University of São Paulo
"Luiz de Queiroz" College of Agriculture

Dissecting expression patterns in the transcriptome of immature sugarcane culms: from methodology to biology

Victor Hugo de Mello Pessoa

Dissertation presented to obtain the degree of Master in Science. Area: Genetics and Plant Breeding

Piracicaba
2020

Victor Hugo de Mello Pessoa
Bachelor of Physical and Biomolecular Sciences

Dissecting expression patterns in the transcriptome of immature sugarcane culms: from methodology to biology

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. PhD. **GABRIEL RODRIGUES ALVES MARGARIDO**

Dissertation presented to obtain the degree of Master in Science. Area: Genetics and Plant Breeding

Piracicaba
2020

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Pessoa, Victor Hugo de Mello

Dissecting expression patterns in the transcriptome of immature sugarcane culms: from methodology to biology / Victor Hugo de Mello Pessoa. – – versão revisada de acordo com a resolução CoPGr 6018 de 2011. – – Piracicaba, 2020 .

86 p.

Dissertação (Mestrado) – – USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Saccharum 2. Partição de carbono 3. Expressão diferencial 4. Réplicas biológicas 5. RNA-Seq . I. Título.

DEDICATÓRIA

Qualquer ser humano que passou pela pós-graduação sabe que, a despeito do que reforçam os setores interessados em desinformação e ausência de pensamento crítico, a atividade científica é séria e essencial para manutenção da sociedade. Nosso trabalho nunca para, porque perpassa nossas vidas em todas as esferas: dentro e fora do laboratório, no bandejão, em casa, em família, com amigos, nas horas de lazer, no banho e similares, muitas vezes nos impedindo de viver uma vida "normal". Não é raro que a pressão torne este período ainda mais difícil do que é. Esta dedicatória se destina a todos que o reconhecem e fornecem os meios para que essa importante profissão seja defendida e difundida, acessível a todes. Dentre estes, destina-se em especial para os que contribuíram para o meu processo.

À Michelle, que lutou muito para que eu estivesse aqui desde que nasci. Você sempre me inspirou com seu exemplo e com seus sonhos, e felizmente incentivou muitas pessoas ao seu redor. É um privilégio imenso ter sido educado com tanto afeto e compreensão. Ao mesmo tempo é um desafio, porque ser crítico passa por discordar, debater, ficar de mal e fazer as pazes. Nisso, eu aprendi com uma especialista. Ah, e ainda te lembro: este documento é um passo mais próximo para o nosso acordo.

À minha família que me defende a unhas e dentes, que me surpreende e que me apoia em minhas (in)decisões. Sempre esteve ao meu lado, presencialmente sempre que possível, mas também compreensíveis quanto a minha ausência. Me encorajou e confiou em mim ainda que eu não confiasse. Logicamente, se estende a minha família consanguínea ou não.

Aos meus amigos de laboratório (inclusive os longínquos agregados que nos visitaram), dentre os quais estendo aos laboratórios adotivos que tentaram me roubar mais de uma vez. Também se estende aos caros roomies, que nutriram minha saúde mental. Me mostraram que às vezes é normal confundir caquis e carambolas, mesmo que você estude muito botânica. Vocês me nortearam, ensinaram, quase sempre sem perceber. Me proporcionaram aprendizados culturais, artísticos, técnicos e humorísticos. Neste último ponto, me ajudaram a regredir também, o que também é um progresso.

Aos meus companheiros de viagem das cidades por onde passei (ou que passaram por mim), que me deixam com saudades constantemente e me fazem prometer a mim mesmo que irei voltar para visitar. Vocês nunca me deixaram ficar só de verdade, mesmo eu sendo relapso. Sei que tenho muitas casas para onde posso correr em caso de um vírus apocalíptico.

Aos que construíram gerações de conhecimento antes de mim e aos que virão. A ciência se faz com trabalho coletivo.

E à Pandora e à Sassá.

ACKNOWLEDGEMENTS

I thank all the people that made this work viable, mainly my supervisor Gabriel Margarido, because of his efforts to clarify my doubts and to create a friendly environment for developing scientific criticism as well as my own research pathways. I also have a special acknowledgment to my laboratory friends and researchers from UFSCar Araras, which have performed an excellent experimental job with the sugarcane plants, providing a flawless dataset for the analysis I performed. I express my gratitude to the Graduate Program of Genetics and Plant Breeding and the USP teachers who have contributed to my scientific formation, besides everyone who encouraged me and allowed the development of university extension activities. Finally, I recognize the funding agencies CAPES and FAPESP, which promoted the scholarship maintenance, and to CNPq, which, together with FAPESP, have funded the projects wherein I participated.

CONTENTS

RESUMO

Dissecando padrões de expressão no transcriptoma de colmos imaturos de cana-de-açúcar: da metodologia à biologia

O genoma da cana-de-açúcar é, de várias formas, o mais complexo dentre as plantas cultivadas, devido à sua alta ploidia, heterozigosidade e histórico de eventos de hibridização. Apesar de esforços nos últimos anos para se construírem três referências genômicas por grupos distintos, estas sequências ainda representem uma informação incompleta sobre os genomas de cana-de-açúcar. A presente dissertação contempla duas análises centrais para explorar o transcriptoma de cana-de-açúcar, visando trazer informações biológicas sobre a expressão gênica em colmos imaturos bem como reflexões metodológicas para o planejamento de experimentos de expressão diferencial. O primeiro capítulo apresenta uma comparação metodológica de duas estratégias visando ilustrar a influência de réplicas biológicas para plantas propagadas vegetativamente, como é o caso da cana-de-açúcar. Estas análises compararam o uso de clones ao uso de um conjunto diverso de genótipos como componentes dos grupos contrastantes de amostras. Os resultados indicam que o uso de clones permitiu a detecção de um maior número de genes diferencialmente expressos, provavelmente incluindo genes de efetivo interesse entre genes induzidos em genótipos específicos. Por outro lado, o uso de genótipos diversos proporcionou menos genes diferencialmente expressos, mas com aparentemente maior proporção de genes biologicamente relevantes. Esta proposição foi corroborada tanto pelos resultados do enriquecimento funcional quanto pelo conjunto de genes detectados em comum pelas estratégias. O segundo capítulo apresenta uma investigação biológica sobre os mecanismos genéticos pelos quais ocorre a partição de carbono em colmos apicais, onde o processo de acúmulo de sacarose não está desenvolvido. Genes diferencialmente expressos foram identificados para metabolismo e transporte de sacarose, tais como os genes de sacarose sintase, invertases e transportador de sacarose. Entretanto, o fenômeno mais notável relativo à partição de carbono foi a biossíntese de componentes da parede celular. Estes estudos podem trazer novas perspectivas para pesquisas sobre genética de cana-de-açúcar, por apresentarem um conjunto de genes de interesse para o metabolismo de açúcares e fibra, bem como conduzindo a uma escolha consciente do delineamento experimental para análises de RNA-Seq.

