E3
**Cidade:** Paulínia
**Estado:** SP
**CEP:** 13148-911
**País:** BRASIL
**Telefone:** (19) 352 16234
**Fax:**
**Email:** paraujo@unicamp.br

### Inventor 2 de 3

**Nome:** MARCELO MENOSSI TEIXEIRA
**CPF:** 65482069915
**Nacionalidade:** Brasileira
**Qualificação Física:** Professor do ensino superior
**Endereço:** Rua Dalva Maria Fernandes Pinheiro, 120
**Cidade:** Campinas
**Estado:** SP
**CEP:** 13085-075
**País:** BRASIL
**Telefone:** (19) 352 16236
**Fax:**
**Email:** menossi@unicamp.br

### Inventor 3 de 3

**Nome:** RAFAEL HENRIQUE GALLINARI

**CPF:** 38825346808

**Nacionalidade:** Brasileira

**Qualificação Física:** Biólogo, biomédico e afins

**Endereço:** Rua Arnaldo Verinaud, 83, Residencial Geneve

**Cidade:** Campinas

**Estado:** SP

**CEP:** 13049-450

**País:** BRASIL

**Telefone:** (19) 352 16234

**Fax:**

**Email:** rafgallinari@gmail.com

## Documentos anexados

| Tipo Anexo | Nome |
|---|---|
| Relatório Descritivo | 1374_RELATORIO DESCRITIVO_290519.pdf |
| Desenho | 1374_DESENHOS_280519.pdf |
| Resumo | 1374_RESUMO_290519.pdf |
| Reivindicação | 1374_REIVINDICACOES_290519.pdf |
| Comprovante de pagamento de GRU 200 | 1374_GRU_211118.pdf |
| Procuração | Procuracao Unicamp INPI_210519.pdf |

## Sequências Biológicas

☑ Declaro que a informação contida na 'Listagem de Sequências' apresentada em formato eletrônico está limitada ao conteúdo da matéria revelada pelas sequências de aminoácidos e/ou de nucleotídeos divulgadas no pedido de patente, conforme depositado

| Tipos de Sequências Biológicas | Nome |
|---|---|
| Listagem de Sequências Biológicas em formato TXT | Sequencia_ST25.txt |

## Acesso ao Patrimônio Genético

☑ Declaração Negativa de Acesso - Declaro que o objeto do presente pedido de patente de invenção não foi obtido em decorrência de acesso à amostra de componente do Patrimônio Genético Brasileiro, o acesso foi realizado antes de 30 de junho de 2000, ou não se aplica.

## Declaração de veracidade

☑ Declaro, sob as penas da lei, que todas as informações acima prestadas são completas e verdadeiras.

# PROCESSO PARA A PRODUÇÃO DE CANA-DE-AÇÚCAR TRANSGÊNICA COM MODIFICAÇÃO NA PAREDE CELULAR

## CAMPO DA INVENÇÃO

[1] A presente invenção se insere no campo da Biologia, mais precisamente na área da modificação genética, e descreve um processo para a produção de cana-de-açúcar com modificação na parede celular apresentando gene que aumenta a sacarificação.

## FUNDAMENTOS DA INVENÇÃO

[2] A hemicelulose e a celulose são polissacarídeos essenciais para o desenvolvimento das plantas e principais componentes da parede celular. Eles também são uma importante fonte de energia para a produção de etanol a partir de biomassa vegetal. No entanto, a sua conversão em açúcares fermentáveis é dificultada pela estrutura complexa das paredes celulares. As enzimas de substituição do ácido glucurônico em Xilano (GUX) liga o ácido glucurônico ao xilano, um dos principais componentes da hemicelulose, diminuindo a eficiência das enzimas utilizadas na produção de etanol. Mutantes de GUX com perda de função na planta *Arabidopsis thaliana* melhoram a acessibilidade de enzimas e a digestão da parede celular sem mostrar fenótipos adversos.

[3] Dados da literatura mostram que o gene AtGUX2 de Arabidopsis é responsável pela deposição de uma molécula (ácido glucurônico) na hemicelulose, conferindo uma maior complexidade na parede celular, aumentando a recalcitrância da mesma, o que dificulta o acesso das enzimas que degradam a biomassa da parede e consequentemente, reduzindo a eficiência na produção de etanol 2G. Esse gene também não causou alterações no desenvolvimento dessa planta modelo (Mortimer et al., 2010). Porém, nenhum estudo foi feito em plantas de interesse agronômico.

[4] O gene usado nas plantas transgênicas de cana, ScGUX2 guarda pouca semelhança com o gene AtGUX2 de Arabidopsis, com o qual temos uma informação sobre um novo gene. Ao nível de sequência de DNA, a similaridade é extremamente baixa. Ao comparar as duas proteínas, há uma identidade de apenas 45%. Também foi verificado que o silenciamento de ScGUX2 em cana não causa impactos para as plantas transgênicas, nem ao desenvolvimento e nem de características agronômicas (ex. tamanho, largura e altura do colmo, Brix, etc.). Assim, esses dados mostram que esta estratégia tem um potencial comercial muito elevado, já que os eventos mostraram um aumento na sacarificação de pelo menos 15%, chegando até 30%, o que em larga escala representa um aumento na produção de etanol muito considerável.

**ESTADO DA TÉCNICA**

[5] Existem no estado da técnica, alguns trabalhos descrevendo a modificação na parede celular de espécies de plantas, incluindo a cana-de-açúcar, através da modificação de alguns genes, por exemplo:

[6] O documento US 2017/0107542 A1 mostra que plantas transgênicas tendo a expressão de certas proteínas alteradas, leva a uma diminuição da recalcitrância, aumento no crescimento, diminuição no conteúdo de lignina e, consequente aumento na produção de biocombustíveis, como etanol, entre outros. Entretanto, a caracterização dos genes IRX10 e IRX10-L em *Populus* é bastante distante do gênero *Saccharum*. Além disso, na presente invenção, não foi observada qualquer alteração fenotípica.

[7] O documento US 2014/0033365 A1 mostra que a inibição de um determinado gene na parede celular em certas espécies de plantas, aumenta a produção de biocombustíveis. Entretanto, tais plantas apresentam fenótipo anão, inviável comercialmente. Na presente invenção, as plantas tiveram incremento de sacarificação sem penalidade no crescimento e desenvolvimento das mesmas.

[8] O documento intitulado AN EVEN PATTERN OF XYLAN SUBSTITUTION IS CRITICAL FOR INTERACTION WITH CELLULOSE IN PLANT CELL WALLS, em nome de Grantham *et al. 2017*, mostra a influência na substituição de um padrão na substituição de xilano para sua interação com celulose na parede celular em plantas. Tal alteração tem um impacto na estrutura da parede celular das plantas e melhora o biorefinamento. Contudo, apresentam um fenótipo prejudicial ao desenvolvimento da planta, o que não acontece na presente invenção, visto que o fenótipo das mesmas não tem nenhuma relação prejudicial ao desenvolvimento das plantas.

[9] O documento intitulado RNAi SUPPRESSION OF LIGNIN BIOSYNTHESIS IN SUGARCANE REDUCES RECALCITRANCE FOR BIOFUEL PRODUCTION FROM LIGNOCELLULOSIC BIOMASS, em nome de Jung *et al. 2012*, mostra que uma redução moderada na lignificação na cana-de-açúcar por interferência de RNA, reduz a recalcitrância da biomassa da cana e consequente aumento na produção de biocombustível. Entretanto, sabe-se que a redução de lignina causa nanismo, maior susceptibilidade à infecções por diferentes patógenos. Além disso, reduções nos parâmetros S/G são indesejáveis para biomassa. A presente invenção, ao contrário, não visa a redução da lignina. Sua via é complexa e com diversas penalidades já conhecidas na literatura.

[10] O documento intitulado REMOVAL OF GLUCURONIC ACID FROM XYLAN IS

A STRATEGY TO IMPROVE THE CONVERSION OF PLANT BIOMASS TO SUGARS FOR BIOENERGY, em nome de Lyczakowski *et al. 2017*, mostra que a remoção do ácido glucurônico do xilano em espécies de plantas, neste caso aqui, a Arabidopsis acarreta em um aumento de 30% de glicose e mais de 700% de xilose liberados durante a sacarificação, e consequente aumento no rendimento na produção de etanol. Todavia, o gene é caracterizado em planta modelo e filogeneticamente distante do gênero *Saccharum*. Além disso, os alinhamentos mostram que a sequência tem baixíssima similaridade com as de cana-de-açúcar o que inviabiliza a condução óbvia de um homólogo funcional. A identificação do gene GUX em cana-de-açúcar, como na presente invenção, não é óbvia através dos alinhamentos. Monocotiledôneas não possuem trabalhos mostrando funcionalmente o gene silenciado.

[11] O documento intitulado REGULATION OF LIGNIN BIOSYNTHESIS THROUGH RNAi IN AID OF BIOFUEL PRODUCTION, em nome de Kumari *et al. 2015*, mostra que a modificação genética da parede celular de diversas espécies, incluindo a cana-de-açúcar pode levar a uma diminuição da recalcitrância e aumento da sacarificação, levando a uma maior produção de biocombustíveis, como o etanol. Entretanto, em tal documento há a utilização de RNAi para o controle genético de lignina. Já na presente invenção, o gene tem atuação em hemicelulose e não lignina. Apesar de utilizar-se RNAi, que é uma técnica amplamente empregada, há ainda a utilização de um vetor específico e uma região ímpar para silenciamento.

[12] O documento intitulado LIGNIN BIOSYNTHESIS PERTURBATIONS AFFECT SECONDARY CELL WALL COMPOSITION AND SACCHARIFICATION YIELD IN *Arabidopsis thaliana*, em nome de Acker *et al. 2013*, relata o efeito na alteração da biossíntese da lignina presente na parede celular da *Arabidopsis* e consequente aumento no rendimento da sacarificação, bem como diminuição da recalcitrância da parede celular da mesma. Percebe-se que são genes completamente diferentes que resultam em aumento da sacarificação. Genes da via da lignina geralmente prejudicam o desenvolvimento da planta ou deixam mais propensas a doenças. Já na presente invenção, as plantas GUX alteram outros componentes da parede celular.

[13] Assim, conforme observado, a presente invenção apresenta a vantagem de se utilizar a cana-de-açúcar em detrimento a outras espécies vegetais comumente utilizadas. Dados mostram que a cana-de-açúcar tem o melhor custo-benefício para a produção de etanol. Cabe ressaltar que diversas modificações na parede celular trazem prejuízo para o desenvolvimento da planta, por exemplo, quanto ao seu crescimento e susceptibilidade a doenças. O gene da presente invenção também atua na parede

celular, a exemplo de outros genes já caracterizados, mas apresenta a vantagem de não alterar características agronômicas importantes. Em *Arabidopsis thaliana*, planta modelo utilizada para provas de conceito em biologia molecular, apresenta 4.000 (15%) dos genes associados à parede celular e 26.500 proteínas traduzidas. Sendo assim, a tarefa de caracterizar e selecionar genes comercialmente importantes, sem prejuízos agronômicos, não é trivial.

[14] Não obstante, sabe-se que a sacarificação é o principal elo/parâmetro para a produção de etanol, onde quanto mais açúcar estiver disponível, maior será a produção do combustível. Podendo sofrer ajustes no processo industrial para a otimização. Os resultados obtidos quando comparados com planta controle mostram aproximadamente 20% mais açúcar fermentável disponível com a mesma quantidade de biomassa. O aumento de açúcares derivados da parede celular e a redução de insumos gastos durante o processo de produção são elementos-chave para viabilizar a produção de etanol de 2ª geração.

**SUMÁRIO DA INVENÇÃO**

[15] A presente invenção tem por objetivo propor um processo para a produção de cana-de-açúcar transgênica com modificação na parede celular apresentando gene que aumenta a sacarificação. Uma abordagem de RNAi foi utilizada para reduzir os níveis de expressão do gene homólogo de GUX em cana-de-açúcar (*ScGUX2*). A recalcitrância da biomassa nas plantas transgênicas foi reduzida e um aumento de pelo menos 15% no rendimento da sacarificação foi observado em plantas transgênicas de cana-de-açúcar. Nenhum efeito negativo foi observado em vários parâmetros agronômicos, como tamanho, peso, número de colmos e Brix. A parede celular das plantas transgênicas não apresentou redução dos níveis de celulose, hemicelulose e pectina. Os dados obtidos mostram que esta estratégia tem um potencial comercial muito elevado, já que os eventos obtidos na concretização da presente invenção mostraram um aumento na sacarificação de pelo menos 15%, chegando até 30%, porém não se limita a essa faixa, o que, em larga escala, representa um aumento significativo na produção de etanol.

[16] Mais especificamente o processo proposto para a produção de cana-de-açúcar transgênica apresenta reduzidos níveis de transcritos do gene *ScGUX2* com modificação na parede celular e compreende ainda as seguintes etapas:

(a) Seleção de um gene de cana-de-açúcar homólogo a um gene que codifica proteína envolvida no depósito de ácido glucurônico na parece celular vegetal;

(b) Manipulação do genoma vegetal que reduza a expressão do gene *ScGUX2*;

(c) Análises de linhagens independentes de cana-de-açúcar com genoma modificado; e

(d) Avaliação da sacarificação da biomassa vegetal.

[17] A seleção na etapa “a” do gene *ScGUX2* de cana-de-açúcar utiliza o gene homólogo de *Arabidopsis thaliana* denominado *AtGUX2* e o gene homólogo de sorgo denominado Sb01g044930. O gene *AtGUX2* da *Arabidopsis thaliana* compreender a SEQ. ID. No. 1, o gene homólogo de sorgo compreender a SEQ. ID. No. 2, o gene da cana-de-açúcar *ScGUX2* compreender a SEQ. ID. No. 3, correspondente a região codificadora.

[18] A redução da expressão do gene *ScGUX2* na etapa “b” compreende as seguintes sub-etapas:

(i) Seleção de região do gene *ScGUX2* para silenciamento gênico;

(ii) Construção de DNA para produção de RNA de interferência composta por promotor ligado operacionalmente a uma região de DNA do gene *ScGUX2* para formação de grampo de RNA, ligada operacionalmente a uma região terminadora;

(iii) Modificação do genoma da cana-de-açúcar mediante transgenia.

[19] A região do gene *ScGUX2* usada na construção de RNA de interferência contem a sequência de 498 pb, representada pela SEQ. ID. No. 4 e a região de DNA contendo grampo de DNA com a região do gene *ScGUX2* contém a SEQ. ID. No. 4 clonada em orientação *sense* e *antisense*, tendo um íntron entre ambas, representada pela SEQ. ID. No. 5.

[20] As plantas apresentam aumento de no mínimo 15% nos níveis de sacarificação da biomassa vegetal comparativamente ao tipo selvagem.

[21] O gene silenciado é homólogo ao gene *ScGUX2* e apresenta a função de modificar a composição e estrutura da parede celular para aumento da sacarificação da biomassa vegetal.

## BREVE DESCRIÇÃO DAS FIGURAS

[22] Para obter uma total e completa visualização do objetivo desta invenção, são apresentadas as figuras as quais se faz referências, conforme segue.

