UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

MARISHANI MARIN CARRASCO

Gene coexpression networks provide deeper insights into sucrose accumulation in sugarcane cultivars.

Redes de coexpressão de genes fornecem informações mais aprofundadas sobre o acúmulo de sacarose em cultivares de cana-de-açúcar.

Campinas

2021

MARISHANI MARIN CARRASCO

Gene coexpression networks provide deeper insights into sucrose accumulation in sugarcane cultivars.

Redes de coexpressão de genes fornecem informações mais aprofundadas sobre o acúmulo de sacarose em cultivares de cana-de-açúcar.

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

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

Orientadora: Profa. Dra. Anete Pereira de Souza

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELA ALUNA MARISHANI MARIN CARRASCO E ORIENTADA PELA PROFESSORA DOUTORA ANETE PEREIRA DE SOUZA.

Campinas

2021

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

C231g

Carrasco, Marishani Marin, 1994-
Gene coexpression networks provide deeper insights into sucrose accumulation in sugarcane cultivars / Marishani Marin Carrasco. – Campinas, SP : [s.n.], 2021.

Orientador: Anete Pereira de Souza.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Redes reguladoras de genes. 2. RNA-seq. 3. Sacarose - Metabolismo. 4. Ontologia genética. I. Souza, Anete Pereira de, 1962-. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Redes de coexpressão de genes fornecem informações mais aprofundadas sobre o acumulo de sacarose em cultivares de cana-de-açúcar
Palavras-chave em inglês:
Gene regulatory networks
RNA-seq
Sucrose - Metabolism
Gene ontology
Área de concentração: Genética Vegetal e Melhoramento
Titulação: Mestra em Genética e Biologia Molecular
Banca examinadora:
Anete Pereira de Souza [Orientador]
Reginaldo Massanobu Kuroshu
Lilian Padilha
Data de defesa: 14-12-2021
Programa de Pós-Graduação: Genética e Biologia Molecular

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-7161-1543
- Currículo Lattes do autor: http://lattes.cnpq.br/6488170810422524

Campinas,14 de Decembro 2021

COMISSÃO EXAMINADORA

Profa. Dra. Anete Pereira de Souza

Prof. Dra. Lilian Padilla

Prof. Dr. Reginaldo Massanobu Kuroshu

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 Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia.

ACKNOWLEDGEMENTS

To God for life, the blessings that each day bestows on me, and for always accompanying me on my path.

To my parents Josefina and Jorge, grandparents Evangelina and Ernestina, and boyfriend Alex for all the love, motivation, effort, and pacient that they give me in order to fulfill my professional goals, who are the fundamental axis and accompanying me on this journey.

To my advisor Prof. Dr. Anete Pereira de Souza for kindly welcoming me into her research group, for all the teachings, for sharing knowledge and for guiding me during the research process of this dissertation, urging me to improve and overcome my limits. Thank you for the scientific-professional contribution in my life, for your patience and dedication as well as for the trust placed in me.

To the doctoral student Alexandre Hild Aono my mentor, and Ricardo Jose Pimenta, for teaching me all the fundamental technical and scientific knowledge in the post-graduation program since the first day, training me on my bioinformatics initiation. In addition to all motivation, advice, and friendship. Thank for your dedication and active collaboration in writing the article and manuscript.

To all colleagues at the Laboratory of Genetic and Molecular Analysis (LAGM) for contributing in some way to my training; Benicio, Felipe Francisco, Melina, Rafaela, Giovani, Michele, Carla, Paulo, Rebeca, Fernanda, Felipe, Prianda and Lucas for friendship, encouragement, support, moments of relaxation and words of comfort.

To the laboratory technicians Danilo, Aline and Patrícia for all their help.

To the Graduate Program in Genetics and Molecular Biology and to the State University of Campinas – UNICAMP.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

CAPES - (Academic Excellence Program - PROEX - 88887.479604/2020-00) for the Master's scholarship granted.

Finally, I thank those who somehow made it possible to carry out my Master's research.

RESUMO

A cana-de-açúcar (*Saccharum* spp.) é a cultura mais importante para a produção de açúcar e etanol de primeira geração. As cultivares modernas de cana-de-açúcar são altamente poliplóides e heterozigotas, apresentando vários alelos diferentes em cada locus, e esse alto nível de complexidade genética cria desafios durante os programas de melhoramento convencional e molecular. Reconhecendo a dificuldade de melhorar o teor de sacarose por meios convencionais, recursos substanciais têm sido direcionados para a compreensão das bases fisiológicas, celulares, bioquímicas e moleculares da produção e acúmulo de açúcar na cana-de-açúcar. O acúmulo de sacarose é descrito como regulado por uma rede de genes modulados durante a maturação do colmo. Nesse contexto, experimentos de RNA-Seq representam uma poderosa ferramenta para estudos de expressão gênica, permitindo obter o perfil do transcriptoma. Usando este tipo de dados, redes de coexpressão gênica podem ser construídas e auxiliar na identificação de genes com padrões de expressão semelhantes biologicamente relevantes para um fenótipo específico. Este projeto teve como objetivo modelar redes de coexpressão para diferentes variedades e tecidos de cana-de-açúcar, a fim de obter uma melhor compreensão da interação gênica envolvida no metabolismo da sacarose da cana-de-açúcar. Usamos uma rede global de coexpressão modelada com dados de RNA-Seq de dezenove diferentes genótipos de cana-de-açúcar e diferentes tecidos (raízes, folhas e colmos). Além disso, uma rede de coexpressão gênica ponderada foi modelada e módulos potencialmente funcionais foram identificados por meio de uma abordagem completa de agrupamento hierárquico. Dentro da estrutura da rede, pudemos identificar 153 grupos diferentes, que foram contrastados com os genes diferencialmente expressos identificados. Por meio de análises de enriquecimento, pudemos estabelecer um total de 1.484 genes relacionados ao metabolismo da sacarose correspondentes a quatro módulos de rede diferentes. Além do metabolismo do amido e da sacarose, as vias do metabolismo da frutose e da manose também foram enriquecidas para este grupo de genes e dois módulos de rede adicionais foram selecionados. Esse grupo coeso apresentou interações gênicas relevantes avaliadas por meio de categorias de Gene Ontology, como resposta ao estresse; processo metabólico de carboidratos; processo metabólico de carboidratos celulares; organização da

parede celular; processo metabólico do polissacarídeo celular; resposta a estímulos e processos metabólicos de polissacarídeos, representando uma importante fonte de dados para o entendimento da dinâmica das atividades genéticas no acúmulo de açúcar. Além disso, realizamos análises de DGE dentro dos seis módulos finais selecionados, mostrando 105 genes regulados positivamente em genótipos com maior teor de sacarose e 78 genes regulados negativamente, englobando diferentes perfis funcionais avaliados através das enzimas associadas. Por fim, para acessar as especificidades funcionais das redes de *Saccharum spontaneum*, *Saccharum officinarum* e da cultivar híbrida SP80-3280, realizamos análises de inferências biológicas para conexões hub e comuns entre esses três genótipos. Nossos resultados mostraram o potencial de reunir um grande número de transcritos com expressão diferencial significativa em categorias de sacarose e auxiliar na compreensão da dinâmica das atividades metabólicas e genéticas em escala global sobre o metabolismo da sacarose em cana-de-açúcar.

Palavras-chave: Redes de coexpressão, RNA-Seq, Metabolismo de Sacarose , Ontologia genética.

Abstract

Sugarcane (*Saccharum* spp.) is the most important crop for sugar and first-generation ethanol production. Modern sugarcane cultivars are highly polyploid and heterozygous, presenting several different alleles at each locus, and this high level of genetic complexity creates challenges during conventional and molecular breeding programs. Recognizing the difficulty of sucrose content improvement through conventional means, substantial resources have been directed to understanding the physiological, cellular, biochemical and molecular basis of sugar production and accumulation in sugarcane. Sucrose accumulation is described as regulated by a network of genes modulated during culm maturation. In this context, RNA-Seq experiments represent a powerful tool for gene expression studies, allowing the transcriptome profile to be obtained. Using this type of data, gene coexpression networks can be constructed and assist the identification of genes with similar expression patterns biologically relevant to a specific phenotype. This project aimed at modelling coexpression networks for different sugarcane varieties and tissues in order to obtain a better comprehension of the gene interaction involved in sugarcane sucrose metabolism. We used a global coexpression network modeled with RNA-Seq data from nineteen different sugarcane genotypes and different tissues (roots, leaves and culms). Additionally, a weighted gene coexpression network was modeled and potentially functional modules were identified through a complete hierarchical clustering approach. Inside the network structure, we could identify 153 different groups, which were contrasted with the differentially expressed genes identified. Through enrichment analyses, we could establish a total of 1,484 genes related to sucrose metabolism corresponding to four different network modules. In addition to starch and sucrose metabolism, fructose and mannose metabolism pathways were also enriched for this group of genes and two additional network modules were selected. This cohesive group presented relevant gene interactions assessed through Gene Ontology categories, such as response to stress; carbohydrate metabolic process; cellular carbohydrate metabolic process; cell wall organization; cellular polysaccharide metabolic process; response to stimulus and polysaccharide metabolic process, representing an important source of data for understanding the dynamics of genetic activities on sugar accumulation. Furthermore, we performed

DGE analyses inside the six final selected modules, showing 105 genes up regulated in genotypes with higher sucrose content and 78 genes as down regulated, encompassing different functional profiles assessed through the associated enzymes. Lastly, in order to access functional specificities of networks of *Saccharum spontaneum*, *Saccharum officinarum* and the SP80-3280 hybrid cultivar, we performed biological inferences analysis for hub and common connections between these three genotypes. Our results showed the potential of assembling a large number of transcripts with significant differential expression in sucrose categories and assisting the comprehension of the dynamics of metabolic and genetic activities on a global scale on sucrose metabolism on sugarcane.

Keywords: coexpression networks, RNA-Seq, sucrose metabolism, Gene Ontology categories.

List of Figures

Contents

1 Introduction

Sugarcane is a major sugar crop, being the source about 80% of the global sugar production (Chen et al., 2019a). Besides, this species is one of the best options for the generation of bioenergy (Silva et al., 2007), due to its capability to store sucrose as a primary energy source, instead of more complex compounds (Silva et al., 2007). For 2021/2022, it is estimated that the Brazilian production will reach 628.1 million tons; this represents a reduction of 4% compared with previous years productions, partially due to the COVID-19 pandemic, which forced companies to make modifications in their routines and suspend contracts and and visits to farms (CONAB, 2021).

Current sugarcane cultivars are derived from interspecific crosses between *Saccharum officinarum* (a species with high sugar production) and *Saccharum spontaneum* (a species highly tolerant to biotic and abiotic stresses), and consecutive backcrosses with *S. officinarum* (Cuadrado et al., 2004). This process of sugarcane breeding caused the formation of a highly complex and polyploid genome (Palhares et al., 2012), with a size of approximately 10 Gb (Piperidis et al., 2010), frequent occurrence of aneuploidies, and great cytogenetic complexity (Daniels and Roach, 1987). The challenge of studying the sugarcane genome forces us to find more precise and specific tools to understand the genetic interactions and metabolic processes for sucrose accumulation. The *S. spontaneum* genome has been recently sequenced (Zhang et al., 2018) and provided a larger comprehension of sugarcane's genomic organization. However, additional strategies are necessary to unravel the specific processes of genetic interactions. Sugarcane transcriptome experiments, for instance, can increase the sets of sequence information and assist the process of deciphering genes interactions.

Sucrose content is the most important commercial trait for sugarcane and a priority characteristic for all sugarcane improvement programs (Jackson, 2005). However, the processes of the transport and metabolism of this carbohydrate are still not fully understood. In sugarcane, it has been described that the stems are the principal sink tissues that store very high concentrations of sugars within the parenchyma cells (Patrick et al., 2013; Bihmidine et al., 2013, 2015a). Sucrose transporters (SUTs) load sucrose into the phloem of leaf minor

veins and also function to retrieve sucrose from the apoplasm during transport (Slewinski and Braun, 2010; Baker et al., 2016; Julius et al., 2017). Moreover, SUTs are potentially responsible for sucrose efflux into the cell wall space from phloem parenchyma cells and the bundle sheath (Jung et al., 2015). There are many unknown mechanisms related to this general process, and novel strategies must be employed so that it is fully understood.

Different methodologies have been applied to understand this cascade of reactions, such as the identification of genes through *in silico* methods (Zinkgraf et al., 2017; Zhang et al., 2018; Gomathi et al., 2018). However, despite extensive efforts, the improvement of sucrose content through plant breeding programs remains very slow globally (Vicentini et al., 2015). Sucrose synthesis and accumulation in sugarcane is highly complex and may involve a large network of interactions operating at different levels (biochemical, molecular and genetic) of organization. Different studies based on gene interactions aimed at understanding the dynamics of gene and metabolic activities associated with sucrose accumulation (Mattiello et al., 2015; Ferreira et al., 2016; Thirugnanasambandam et al., 2018b; Chen et al., 2019a) and the most promising strategies have been based on RNA-Seq (Cardoso-Silva et al., 2014; Ma et al., 2020).

RNA-Seq experiments allow the transcriptome profile of an organism to be obtained and are a starting point for the identification of new and rare transcripts. There are several advantages on using RNA-Seq, including the more precise measurement of transcript levels and their isoforms (Marguerat and Bähler, 2010). The great volume and the complexity of data generated by RNA-Seq experiments requires scalable, fast and mathematically principled analysis tools, such as gene coexpression networks, which can capture biologically important patterns in gene expression data, enabling functional analyses of genes and interpretation of genetic variants (Saha et al., 2017). Computational methods provide powerful approaches for understanding complex biological systems (Ficklin and Feltus, 2011a) and biological networks can assist on unravelling the inherent structure of molecular interactions, what can lead to the discovery of driver genes and meaningful pathways (Mall et al., 2017a).

A network is a graph that can be used to model how members of a discrete set can be

related to another one according to some relationship (Ferlic and Tracy, 2016). Correlation networks have increasingly been used in bioinformatics applications and have been found useful for describing the pairwise relationships among gene transcripts (Zhou et al., 2002). A common method used to model these networks is the Weighted Correlation Network Analysis (WGCNA), which can be used together with different approaches, e.g. finding modules of highly correlated genes, identifying intramodular hub genes, relating modules to one another, externalizing sample traits, and calculating module membership measures (Langfelder and Horvath, 2008a). The detection of community structure is a widely accepted approach for investigating the principles governing biological systems (Mall et al., 2017a). These communities, also known as modules, have been shown to comprise groups of biomolecules that physically interact, are functionally cohesive, co-regulated or correspond to similar biological pathways (Mitra et al., 2013a). The notion of communities means that nodes within one module are densely connected to each other and sparsely connected to nodes outside that community (Mall et al., 2017a). Defining appropriate methods to assess and detect communities in coexpression networks is essential and allows a specific treatment on RNA-Seq data.

In this context, this project aimed at identifying a large number of genes involved in the metabolism of sucrose by analyzing sugarcane transcripts through coexpression networks. A global network was constructed, and a subsequent selection of modules containing a great number of genes related to sucrose metabolism was performed. Additionally, DGE analyses were carried out for all the genotypes under analysis based on their contrasting sugar content. Likewise, specific networks were constructed for representative genotypes and hybrids cultivars. We observed that, through different *in silico* strategies, we could access novel genes involved in these processes, which may assist sugarcane breeding programs through molecular strategies.

2 Bibliographic Review

2.1 Sugarcane

Sugarcane species (*Saccharum* spp) have China and India as center of origin and New Guinea as the center of diversity (James, 2004). In the colonial period, sugarcane was introduced in Brazil and thenceforth, the cultivation of sugarcane has had prominent importance on the Brazilian economy (Figueiredo, 2008). Because of favorable climate factors in this country, its cultivation expanded rapidly to the Northeast and Southeast of the country, and today Brazil is the principal producer of sugarcane in the world (CONAB, 2021).

Sugarcane comprises a group of polyploid species belonging to to tribe Andropogoneae, family Poaceae, known as the *Saccharum* complex. The genus *Saccharum* is composed by six species: *S. officinarum* (2n = 80), *Saccharum barberi* (2n = 81-124), *Saccharum robustum* (2n = 60-250), *S. spontaneum* (2n = 40-128), *Saccharum sinense* (2n = 111-120) and *Saccharum edule* (2n = 60-80). This variability in the number of chromosomes strongly influences hybridization within the genus; modern sugarcane derived from interspecific crosses between *S. officinarum* and *S. spontaneum* followed by successive backcrosses with *S. officinarum* (Cuadrado et al., 2004). This process, historically known as "nobilization", resulted in modern sugarcane being highly polyploid and aneuploid (Palhares et al., 2012), heterozygous, with a genome of approximately 10 Gb (Piperidis et al., 2010), with 40-50% of repetitions, and variable ploidy among different loci. This process further deepened the genetic complexity of these species (Daniels and Roach, 1987; Garsmeur et al., 2018) as schematized on Figure 1.

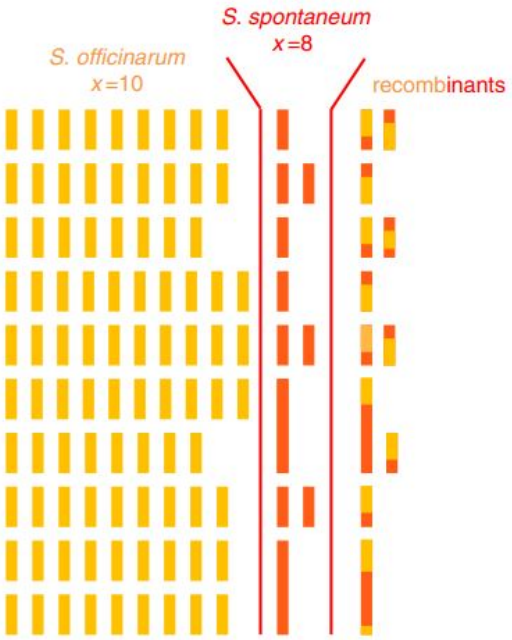

Figure 1: Schematic representation of the genome of a typical modern sugarcane cultivar (Garsmeur et al., 2018).