Palavras-chave: Saccharum, Partição de carbono, Expressão diferencial, Réplicas biológicas, RNA-Seq

ABSTRACT

Dissecting expression patterns in the transcriptome of immature sugarcane culms: from methodology to biology

The sugarcane genome is, by all accounts, the most complex among the cultivated crops due to its high ploidy, heterozygosity and history of hybridization events. Despite substantial efforts in the past years to obtain three genomic references by different groups, these sequences still represent incomplete information about sugarcane genomes. The current master thesis presents two core analyses to explore the sugarcane transcriptome, with the goal of bringing both biological insights about gene expression in immature culms as well as methodological considerations for the planning of differential expression experiments. The first chapter presents a methodological comparison of two strategies aiming to illustrate the influence of biological replication for vegetatively propagated plants, such as sugarcane. These analyses compared the use of clones and a diverse set of genotypes as components of contrasting groups of samples. The results indicate that the use of clones yielded an increased number of differentially expressed genes, which likely include genes of actual biological interest amidst genotype-specific significant tests. On the other hand, the use of diverse genotypes provided fewer differentially expressed genes, but the proportion of biologically relevant genes was seemingly higher. This statement was supported by evidence from both functional enrichment tests as well as the set of shared genes detected between the strategies. The second chapter presents a biological inquiry about the genetic mechanics regarding carbon partitioning in apical culms, where the sucrose accumulation process has not yet unfolded. Differentially expressed genes were identified for sucrose metabolism and transport, such as sucrose synthase, invertases, and sucrose transporter. However, the most apparent phenomenon with regard to carbon partitioning was the biosynthesis of cell wall components. These studies could drive new insights into sugarcane genetic investigations, by providing a set of important genes for early fiber and sugar metabolism in sugarcane, as well as aid researchers in making a more careful choice of experimental design for RNA-Seq essays.

Keywords: Saccharum, Carbon partitioning, Differential expression, Biological replicates, RNA-Seq

GENERAL INTRODUCTION

Sugarcane is one of the most valuable crops worldwide due to its importance for sugar, ethanol, and, more recently, biomass production. However, sugarcane has a highly complex genome. Modern cultivars result from hybridization between *Saccharum officinarum*, a high-sugar domesticated grass, and *S. spontaneum*, a fibrous plant tolerant to a broad range of biotic and abiotic stresses. Both parental species are autopolyploids (Bremer, 1925; Panje & Babu, 1960) — *S. officinarum* likely has eight sets of ten chromosomes (D'Hont *et al.*, 1998), while the number of chromosomes in *S. spontaneum* ranges from 40 to 128 —, such that sugarcane hybrids are auto-allopolyploids with frequent aneuploidy. Moreover, the number of chromosomes varies among genotypes (Piperidis *et al.*, 2010), as well as within the same genotype (D'Hont *et al.*, 1996).

Among the species of the genus *Saccharum*, *S. spontaneum* includes accessions with the most variable morphological features and chromosome number, as well as with a broader geographic span, from Northeast Africa to the Pacific Islands. Ecological and morphological adaptations allowed the species to thrive in widely diverse habitats. For instance, plants exhibit a great variation in size, ratooning capacity, amount of juice, stalk color, as well as tolerance to growing in dry soil or submerged into river waters (Mary *et al.*, 2006). When compared to modern cultivars, it has lower sucrose content, higher fiber yield, increased ratooning performance, and thinner stalks. Its high tolerance to biotic and abiotic stresses led to efforts of hybridization at the end of the 19th century, to introgress these traits to the high-sugar *S. officinarum*.

The narrow genetic base of the hybrids from the first decades of the 20th century has driven to issues such as the need to introgress new traits and the decreased rate of genetic gain, which led to new attempts of crossing with the parental species. However, a study using molecular markers showed that the diversity currently captured by breeding programs is still low, when considering the contribution of *S. spontaneum* (Aitken *et al.*, 2018). The use of new accessions as sources of alleles can be particularly relevant for introducing desirable traits for improving the so-called energy cane, given the growing allocation of sugarcane resources to the production of ethanol.

Besides classical plant breeding programs, both public and private initiatives have been developing genetically modified sugarcane aiming to tackle several agronomical issues. Transgenic sugarcane harboring genes for resistance to insects (Gao *et al.*, 2016; Cristofoletti *et al.*, 2018) and viruses (Yao *et al.*, 2017), for conferring drought tolerance (Zhao *et al.*, 2020), and for increasing sugar yield (Anur *et al.*, 2020) are examples of these. Yet, the lack of detailed information about the sugarcane genome hinders the understanding of how molecular mechanisms happen and can be leveraged for sugarcane breeding.

Genomic or transcriptomic assays are two approaches that can be used to acquire data for this purpose. Genomic studies are particularly complex to be performed for sugarcane, due to the high ploidy numbers and heterozygosity. There are three major scientific studies for presenting a comprehensive view of *Saccharum* genomes. The hybrids R570 and SP80-3280, as well as the haploid *S. spontaneum* accession AP85-441, had their genome sequenced using different methodologies, providing our best knowledge of the sugarcane genome to date (Garsmeur *et al.*, 2018; Souza *et al.*, 2019; Zhang *et al.*, 2018). Despite the advance brought by these studies, they still represent incomplete assemblies and pose obstacles when used as references. On the other hand, transcriptomic studies can be performed with fewer complications when compared to genomic ones. In particular for RNA-Seq analyses in non-model species, the currently used de novo assemblers do not necessarily require genomic references and are able to deal with polymorphisms present in different alleles (at least for highly expressed genes).

One suitable approach for detecting genes involved in biological processes is using differential expression analyses to identify up and downregulated genes in comparisons of interest. For instance, this strategy has been used to identify genes related to sugar yield (Papini-Terzi *et al.*, 2009; Thirugnanasam-

bandam *et al.*, 2017), fiber content (Vicentini *et al.*, 2015), drought stress (Li *et al.*, 2016), and resistance to smut (Rody *et al.*, 2019). Because gene expression is cell, tissue, and organ-dependent, the part of the plant chosen to be sampled provides data to answer different biological questions. Historically, sucrose yield is at the spotlight of studies regarding not only gene expression, but also enzymatic activity, plant physiology, and cellular biochemistry. However, these studies frequently focus on mature internodes or immature internodes at a late developmental stage. The current dissertation comprises two projects using sampled immature internodes in the earliest stage of development, which is right below the apical meristem.