[23] A Figura 1 apresenta o perfil de expressão *northern* virtual ScGUX2. Meristema apical (AM), flores em diferentes estágios de desenvolvimento (FL1, FL3, FL4, FL5, FL8), sementes (SD), rolo de folhas de plantas imaturas (LR1, LR2), folhas estioladas de plântulas cultivadas *in vitro* (LV1), brotos laterais de plantas maduras (LB1, LB2), tronco (ST1, ST3), casca de caule de plantas

adultas de cana-de-açúcar (SB1), raízes de plantas jovens (RZ1, RZ2, RZ3), raízes de plantas maduras (RT1, RT2, RT3), calo (CL6), mistura de tecidos da raiz à zona de broto, caule e meristema apical de plântulas cultivadas *in vitro* e infectadas com *Gluconacetobacter diazotroficans* (AD1), mistura de tecidos da raiz à zona de broto, caule e meristema apical de plântulas cultivadas *in vitro* e infectadas com *Herbaspirilum diazotroficans* (HR1);

[24] A Figura 2 apresenta os resultados do alinhamento de MAFFT entre sequências de nucleotídeos de AtGUX2 (SEQ. ID. No. 1) e ScGUX2 (SEQ. ID. No. 3) usando o método de refinamento iterativo L-INS-I;

[25] A Figura 3 apresenta os resultados do alinhamento de MAFFT entre sequências de proteínas AtGUX2 (SEQ. ID. No. 1) e ScGUX2 (SEQ. ID. No. 3) usando o método de refinamento iterativo L-INS-I;

[26] A Figura 4 apresenta a árvore filogenética GUX2. Os números indicam os valores de clado de suporte de *bootstrap* de máxima verossimilhança/probabilidades posteriores bayesianas. A planta *Marchantya polymorpha* foi usada como um grupo externo (não mostrado);

[27] A Figura 5 apresenta o esquema da construção de DNA visando a produção de um transcrito que forme um grampo (também denominado "*hairpin"* em inglês), contendo os fragmentos *sense* e *antisense* de *ScGUX2*. Os sítios de restrição para asGUX2 são BamHI e EcoRV e para sGUX2 são MluI e KpnI. Ambas as sequências (asGUX2 e sGUX2) têm 498 pb;

[28] A Figura 6 apresenta o esquema do vetor pCR8_GUX2_RNAi. Os sítios de restrição BamHI e KpnI foram adicionados no vetor pCR8 para permitir a ligação com o presente vetor de silenciamento (Figura 3; asGUX2 e sGUX2);

[29] A Figura 7 apresenta a construção pGVG_GUX2_RNAi. A região ampliada corresponde ao *grampo* contendo o fragmento *ScGUX2* de 498 pb nas orientações *antisense* e *sense*. A expressão é conduzida pelo promotor ZmUbi1 (incluindo o *exon* 5' não traduzido e o primeiro *intron*) e o terminador 35M do CaMV. O vetor contém o gene NPTII como marcador selecionável de plantas sob controle do promotor CaMV 35S aprimorado;

[30] A Figura 8 apresenta a expressão gênica de *ScGUX2* em eventos silenciados e controle por qPCR. Tipo selvagem (C+), vetor vazio pGVG (VV) e eventos independentes silenciados para *ScGUX2* (I.10, I.12, I.15, I.2, I.3, I.4, I.5 e I. 9). Valores representam a expressão relativa Δct;

[31] A Figura 9 apresenta o esquema representando a linha do tempo durante o período

de crescimento (7 meses) e rebrotamento (9 meses) das plantas de cana-de-açúcar;

[32] A Figura 10 apresenta o esquema de cana-de-açúcar e os colmos utilizados nas análises. O "x" representa os entrenós dos colmos coletados;

[33] A Figura 11 apresenta o teor de celulose em plantas de cana-de-açúcar. Tipo selvagem (C+), vetor vazio pGVG (VV) e eventos independentes silenciados para *ScGUX2* (I.10, I.12, I.15, I.2, I.3, I.4, I.5 e I. 9). As barras de erro representam o desvio padrão;

[34] A Figura 12 apresenta o teor de pectina em plantas de cana-de-açúcar. Tipo selvagem (C+), vetor vazio pGVG (VV) e eventos independentes silenciados para *ScGUX2* (I.10, I.12, I.15, I.2, I.3, I.4, I.5 e I. 9). As barras de erro representam o desvio padrão;

[35] A Figura 13 apresenta o teor de hemicelulose na cana-de-açúcar. Tipo selvagem (C+), vetor vazio pGVG (VV) e eventos independentes silenciados para *ScGUX2* (I.10, I.12, I.15, I.2, I.3, I.4, I.5 e I. 9). As barras de erro representam o desvio padrão;

[36] A Figura 14 apresenta a quantificação da sacarificação na cana-de-açúcar em porcentagem (%) em relação a massa inicial de 5 mg de celulose. Os dados do tipo selvagem (C+), do vetor vazio pGVG (VV) e dos eventos independentes silenciados para *ScGUX2* são mostrados. Os eventos estatisticamente diferentes são representados com asterisco. As barras de erro representam o desvio padrão. (n = 8 para C e VV; n = 7 para I.15; n = 5 para I.10, I.12, I.3 e I.5; n = 4 para I.2 e I.9, ANOVA);

[37] A Figura 15 apresenta o número de colmos das plantas usadas em todos os experimentos. Os dados do tipo selvagem (C+), do vetor vazio pGVG (VV) e dos eventos independentes silenciados para *ScGUX2* são mostrados. As barras de erro representam o desvio padrão. (n = 8 para C+ e VV; n = 7 para I.15; n = 5 para I.10, I.12, I.3 e I.5; n = 4 para I.2 e I.9);

[38] A Figura 16 apresenta o diâmetro do colmo maduro das plantas. Os dados do tipo selvagem (C+), do vetor vazio pGVG (VV) e dos eventos independentes silenciados para *ScGUX2* são mostrados. As barras de erro representam o desvio padrão. (n = 8 para C+ e VV; n = 7 para I.15; n = 5 para I.10, I.12, I.3 e I.5; n = 4 para I.2 e I.9);

[39] A Figura 17 apresenta o grau Brix de colmos maduros. Os dados do tipo selvagem (C+), do vetor vazio pGVG (VV) e dos eventos independentes silenciados para *ScGUX2* são mostrados. As barras de erro representam o desvio padrão. (n = 8 para C+ e VV; n = 7 para I.15; n = 5 para I.10, I.12, I.3 e I.5; n = 4 para I.2 e I.9);

[40] A Figura 18 apresenta a altura das plantas de cana-de-açúcar. Os dados do tipo selvagem (C+), do vetor vazio pGVG (VV) e dos eventos independentes silenciados para *ScGUX2* são mostrados. As barras de erro representam o desvio padrão. (n = 8 para C+ e VV; n = 7 para I.15; n = 5 para I.10, I.12, I.3 e I.5; n = 4 para I.2 e I.9); e

[41] A Figura 19 apresenta o peso fresco dos colmos. Os dados do tipo selvagem (C+), do vetor vazio pGVG (VV) e dos eventos independentes silenciados para *ScGUX2* são mostrados. As barras de erro representam o desvio padrão. (n = 8 para C+ e VV; n = 7 para I.15; n = 5 para I.10, I.12, I.3 e I.5; n = 4 para I.2 e I.9).

**DESCRIÇÃO DETALHADA DA INVENÇÃO**

[42] A presente invenção refere-se a um processo para a produção de cana-de-açúcar com modificação na parede celular apresentando gene que aumenta a sacarificação, compreendendo as seguintes etapas:

(a) Seleção de um gene homólogo da cana-de-açúcar;

(b) Análise *in silico* e Filogenética;

(c) Seleção da sequência para silenciamento gênico;

(d) Construção de vetor para silenciar o gene *ScGUX2*;

(e) Transformação da variedade de cana-de-açúcar SP80-3280;

(f) Análises de linhagens independentes de cana-de-açúcar transgênica; e

(g) Avaliação da composição da parede celular e sacarificação.

[43] As referidas etapas serão descritas em detalhes a seguir.

**(a) Seleção de um gene homólogo da cana-de-açúcar**

[44] Primeiramente, utilizou-se o gene *AtGUX2* da *Arabidopsis thaliana* (SEQ. ID. No. 1 - acesso At4G33330) como isca para selecionar um gene homólogo de sorgo com pelo menos 70% de similaridade com a sequência usando o banco de dados NCBI. A partir do gene homólogo de sorgo (SEQ. ID. No. 2) foi feita uma busca usando o banco de dados SUCEST (Vettore et al., 2001) para encontrar o homólogo de cana. O gene da cana-de-açúcar foi denominado *ScGUX2* e a sequência SEQ. ID. No. 3 corresponde à região codificadora, obtida da sequência SCCCST3006B11.g da cana-de-açúcar do banco de dados SUCEST.

[45] Um perfil de expressão de *Northern blot* virtual do gene *ScGUX2* mostrou uma expressão mais alta no caule (Figura 1). Como as hastes fornecem a maior parte da biomassa da cana-de-

açúcar, a manipulação de um gene com esse perfil de expressão pode ter um impacto importante na sacarificação.

**(b) Análise Filogenética com as sequências presentes no mesmo clado**

[46] A fim de identificar ambas as sequências de nucleotídeos GUX para Arabidopsis e cana-de-açúcar, as mesmas foram alinhadas com o MAFFT (Katoh & Standley 2013) usando o método de refinamento iterativo L-INS-I. Os alinhamentos foram feitos para nucleotídeo (Figura 2) e sequências de proteínas (Figura 3).

[47] Para caracterizar ainda mais o *ScGUX2* como o homólogo de cana-de-açúcar de *AtGUX2*, cinco sequências de proteínas GUX para *Arabidopsis thaliana* foram recuperadas (Mortimer et al., 2010; Reenie et al., 2012) do GenBank (https://www.ncbi.nlm.nih.gov)). Assim, identificou-se a GUX em 15 outras espécies de angiospermas (incluindo seis monocotiledôneas e nove dicotiledôneas): bromopeia roxa (*Brachypodium distachyon*), repolho selvagem (*Brassica oleraceae*), nabo (*Brassica rapa*), laranja doce (*Citrus sinensis*), eucalipto (*Eucalyptus grandis*), soja (*Glycine max*), arroz (*Oryza sativa*), algodoeiro preto (*Populus trichocarpa*), cana-de-açúcar (*Saccharum spp.*), painço (*Setaria italica*), batata (*Solanum tuberosum*), sorgo (*Sorghum bicolor*), cacau (*Theobroma cacao*), uva (*Vitis vinifera*) e milho (*Zea mays*). Também se identificou a GUX em duas briófitas (o musgo *Physcomitrella patens* e a hepática comum *(Marchantia polymorpha*) para servir como grupos externos para análises filogenéticas adicionais.

[48] Todas as sequências de proteínas (somente a partir de transcritos primários) foram recuperadas da versão mais recente do genoma de referência disponível no Phytozome v12 (https://phytozome.jgi.doe.gov). Todas as enzimas GUX possuem o domínio da família da glicosil-transferase 8 (GT8), responsável pela adição de substituições de glucuronosil no esqueleto de xilano(Reenie et al., 2012). Portanto, todas as sequências de proteínas foram rastreadas com uma pesquisa oculta do modelo de Markov (HMM) (*hmmsearch* do HMMER v3.1b2; http://hmmer.org/) usando o GT8 HMM disponível no PFAM (PF01501; http://pfam.xfam.org/family/PF01501). Como nem todas as proteínas que possuem o domínio GT8 são proteínas GUX, preocupou-se em identificar um motivo específico de GUX. Para isso, a análise de MEME (Bailey et al. 2009) foi realizada usando as cinco sequências de proteínas GUX descritas para Arabidopsis (Mortimer et al. 2010; Rennie et al. 2012) e duas sequências de arroz identificadas pela HomoloGene (coordenador de recursos do NCBI, 2016) como entrada. O motivo presente em todas aquelas sequências Gux foi utilizado para rastrear todas as

sequências proteicas GT8 em uma análise HMMER subsequente. Finalmente, sequências putativas de GUX foram definidas para cada espécie quando o domínio GT8 e o motivo específico da GUX estavam presentes.

[49] Dentre as 18 espécies pesquisadas, a cana-de-açúcar é a única que não possui um genoma de referência disponível no Phytozome. Assim, suas proteínas GUX foram identificadas realizando buscas no BLAST no banco de dados do SUCEST (Vettore et al. 2003) usando os homólogos do sorgo como referências. Em seguida, o programa de montagem contig CAP3 (Huang & Madan 1999) foi usado com as tags de sequência expressa (ESTs) obtidas da pesquisa BLAST para montar contigs para cada gene GUX na cana-de-açúcar. Para contigs com transcritos incompletos, o homólogo de sorgo mais próximo foi usado para completar a sequência.

[50] Após identificar as sequências de proteínas GUX para cada espécie, as mesmas foram alinhadas com o MAFFT (Katoh & Standley 2013) usando o método de refinamento iterativo L-INS-i. A análise filogenética de máxima verossimilhança do alinhamento de múltiplas sequências da GUX foi realizada usando o IQ-Tree v1.6.1 (Trifinopoulos et al. 2016). Suporte de filiais foi adquirido por 1.000 pseudoreplicados de bootstraps ultra-rápidos (Minh et al. 2013), sob o modelo GTR + I + G4 identificado por ModelFinder (Kalyaanamoorthy et al. 2017). Para a análise filogenética Bayesiana, utilizou-se o MrBayes v3.2.6 (Ronquist & Huelsenbeck 2003), com 1.000.000 de gerações, frequência de amostragem de 500 e frequência de diagnóstico de 5.000, sob o modelo de evolução GTR + I + G. Árvores filogenéticas foram visualizadas e editadas no FigTree v1.4.3 (Rambaut 2009). Além disso, a robustez dos arranjos de clado foi testada executando as mesmas análises com sequências de codificação, sequências completas de genes (incluindo íntrons, exons e UTRs) e sequências de genes contendo a região codificante mais 1 kb das regiões flanqueadoras 5 'e 3'.

[51] Interessantemente, a proximidade filogenética às monocotiledôneas indica que as proteínas GUX2 têm o clado mais conservado e estabelecido entre monocotiledôneas e dicotiledôneas (Figura 4) em comparação com as GUX1, GUX3 e GUX4 (dados não mostrados). A evolução dos genes da GUX parece ser conservada entre as plantas, uma vez que foi originada durante a colonização terrestre (Jensen et al., 2018).

[52] Todas estas evidências *in silico* contribuem para o suporte de que a SEQ. ID. 3 corresponde ao homólogo AtGUX2 na cana-de-açúcar.

**(c) <u>Seleção da sequência para silenciamento gênico</u>**

[53] A ferramenta on-line SiRNA do Galaxy Pasteur (https://galaxy.pasteur.fr/#forms::sirna) foi utilizada para obter as sequências mais prováveis de formar um grampo do gene da cana *ScGUX2*. O limiar para as sequências foi estabelecido como uma pontuação superior a 6. Todas as sequências obtidas foram alinhadas utilizando BioEdit v.7.0.5 com CAP3. Dentre as sequências com alta pontuação para formação de grampo, a sequência de 498 pb (SEQ. ID. No. 4) foi escolhida devido à sua especificidade para o *ScGUX2*, pois não é complementar a outros genes da GUX identificados na cana-de-açúcar, incluindo o gene *ScGUX1*, sequência montada de cana-de-açúcar SCCCST3006B11.g (dados não mostrados).

**(d) Construção de vetor para silenciar o gene ScGUX2**

[54] Os iniciadores projetados para amplificar os fragmentos *sense* e *antisense* a partir do gene ScGUX2 são mostrados na Tabela 1. Estes iniciadores contêm adicionalmente às sequências *ScGUX2*, sequências de locais de restrição para permitir a clonagem das sequências amplificadas *sense* e *antisense* como indicado na Figura 5. A região sublinhada corresponde aos locais de restrição para a construção do vetor (pUbi_hp_GAI backbone). **BamHI**: ggatcc; **EcoRV**: gatatc; Mlul: acgcgt e **Kpnl**: ggtacc.

**Tabela 1:** Iniciadores projetados para construir o vetor pUBI_hp_GUX2.

| Iniciador | Sequência |
|---|---|
| gux2_rnai_sFw | cgacgcgtcGTGGCAGCTGTCCGACTAC |
| gux2_rnai_sRv | ggggtacccGTCTCCCCCATCTGGTCGTA |
| gux2_rnai_asFw | cgggatccGTCTCCCCCATCTGGTCGTA |
| gux2_rnai_asRv | ggccagatatcGTGGCAGCTGTCCGACTAC |

[55] Os fragmentos correspondentes ao *ScGUX2 sense* (sGUX2) e ScGUX2 *antisense* (asGUX2) foram amplificados a partir de DNA genômico de cana-de-açúcar, purificados a partir de eletroforese em gel (Wizard® SV Gel e PCR Clean-Up System) e clonados no vetor pGEMT-easy (Promega, EUA). Após uma série de digestões com enzimas de restrição e ligações com DNA-ligase T4, estes fragmentos foram combinados para formar um grampo como mostrado na Figura 5.

[56] Então, o fragmento contendo a construção de grampo ScGUX2 (asGUX2 + intron + sGUX2, Figura 5) foi isolado com as enzimas de restrição BamHI e KpnI e ligado a um vetor modificado pCR8, dando origem à construção mostrada na Figura 6.

[57] O grampo mostrado na Figura 6 foi então clonado no vector pGVG (Guidelli et al.,

2018) seguido por uma reação de recombinação utilizando a mistura Enzima Gateway™ LR Clonase™. A construção final, pGVG_GUX2_RNAi, utilizada para silenciamento em cana-de-açúcar é mostrada na Figura 7. A descrição do vetor pGVG é detalhada em Guidelli et al (2018).

[58] A sequência utilizando as enzimas de restrição foi confirmada e o sequenciamento de DNA confirmou que nenhuma mutação foi observada no cassete pGVG_GUX2_RNAi.