However, these crosses, performed by genetic improvement programs, resulted in these modern varieties acquiring interesting agronomic characteristics such as a high weight of plants and canes, high sucrose content, the ability of the stalks to drop old leaves, high fiber content, disease resistance and, not less importantly, a high adaptability to different climates (James, 2004). Plant breeding programs use these agronomic characteristics to obtain optimal commercial hybrid cultivars, which normally takes a long time - between 10 and 12 years. Sugarcane has an enormous economic importance because of their primordial derivatives products such as sugar and ethanol (James, 2004; Palhares et al., 2012). In Brazil, the sugarcane industry is one of the sectors with the greatest generation of direct and indirect jobs, and represents an important percentage of the agricultural PIB of the country, as observed on Figure 2. It is estimated that 628.1 million tons of sugarcane will be harvested during the 2021/2022 harvest (CONAB, 2021).

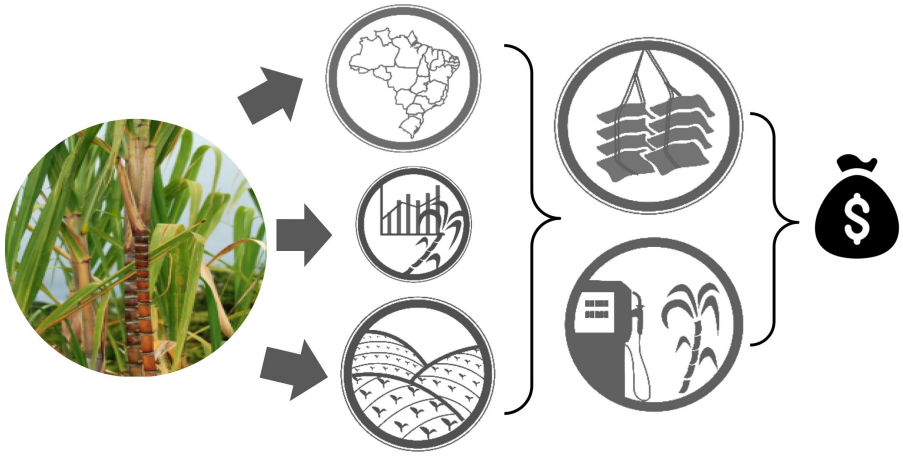

Figure 2: Schematic representation of the big sugarcane manufacturing and economic importance for Brazil, in which each circle represents the basic pyramid of sugarcane derived products and the economic relationship between them (CONAB, 2021).

Thus, understanding this huge economic importance, sugarcane breeding programs look for approaches that can reduce the time to obtain more productive cultivars, betting on efficient molecular techniques capable of predicting genotypes of modern cultivars in shorter periods and and at cheaper rates. Sugarcane breeding programs have special attention on sugar content and consider it a valuable resource, based on the fact that high-sugar cultivars allow the increase of productivity and consequently of profit (Jackson, 2005).

2.2 Sucrose

Sugarcane stores great amounts of sucrose - the most abundant disaccharide on the world, used as a food sweetener since 2000 BC (Patrick et al., 2013). Furthermore, sucrose is the main photo-assimilate of photosynthesis in plants. As a carbohydrate that is transported at long distances (Julius et al., 2017), sucrose is vital for the normal functioning of plants's energy machinery; for that reason, a plant without sucrose is not conceivable (Salerno and Curatti, 2003).

In plants, sucrose metabolism is composed by its synthesis, transport and storage. Leaves are photosynthetic regions, where sucrose is synthesized, and from which it needs to be trans-

ported to non-photosynthetic regions, such as the culm and roots. This is possible thanks to sugar transporters, which transfer sucrose through phloem tissues. After transport, sucrose is stored in mature culms (Slewinski and Braun, 2010; Bihmidine et al., 2015a). In the source tissues, the loading of sucrose into the phloem can involve two ways: the symplasmic and apoplasmic pathways (Patrick et al., 2013). On the symplasmic pathway, sucrose migrates through sieve tubes and phloem parenchyma to enter parenchyma cells by interconnecting plasmodesmata, where it is stored. A main characteristic of this mechanism is the presence of the apoplasmic barrier, which divides vascular bundles from sugar-storage cells and allows osmotic potentials of their separate apoplasms to be independently regulated (Patrick et al., 2013; Julius et al., 2017). On the apoplasmic pathway, sucrose amounts are larger and the presence of transporters is fundamental, as these show specific membrane affinity by easily moving sucrose across all organelles membranes to reach to the vacuoles (Patrick et al., 2013; Bihmidine et al., 2015a).

2.3 Genes and Enzymes Related to Sugarcane Sucrose Accumulation

The metabolic pathways of sucrose and starch are closely related; while sucrose is metabolized, transported and accumulated in the presence of solar energy, starch is metabolized in the absence of light energy (Kanehisa and Goto, 2000). Both these carbohydrates, which have glucose in their composition, are degraded and are used as oxidizable substrate for mitochondrial respiration, which is necessary for several plant developmental processes (Douce and Day, 1985). Starch metabolism is determined by the amount of degradative enzyme activity and the accessibility of the particulate substrate as pyrophosphate, which is essential for sucrose cleavage; thus, sucrose can be synthesised from starch on temporal circumstances controlled by circadian rhythm (Zhu et al., 2017).

In sugarcane, photoassimilates need to be transported to sink organs, where they are stored in the parenchymal cells in form of sucrose (Wang et al., 2017; Moore, 1995; Rae et al., 2005a; Chen et al., 2012). Several studies indicate that two types of sucrose unloading exist

in plants (Moore, 1995; Patrick et al., 2013; Wang et al., 2017) (Fig. 3): (a) the symplastic, and (b) the apoplasmatic unloading. In (a), when the sucrose transporters are available in sieve tubes, sucrose with molecules of water and energy enzymes migrate to phloem parenchyma, where sucrose moves to parenchymal cells (Moore and Cosgrove, 1991; Welbaum and Meinzer, 1990) and goes inside the vacuole by the activity of proton symporters and aquaporines (Moore and Cosgrove, 1991; Welbaum and Meinzer, 1990; Carpaneto et al., 2005; Bihmidine et al., 2015b). A main characteristic of this symplastic unloading is the presence of an apoplasmic barrier that divides vascular bundles, generating differences on the pressure in cells (Patrick et al., 2013). On the other hand, in (b), the unloading happens when there is enough sucrose in the sink organs (Carpaneto et al., 2005; Wang et al., 2017; Patrick et al., 2013), it migrates to the plasma membrane that contain available invertases that hydrolyses the sucrose into hexoses (Chen et al., 2012, 2015; Patrick et al., 2013). Such activities could explain why sucrose is present in immature internodes and has a greater presence of enzyme activities in this tissue (Verma et al., 2011; Bindon and Botha, 2002). Thus, the sink organ determines, together with several enzymes, the content and abundance of sucrose in the stems (Smith and Stitt, 2007; McCormick et al., 2006a).

Sucrose metabolism is a complex physiological process controlled by a wide range of genes in different plant tissues (Wang et al., 2017). In general, the sucrose synthesized in the leaves will be transported by the phloem to the heterotrophic organs of the sink to supply energy for the growth and development of the plant (McCormick et al., 2006b). Therefore, sugar transporters represent an important element in plants' energy supply, regulating the transport, compartmentalization and storage of sucrose. These proteins can be divided into monosaccharide transporters, SUTs, and hexose and sucrose transporters (SWEETs), being predominantly expressed in mature leaves of sugarcane for exporting sucrose and in internodes for accumulating it (Rae et al., 2005b). Several sugar transporters, such as ShSUT1, ShPST2a, ShPST2b and ShSUT4, have already been demonstrated as having differential expression in plant tissues (Casu et al., 2015a), showing the need of evaluating these biological processes across different parts of plants for unravelling the entire cascade of metabolic reactions.

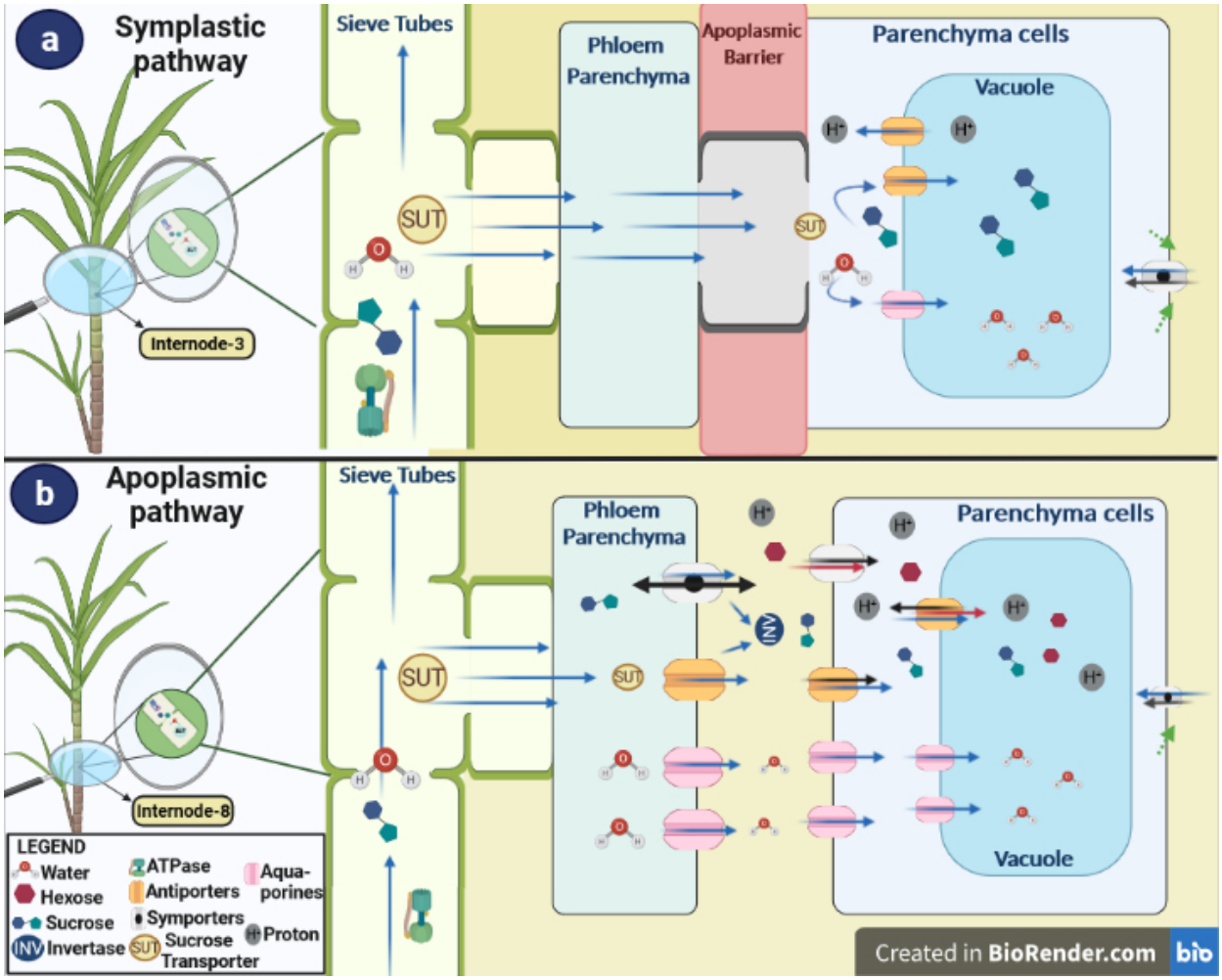

Figure 3: Schematic representation of sucrose storage and transport in sugarcane, separated into: (a) the symplastic, and (b) the apoplasmatic pathways.

In addition, SNrk1 kinase has been described as a regulatory gene for sucrose accumulation, inactivating several key metabolic enzymes and activating transcription of sucrose synthase and α-amylase (Polge and Thomas, 2007). Together with this kinase, the transcription factor bZIP11 is a transcriptional repressor in response to sucrose (Rahmani et al., 2009a). Enzymes of sucrose metabolism also play an important role in controlling the sucrose content in sugarcane stalks (Rae et al., 2005b). Likewise, sucrose synthase, which works by catalyzing the cleavage reaction of sucrose, degrades sucrose in the presence of UDP into UDP-glucose and fructose (Vargas and Salerno, 2010a). Another interesting group of proteins that are related to sucrose metabolism are invertases, which hydrolyses sucrose into glucose and fructose, and sucrose phosphate phosphatases, encoded by multiple gene

families (Lutfiyya et al., 2007).

Even though several genes were already described as part of accumulation sucrose processes in sugarcane, the entire set of biochemical reactions and related proteins is not fully understood. There is still a lack of knowledge into these biological mechanisms, which we intend to assess through RNA-Seq-based strategies.

2.4 RNA-Seq Experiments

In recent years, RNA sequencing through high-throughput, next-generation technologies (RNA-Seq) has proven to be a powerful tool for whole transcriptome profiling with enhanced sensitivity for the discovery of new transcripts and the examination of allele-specific expression profiles (Serin et al., 2016). RNA-Seq data is generally employed in order to provide comprehensive transcriptome resources (Cardoso-Silva et al., 2014; Vicentini et al., 2015). Through the library preparation process, it is possible to determine how closely cDNA sequence data reflect the original RNA population (Marguerat and Bähler, 2010). The most straightforward approach for library preparation is to synthesize double-stranded cDNA, to which adapters can be ligated for multiplexing (Cardoso-Silva et al., 2014).

The generation of reliable RNA-Seq data for biological inferences relies heavily on proper alignments of the sequencing reads to a coding genomic region obtained using a reference genome or assembled transcripts. Thus, it is necessary to perform several steps of quality filtering and quality control on all the obtained reads (Marguerat and Bähler, 2010; Gomathi et al., 2018) in order to produce reliable correspondences (Cardoso-Silva et al., 2014). The countable, almost digital, nature of RNA-Seq data makes them particularly attractive for the quantitative analysis of transcript expression levels (Gomathi et al., 2018).

Besides this quantitative aspect, RNA-Seq studies enable researchers to refine transcript annotations, providing for instance accurate maps of transcript start and end sites (Marguerat and Bähler, 2010). The annotation process of a transcript is the description of the related gene or protein in relation to its function performed in an organism, with evidence of sequence homology with previously annotated sequences and known functions (Koonin and

Galperin, 2003). These comparisons are performed using different biological databases, e.g. GenBank (Benson et al., 2000), KEGG (Kanehisa and Goto, 2000), UniProt (Consortium, 2016), Ensembl (Cunningham et al., 2014), Phytozome (Goodstein et al., 2011), PLAZA (Proost et al., 2014), and STRING (Szklarczyk et al., 2016). Such analyses are essentially the most commonly used methods for performing sequence annotations, but there are alternatives based on domain fusion, expression patterns between orthologs, and gene groups (Koonin and Galperin, 2003).

2.5 Gene coexpression Networks

Different computational methodologies can provide deeper insights into complex biological systems (Ficklin and Feltus, 2011a). In cases of grass species with poorly resolved, polyploid genomes such as sugarcane, genomic resources are not as far progressed as in other crops. A current efficient strategy is the modelling of such systems through complex networks, which are a set of elements organized in the form of nodes and connected through edges representing their relationship through some similarity or dissimilarity metric (Newman, 2003).

A large number of real systems can be converted to complex networks using mapping procedures (e.g. correlation measures) and are present in different areas of knowledge (Costa et al., 2011; Strogatz, 2001). There are several types of networks for modelling biological interactions, with metabolites, enzymes, proteins, and genes as vertices, connected based on some specific interaction (Costa et al., 2011). Complex networks have several factors that make their understanding difficult, such as the structure in which it is modeled, its evolution, its diversity of connections and nodes, and its dynamic complexity (Strogatz, 2001). However, massive analyses have been performed in different domains of knowledge (Boccaletti et al., 2006) and improved their comprehension, especially on its topology, organization, properties and possible applications to different areas. In the case of metabolic networks, for example, there are several associations that have been made between network topology and related biological functions (Jeong et al., 2000).

One method for identifying interacting gene sets is the construction of a gene coexpression network, modelled through the discovery of nonrandom gene-gene expression dependencies measured across multiple transcript perturbations (Ficklin and Feltus, 2011a). In these networks, there is a tendency of transcripts with similar expression patterns to be grouped into the same module, assisting the comprehension of related biological categories. In the case where the dependency is determined via correlation metrics, a correlation matrix can be easily used to represent expression similarities and model a network. Computational methods are then applied to circumscribe groups of network nodes that are highly connected (Langfelder and Horvath, 2008a) as observed on Figure 3.

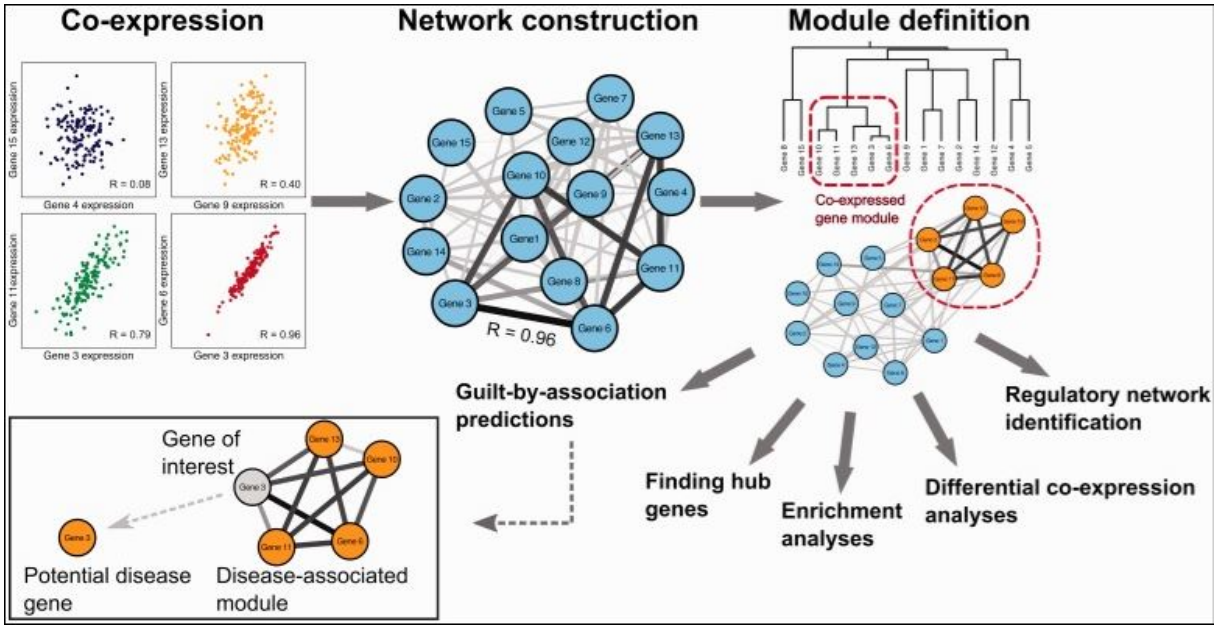

Figure 4: Workflow of a coexpression network analysis (Van Dam et al., 2018).