The RNA-Seq data used as input to identify differentially expressed genes often fits into three sources of comparison: different tissues (organs) of the same plant; plants under different treatment levels; or plants with different genotypes, often selected based on their phenotypes. The first two examples do not necessarily depend on the choice of genotype, because the genomic composition of the contrasting groups is often identical. On the other hand, the latter case has a marginal effect of the combination of genotypes, regardless of the experimental design, such as in comparisons of high versus low sugar genotypes or susceptible versus tolerant plants to a pathogen. Also, especially for sugarcane and other vegetatively propagated crops, the use of clones or elite lines as biological replicates is a frequently adopted sampling strategy for the comparison of expression patterns. Another equally valid strategy and also used in gene expression assays is selecting different genotypes, grouped by a shared phenotypic trait. In chapter one, we analyze these different sampling strategies for biological replication, in which the group of interest is formed by clones or by a diverse set of genotypes (Figure 1). This study aims to compare the outcomes of differential expression analyses corresponding to these strategies and to provide a reference for the experimental design of future researches using RNA-Seq data under these conditions.

Little is known about the immature internode +1, which is the youngest part of sugarcane stalk, and remains largely unexplored with regard to its transcriptome. Chapter two presents an investigation of expression patterns in this organ for plants contrasting in sucrose levels, presenting a set of differentially expressed genes and enriched biological functions. Here, we identified putative markers for carbon partitioning before the start of sucrose accumulation.

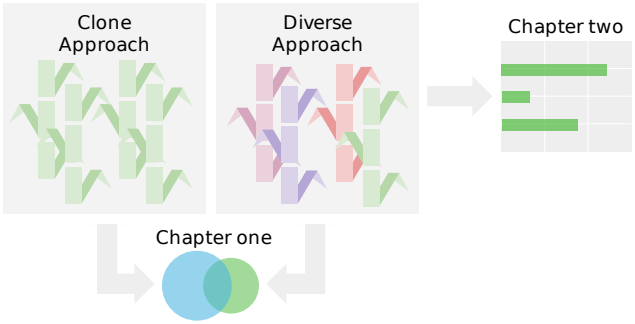

Figure 1: Scheme representing the use of biological data in the analyses performed in chapters one and two.

REFERENCES

Aitken, K., Li, J., Piperidis, G., Qing, C., Yuanhong, F., & Jackson, P. (2018). Worldwide genetic diversity of the wild species Saccharum spontaneum and level of diversity captured within sugarcane breeding programs. *Crop Science, 58*(1), 218-229.

Anur, R. M., Mufithah, N., Sawitri, W. D., Sakakibara, H., & Sugiharto, B. (2020). Overexpression of Sucrose Phosphate Synthase Enhanced Sucrose Content and Biomass Production in Transgenic Sugarcane. *Plants, 9*(2), 200.

Bremer, G. (1925). The cytology of the sugarcane. *Genetica, 7*(3), 293-322.

Cristofoletti, P. T., Kemper, E. L., Capella, A. N., Carmago, S. R., Cazoto, J. L., Ferrari, F., ... & Santos, N. Z. (2018). Development of transgenic sugarcane resistant to sugarcane borer. *Tropical Plant Biology, 11*(1-2), 17-30.

D'Hont, A., Grivet, L., Feldmann, P., Glaszmann, J. C., Rao, S., & Berding, N. (1996). Characterisation of the double genome structure of modern sugarcane cultivars (Saccharum spp.) by molecular cytogenetics. *Molecular and General Genetics MGG, 250*(4), 405-413.

D'Hont, A., Ison, D., Alix, K., Roux, C., & Glaszmann, J. C. (1998). Determination of basic chromosome numbers in the genus Saccharum by physical mapping of ribosomal RNA genes. *Genome, 41*(2), 221-225.

Gao, S., Yang, Y., Wang, C., Guo, J., Zhou, D., Wu, Q., ... & Que, Y. (2016). Transgenic sugarcane with a cry1Ac gene exhibited better phenotypic traits and enhanced resistance against sugarcane borer. *PLoS One, 11*(4), e0153929.

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., ... & Costet, L. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature communications, 9*(1), 2638.

Li, C., Nong, Q., Solanki, M. K., Liang, Q., Xie, J., Liu, X., ... & Li, Y. (2016). Differential expression profiles and pathways of genes in sugarcane leaf at elongation stage in response to drought stress. *Scientific reports, 6*, 25698.

Mary, S., Nair, N. V., Chaturvedi, P. K., & Selvi, A. (2006). Analysis of genetic diversity among Saccharum spontaneum L. from four geographical regions of India, using molecular markers. *Genetic Resources and Crop Evolution, 53*(6), 1221-1231.

Panje, R. R., & Babu, C. N. (1960). Studies in Saccharum spontaneum distribution and geographical association of chromosome numbers. *Cytologia, 25*(2), 152-172.

Papini-Terzi, F. S., Rocha, F. R., Vêncio, R. Z., Felix, J. M., Branco, D. S., Waclawovsky, A. J., ... & Vicentini, R. (2009). Sugarcane genes associated with sucrose content. *BMC genomics, 10*(1), 120.

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

Rody, H. V., Bombardelli, R. G., Creste, S., Camargo, L. E., Van Sluys, M. A., & Monteiro-Vitorello, C. B. (2019). Genome survey of resistance gene analogs in sugarcane: genomic features and differential expression of the innate immune system from a smut-resistant genotype. *BMC genomics, 20*(1), 809.

Souza, G. M., Van Sluys, M. A., Lembke, C. G., Lee, H., Margarido, G. R. A., Hotta, C. T., ... & Nishiyama Jr, M. Y. (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.

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

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

Yao, W., Ruan, M., Qin, L., Yang, C., Chen, R., Chen, B., & Zhang, M. (2017). Field performance of transgenic sugarcane lines resistant to sugarcane mosaic virus. *Frontiers in Plant Science, 8*, 104.

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., ... & Wai, C. M. (2018). Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. *Nature genetics, 50*(11), 1565-1573.

Zhao, X., Jiang, Y., Liu, Q., Yang, H., Wang, Z., & Zhang, M. (2020). Effects of drought-tolerant Ea-DREB2B transgenic sugarcane on bacterial communities in soil. *Frontiers in Microbiology, 11*, 704.