**(e) Transformação da variedade de cana-de-açúcar**

[59] Para produzir linhagens transgênicas, plantas de cana-de-açúcar (SP80-3280) foram cultivadas em estufa por seis meses e a região meristemática do ápice da parte aérea foi usada para gerar explantes. Este material foi cultivado em meio de manutenção do MS de sais MS (Murashige e Skoog, 1962), a 26 °C no escuro, até a geração de calos embriogênicos. O pGVG_GUX2_RNAi foi transferido para *Agrobacterium tumefaciens* (cepa EHA105) por choque térmico. As culturas bacterianas foram incubadas com calos de cana-de-açúcar sob vácuo por cinco minutos e transferidas para meio de co-cultivo a 22 °C, no escuro por 3 dias. Depois disso, os calos foram mantidos em meio de repouso a 26 °C, no escuro por 6 dias. Após a fase de repouso, os calos transformados foram transferidos para um meio de regeneração seletivo a 26 ° C, durante 14 dias com 16 h de fotoperíodo. Os eventos transgênicos foram mantidos em meio sem hormônios para induzir crescimento e enraizamento. As plantas transformadas com o vector vazio pGVG e plantas do tipo selvagem foram utilizadas como controles negativos. Em seguida, os eventos foram transplantados para um vaso com substrato (Tropstrato), onde foram aclimatizados durante 15 dias (dentro de uma câmara com alta concentração de umidade). Então os vasos foram transferidos para a estufa. Após 30 dias, as plantas foram transplantadas para vasos maiores (25 L), e a cada dois meses foram adicionados NPK granulados aos vasos, onde cresceram até os 7 meses de idade e foram colhidos.

**(f) Análises de linhagens independentes de cana-de-açúcar transgênica**

[60] Plantas transgênicas putativas foram analisadas para avaliar o silenciamento do gene *ScGUX2*. Caules de cana-de-açúcar foram coletados de plantas com sete meses de idade e congelados em nitrogênio líquido. Em seguida, os caules foram homogeneizados com almofariz e pilão com o uso de nitrogênio líquido. Aproximadamente 150-200 mg de tecido homogeneizado de cana-de-açúcar foram usados para extrair RNA usando TRIZOL de acordo com o protocolo do fabricante (Invitrogen, EUA). O RNA foi tratado com DNAse e o cDNA foi sintetizado usando Biorad iSCRIPT. O qPCR foi realizado usando SYBR seguindo as diretrizes do fabricante (Biorad, EUA) no aparelho Applied Biosystems 7500 (Thermofisher, EUA). A ubiquitina foi utilizada como controle interno para normalização do gene

(Böttcher et al, 2013). Inicialmente, plantas do tipo selvagem foram analisadas para identificar a expressão tecidual preferencial de *ScGUX2* usando iniciadores gene específico. Então, plantas transgênicas putativas para *ScGUX2*, plantas vetor vazio (controle de transformação) e WT (selvagem, mas passaram pela cultura de tecidos) foram usadas para determinar a expressão de *ScGUX2* por qRT-PCR para confirmar o silenciamento dos eventos independentes. Em vários deles, os transcritos tiveram baixa ou nenhuma detecção (Figura 8). As análises Δct mostram a expressão relativa de cada evento, composto por um conjunto de três plantas. Foram utilizados os seguintes materiais: material do tipo selvagem cultivado *in vitro* sem contato com *Agrobacterium tumefaciens* (C+); plantas transgênicas transformadas com o vetor pGVG, sem a construção de grampo ScGUX2 (VV) e os eventos transgênicos independentes silenciados para o gene *ScGUX2* (transformados com pGVG_GUX2_RNAi), identificados como I.10, I.12, I.15, I.2, I.3, I.4, I.5 e I.9 (Figura 8). Todos os eventos pGVG_GUX2_RNAi tiveram menor expressão do gene *ScGUX2* em comparação com as plantas C+ e VV. Estes dados foram utilizados para plantas selecionadas para experimentos adicionais, como descrito abaixo.

**(g) <u>Avaliação da composição da parede celular e sacarificação</u>**

[61] Todas as plantas foram podadas e rebrotadas por 9 meses para reduzir as heterogeneidades no estágio de desenvolvimento para evitar a variação nas análises da parede celular, conforme ilustrado na Figura 9.

[62] Os seguintes dados foram medidos em plantas com nove meses de idade: número de caules, altura, diâmetro, graus Brix (topo e base) e peso fresco. Após as medidas, os caules foram congelados rapidamente em nitrogênio líquido. Três partes de cada planta foram coletadas com base em sua fase de desenvolvimento: caules jovens - tecidos menos lignificados (entrenós 2 e 3), caule intermediário (entrenós 4 e 5) e caule maduro – altamente lignificado (entrenós 10 e 11), como mostrado Figura 10.

[63] Em seguida, os tecidos do caule foram homogeneizados com almofariz e pilão com o uso de nitrogênio líquido e armazenados em freezer a -72ºC. A composição da parede celular - celulose, hemicelulose e pectina - foi analisada por amostras de moagem e lavagem sequencial com $H_2O$ destilada a frio, acetona e metanol/clorofórmio (1:1, v/v). Subsequentemente, o amido foi removido usando 2 unidades.$mL^{-1}$ de amilase pancreática por 3 horas a 37ºC em acetato de sódio 0,1 M (pH 6.5). Em seguida, as amostras foram centrifugadas e o material sólido foi recuperado. O *pellet* foi extraído com oxalato de amônio 20 mM (pH 4.0) para obter a fração de pectina. Após centrifugação, o *pellet* foi

extraído sequencialmente sob vácuo à temperatura ambiente no escuro com NaOH 0,1 M (fração de sobrenadante 1) e NaOH a 17,5% (fração de sobrenadante 2). A fração hemicelulósica total (1 e 2) foi obtida agrupando ambas as frações sobrenadantes. O *pellet* resultante foi tratado com 72% de $H_2SO_4$ para obter a fração de celulose. O conteúdo total de açúcar em cada fração foi determinado com o reagente fenol-sulfúrico, usando glicose como padrão. Os resultados são compilados nas Figuras 11 (celulose), 12 (pectina) e 13 (hemicelulose).

[64] Estudos anteriores mostraram que os níveis de pectina de celulose, hemicelulose em "nocautes" de *AtGUX2* de Arabidopsis não diferiram do tipo selvagem (Mortimer et al., 2010). Os dados dos presentes ensaios em cana-de-açúcar (Figuras 11, 12 e 13) mostraram que a maioria dos eventos silenciados para o gene *ScGUX2* não diferiu das plantas de tipo selvagem e vetora vazia.

[65] Para avaliar os efeitos do silenciamento do ScGUX2 nos níveis de sacarificação, os seguintes procedimentos foram empregados com as mesmas amostras utilizadas nos experimentos anteriores. O ensaio de sacarificação foi realizado lavando as amostras antes dos ensaios enzimáticos para evitar a inibição devido ao excesso de açúcar presente nas amostras. 30 mg da amostra foram lavadas com 1 mL de água destilada fria, agitada em vórtice, centrifugadas a 14.000 rpm durante 10 min, sendo o sobrenadante eliminado. Essa etapa de lavagem foi repetida mais uma vez e as amostras foram agitadas em vórtice com 1 mL de acetona PA (Sigma-Aldrich, EUA). Após centrifugação a 14.000 rpm por 10 min, o sobrenadante foi descartado. O material residual foi lavado em vórtex com 1 mL de metanol/clorofórmio 1:1 (v/v) (Sigma-Aldrich, EUA) e depois centrifugado a 14.000 rpm durante 10 min. O sobrenadante foi descartado e as amostras foram secas em temperatura ambiente (RT) por 24h.

[66] Uma quantidade equivalente a 5 mg de celulose foi misturada com uma mistura de celobiose (Sigma-Aldrich, EUA, catálogo # C1184) e celulase de *Trichoderma reesei* (Sigma-Aldrich, EUA, catálogo # C2730) durante 5 dias a 50 ° C em um agitador de incubadora Innova 4430 (Marshall Scientific, EUA) a 200 rpm. As amostras foram então centrifugadas a 14.000 rpm durante 10 min e o sobrenadante foi recolhido. A amostra foi diluída 150x em água e 100 µL foram misturados com 400 µL de $H_2SO_4$ PA (Sigma-Aldrich, EUA) e 100 µL de fenol a 5% (Sigma-Aldrich, EUA). As amostras foram deixadas a arrefecer à temperatura ambiente durante 15 minutos e as leituras foram realizadas num espectrofotómetro SpectraMax M3 (Molecular Devices, EUA) a 490 nm. Uma curva padrão para glicose foi usada para estimar a porcentagem de sacarificação.

[67] Oito eventos independentes foram avaliados e quatro deles (I.15, I.10, I.4 e I.5)

aumentaram a sacarificação em comparação com plantas transformadas com o vetor vazio e plantas do tipo selvagem (Figura 14). Este aumento está de acordo com os dados das plantas de Arabidopsis com mutação no gene *Atgux2* (Mortimer et al., 2010). Em comparação com plantas do tipo selvagem, o mutante *Atgux2* de Arabidopsis apresenta uma redução de 30% no padrão de decoração xilose e um aumento na sacarificação, sem alterar significativamente as características fisiológicas da planta. Esses resultados mostraram claramente que o silenciamento do gene *ScGUX2* na cana-de-açúcar aumenta a sacarificação sem comprometer os componentes da parede celular (conteúdo de pectina, hemicelulose e celulose) e o desenvolvimento regular da cana-de-açúcar.

[68] Os dados agronômicos coletados foram usados para identificar qualquer diferença durante o desenvolvimento da planta (Figura 15 - 19). Não houve diferenças estatísticas entre os eventos transgênicos e as plantas C + ou VV para esses parâmetros. É importante ressaltar que os eventos apresentaram valores superiores a 18 ºBrix (entrenós 10 e 11), indicando plantas maduras.

[69] Embora a invenção tenha sido amplamente descrita, é óbvio para aqueles versados na técnica que várias alterações e modificações podem ser feitas visando aprimoramento do projeto sem que as referidas alterações não estejam cobertas pelo escopo da invenção.

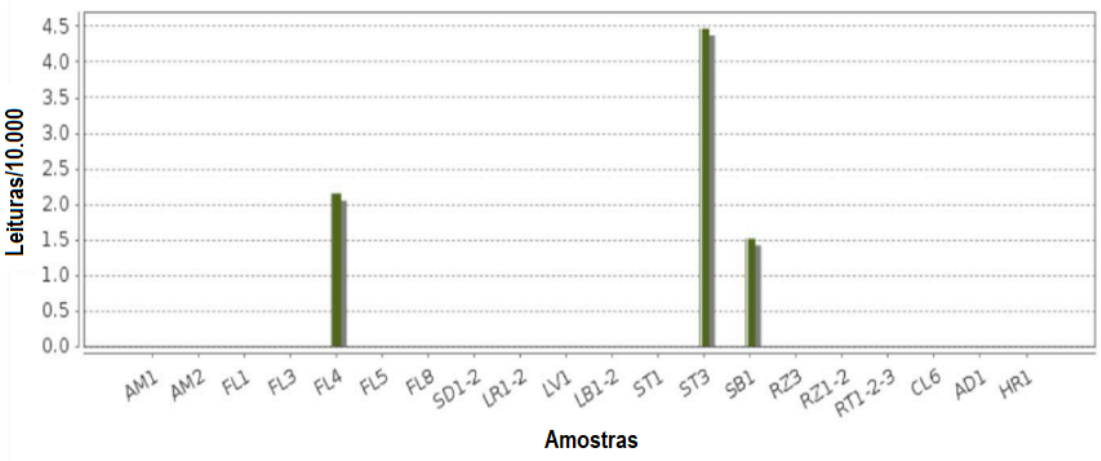

Leituras/10.000
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
AM1
AM2
FL1
FL3
FL4
FL5
FL8
SD1-2
LR1-2
LV1
LB1-2
ST1
ST3
SB1
RZ3
RZ1-2
RT1-2-3
CL6
AD1
HR1
Amostras