2.6 Complex Networks Analyses

Considering the process of modelling biological networks, different approaches may be applied in order to extract relevant information. Even though their comprehension represents a difficult task due to their dynamic complexity (Strogatz, 2001), this type of modeling enables the application of several techniques based on Graph Theory (Boccaletti et al., 2006). One of the most common methods in such structures to infer novel relationships

is the identification of association groups through community detection algorithms such as Louvain (Blondel et al., 2008) and Infomap (Rosvall and Bergstrom, 2007). These modules of a network are groups of vertices that share properties or perform a common function in the context of the network (Fortunato, 2010). With real applications, they can provide insights into the subset organization characteristics of a network and inferences about the hierarchical organization of these communities (Fortunato, 2010; Zhang et al., 2018). For example, the Louvain algorithm has been used to identify functional modules in coexpression networks for inferring gene relevance in SI-NETsSmall intestinal neuroendocrine tumors (Drozdov et al., 2011).

2.7 Biological Inferences through Coexpression Networks

Different methodologies for inferring novel gene relationships through coexpression networks have already been employed in several plant species, such as *Arabidopsis thaliana* (Feltus et al., 2013), *Populus trichocarpa* (Huang et al., 2020), *Zea mays* (Ficklin and Feltus, 2011a), *Chamaemelum mobile* (Tai et al., 2020), *Cicer arietinum* (Yadav et al., 2019), *Oryza sativa* (Childs et al., 2011), *Bombyx mandarina* (Zhou et al., 2020), and also humans (Liao et al., 2020a). The most common strategy applied for studying these biological systems is the construction of a coexpression network, usually employing a WGCNA methodology (Horvath, 2011), followed by an identification of functional modules in the network. The definition of these groups is generally based on a hierarchical clustering approach together with a cut-off point in the estimated dendrogram with the Dynamic Tree Cut algorithm (Langfelder et al., 2013). However, other methodologies have already been employed in more complex scenarios, such as the Infomap algorithm for detecting communities in breast cancer datasets (de Anda-Jáuregui et al., 2019) and the Link Communities method for clustering gene associations in cancer data (Ficklin et al., 2017). Besides, the hierarchical clustering strategy does not have the capability of identifying precise groups, being complemented by other approaches (Borin et al., 2018; Di et al., 2019; Zhong et al., 2020), such as the MCODE algorithm (Bader and Hogue, 2003).

The detection of cohesive clusters inside a network indicates the presence of genes with substantial coexpression (Uygun et al., 2016), representing a method for investigating the structure of molecular interactions (Mall et al., 2017a) and for identifying genes present in the same or associated metabolic pathways (Mitra et al., 2013a). Therefore, this identification of modules have been suggested as an effective method for inferring novel biological relationships among transcripts (Zhang and Yin, 2020) and can be used together with differential expression results, expanding the findings of a small set of genes to functional modules with associated topologies and connections (Poehlman et al., 2019).

The use of module detection inside a coexpression network can maximize the obtained information from gene expression datasets (Uygun et al., 2016). Besides, by using multiple and different statistical approaches for modelling this system, such analyses can allow the recognition of regulatory relationships (Saha et al., 2017). Even though the most used approach for defining these modules is hierarchical clustering, the introduction and use of more techniques, such as community detection methods, can increase the sensitivity and accuracy of the search for associated genes (Poehlman et al., 2019; Liao et al., 2020a; Huang et al., 2020; Tai et al., 2020).

Various metabolic pathways have already been studied with such methodologies, e.g. the metabolism of riboflavin, phenylpropanoid, biosynthesis of amino acids, isoflavonoids and indole alkaloid (Yadav et al., 2019); pathways related to response to stress (Fait et al., 2020); lipid biosynthesis (Poehlman et al., 2019); plant-pathogen interactions and the circadian rhythm (Tai et al., 2020); pathways related to the cell cycle (Liao et al., 2020a); photosynthesis (Zhang and Yin, 2020); and light reactions (Ficklin and Feltus, 2011a). All of these biological reactions could be complemented by coexpression network methodologies and module identification. Diverse important groups could be identified and several inferences could be performed. While investigating lipid biosynthesis metabolic pathways, for example, (Poehlman et al., 2019) found a set of genes probably related to lipid modulation through coexpression networks. (Tai et al., 2020) could identify modules related to plant-pathogen interactions, detecting sets of genes strongly related to circadian rhythm

pathway.

Recognizing the big challenge represented by sugarcane's genetic complexity, we used interdisciplinary approaches based on RNA-Seq and coexpression networks to access genes related to sucrose metabolism. With this, we seek to elucidate the sugarcane genetic networks controlling sucrose metabolism pathways.

3 Objectives

3.1 Main

Building coexpression networks as a tool to unravel potential genes and expression patterns related to sucrose metabolism of sugarcane.

3.2 Specific

- Building a global coexpression network for all experiments;
- Identifying communities in the networks containing key genes already associated with sucrose metabolic pathways and studying the role of the other genes found in these communities into these biological categories;
- Selecting modules related to sucrose categories for all varieties and species of sugarcane used in this work;
- Performing differential expression analyses among sugarcane cultivars and tissues based on sugar contrasting content;
- Performing an enrichment GO terms and pathway enrichment analyzed for selected modules related to sucrose metabolism;
- Performing additional differential gene expression analyses inside final selected modules;

- Identifying hubs and common connections for specific networks in order to analyze inferences.

4 Article

Gene coexpression networks provide deeper insights into sucrose accumulation in sugarcane cultivars

Abstract

Sugarcane is the most important crop for sugar and first-generation ethanol production, with the sucrose obtained from this plant being a fundamental alimentary and economic resource worldwide. This photoassimilate, a result of the photosynthesis, is transported through the phloem from source to sink organs to be stored in the parenchyma of the stems. Its metabolism and transport are regulated by a network of genes modulated during culm maturation, but the mechanisms involved in this process are not yet fully understood. Here, we constructed a global coexpression network using RNA-Seq data from different sugarcane genotypes and tissues. Additionally, a weighted gene coexpression network was modeled and potentially functional modules were identified through a complete hierarchical clustering approach. Inside the network structure, we could identify 153 different groups, which were contrasted with the differentially expressed genes identified. Through enrichment analyses and gene annotations, we could establish a total of 1,484 genes related to sucrose metabolism corresponding to 6 different network modules. This cohesive group presented relevant gene interactions with important biological processes, including response to stress, carbohydrate metabolic process, cellular carbohydrate metabolic process, cell wall organization, cellular polysaccharide metabolic process, and response to stimulus, representing an important source of data for understanding the dynamics of genetic activities on sugar accumulation. Through biological network strategies, our findings show the potential of coexpression networks to unravel new promising targets for additional studies on sugarcane sucrose metabolism and molecular breeding.

Introduction

Sugarcane is a worldwide important species for sugar and ethanol production, representing one of the main bioenergy sources available today (Chen et al., 2019b). Despite its importance, the comprehension of the genetic control mechanisms of sucrose accumulation remains slow and provides an incomplete and imperfect picture (Patrick et al., 2013; Cardoso-Silva

et al., 2014; Wang et al., 2017). This fact is mainly explained by the singular high genomic complexity of this crop, generated through its breeding process (Palhares et al., 2012). Modern sugarcane cultivars are a result of interspecific crosses between *Saccharum officinarum* (a species with high sugar production) and *Saccharum spontaneum* (a species highly tolerant to biotic and abiotic stresses), and consecutive backcrosses with *S. officinarum* (Cuadrado et al., 2004; Zhang et al., 2018). This hybridization caused the formation of a highly polyploid and aneuploid genome, with a size of approximately 10 Gb and abundant in repetitive elements, such as transposons, which can represent about 50% of all sugarcane sequences (D'Hont and Glaszmann, 2001; Piperidis et al., 2010; Mancini et al., 2018; Sforça et al., 2019).

Sucrose is one of the most representative sugars, because of its multiple activity on plants, including growth, development processes, cell division, cell differentiation and the accumulation of important metabolites (Chen et al., 2019b; Lastdrager et al., 2014; Mirajkar et al., 2016). In sugarcane, this carbohydrate is synthesized in the leaves, reaching the sink organs through the phloem for storage; its metabolism is regulated by a network of genes induced during culm maturation (Moore, 1995). The most important regulation reactions for sucrose metabolism in sugarcane have been described as performed by sucrose synthase (SS), sucrose phosphate synthase (SPS), soluble acid (SAI), neutral invertases (NI), sucrose transporters (SUTs), SWEET and diverse transcriptional factors (Rae et al., 2005a; Chen et al., 2012; Rahmani et al., 2009b; Mirajkar et al., 2016), where different levels of sucrose concentration represent an important factor in the development of mature and immature internodes in the culms of sugarcane (Bindon and Botha, 2002; Thirugnanasambandam et al., 2017).

One of the most common approaches for assisting the comprehension of sugarcane genes impacting sucrose metabolism is the use of RNA sequencing (RNA-Seq). In addition to providing the transcriptome profile of this crop (Marguerat and Bähler, 2010), these experiments have enabled the identification of new and rare transcripts directly or indirectly related to sucrose metabolism (Thirugnanasambandam et al., 2017; Kasirajan et al., 2018;

Chen et al., 2019b). In order to take advantage of the great volume of complex data that RNA-Seq experiments offer, network-based methodologies offer ways of integrating these data in a scalable and dynamic approach (Grabherr et al., 2011; Ferreira et al., 2016).

Gene coexpression networks are graph structures used to model how members of a gene set can be related to others, using nonrandom gene-gene expression dependencies measured across multiple transcriptome perturbations (Strogatz, 2001; Newman, 2003; Ficklin and Feltus, 2011b; Ferlic and Tracy, 2016). In such a network, there is a tendency of transcripts with similar expression patterns to be grouped in a coexpressed network module (Langfelder et al., 2008; Mall et al., 2017b), comprising groups of genes that are functionally cohesive, co-regulated or correspond to similar biological pathways (Mitra et al., 2013b). Such network categorization into groups of highly connected network nodes can enable the identification of novel gene associations (Langfelder et al., 2008; Zhang et al., 2018; Poehlman et al., 2019) which, in sugarcane, has shown promising results (Ferreira et al., 2016; Hosaka et al., 2021).

In this context, this study aims to identify a large number of key genes involved on sucrose metabolism and accumulation through RNA-Seq based gene coexpression networks. Firstly, we modeled a global weighted gene coexpression network using multiple RNA-Seq experiments encompassing three different tissues across nineteen contrasting sugarcane genotypes. Through differential expression analyses and bioinformatic approaches for categorizing sucrose related genes, we could assess coexpressed network modules with putative association with these processes. These groups of genes were used for novel inferences on such metabolic mechanisms by coupling known biological categories with novel associations, assessed through complex network approaches.

5 Material and Methods

5.1 Plant Material and Sequencing

This study comprises four independent RNA-Seq experiments (Exp1 to Exp4), which include root (Exp1), leaf (Exp2) and culm (Exp3 and Exp4) samples. For Exp1, several hybrid

cultivars (RB92579, RB855113, RB855536, SP79-1011, and SP80-3280) were planted in a greenhouse with three replicates. Plants were grown for 95 days and root samples were collected avoiding tiller roots, as described by Aono et al. (2021). For Exp2, in addition to hybrids (SP80-3280, RB72454, RB855156, TUC71-7 and US85-1008), *S. officinarum* (White Transparent and Criolla Rayada), *S. spontaneum* (IN84-58, IN84-88, Krakatau and SES205A), and *Saccharum robustum* (IJ76-318) genotypes were used. The middle section of top visible dewlap leaves were sampled from six-month-old sugarcane plants in the field in three replicates, as described by Correr et al. (2020). Exp3 included hybrids (F36-819, SP80-3280 and R570) and a *S. spontaneum* genotype (IN84-58). Culm samples from internode 1 were collected in four different time points, corresponding to 6-, 8-, 10- and 12-month-old plants in the field with three replicates (Hosaka et al., 2021). In Exp4, we used hybrid cultivars (SP80-3280 and IACSP93-3046), a *S. spontaneum* (Krakatau) and a *S. officinarum* genotype (Badilla De Java). Culm samples were collected from the internodes 3 and 8 of plants in the field in three replicates (Aono et al., 2021). For all experiments, total RNA was extracted from tissues with the RNeasy Plant Mini Kit (Qiagen, Valencia, CA, United States), quantified and and used for the preparation of cDNA libraries using the TruSeq Stranded mRNA Kit (Illumina, San Diego, USA), which were sequenced on an Illumina HiSeq 2500 platform.

5.2 Bioinformatics

RNA-Seq data quality was evaluated with FastQC (Wingett and Andrews, 2018) and clean sequence reads were obtained by removing adapter sequences and reads with low quality (Phred score below 20, reads shorter than 75 bp), and cutting the first 12 bases using Trimmomatic v.0.39 software (Bolger et al., 2014). SortMeRNA (Kopylova et al., 2012) was also used for removing residual rRNA. For gene quantification, we used *S. spontaneum* DNA coding sequences (CDS) (Zhang et al., 2018) and the Salmon v.1.1.0 software (Patro et al., 2015). CDS sequences were annotated with Phytozome v.13 database (Goodstein et al., 2012) considering only species from the Poaceae family and retrieving Gene Ontology (GO)

terms (Consortium, 2006) and KEGG pathways (Kanehisa and Goto, 2000).

5.3 Differential Expression Analyses

To identify differentially expressed genes (DEGs), we retrieved gene quantifications from the Salmon output using the tximport package (Soneson et al., 2015) in R software (R Core Team, 2021). Only the quantifications of the longest form of each CDS were used, considering a minimum quantity of 1 count per million. With the edgeR package (Robinson et al., 2010), DEGs were identified in each experiment by comparing genotypes with contrasting characteristics on sugar accumulation (Supplementary Table 1). For differential expression, we used a gene length corrected trimmed mean of M-values (TMMs) normalization (Smid et al., 2018), and considered an adjusted p-value of 0.05 according to the false discovery rate (FDR) correction together with a minimum absolute fold-change of 1.5. To evaluate enriched GO terms in groups of DEGs, we used the R package topGO (Alexa and Rahnenführer, 2009) with a Fisher's test (p-value of 0.01).

5.4 Global Weighted Gene Coexpression Network

We modeled a global weighted gene coexpression network (GWGCN) with the WGCNA R package (Langfelder and Horvath, 2008b). For building a scale-free GWGCN, we used pairwise Pearson correlations on transcripts per million (TPM) values from Salmon output considering a power function in order to fit a scale-free independence. For that, we estimated a soft threshold power value with a generated scale-free topology model fit $R^2 > 0.8$ and the largest mean connectivity possible. Genes with zero variance across quantifications were excluded using the WGCNA package. We defined the functional modules using an average linkage hierarchical clustering with euclidean distances together with an adaptive branch pruning implemented in cutreeDynamic R package (Langfelder et al., 2008).

Selected modules were analysed for enriched GO terms in the same way as DEG analyses, and for enriched KEGG pathways, we used a Fisher's test (p-value of 0.05) implemented in

R software (Team et al., 2013). For graphical inspections, we used the ggplot2 R package (Wickham, 2011).

5.5 Gene Coexpression Networks' Comparisons

Using the genes from the modules selected, we calculated pairwise Pearson correlations to estimate intra-modular co-expression, discarding correlations with an absolute coefficient below 0.8. All these measures were used for modeling specific networks for genotypes of *S. spontaneum*, *S. officinarum* and the SP80-3280 hybrid cultivar, which was present on all experiments. We employed a highest reciprocal rank (HRR) approach (Mutwil et al., 2010) for retrieving the strongest edges limited to 30 connections. The obtained network had its connections codified as 1/5 (top 10), 1/15 (top 20), and 1/30 (top 30).

We used R statistical software for network modeling together with igraph package (Csardi et al., 2006). For each network, we evaluated network hubs considering Kleinberg's hub centrality score (Kleinberg et al., 2011) and compared networks' connections through igraph.

6 Results

6.1 Bioinformatics

Following the pipeline established for filtering the RNA-Seq data across the 234 samples, and quantifying the 35,519 *S. spontaneum* CDSs, we could generate a large reservoir of sugarcane expression data, encompassing a wide range of different genotypes and tissues. All of these CDSs were aligned against the Phytozome database, and we could find positive correspondences on 21,828 sequences (Supplementary Table 2).

To perform the DGE analyses, we separated the raw gene count data according to the experiments performed, normalizing the subsets by TMMs and identifying DEGs for each condition considering the filters established (Supplementary Table 3). Comparing genotypes with contrasting phenotypes on sugar accumulation led to the retrieval of different num-

bers of DEGs, ranging from 34 in Exp2 (IN84-58 vs RB72454) to 3,723 in Exp3 (IN84-58 vs F36-819 on time point 4). For each one of the comparisons established, we selected the enriched GO terms (Supplementary Table 2) and evaluated the presence of the following terms: (i) carbohydrate biosynthetic process; (ii) trehalose biosynthetic process; (iii) trehalose metabolic process; (iv) disaccharide biosynthetic process; (v) oligosaccharide biosynthetic process; (vi) polysaccharide metabolic process; (vii) monosaccharide metabolic process; (viii) cellular polysaccharide biosynthetic process; and (ix) carbohydrate derivative biosynthetic process. As these terms represent the major categories associated with sucrose metabolism, differential expression results obtained on genotypes without any of these terms were discarded, leading to the retention of 18 and removal of 17 comparisons (Supplementary Table 3). All comparisons from Exp1 were discarded. From Exp2, we discarded SES205A vs RB72454, IN84-58 vs White Transparent, and IN84-58 vs RB72454; from Exp3 IN84-58 vs SP80-3280 on time point 2, SP80-3280 vs R570 on time point 2, IN84-58 vs F36-819 on time point 2, SP80-3280 vs R570 on time point 4, and F36-819 vs R570 on time point 4; and from Exp4 Krakatau vs SP80-3280 on internodes 3 and 8, Krakatau vs IACSP93-3046 on internodes 3 and 8, BadilaDeJava vs Krakatau on internode 3, and Krakatau vs SP80-3280 on internode 3.