CHAPTER ONE: AN RNA-SEQ-BASED COMPARISON OF BIOLOGICAL REPLICATION STRATEGIES FOR DIFFERENTIAL GENE EXPRESSION IN *SACCHARUM*

1 INTRODUCTION

The genus *Saccharum* comprises six species, of which *S. spontaneum* and *S. robustum* are the only wild representatives, spread over a large area in Asia and Indonesia, and the others are domesticated species — *S. officinarum*, *S. barberi*, *S. sinense*, and *S. edule*. Modern sugarcane cultivars are mainly descendant from the crossing of *S. officinarum* and *S. spontaneum*, followed by backcrossing to *S. officinarum*, such that they inherit the high sugar yield from the former species and the pathogen resistance, adaptability, and increased vigor of the latter (Irvine, 1999; Piperidis *et al.*, 2010). Sugarcane cultivation accounts for 86% of the worldwide production of sugar, despite the increasing allocation of its juice for ethanol production. Moreover, the sugarcane residue after juice extraction, called bagasse, is a byproduct that can be used for energy generation and production of bioplastics (OECD/FAO, 2019; Aguilar *et al.*, 2019). The crop is a renewable source of fuel and presents a significant advantage over fossil fuels due to the reduced emission of greenhouse gases (Goldemberg, 2008).

Sugarcane breeding programs usually rely on a few recurrent crosses between elite parents or wild germplasm to produce genotypes with desired traits, such as sugar or fiber yield and resistance to abiotic and biotic stresses (Heinz & Tew, 1987; de Souza Barbosa *et al.*, 2002; Jackson, 2005). As a consequence, scientific investigations in sugarcane are often based on elite lines as a source of plant material, such as two genome assemblies for the hybrids R570 — a major model in sugarcane genomic studies — and SP80-3280 (Garsmeur *et al.*, 2018; Souza *et al.*, 2019).

Also, these hybrids show a large variation in chromosome number and genome constitution. *S. officinarum* (2n = 8x = 80) and *S. spontaneum* (2n = 40-128), the parental species, have high levels of ploidy and complex genomes *per se* (Bremer, 1925; Panje & Babu, 1960). Chromosome number multiplicity and molecular evidence have led to the acceptance of the basic number of x = 8 for *S. spontaneum* (Liu *et al.*, 2016); however, the description of a wild accession with x = 10 brought a new panorama to the evolutionary history of the genus (Meng *et al.*, 2020). These facts reveal an intricate set of hurdles concerning the understanding of sugarcane genomics, which must be considered for data-driven experiments.

More specifically, the use of phenotypic trait variation between genotypes is a common approach found in differential expression studies. In the literature of sugarcane gene expression research, there are analyses conducted with a single genotype representing each group of interest (Casu *et al.*, 2007; Papini-Terzi *et al.*, 2009; Casu *et al.*, 2015; Vicentini *et al.*, 2015; Dharshini *et al.*, 2016), as well as with multiple genotypes per group (Papini-Terzi *et al.*, 2009; Ferreira *et al.*, 2016; Thirugnanasambandam *et al.*, 2017; Hoang *et al.*, 2017). Biological replicates provide more accurate estimates of transcript abundances when comparing samples from two treatment levels. Clones from the same genotype are subject to variability in their expression levels due to factors such as interactions with the environment and other organisms. Still, the transcriptional variation within clones is expected to be smaller when compared to plants from different genotypes, which increases the dispersion of gene quantification estimates. Statistical parameters such as means of expression levels and their residual variances are the main variables considered in modern differential expression tests, which highlights the relevance of the choice of approach for performing these studies. While the use of clones renders a more homogeneous set of samples, and consequently more statistical power to detect differences of expression between groups, it also restrains the set of samples to a limited number of genotypes.

Here, we evaluate the influence of using clonal replicates or multiple genotypes in differential gene expression analysis between contrasting groups. The comparison of approaches we propose relies both on quantitative estimates of differentially expressed genes and qualitative functional enrichment tests. We aim to present an information-based criterion for selecting biological replicates for further experiments using RNA-Seq, particularly for sugarcane, whose genomic properties can deviate dramatically among genotypes.

2 CONCLUSIONS

With the increasing application of next generation sequencing to investigate complex transcriptomes, such as that of sugarcane, recent studies aim to apply these techniques to unravel the molecular mechanisms controlling several phenotypic traits. However, a single biological replicate in each contrasting group is not enough for performing this sort of analysis, leaving for the researcher the choice of a suitable experimental design. Our present study intended to illustrate the strengths and caveats inherent to two sampling strategies for biological replication, namely by using a diverse group of genotypes with common phenotypic characteristics or clones from the same genotype, chosen to be representative of this group. The results have provided evidence of discrepancies in *(i)* quantitative terms, regarding the number of genes detected as differentially expressed, *(ii)* consistency, when subjected to self-validation using subsampling, and *(iii)* inferred biological conclusions from the functional annotation of differentially expressed genes. These analyses suggest that the use of clones as biological replicates may yield somewhat restricted results, biased by the particular choice of genotypes. Regardless of these concerns, the direct comparison of two genotypes can still be useful in particular situations. On the other hand, the presence of a representative set of genotypes within the same group can lead to more reasonable biologic outcomes. In any case, it is possible to combine these strategies to refine the level of details, if economically viable. This research offers support to the experimental design of new studies using differential expression as a method of investigation in sugarcane and other plants with high genomic complexity.

REFERENCES

Aguilar, N. M., Arteaga-Cardona, F., de Anda Reyes, M. E., Gervacio-Arciniega, J. J., & Salazar-Kuri, U. (2019). Magnetic bioplastics based on isolated cellulose from cotton and sugarcane bagasse. *Materials Chemistry and Physics, 238*, 121921.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology, 215*(3), 403-410.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological), 57*(1), 289-300.

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

Bremer, G. (1925). The cytology of the sugarcane. *Genetica, 7*(3), 293-322.

Casu, R. E., Jarmey, J. M., Bonnett, G. D., & Manners, J. M. (2007). Identification of transcripts associated with cell wall metabolism and development in the stem of sugarcane by Affymetrix GeneChip Sugarcane Genome Array expression profiling. *Functional & integrative genomics, 7*(2), 153-167.

Casu, R. E., Rae, A. L., Nielsen, J. M., Perroux, J. M., Bonnett, G. D., & Manners, J. M. (2015). 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.

de Souza Barbosa, G. V., de Menezes Cruz, M., Soares, L., Rocha, A. M. C., Ribeiro, C. A. G., Sousa, A. J. R., ... & dos Santos, A. V. P. (2002). A brief report on sugarcane breeding program in Alagoas, Brazil. *Crop Breeding and Applied Biotechnology, 2*(4).

Dharshini, S., Chakravarthi, M., Manoj, V. M., Naveenarani, M., Kumar, R., Meena, M., ... & Appunu, C. (2016). De novo sequencing and transcriptome analysis of a low temperature tolerant Saccharum spontaneum clone IND 00-1037. *Journal of biotechnology, 231*, 280-294.