Figura 1

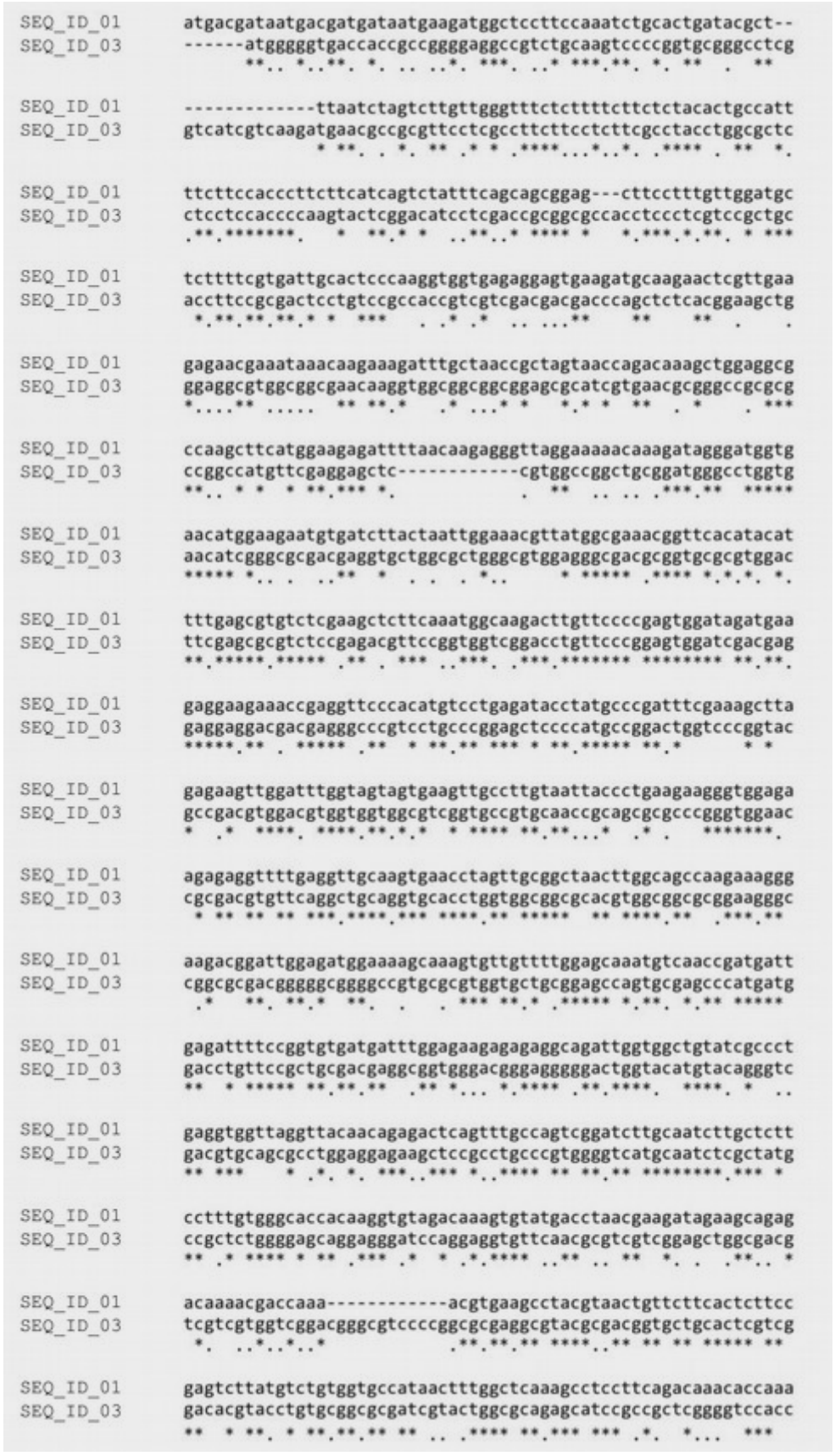

SEQ_ID_01 atgacgataatgacgatgataatgaagatggctccttccaaatctgcactgatacgct--
SEQ_ID_03 ------atgggggtgaccaccgccggggaggccgtctgcaagtccccggtgcgggcctcg
SEQ_ID_01 ------------ttaatctagtcttgttgggtttctcttttcttctctacactgccatt
SEQ_ID_03 gtcatcgtcaagatgaacgccgcgttcctcgccttcttcctcttcgcctacctggcgctc
SEQ_ID_01 ttcttccacccttcttcatcagtctatttcagcagcggag---cttcctttgttggatgc
SEQ_ID_03 ctcctccaccccaagtactcggacatcctcgaccgcggcgccacctccctcgtccgctgc
SEQ_ID_01 tcttttcgtgattgcactcccaaggtggtgagaggagtgaagatgcaagaactcgttgaa
SEQ_ID_03 accttccgcgactcctgtccgccaccgtcgtcgacgacgacccagctctcacggaagctg
SEQ_ID_01 gagaacgaaataaacaagaaagatttgctaaccgctagtaaccagacaaagctggaggcg
SEQ_ID_03 ggaggcgtggcggcgaacaaggtggcggcggcggagcgcatcgtgaacgcgggccgcgcg
SEQ_ID_01 ccaagcttcatggaagagattttaacaagagggttaggaaaaacaaagatagggatggtg
SEQ_ID_03 ccggccatgttcgaggagctc------------cgtggccggctgcggatgggcctggtg
SEQ_ID_01 aacatggaagaatgtgatcttactaattggaaacgttatggcgaaacggttcacatacat
SEQ_ID_03 aacatcgggcgcgacgaggtgctggcgctgggcgtggaggggcgacgcggtgcgcgtggac
SEQ_ID_01 tttgagcgtgtctcgaagctcttcaaatggcaagacttgttccccgagtggatagatgaa
SEQ_ID_03 ttcgagcgcgtctccgagacgttccggtggtcggacctgttcccggagtggatcgacgag
SEQ_ID_01 gaggaagaaaccgaggttcccacatgtcctgagatacctatgcccgatttcgaaagctta
SEQ_ID_03 gaggaggacgacgagggcccgtcctgcccggagctccccatgccggactggtcccggtac
SEQ_ID_01 gagaagttggatttggtagtagtgaagttgccttgtaattaccctgaagaagggtggaga
SEQ_ID_03 gccgacgtggacgtggtggtggcgtcggtgccgtgcaaccgcagcgcgcccgggtggaac
SEQ_ID_01 agagaggttttgaggttgcaagtgaacctagttgcggctaacttggcagccaagaaaggg
SEQ_ID_03 cgcgacgtgttcaggctgcaggtgcacctggtggcggcgcacgtggcggcgcggaagggc
SEQ_ID_01 aagacggattggagatggaaaagcaaagtgttgttttggagcaaatgtcaaccgatgatt
SEQ_ID_03 cggcgcgacgggggcggggccgtgcgcgtggtgctgcggagccagtgcgagcccatgatg
SEQ_ID_01 gagattttccggtgtgatgatttggagaagagagaggcagattggtggctgtatcgccct
SEQ_ID_03 gacctgttccgctgcgacgaggcggtgggacgggaggggactggtacatgtacagggtc
SEQ_ID_01 gaggtggttaggttacaacagagactcagtttgccagtcggatcttgcaatcttgctctt
SEQ_ID_03 gacgtgcagcgcctggaggagaagctccgcctgcccgtggggtcatgcaatctcgctatg
SEQ_ID_01 cctttgtgggcaccacaaggtgtagacaaagtgtatgacctaacgaagatagaagcagag
SEQ_ID_03 ccgctctggggagcaggagggatccaggaggtgttcaacgcgtcgtcggagctggcgacg
SEQ_ID_01 acaaaacgaccaaa------------acgtgaagcctacgtaactgttcttcactcttcc
SEQ_ID_03 tcgtcgtggtcggacgggcgtccccggcgcgaggcgtacgcgacggtgctgcactcgtcg
SEQ_ID_01 gagtcttatgtctgtggtgccataactttggctcaaagcctccttcagacaaacaccaaa
SEQ_ID_03 gacacgtacctgtgcggcgcgatcgtactggcgcagagcatccgccgctcggggtccacc

Figura 2


FIG. 2 (Continuação)

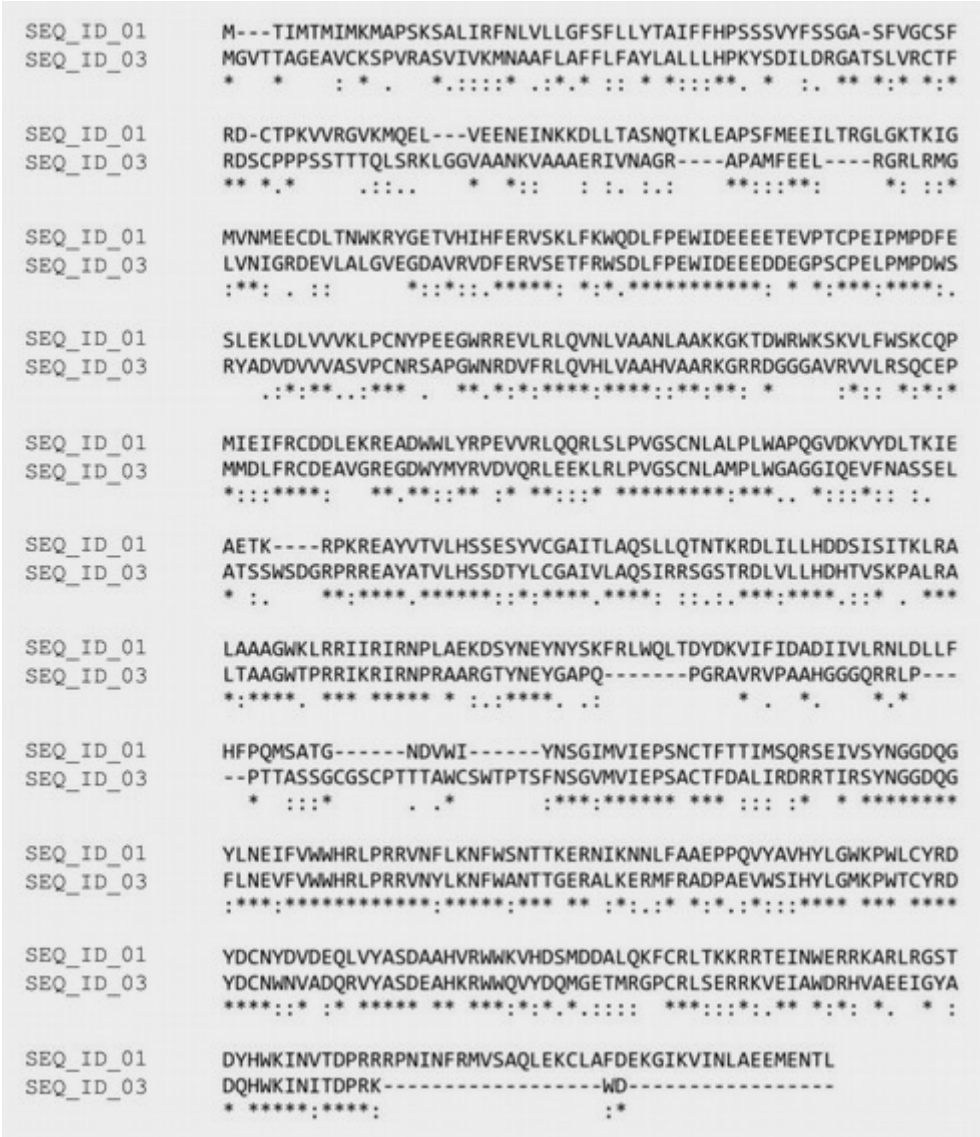



Figura 3

>SEQ. ID. No. 1

```
atgacgataa tgacgatgat aatgaagatg gctccttcca aatctgcact gatacgcttt        60
aatctagtct tgttgggttt ctcttttctt ctctacactg ccattttctt ccacccttct       120
tcatcagtct atttcagcag cggagcttcc tttgttggat gctcttttcg tgattgcact       180
cccaaggtgg tgagaggagt gaagatgcaa gaactcgttg aagagaacga aataaacaag       240
aaagatttgc taaccgctag taaccagaca aagctggagg cgccaagctt catggaagag       300
attttaacaa gagggttagg aaaaacaaag atagggatgg tgaacatgga agaatgtgat       360
cttactaatt ggaaacgtta tggcgaaacg gttcacatac attttgagcg tgtctcgaag       420
ctcttcaaat ggcaagactt gttccccgag tggatagatg aagaggaaga aaccgaggtt       480
cccacatgtc ctgagatacc tatgcccgat ttcgaaagct tagagaagtt ggatttggta       540
gtagtgaagt tgccttgtaa ttaccctgaa gaagggtgga gaagagaggt tttgaggttg       600
caagtgaacc tagttgcggc taacttggca gccaagaaag ggaagacgga ttggagatgg       660
```

```
aaaagcaaag tgttgttttg gagcaaatgt caaccgatga ttgagatttt ccggtgtgat       720
gatttggaga agagagaggc agattggtgg ctgtatcgcc ctgaggtggt taggttacaa       780
cagagactca gtttgccagt cggatcttgc aatcttgctc ttcctttgtg ggcaccacaa       840
ggtgtagaca aagtgtatga cctaacgaag atagaagcag agacaaaacg accaaaacgt       900
gaagcctacg taactgttct tcactcttcc gagtcttatg tctgtggtgc cataactttg       960
gctcaaagcc tccttcagac aaacaccaaa cgcgatctta tccttctcca cgatgactcc      1020
atctccatta ccaaacttcg agctctcgcc gccgcaggat ggaagcttcg tcggatcatt      1080
cgaatcagaa acccacttgc ggagaaggac tcgtacaatg aatacaacta cagcaagttt      1140
cgactctggc aattgacaga ttacgacaaa gtgatcttca ttgatgccga catcatcgtc      1200
ttacgtaacc ttgatcttct cttccatttt cctcagatgt cggccaccgg aaatgatgta      1260
tggatatata attcaggcat catggtcatc gagccttcta attgtacgtt tactacaatc      1320
atgagccagc gaagcgagat cgtttcatac aacggtggag atcaagggta cctaaacgag      1380
atatttgtgt ggtggcaccg attgcctcga cgagtaaact ttctgaagaa cttctggtcg      1440
aacacaacca aagaaagaaa catcaagaac aacctcttcg ccgcggagcc gcctcaggtc      1500
tacgcggtcc actacttagg ttggaaacca tggctttgct atagggacta cgattgcaac      1560
tacgacgtgg acgagcagtt ggtgtacgct agtgatgcgg ctcacgttag gtggtggaaa      1620
gtgcacgact ccatggacga tgcattgcaa aagttttgca ggctgacgaa aaagaggaga      1680
acggagatca actgggagag gaggaaagca aggcttagag gttccactga ttatcattgg      1740

aagatcaatg tcactgatcc aagacgacgt cgttcttatt tgattggtta a               1791
```

>SEQ. ID. No. 2

```
atgggggtga ccaccgccgg tgaagcggcc tgcaagtccc cggcggtgcg ggcctcggtc         60
atcgtcaggc tcaacgccgc gttcctcgcc ttcttcctct tcgcctacct ggcgctcctc        120
ctccacccca agtactcgga catcctcgac cgcggcgcca cctccctcgt ccgctgcacc        180
ttccgcgact cctgcccacc gccggcgccg tcgtcgtcgt cgtcgaccca gctctcacgg        240
aagctgggag gcgtggcggc gaacaaggtg gcggcggcgg cggagcgcat cgtgatcgtg        300
aacgcggggc gcgcgccggc catgttcgac gacctccgtg gccgtcttcg gatgggcctg        360
gtgaacatcg ggcgcgacga ggtggtggcg ctgggcgtgg agggcgacgc ggtgcgcgtg        420
gacttcgacc gcgtctccga gacgttccgg tggtcggacc tgttcccgga gtggatcgac        480
gaggaggagg acgacgaggg cccgtcctgc ccggagctcc ccatgccgga ctggtcccgg        540
tacggcgacg tggacgtggt ggtggcgtcg ctgccgtgca accgcagcgc gaccgggtgg        600
aaccgcgacg tgttcaggct gcaggtgcac ctggtggcgg cgcaggtggc ggcacggaag        660
ggccggcgca acggggccgg cgccggggcc gtgcgcgtgg tgctgcggag ccagtgcgag        720
cccatgatgg acctgttccg ctgcgacgag gcggtgggac gggaggggga ctggtacatg        780
tacagagtcg acgtgcagcg cctggaggag aagctccgcc tgcccgtggg ctcctgcaac        840
ctcgccatgc cgctctgggg agcaggaggg atccaggagg tgttcaacgc gtcgtcggag        900
ctggcgacgt cgccgtcgtg gtcgggcggg cgtccccggc gtgaggcgta cgcgacggtg        960
ctgcactcgt cggacacgta cctgtgcggc gcgatcgtgc tggcgcagag catccggcgc       1020
tcggggtcca cccgggacct gatcctcctc cacgaccaca cggtgtcgaa gccggcgctc       1080
cgggcgctga cggcggccgg gtggacgcct cgccggatca agcggatccg caacccgcgc       1140
gcggcgcggg gcacctacaa cgagtacaac tacagcaagt tccggctgtg gcagctgacc       1200
gactacgacc gcgtggtgtt cgtggacgcc gacatcctgg tgctccgcag cctggacgcg       1260
ctgttcgcgt tcccgcagct cacggcggtg ggcaacgacg ggtccctctt caactccggt       1320
gtgatggtga tcgagccgtc ggcgtgcacg ttcgacgcgc tgatccggga ccgccggacc       1380
atccggtcgt acaacggcgg cgaccagggg ttcctgaacg aggtgttcgt gtggtggcac       1440
cggctgccgc ggcgggtgaa ctacctcaag aacttctggg caaacaccac cggggagcgc       1500
gcgctcaagg agcggatgtt ccgggcggac ccggccgagg tgtggtccat ccactacctg       1560
gggatgaagc catggacgtg ctacagggac tacgactgca actggaacgt ggcggaccag       1620
cgggtgtacg ccagcgacga ggcgcacagg cgctggtggc aggtgtacga ccagatgggg       1680
gagaccatgc gcgggccatg ccgcctctcc gagcggagga aggtggagat cgcatgggac       1740
aggcacgtcg ccgaggaggt cgggtacgcc gaccagcact ggaagatcaa catcacggac       1800
ccaaggaaat gggactga                                                     1818
```

>SEQ. ID. No. 3

```
atgggggtga ccaccgccgg ggaggccgtc tgcaagtccc cggtgcgggc ctcggtcatc         60
gtcaagatga acgccgcgtt cctcgccttc ttcctcttcg cctacctggc gctcctcctc        120
caccccaagt actcggacat cctcgaccgc ggcgccacct ccctcgtccg ctgcaccttc        180
cgcgactcct gtccgccacc gtcgtcgacg acgacccagc tctcacggaa gctgggaggc        240
gtggcggcga acaaggtggc ggcggcggag cgcatcgtga acgcgggccg cgcgccggcc        300
atgttcgagg agctccgtgg ccggctgcgg atgggcctgg tgaacatcgg gcgcgacgag        360
gtgctggcgc tgggcgtgga gggcgacgcg gtgcgcgtgg acttcgagcg cgtctccgag        420
acgttccggt ggtcggacct gttcccggag tggatcgacg aggaggagga cgacgagggc        480
ccgtcctgcc cggagctccc catgccggac tggtcccggt acgccgacgt ggacgtggtg        540
gtggcgtcgg tgccgtgcaa ccgcagcgcg cccgggtgga accgcgacgt gttcaggctg        600
caggtgcacc tggtggcggc gcacgtggcg gcgcggaagg gccggcgcga cgggggcggg        660
gccgtgcgcg tggtgctgcg gagccagtgc gagcccatga tggacctgtt ccgctgcgac        720
gaggcggtgg gacgggaggg ggactggtac atgtacaggg tcgacgtgca gcgcctggag        780
gagaagctcc gcctgcccgt ggggtcatgc aatctcgcta tgccgctctg gggagcagga        840
gggatccagg aggtgttcaa cgcgtcgtcg gagctggcga cgtcgtcgtg gtcggacggg        900
cgtccccggc gcgaggcgta cgcgacggtg ctgcactcgt cggacacgta cctgtgcggc        960
```

```
gcgatcgtac tggcgcagag catccgccgc tcggggtcca cccgcgacct ggtcctcctc      1020
cacgaccaca cggtgtcgaa gccggcgctc cgcgcgctga cggcggccgg gtggacgccg      1080
cgccggatca agcgcatccg caaccctcgc gcggcgcggg gcacctacaa cgagtatggt      1140
gctccgcaac ctggacgcgc tgttcgcgtt cccgcagctc acggcggtgg gcaacgacgg      1200
ctccctccaa ctacagcaag ttccggctgt ggcagctgtc cgactacgac cgcgtggtgt      1260
tcgtggacgc cgacatcctt caactccggg gtgatggtga tcgagccgtc ggcatgcacg      1320
ttcgacgcgc tgatccggga ccgccggacc atccggtcgt acaacggcgg cgaccagggg      1380
ttcctgaacg aggtgttcgt gtggtggcac cggctgccgc ggcgggtgaa ctacctcaag      1440
aacttctggg ccaacaccac cggggagcgc gcgctcaagg agcggatgtt ccgggcggac      1500
ccggccgagg tgtggtccat ccactacctg gggatgaagc cctggacgtg ctacagggac      1560
tacgactgca actggaacgt ggcggaccag cgggtgtacg ccagcgacga ggcgcacaag      1620
cgctggtggc aggtgtacga ccagatgggg gagaccatgc gcgggccatg ccgcctctcg      1680
gagcggagga aggtggagat cgcctgggac aggcacgtcg ccgaggagat cgggtacgcg      1740
gaccagcact ggaagatcaa catcacggac ccaaggaaat gggactga                  1788
```