6.2 GWGCN

Quantifications with zero variance were excluded from the initial dataset; consequently, the GWGCN was modeled with 34,085 genes. In order to fit the network into a scale-free topology, we selected a soft threshold power of 7 ($R^2 \cong 0.81$ and mean connectivity of about 164) and calculated the topological overlap-based dissimilarity matrix, which was used for estimating putative functional modules within the network structure (Supplementary Table 4). We defined 153 groups with sizes ranging from 53 to 4,182 genes, which had a mean size of 223.

The functional modules in the network putatively associated with sucrose accumulation were selected based on the presence of DEGs within the network structure. From each

differential expression comparison, we selected the network groups containing these DEGs. However, only the groups enriched for the GO terms indicated as associated with sucrose metabolism (see section above) were selected. From this approach, we could establish 13 groups containing DEGs obtained across the 18 retained differential expression analyses and also identify important GO terms enriched (Supplementary Tables 5-6).

In order to restrict the quantity of genes observed and supply narrow associations into sucrose-associated processes, we performed an additional step of restricting the set of 13 modules to only four (3, 87, 17 and 93). The other nine groups were excluded because they only contained DEGs from the comparisons enriched in sucrose-related GO terms of one experiment; by contrast, the remaining four groups had DEGs related to comparisons enriched in at least two experiments. Finally, for each WGCNA module we tested the enrichment of the following KEGG pathways: (i) starch and sucrose metabolism and (ii) fructose and mannose metabolism. As only three modules were enriched for these pathways (76, 84 and 93), we included these groups in the modules selected through the filters established for differential expression analyses.

Ultimately, we generated a final set of genes composed of 1,689 elements and separated into six different modules: 3 (801 genes), 17 (337), 76 (155), 84 (144), 87 (128), and 93 (124). The mean proportion of DEGs in each group was 40%, with DEGs spread according to the three experiments performed (Supplementary Table 7). This final set was then selected to supply novel insights into sugarcane sucrose accumulation processes.

6.3 Functional Modules' Characterization

We performed a functional characterization using the final set of six modules selected (Supplementary Table 8). We found a great quantity of genes encoding transporters related with sucrose metabolism (21 genes), including sucrose transporter 2, sugar transporter 1, sugar transporter protein 7, monosaccharide transporter 5, monosaccharide transporter 1, a nucleotide-sugar transporter family protein, and a major facilitator superfamily protein. Likewise, we found several proteins involved in sucrose metabolism, e.g.,

trehalose-phosphatase synthase, glucose-6-phosphate, sucrose synthase, sucrose phosphate synthase, carbohydrate binding protein, glucan endo-1,3-beta-glucosidase, beta glucosidase, a nucleotide-diphospho-sugar transferase family protein, sugar isomerase, phosphofructokinase, and glucose-6-phosphate dehydrogenase.

Regarding the gene ontology categories, we observed important categories on sugar accumulation present among the genes of the selected groups, including the metabolic processes of carbohydrates, glucose and sucrose, the biosynthetic process of cellulose, and sucrose and sugar transport (Figure 5C). Some of these terms were also enriched when compared to the entire set of GO terms within the transcriptome (Supplementary Table 9), with the most pronounced terms being the response to oxidative stress, oxidation-reduction process, response to stress, carbohydrate metabolic process, and cell wall biogenesis.

By selecting the KEGG pathways related to at least five genes (Fig. 5A) of these groups, we found starch and sucrose metabolism, glycolysis/gluconeogenesis, fructose and mannose metabolism, and amino sugar and nucleotide sugar metabolism among the identified pathways. Interestingly, the most pronounced pathway was phenylpropanoid biosynthesis. By organizing these enzymes into categories (kinases, sugar enzymes, transcription factors, and transporters), it is also possible to notice their presence in almost all the GWGCN modules selected (Fig. 5B).

Considering all the genes within this group, we could evaluate their expression as up- and down-regulated taking into account the differential expression analyses performed (Supplementary Table 10). Although several genes did not present a uniform behavior among the tests established, we could identify 105 genes as up-regulated in genotypes with higher sucrose content and 78 genes as down-regulated, encompassing different functional profiles assessed through the associated enzymes.

6.4 Gene Coexpression Networks' Comparisons

In order to assess the functional specificities on *S. spontaneum*, *S. officinarum* and the SP80-3280 hybrid network of genes, we created individual networks for each genotype considering

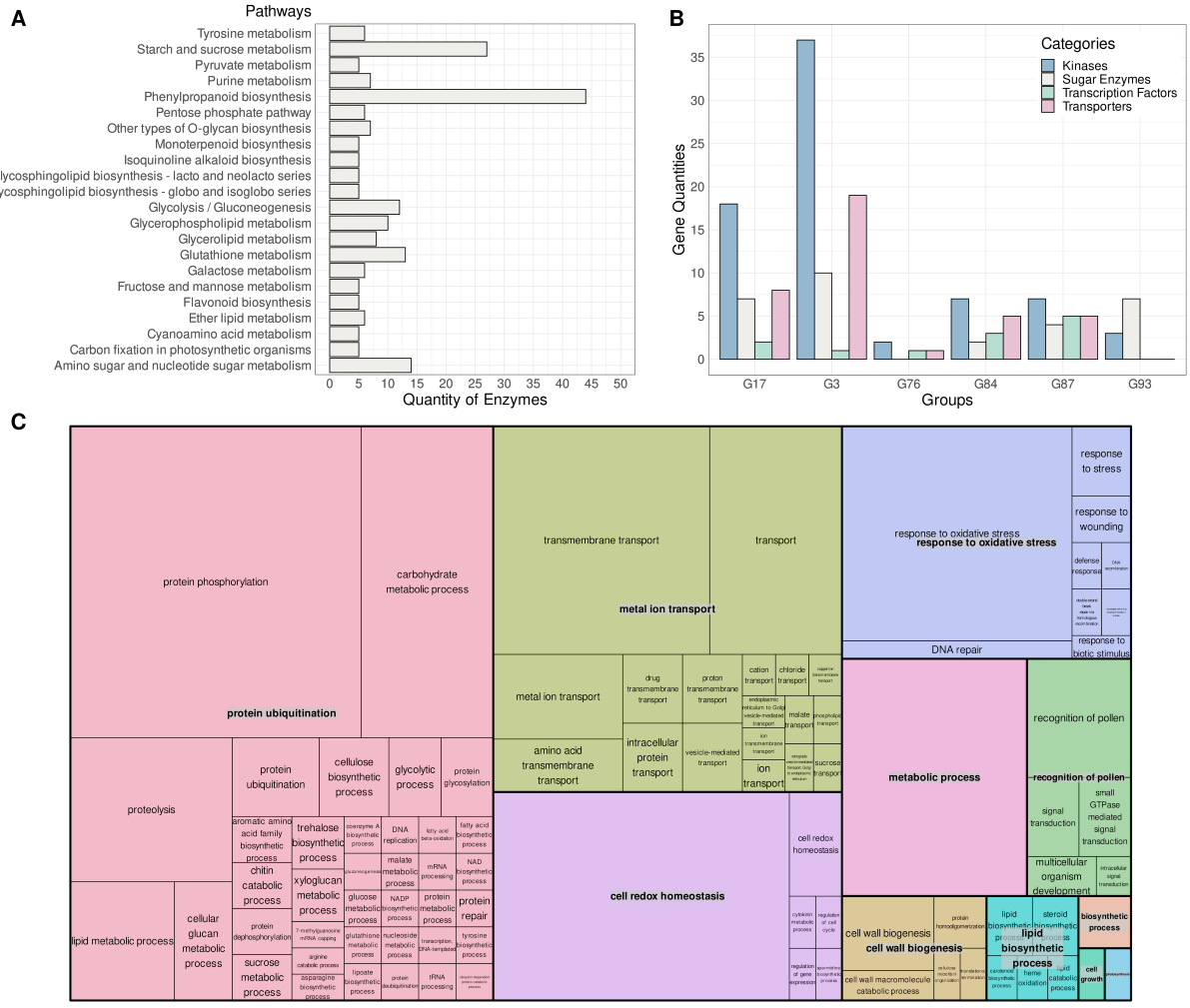

Figure 5: Functional profile of genes within the groups selected as associated with sucrose accumulation in the global weighted gene coexpression network modeled: (A) Quantity of genes associated with KEGG metabolic pathways (minimum quantity of five to be included); (B) Quantity of genes per group (G) classified according to the established categories; and (C) Gene Ontology (GO) terms found for the genes selected and summarized into a treemap.

the selected group of genes established as associated with sucrose accumulation (Fig. 6). We could observe a distinct profile on the network topologies, with *S. officinarum* corresponding to a cohesive group of gene interactions. Interestingly, the *S. spontaneum* and SP80-3280 networks were characterized by sparse node connections, except for a few specific subgroups of genes.

Such differences can be corroborated with the hub scores calculated for each network (Supplementary Tables 11-13). For *S. spontaneum*, we could find only one gene densely

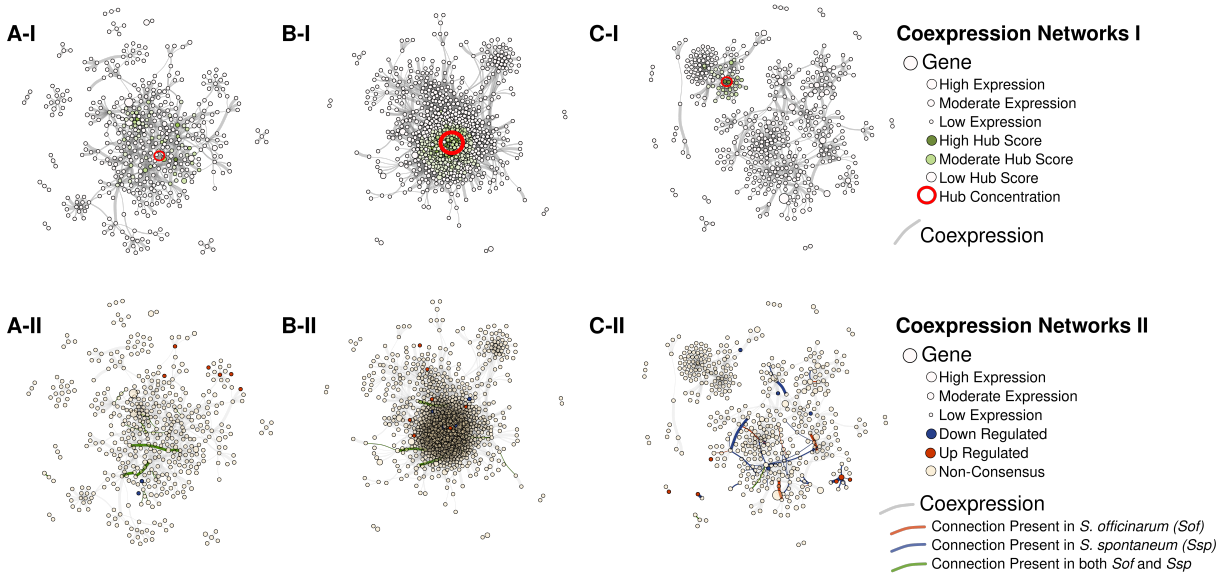

Figure 6: Gene coexpression networks modeled with the groups of genes associated with sucrose accumulation for (A) *Saccharum spontaneum*; (B) *Saccharum officinarum*; and (C) sugarcane hybrid SP80-3280. In (I) node colors are selected based on hub scores calculated (Supplementary Tables 11-13); and in (II) according to up- and down-regulated conditions found (Supplementary Table 10).

connected (hub score larger than 0.8), contrasted to seven in *S. officinarum* and two in SP80-3280. Evaluating the connections in common among the three networks, 16 connections were found in common between *S. officinarum* and *S. spontaneum* (Supplementary Table 14), 35 between SP80-3280 and *S. spontaneum*, and 10 between SP80-3280 and *S. officinarum* (Supplementary Table 15).

7 Discussion

Modern sugarcane cultivars are known by their complex genome (Cuadrado et al., 2004; Mancini et al., 2018; Souza et al., 2019) embracing a variable number of chromosomes (Cuadrado et al., 2004; Piperidis et al., 2010; Zhang et al., 2018), a considerable high genome size (Sforça et al., 2019), and the common presence of duplicated regions (Palhares et al., 2012; Mancini et al., 2018). This singular crop complexity forces sugarcane researchers to employ many multidisciplinary efforts to understand the molecular basis of complex traits

(Wang et al., 2017; Chen et al., 2019b). These approaches take advantage of comparing sugarcane genotypes with contrast regarding relevant characteristics (Wang et al., 2017; Chen et al., 2019b), e.g., the ability to store high levels of sucrose in culm tissues (Casu et al., 2015b; Pereira et al., 2017; Thirugnanasambandam et al., 2017), elevated fiber content (Vicentini et al., 2015; Kasirajan et al., 2018), and biotic/abiotic stress tolerance (Schaker et al., 2016; Selvarajan et al., 2018; Wu et al., 2018). Although all these characteristics are important for sugarcane production, the major interest in this crop is its sugar content, which needs to be increased for reaching the worldwide demand (Patrick et al., 2013; Pereira-Santana et al., 2017; Chen et al., 2019b; FAO, 2020). In this sense, our work was developed for providing molecular resources associated with sucrose metabolism, providing insights into specific sugarcane mechanisms for storing sugar.

Despite the economic importance of the sugar produced by sugarcane, the molecular genetic mechanisms that comprise the metabolism of sucrose are not fully understood, and the establishment of specific targets for molecular breeding is hindered. Therefore, there is a need to find scalable, dynamic and interdisciplinary approaches that could be applied in an efficient way into sugarcane research (Watt et al., 2005). For that, we opted to use RNA-Seq-based gene coexpression networks with a wide variety of relevant sugarcane genotypes, embracing several commercial cultivars and diverse accessions of interest for breeding and research. From such structures, functional inferences could be performed based on gene relationships measured through correlation metrics (Langfelder et al., 2008; Ballouz et al., 2015; Ferreira et al., 2016). Although the annotation process can indicate sets of genes related to their putative molecular functions (Marguerat and Bähler, 2010; Koonin and Galperin, 2013), several sugarcane sequences could not be fully characterized in our dataset, and the identification of network modules represents a powerful tool for depicting novel gene relationships and inferring unknown gene roles (Ferreira et al., 2016; Zhang et al., 2018; Poehlman et al., 2019).

Differential expression analyses between contrasting sugarcane genotypes have already provided important inferences for specific sugarcane traits (Correr et al., 2020; Hosaka et al.,

2021). Using such an approach, we performed several comparisons among genotypes and tissues with contrasting sugar accumulation, forming a set of DEGs with high potential to be involved in sucrose metabolism, which were then evaluated considering the network structure and the functional modules established. By using different filtering criteria, we formed a conservative group of genes with potential participation in sucrose-related mechanisms. In addition to being functionally enriched for sucrose related categories, this set of modules was used as a predictable structure for complementing our comprehension into the multiple processes related to sucrose storage, metabolism and transport.

7.1 Sugarcane Sucrose Metabolism

Even though sucrose is present in all plant tissues and cell environments, this carbohydrate is more abundant in specific tissues, specially in sink and source tissues (Casu et al., 2015b; Thirugnanasambandam et al., 2017; Verma et al., 2019). From the complex set of molecular mechanisms related to sucrose metabolism, two main sets of metabolic reactions can be established, based on enzymatic interactions and final sugar products of each group (Fig. 7). The metabolic reactions in first group (a) involve invertases, leading to the synthesis of fructose and glucose; the reactions in second group (b) involve sucrose synthases, which lead to the synthesis of fructose and UDP-glucose. The importance of the enzymes in (a) and (b) is fundamental in the metabolism of sucrose, and, as expected, we could find a significant subset of these proteins in the modules selected.

In (a), there is a primordial group of enzymes known as invertases (Vargas and Salerno, 2010b; Verma et al., 2019) such as the beta-fructofuranosidase, which hydrolyzes sucrose into fructose and glucose (Wan et al., 2018). We only found the incidence of an invertase in one of the selected modules (module 3). Although this associated gene was not differentially expressed, its inclusion on the module corroborates the multiple roles of the selected group of genes found by the GO terms identified (Mirajkar et al., 2016), potentially indicating processes related to partition and storage of sucrose (Wang et al., 2019).