D'Hont, A., Grivet, L., Feldmann, P., Glaszmann, J. C., Rao, S., & Berding, N. (1996). Characterisation of the double genome structure of modern sugarcane cultivars (Saccharum spp.) by molecular cytogenetics. *Molecular and General Genetics MGG, 250*(4), 405-413.

Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. In *Genome Informatics 2009: Genome Informatics Series Vol. 23* (pp. 205-211).

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

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., ... & Costet, L. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature communications, 9*(1), 1-10.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Chen, Z. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology, 29*(7), 644.

Goldemberg, J. (2008). The Brazilian biofuels industry. *Biotechnology for biofuels, 1*(1), 6.

Heinz, D. J., & Tew, T. L. (1987). Hybridization procedures. In *Developments in crop science* (Vol. 11, pp. 313-342). Elsevier.

Hoang, N. V., Furtado, A., O'Keeffe, A. J., Botha, F. C., & Henry, R. J. (2017). Association of gene expression with biomass content and composition in sugarcane. *PLoS One, 12*(8).

Irvine, J. E. (1999). Saccharum species as horticultural classes. *Theoretical and Applied Genetics, 98*(2), 186-194.

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

Liu, X., Li, X., Liu, H., Xu, C., Lin, X., Li, C., & Deng, Z. (2016). Phylogenetic analysis of different ploidy Saccharum spontaneum based on rDNA-ITS sequences. *PloS one, 11*(3).

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal, 17*(1), 10-12.

Meng, Z., Han, J., Lin, Y., Zhao, Y., Lin, Q., Ma, X., ... & Wang, K. (2020). Characterization of a Saccharum spontaneum with a basic chromosome number of x= 10 provides new insights on genome evolution in genus Saccharum. *Theoretical and Applied Genetics, 133*(1), 187-199.

Panje, R. R., & Babu, C. N. (1960). Studies in Saccharum spontaneum distribution and geographical association of chromosome numbers. *Cytologia, 25*(2), 152-172.

Papini-Terzi, F. S., Rocha, F. R., Vêncio, R. Z., Felix, J. M., Branco, D. S., Waclawovsky, A. J., ... & Vicentini, R. (2009). Sugarcane genes associated with sucrose content. *BMC genomics, 10*(1), 120.

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

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31*(19), 3210-3212.

Souza, G. M., Van Sluys, M. A., Lembke, C. G., Lee, H., Margarido, G. R. A., Hotta, C. T., ... & Nishiyama Jr, M. Y. (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.

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

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

CHAPTER TWO: IDENTIFYING EXPRESSION PROFILES ASSOCIATED WITH DIVERSE CARBON PARTITIONING PHENOTYPES IN YOUNG SUGARCANE STEMS

1 INTRODUCTION

Sugarcane cultivation has a large contribution to Brazilian agribusiness in terms of production, economic importance, and cultivated area. Its relevance is mainly due to the use of culms to produce sugar and ethanol. Brazil has over 8 million ha cultivated with sugarcane, or roughly 1% of the total land area, with an estimated production of 642.7 million tons in 2019/2020 (`https://www.conab.gov.br/`). It is thus the major worldwide producer of this crop. Using vegetable biomass from sugarcane represents an alternative source of biofuel generation, which has a lower environmental impact than fossil fuels, due to carbon sequestration. Therefore, concerns involving environmental pollution and long-term fuel reserves are responsible for the interest in using fuels from renewable energy sources. This fact drives genetic breeding of this crop aiming to increase production with phenotypes of agronomic and economic interest, such as drought resistance, sugar yield, biomass content, high tillering, and ratooning (Rooney *et al.*, 2007; Morris *et al.*, 2013).

The complexity of the sugarcane genome is one of the main challenges for performing computational analyses based on molecular information. Panje & Babu (1960) performed an extensive manual characterization of the number of chromosomes in more than three hundred *S. spontaneum* accessions, concluding that the diploid number varies from 40 to 128 for this species. For *S. officinarum*, the karyotype 2n = 80 is found among most of the accessions since the first evaluation by Bremer (1925), with the exception of aneuploid variation from atypical clones (Heinz, 1991). The previous interspecific hybridization events between *S. spontaneum* and *S. officinarum* aimed to create crops with high sugar yield in the culms and resistance to pathogens. From a genomic standpoint, these hybrids resulted in polyploid and aneuploid crops with 2n = 100-130, with chromosomes from both parental species as well as recombinants (D'Hont *et al.*, 1996). High ploidy levels increase the number of gene copies, which can lead to the emergence of new alleles by sequence divergence, pseudogenes by loss of function, and neofunctionalization in exceptional cases. Aneuploidy is also a determinant factor on phenotypic variation, because genic dosage ratios are affected for genes in different chromosomes (Makarevitch & Harris, 2010).

These obstacles have hindered advances into understanding the roles of key genes related to physiological characteristics of sugarcane, such as sucrose accumulation, one of its most important agronomic traits. The molecular mechanisms of sucrose concentration in mature sugarcane culms depend on sucrose balance and availability, passing through several biological processes, such as photosynthesis, cellular growth, respiration, and sugar transport. Sugar accumulation in sugarcane works differently from most of the plants, because its storage carbohydrate is sucrose, instead of complex and insoluble polysaccharides, such as starch, and it occurs in stem parenchymal tissue (McCormick *et al.*, 2008 b). Sacher *et al.* (1963) described that sucrose undergoes conversion to hexoses by acid invertases acting at the apoplast of parenchyma cells in young culms. By doing so, these enzymes can control the uptake of sugars for posterior degradation or resynthesis for accumulation. Moore (1995) evaluated the metabolite dynamics through each compartment — apoplast, cytosol, and vacuole of storage parenchyma cells in culms. In his work, the main hypothesis is that the maintenance of low turgor gradient of solutes could regulate the transport of metabolites, mediated by a turgor-sensing system, which has been endorsed by Wang *et al.* (2013). However, Wu *et al.* (2007) were able to develop transgenic sugarcane carrying a gene for sucrose isomerase (SI) that can double its sugar content by converting sucrose to isomaltulose, with no decrease in sucrose concentration and increased photosynthetic rates. In this experiment, the authors showed that signaling pathways in the source-sink system can regulate photosynthetic activity on leaves

and that osmotic restraints are not the main limiters to the sucrose accumulation process.