>SEQ. ID. No. 4

```
gtggcagctg tccgactacg accgcgtggt gttcgtggac gccgacatcc tggtgctccg        60
caacctggac gcgctgttcg cgttcccgca gctcacggcg gtgggcaacg acggctccct       120
cttcaactcc ggggtgatgg tgatcgagcc gtcggcatgc acgttcgacg cgctgatccg       180
ggaccgccgg accatccggt cgtacaacgg cggcgaccag gggttcctga acgaggtgtt       240
cgtgtggtgg caccggctgc cgcggcgggt gaactacctc aagaacttct gggccaacac       300
caccggggag cgcgcgctca aggagcggat gttccgggcg gacccggccg aggtgtggtc       360
catccactac ctggggatga agccctggac gtgctacagg gactacgact gcaactggaa       420
cgtggcggac cagcgggtgt acgccagcga cgaggcgcac aagcgctggt ggcaggtgta       480
cgaccagatg ggggagac                                                    498
```

>SEQ. ID. No. 5

```
gtctccccca tctggtcgta cacctgccac cagcgcttgt gcgcctcgtc gctggcgtac        60
acccgctggt ccgccacgtt ccagttgcag tcgtagtccc tgtagcacgt ccagggcttc       120
atccccaggt agtggatgga ccacacctcg gccgggtccg cccggaacat ccgctccttg       180
agcgcgcgct ccccggtggt gttggcccag aagttcttga ggtagttcac ccgccgcggc       240
agccggtgcc accacacgaa cacctcgttc aggaacccct ggtcgccgcc gttgtacgac       300
cggatggtcc ggcggtcccg gatcagcgcg tcgaacgtgc atgccgacgg ctcgatcacc       360
atcaccccgg agttgaagag ggagccgtcg ttgcccaccg ccgtgagctg cgggaacgcg       420
aacagcgcgt ccaggttgcg gagcaccagg atgtcggcgt ccacgaacac cacgcggtcg       480
tagtcggaca gctgccacgg ccagatatcg atatcatgcg gtaactgatc tgaattctcc       540
catcagttgg cctttagttc cttggcagtg taccatgtaa ttttacctgt tatctacaca       600
ctctgcaggt tgacgcgtcg acgcgtcgtg gcagctgtcc gactacgacc gcgtggtgtt       660
cgtggacgcc gacatcctgg tgctccgcaa cctggacgcg ctgttcgcgt tcccgcagct       720
cacggcggtg ggcaacgacg gctccctctt caactccggg gtgatggtga tcgagccgtc       780
ggcatgcacg ttcgacgcgc tgatccggga ccgccggacc atccggtcgt acaacggcgg       840
cgaccagggg ttcctgaacg aggtgttcgt gtggtggcac cggctgccgc ggcgggtgaa       900
ctacctcaag aacttctggg ccaacaccac cggggagcgc gcgctcaagg agcggatgtt       960
ccgggcggac ccggccgagg tgtggtccat ccactacctg gggatgaagc cctggacgtg      1020
ctacagggac tacgactgca actggaacgt ggcggaccag cgggtgtacg ccagcgacga      1080
ggcgcacaag cgctggtggc aggtgtacga ccagatgggg gaga                        1124
```

ScGUX2 antisense

ScR1MYB1 - Hairpin

ScGUX2 sense

Figura 3

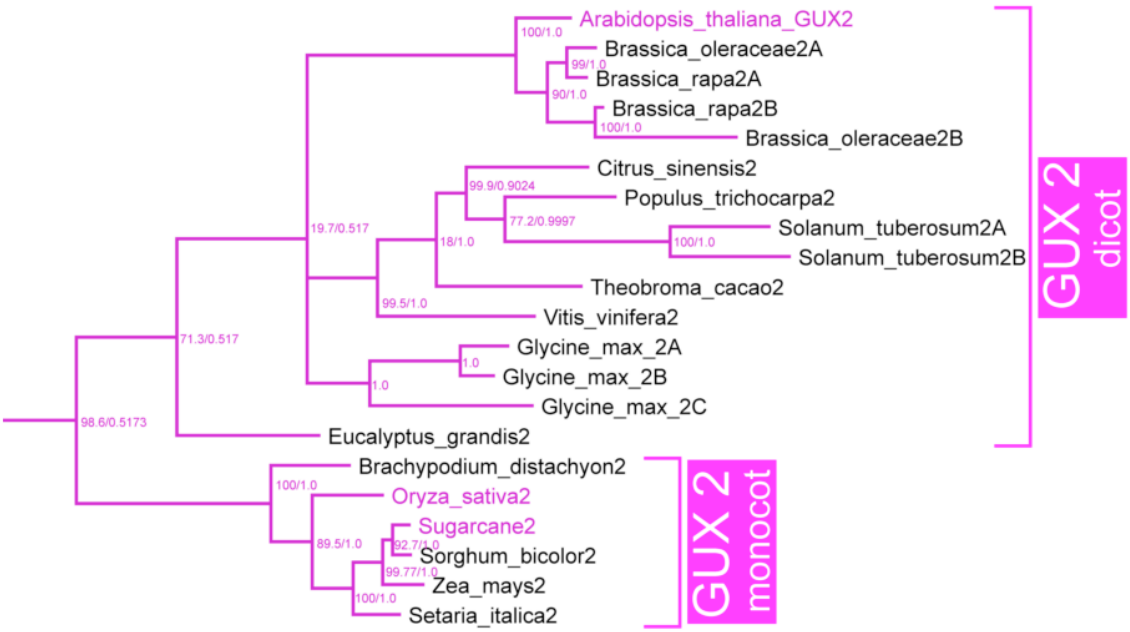

Arabidopsis_thaliana_GUX2
100/1.0
Brassica_oleraceae2A
99/1.0
Brassica_rapa2A
90/1.0
Brassica_rapa2B
100/1.0
Brassica_oleraceae2B
Citrus_sinensis2
99.9/0.9024
Populus_trichocarpa2
77.2/0.9997
Solanum_tuberosum2A
100/1.0
Solanum_tuberosum2B
19.7/0.517
18/1.0
Theobroma_cacao2
99.5/1.0
Vitis_vinifera2
71.3/0.517
Glycine_max_2A
1.0
Glycine_max_2B
1.0
Glycine_max_2C
98.6/0.5173
Eucalyptus_grandis2
Brachypodium_distachyon2
100/1.0
Oryza_sativa2
Sugarcane2
89.5/1.0
92.7/1.0
Sorghum_bicolor2
99.77/1.0
Zea_mays2
100/1.0
Setaria_italica2
GUX 2
dicot
GUX 2
monocot

Figura 4

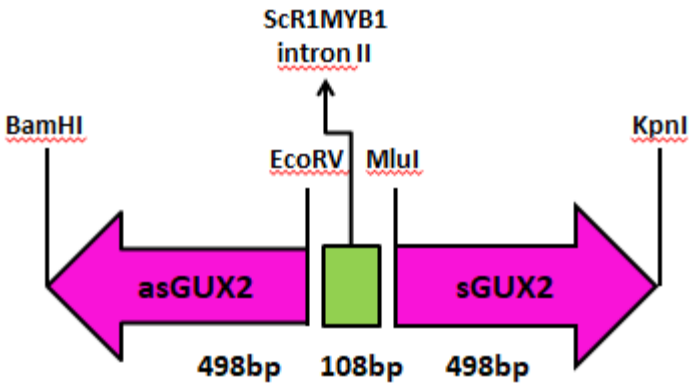

ScR1MYB1
intron II
BamHI
EcoRV
MluI
KpnI
asGUX2
sGUX2
498bp
108bp
498bp

Figura 5

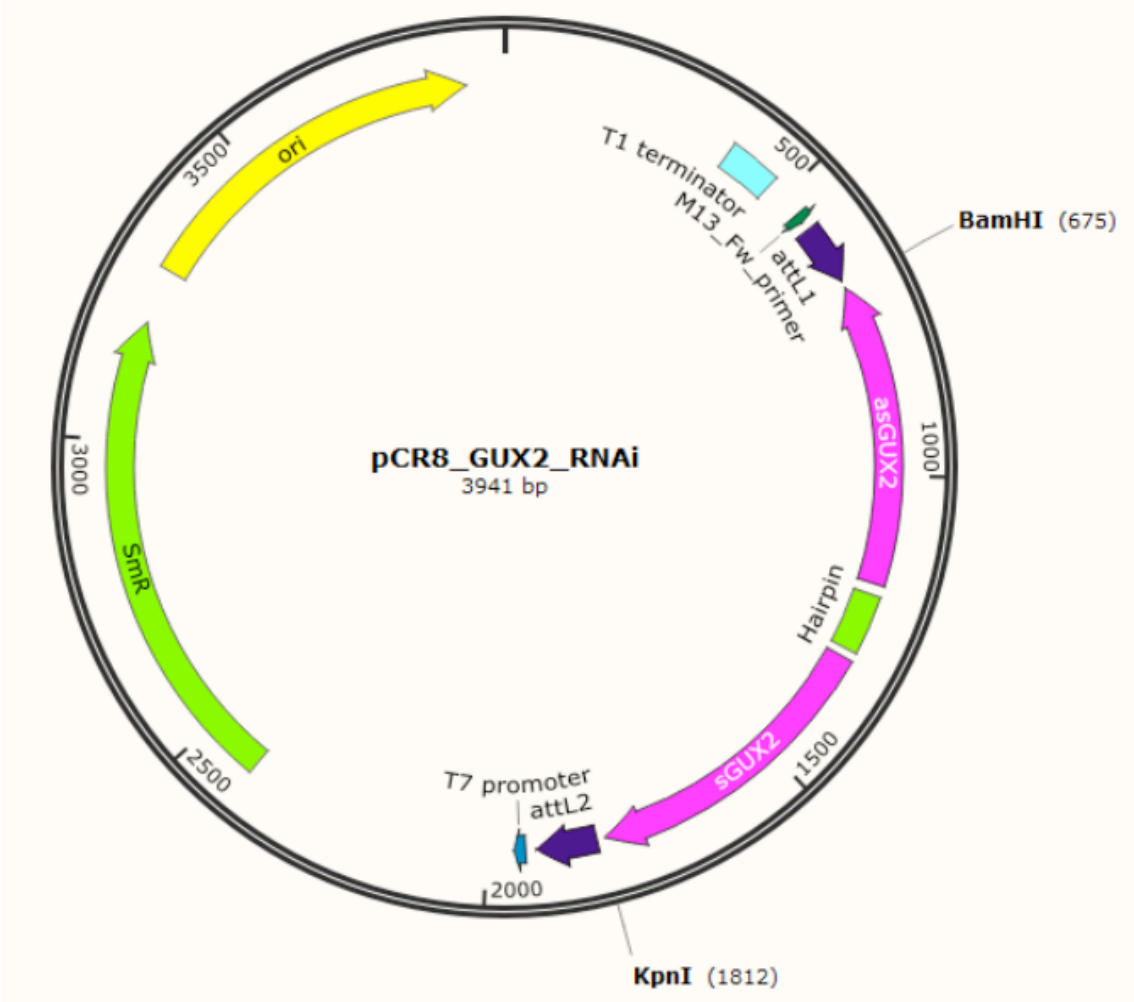

pCR8_GUX2_RNAi
3941 bp
ori
T1 terminator
M13_Fw_primer
attL1
BamHI (675)
asGUX2
Hairpin
sGUX2
T7 promoter
attL2
KpnI (1812)
SmR
500
1000
1500
2000
2500
3000
3500

Figura 6

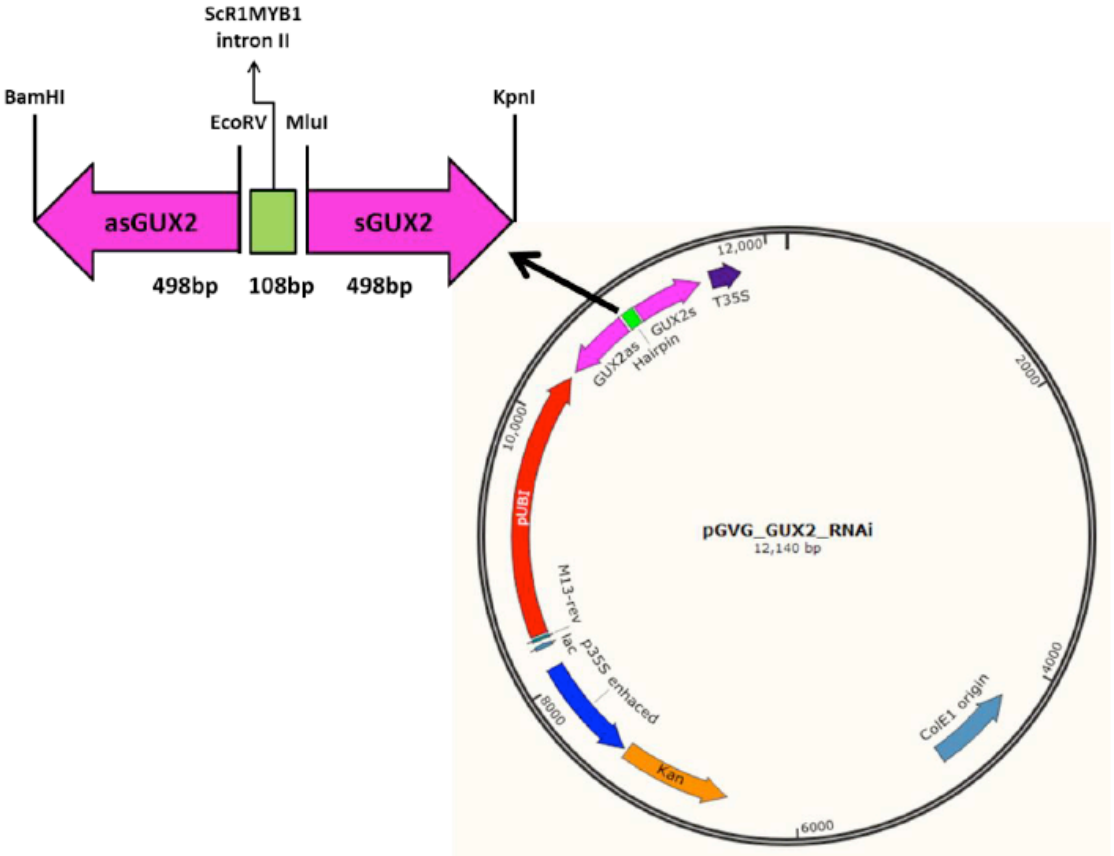

ScR1MYB1
intron II
BamHI
EcoRV
MluI
KpnI
asGUX2
sGUX2
498bp
108bp
498bp
12,000
T35S
GUX2s
GUX2as
Hairpin
2000
10,000
pUBI
pGVG_GUX2_RNAi
12,140 bp
M13-rev
lac
p35S enhaced
4000
8000
ColE1 origin
kan
6000

Figura 7

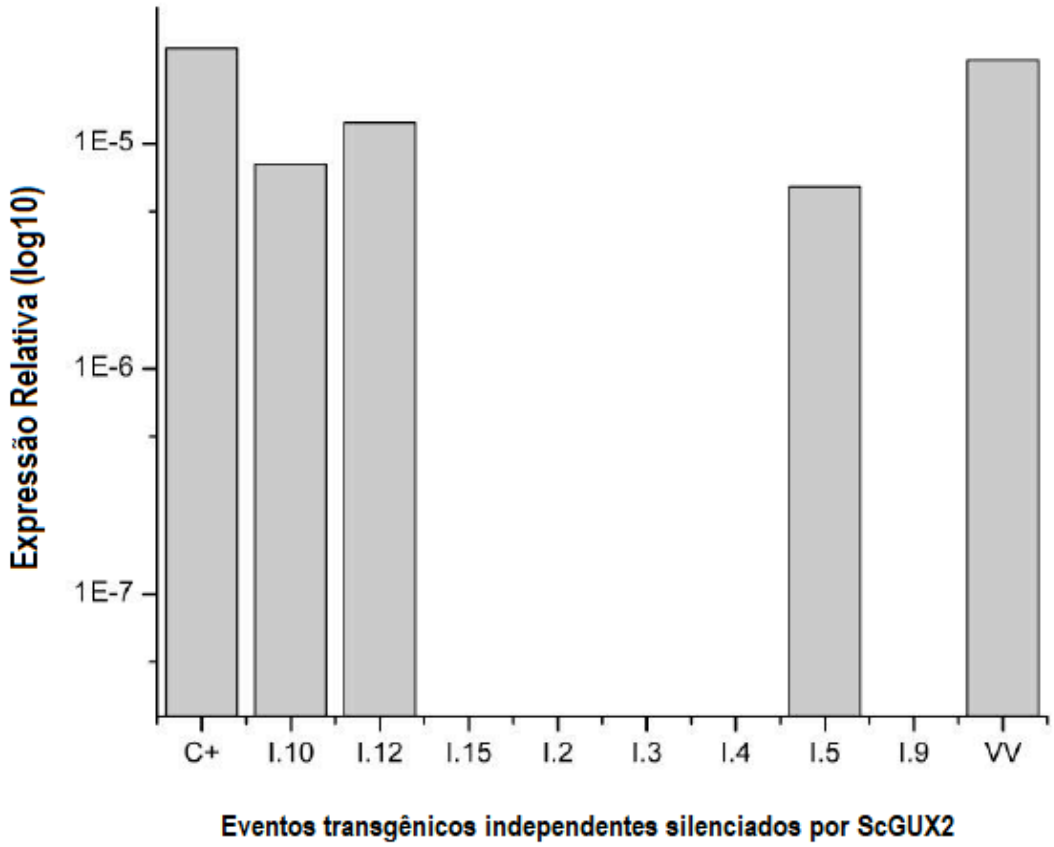

Expressão Relativa (log10)
1E-5
1E-6
1E-7
C+
I.10
I.12
I.15
I.2
I.3
I.4
I.5
I.9
VV
Eventos transgênicos independentes silenciados por ScGUX2