Fructose derived from invertase reactions can be phosphorylated by enzymes such as

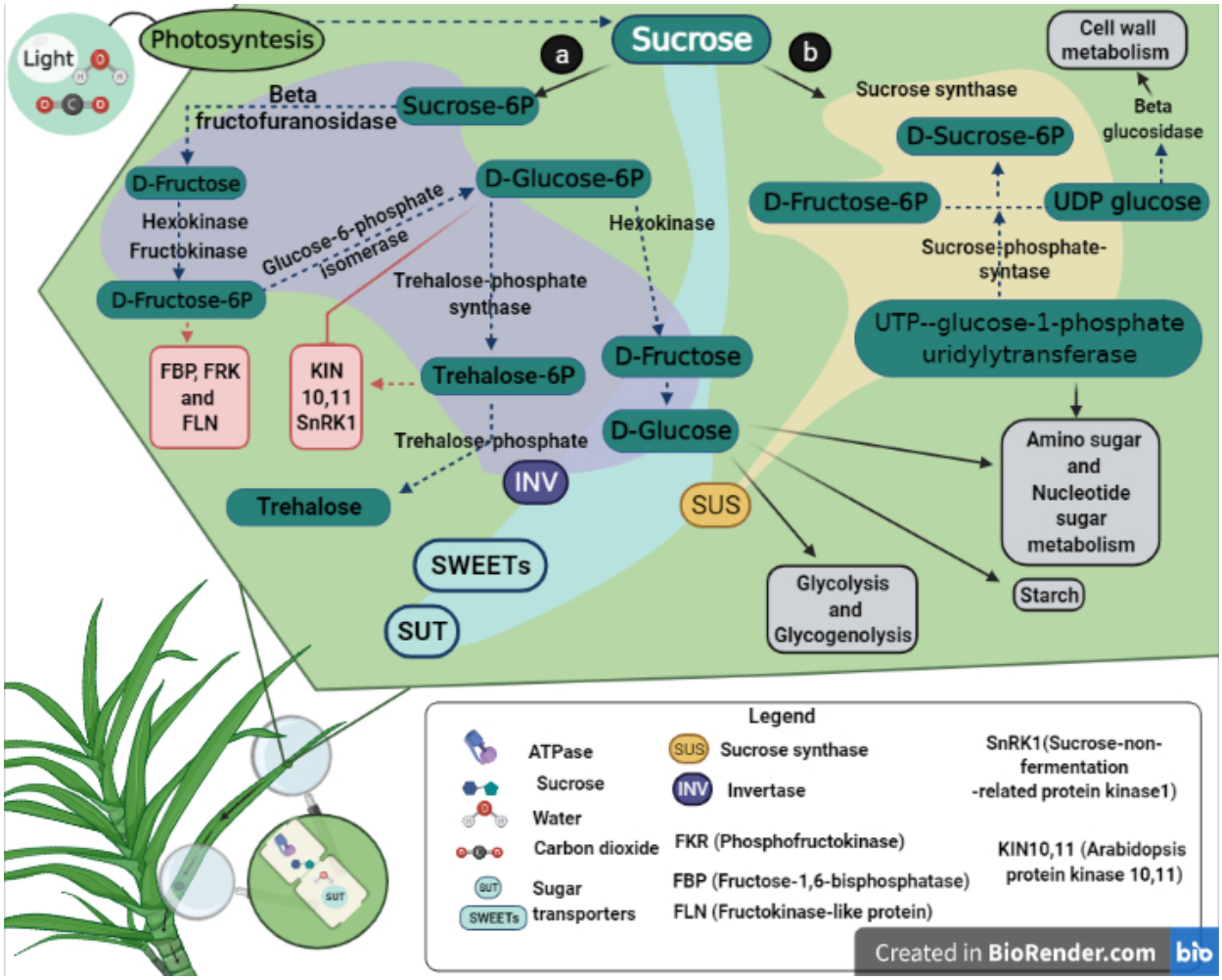

Figure 7: Metabolic reactions associated with sucrose metabolism and separated into (a) and (b) according to the enzyme activities.

fructokinase and hexokinase. These kinases act as signals for the activation of a big energy complex, encompassing the fructokinase-like 1 (FL1), phosphofructokinase (FKR) and fructose-1,6-bisphosphatase (FBP) enzymes (Wang et al., 2019; Stein and Granot, 2018). This important complex is part of many biological processes, including the generation of precursor metabolites and energy process, carbohydrate catabolic and metabolic processes, organic substance biosynthesis and the small molecule metabolism, which were part of the criteria for selecting the functional modules enriched for categories associated with sucrose metabolism. The presence of a phosphofructokinase was observed in the selected modules as an up-regulated DEG, being enriched on genotypes with high sugar content, which is in accordance with its regulation aspect on gluconeogenesis (Wang et al., 2019; Stein and

Granot, 2018).

Furthermore, fructose-6-phosphate by a reversible isomerization by glucose-6-phosphate isomerase can produce a glucose-6-phosphate (Stein and Granot, 2018), which is essential for the carbohydrate catabolic and biosynthesis, carbohydrate phosphorylation, hexose catabolic, and ATP biosynthetic processes, possessing a primordial molecular function of glucose binding and acting as an activator of other pathways. Additionally, the glucose-6-phosphate together with UDP-glucose and trehalose-6-phosphate synthase synthesize trehalose-6-phosphate, an enzyme with a vital function on maintaining the sucrose concentration in a homeostatic way throughout the plant (Hu et al., 2020), coordinating the synthesis of sucrose on source organs and regulating its content on sink organs (Ponnu et al., 2011; Figueroa and Lunn, 2016; Fichtner and Lunn, 2021; Wu and Birch, 2010). Trehalose-phosphate phosphatase is further dephosphorylated into trehalose, forming a sucrose signaling complex also linked with growth and plant development (Ponnu et al., 2011; Lunn et al., 2014). Interestingly, we found gene representations of this enzyme throughout the module established, being DEGs with a more pronounced expression in low-sugar, high-fiber genotypes. This can be seen as a negative or inhibition mechanism with fundamental roles in the sucrose accumulation (Figueroa and Lunn, 2016; Fichtner and Lunn, 2021).

Futhermore, trehalose 6-phosphate acts as an inhibitor of sucrose-non-fermenting-1-related protein kinase 1 (SnRK1), a central integrator of energy signaling (Nukarinen et al., 2016; Wurzinger et al., 2018; Li et al., 2017). SnRK1 is indicated as an activator of sucrose synthase (SUS) breakage activity while down-regulating sucrose phosphate synthase (Wu and Birch, 2010; Figueroa and Lunn, 2016). Although we did not find SnRK1 on the modules selected, it was found on other network modules (4, 6 and 55), indicating a joint mechanism of repression that still need to be investigated for further studies.

In (b) (Fig. 7), there is a primordial group of enzymes, including sucrose synthase (SUS), a key enzyme on the synthesis and breakdown of sucrose in sugarcane, with functional influence in sucrose utilization to support growth (Verma et al., 2011; Thirugnanasambandam et al., 2019). We could find SUS-related genes in the selected groups (module 3). Inter-

estingly, we did not observe differential expression of SUS genes on the sugar-contrasting genotypes, which corroborates previous reports indicating these gene's moderate expression levels (Thirugnanasambandam et al., 2019). The SUS sucrolytic activity, cleaving the UDP-glucose and fructose-6-phosphate enzymes, has a reversible behavior because SUS keeps the glycosidic link catalyzed by the sucrose phosphate synthase (SPS) (Anur et al., 2020; Verma et al., 2019; ElSayed et al., 2017; Patrick et al., 2013; Stein and Granot, 2018; Verma et al., 2011; Kurniah et al., 2021). SPS is mainly reported as involved in the synthesis of sucrose on leaves and capable of controlling sucrose content (Verma et al., 2011; Anur et al., 2020; Kurniah et al., 2021). We observed this enzyme on module 17, but it was not diferentially expressed, which might be a result of its multiple role on several simultaneous sucrose processes, including accumulation, transport, and synthesis. For that reason, there are several patterns of expression for the same enzyme at the entire plant, which are influenced by other biological processes, such as the circadian rhythm and the light assimilation process on each cultivar (Anur et al., 2020; Ma et al., 2020).

All sucrose enzymes described on (b) interact to form UTP-glucose-1-phosphate uridylyltransferase, which catalyzes the interconversion of glucose and fructose, acting as a signal for the production of other nucleotide-sugars and being a primordial glycosyl contributor for polysaccharides biosynthesis (Hosaka et al., 2021; Wang et al., 2019). This enzyme was found in other groups of the network (10, 35, 115), indicating a putative functional connection of these modules that should be a target for further investigations. Additionally, another important enzyme in (b) is beta glucosidase, which is involved in the conversion of glucose into cellulose (Cao et al., 2015). This enzyme was found among the selected modules and according to its retrieved GO terms, it may act as a signal for the beginning of cell wall polysaccharide metabolic process, cellular glucan metabolic process and carbohydrate metabolic process. However, this enzyme was not found among the DEGs, which could be explained by its expression profile, which is closely related to specific tissue expression and extremely important for sucrose accumulation in sugarcane (Qin et al., 2021; Li et al., 2017).

All these enzymes described in (a) and (b) sets had an indisputable affinity with several types of sugar, including sucrose, glucose and fructose (Zhang et al., 2020; ElSayed et al., 2017). This set of proteins represent the known elements into sucrose metabolism in sugarcane and, as expected, could be almost entirely reconstructed by the modules selected in this work, clearly corroborating the importance of the set of genes established throughout the manuscript. Additionally, the final products of these enzymatic reactions described above are transported to the entire plant in order to support several vital processes, which was also identified as key elements in the module selection through the definition of monosaccharide, sucrose, and sugar transporters. These transportes are essential for sucrose transport and control in sugarcane (Chen et al., 2015; Zhang et al., 2020; Khan, Chen, Zeng, Qin, Guo, Mahmood, Yang, Liang, Song, Xing, et al., Khan et al.; Correr et al., 2020), having high affinity with organelle membranes, such as vacuolar membrane, and are integral components of plasma membranes, ratifying the associated GO terms found for these transporters, including proton transmembrane transport, proton-exporting ATPase, sucrose metabolic process, sucrose transports, and phosphorylative mechanism activity (Rae et al., 2005a; Zhang et al., 2020; ElSayed et al., 2017).

A high abundance of transporters is generally linked to the developmental stages on tissues differentiation of sugarcane (Perlo et al., 2020), which is in agreement with the great amount of transporters observed in the selected modules, which were most up-regulated for high-sugar content genotypes. This abundance of transporters corroborates the propagation of this intense enzymatic activity across the network. Additionally, some types of transporters showed notable distribution on the selected modules, including phosphate transporter, potassium and sodium symporters, and cationic amino acid transporters. These transporters are responsible by crucial ion homeostasis processes on sugarcane, and previous studies have already indicated an association between the differential accumulation of Na^+ and K^+ ions in contrasting sugarcane genotypes (Brindha et al., 2021). The transport protein particle (TRAPP) was also observed within the selected groups, being fundamental for exocytosis, endocytosis, and protein sorting, and associated to GTPases that may be involved on the operation of plant membrane traffic (Kalde et al., 2019). This is in agreement

with some of the GO terms found (vesicle-mediated transport and small GTPase mediated signal transduction), and also directly linked to the sucrose metabolism GO category. From that, we may speculate that TRAPP might be involved in the transmembrane transport of several compounds in sugarcane, including sucrose.

7.2 Novel Inferences

The selected modules related to sucrose metabolism are the axis of our approach. In addition to the described direct relationship of these genes into the sucrose metabolic mechanisms, we also observed a large set of genes encoding enzymes that are part of diverse biological processes in plants, which could be closely related to sugar accumulation process in modern sugarcane cultivars. In a few of the selected modules (76 and 17), we identified DEGs that encode proteins containing pentatricopeptide repeat (PPR) and tetratricopeptide repeat (TPR), which have been described to influence transcription, transport, folding, control cycle control and organelle RNA edit in other plants (Sun et al., 2018; Wei and Han, 2017). These regulation proteins have direct effects on plant development and photosynthesis (Barkan and Small, 2014), and their inclusion in the established groups may indicate a different source of sucrose regulation in sugarcane, with high potential to be a valuable source for biotechnological studies.

A noteworthy group of genes involved in plant defense was observed on some modules (84, 3 and 17), which included: a mitogen-activated protein kinase, which is crucial for establishing disease resistance (Bi et al., 2018); an AAA-type ATPase family protein, which are involved on specific defense-related programmed cell death in rice (Fekih et al., 2015); a target of rapamycin, involved in early pathogen recognition pathway in sugarcane (Oloriz et al., 2021); a GTP-binding family protein strongly involved in plant defense in *Arabidopsis* (Maruta et al., 2021); and a NBS-LRR protein, involved in response to disease resistance in plants, being capable to identify compounds secreted by pathogens during infection (Singh and Singh, 2018). Interestingly, these genes encoding such proteins did not present an up-regulation in expression in *S. spontaneum* genotypes. It is clearly observed that sugarcane

has a very complex defense system, directly influenced by the sucrose associated pathways and which needs deeper investigations for unravelling how wild and modern sugarcane genotypes have different mechanisms for dealing with such combined processes.

Other important groups of genes involved in plant development were observed on modules 76 and 3, and especially on module 84 (Supplementary Table 8). This is the case of a Gibberellin 2-oxidase 8, which plays a direct role in determining the levels of bioactive gibberellins, hormones related to plant growth and development (Bai et al., 2014; Hu et al., 2017); and a U2 snRNP auxiliary factor, which is fundamental for the splicing of key flowering genes in *Arabidopsis* (Wang et al., 2020). Other examples include: (i) the Regulator of Chromatin Condensation 1 (RCC1) family protein, which shows an opposite response to abcisic acid related to early seeding growth in *Arabidopsis* (Ji et al., 2019); an ethylene-responsive element binding factor related to a fundamental role in plant growth, development, metabolism, and responses to biotic and abiotic stresses (Chen et al., 2019b; Li et al., 2020); and a senescence-associated gene described as induced under abiotic stress conditions (Carrillo-Bermejo et al., 2020). Being so tightly associated with sucrose metabolic processes, these enzymes might represent important factors linking sucrose metabolism to developmental processes long known to interfere with sugar storage in sugarcane, including growth, flowering, abiotic stresses (such as cold) and senescence (Yoon et al., 2021).

Finally, a diferentially expressed WRKY DNA-binding protein has fundamental roles of developmental and physiological processes in *S. spontaneum* (Li et al., 2020), probably regulating the abundance of sucrose (Rushton et al., 2010), and PLATZ transcription factor family protein was also described as a positively regulator of starch synthesis (Li et al., 2021). From the different genes found as belonging to these groups established, we could not only provide groups of genes corresponding to and controlling this metabolic route, but also elucidate different regulation mechanisms affecting other sugarcane characteristics.

Additionally, we identified up and down-regulated genes associated with sucrose metabolism inside the selected modules. Leucine-rich repeat (LRR) protein genes were found up-regulated on at least two experiments from different tissues. This corroborates previous

indications that these proteins are associated with the synthesis of sucrose in several sugarcane tissues and genotypes (Vicentini et al., 2009). Also, we observed calmodulin domain protein gene as up-regulated on genotypes with higher sugar accumulation; these proteins are sensors of calcium and were reported to interlink calcium and sugar signaling in plants. In mutant wheat lines, for instance, it showed important correlation with water-soluble carbohydrate as sucrose-1-fructosyltransferase (Kalaipandian et al., 2019). Additionally, calcium proteins are involved in drought tolerance pathways in sugarcane (Selvi et al., 2021).

Furthermore, MYB family transcription factor genes were up-regulated. These transcription factors are involved in sugar and hormone-regulated alfa-amylase gene expression, also encoding a monosaccharide transporter on rice (Zhang et al., 2010). In *Arabidopsis*, it was observed that a mutant MYB gene showed an increased expression of a hexokinase gene, which is related to the response to glucose (Chen et al., 2017); this suggests that the mutant gene was expressed in the anther tapetal cells and in sugar-transporting vascular tissues (Zhang et al., 2010). Additionally, glucosyltransferases were observed to be up- and down-regulated in relation to sucrose metabolism. They are known to transfer sugar and several acceptor molecules; a mutant, overexpressed glucosyltransferase gene increased glucose tolerance on *Arabidopsis* (Yan et al., 2010). Lastly, we observed an up-regulated glucan endo-1,3-beta-glucosidase on module 3 with their GO terms enrichment for sucrose metabolism, it was previously associated with response to drought stress in wheat, showing major expression on the leaves and roots of tolerant genotypes (Faghani et al., 2015). As expected of differentially expressed genes, all of them are strongly related with sucrose metabolism according to the inferences. This means that these genes could act as markers for the sucrose pathway in sugarcane, because of their interactions that are associated. While most of these genes were not known to act directly in this metabolic pathway, our inferences allowed us to identify their links to sucrose metabolism.

While accessing functional specificities of the networks modeled, we observed that two hubs in the *S. officinarum* network showed major cohesive behavior (Supplementary Table 12). One of them is a bZIP transcription factor previously reported to be specifically

repressed by sucrose on a reversible way (Kang et al., 2010; Thalor et al., 2012) and named as a sugar sensing and signaling system led by sucrose in *Arabidopsis* (Rook et al., 1998). The other one is CPR5, responsible for the positive modulation of growth under stress conditions and acting as an inhibitor of bZIP (Meng et al., 2017). However, in the SP80-3280 network, we observed only one hub (Supplementary Table 11), which encodes a basic helix-loop-helix (bHLH), part of the family of transcription factors that may have different roles in plant cell, tissue development and plant metabolism (Heim et al., 2003). Also, basic helix-loop-helix (bHLH) were reported to act as regulators in the control of ethylene synthesis in sugarcane, with its abundance being influenced by photoperiods and ripening (Alessio et al., 2018). Furthermore, Basic helix-loop-helix (bHLH) proteins were a key piece to show the direct link between the phenylpropanoid metabolism and sucrose. In *Solanum tuberosum* under on a treatment induced by sucrose, bHLH showed higher expression and was correlated with a greater presence of enzymes involved in the metabolism of sucrose, such as invertases and sucrose synthases (Payyavula et al., 2013).The relationship between these both pathways could explain why we observed large amount of genes on phenylpropanoid pathway on selected groups (Fig. 5A). Furthermore, hubs were used to access genes densely connected by the inferences; with this, we identified some links between sucrose metabolism and other pathways, helping us to understand what is happening on network structure. Using the highest hub score and annotations, we identified interesting genes described above. These functional specifies complement the novel inferences described on first part, because specific relationships were visualized, such as the case of SP80-3280 network supporting the fact that sucrose metabolism at plant is a connected network of reactions.

Additionally, common connections were found between the *S. spontaneum* and *S. officinarum* networks. Some of them showed biological significance, such as a connection between Ribosomal protein S4 and Ribosomal protein L6, both of which have important roles in the translation of key proteins involved in chloroplast development, photosynthesis and response to salt stress tolerance in rice (Qiu et al., 2018; Moin et al., 2021). Another connection can be established between the ribosomal protein S9 and quiescin-sulfhydryl oxidase 1. The latter was described as a novel regulator of ion homeostasis on a mechanism that consists

in the activation of root systems loading K^+ into xylem on mutant *Arabidopsis* (Alejandro et al., 2007). Interesting common connections were also found between the SP80-3280, *S. spontaneum* and *S. officinarum* networks. This is the case of the UDP-glucosyl transferase 72B3 with major facilitator superfamily proteins, which are involved in the transport sugar and several acceptor molecules (Deepak et al., 2010) and in loading sucrose into the phloem and sink tissues on sugarcane (Reinders et al., 2006). Likewise, another interesting connection between ribosomal protein L6 with hydroxyproline-rich glycoprotein family both involved on abiotic and biotic stress response, such as salt stress tolerance (Moin et al., 2021), plant defense reactions, root elongation zone on roots and cell wall extensibility regulation in rice (Lucob-Agustin et al., 2020). These common connections between all networks helped us more than we expected because we identified one connection that is strongly involved in sucrose metabolism. Although this connection has not been described previously for sugarcane, this shows the efficiency and flexibility of computational methods such as coexpression networks to elucidate more relationship at sugarcane that could be important to be investigated by additional studies.