Studies indicate invertases as key regulators of sugar levels in sugarcane culms (Wang *et al.*, 2013). Three invertases synthesized in storage parenchyma — cell wall bound invertase (CWI), neutral invertase (NI), and soluble acid invertase (SAI) — are redirected to different cell compartments and prevail in different stages of culm maturation (Wang *et al.*, 2013). Alongside with the invertases, sucrose synthase (SuSy) and sucrose phosphate synthase (SPS) have been identified as putative markers of sucrose accumulation. SuSy catalyzes a reversible reaction for cleaving sucrose into fructose and UDP-glucose. The breakdown/synthesis ratio of SuSy increases with internode maturity, with activity in young internodes mainly in the synthetic direction (Schäfer *et al.*, 2004). Lower sucrose concentration in immature culms also corroborates these results, supporting sucrose synthesis instead of breakdown by SuSy. Overall sucrose synthesis occurs by SuSy and SPS catalysis (Botha & Black, 2000). However, other factors might be related to sucrose levels in sugarcane, such as redirecting of UDP-glucose to cell wall synthesis by cellulose synthase A (CesA) complexes and intricate regulation of sugar trafficking through the phloem, apoplast, storage parenchyma cytoplasm, and vacuole, via apoplasmic or symplasmic paths (Casu *et al.*, 2015).

RNA-Seq and EST sequencing projects have been performed in sugarcane to understand genes in terms of their tissue-specificity, to determine differences between mature and immature culms, and to detect genes affected by abiotic stresses. Several of these studies also combined sequencing and hybridization approaches to identify genes of interest and estimate the abundance of their corresponding transcripts (de Araujo *et al.*, 2005; Rae *et al.*, 2005; Papini-Terzi *et al.*, 2009; Casu *et al.*, 2015; Thirugnanasambandam *et al.*, 2019). Previous studies, using sugar-contrasting genotypes at distinct levels of stem maturity, reported several classes of genes as differentially expressed. Genes related to signaling such as kinases, phosphatases, and transcription factors, and to cell wall biosynthesis, as well as SuSy, SPS, and bidirectional sugar transporters (SWEETs) were either up or downregulated in high sugar genotypes (Papini-Terzi *et al.*, 2009; Thirugnanasambandam *et al.*, 2017). Moreover, genes whose expression was affected by abiotic stresses, such as drought, were also related to sugar accumulation processes in the case of abscisic acid (ABA) signaling and biosynthesis, for example (Papini-Terzi *et al.*, 2009). This observation may be due to the fact that sugarcane shows increased sugar levels in response to abiotic stresses. Genotypes contrasting in lignin content also exhibited a consistent differential expression pattern of genes in the phenylpropanoid biosynthetic pathway and cell wall proteins (Vicentini *et al.*, 2015). Regardless of efforts to identify genes associated with sucrose accumulation in sugarcane culms, little is known about how carbon partitioning takes place in the apical section of its stalks.

In this work, we investigate gene expression in internode +1 of different sugarcane genotypes contrasting in sugar levels, representing a wide variety of phenotypes and origins of *Sachharum* accessions. Because this section of the culm is directly below the apical meristem, the sugar accumulation process has not yet taken place. The main biological activity in this section of the stem is the expanding of cell wall surface for cellular growth (Rose & Botha, 2000). We used RNA-seq data to analyze expressed transcripts in sugarcane to understand the mechanisms involved in carbon partitioning at an early stage of development. Gene expression studies using immature internodes often involve comparisons to other organs (Casu *et al.*, 2003, Casu *et al.*, 2004; Papini-Terzi *et al.*, 2013; de Barros Dantas *et al.*, 2020) or between culms at a later developmental stage and low levels of sucrose accumulation (Thirugnanasambandam *et al.*, 2017). However, to the best of our knowledge there is no study using RNA-Seq data to assess the expression patterns in internodes at the most immature stage of development. Internode +1 may provide useful information regarding the process of carbon partitioning in young sucrose storage cells. Using a diverse set of genotypes with a broad range of sugar yield, we could assess whether there was consistency in the expression patterns of sugar-related genes. Instead of using a single accession per group, this approach is suitable to reveal alternative routes to achieve similar phenotypes, because

we can evaluate genotypic specificities regarding the expression levels of sugar regulators. Here, our main purpose was to characterize putative key regulators of carbon partitioning that could initiate the phenotypical differentiation between high and low sugar sugarcane genotypes.

2 CONCLUSIONS

In this study, we investigated possible genetic reasons that explain sugarcane variation in terms of sucrose accumulation, which can reach the highest concentration known in a plant. More specifically, we focused on understanding how regulation of gene expression in apical culms could be responsible for differences in sugar storage in the target tissue. With the information acquired from the set of differentially expressed genes, we observed that several genes pointed as main regulators of sucrose yield were not significantly upregulated in the group of sugar-rich genotypes. Despite the complexity of factors that might contribute to this phenomenon, our analyses suggest that even at this early development stage, cellulose synthesis plays a vital role in the differentiation of sugarcane genotypes with lower and higher levels of sucrose and fiber. Also, SuSy2 was among the few differentially expressed genes responsible for sucrose synthesis/breakdown, and it might be useful to provide the required substrate for cellulose synthetic activity. The transport of sugars by transmembrane proteins was a noticeable process that distinguished samples with very low sugar concentration from the others. We also found evidence of a novel alternative splicing form in two SWEET genes expressed in the internode +1, which encodes a protein isoform with a different number of transmembrane helices. This discovery could bring new information to the sugar transport process in sugarcane. However, for a better glimpse of sugar transporters in general, it is necessary to acquire both tissue-specific expression patterns and cellular localization of these proteins to understand the exact trafficking of photoassimilates from source to sink. As a whole, gene expression patterns indicate a route for carbon partitioning in these young culms, provided by *i)* driving sucrose to the cytoplasm instead of apoplasm or vacuole; *ii)* breaking down of sucrose according to the metabolic demand; and *iii)* synthesizing cellulose for cell wall expansion. This molecular workflow may partially explain the relationship between fiber and sucrose yield that distinguishes *S. officinarum* and *S. spontaneum* phenotypes. Remarkably, our study pointed to several particularities possessed by each genotype in expression levels. This result simultaneously shows that sugar yield depends on multiple genes, and that genotypes with similar phenotypes might not have common grounds when it comes to expression profiles of well-established genes as the main regulators of the sugar accumulation process, at an early stage of development.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology, 215*(3), 403-410.

de Araujo, P. G., Rossi, M., de Jesus, E. M., Saccaro Jr, N. L., Kajihara, D., Massa, R., ... & Ulian, E. C. (2005). Transcriptionally active transposable elements in recent hybrid sugarcane. *The Plant Journal, 44*(5), 707-717.