Figura 8

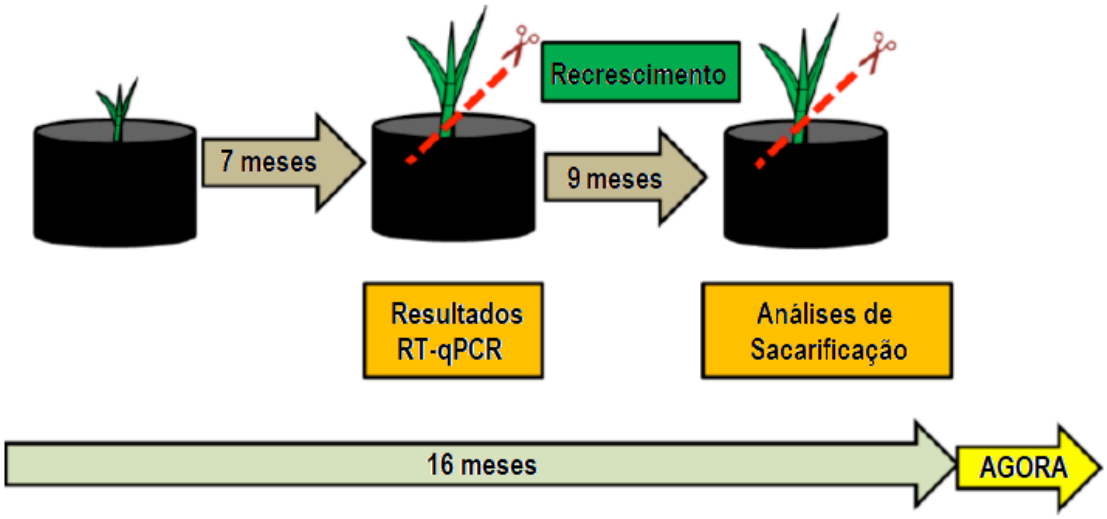

Recrescimento
7 meses
9 meses
Resultados
RT-qPCR
Análises de
Sacarificação
16 meses
AGORA

Figura 9

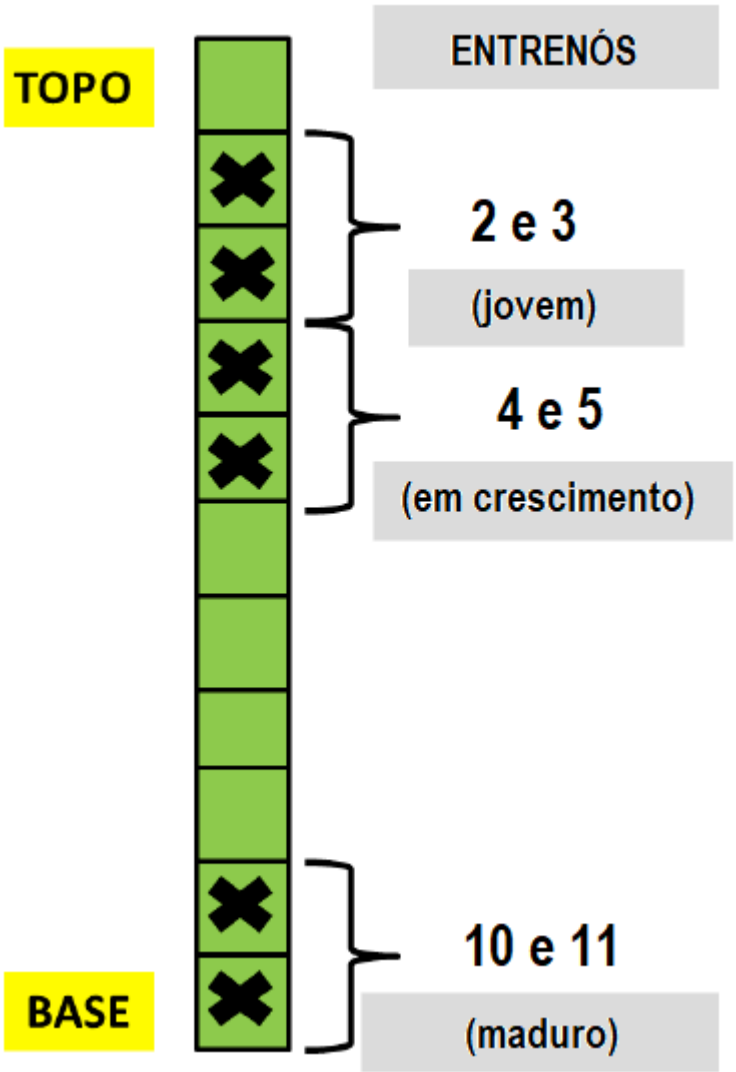

TOPO
ENTRENÓS
2 e 3
(jovem)
4 e 5
(em crescimento)
10 e 11
(maduro)
BASE

Figura 10

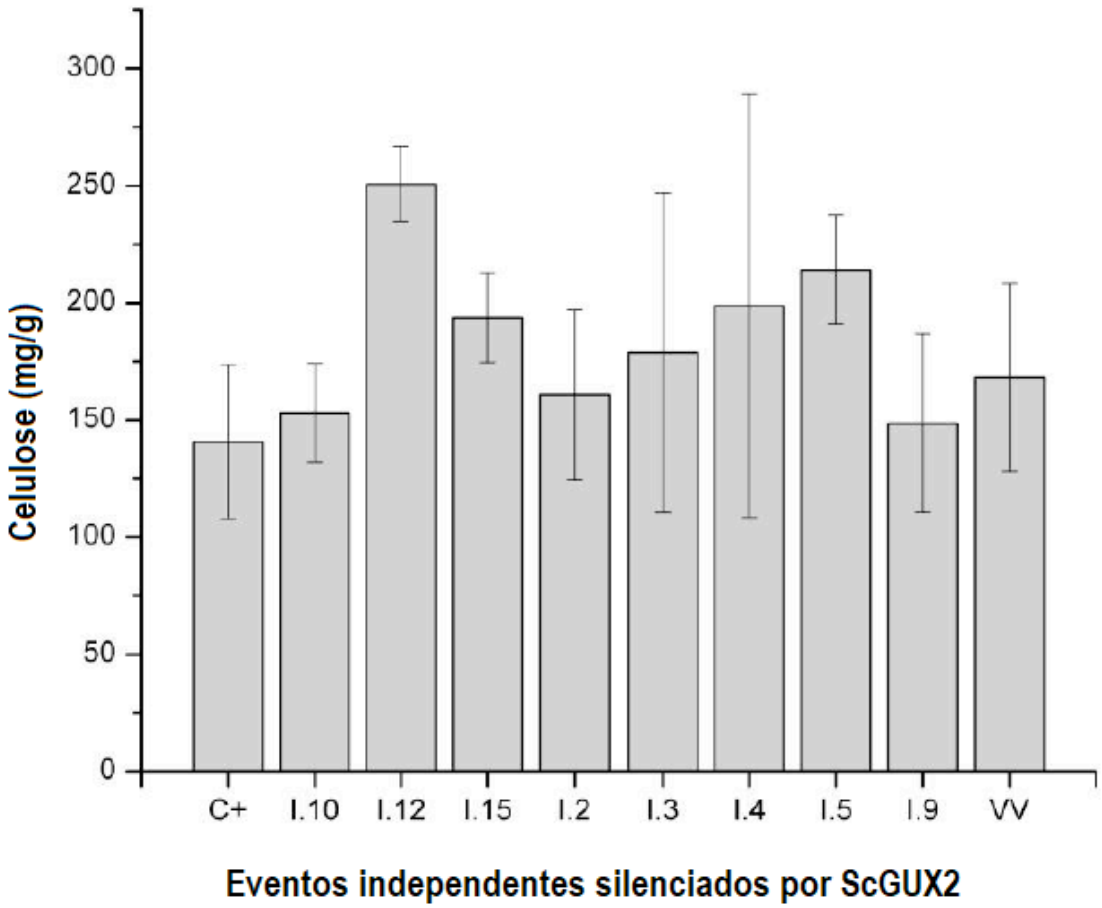

Celulose (mg/g)
300
250
200
150
100
50
0
C+
I.10
I.12
I.15
I.2
I.3
I.4
I.5
I.9
VV
Eventos independentes silenciados por ScGUX2

Figura 11

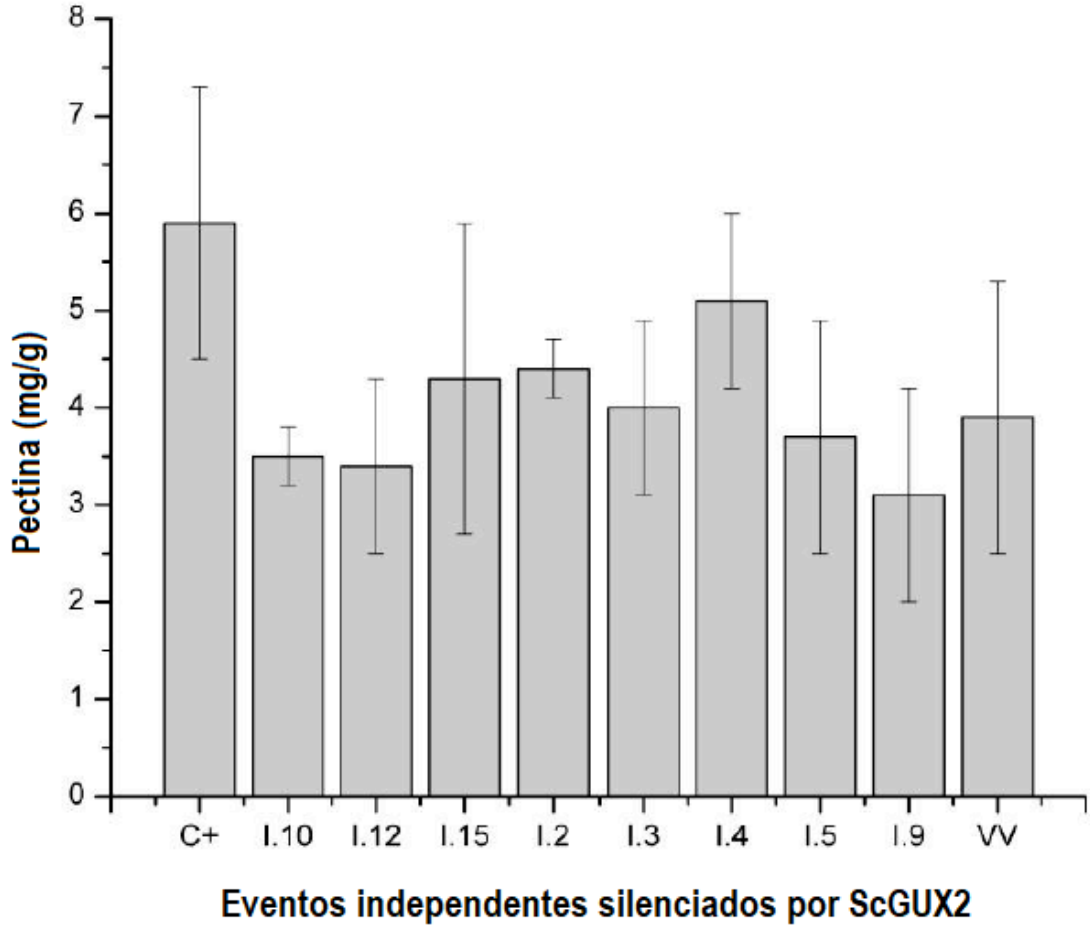

Pectina (mg/g)
0
1
2
3
4
5
6
7
8
C+
I.10
I.12
I.15
I.2
I.3
I.4
I.5
I.9
VV
Eventos independentes silenciados por ScGUX2

Figura 12

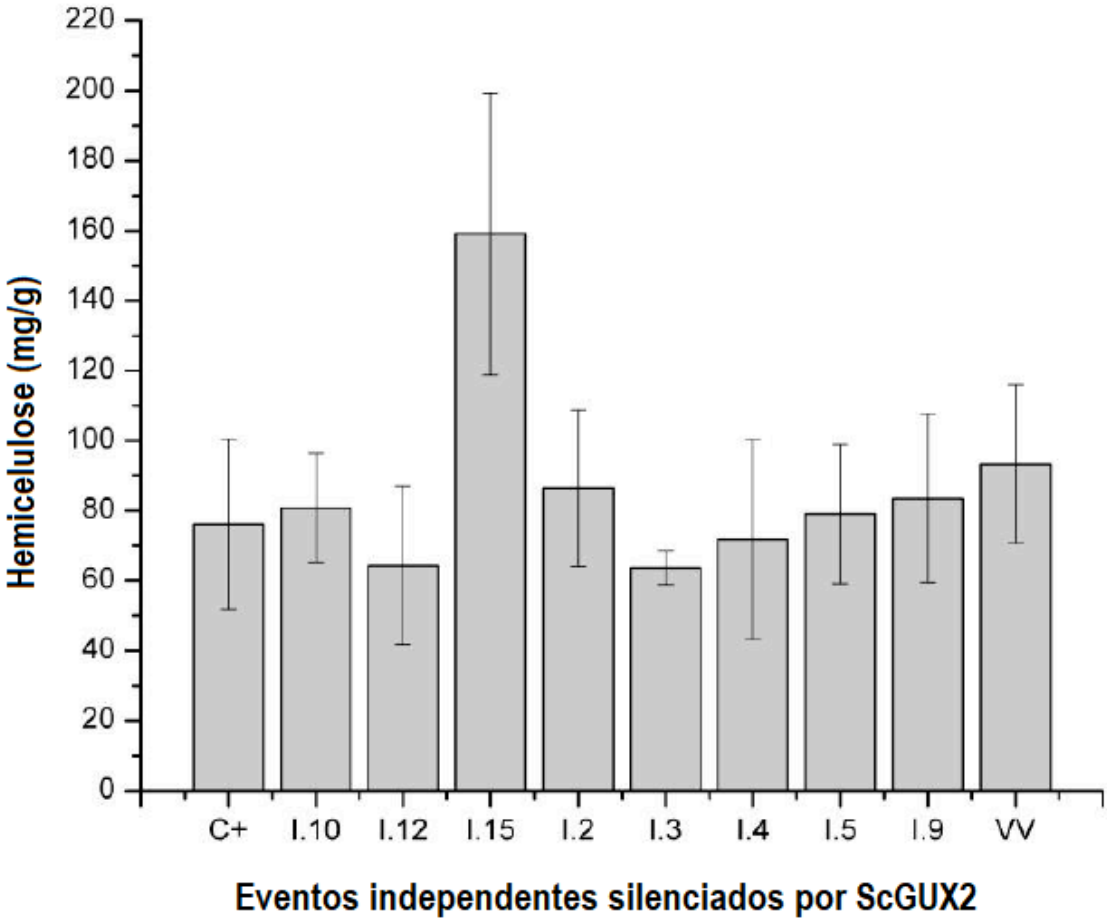

Hemicelulose (mg/g)
220
200
180
160
140
120
100
80
60
40
20
0
C+
I.10
I.12
I.15
I.2
I.3
I.4
I.5
I.9
VV
Eventos independentes silenciados por ScGUX2