Interestingly, we found many genes that had no known annotation or lack inferences for other species and were not studied yet. This is unfortunately expected for species such as sugarcane, which do not have a complete reference genome due to its high genetic complexity. Consequently, this situation was evidenced when we analyzed the six selected modules which are enriched and associated with sucrose metabolism, as previously described. In detail, in hubs and common connections analyzed (Supplementary Tables 11,12,13,14,15), we observed several genes with all troubles described above but, according to network structure, could have provided important information about the links between sucrose metabolism and other pathways. We believe that these genes might belong specifically to sugarcane; additionally, we understand that a complete annotation of sugarcane could represent a valuable resource of information on gene structure, function, and regulatory networks related to sucrose metabolism, allowing the identification of specific sugar-controlling regions in the sugarcane genome (Barnabas et al., 2015; Thirugnanasambandam et al., 2018a).

8 Summary of results

8.1 Bioinformatics

- Using the *S. spontaneum* CDSs as a reference, gene expression was quantified in all samples and genes were aligned against the Phytozome database.
- These data were used to perform DGE analyses between phenotypes constrating on sugar accumulation.

8.2 GWGCN

- A global coexpression network was modelled with 34,085 genes, followed by the selection of modules related to sucrose categories for all varieties and species of sugarcane used in this work.
- On these selected modules, we identified key genes already associated with sucrose metabolic pathways. Studying the role of these genes, we reduced the number of modules based on GO terms, KEGG pathways enrichment, and DEG analyses.

8.3 Functional Modules' Characterization

- We found a great quantity of genes encoding transporters related with sucrose metabolism.
- The gene ontology categories showed important categories associated with sugar accumulation; inside the modules, we evaluated these genes based on their differential expression.

8.4 Gene Coexpression Networks' Comparisons

- Individual networks were modelled in order to assess the functional specificities of *S. spontaneum*, *S. officinarum* and the SP80-3280 hybrid. Additionally, we evaluated hubs with high scores and common connections between networks.

9 Conclusions

We provided novel cues into genetic associations with sugarcane sugar accumulation. Through coexpression networks and DEGs obtained with comparisons between genotypes with contrasting agronomic characteristics, we could identify a rich set of sucrose-related genes. We could also model the genotype-specific dynamic responses to sucrose metabolism, with an additional degree of gene enrichment to understand the molecular mechanisms involved in the metabolic pathways related to sucrose metabolism. Thus, with WGCNA, we were able to identify important modules and hub genes. Through these modules, we obtained information about biological pathways related to sucrose accumulation that could be strongly linked with genomics regions controlling this process.

10 Perspectives

These modules showed genes that could be treated as sugar markers, that could be employed to manage agronomic characteristics in sugarcane cultivars. We may suggest this based on studies that applied this approach to identify biomarkers associated with various processes in other species, such as the development of cancer and other diseases in humans (Liao et al., 2020b; Shi et al., 2020; Niu et al., 2019). Further research should be performed to study these genes related to sucrose metabolism separately, in order to better understand their specific genetic mechanisms on sucrose metabolism pathways (synthesis, transport, and storage). Due to the genetic complexity of sugarcane, it is difficult to obtain biological annotations and inferences, and there is a large amount of information still undeciphered. Despite the efforts of development of new computational methods to supply these deficiencies, more traditional approaches that allow more precise identification of the chromosomal regions related to sucrose metabolism, such as quantitative trait loci (QTL) mapping (Kugler et al., 2013; Springer et al., 2019) and genome-wide association studies (GWAS), are also valuable. QTL mapping can identify larger QTL regions containing potential candidate genes, which can be narrowed down with GWAS. These approaches can be complemented with output

data of gene expression networks in order to avoid potential false-positive errors and support the validation of their results, helping to establish a global approach for studying sucrose metabolism in sugarcane.

References

Alejandro, S., P. L. Rodríguez, J. M. Belles, L. Yenush, M. J. García-Sanchez, J. A. Fernandez, and R. Serrano (2007). An arabidopsis quiescin-sulfhydryl oxidase regulates cation homeostasis at the root symplast–xylem interface. *The EMBO journal 26*(13), 3203–3215.

Alessio, V. M., N. Cavaçana, L. L. d. B. Dantas, N. Lee, C. T. Hotta, T. Imaizumi, and M. Menossi (2018). The fbh family of bhlh transcription factors controls acc synthase expression in sugarcane. *Journal of experimental botany 69*(10), 2511–2525.

Alexa, A. and J. Rahnenführer (2009). Gene set enrichment analysis with topgo. *Bioconductor Improv 27*, 1–26.

Anur, R. M., N. Mufithah, W. D. Sawitri, H. Sakakibara, and B. Sugiharto (2020). Overexpression of sucrose phosphate synthase enhanced sucrose content and biomass production in transgenic sugarcane. *Plants 9*(2), 200.

Aono, A. H., R. J. G. Pimenta, A. L. B. Garcia, F. H. Correr, G. K. Hosaka, M. M. Carrasco, C. B. Cardoso-Silva, M. C. Mancini, D. A. Sforça, L. B. Dos Santos, et al. (2021). The wild sugarcane and sorghum kinomes: Insights into expansion, diversification, and expression patterns. *Frontiers in Plant Science 12*.

Bader, G. D. and C. W. Hogue (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics 4*(1), 2.

Bai, W.-Q., Y.-H. Xiao, J. Zhao, S.-Q. Song, L. Hu, J.-Y. Zeng, X.-B. Li, L. Hou, M. Luo, D.-M. Li, et al. (2014). Gibberellin overproduction promotes sucrose synthase expression and secondary cell wall deposition in cotton fibers. *PLoS One 9*(5), e96537.

Baker, R. F., K. A. Leach, N. R. Boyer, M. J. Swyers, Y. Benitez-Alfonso, T. Skopelitis, A. Luo, A. Sylvester, D. Jackson, and D. M. Braun (2016). Sucrose transporter ZmSut1 expression and localization uncover new insights into sucrose phloem loading. *Plant Physiology 172*(3), 1876–1898.

Ballouz, S., W. Verleyen, and J. Gillis (2015). Guidance for rna-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics 31*(13), 2123–2130.

Barkan, A. and I. Small (2014). Pentatricopeptide repeat proteins in plants. *Annual review of plant biology 65*, 415–442.

Barnabas, L., A. Ramadass, R. S. Amalraj, M. Palaniyandi, and V. Rasappa (2015). Sugarcane proteomics: An update on current status, challenges, and future prospects. *Proteomics 15*(10), 1658–1670.

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, B. A. Rapp, and D. L. Wheeler (2000). Genbank. *Nucleic acids research 28*(1), 15–18.

Bi, G., Z. Zhou, W. Wang, L. Li, S. Rao, Y. Wu, X. Zhang, F. L. Menke, S. Chen, and J.-M. Zhou (2018). Receptor-like cytoplasmic kinases directly link diverse pattern recognition receptors to the activation of mitogen-activated protein kinase cascades in arabidopsis. *The Plant Cell 30*(7), 1543–1561.

Bihmidine, S., R. F. Baker, C. Hoffner, and D. M. Braun (2015a). Sucrose accumulation in sweet sorghum stems occurs by apoplasmic phloem unloading and does not involve differential Sucrose transporter expression. *BMC Plant Biology 15*(1), 1–22.

Bihmidine, S., R. F. Baker, C. Hoffner, and D. M. Braun (2015b). Sucrose accumulation in sweet sorghum stems occurs by apoplasmic phloem unloading and does not involve differential sucrose transporter expression. *BMC plant biology 15*(1), 186.

Bihmidine, S., C. T. Hunter, C. E. Johns, K. E. Koch, and D. M. Braun (2013). Regulation of assimilate import into sink organs: Update on molecular drivers of sink strength. *Frontiers in Plant Science 4*(JUN), 1–15.

Bindon, K. A. and F. C. Botha (2002). Carbon allocation to the insoluble fraction, respiration and triose-phosphate cycling in the sugarcane culm. *Physiologia Plantarum 116*(1), 12–19.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008). Fast 379 unfolding of communities in large networks. *J. Stat. Mech. Theory Exp. P 10008*.

Boccaletti, S., V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang (2006). Complex networks: Structure and dynamics. *Physics Reports 424*(4-5), 175–308.

Bolger, A. M., M. Lohse, and B. Usadel (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics 30*(15), 2114–2120.

Borin, G. P., M. F. Carazzolle, R. A. C. dos Santos, D. M. Riaño-Pachón, and J. V. d. C. Oliveira (2018). Gene co-expression network reveals potential new genes related to sugarcane bagasse degradation in trichoderma reesei rut-30. *Frontiers in Bioengineering and Biotechnology 6*, 151.

Brindha, C., S. Vasantha, A. K. Raja, and A. S. Tayade (2021). Characterization of the salt overly sensitive pathway genes in sugarcane under salinity stress. *Physiologia Plantarum 171*(4), 677–687.

Cao, L.-c., Z.-j. Wang, G.-h. Ren, W. Kong, L. Li, W. Xie, and Y.-h. Liu (2015). Engineering a novel glucose-tolerant β-glucosidase as supplementation to enhance the hydrolysis of sugarcane bagasse at high glucose concentration. *Biotechnology for biofuels 8*(1), 1–12.

Cardoso-Silva, C. B., E. A. Costa, M. C. Mancini, T. W. A. Balsalobre, L. E. C. Canesin, L. R. Pinto, M. S. Carneiro, A. A. F. Garcia, A. P. de Souza, and R. Vicentini (2014). De novo assembly and transcriptome analysis of contrasting sugarcane varieties. *PloS one 9*(2).

Carpaneto, A., D. Geiger, E. Bamberg, N. Sauer, J. Fromm, and R. Hedrich (2005). Phloem-localized, proton-coupled sucrose carrier zmsut1 mediates sucrose efflux under the control of the sucrose gradient and the proton motive force. *Journal of Biological Chemistry 280*(22), 21437–21443.

Carrillo-Bermejo, E. A., S. D. Gamboa-Tuz, A. Pereira-Santana, M. A. Keb-Llanes, E. Castaño, L. J. Figueroa-Yañez, and L. C. Rodriguez-Zapata (2020). The sonap gene from sugarcane (saccharum officinarum) encodes a senescence-associated nac transcription factor involved in response to osmotic and salt stress. *Journal of plant research 133*(6), 897–909.

Casu, R. E., A. L. Rae, J. M. Nielsen, J. M. Perroux, G. D. Bonnett, and J. M. Manners (2015a). Tissue-specific transcriptome analysis within the maturing sugarcane stalk reveals spatial regulation in the expression of cellulose synthase and sucrose transporter gene families. *Plant Molecular Biology 89*(6), 607–628.

Casu, R. E., A. L. Rae, J. M. Nielsen, J. M. Perroux, G. D. Bonnett, and J. M. Manners (2015b). Tissue-specific transcriptome analysis within the maturing sugarcane stalk reveals spatial regulation in the expression of cellulose synthase and sucrose transporter gene families. *Plant molecular biology 89*(6), 607–628.

Chen, L.-Q., L. S. Cheung, L. Feng, W. Tanner, and W. B. Frommer (2015). Transport of sugars. *Annual Review of Biochemistry 84*.

Chen, L.-Q., X.-Q. Qu, B.-H. Hou, D. Sosso, S. Osorio, A. R. Fernie, and W. B. Frommer (2012). Sucrose efflux mediated by sweet proteins as a key step for phloem transport. *Science 335*(6065), 207–211.

Chen, Y.-S., Y.-C. Chao, T.-W. Tseng, C.-K. Huang, P.-C. Lo, and C.-A. Lu (2017). Two myb-related transcription factors play opposite roles in sugar signaling in arabidopsis. *Plant molecular biology 93*(3), 299–311.

Chen, Z., C. Qin, M. Wang, F. Liao, Q. Liao, X. Liu, Y. Li, P. Lakshmanan, M. Long, and D. Huang (2019a). Ethylene-mediated improvement in sucrose accumulation in ripening sugarcane involves increased sink strength. *BMC Plant Biology 19*(1), 1–17.

Chen, Z., C. Qin, M. Wang, F. Liao, Q. Liao, X. Liu, Y. Li, P. Lakshmanan, M. Long, and D. Huang (2019b). Ethylene-mediated improvement in sucrose accumulation in ripening sugarcane involves increased sink strength. *BMC plant biology 19*(1), 285.

Childs, K. L., R. M. Davidson, and C. R. Buell (2011). Gene coexpression network analysis as a source of functional annotation for rice genes. *PloS One 6*(7), e22196.

CONAB (2021). Acompanhamento da safra brasileira de cana-de-açúcar, brasília, df, n. 1-primeiro levantamento, brasília v. 8 – safra 2021-22. *8*, 56.

Consortium, G. O. (2006). The gene ontology (go) project in 2006. *Nucleic acids research 34*(suppl_1), D322–D326.

Consortium, U. (2016). Uniprot: the universal protein knowledgebase. *Nucleic Acids Research 45*(D1), D158–D169.

Correr, F. H., G. K. Hosaka, F. Z. Barreto, I. B. Valadão, T. W. A. Balsalobre, A. Furtado, R. J. Henry, M. S. Carneiro, and G. R. A. Margarido (2020). Differential expression in leaves of saccharum genotypes contrasting in biomass production provides evidence of genes involved in carbon partitioning. *BMC genomics 21*(1), 1–16.

Costa, L. d. F., O. N. Oliveira Jr, G. Travieso, F. A. Rodrigues, P. R. Villas Boas, L. Antiqueira, M. P. Viana, and L. E. Correa Rocha (2011). Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics 60*(3), 329–412.

Csardi, G., T. Nepusz, et al. (2006). The igraph software package for complex network research. *InterJournal, complex systems 1695*(5), 1–9.

Cuadrado, A., R. Acevedo, S. Moreno Diaz de la Espina, N. Jouve, and C. De La Torre (2004). Genome remodelling in three modern s. officinarum× s. spontaneum sugarcane cultivars. *Journal of experimental botany 55*(398), 847–854.

Cunningham, F., M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, et al. (2014). Ensembl 2015. *Nucleic acids research 43*(D1), D662–D669.

Daniels, J. and B. T. Roach (1987). Taxonomy and evolution. In *Developments in crop science*, Volume 11, pp. 7–84. Elsevier.

de Anda-Jáuregui, G., S. A. Alcalá-Corona, J. Espinal-Enríquez, and E. Hernández-Lemus (2019). Functional and transcriptional connectivity of communities in breast cancer co-expression networks. *Applied Network Science 4*(1), 22.

Deepak, S., S. Shailasree, R. K. Kini, A. Muck, A. Mithöfer, and S. H. Shetty (2010). Hydroxyproline-rich glycoproteins and plant defence. *Journal of Phytopathology 158*(9), 585–593.

Di, Y., D. Chen, W. Yu, and L. Yan (2019). Bladder cancer stage-associated hub genes revealed by wgcna co-expression network analysis. *Hereditas 156*(1), 7.

Douce, R. and D. A. Day (1985). Higher plant cell respiration.

Drozdov, I., B. Svejda, B. I. Gustafsson, S. Mane, R. Pfragner, M. Kidd, and I. M. Modlin (2011). Gene network inference and biochemical assessment delineates gpcr pathways and creb targets in small intestinal neuroendocrine neoplasia. *PLoS One 6*(8), e22457.

D'Hont, A. and J. C. Glaszmann (2001). Sugarcane genome analysis with molecular markers, a first decade of research. In *Proc Int Soc Sugarcane Technol*, Volume 24, pp. 556–559.

ElSayed, A. I., A. Lehrer, M. Ebrahim, A. H. Mohamed, and E. Komor (2017). Assessment of sucrose transporters, metabolites and sucrose phosphate synthase in different sugarcane tissues. *Physiology and Molecular Biology of Plants 23*(3), 703–712.

Faghani, E., J. Gharechahi, S. Komatsu, M. Mirzaei, R. A. Khavarinejad, F. Najafi, L. K. Farsad, and G. H. Salekdeh (2015). Comparative physiology and proteomic analysis of two wheat genotypes contrasting in drought tolerance. *Journal of proteomics 114*, 1–15.

Fait, A., A. Batushansky, V. Shrestha, A. Yobi, and R. Angelovici (2020). Can metabolic tightening and expansion of co-expression network play a role in stress response and tolerance? *Plant Science 293*, 110409.

FAO (2020). Faostat: Production sheet. URL http://www.fao.org/faostat/en.

Fekih, R., M. Tamiru, H. Kanzaki, A. Abe, K. Yoshida, E. Kanzaki, H. Saitoh, H. Takagi, S. Natsume, J. R. Undan, et al. (2015). The rice (oryza sativa l.) lesion mimic resembling, which encodes an aaa-type atpase, is implicated in defense response. *Molecular Genetics and Genomics 290*(2), 611–622.

Feltus, F. A., S. P. Ficklin, S. M. Gibson, and M. C. Smith (2013). Maximizing capture of gene co-expression relationships through pre-clustering of input expression samples: an arabidopsis case study. *BMC Systems Biology 7*(1), 1–12.

Ferlic, J. and S. Tracy (2016). Wgcna: Weighted gene co-expression network analysis.

Ferreira, S. S., C. T. Hotta, V. G. de Carli Poelking, D. C. C. Leite, M. S. Buckeridge, M. E. Loureiro, M. H. P. Barbosa, M. S. Carneiro, and G. M. Souza (2016). Co-expression network analysis reveals transcription factors associated to cell wall biosynthesis in sugarcane. *Plant molecular biology 91*(1-2), 15–35.

Ferreira, S. S., C. T. Hotta, V. G. d. C. Poelking, D. C. C. Leite, M. S. Buckeridge, M. E. Loureiro, M. H. P. Barbosa, M. S. Carneiro, and G. M. Souza (2016). Co-expression network analysis reveals transcription factors associated to cell wall biosynthesis in sugarcane. *Plant Molecular Biology 91*(1-2), 15–35.

Fichtner, F. and J. E. Lunn (2021). The role of trehalose 6-phosphate (tre6p) in plant metabolism and development. *Annual Review of Plant Biology 72*.

Ficklin, S. P., L. J. Dunwoodie, W. L. Poehlman, C. Watson, K. E. Roche, and F. A. Feltus (2017). Discovering condition-specific gene co-expression patterns using gaussian mixture models: a cancer case study. *Scientific Reports 7*(1), 1–11.

Ficklin, S. P. and F. A. Feltus (2011a). Gene coexpression network alignment and conservation of gene modules between two grass species: Maize and rice. *Plant Physiology 156*(3), 1244–1256.