Barreto, F. Z., Rosa, J. R. B. F., Balsalobre, T. W. A., Pastina, M. M., Silva, R. R., Hoffmann, H. P., ... & Carneiro, M. S. (2019). A genome-wide association study identified loci for yield component traits in sugarcane (Saccharum spp.). *PloS one, 14*(7), e0219843.

de Barros Dantas, L. L., Almeida-Jesus, F. M., de Lima, N. O., Alves-Lima, C., Nishiyama-Jr, M. Y., Carneiro, M. S., ... & Hotta, C. T. (2020). Rhythms of transcription in field-grown sugarcane are highly organ specific. *Scientific reports, 10*(1), 1-12.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological), 57*(1), 289-300.

Bihmidine, S., Baker, R. F., Hoffner, C., & Braun, D. M. (2015). 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.

Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C., & Hochreiter, S. (2015). msa: an R package for multiple sequence alignment. *Bioinformatics, 31*(24), 3997-3999.

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

Botha, F. C., & Black, K. G. (2000). Sucrose phosphate synthase and sucrose synthase activity during maturation of internodal tissue in sugarcane. *Functional Plant Biology, 27*(1), 81-85.

Botha, F. C., Sawyer, B. J. B., & Birch, R. G. (2001). Sucrose metabolism in the culm of transgenic sugarcane with reduced soluble acid invertase activity. In *Proceedings of the International Society of Sugar Cane Technologists* (Vol. 24, pp. 588-591).

Bremer, G. (1925). The cytology of the sugarcane. *Genetica, 7*(3), 293-322.

Casu, R. E., Dimmock, C. M., Chapman, S. C., Grof, C. P., McIntyre, C. L., Bonnett, G. D., & Manners, J. M. (2004). Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling. *Plant molecular biology, 54*(4), 503-517.

Casu, R. E., Grof, C. P., Rae, A. L., McIntyre, C. L., Dimmock, C. M., & Manners, J. M. (2003). Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. *Plant molecular biology, 52*(2), 371-386.

Casu, R. E., Rae, A. L., Nielsen, J. M., Perroux, J. M., Bonnett, G. D., & Manners, J. M. (2015). 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., Hou, B. H., Lalonde, S., Takanaga, H., Hartung, M. L., Qu, X. Q., ... & Chermak, D. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature, 468*(7323), 527-532.

Ding, S. Y., & Himmel, M. E. (2006). The maize primary cell wall microfibril: a new model derived from direct visualization. *Journal of Agricultural and Food Chemistry, 54*(3), 597-606.

D'Hont, A., Grivet, L., Feldmann, P., Glaszmann, J. C., Rao, S., & Berding, N. (1996). Characterisation of the double genome structure of modern sugarcane cultivars (Saccharum spp.) by molecular cytogenetics. *Molecular and General Genetics MGG, 250*(4), 405-413.

Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. In *Genome Informatics 2009: Genome Informatics Series Vol. 23* (pp. 205-211).

Endler, A., Meyer, S., Schelbert, S., Schneider, T., Weschke, W., Peters, S. W., ... & Schmidt, U. G. (2006). Identification of a vacuolar sucrose transporter in barley and Arabidopsis mesophyll cells by a tonoplast proteomic approach. *Plant Physiology, 141*(1), 196-207.

Eom, J. S., Chen, L. Q., Sosso, D., Julius, B. T., Lin, I. W., Qu, X. Q., ... & Frommer, W. B. (2015). SWEETs, transporters for intracellular and intercellular sugar translocation. *Current Opinion in Plant Biology, 25*, 53-62.

Eom, J. S., Cho, J. I., Reinders, A., Lee, S. W., Yoo, Y., Tuan, P. Q., ... & Bhoo, S. H. (2011). Impaired function of the tonoplast-localized sucrose transporter in rice, OsSUT2, limits the transport of vacuolar reserve sucrose and affects plant growth. *Plant Physiology, 157*(1), 109-119.

Fujii, S., Hayashi, T., & Mizuno, K. (2010). Sucrose synthase is an integral component of the cellulose synthesis machinery. *Plant and cell physiology, 51*(2), 294-301.

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., ... & Costet, L. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature communications, 9*(1), 2638.

Gazaffi, R., Cursi, D. E., Chapola, R. G., Santos, J. D., Fernandes, A. R., Carneiro, M. S., & Charoenwong, S. (2016). RB varieties: A major contribution to the sugarcane industry in Brazil. *Proceedings of the 29th International Society of Sugar Cane Technologists, Chang Mai, Thailand. Dec. 2016*, 1677-1682.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Chen, Z. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology, 29*(7), 644.

Hedrich, R., Sauer, N., & Neuhaus, H. E. (2015). Sugar transport across the plant vacuolar membrane: nature and regulation of carrier proteins. *Current opinion in plant biology, 25*, 63-70.

Heinz, D. J. (1991). Sugarcane cytogenetics. *Chromosome engineering in plants: Genetics, Breeding, Evolution, Part B. Elsevier, Amsterdam*, 279-293.

Hu, Y. B., Sosso, D., Qu, X. Q., Chen, L. Q., Ma, L., Chermak, D., ... & Frommer, W. B. (2016). Phylogenetic evidence for a fusion of archaeal and bacterial SemiSWEETs to form eukaryotic SWEETs and identification of SWEET hexose transporters in the amphibian chytrid pathogen Batrachochytrium dendrobatidis. *The FASEB Journal, 30*(10), 3644-3654.

Jacobsen, K. R., Fisher, D. G., Maretzki, A., & Moore, P. H. (1992). Developmental changes in the anatomy of the sugarcane stem in relation to phloem unloading and sucrose storage. *Botanica Acta, 105*(1), 70-80.

Jain, R., Singh, S. P., Singh, A., Singh, S., Kishor, R., Singh, R. K., ... & Solomon, S. (2017). Soluble Acid Invertase (SAI) Activity and Gene Expression Controlling Sugar Composition in Sugarcane. *Sugar Tech, 19*(6), 669-674.

Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., & Gao, G. (2016). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic acids research*, gkw982.

Khaled, S. K., Abdel-Tawab, F. M., Fahmy, E. M., Amer, E. A. M., & Khaled, K. A. (2018). The siRNA efficacy of soluble acid invertase down-regulation in Sugarcane (Saccharum spp.). *Arab Universities Journal of Agricultural Sciences, 26*(Special issue (2C)), 2011-2017.

Lembke, C. G., Nishiyama, M. Y., Sato, P. M., de Andrade, R. F., & Souza, G. M. (2012). Identification of sense and antisense transcripts regulated by drought in sugarcane. *Plant molecular biology, 79*(4-5), 461-477.

Lingle, S. E., & Irvine, J. E. (1994). Sucrose synthase and natural ripening in sugarcane. *Crop Science, 34*(5), 1279-1283.