Figura 13

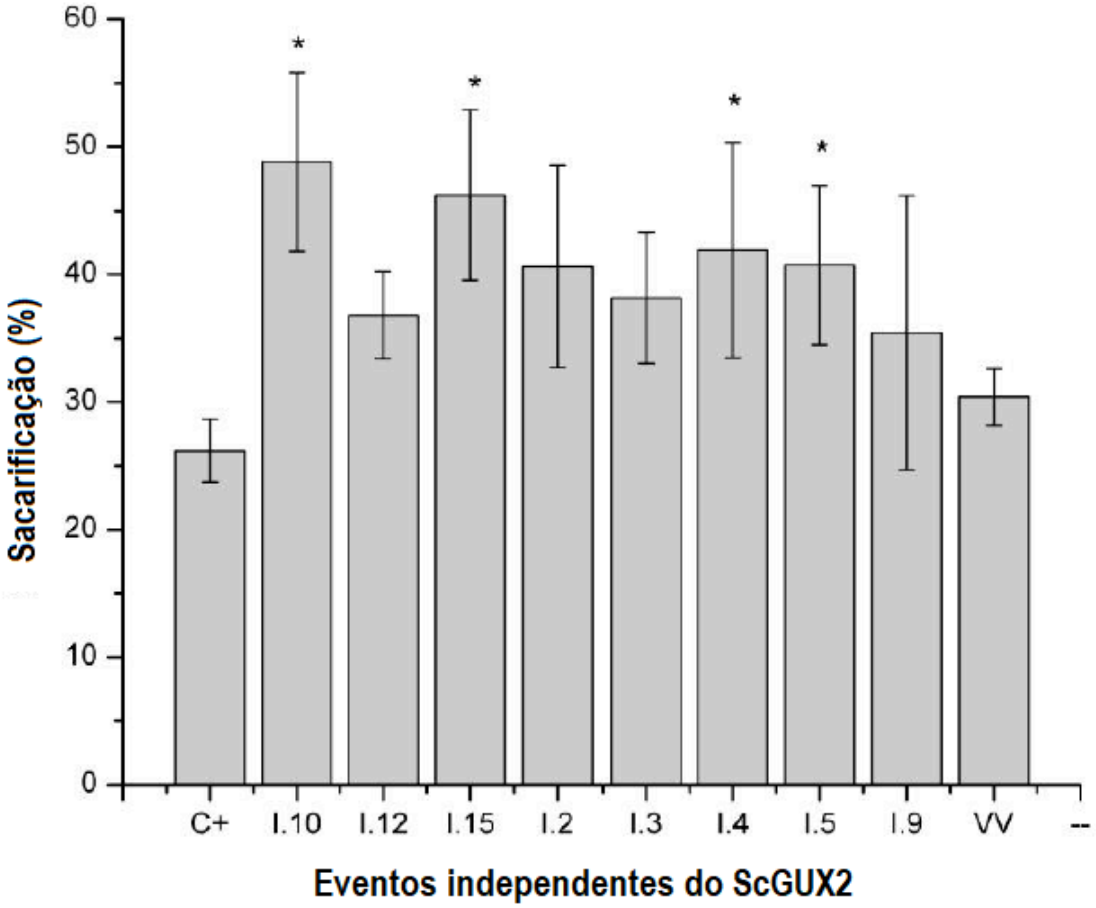

60
50
40
30
20
10
0
Sacarificação (%)
*
*
*
*
C+
I.10
I.12
I.15
I.2
I.3
I.4
I.5
I.9
VV
--
Eventos independentes do ScGUX2

Figura 14

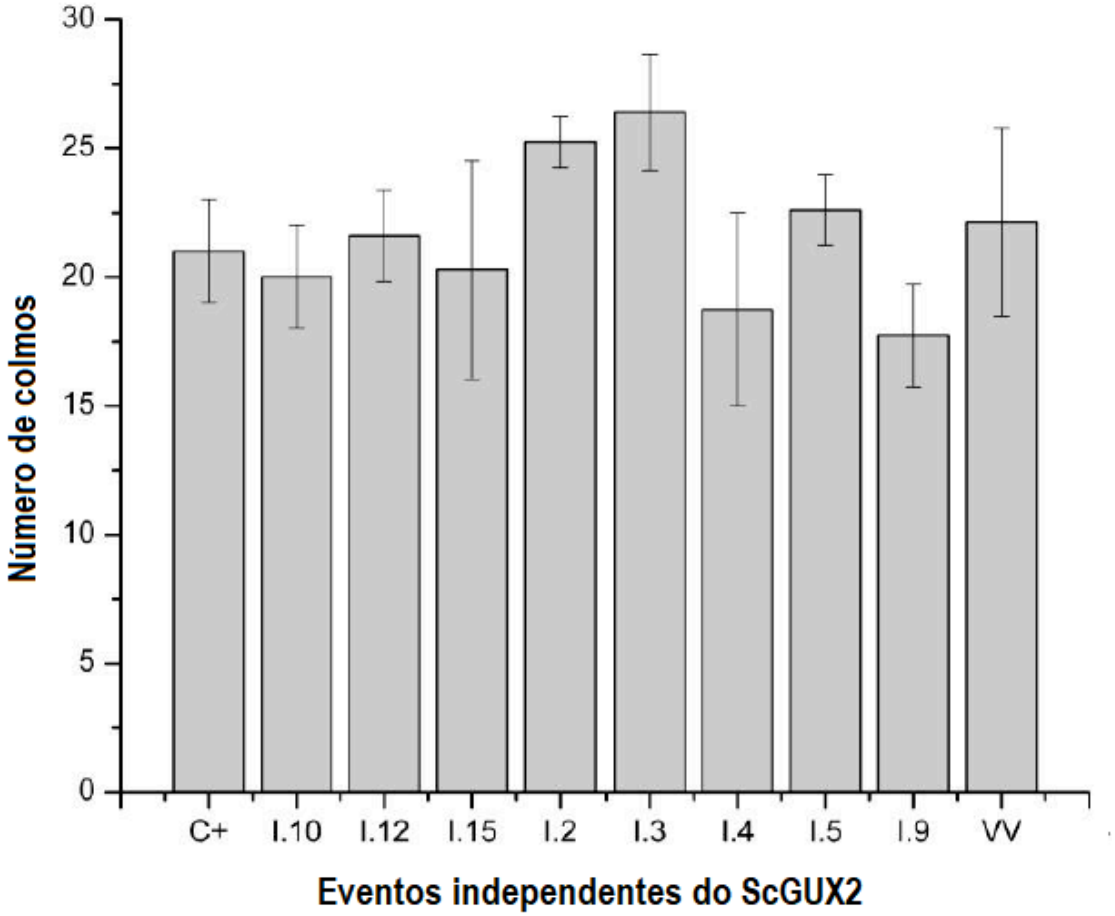

30
25
20
15
10
5
0
Número de colmos
C+
I.10
I.12
I.15
I.2
I.3
I.4
I.5
I.9
VV
Eventos independentes do ScGUX2

Figura 15

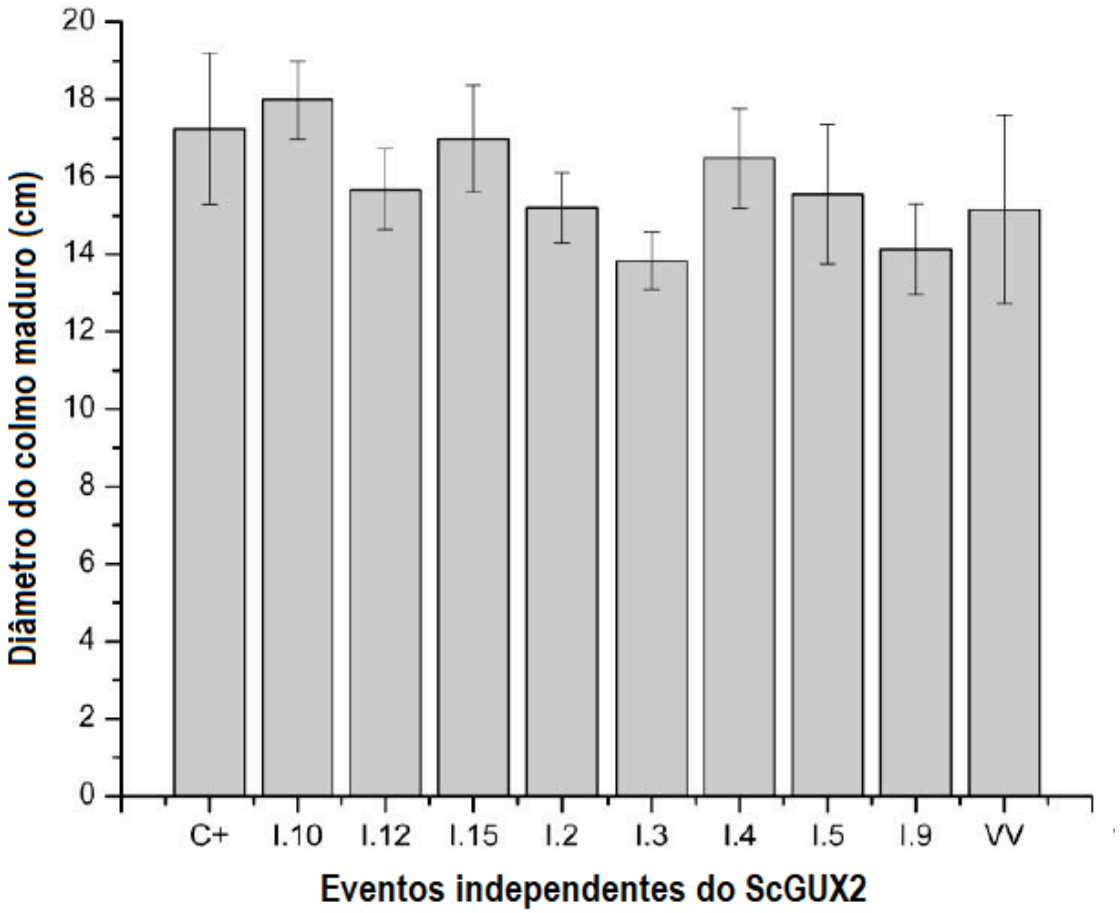

Diâmetro do colmo maduro (cm)
0
2
4
6
8
10
12
14
16
18
20
C+
I.10
I.12
I.15
I.2
I.3
I.4
I.5
I.9
VV
Eventos independentes do ScGUX2

Figura 16

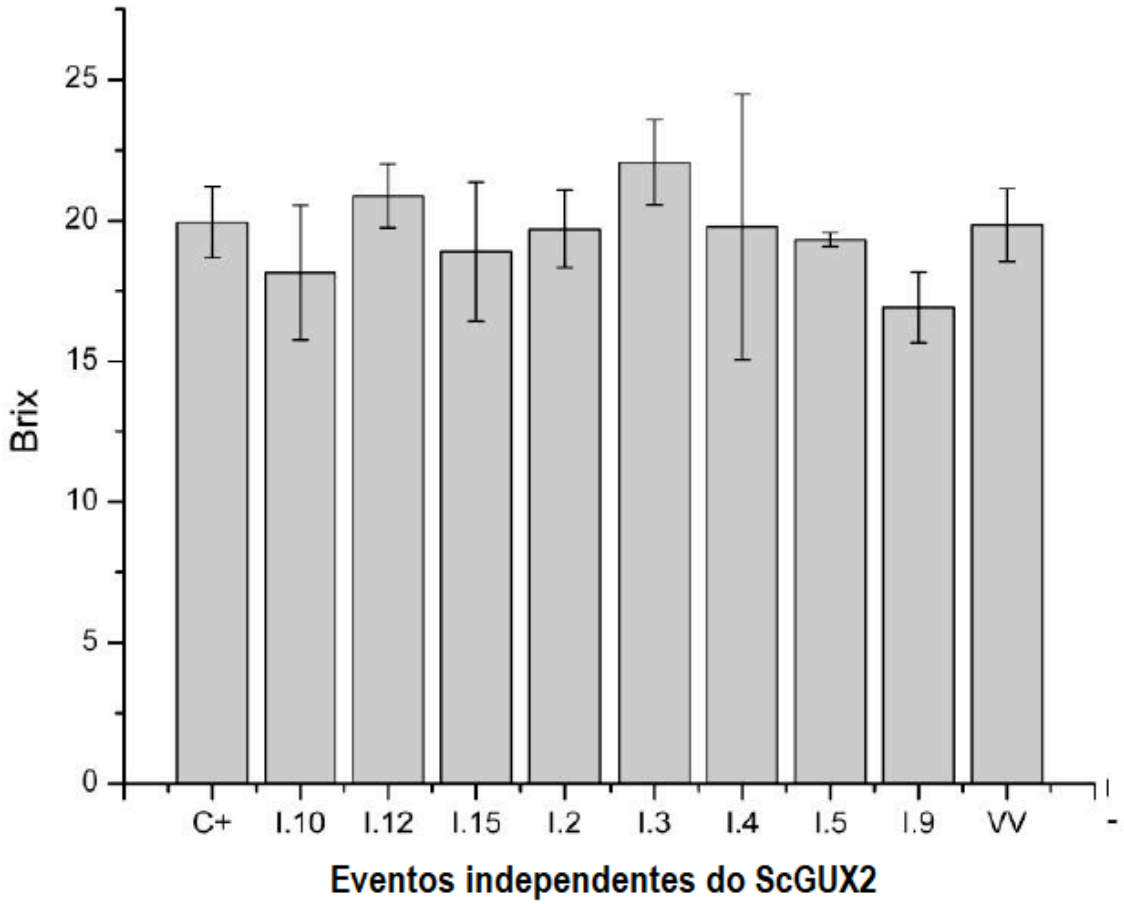

25
20
15
10
5
0
Brix
C+
I.10
I.12
I.15
I.2
I.3
I.4
I.5
I.9
VV
-
Eventos independentes do ScGUX2

Figura 17

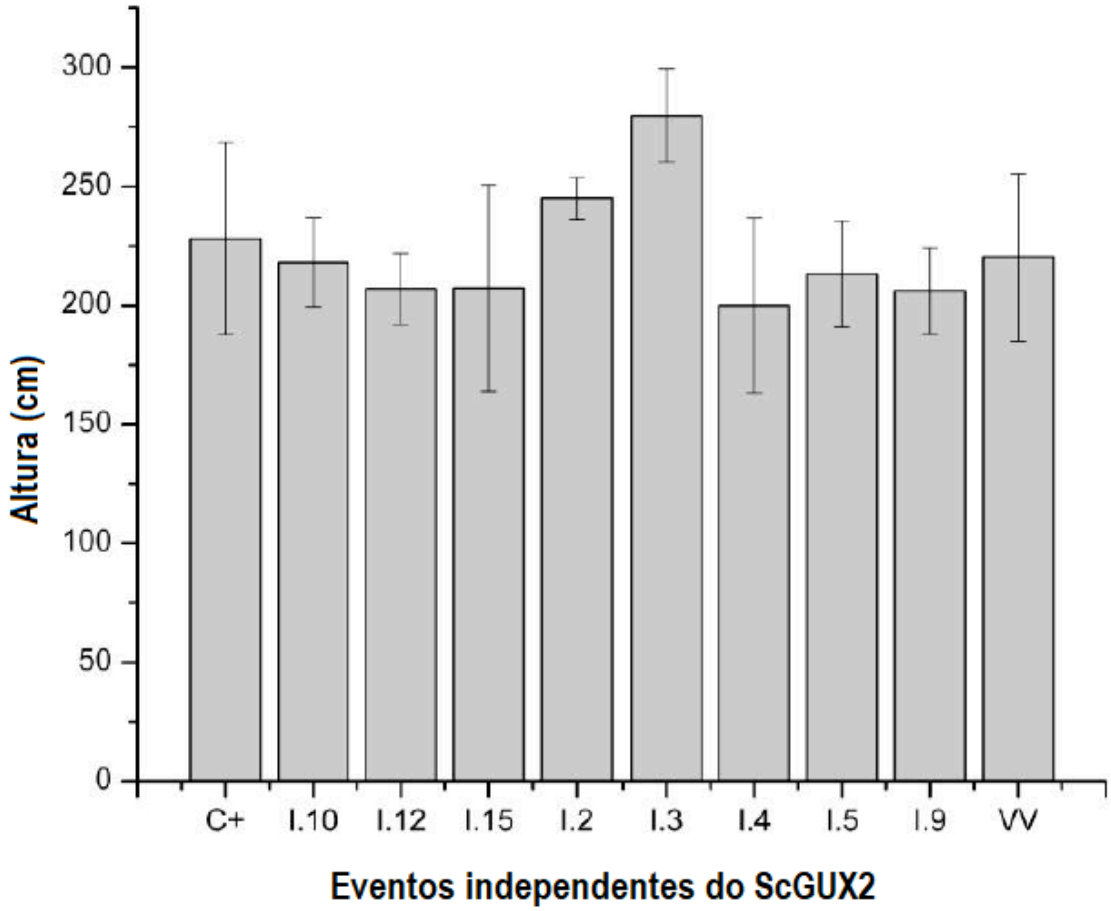

Altura (cm)
300
250
200
150
100
50
0
C+
I.10
I.12
I.15
I.2
I.3
I.4
I.5
I.9
VV
Eventos independentes do ScGUX2