Ficklin, S. P. and F. A. Feltus (2011b). Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. *Plant Physiology 156*(3), 1244–1256.

Figueiredo, P. (2008). *Breve história da cana-de-açúcar e do papel do Instituto Agronômico no seu estabelecimento no Brasil*. Cana-de-açúcar. Campinas: Instituto Agronomico.

Figueroa, C. M. and J. E. Lunn (2016). A tale of two sugars: trehalose 6-phosphate and sucrose. *Plant Physiology 172*(1), 7–27.

Fortunato, S. (2010). Community detection in graphs. *Physics reports 486*(3-5), 75–174.

Garsmeur, O., G. Droc, R. Antonise, J. Grimwood, B. Potier, K. Aitken, J. Jenkins, G. Martin, C. Charron, C. Hervouet, et al. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature Communications 9*(1), 1–10.

Gomathi, R., S. Kohila, and L. Kasirajan (2018). Differential transcript expression profiling of sugarcane (co 99004) for elevated temperature stress. *Madras Agricultural Journal 105*.

Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, et al. (2011). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research 40*(D1), D1178–D1186.

Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic acids research 40*(D1), D1178–D1186.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, et al. (2011). Full-length transcriptome assembly from rna-seq data without a reference genome. *Nature biotechnology 29*(7), 644.

Heim, M. A., M. Jakoby, M. Werber, C. Martin, B. Weisshaar, and P. C. Bailey (2003). The basic helix–loop–helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Molecular biology and evolution 20*(5), 735–747.

Horvath, S. (2011). *Weighted network analysis: applications in genomics and systems biology*. Springer Science & Business Media.

Hosaka, G. K., F. H. Correr, C. C. Silva, D. A. Sforça, F. Z. Barreto, T. W. A. Balsalobre, A. Pasha, A. P. de Souza, N. J. Provart, M. S. Carneiro, et al. (2021). Temporal gene expression in apical culms shows early changes in cell wall biosynthesis genes in sugarcane. *bioRxiv*.

Hu, X., Z.-D. Wu, Z.-Y. Luo, D. M. Burner, Y.-B. Pan, and C.-W. Wu (2020). Genome-wide analysis of the trehalose-6-phosphate synthase (tps) gene family and expression profiling of sctps genes in sugarcane. *Agronomy 10*(7), 969.

Hu, Y.-X., Y.-B. Tao, and Z.-F. Xu (2017). Overexpression of jatropha gibberellin 2-oxidase 6 (jcga2ox6) induces dwarfism and smaller leaves, flowers and fruits in arabidopsis and jatropha. *Frontiers in plant science 8*, 2103.

Huang, Z., A. Ma, S. Yang, X. Liu, and T. Zhao (2020). Comparative Biochemistry and Physiology - Part D Transcriptome analysis and weighted gene co-expression network reveals potential genes responses to heat stress in turbot Scophthalmus maximus. *Comparative Biochemistry and Physiology - Part D 33*(106), 100632.

Jackson, P. A. (2005). Breeding for improved sugar content in sugarcane. *Field Crops Research 92*(2-3), 277–290.

James, G. (2004). An introduction to sugarcane. sugarcane.

Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási (2000). The large-scale organization of metabolic networks. *Nature 407*(6804), 651–654.

Ji, H., S. Wang, C. Cheng, R. Li, Z. Wang, G. I. Jenkins, F. Kong, and X. Li (2019). The rcc 1 family protein sab 1 negatively regulates abi 5 through multidimensional mechanisms during postgermination in arabidopsis. *New Phytologist 222*(2), 907–922.

Julius, B. T., K. A. Leach, T. M. Tran, R. A. Mertz, and D. M. Braun (2017). Sugar transporters in plants: new insights and discoveries. *Plant and Cell Physiology 58*(9), 1442–1460.

Jung, B., F. Ludewig, A. Schulz, G. Meissner, N. Woestefeld, U.-I. Fluegge, B. Pommerrenig, P. Wirsching, N. Sauer, W. Koch, et al. (2015). Identification of the transporter responsible for sucrose accumulation in sugar beet taproots. *Nature Plants 1*(1), 1–6.

Kalaipandian, S., G.-P. Xue, A. L. Rae, D. Glassop, G. D. Bonnett, and L. C. McIntyre (2019). Overexpression of tacml20, a calmodulin-like gene, enhances water soluble carbohydrate accumulation and yield in wheat. *Physiologia plantarum 165*(4), 790–799.

Kalde, M., L. Elliott, R. Ravikumar, K. Rybak, M. Altmann, S. Klaeger, C. Wiese, M. Abele, B. Al, N. Kalbfuß, et al. (2019). Interactions between transport protein particle (trapp) complexes and rab gtp ases in arabidopsis. *The Plant Journal 100*(2), 279–297.

Kanehisa, M. and S. Goto (2000). Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research 28*(1), 27–30.

Kang, S. G., J. Price, P.-C. Lin, J. C. Hong, and J.-C. Jang (2010). The arabidopsis bzip1 transcription factor is involved in sugar signaling, protein networking, and dna binding. *Molecular plant 3*(2), 361–373.

Kasirajan, L., N. V. Hoang, A. Furtado, F. C. Botha, and R. J. Henry (2018). Transcriptome analysis highlights key differentially expressed genes involved in cellulose and lignin biosynthesis of sugarcane genotypes varying in fiber content. *Scientific reports 8*(1), 1–16.

Khan, Q., J.-Y. Chen, X.-P. Zeng, Y. Qin, D.-J. Guo, A. Mahmood, L.-T. Yang, Q. Liang, X.-P. Song, Y.-X. Xing, et al. Transcriptomic exploration of a high sucrose mutant in comparison with the low sucrose mother genotype in sugarcane during sugar accumulating stage. *GCB Bioenergy*.

Kleinberg, J. M., M. Newman, A.-L. Barabási, and D. J. Watts (2011). *Authoritative sources in a hyperlinked environment*. Princeton University Press.

Koonin, E. V. and M. Galperin (2013). *Sequence—evolution—function: computational approaches in comparative genomics*. Springer Science & Business Media.

Koonin, E. V. and M. Y. Galperin (2003). Genome annotation and analysis. In *Sequence—Evolution—Function*, pp. 193–226. Springer.

Kopylova, E., L. Noé, and H. Touzet (2012). Sortmerna: fast and accurate filtering of ribosomal rnas in metatranscriptomic data. *Bioinformatics 28*(24), 3211–3217.

Kugler, K. G., G. Siegwart, T. Nussbaumer, C. Ametz, M. Spannagl, B. Steiner, M. Lemmens, K. F. Mayer, H. Buerstmayr, and W. Schweiger (2013). Quantitative trait loci-dependent analysis of a gene co-expression network associated with fusarium head blight resistance in bread wheat (triticum aestivum l.). *BMC genomics 14*(1), 1–15.

Kurniah, N. I., W. D. Sawitri, M. S. Rohman, Y. Nugraha, T. Hase, and B. Sugiharto (2021). Mutation of udp-glucose binding motif residues lead to increased affinity for adp-glucose in sugarcane sucrose phosphate synthase. *Molecular Biology Reports 48*(2), 1697–1706.

Langfelder, P. and S. Horvath (2008a). Wgcna: an r package for weighted correlation network analysis. *BMC Bioinformatics 9*(1), 1–13.

Langfelder, P. and S. Horvath (2008b). Wgcna: an r package for weighted correlation network analysis. *BMC bioinformatics 9*(1), 1–13.

Langfelder, P., P. S. Mischel, and S. Horvath (2013). When is hub gene selection better than standard meta-analysis? *PloS One 8*(4), e61505.

Langfelder, P., B. Zhang, and S. Horvath (2008). Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for r. *Bioinformatics 24*(5), 719–720.

Lastdrager, J., J. Hanson, and S. Smeekens (2014). Sugar signals and the control of plant growth and development. *Journal of Experimental Botany 65*(3), 799–807.

Li, C., Q. Nong, J. Xie, Z. Wang, Q. Liang, M. K. Solanki, M. K. Malviya, X. Liu, Y. Li, R. Htun, et al. (2017). Molecular characterization and co-expression analysis of the snrk2 gene family in sugarcane (saccharum officinarum l.). *Scientific reports 7*(1), 1–12.

Li, H., Y. Wang, Q. Xiao, L. Luo, C. Zhang, C. Mao, J. Du, T. Long, Y. Cao, Q. Yi, et al. (2021). Transcription factor zmplatz2 positively regulate the starch synthesis in maize. *Plant Growth Regulation 93*(3), 291–302.

Li, P., Z. Chai, P. Lin, C. Huang, G. Huang, L. Xu, Z. Deng, M. Zhang, Y. Zhang, and X. Zhao (2020). Genome-wide identification and expression analysis of ap2/erf transcription factors in sugarcane (saccharum spontaneum l.). *BMC genomics 21*(1), 1–17.

Liao, Y., Y. Wang, M. Cheng, C. Huang, and X. Fan (2020a). Weighted gene coexpression network analysis of features that control cancer stem cells reveals prognostic biomarkers in lung adenocarcinoma.

Liao, Y., Y. Wang, M. Cheng, C. Huang, and X. Fan (2020b). Weighted gene coexpression network analysis of features that control cancer stem cells reveals prognostic biomarkers in lung adenocarcinoma. *Frontiers in Genetics 11*, 311.

Lucob-Agustin, N., T. Kawai, M. Takahashi-Nosaka, M. Kano-Nakata, C. M. Wainaina, T. Hasegawa, M. Inari-Ikeda, M. Sato, H. Tsuji, A. Yamauchi, et al. (2020). Weg1, which encodes a cell wall hydroxyproline-rich glycoprotein, is essential for parental root elongation controlling lateral root formation in rice. *Physiologia plantarum 169*(2), 214–227.

Lunn, J. E., I. Delorge, C. M. Figueroa, P. Van Dijck, and M. Stitt (2014). Trehalose metabolism in plants. *The Plant Journal 79*(4), 544–567.

Lutfiyya, L. L., N. Xu, L. Robert, J. A. Morrell, P. W. Miller, and S. M. Duff (2007). Phylogenetic and expression analysis of sucrose phosphate synthase isozymes in plants. *Journal of Plant Physiology 164*(7), 923–933.

Ma, P., X. Zhang, L. Chen, Q. Zhao, Q. Zhang, X. Hua, Z. Wang, H. Tang, Q. Yu, M. Zhang, et al. (2020). Comparative analysis of sucrose phosphate synthase (sps) gene family between saccharum officinarum and saccharum spontaneum. *BMC plant biology 20*(1), 1–15.

Mall, R., E. Ullah, K. Kunji, F. D'Angelo, H. Bensmail, and M. Ceccarelli (2017a). Differential community detection in paired biological networks. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 330–339.

Mall, R., E. Ullah, K. Kunji, F. D'Angelo, H. Bensmail, and M. Ceccarelli (2017b). Differential community detection in paired biological networks. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 330–339.

Mancini, M. C., C. B. Cardoso-Silva, D. A. Sforça, and A. Pereira de Souza (2018). "targeted sequencing by gene synteny," a new strategy for polyploid species: Sequencing and physical structure of a complex sugarcane region. *Frontiers in plant science 9*, 397.

Marguerat, S. and J. Bähler (2010). Rna-seq: from technology to biology. *Cellular and molecular life sciences 67*(4), 569–579.

Maruta, N., Y. Trusov, D. Urano, D. Chakravorty, S. M. Assmann, A. M. Jones, and J. R. Botella (2021). Gtp binding by arabidopsis extra-large g protein 2 is not essential for its functions. *Plant Physiology 186*(2), 1240–1253.

Mattiello, L., D. M. Riaño-Pachón, M. C. M. Martins, L. P. da Cruz, D. Bassi, P. E. R. Marchiori, R. V. Ribeiro, M. T. Labate, C. A. Labate, and M. Menossi (2015). Physiological and transcriptional analyses of developmental stages along sugarcane leaf. *BMC Plant Biology 15*(1), 1–21.

McCormick, A., M. Cramer, and D. Watt (2006a). Sink strength regulates photosynthesis in sugarcane. *New Phytologist 171*(4), 759–770.

McCormick, A. J., M. D. Cramer, and D. A. Watt (2006b). Sink strength regulates photosynthesis in sugarcane. *New Phytologist 171*(4), 759–770.

Meng, Z., C. Ruberti, Z. Gong, and F. Brandizzi (2017). Cpr 5 modulates salicylic acid and the unfolded protein response to manage tradeoffs between plant growth and stress responses. *The Plant Journal 89*(3), 486–501.

Mirajkar, S. J., P. Suprasanna, and E. R. Vaidya" (2016). Spatial distribution and dynamics of sucrose metabolising enzymes in radiation induced mutants of sugarcane. *Plant Physiology and Biochemistry 100*, 85 – 93.

Mitra, K., A.-R. Carvunis, S. K. Ramesh, and T. Ideker (2013a). Integrative approaches for finding modular structure in biological networks. *Nature Reviews Genetics 14*(10), 719–732.

Mitra, K., A.-R. Carvunis, S. K. Ramesh, and T. Ideker (2013b). Integrative approaches for finding modular structure in biological networks. *Nature Reviews Genetics 14*(10), 719–732.

Moin, M., A. Saha, A. Bakshi, M. Madhav, and P. Kirti (2021). Constitutive expression of ribosomal protein l6 modulates salt tolerance in rice transgenic plants. *Gene 789*, 145670.

Moore, P. (1995). Temporal and spatial regulation of sucrose accumulation in the sugarcane stem. *22*(4), 661–679.

Moore, P. H. and D. J. Cosgrove (1991). Developmental changes in cell and tissue water relations parameters in storage parenchyma of sugarcane. *96*(3), 794–801.

Mutwil, M., B. Usadel, M. Schutte, A. Loraine, O. Ebenhoh, and S. Persson (2010). Assembly of an interactive correlation network for the arabidopsis genome using a novel heuristic clustering algorithm. *Plant physiology 152*(1), 29–43.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review 45*(2), 167–256.

Niu, X., J. Zhang, L. Zhang, Y. Hou, S. Pu, A. Chu, M. Bai, and Z. Zhang (2019). Weighted gene co-expression network analysis identifies critical genes in the development of heart failure after acute myocardial infarction. *Frontiers in genetics 10*, 1214.

Nukarinen, E., T. Nägele, L. Pedrotti, B. Wurzinger, A. Mair, R. Landgraf, F. Börnke, J. Hanson, M. Teige, E. Baena-Gonzalez, et al. (2016). Quantitative phosphoproteomics

reveals the role of the ampk plant ortholog snrk1 as a metabolic master regulator under energy deprivation. *Scientific Reports 6*, 31697.

Oloriz, M. I., O. Portal, L. Rojas, V. Gil, B. Ocana, N. Machado, M. Hernández, E. Jiménez, and M. Höfte (2021). Puccinia melanocephala and salt stress modified the inositol phosphate and tor signalling pathway in sugarcane. *Sugar Tech 23*(2), 407–414.

Palhares, A. C., T. B. Rodrigues-Morais, M.-A. Van Sluys, D. S. Domingues, W. Maccheroni, H. Jordão, A. P. Souza, T. G. Marconi, M. Mollinari, R. Gazaffi, et al. (2012). A novel linkage map of sugarcane with evidence for clustering of retrotransposon-based markers. *BMC genetics 13*(1), 51.

Patrick, J. W., F. C. Botha, and R. G. Birch (2013). Metabolic engineering of sugars and simple sugar derivatives in plants. *Plant biotechnology journal 11*(2), 142–156.

Patro, R., G. Duggal, and C. Kingsford (2015). Salmon: accurate, versatile and ultrafast quantification from rna-seq data using lightweight-alignment. *Biorxiv*, 021592.

Payyavula, R. S., R. K. Singh, and D. A. Navarre (2013). Transcription factors, sucrose, and sucrose metabolic genes interact to regulate potato phenylpropanoid metabolism. *Journal of experimental botany 64*(16), 5115–5131.

Pereira, L. F., V. M. Ferreira, N. G. OLIVEIRA, P. L. Sarmento, L. Endres, and I. Teodoro (2017). Sugars levels of four sugarcane genotypes in different stem portions during the maturation phase. *Anais da Academia Brasileira de Ciências 89*(2), 1231–1242.

Pereira-Santana, A., E. J. Alvarado-Robledo, J. A. Zamora-Briseno, J. T. Ayala-Sumuano, V. M. Gonzalez-Mendoza, F. Espadas-Gil, L. D. Alcaraz, E. Castano, M. A. Keb-Llanes, F. Sanchez-Teyer, et al. (2017). Transcriptional profiling of sugarcane leaves and roots under progressive osmotic stress reveals a regulated coordination of gene expression in a spatiotemporal manner. *PloS one 12*(12).

Perlo, V., F. C. Botha, A. Furtado, K. Hodgson-Kratky, and R. J. Henry (2020). Metabolic changes in the developing sugarcane culm associated with high yield and early high sugar content. *Plant direct 4*(11), e00276.

Piperidis, G., N. Piperidis, and A. D'Hont (2010). Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Molecular Genetics and Genomics 284*(1), 65–73.

Poehlman, W., E. Schnabel, S. Chavan, J. A. Frugoli, and F. A. Feltus (2019). Identifying temporally regulated root nodulation biomarkers using time series gene co-expression network analysis. *Frontiers in plant science 10*, 1409.

Polge, C. and M. Thomas (2007). Snf1/ampk/snrk1 kinases, global regulators at the heart of energy control? *Trends in plant science 12*(1), 20–28.

Ponnu, J., V. Wahl, and M. Schmid (2011). Trehalose-6-phosphate: connecting plant metabolism and development. *Frontiers in Plant Science 2*, 70.

Proost, S., M. Van Bel, D. Vaneechoutte, Y. Van de Peer, D. Inzé, B. Mueller-Roeber, and K. Vandepoele (2014). Plaza 3.0: an access point for plant comparative genomics. *Nucleic Acids Research 43*(D1), D974–D981.

Qin, C.-X., Z.-L. Chen, M. Wang, A.-M. Li, F. Liao, Y.-R. Li, M.-Q. Wang, M.-H. Long, P. Lakshmanan, and D.-L. Huang (2021). Identification of proteins and metabolic networks associated with sucrose accumulation in sugarcane (saccharum spp. interspecific hybrids). *Journal of Plant Interactions 16*(1), 166–178.