Makarevitch, I., & Harris, C. (2010). Aneuploidy causes tissue-specific qualitative changes in global gene expression patterns in maize. *Plant physiology, 152*(2), 927-938.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal, 17*(1), 10-12.

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic acids research, 40*(10), 4288-4297.

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

McCormick, A. J., Cramer, M. D., & Watt, D. A. (2008). Regulation of photosynthesis by sugars in sugarcane leaves. *Journal of plant physiology, 165*(17), 1817-1829.

McCormick, A. J., Watt, D. A., & Cramer, M. D. (2008). Supply and demand: sink regulation of sugar accumulation in sugarcane. *Journal of experimental botany, 60*(2), 357-364.

Milne, R. J., Grof, C. P., & Patrick, J. W. (2018). Mechanisms of phloem unloading: shaped by cellular pathways, their conductances and sink function. *Current opinion in plant biology, 43*, 8-15.

Milne, R. J., Offler, C. E., Patrick, J. W., & Grof, C. P. (2015). Cellular pathways of source leaf phloem loading and phloem unloading in developing stems of Sorghum bicolor in relation to stem sucrose storage. *Functional plant biology, 42*(10), 957-970.

Mizuno, H., Kasuga, S., & Kawahigashi, H. (2016). The sorghum SWEET gene family: stem sucrose accumulation as revealed through transcriptome profiling. *Biotechnology for biofuels, 9*(1), 127.

Moore, P. H. (1995). Temporal and spatial regulation of sucrose accumulation in the sugarcane stem. *Functional Plant Biology, 22*(4), 661-679.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic acids research, 35*(suppl 2), W182-W185.

Morris, D. R., Gilbert, R. A., Rainbolt, C. R., Perdomo, R. E., Powell, G., Eiland, B., & Montes, G. (2007). Sugarcane yields and soil chemical properties due to mill mud applications to a sandy soil. In *Proc. Int. Soc. Sugar Cane Technol* (Vol. 26, pp. 444-448).

Panje, R. R., & Babu, C. N. (1960). Studies in Saccharum spontaneum distribution and geographical association of chromosome numbers. *Cytologia, 25*(2), 152-172.

Papini-Terzi, F. S., Rocha, F. R., Vêncio, R. Z., Felix, J. M., Branco, D. S., Waclawovsky, A. J., ... & Vicentini, R. (2009). Sugarcane genes associated with sucrose content. *BMC genomics, 10*(1), 120.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods, 14*(4), 417.

Poschet, G., Hannich, B., Raab, S., Jungkunz, I., Klemens, P. A., Krueger, S., ... & Büttner, M. (2011). A novel Arabidopsis vacuolar glucose exporter is involved in cellular sugar homeostasis and affects the composition of seed storage compounds. *Plant physiology, 157*(4), 1664-1676.

Rae, A. L., Perroux, J. M., & Grof, C. P. (2005). 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.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140.

Rooney, W. L., Blumenthal, J., Bean, B., & Mullet, J. E. (2007). Designing sorghum as a dedicated bioenergy feedstock. *Biofuels, Bioproducts and Biorefining, 1*(2), 147-157.

Rose, S., & Botha, F. C. (2000). Distribution patterns of neutral invertase and sugar content in sugarcane internodal tissues. *Plant Physiology and Biochemistry, 38*(11), 819-824.

Sacher, J. A., Hatch, M. D., & Glasziou, K. T. (1963). Sugar accumulation cycle in sugar cane. III. Physical & metabolic aspects of cycle in immature storage tissues. *Plant Physiology, 38*(3), 348.

Schäfer, W. E., Rohwer, J. M., & Botha, F. C. (2004). Protein-level expression and localization of sucrose synthase in the sugarcane culm. *Physiologia Plantarum, 121*(2), 187-195.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31*(19), 3210-3212.

Souza, G. M., Van Sluys, M. A., Lembke, C. G., Lee, H., Margarido, G. R. A., Hotta, C. T., ... & Nishiyama Jr, M. Y. (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.

Tanaka, K., Murata, K., Yamazaki, M., Onosato, K., Miyao, A., & Hirochika, H. (2003). Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. *Plant physiology, 133*(1), 73-83.

Tao, Y., Cheung, L. S., Li, S., Eom, J. S., Chen, L. Q., Xu, Y., ... & Feng, L. (2015). Structure of a eukaryotic SWEET transporter in a homotrimeric complex. *Nature, 527*(7577), 259-263.

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

Thirugnanasambandam, P. P., Mason, P. J., Hoang, N. V., Furtado, A., Botha, F. C., & Henry, R. J. (2019). Analysis of the diversity and tissue specificity of sucrose synthase genes in the long read transcriptome of sugarcane. *BMC plant biology, 19*(1), 160.

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

Wang, J., Nayak, S., Koch, K., & Ming, R. (2013). Carbon partitioning in sugarcane (Saccharum species). *Frontiers in plant science, 4*, 201.

Wang, L., Zheng, Y., Ding, S., Zhang, Q., Chen, Y., & Zhang, J. (2017). Molecular cloning, structure, phylogeny and expression analysis of the invertase gene family in sugarcane. *BMC plant biology, 17*(1), 109.

Wu, L., & Birch, R. G. (2007). Doubled sugar content in sugarcane plants modified to produce a sucrose isomer. *Plant biotechnology journal, 5*(1), 109-117.

Xu, Y., Tao, Y., Cheung, L. S., Fan, C., Chen, L. Q., Xu, S., ... & Feng, L. (2014). Structures of bacterial homologues of SWEET transporters in two distinct conformations. *Nature, 515*(7527), 448-452.

Yamada, K., Osakabe, Y., Mizoi, J., Nakashima, K., Fujita, Y., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2010). Functional analysis of an Arabidopsis thaliana abiotic stress-inducible facilitated diffusion transporter for monosaccharides. *Journal of Biological Chemistry, 285*(2), 1138-1146.

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology, 11*(2), R14.

Zhang, J., Sharma, A., Yu, Q., Wang, J., Li, L., Zhu, L., ... & Ming, R. (2016). Comparative structural analysis of Bru1 region homeologs in Saccharum spontaneum and S. officinarum. *BMC genomics, 17*(1), 446.

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., ... & Wai, C. M. (2018). Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. *Nature genetics, 50*(11), 1565.

Zhang, Q., Hu, W., Zhu, F., Wang, L., Yu, Q., Ming, R., & Zhang, J. (2016). Structure, phylogeny, allelic haplotypes and expression of sucrose transporter gene families in Saccharum. *BMC genomics, 17*(1), 88.