Figura 18

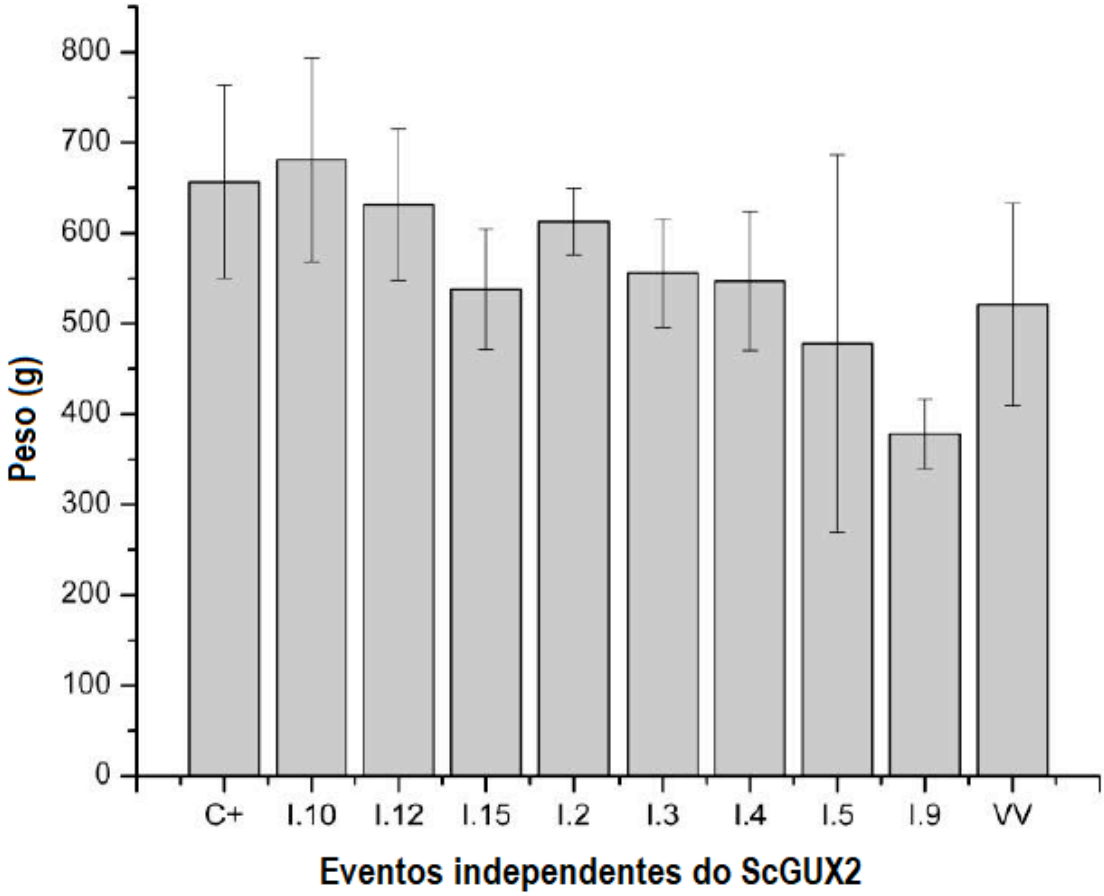

800
700
600
500
400
300
200
100
0
Peso (g)
C+
I.10
I.12
I.15
I.2
I.3
I.4
I.5
I.9
VV
Eventos independentes do ScGUX2

Figura 19

**RESUMO**

**PROCESSO PARA A PRODUÇÃO DE CANA-DE-AÇÚCAR COM MODIFICAÇÃO NA PAREDE CELULAR**

A presente invenção refere-se a um processo para a produção de cana-de-açúcar com modificação na parede celular, utilizando uma abordagem de RNAi para reduzir os níveis de expressão do gene homólogo de GUX em cana-de-açúcar (*ScGUX2*). Foi observado que a recalcitrância da biomassa nas plantas transgênicas foi reduzida e um aumento de pelo menos 15% no rendimento da sacarificação foi observado em plantas transgênicas de cana-de-açúcar. Descobriu-se também que nenhum efeito negativo foi observado em vários parâmetros agronômicos, como tamanho, peso, número de colmos e Brix. A parede celular das plantas transgênicas não apresentou redução dos níveis de celulose, hemicelulose e pectina. Os dados obtidos mostram que o processo sugerido na presente invenção apresenta um potencial comercial muito elevado, já que os eventos mostraram um aumento na sacarificação de pelo menos 15%, chegando até 30%, o que em larga escala representa um aumento na produção de etanol muito considerável.

**REIVINDICAÇÕES**

1. Processo para a produção de cana-de-açúcar transgênica, **caracterizado** por apresentar reduzidos níveis de transcritos do gene *ScGUX2* com modificação na parede celular e compreender ainda as seguintes etapas:

(a) Seleção de um gene de cana-de-açúcar homólogo a um gene que codifica proteína envolvida no depósito de ácido glucurônico na parece celular vegetal;

(b) Manipulação do genoma vegetal que reduza a expressão do gene *ScGUX2*;

(c) Análises de linhagens independentes de cana-de-açúcar com genoma modificado; e

(d) Avaliação da sacarificação da biomassa vegetal.

2. Processo, de acordo com a reivindicação 1, **caracterizado** pelo fato de utilizar no processo de seleção (etapa a) do gene *ScGUX2* de cana-de-açúcar o gene homólogo de *Arabidopsis thaliana* denominado *AtGUX2* e o gene homólogo de sorgo denominado Sb01g044930.

3. Processo, de acordo com a reivindicação 2, **caracterizado** pelo fato de o gene *AtGUX2* da *Arabidopsis thaliana* compreender a SEQ. ID. No. 1.

4. Processo, de acordo com a reivindicação 2, **caracterizado** pelo fato de o gene homólogo de sorgo compreender a SEQ. ID. No. 2.

5. Processo, de acordo com a reivindicação 1, **caracterizado** pelo fato de o gene da cana-de-açúcar *ScGUX2* compreender a SEQ. ID. No. 3, correspondente a região codificadora.

6. Processo, de acordo com a reivindicação 1, **caracterizado** pelo fato de que a redução da expressão do gene *ScGUX2* na etapa b compreender as seguintes sub-etapas:

(i) Seleção de região do gene *ScGUX2* para silenciamento gênico;

(ii) Construção de DNA para produção de RNA de interferência composta por promotor ligado operacionalmente a uma região de DNA do gene *ScGUX2* para formação de grampo de RNA, ligada operacionalmente a uma região terminadora;

(iii) Modificação do genoma da cana-de-açúcar mediante transgenia.

7. Processo, de acordo com a reivindicação 6, **caracterizado** pela região do gene *ScGUX2* usada na construção de RNA de interferência conter a sequência de 498 pb, representada pela SEQ. ID. No. 4.

8. Processo, de acordo com a reivindicação 6, **caracterizado** pelo fato de que a região de DNA contendo grampo de DNA com a região do gene *ScGUX2* contém a SEQ. ID. No. 4 clonada em

orientação *sense* e *antisense*, tendo um íntron entre ambas, representada pela SEQ. ID. No. 5.

9. Processo, de acordo com a reivindicação 1, **caracterizado** pelo fato de que as plantas apresentam aumento de no mínimo 15% nos níveis de sacarificação da biomassa vegetal comparativamente ao tipo selvagem.

10. Processo, de acordo com a reivindicação 1, **caracterizado** pelo fato de que o gene silenciado seja homólogo ao gene *ScGUX2* e apresente a função de modificar a composição e estrutura da parede celular para aumento da sacarificação da biomassa vegetal.

[bb.com.br] - Boleto gerado pelo sistema MPAG. 12/11/2018 14:11:03

***INSTRUÇÕES:***

***A data de vencimento não prevalece sobre o prazo legal. O pagamento deve ser efetuado antes do protocolo. Órgãos públicos que utilizam o sistema SIAFI devem utilizar o número da GRU no campo Número de Referência na emissão do pagamento. Serviço: 200-Pedido nacional de Invenção, Modelo de Utilidade, Certificado de Adição de Invenção e entrada na fase nacional do PCT***

***Clique aqui e pague este boleto através do Auto Atendimento Pessoa Física.***
***Clique aqui e pague este boleto através do Auto Atendimento Pessoa Jurídica.***

Recibo do Pagador

**BANCO DO BRASIL** | **001-9** | **00190.00009 02940.916188 12001.036172 1 77350000007000**

Nome do Pagador/CPF/CNPJ/Endereço
UNIVERSIDADE ESTADUAL DE CAMPINAS UNICAMP CPF/CNPJ: 46068425000133
CIDADE UNIVERSITARIA ZEFERINO VAZ, CAMPINAS -SP CEP:13084971
Sacador/Avalista

| Nosso-Número | Nr. Documento | Data de Vencimento | Valor do Documento | (=) Valor Pago |
|---|---|---|---|---|
| 29409161812001036 | 29409161812001036 | 11/12/2018 | 70,00 | |

Nome do Beneficiário/CPF/CNPJ/Endereço
INSTITUTO NACIONAL DA PROPRIEDADE INDUST CPF/CNPJ: 42.521.088/0001-37
RUA MAYRINK VEIGA 9 24 ANDAR ED WHITE MARTINS , RIO DE JANEIRO - RJ CEP: 20090910

Agência/Código do Beneficiário
2234-9 / 333028-1

Autenticação Mecânica

**BANCO DO BRASIL** | **001-9** | **00190.00009 02940.916188 12001.036172 1 77350000007000**

Local de Pagamento
**PAGÁVEL EM QUALQUER BANCO ATÉ O VENCIMENTO**

Data de Vencimento
11/12/2018

Nome do Beneficiário/CPF/CNPJ
INSTITUTO NACIONAL DA PROPRIEDADE INDUST CPF/CNPJ: 42.521.088/0001-37

Agência/Código do Beneficiário
2234-9 / 333028-1

| Data do Documento | Nr. Documento | Espécie DOC | Aceite | Data do Processamento |
|---|---|---|---|---|
| 12/11/2018 | 29409161812001036 | DS | N | 12/11/2018 |

| Uso do Banco | Carteira | Espécie | Quantidade | xValor |
|---|---|---|---|---|
| 29409161812001036 | 17 | R$ | | |

Nosso-Número
29409161812001036

(=) Valor do Documento
70,00

Informações de Responsabilidade do Beneficiário

A data de vencimento não prevalece sobre o prazo legal.
O pagamento deve ser efetuado antes do protocolo.
Órgãos públicos que utilizam o sistema SIAFI devem utilizar o número da GRU no campo Número de Referência na emissão do pagamento.
Serviço: 200-Pedido nacional de Invenção, Modelo de Utilidade, Certificado de Adição de Invenção e entrada na fase nacional do PCT

(-) Desconto/Abatimento

(+) Juros/Multa

(=) Valor Cobrado

Nome do Pagador/CPF/CNPJ/Endereço
UNIVERSIDADE ESTADUAL DE CAMPINAS UNICAMP CPF/CNPJ: 46068425000133
CIDADE UNIVERSITARIA ZEFERINO VAZ,
CAMPINAS-SP CEP:13084971
Sacador/Avalista

Código de Baixa

Autenticação Mecânica - **Ficha de Compensação**

21/11/2018 - BANCO DO BRASIL - 13:08:00
420304203 0017

COMPROVANTE DE PAGAMENTO DE TITULOS

CLIENTE: UNICAMP-MOVIMENTO TESOURO
AGENCIA: 4203-X CONTA: 131.009-7
==========
BANCO DO BRASIL
---
00190000090294091618812001036172177350000007000
BENEFICIARIO:
INSTITUTO N P I - INPI
NOME FANTASIA:
INSTITUTO NACIONAL DA PROPRIEDADE I
CNPJ: 42.521.088/0001-37
PAGADOR:
UNIVERSIDADE ESTADUAL DE CAMPINAS U
CNPJ: 46.068.425/0001-33
---
NR. DOCUMENTO 112.169
NOSSO NUMERO 29409161812001036
CONVENIO 02940916
DATA DE VENCIMENTO 11/12/2018
DATA DO PAGAMENTO 21/11/2018
VALOR DO DOCUMENTO 70,00
VALOR COBRADO 70,00
==========
NR.AUTENTICACAO 0.D1E.E46.536.CDA.5BD

UNICAMP

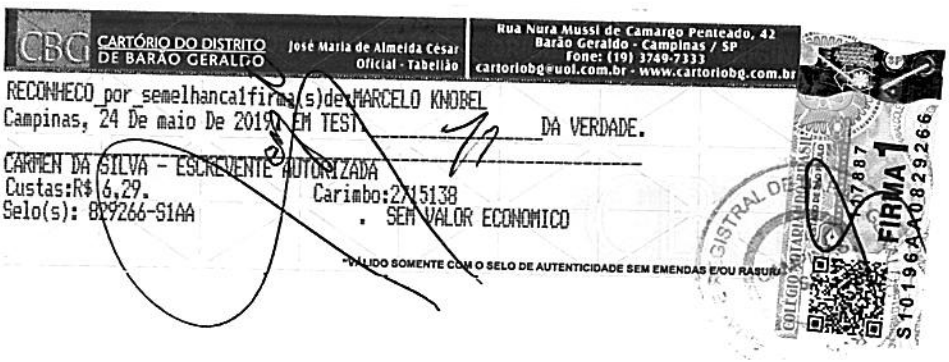
# *UNIVERSIDADE ESTADUAL DE CAMPINAS*

## PROCURAÇÃO

**UNIVERSIDADE ESTADUAL DE CAMPINAS**, autarquia em regime especial, inscrita no CNPJ/MF nº 46.068.425/0001-33, com sede na Cidade Universitária "Zeferino Vaz", Distrito de Barão Geraldo, Campinas, Estado de São Paulo, neste ato representada por seu **Magnífico Reitor Professor Doutor Marcelo Knobel**, constitui e nomeia **PATRÍCIA FRANCO LEAL GESTIC**, brasileira, casada, engenheira de alimentos, portadora do RG nº 32.372.385-8, inscrita no CPF nº 219.274.108-19, matrícula 30.418-9; **CIRO DE LA CERDA**, argentino, solteiro, engenheiro de produção mecânica, portador do RG nº W255248-9, inscrito no CPF nº 045.845.188-66, matrícula 17.857-8; e **NEWTON CESARIO FRATESCHI**, brasileiro, casado, físico, portador do RG nº 13771984, inscrito no CPF 059.239.518-94, matrícula 265942; **RAQUEL MOUTINHO BARBOSA**, brasileira, casada, bióloga, portadora do RG nº 28.234.848-7, inscrita no CPF: 305.798.788-56, matrícula 315944; todos lotados na Agência de Inovação Inova Unicamp, com endereço na Av. Roxo Moreira, 1831, Caixa Postal 6131, Cidade Universitária "Zeferino Vaz", Distrito de Barão Geraldo, em Campinas-SP; a quem confere poderes para, nos termos dos atos normativos, das portarias e das resoluções baixadas pelo INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL - INPI, requerer pedidos de proteção de propriedade industrial desenvolvidos pela Unicamp, junto ao **INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL** através do sistema e-INPI, com poderes para praticar todos os atos que se fizerem necessários aos pedidos e manutenção da proteção de direitos de propriedade intelectual em benefício da outorgante junto ao INPI, sendo vedado o substabelecimento dos poderes ora concedidos. Esta procuração terá vigência de 2 (dois anos), a partir da data de sua assinatura.

Campinas, 21 de maio de 2019.

**UNIVERSIDADE ESTADUAL DE CAMPINAS**
**Professor Doutor Marcelo Knobel**
**Reitor - UNICAMP**


CBG CARTÓRIO DO DISTRITO DE BARÃO GERALDO
José Maria de Almeida César
Oficial - Tabelião
Rua Nura Mussi de Camargo Penteado, 42
Barão Geraldo - Campinas / SP
Fone: (19) 3749-7333
cartoriobg@uol.com.br - www.cartoriobg.com.br
RECONHECO por semelhança 1 firma(s) de: MARCELO KNOBEL
Campinas, 24 De maio De 2019. EM TEST DA VERDADE.
CARMEN DA SILVA - ESCREVENTE AUTORIZADA
Custas:R$ 6,29. Carimbo:2715138
Selo(s): BC9266-S1AA . SEM VALOR ECONOMICO
VÁLIDO SOMENTE COM O SELO DE AUTENTICIDADE SEM EMENDAS E/OU RASURA
FIRMA 1
S1019AA0829266

Este anexo apresenta o código de controle da listagem de sequências biológicas.

**Código de Controle**

Campo 1


Campo 2


Outras Informações:

- Nome do Arquivo: Sequencia_ST25.txt
- Data de Geração do Código: 31/05/2019
- Hora de Geração do Código: 12:04:05
- Código de Controle:
  - Campo 1: 02D9FB3405D01B52
  - Campo 2: 13AEF2136333525A