Qiu, Z., D. Chen, L. He, S. Zhang, Z. Yang, Y. Zhang, Z. Wang, D. Ren, Q. Qian, L. Guo, et al. (2018). The rice white green leaf 2 gene causes defects in chloroplast development and affects the plastid ribosomal protein s9. *Rice 11*(1), 1–12.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rae, A. L., J. M. Perroux, and C. P. Grof (2005a). Sucrose partitioning between vascular bundles and storage parenchyma in the sugarcane stem: a potential role for the shsut1 sucrose transporter. *Planta 220*(6), 817–825.

Rae, A. L., J. M. Perroux, and C. P. Grof (2005b). Sucrose partitioning between vascular bundles and storage parenchyma in the sugarcane stem: a potential role for the shsut1 sucrose transporter. *Planta 220*(6), 817–825.

Rahmani, F., M. Hummel, J. Schuurmans, A. Wiese-Klinkenberg, S. Smeekens, and J. Hanson (2009a). Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. *Plant Physiology 150*(3), 1356–1367.

Rahmani, F., M. Hummel, J. Schuurmans, A. Wiese-Klinkenberg, S. Smeekens, and J. Hanson (2009b). Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. *Plant physiology 150*(3), 1356–1367.

Reinders, A., A. B. Sivitz, A. Hsi, C. P. Grof, J. M. Perroux, and J. M. Ward (2006). Sugarcane shsut1: analysis of sucrose transport activity and inhibition by sucralose. *Plant, cell & environment 29*(10), 1871–1880.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth (2010). edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics 26*(1), 139–140.

Rook, F., N. Gerrits, A. Kortstee, M. Van Kampen, M. Borrias, P. Weisbeek, and S. Smeekens (1998). Sucrose-specific signalling represses translation of the arabidopsis atb2 bzip transcription factor gene. *The Plant Journal 15*(2), 253–263.

Rosvall, M. and C. T. Bergstrom (2007). Maps of information flow reveal community structure in complex networks. *ArXiv Preprint Physics.soc-ph/0707.0609*.

Rushton, P. J., I. E. Somssich, P. Ringler, and Q. J. Shen (2010). Wrky transcription factors. *Trends in plant science 15*(5), 247–258.

Saha, A., Y. Kim, A. D. Gewirtz, B. Jo, C. Gao, I. C. McDowell, B. E. Engelhardt, A. Battle, F. Aguet, K. G. Ardlie, et al. (2017). Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Research 27*(11), 1843–1858.

Salerno, G. L. and L. Curatti (2003). Origin of sucrose metabolism in higher plants: when, how and why? *Trends in Plant Science 8*(2), 63–69.

Schaker, P. D., A. C. Palhares, L. M. Taniguti, L. P. Peters, S. Creste, K. S. Aitken, M.-A. Van Sluys, J. P. Kitajima, M. L. Vieira, and C. B. Monteiro-Vitorello (2016). Rnaseq transcriptional profiling following whip development in sugarcane smut disease. *PLoS One 11*(9), e0162237.

Selvarajan, D., C. Mohan, V. Dhandapani, G. Nerkar, A. N. Jayanarayanan, M. V. Mohanan, N. Murugan, L. Kaur, M. Chennappa, R. Kumar, et al. (2018). Differential gene expression profiling through transcriptome approach of saccharum spontaneum l. under low temperature stress reveals genes potentially involved in cold acclimation. *3 Biotech 8*(4), 195.

Selvi, A., K. Devi, R. Manimekalai, P. T. Prathima, R. Valiyaparambth, and K. Lakshmi (2021). High-throughput mirna deep sequencing in response to drought stress in sugarcane. *3 Biotech 11*(7), 1–18.

Serin, E. A., H. Nijveen, H. W. Hilhorst, and W. Ligterink (2016). Learning from co-expression networks: possibilities and challenges. *Frontiers in plant science 7*, 444.

Sforça, D. A., S. Vautrin, C. B. Cardoso-Silva, M. C. Mancini, M. V. R. d. Cruz, G. S. Pereira, M. Conte, A. Bellec, N. Dahmer, J. Fourment, et al. (2019). Gene duplication in the sugarcane genome: A case study of allele interactions and evolutionary patterns in two genic regions. *Frontiers in plant science 10*, 553.

Shi, G., Z. Shen, Y. Liu, and W. Yin (2020). Identifying biomarkers to predict the progression and prognosis of breast cancer by weighted gene co-expression network analysis. *Frontiers in Genetics 11*, 1614.

Silva, M. D. A., J. L. Jifon, J. A. Da Silva, and V. Sharma (2007). Use of physiological parameters as fast tools to screen for drought tolerance in sugarcane. *Brazilian Journal of Plant Physiology 9*(7), 193–201.

Singh, A. and I. K. Singh (2018). *Molecular aspects of plant-pathogen interaction*. Springer.

Slewinski, T. L. and D. M. Braun (2010). Current perspectives on the regulation of whole-plant carbohydrate partitioning. *Plant Science 178*(4), 341–349.

Smid, M., R. R. C. van den Braak, H. J. van de Werken, J. van Riet, A. van Galen, V. de Weerd, M. van der Vlugt-Daane, S. I. Bril, Z. S. Lalmahomed, W. P. Kloosterman, et al. (2018). Gene length corrected trimmed mean of m-values (getmm) processing of rna-seq data performs similarly in intersample analyses while improving intrasample comparisons. *BMC bioinformatics 19*(1), 1–13.

Smith, A. M. and M. Stitt (2007). Coordination of carbon supply and plant growth. *Plant, cell & environment 30*(9), 1126–1149.

Soneson, C., M. I. Love, and M. D. Robinson (2015). Differential analyses for rna-seq: transcript-level estimates improve gene-level inferences. *F1000Research 4*.

Souza, G. M., M.-A. Van Sluys, C. G. Lembke, H. Lee, G. R. A. Margarido, C. T. Hotta, J. W. Gaiarsa, A. L. Diniz, M. d. M. Oliveira, S. d. S. Ferreira, 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), giz129.

Springer, N., N. de León, and E. Grotewold (2019). Challenges of translating gene regulatory information into agronomic improvements. *Trends in plant science 24*(12), 1075–1082.

Stein, O. and D. Granot (2018). Plant fructokinases: evolutionary, developmental, and metabolic aspects in sink tissues. *Frontiers in plant science 9*, 339.

Strogatz, S. H. (2001). Exploring complex networks. *nature 410*(6825), 268–276.

Sun, Y., J. Huang, S. Zhong, H. Gu, S. He, and L.-J. Qu (2018). Novel dyw-type pentatricopeptide repeat (ppr) protein blx controls mitochondrial rna editing and splicing essential for early seed development of arabidopsis. *Journal of Genetics and Genomics 45*(3), 155–168.

Szklarczyk, D., J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, et al. (2016). The string database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research*, gkw937.

Tai, Y., C. Ling, C. Wang, H. Wang, L. Su, L. Yang, W. Jiang, X. Yu, L. Zheng, Z. Feng, et al. (2020). Analysis of terpenoid biosynthesis pathways in german chamomile (matricaria recutita) and roman chamomile (chamaemelum nobile) based on co-expression networks. *Genomics 112*(2), 1055–1064.

Team, R. C. et al. (2013). R: A language and environment for statistical computing.

Thalor, S. K., T. Berberich, S. S. Lee, S. H. Yang, X. Zhu, R. Imai, Y. Takahashi, and T. Kusano (2012). Deregulation of sucrose-controlled translation of a bzip-type transcription factor results in sucrose accumulation in leaves. *PLoS One 7*(3), e33111.

Thirugnanasambandam, P. P., N. V. Hoang, A. Furtado, F. C. Botha, and R. J. Henry (2017). Association of variation in the sugarcane transcriptome with sugar content. *BMC genomics 18*(1), 909.

Thirugnanasambandam, P. P., N. V. Hoang, and R. J. Henry (2018a). The challenge of analyzing the sugarcane genome. *Frontiers in plant science 9*, 616.

Thirugnanasambandam, P. P., N. V. Hoang, and R. J. Henry (2018b). The challenge of analyzing the sugarcane genome. *Frontiers in Plant Science 9*(May), 1–18.

Thirugnanasambandam, P. P., P. J. Mason, N. V. Hoang, A. Furtado, F. C. Botha, and R. J. Henry (2019). Analysis of the diversity and tissue specificity of sucrose synthase genes in the long read transcriptome of sugarcane. *BMC plant biology 19*(1), 1–14.

Uygun, S., C. Peng, M. D. Lehti-Shiu, R. L. Last, and S.-H. Shiu (2016). Utility and limitations of using gene expression data to identify functional associations. *PLoS Computational Biology 12*(12), e1005244.

Van Dam, S., U. Vosa, A. van der Graaf, L. Franke, and J. P. de Magalhaes (2018). Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings in Bioinformatics 19*(4), 575–592.

Vargas, W. A. and G. L. Salerno (2010a). The cinderella story of sucrose hydrolysis: alkaline/neutral invertases, from cyanobacteria to unforeseen roles in plant cytosol and organelles. *Plant Science 178*(1), 1–8.

Vargas, W. A. and G. L. Salerno (2010b). The cinderella story of sucrose hydrolysis: alkaline/neutral invertases, from cyanobacteria to unforeseen roles in plant cytosol and organelles. *Plant Science 178*(1), 1–8.

Verma, A. K., S. Upadhyay, P. C. Verma, S. Solomon, and S. B. Singh (2011). Functional analysis of sucrose phosphate synthase (sps) and sucrose synthase (ss) in sugarcane (saccharum) cultivars. *Plant Biology 13*(2), 325–332.

Verma, I., K. Roopendra, A. Sharma, A. Chandra, and A. Kamal (2019). Expression analysis of genes associated with sucrose accumulation and its effect on source–sink relationship in high sucrose accumulating early maturing sugarcane variety. *Physiology and Molecular Biology of Plants 25*(1), 207–220.

Vicentini, R., A. Bottcher, M. Dos Santos Brito, A. B. Dos Santos, S. Creste, M. G. De Andrade Landell, I. Cesarino, and P. Mazzafera (2015). Large-scale transcriptome analysis of two sugarcane genotypes contrasting for lignin content. *PLoS One 10*(8), 1–19.

Vicentini, R., A. Bottcher, M. dos Santos Brito, A. B. dos Santos, S. Creste, M. G. de Andrade Landell, I. Cesarino, and P. Mazzafera (2015). Large-scale transcriptome analysis of two sugarcane genotypes contrasting for lignin content. *PloS one 10*(8), e0134909.

Vicentini, R., J. de Maria Felix, M. C. Dornelas, and M. Menossi (2009). Characterization of a sugarcane (saccharum spp.) gene homolog to the brassinosteroid insensitive1-associated receptor kinase 1 that is associated to sugar content. *Plant cell reports 28*(3), 481–491.

Wan, H., L. Wu, Y. Yang, G. Zhou, and Y.-L. Ruan (2018). Evolution of sucrose metabolism: the dichotomy of invertases and beyond. *Trends in plant science 23*(2), 163–177.

Wang, J., T. Zhao, B. Yang, and S. Zhang (2017). Sucrose metabolism and regulation in sugarcane. *Journal of Plant Physiology & Pathology 5*(4).

Wang, J.-g., T.-t. Zhao, W.-z. Wang, C.-l. Feng, X.-y. Feng, G.-r. Xiong, L.-b. Shen, S.-z. Zhang, W.-q. Wang, and Z.-x. Zhang (2019). Culm transcriptome sequencing of badila (saccharum officinarum l.) and analysis of major genes involved in sucrose accumulation. *Plant Physiology and Biochemistry 144*, 455–465.

Wang, L., L. Wu, Q. Chen, S. Li, Y. Zhu, J. Wu, J. Chu, and S. Wu (2019). Development of sugarcane resource for efficient fermentation of exopolysaccharide by using a novel strain of kosakonia cowanii lt-1. *Bioresource technology 280*, 247–254.

Wang, Y.-Y., F. Xiong, Q.-P. Ren, and X.-L. Wang (2020). Regulation of flowering transition by alternative splicing: the role of the u2 auxiliary factor. *Journal of experimental botany 71*(3), 751–758.

Watt, D., A. McCormick, C. Govender, D. Carson, M. Cramer, B. Huckett, and F. Botha (2005). Increasing the utility of genomics in unravelling sucrose accumulation. *Field Crops Research 92*(2-3), 149–158.

Wei, K. and P. Han (2017). Comparative functional genomics of the tpr gene family in arabidopsis, rice and maize. *Molecular breeding 37*(12), 152.

Welbaum, G. E. and F. C. Meinzer (1990). Compartmentation of solutes and water in developing sugarcane stalk tissue. *Plant Physiology 93*(3), 1147–1153.

Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics 3*(2), 180–185.

Wingett, S. W. and S. Andrews (2018). Fastq screen: A tool for multi-genome mapping and quality control. *F1000Research 7*.

Wu, K.-C., L.-P. Wei, C.-M. Huang, Y.-W. Wei, H.-Q. Cao, L. Xu, H.-B. Luo, S.-L. Jiang, Z.-N. Deng, and Y.-R. Li (2018). Transcriptome reveals differentially expressed genes in saccharum spontaneum gx83-10 leaf under drought stress. *Sugar Tech 20*(6), 756–764.

Wu, L. and R. G. Birch (2010). Physiological basis for enhanced sucrose accumulation in an engineered sugarcane cell line. *Functional Plant Biology 37*(12), 1161–1174.

Wurzinger, B., E. Nukarinen, T. Nägele, W. Weckwerth, and M. Teige (2018). The snrk1 kinase as central mediator of energy signaling between different organelles. *Plant Physiology 176*(2), 1085–1094.

Yadav, B. S., S. Singh, S. Srivastava, and A. Mani (2019). Analysis of chickpea gene co-expression networks and pathways during heavy metal stress. *Journal of Biosciences 44*(4), 1–10.

Yan, J., H. He, J. Zhang, Z. Liu, J. Wang, X. Li, and Y. Yang (2010). Overexpression of udp-glucosyltransferase 71c5 increases glucose tolerance in arabidopsis thaliana. *Russian Journal of Plant Physiology 57*(6), 875–881.

Yoon, J., L.-H. Cho, W. Tun, J.-S. Jeon, and G. An (2021). Sucrose signaling in higher plants. *Plant Science 302*, 110703.

Zhang, H., W. Liang, X. Yang, X. Luo, N. Jiang, H. Ma, and D. Zhang (2010). Carbon starved anther encodes a myb domain protein that regulates sugar partitioning required for rice pollen development. *The Plant Cell 22*(3), 672–689.

Zhang, H. and T. Yin (2020). Analysis of topology properties in different tissues of poplar based on gene co-expression networks. *Tree Genetics & Genomes 16*(1), 1–12.

Zhang, J., Z. Ma, Q. Sun, and J. Yan (2018). Research review on algorithms of community detection in complex networks. In *Journal of Physics: Conference Series*, Volume 1069, pp. 012124. IOP Publishing.

Zhang, J., X. Zhang, H. Tang, Q. Zhang, X. Hua, X. Ma, F. Zhu, T. Jones, X. Zhu, J. Bowers, et al. (2018). Allele-defined genome of the autopolyploid sugarcane saccharum spontaneum l. *Nature genetics 50*(11), 1565.

Zhang, Q., X. Hua, H. Liu, Y. Yuan, Y. Shi, Z. Wang, M. Zhang, R. Ming, and J. Zhang (2020). Evolutionary expansion and functional divergence of sugar transporters in saccharum (s. spontaneum and s. officinarum). *The Plant Journal*.

Zhong, D., C. Wu, J. Bai, D. Xu, X. Zeng, and Q. Wang (2020). Co-expression network analysis reveals the pivotal role of mitochondrial dysfunction and interferon signature in juvenile dermatomyositis. *PeerJ 8*, e8611.

Zhou, Q.-Z., P. Fu, S.-S. Li, C.-J. Zhang, Q.-Y. Yu, C.-Z. Qiu, H.-B. Zhang, and Z. Zhang (2020). A comparison of co-expression networks in silk gland reveals the causes of silk yield increase during silkworm domestication.

Zhou, X., M. C. J. Kao, and W. H. Wong (2002). Transitive functional annotation by shortest-path analysis of gene expression data. *Proceedings of the National Academy of Sciences of the United States of America 99*(20), 12783–12788.

Zhu, X., C. Zhang, W. Wu, X. Li, C. Zhang, and J. Fang (2017). Enzyme activities and gene expression of starch metabolism provide insights into grape berry development. *Horticulture Research 4*(1), 1–16.

Zinkgraf, M., L. Liu, A. Groover, and V. Filkov (2017). Identifying gene coexpression networks underlying the dynamic regulation of wood-forming tissues in Populus under diverse environmental conditions. *New Phytologist 214*(4), 1464–1478.

11 Attachment

11.0.1 CIBio declaration

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA, www.ib.unicamp.br
Rua Monteiro Lobato, 255 CEP: 13083-862 Campinas-SP

DECLARAÇÃO

Em observância ao **§5º do Artigo 1º da Informação CCPG-UNICAMP/001/15**, referente a Bioética e Biossegurança, declaro que o conteúdo de minha Tese de Mestrado, intitulada:
Gene coexpression networks provide deeper insights into sucrose accumulation in sugarcane cultivars / Redes de coexpressão de genes fornecem informações mais aprofundadas sobre o acúmulo de sacarose em cultivares de cana-de-açúcar.
Desenvolvida no Programa de Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia da Unicamp, não versa sobre pesquisa envolvendo seres humanos, animais ou temas afetos a Biossegurança.

Campinas 25 de janeiro de 2022

Assinatura:
Nome do(a) autor(a) : **Marishani Marin Carrasco**
RG nº: F133072-0

Assinatura:
Nome do(a) orientador(a) : **Anete Pereira de Souza**
RG nº: 8680325-6

11.0.2 Authorship declaration

ANEXO II – DECLARAÇÃO DE AUTORIA

Declaração

As cópias de artigos de minha autoria ou de minha coautoria, publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **Redes de coexpressão gênica, fornecem informações mais profundas sobre acúmulo de açúcar em cultivares de cana-de-açúcar**. Não infringem o disposto na Lei nº 9.610/98 nem o direito autoral de qualquer editora.

Campinas 25 de janeiro de 2022.

Assinatura:
Nome do(a) autor(a) : **Marishani Marin Carrasco**
RG nº: F133072-0

Assinatura:
Nome da orientadora: **Anete Pereira de Souza**
RG nº: 8680325-6